

Supplementary Material

POSSMs: a parsimonious speciation model for metals in soils

Stephen Lofts^{A,*}

^AAUK Centre for Ecology and Hydrology, Lancaster Environment Centre, Lancaster, LA1 4AP, UK.

*Correspondence to: Email: stlo@ceh.ac.uk

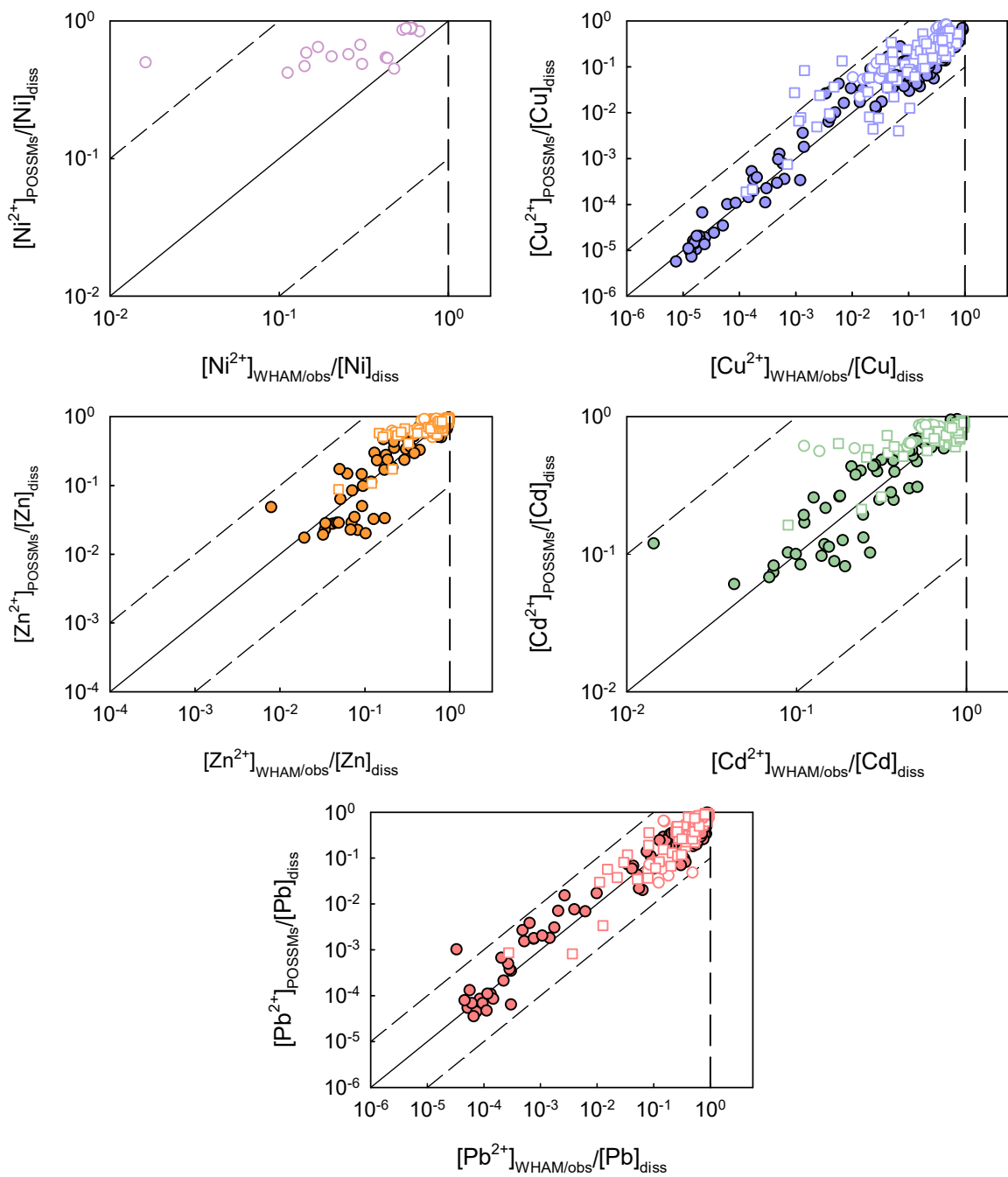


Figure S1. Free ion concentrations of Ni, Cu, Zn, Cd and Pb obtained by solution only application of POSSMs, compared to concentrations obtained by WHAM/Model VII speciation (UK1, UK2) or DMT measurement (NL), expressed as a proportion of the observed solution concentration. Filled circles: UK1 dataset; open triangles: UK2 dataset; open squares: NL dataset.

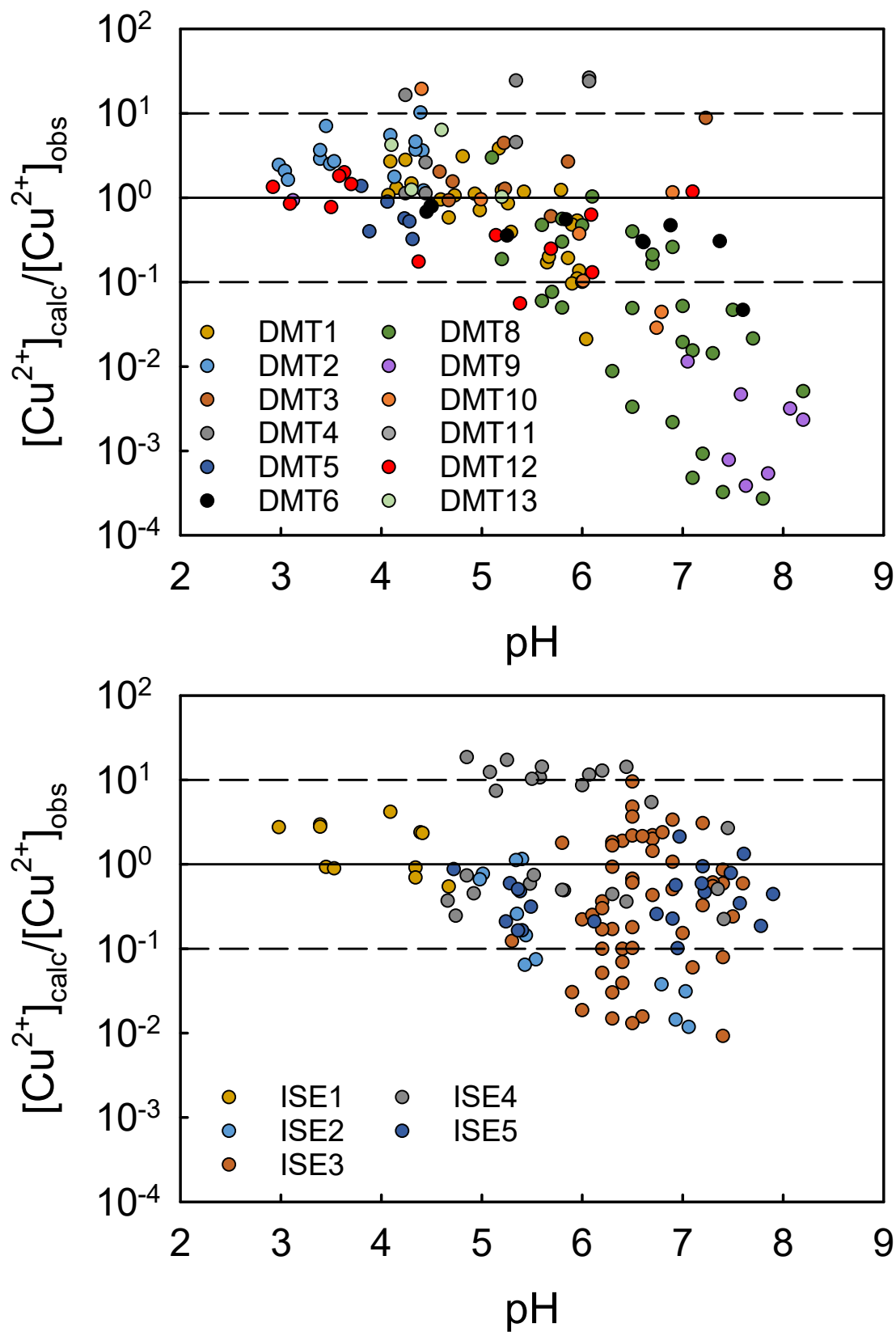


Figure S2. Ratio of predicted to observed $[Cu^{2+}]$ in solution phase only speciation, for studies using DMT (top) and ISE (bottom).

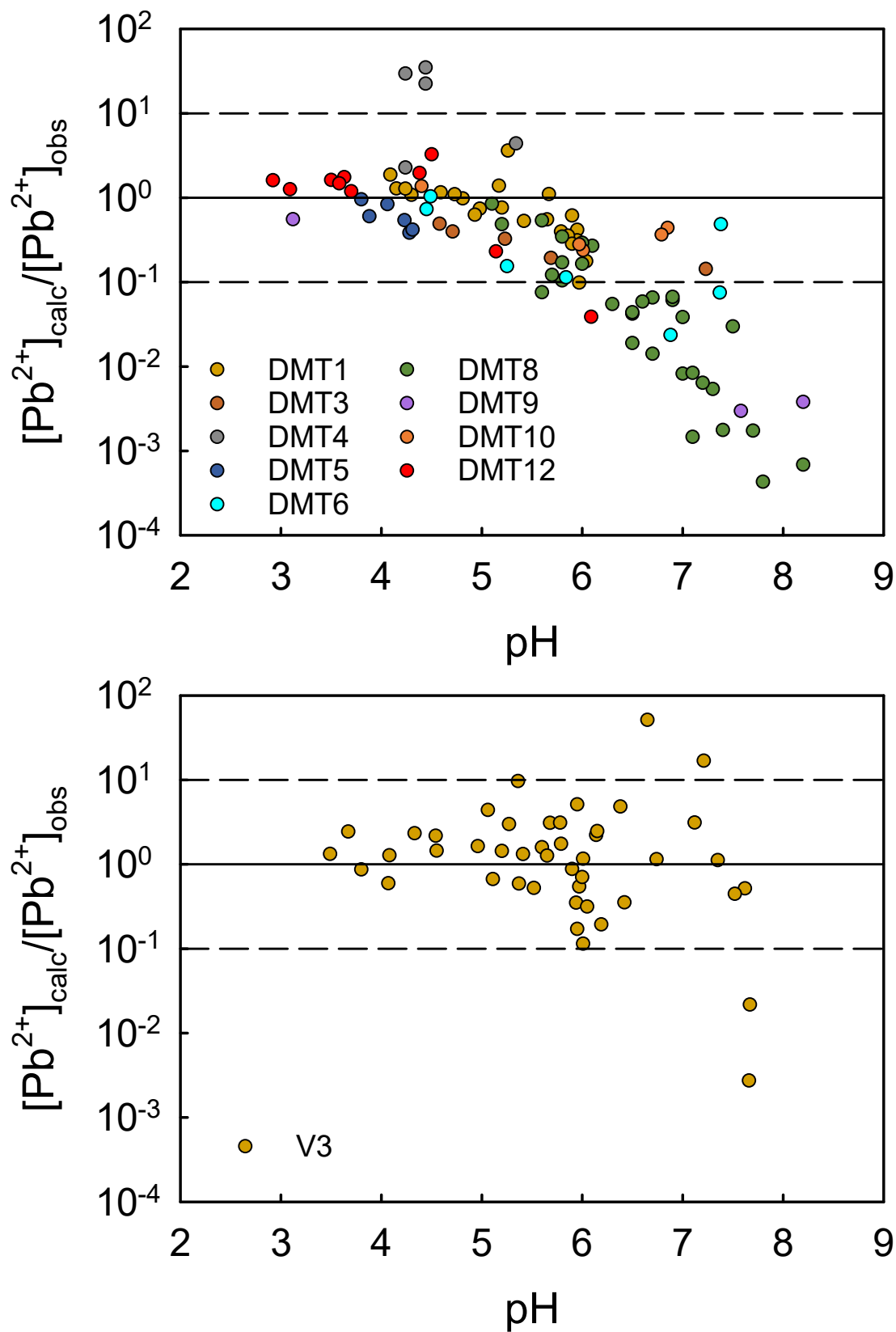


Figure S3. Ratio of predicted to observed $[Pb^{2+}]$ in solution phase only speciation, for studies using DMT (top) and voltammetry (bottom).

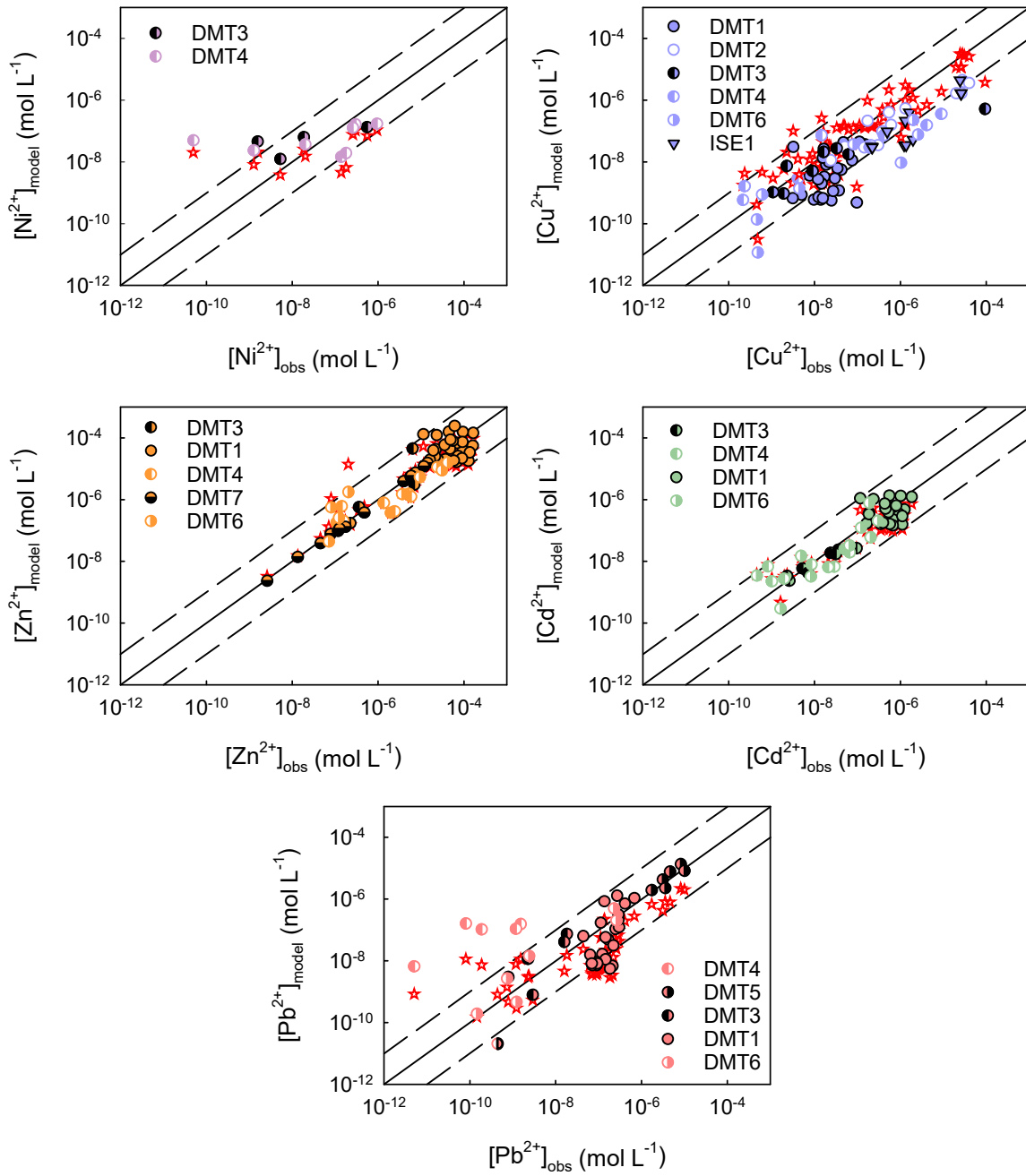


Figure S4. Free ion concentrations in whole soil studies, predicted using the empirical model of Groenenberg et al. (2010) and POSSMs. Predictions of POSSMs are shown as red stars.

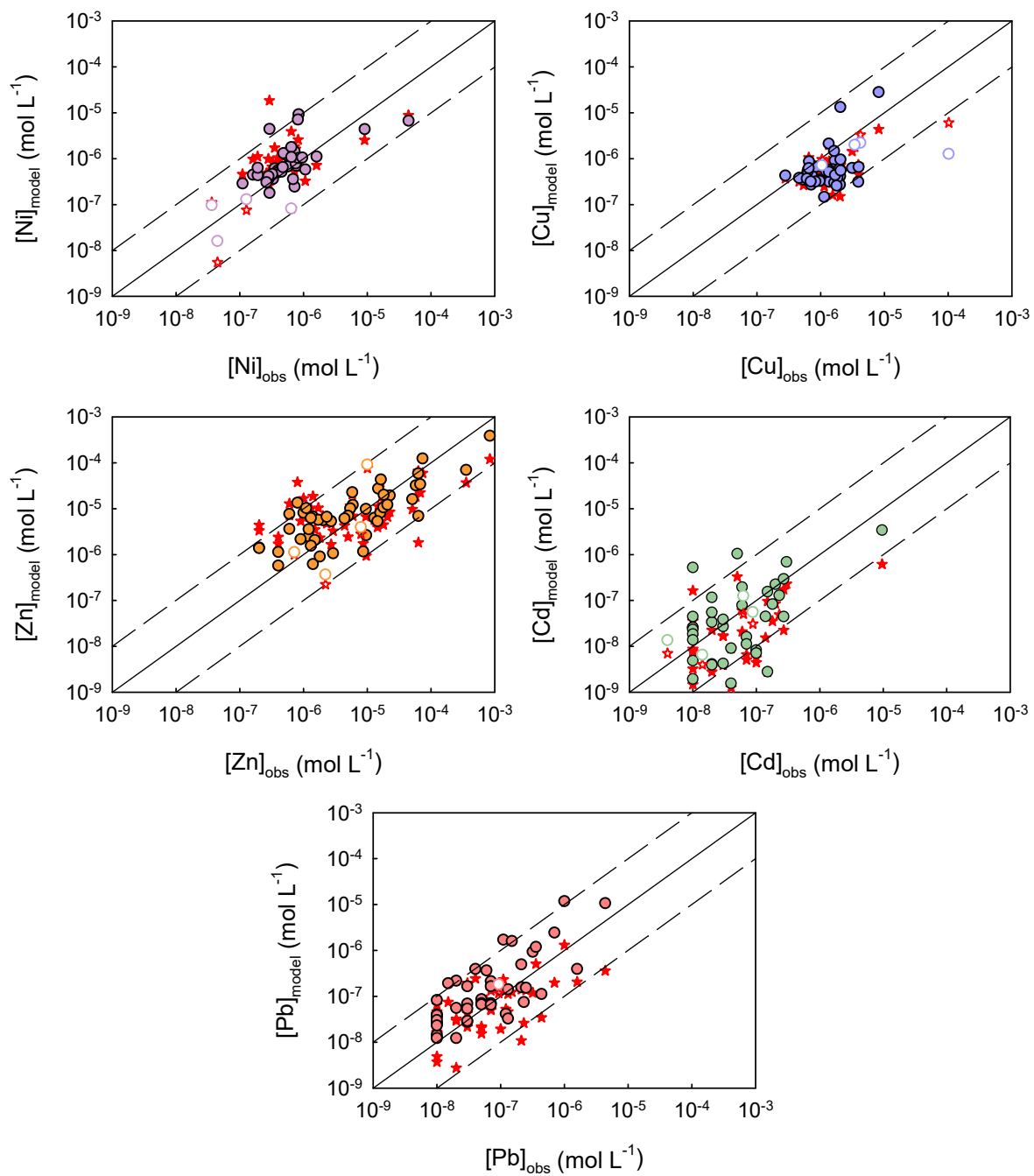


Figure S5. Solution metal concentrations in datasets DMT3 (open symbols) and KD1 (closed symbols) predicted using the empirical solubility model of Groenenberg et al. (2012) and POSSMs. POSSMs predictions are shown as red stars.

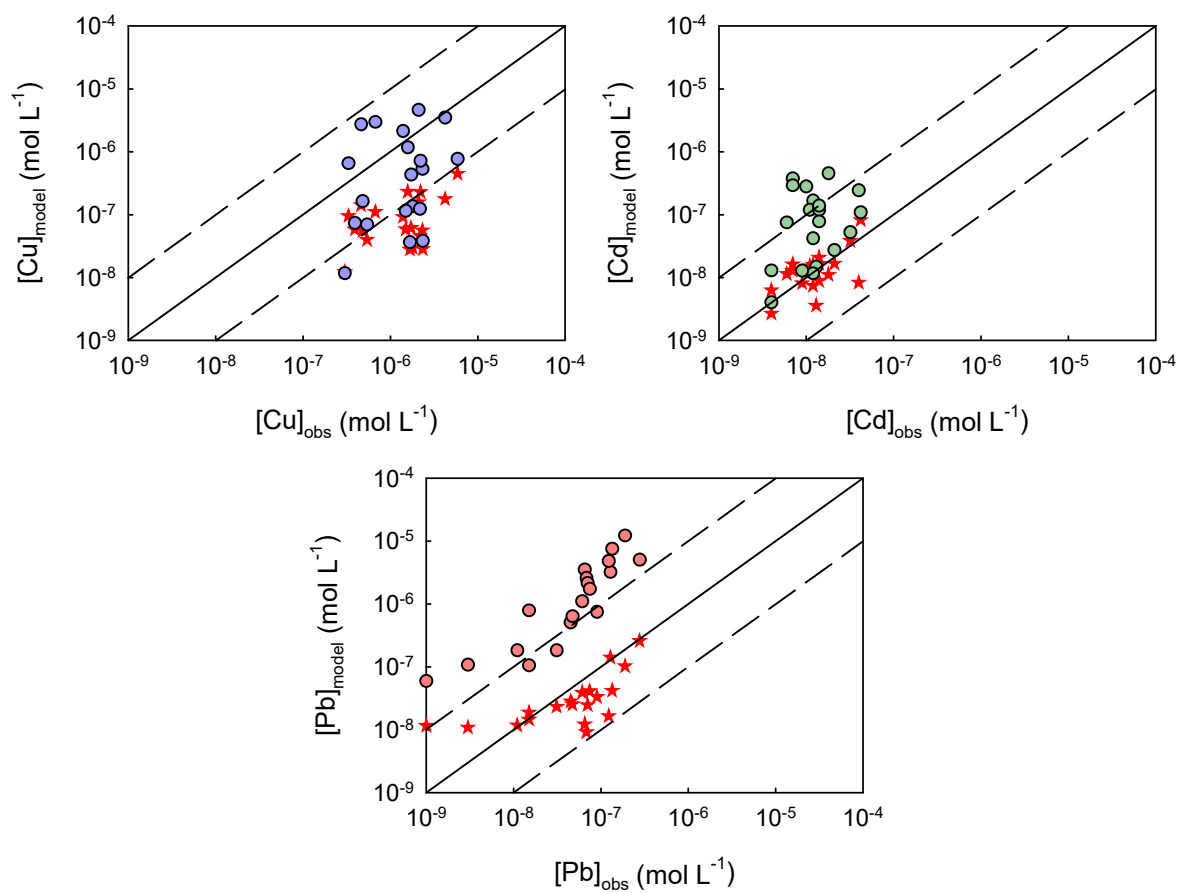


Figure S6. Solution metal concentrations in dataset KD2 predicted using an empirical solubility model (Groenenberg et al. 2012) and POSSMs. POSSMs predictions are shown as red stars.

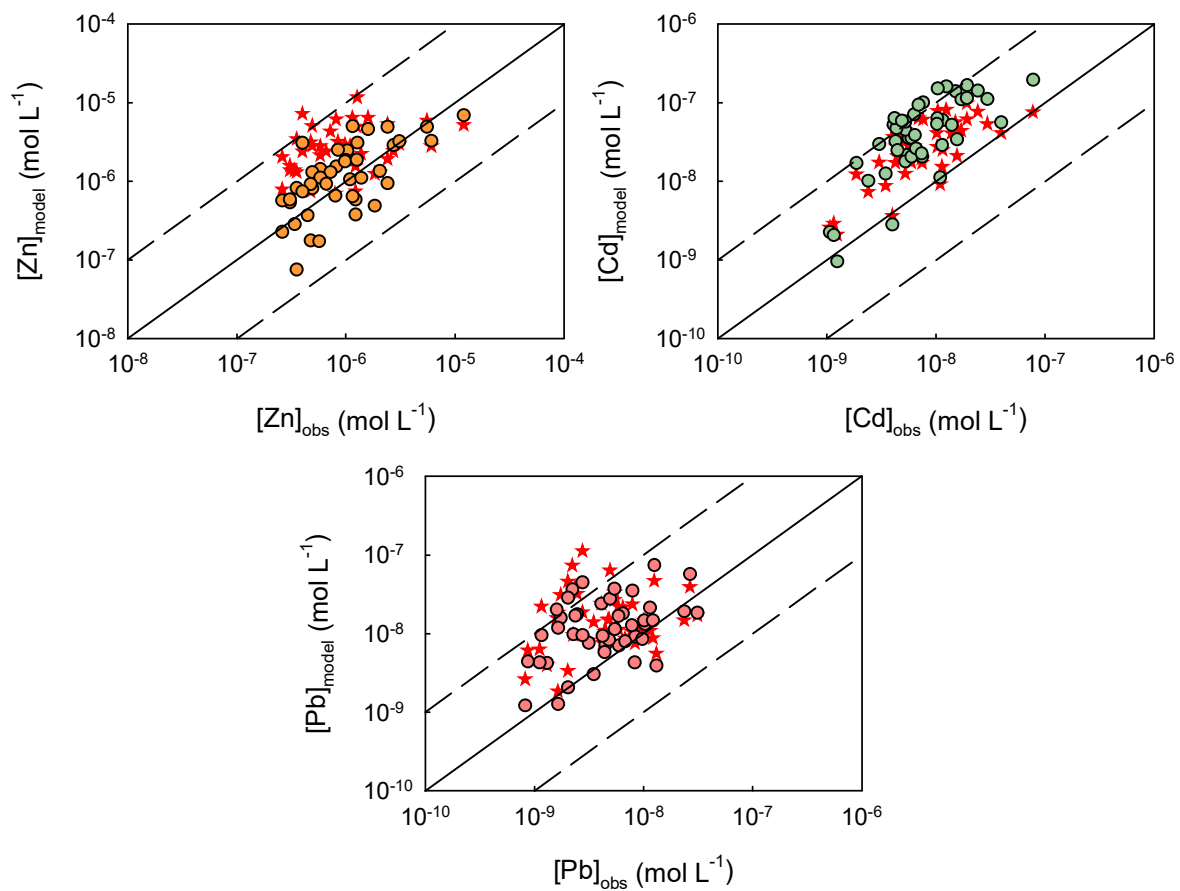


Figure S7. Solution metal concentrations in dataset KD3 predicted using an empirical solubility model (Groenenberg et al. 2012) and POSSMs. POSSMs predictions are shown as red stars.

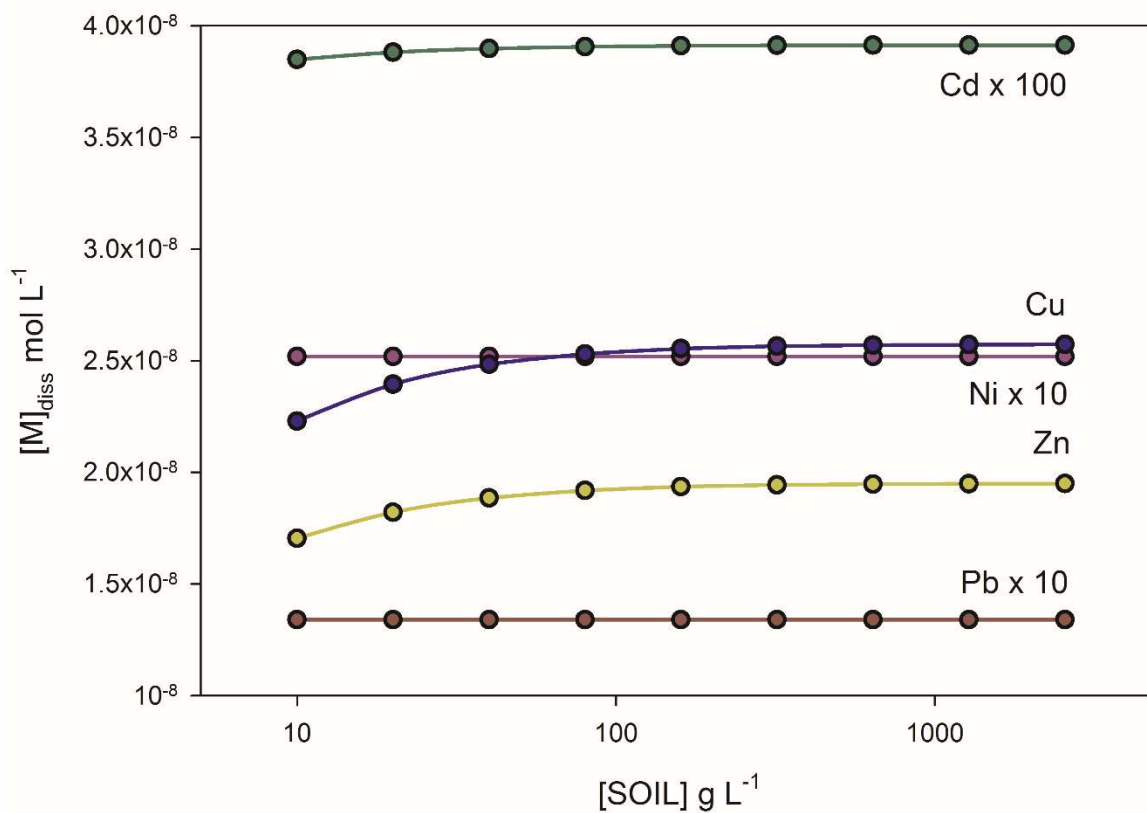


Figure S8. Influence of soil concentration on POSSMs predictions of metal solubility. Computations at pH = 5, SOM = 1%, [DOC] = 100 mg L^{-1} , [SOIL] from 10 to 2560 g L^{-1} . Labile metal concentrations (mol g^{-1}), Ni = 10^{-9} , Cu = 10^{-8} , Zn = 10^{-8} , Cd = 10^{-9} , Pb = 10^{-8} .

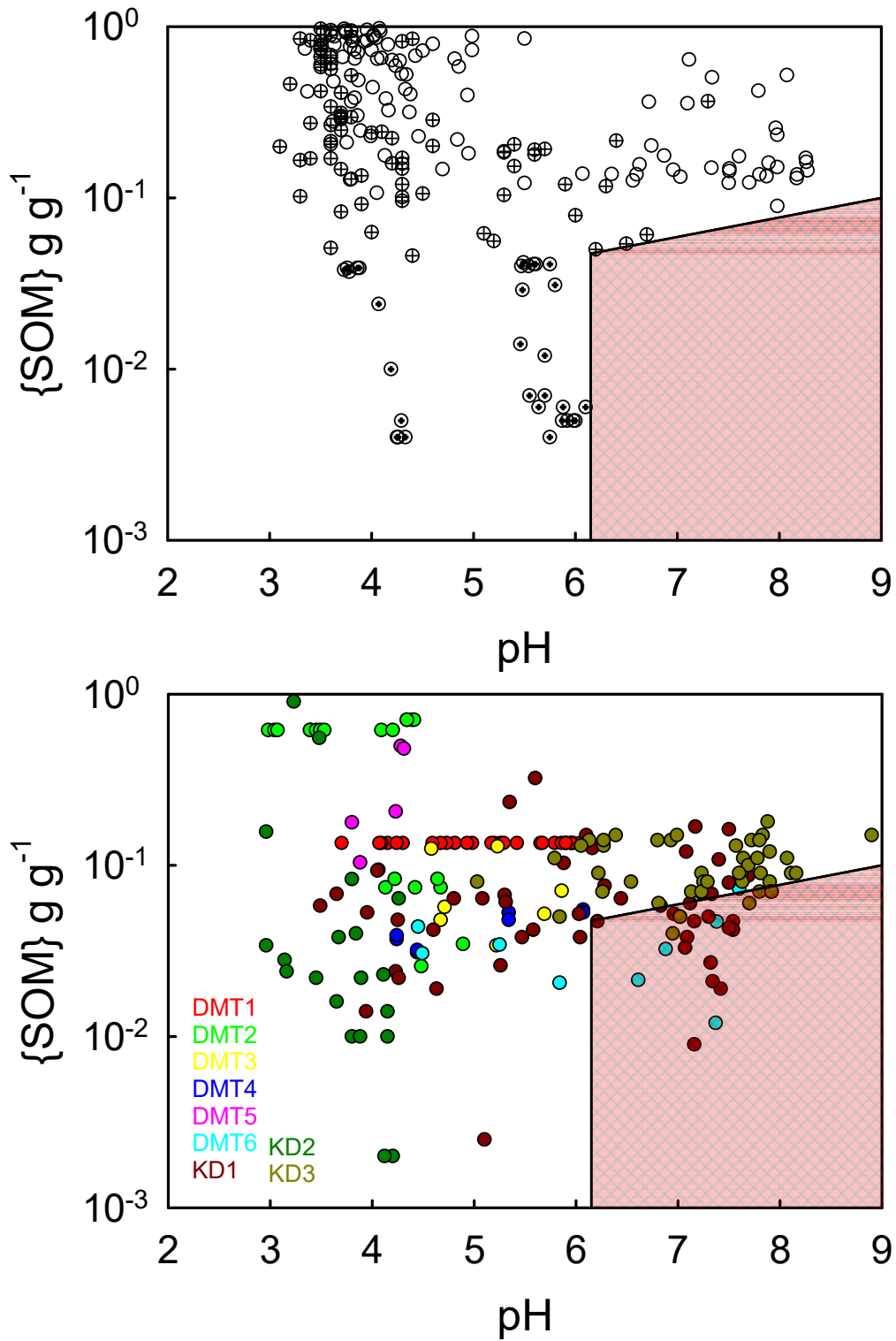


Figure S9. Porewater pH-SOM relationships for the calibration dataset (top; open points = GB1, crossed points = GB2, dotted points = NL) and the literature datasets for which K_d computation is possible. The shaded area indicates the range of pH and {SOM} for which a comparison was made with model goodness of prediction with the remaining points.

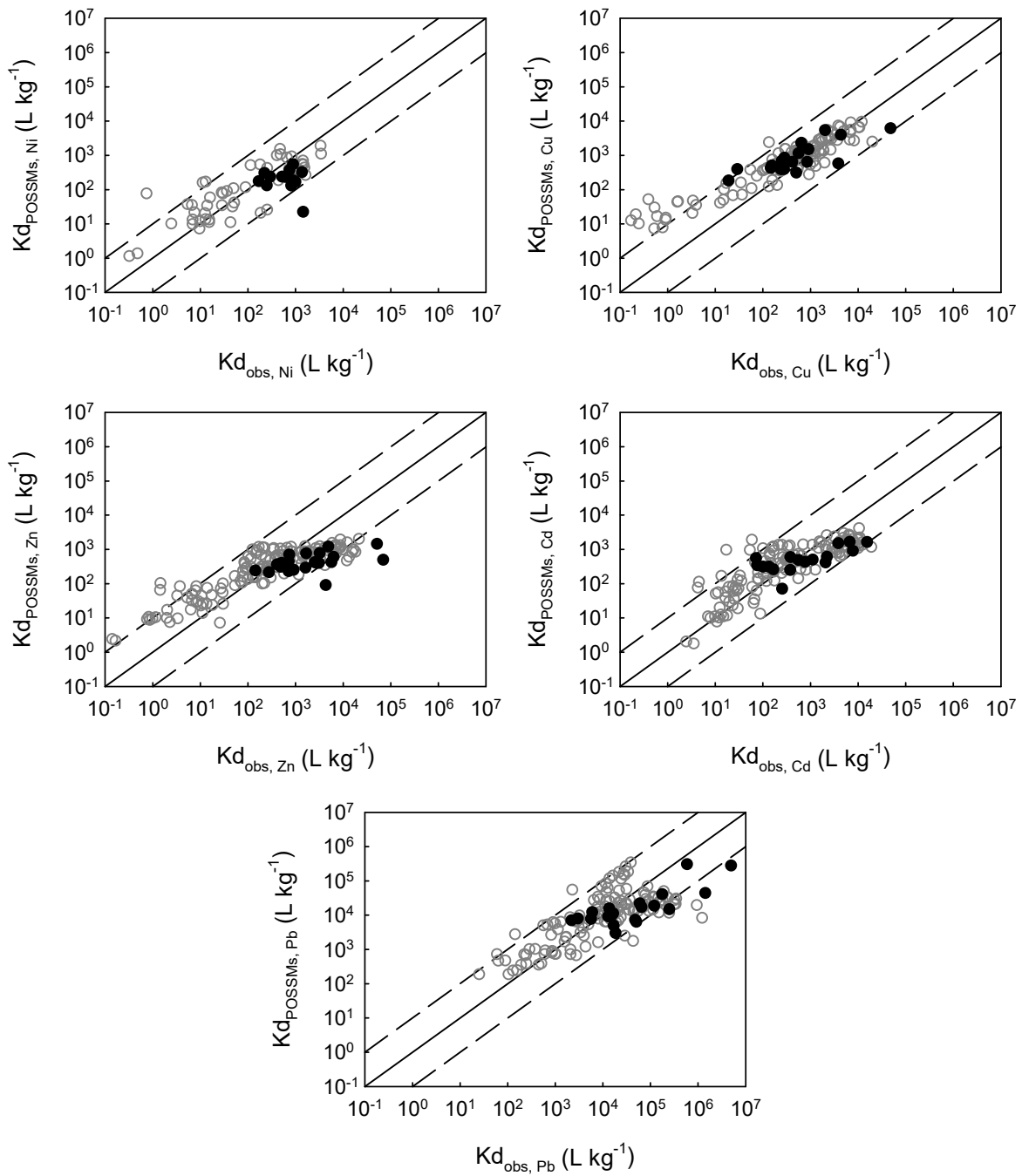


Figure S10. Observed and POSSMs-predicted K_d values, with points outside the calibration dataset pH-SOM range (Figure S9) highlighted.

Summaries of literature datasets

Solid and solution phase studies

Dataset DMT1 (Cancès et al. 2003). Measurements of soluble and free Cu, Zn, Cd and Pb on a soil of Northern France impacted by past metal smelting activities. Labile metal concentrations were estimated with $0.43 \text{ mol L}^{-1} \text{ HNO}_3$ extraction. Soluble and free metal concentrations were measured using the soil column DMT method with $0.002 \text{ mol L}^{-1} \text{ Ca(NO}_3)_2$ as the solution phase, using varying soil:solution ratio and adjusting soil pH to give $n = 26$ for Cu, $n = 27$ for Zn, $n = 22$ for Cd and $n = 23$ for Pb.

Datasets DMT2 and ISE1 (Pampura et al. 2006). Measurements of soluble Cu and Cu^{2+} activities on a set of 21 soils from the Kola Peninsula, Russia, by both the soil column DMT technique and by ion selective electrode (ISE). A dry soil mass of 100g and 193ml of $0.002 \text{ mol L}^{-1} \text{ Ca(NO}_3)_2$ as the solution phase were used in the determinations. Ion selective electrode measurements were made directly in the donor portion of the solution phase. Soil labile metal concentrations were estimated by extraction with $0.43 \text{ mol L}^{-1} \text{ HNO}_3$. Quoted Cu^{2+} activities were converted to true solution concentrations using estimated ionic strengths based on ionic compositions of saturation extracts of the soils plus $0.002 \text{ mol L}^{-1} \text{ Ca(NO}_3)_2$. Reliable DMT and ISE measurements were made on 15 and 12 soils, respectively. For modelling of free Cu^{2+} from the whole soil and the solution phase, $n = 15$ for DMT and $n = 11$ for ISE, and for modelling of solution Cu from the soil, $n = 19$.

Dataset DMT3 (Koopmans et al. 2008). Measurements of soluble Ni, Cu, Zn, Cd and Pb, and Ni^{2+} , Cu^{2+} , Zn^{2+} , Cd^{2+} and Pb^{2+} concentrations, on four soils from Bulgaria, the Netherlands, and China. Metal concentrations were measured on soil extracts obtained using $0.002 \text{ mol L}^{-1} \text{ Ca(NO}_3)_2$ at a soil:solution ratio of 100. Free ion concentrations were measured using DMT. Extended (Groenenberg et al. 2010) with further measurements on four Netherlands soils, using the same experimental approach, and included $0.43 \text{ mol L}^{-1} \text{ HNO}_3$ -extractable metal concentrations for all eight soils. For some of the dataset, results presented in each paper (soluble metal and free ion concentration) differ, where this was encountered the geometric mean was taken. For Ni and Zn, $n = 4$, for Cu and Cd, $n = 8$, and for Pb, $n = 5$.

Dataset DMT4 (Fest et al. 2008). Measurements of labile ($0.43 \text{ mol L}^{-1} \text{ HNO}_3$ -extractable), dissolved and free Ni, Cu, Zn, Cd and Pb from eight samples of Wildekamp, the Netherlands. Free metals were measured using soil column DGT at a soil concentration of 500 g L^{-1} . For Ni, Cu, Zn and Cd, $n = 8$, for Pb $n = 5$.

Dataset DMT5 (Schneider et al. 2016a; Schneider et al. 2016b). Measurements of dissolved Cu, Zn, Cd and Pb, and Cu^{2+} , Zn^{2+} , Cd^{2+} and Pb^{2+} concentrations in each of two subsamples (topsoil and subsoil) of soils sampled at three locations in France, using the soil column DMT

technique with $0.002 \text{ mol L}^{-1} \text{ Ca(NO}_3)_2$ as the background electrolyte and a soil concentration of 66.7 g L^{-1} . Additionally, sequential extraction was used to probe the solid phase speciation of Pb. Whole soil speciation of Pb was done by assuming the fractions F1-F5 (Exchangeable, acid soluble, bound to Mn oxides, easily oxidisable, bound to amorphous Fe oxides) to comprise the labile fraction. For the other metals, only solution speciation calculations were possible. For all metals $n = 6$.

Dataset DMT6 (Ren et al. 2015a; Ren et al., 2017). Measurements of Cu, Zn, Cd and Pb in nine soils of France and China. Concentrations of Cu^{2+} and Cd^{2+} obtained using the soil column–DMT technique with solution electrolyte $0.002 \text{ mol L}^{-1} \text{ Ca(NO}_3)_2$ and a soil concentration of 100 g L^{-1} . Labile Cu and Cd were measured using both 0.43 mol L^{-1} nitric acid extraction and isotopic dilution. The nitric acid extractable concentrations were used in computations as isotopic dilution did not always yield detectable values. Free metal ion concentrations were obtained by averaging of the data presented in this work and that of Ren et al. (2015b). Concentrations of DOM were taken from Ren et al. (2015b). For Cu, $n = 8$, for Zn and Cd, $n = 9$, and for Pb, $n = 8$.

Dataset KD1 (de Groot et al. 1998). Measurements of labile and porewater metal in 49 Dutch soils of varying chemical composition. Porewaters were extracted by addition of $0.002 \text{ mol L}^{-1} \text{ Ca(NO}_3)_2$ to obtain a pF of 2, followed by centrifugation and filtration ($0.45\mu\text{m}$). Labile metals were estimated by extraction with hot concentrated nitric acid. Accounting for concentrations below detection limits, the number of usable lines for Ni, Cu, Zn, Cd and Pb were 34, 39, 49, 36 and 47 respectively.

Dataset KD2 (Goody et al. 1995). Measurements of labile and porewater metal in acidic sandy soils from two profiles from Southern England, comprising 20 samples in all. Soil organic matter content was estimated by elemental analysis of carbon, assuming all C to be organic, and assuming organic matter to be 50%C. Labile metals were estimated using $0.43 \text{ mol L}^{-1} \text{ HNO}_3$ extraction. Soil water samples were obtained by centrifugation of field–moist soils and filtration of the extracted water at $0.45\mu\text{m}$. One sample had a Pb concentration below the detection limit, making the sample number 19 for this element.

Dataset KD3 (Izquierdo et al. 2013). Measurements of labile and porewater metal in alluvial topsoils (0–15cm depth) and subsoils (35–50cm depth) of the catchment of the River Trent (England). In all 48 soil samples were used, 27 topsoils and 19 subsoils. Soil organic matter was estimated by loss on ignition. Labile metals were estimated by isotopic dilution analysis. Soil waters were obtained by liquid displacement using trifluoroethane and centrifugation. Determination of labile Zn was not possible in three soils, so $n = 45$ for Zn, $n = 48$ for Cd and Pb. For the application of the model of Groenenberg et al. (2012), oxalate–extractable Fe in

the soils was obtained from Izquierdo et al. (2012). An estimate of oxalate–extractable Al was made by scaling measurements of Al made by citrate/dithionate extraction to the soil–specific ratio of oxalate–extractable Fe to citrate/dithionate–extracted Fe.

Dataset F1 (Rennert et al. 2017). Measurements of labile metals, using $0.43 \text{ mol L}^{-1} \text{ HNO}_3$ extraction, at two locations in the floodplain of the River Elbe (Germany). At one location a single soil horizon was sampled, at the other a profile of three horizons was measured. They measured pH, DOC and dissolved metals in soil porewater samples (between 15 and 31 samples per soil, sampled fortnightly) obtained using suction cups. Ranges of pH, DOC and metal concentrations in all porewaters were provided. For modelling purposes, central estimates were made by taking the mean (for pH) and geometric mean (for DOC and metals) of the extreme values. For all metals $n = 6$.

Dataset DMT7 (Duffner et al. 2014). Measurements of Zn^{2+} concentrations on 15 samples comprising a mixture of Zn–deficient and Zn–contaminated soils. A soil concentration of 500 g/dm^3 was used and free Zn^{2+} was determined by DMT using a $0.002 \text{ mol L}^{-1} \text{ Ca}(\text{NO}_3)_2$ electrolyte. Soil labile Zn concentration was estimated by extraction with $0.43 \text{ mol L}^{-1} \text{ HNO}_3$. Calculations were done with both estimates of labile soil Zn. Solution phase Zn was not measured, so only predicted free ion concentrations can be compared with observations.

Solution phase only

Dataset DMT8 (Kim and Owens 2009). Measurements of Cu^{2+} , Zn^{2+} , Cd^{2+} and Pb^{2+} concentrations in solutions extracted from smelter–contaminated soils from Australia, using DMT. Solutions were obtained by equilibrating soils with deionised water for 24h at 70% of soil water holding capacity. For all metals $n = 31$.

Dataset DM9 (Nolan et al. 2003). Measurements of Cu^{2+} , Zn^{2+} , Cd^{2+} and Pb^{2+} concentrations in solutions extracted from a set of 27 Australian soils using DMT. Solutions were obtained by rewetting dried soils to their water holding capacity, incubating for 16 hours and separating the solution phase by centrifugation and filtering. Free metal ion concentrations in the acceptor solution of the DMT were converted to activities using the Davies equation, so the quoted activities were converted back to concentrations using the same approach, and using quoted ionic strengths estimated from the conductivity of the soil solutions. Free ions could not be quantified in all solutions because of detection limits or limited sample volume, so $n = 8, 26, 16$ and 3 for Cu, Zn, Cd and Pb respectively.

Dataset DMT10 (Ren et al. 2015b). Measurements of Ni^{2+} , Cu^{2+} , Zn^{2+} , Cd^{2+} and Pb^{2+} concentrations in solutions obtained from nine soils sampled in France and China, using the soil column–DMT technique with solution electrolyte $0.002 \text{ mol L}^{-1} \text{ Ca}(\text{NO}_3)_2$ and a soil

concentration of 10 g L^{-1} . An additional set of Ni^{2+} measurements were obtained from seven further soils using a soil concentration of 100 g L^{-1} . Not all solutions were analysed for all metals, so $n = 14$ for Ni, $n = 7$ for Cu, $n = 8$ for Zn, $n = 8$ for Cd and $n = 5$ for Pb.

Dataset name DMT11 (Salam and Helmke 1998). Measurements of Cu^{2+} , Zn^{2+} and Cd^{2+} in solutions of two soils of the U.S.A., using Donnan analysis. Solutions were obtained by shaking soils with deionised water (1:1 w/v ratio), centrifugation and filtration. Quoted free ion activities were converted to concentrations using activity coefficients computed with the extended Debye–Hückel equation. Ionic strength for the activity correction was estimated by speciation of the quoted solution ionic compositions, using WHAM/Model VII and balancing the charge by adjusting the solution Cl concentration.

Dataset name DMT12 (Weng et al. 2001). Measurements of Cu^{2+} , Zn^{2+} , Cd^{2+} and Pb^{2+} concentrations using 15 soils from the Netherlands Belgium and the soil column DMT technique with $0.002 \text{ mol L}^{-1} \text{ Ca}(\text{NO}_3)_2$ as the background electrolyte. Free metal could not be quantified in some cases due to limits of detection, so $n = 14$, 15, 10 and 10 for Cu, Zn, Cd and Pb respectively.

Dataset name DMT13 (Koopmans and Groenenberg 2011). Measurements of dissolved Ni, Cu, Zn and Cd, and Ni^{2+} , Cu^{2+} , Zn^{2+} and Cd^{2+} concentrations, in four soils of the Netherlands. Soils were extracted using $0.002 \text{ mol L}^{-1} \text{ Ca}(\text{NO}_3)_2$ at a concentration of 100 g L^{-1} and the free ion measured by DMT. For all metals $n = 4$.

Datasets DMT14 and AGNES1 (Chito et al. 2012). Measurements of concentrations of Zn^{2+} in CaCl_2 extractions (soil:solution ratio 1:10) of four soils of the Rhine catchment. Measurements of the free ion concentration were made using both DMT and AGNES (Absence of Gradients and Nernstian Equilibrium Stripping).

Dataset ISE2 (Cloutier–Hurteau et al. 2007). Measurements of Cu^{2+} activities, using an ISE, in extracts of 18 bulk and 18 rhizosphere soils obtained by water extraction at a soil:solution ratio of 1:10 (w/v). Soil and speciation quoted are means of three replicates for each soil, so $n = 12$. The measured activities were converted to estimated concentrations using activity coefficients computed using the extended Debye–Hückel equation with solution ionic strength computed from measured electrical conductivity using the linear relationship given by Simón and García (1999).

Dataset ISE3 (Djae et al. 2017). Measurements of Cu^{2+} activities, using an ion selective electrode, on solutions extracted from a set of soils sampled in France, Réunion and New Caledonia ($n = 55$). Soil solutions were obtained by incubation at 70% water holding capacity with a nutrient salt solution for 22 days followed by extraction with the same solution at a

soil:liquid ratio of 1:10 for 2 hours. The measured activities were converted to estimated concentrations using activity coefficients computed with the extended Debye–Hückel equation, using solution ionic strengths computed from the supplied ionic compositions.

Dataset ISE4 (Sanders 1982). Measurements of Cu^{2+} concentrations in eight soil solutions using an ISE, using soils from a UK site subject to a gradient of long term liming. The most acidic soil was also used to produce a gradient of pH by addition of varying amounts of $\text{Ca}(\text{OH})_2$. Soil solutions were obtained by incubation of the soils at 90% of field capacity for four or eight weeks, followed by addition of water to displace 1/3rd of the initially added water.

Dataset ISE5. (Vulkan et al. 2000). Measurements of Cu^{2+} activities, using an ion selective electrode, on solutions extracted using Rhizon samplers from soils sampled from the UK, Chile and China, incubated at 75% of their water holding capacity for four weeks ($n = 22$). The measured activities were converted to estimated concentrations using activity coefficients computed using the extended Debye–Hückel equation with solution ionic strength computed from measured electrical conductivity using the linear relationship given by Simón and García (1999).

Dataset name CE1 (Chaudri et al. 2000). Measurements of Zn^{2+} and Cd^{2+} concentrations in solutions obtained from a set of UK agricultural soils subject to a range of past amendments with metals. Solutions were obtained by extraction from soils incubated initially at 50% water holding capacity then for two weeks at 75% water hold capacity, using Rhizon soil moisture samplers. Free ion concentrations were determined using the cation exchange method detailed in Holm et al. (1995).

Dataset CE2 (Knight et al. 1998). Measurements of Zn^{2+} and Cd^{2+} concentrations in solutions obtained from a long term soil liming experiment and from sites across Europe with a history of contamination. Solutions were obtained by extraction from soils incubated initially at 50% water holding capacity then for two weeks at 75% water holding capacity, using Rhizon soil moisture samplers. Free ion concentrations were determined using the cation exchange method detailed in Holm et al. (1995).

Dataset CE3 (McGrath et al. 1999). Measurements of Zn^{2+} and Cd^{2+} concentrations in solutions obtained from a long term soil liming experiment and from sites across Europe with a history of contamination. Solutions were obtained by extraction from soils incubated initially at 50% water holding capacity then for two weeks at 75% water hold capacity, using Rhizon soil moisture samplers. Free ion concentrations were determined using the cation exchange method detailed in Holm et al. (1995).

Dataset CE4 (Sanders 1983). Measurements of Zn^{2+} concentrations in eight soil solutions using a cation exchange method, using soils from a UK site subject to a gradient of long term liming. The most acidic soil was also used to produce a gradient of pH by addition of varying amounts of $Ca(OH)_2$. Soil solutions were obtained by incubation of the soils at 90% of field capacity for 88 days, followed by addition of water to displace $1/3^{rd}$ of the initially added water.

Dataset name V1 (Stephan et al. 2008). Measurements of Zn speciation in extracts of 66 contaminated soils obtained by 0.01M KNO_3 extraction at a soil:solution ratio of 1:2 (w/v). Labile Zn was measured by anodic stripping voltammetry, assumed to be free and inorganically complexed species, and the free Zn^{2+} activity computed by speciation modelling. The measured free activities were converted to concentrations using activity coefficients computed with the extended Debye–Hückel equation, assuming the solution ionic strength to be 0.01 mol L^{-1} . Labile Zn was undetectable in two samples, so $n = 64$.

Dataset V2 (Sauvé et al. 2000). Measurements of Cd speciation in extracts of 64 soils of Canada and the U.S.A., obtained by $0.01 \text{ mol L}^{-1} KNO_3$ extraction at a soil:solution ratio of 1:2 (w/v). Labile Cd was measured by differential pulse anodic stripping voltammetry, assumed to comprise free and inorganically complexed species, and the Cd^{2+} activity computed by speciation modelling. The measured free activities were converted to concentrations using activity coefficients computed with the extended Debye–Hückel equation, assuming the solution ionic strength to be 0.01 mol L^{-1} .

Dataset name V3 (Sauvé et al. 1997). Measurements of Pb speciation in extracts of soils of Canada and the U.S.A., obtained by 0.01M KNO_3 extraction at a soil:solution ratio of 1:2 (w/v). Labile Pb was measured by anodic stripping voltammetry, assumed to comprise free and inorganically complexed species, and the Pb^{2+} activity computed by speciation modelling. Soil extract DOC concentrations were not quoted, so analysis was confined to those soils ($n = 45$) for which DOC concentrations were available (Sauvé et al. 2000). The measured free activities were converted to concentrations using activity coefficients computed with the extended Debye–Hückel equation, assuming the solution ionic strength to be 0.01 mol L^{-1} .

Computation of soil concentration

The soil concentration, [SOIL], is given by

$$[SOIL](\text{g L}^{-1}) = \frac{\rho_{\text{bulk}}\rho_{\text{particle}}}{\%WHC \cdot (\rho_{\text{particle}} - \rho_{\text{bulk}})} \cdot 10^5$$

where ρ_{particle} and ρ_{bulk} are the particle and bulk densities of the soil, respectively (g cm^{-3}) and %WHC is the soil moisture content as a percentage of its water holding capacity. The particle density is given by

$$\rho_{\text{particle}} = \frac{\rho_{\text{IM}} \cdot (100 - \text{OM}) + \rho_{\text{OM}} \cdot \text{OM}}{100}$$

where ρ_{IM} and ρ_{OM} are densities of inorganic (mineral) and organic matter, respectively (g cm^{-3}). Values of $\rho_{\text{IM}} = 2.6$ and $\rho_{\text{OM}} = 1.5$ have been used in this work. The bulk density (g cm^{-3}) may be given by (Tipping et al. 2003):

$$\rho_{\text{bulk}} = 1.38 - 0.29 \cdot \ln(\text{OM}/2)$$

Using this approach only a single variable, %WHC, needs to be known or estimated in order to compute the soil concentration.

Table S1. Literature datasets used for POSSMs evaluation.

Code and reference	pH	SOM (g/g)	[SOIL] (g L ⁻¹)	[DOM] (g L ⁻¹)	$p\{M\}_{labile,obs}$ (mol g soil ⁻¹)					$p[M]_{total,aq}$ (mol L ⁻¹)				
					Ni	Cu	Zn	Cd	Pb	Ni	Cu	Zn	Cd	Pb
Solid phase and solution														
DMT1 Cancès et al. 2003	3.70– 6.04	0.135	2– 80	0.0044– 0.0358		6.39	4.74	6.71	5.44		7.69– 6.36	4.85– 3.72	6.78– 5.66	7.03– 5.80
DMT2, ISE1 Pampura et al. 2006	2.98– 4.67	0.074 – 0.708	125– 333	0.0152– 0.702		5.93– 4.74					5.83– 4.04			
DMT3 Koopmans et al. 2008, Groenenberg et al. 2010a	4.58– 7.23	0.034 – 0.129	100	0.0188– 0.0722	9.03– 6.91	6.98– 4.83	7.60– 3.91	9.25– 7.16	7.19– 5.09	7.44– 6.19	7.28– 3.89	6.15– 5.00	8.01– 7.06	8.22– 7.03
DMT4 Fest et al. 2008	3.78– 5.27	0.031 – 0.053	500	0.024– 0.332	9.40– 7.48	7.05– 5.54	7.69– 6.77	9.22– 8.59	7.10– 6.98	7.28– 5.86	6.43– 4.54	6.43– 5.12	8.57– 7.06	8.62– 7.18
DMT5 Schneider et al 2016a; Schneider et al 2016b	3.80– 4.31	0.047 – 0.249	80	0.0486– 0.216					5.44– 4.38	6.35– 6.01	7.12– 6.42	5.47– 4.74	6.54– 5.80	5.48– 4.81
DMT6 Ren et al. 2015a; Ren et al. 2017	4.45– 7.60	0.012 – 0.073	100	0.0036– 0.1196		6.98– 4.98	7.19– 4.19	9.64– 6.41	6.59– 5.22		7.26– 5.28	7.30– 4.38	9.21– 6.35	8.92– 6.43

Code and reference	pH	SOM (g/g)	[SOIL] (g L ⁻¹)	[DOM] (g L ⁻¹)	$p\{M\}_{labile,obs}$ (mol g soil ⁻¹)					$p[M]_{total,aq}$ (mol L ⁻¹)				
					Ni	Cu	Zn	Cd	Pb	Ni	Cu	Zn	Cd	Pb
KD1 de Groot et al. 1998	3.49– 7.69	0.009 – 0.323	770– 8385	0.0026– 0.0892	8.00– 5.98	7.52– 5.10	7.40– 3.85	9.30– 6.72	7.70– 5.10	6.96– 4.36	6.65– 5.09	6.70– 2.90	8.00– 5.02	8.00– 5.36
KD2 Goody et al. 1995	2.96– 4.26	0.002 – 0.904	592– 33333	0.0206– 0.532	9.60– 7.14	9.62– 7.05	8.51– 6.02	10.1– 8.27	8.51– 6.36	7.70– 5.92	6.52– 5.23	5.80– 4.66	8.40– 7.38	>9.00 –6.56
KD3 Izquierdo et al. 2013	5.03– 8.90	0.04– 0.18	2151– 4377 ^A	0.028– 0.126			6.42 _c 5.04	8.35– 6.90	7.28– 5.82			6.59– 4.92	8.97– 7.11	9.02– 7.50
Solid phase only														
DMT7 Duffner et al. 2014	4.2– 7.5	0.008 – 0.056	459	0.0034– 0.149			8.51– 5.06							
Solution phase only														
DMT8 Kim and Owens 2009	5.1– 8.2			0.058– 0.814							6.55– 5.37	5.96– 5.38	8.28– 5.61	7.42– 5.67
DMT9 Nolan et al. 2003	3.12– 8.52			0.0362– 2.28							6.53– 3.10	7.82– 1.39	8.47– 4.03	8.15– 5.70
DMT10 Ren et al. 2015 ^b	4.40– 6.90			0.0042– 0.0158						8.92– 6.29	7.94– 6.40	7.70– 5.55	9.24– 7.37	9.25– 8.12
DMT11 Salam and Helmke 1998	6.14– 6.97			0.204– 0.366							6.38– 6.05	5.02– 4.91	8.77– 8.01	

Code and reference	pH	SOM (g/g)	[SOIL] (g L ⁻¹)	[DOM] (g L ⁻¹)	$p\{M\}_{labile,obs}$ (mol g soil ⁻¹)					$p[M]_{total,aq}$ (mol L ⁻¹)				
					Ni	Cu	Zn	Cd	Pb	Ni	Cu	Zn	Cd	Pb
DMT12 Weng et al. 2001	2.92– 7.10			0.0078– 0.064							7.64– 5.99	6.47– 4.37	8.05– 6.24	7.77– 5.29
DMT13 Koopmans and Groenenberg 2011	4.1– 5.2			0.0244– 0.0382						6.02– 5.15	7.40– 6.30	5.42– 4.43	8.00– 7.22	
DMT14, AGNES1 Chito et al. 2012	4.5– 6.2			0.0356– 0.0428								5.95– 5.50		
ISE2 Cloutier– Hurteau et al. 2007	4.98– 7.96			0.0536– 0.1836							8.40– 5.79			
ISE3 Djae et al. 2017	5.3– 7.6			0.0162– 0.6							8.96– 5.43			
ISE4 Sanders 1982	4.85– 7.45			0.096– 0.238							6.19– 5.97			
ISE5 Vulkan et al. 2000	4.72– 7.90			0.0136– 0.1422							6.85– 3.58			
CE1 Chaudri et al. 2000	5.3– 6.5			0.044– 0.074								6.04– 3.52	9.05– 6.89	

Code and reference	pH	SOM (g/g)	[SOIL] (g L ⁻¹)	[DOM] (g L ⁻¹)	$p\{M\}_{labile,obs}$ (mol g soil ⁻¹)					$p[M]_{total,aq}$ (mol L ⁻¹)				
					Ni	Cu	Zn	Cd	Pb	Ni	Cu	Zn	Cd	Pb
CE2 Knight et al. 1998	3.8– 7.5			0.016– 0.308								6.26– 4.71	7.65– 5.52	
CE3 McGrath et al. 1999	4.4– 6.9			20.6– 49.0								6.44– 3.42	8.75– 6.00	
CE4 Sanders 1983	4.75– 6.60			35– 75								7.10– 5.60		
V1 Stephan et al. 2008	4.09– 7.96			15– 743								6.82– 3.47		
V2 Sauvé et al. 2000	3.49– 8.14			1.5– 11.9									9.57– 5.79	
V3 Sauvé et al. 1997	3.49– 7.67			1.5– 11.9										8.50– 6.22

^A assuming a soil porewater saturation of 50%

Table S2. Root mean squared errors (RMSEs) in $\log[M^{2+}]$ and $\log K_d$ for predictions of whole soil speciation.

Figures in brackets are the number of points. Italicised figures refer to datasets with points having variables outside POSSMs calibration ranges removed.

Dataset	Reference	RMSE in $\log[M^{2+}]_{aq}$					RMSE in $\log K_d$				
		Ni	Cu	Zn	Cd	Pb	Ni	Cu	Zn	Cd	Pb
DMT1	Cancès et al. 2003	–	0.702 (26)	0.445 (27)	0.493 (22)	1.04 (23)	–	0.271 (26)	0.495 (27)	0.513 (22)	0.733 (23)
		–	<i>0.635</i> (21)	<i>0.458</i> (22)	<i>0.515</i> (18)	<i>1.08</i> (19)	–	<i>0.196</i> (21)	<i>0.506</i> (22)	<i>0.530</i> (18)	<i>0.742</i> (19)
DMT2	Pampura et al. 2006	–	0.335 (15)	–	–	–	–	0.393 (19)	–	–	–
		–	<i>0.430</i> (5)	–	–	–	–	<i>0.339</i> (5)	–	–	–
DMT3	Koopmans et al. 2008, Groenenberg et al. 2010	0.725 (4)	0.734 (8)	0.479 (4)	0.260 (8)	0.419 (5)	0.848 (4)	0.524 (8)	1.06 (4)	0.366 (7)	0.335 (5)
		–	<i>0.644</i> (5)	<i>0.279</i> (3)	–	–	–	<i>0.327</i> (5)	<i>1.12</i> (3)	–	–
DMT4	Fest et al. 2008	1.28 (8)	0.817 (8)	0.560 (8)	0.491 (8)	1.65 (5)	0.905 (7)	0.446 (8)	0.557 (8)	0.358 (8)	0.782 (5)
DMT5	Schneider et al 2016a; Schneider et al 2016b	–	–	–	–	0.666 (6)	–	–	–	–	0.468 (6)
		–	–	–	–	–	–	–	–	–	–

Table S2 (contd.)

Dataset	Reference	RMSE in log[M ²⁺] _{aq}					RMSE in log K _d				
		Ni	Cu	Zn	Cd	Pb	Ni	Cu	Zn	Cd	Pb
DMT6	Ren et al. 2015a; Ren et al. 2017	–	0.731 (8)	0.784 (9)	0.494 (9)	0.532 (7)	–	0.361 (9)	1.04 (9)	0.544 (9)	1.17 (9)
		–	0.546 (5)	0.341 (7)	0.381 (7)	0.617 (5)	–	0.173 (6)	0.615 (7)	0.417 (7)	1.24 (7)
DMT7	Duffner et al. 2014	–	–	0.503 (15)	–	–	–	–	–	–	–
		–	–	0.531 (10)	–	–	–	–	–	–	–
ISE1	Pampura et al. 2006	–	0.495 (11)	–	–	–	–	0.393 (19)	–	–	–
		–	0.631 (4)	–	–	–	–	0.339 (9)	–	–	–
KD1	de Groot et al. 2008	–	–	–	–	–	0.580 (34)	0.475 (39)	0.762 (49)	0.803 (36)	0.538 (47)
		–	–	–	–	–	0.575 (31)	0.484 (36)	0.742 (46)	0.798 (33)	0.531 (44)
KD2	Goody et al. 1995	–	–	–	–	–	0.343 (20)	1.42 (20)	0.826 (20)	0.270 (20)	0.484 (19)
		–	–	–	–	–	0.318 (16)	0.696 (4)	0.757 (16)	0.231 (16)	0.514 (15)
KD3	Izquierdo et al. 2013	–	–	–	–	–	–	–	0.587 (43)	0.606 (44)	0.719 (44)
		–	–	–	–	–	–	–	0.587 (42)	0.612 (43)	0.721 (43)

Table S2 (contd.)

Dataset	Reference	RMSE in $\log[M^{2+}]_{\text{aq}}$					RMSE in $\log K_d$				
		Ni	Cu	Zn	Cd	Pb	Ni	Cu	Zn	Cd	Pb
F1	Rennert et al. 2017	–	–	–	–	–	0.484 (4)	0.641 (4)	0.389 (4)	0.363 (4)	0.917 (4)
		–	–	–	–	–	0.543 (3)	–	–	–	–

Table S3. Root mean squared errors (RMSEs) in log[M²⁺] for predictions of solution phase speciation.

Italicised figures refer to datasets with points having variables outside POSSMs calibration ranges removed.

Dataset	Reference	RMSE in log[M ²⁺]				
		Ni	Cu	Zn	Cd	Pb
DMT1	Cancès et al. 2003	–	0.574 (26)	0.048 (27)	0.060 (22)	0.381 (23)
		–	<i>0.562 (21)</i>	<i>0.046 (22)</i>	<i>0.062 (18)</i>	<i>0.415 (19)</i>
DMT2	Pampura et al. 2006	–	0.554 (15)	–	–	–
		–	<i>0.652 (5)</i>	–	–	–
DMT3	Koopmans et al. 2008; Groenenberg et al. 2010	0.533 (4)	0.459 (8)	0.449 (4)	0.260 (8)	0.585 (5)
		–	<i>0.350 (5)</i>	<i>0.510 (3)</i>	–	–
DMT4	Fest et al. 2008	1.09 (8)	0.996 (8)	0.165 (8)	0.180 (8)	1.18 (5)
		–	–	–	–	–
DMT5	Schneider et al. 2016a; Schneider et al. 2016b	0.209 (6)	0.306 (6)	0.102 (6)	0.160 (6)	0.272 (6)
		–	–	–	–	–
DMT6	Ren et al. 2015a; Ren et al. 2017	–	0.575 (8)	0.468 (9)	0.188 (9)	0.897 (7)
		–	<i>0.612 (5)</i>	<i>0.456 (7)</i>	<i>0.206 (7)</i>	<i>0.921 (5)</i>
DMT8	Kim and Owens 2009	–	1.74 (31)	0.406 (31)	0.287 (31)	1.71 (31)
		–	<i>1.73 (25)</i>	<i>0.400 (25)</i>	<i>0.294 (25)</i>	<i>1.73 (25)</i>
DMT9	Nolan et al. 2003	–	2.61 (8)	1.11 (26)	0.847 (16)	2.03 (3)
		–	<i>2.20 (3)</i>	<i>1.20 (10)</i>	<i>0.559 (5)</i>	<i>1.32 (1)</i>
DMT10	Ren et al. 2015b	0.228 (14)	1.00 (7)	0.240 (8)	0.124 (8)	0.454 (5)
		<i>0.125 (1)</i>	<i>1.54 (1)</i>	<i>0.505 (1)</i>	<i>no data</i>	<i>no data</i>

Table S3 (contd.)

Dataset	Reference	RMSE in log[M ²⁺]				
		Ni	Cu	Zn	Cd	Pb
DMT11	Salam and Helmke 1998	–	1.64 (2)	0.731 (2)	0.271 (2)	–
		–	–	–	–	–
DMT12	Weng et al. 2001	–	0.519 (14)	0.105 (15)	0.103 (10)	0.482 (10)
		–	<i>0.537 (13)</i>	<i>0.110 (13)</i>	<i>0.109 (9)</i>	<i>0.524 (8)</i>
DMT13	Koopmans and Groenenberg 2011	0.173 (4)	0.511 (4)	0.252 (4)	0.098 (4)	–
DMT14	Chito et al. 2012	–	–	0.087 (4)	–	–
		–	–	–	–	–
ISE1	Pampura et al. 2006	–	0.353 (11)	–	–	–
		–	<i>0.232 (4)</i>	–	–	–
ISE2	Cloutier–Hurteau et al. 2007	–	1.13 (12)	–	–	–
		–	–	–	–	–
ISE3	Djae et al. 2017	–	0.892 (52)	–	–	–
		–	–	–	–	–
ISE4	Sanders 1982	–	0.767 (25)	–	–	–
		–	–	–	–	–
ISE5	Vulkan et al. 2000	–	0.506 (22)	–	–	–
		–	–	–	–	–

Table S3 (contd.)

Dataset		Reference	RMSE in log[M ²⁺]				
			Ni	Cu	Zn	Cd	Pb
CE1	Chaudri et al. 2000	–	–	0.225 (14)	0.125 (14)	–	
		–	–	–	–	–	
CE2	Knight et al. 1998	–	–	0.309 (9)	0.455 (9)	–	
		–	–	–	–	–	
CE3	McGrath et al. 1999	–	–	0.175 (10)	0.112 (10)	–	
		–	–	–	–	–	
CE4	Sanders 1983	–	–	0.236 (8)	–	–	
		–	–	–	–	–	
V1	Stephan et al. 2008	–	–	0.755 (64)	–	–	
		–	–	0.750 (60)	–	–	
V2	Sauvé et al. 2000	–	–	–	0.735 (64)	–	
		–	–	–	0.689 (41)	–	
V3	Sauvé et al. 1997	–	–	–	–	0.686 (45)	
		–	–	–	–	0.776 (29)	
AGNES1	Chito et al. 2012	–	–	0.114 (4)	–	–	
		–	–	–	–	–	

Table S4. Summary of root mean squared errors in POSSMs prediction of free ion from whole soil, and of partition coefficients. Italicised figures refer to datasets with points having variables outside POSSMs calibration ranges removed.

Metal	log [M ²⁺] – DMT			log [M ²⁺] – ISE			log K _d		
	<i>d</i>	<i>n</i>	RMSE	<i>d</i>	<i>n</i>	RMSE	<i>d</i>	<i>n</i>	RMSE
Ni	2	12	1.13	–	–	–	4	65	0.58
		<i>12</i>	<i>1.13</i>	–	–	–		<i>54</i>	<i>0.57</i>
Cu	5	65	0.66	1	11	0.50	7	85	0.83
		<i>44</i>	<i>0.65</i>		<i>4</i>	<i>0.63</i>		<i>60</i>	<i>0.46</i>
Zn	5	63	0.54	–	–	–	7	134	0.77
		<i>50</i>	<i>0.47</i>	–	–	–		<i>123</i>	<i>0.68</i>
Cd	4	47	0.46	–	–	–	7	125	0.64
		<i>41</i>	<i>0.45</i>	–	–	–		<i>108</i>	<i>0.64</i>
Pb	5	46	0.98	–	–	–	8	136	0.72
		<i>40</i>	<i>1.02</i>	–	–	–		<i>121</i>	<i>0.73</i>

Table S5. Summary of root mean squared errors in POSSMs prediction of free metal ions in soil solutions, by analysis type.

Italicised figures refer to datasets with points having variables outside POSSMs calibration ranges removed.

Metal	DMT			ISE			CE			V			AGNES		
	<i>d</i>	<i>n</i>	RMSE	<i>d</i>	<i>n</i>	RMSE	<i>d</i>	<i>n</i>	RMSE	<i>d</i>	<i>n</i>	RMSE	<i>d</i>	<i>n</i>	RMSE
Ni	5	36	0.57	–	–	–	–	–	–	–	–	–	–	–	–
		23	0.69	–	–	–	–	–	–	–	–	–	–	–	–
Cu	13	137	1.18	5	122	0.80	–	–	–	–	–	–	–	–	–
		74	1.28		115	0.82	–	–	–	–	–	–	–	–	–
Zn	11	144	0.54	–	–	4	31	0.24	1	64	0.76	1	4	0.11	
		83	0.52	–	–		–	–		60	0.75		–	–	
Cd	11	124	0.36	–	–	3	23	0.30	1	64	0.74	–	–		
		82	0.26	–	–		–	–		41	0.69	–	–		
Pb	10	95	1.14	–	–	–	–	–	1	45	0.69	–	–		
		60	1.27	–	–	–	–	29		0.78	–	–			

References

- Cancès B, Ponthieu M, Castrec-Rouelle M, Aubry E, Benedetti MF (2003). Metal ions speciation in a soil and its solution: experimental data and model results. *Geoderma* **113**, 341–355. doi:10.1016/S0016-7061(02)00369-5
- Chaudri AM, Allain CMG, Barbosa-Jefferson VL, Nicholson FA, Chambers BJ, McGrath SP (2000). A study of the impacts of Zn and Cu on two rhizobial species in soils of a long-term field experiment. *Plant and Soil* **221**, 167–179. doi:10.1023/a:1004735705492
- Cloutier-Hurteau B, Sauvé S, Courchesne F (2007). Comparing WHAM 6 and MINEQL+ 4.5 for the Chemical Speciation of Cu²⁺ in the Rhizosphere of Forest Soils. *Environmental Science and Technology* **41**, 8104–8110. doi:10.1021/es0708464
- de Groot AC, Peijnenburg WJGM, van den Hoop MAGT, Ritsema R, van Veen RPM (1998). Heavy metals in Dutch field soils: an experimental and theoretical study on equilibrium partitioning. Report no. 607220001. RIVM, Bilthoven, The Netherlands.
- Djae T, Bravin MN, Garnier C, Doelsch E (2017). Parameterizing the binding properties of dissolved organic matter with default values skews the prediction of copper solution speciation and ecotoxicity in soil. *Environmental Toxicology and Chemistry* **36**, 898–905. doi:10.1002/etc.3622
- Duffner A, Weng L, Hoffland E, van der Zee SEATM (2014). Multi-surface Modeling To Predict Free Zinc Ion Concentrations in Low-Zinc Soils. *Environmental Science and Technology* **48**, 5700–5708. doi:10.1021/es500257e
- Fest EPMJ, Temminghoff, EJM, Comans RNJ, van Riemsdijk WH (2008). Partitioning of organic matter and heavy metals in a sandy soil: Effects of extracting solution, solid to liquid ratio and pH. *Geoderma* **146**, 66–74. doi:10.1016/j.geoderma.2008.05.00
- Groenenberg JE, Koopmans GF, Comans RNJ (2010). Uncertainty Analysis of the Nonideal Competitive Adsorption–Donnan Model: Effects of Dissolved Organic Matter Variability on Predicted Metal Speciation in Soil Solution. *Environmental Science and Technology* **44**, 1340–1346. doi:10.1021/es902615w
- Groenenberg JE, Dijkstra JJ, Bonten LTC, de Vries W, Comans RNJ (2012). Evaluation of the performance and limitations of empirical partition–relations and process based multisurface models to predict trace element solubility in soils. *Environmental Pollution* **166**, 98–107. doi:10.1016/j.envpol.2012.03.011

- Holm PE, Christensen TH, Tjell JC, McGrath SP (1995). Speciation of cadmium and zinc with application to soil solutions. *Journal of Environmental Quality* **24**, 183–190.
doi:10.2134/jeq1995.00472425002400010025x
- Izquierdo M, Tye AM, Chenery SR (2012). Sources, lability and solubility of Pb in alluvial soils of the River Trent catchment, UK. *Science of the Total Environment* **433**, 110–122.
doi:10.1016/j.scitotenv.2012.06.039
- Izquierdo M, Tye AM, Chenery SR (2013). Lability, solubility and speciation of Cd, Pb and Zn in alluvial soils of the River Trent catchment UK. *Environmental Science – Processes & Impacts* **15**, 1844–1858. doi:10.1039/c3em00370a
- Kim KR, Owens G (2009). Chemodynamics of heavy metals in long-term contaminated soils: Metal speciation in soil solution. *Journal of Environmental Science* **21**, 1532–1540.
doi:10.1016/S1001-0742(08)62451-1
- Koopmans GF, Schenkeveld WDC, Song J, Luo Y, Japenga J, Temminghoff EJM (2008). Influence of EDDS on Metal Speciation in Soil Extracts: Measurement and Mechanistic Multicomponent Modeling. *Environmental Science and Technology* **42**, 1123–1130.
doi:10.1021/es071694f
- Koopmans GF, Groenenberg JE (2011). Effects of soil oven-drying on concentrations and speciation of trace metals and dissolved organic matter in soil solution extracts of sandy soils. *Geoderma* **161**, 147–158. doi:10.1016/j.geoderma.2010.12.012
- Knight BP, Chaudri AM, McGrath SP, Giller KE (1998). Determination of chemical availability of cadmium and zinc in soils using inert soil moisture samplers. *Environmental Pollution* **99**, 293–298. doi:10.1016/S0269-7491(98)00021-9
- McGrath SP, Knight B, Killham K, Preston S, Paton GI (1999). Assessment of the toxicity of metals in soils amended with sewage sludge using a chemical speciation technique and a lux-based biosensor. *Environmental Toxicology and Chemistry* **18**, 659–663.
doi:10.1002/etc.5620180411
- Nolan AL, McLaughlin MJ, Mason SD (2003). Chemical speciation of Zn, Cd, Cu, and Pb in pore waters of agricultural and contaminated soils using Donnan dialysis. *Environmental Science and Technology* **37**, 90–98. doi:10.1021/es025966k
- Pampura T, Groenenberg JE, Rietra RPJJ (2006). Comparison of methods for copper free ion activity determination in soil solutions of contaminated and background soils. *Forest Snow and Landscape Research* **80**, 305–322.

- Ren Z-L, Sivry Y, Dai J, Tharaud M, Cordier L, Benedetti MF (2015a). Multi-element stable isotopic dilution and multi-surface modelling to assess the speciation and reactivity of cadmium and copper in soil. *European Journal of Soil Science* **66**, 973–982. doi: 10.1111/ejss.12298
- Ren Z-L, Tella M, Bravin MN, Comans RNJ, Dai J, Garnier J-M, Sivry Y, Doelsch E, Straathof A, Benedetti MF (2015b). Effect of dissolved organic matter composition on metal speciation in soil solutions. *Chemical Geology* **398**, 61–69. doi:10.1016/j.chemgeo.2015.01.020
- Ren ZL, Sivry Y, Tharaud M, Cordier L, Li YT, Dai J, Benedetti MF (2017). Speciation and reactivity of lead and zinc in heavily and poorly contaminated soils: Stable isotope dilution, chemical extraction and model views. *Environmental Pollution* **225**, 654–662. doi: 10.1016/j.envpol.2017.03.051
- Rennert T, Rabus W, Rinklebe J (2017). Modelling the concentrations of dissolved contaminants (Cd, Cu, Ni, Pb, Zn) in floodplain soils. *Environmental Geochemistry and Health* **39**, 331–344. doi:10.1007/s10653-106-9859-4
- Salam AK, Helmke PA (1998). The pH dependence of free ionic activities and total dissolved concentrations of copper and cadmium in soil solution. *Geoderma* **83**, 281–291. doi:10.1016/S0016-7061(98)00004-4
- Sanders JR (1982). The effect of pH upon the copper and cupric ion concentrations in soil solutions. *Journal of Soil Science* **33**, 679–689. doi:10.1111/j.1365-2389.1982.tb01799.x
- Sanders JR (1983). The effect of pH on the total and free ionic concentrations of manganese, zinc and cobalt in soil solutions. *Journal of Soil Science* **34**, 315–323. doi:10.1111/j.1365-2389.1983.tb01037.x
- Sauve S, McBride MB, Hendershot W (1997). Speciation of lead in contaminated soils. *Environmental Pollution* **98**, 149–155. doi:10.1016/S0269-7491(97)00139-5
- Sauvé S, Norvell WA, McBride M, Hendershot W, (2000). Speciation and Complexation of Cadmium in Extracted Soil Solutions. *Environmental Science and Technology* **34**, 291–296. doi:10.1021/es990202z
- Stephan CH, Courchesne F, Hendershot WH, McGrath SP, Chaudri AM, Sappin-Didier V, Sauvé S, (2008). Speciation of zinc in contaminated soils. *Environmental Pollution* **155**, 208–216. doi:10.1016/j.envpol.2007.12.006

Schneider AR, Cancès B, Ponthieu M, Sobanska S, Benedetti MF, Pourret O, Conreux A, Calandra I, Martinet B, Morvan X, Gommeaux M, Marin B, (2016a). Lead distribution in soils impacted by a secondary lead smelter: Experimental and modelling approaches. *Science of the Total Environment* **568**, 155–163. doi:10.1016/j.scitotenv.2016.06.001

Schneider AR, Ponthieu M, Cancès B, Conreux A, Morvan X, Gommeaux M, Marin B, Benedetti MF, (2016b). Influence of dissolved organic matter and manganese oxides on metal speciation in soil solution: A modelling approach. *Environmental Pollution* **213**, 618–627. doi:10.1016/j.envpol.2016.03.010

Simón M, García I, (1999). Physico–chemical properties of the soil–saturation extracts: estimation from electrical conductivity. *Geoderma* **90**, 99–109. doi:10.1016/S0016-7061(98)00098-6

Tipping E, Rieuwerts J, Pan G, Ashmore MR, Lofts S, Hill MTR, Farago ME, Thornton I, (2003). The solid-solution partitioning of heavy metals (Cu, Zn, Cd, Pb) in upland soils of England and Wales. *Environmental Pollution* **125**, 213–225. doi:10.1016/s0269-7491(03)00058-7

Vulkan R, Zhao F-J, Barbosa-Jefferson V, Preston S, Paton GI, Tipping E, McGrath SP (2000). Copper Speciation and Impacts on Bacterial Biosensors in the Pore Water of Copper-Contaminated Soils. *Environmental Science and Technology* **34**, 5115–5121. doi:10.1021/es0000910

Weng L, Temminghoff EJM, Van Riemsdijk WH (2001). Determination of the free ion concentration of trace metals in soil solution using a soil column Donnan membrane technique. *European Journal of Soil Science* **52**, 629–637. doi:10.1046/j.1365-2389.2001.00416.x