

Solid-phase cadmium speciation in soil using L_3 -edge XANES spectroscopy with partial least-squares regression

Nina Siebers,^{a*} Jens Kruse,^a Kai-Uwe Eckhardt,^a Yongfeng Hu^b and Peter Leinweber^a

^aInstitute for Land Use, University of Rostock, Justus-von-Liebig-Weg 6, 18051 Rostock, Germany, and ^bCanadian Light Source Inc., University of Saskatchewan, 101 Perimeter Rd, Saskatoon, SK, Canada S7N 0X4. E-mail: nina.siebers@uni-rostock.de

Cadmium (Cd) has a high toxicity and resolving its speciation in soil is challenging but essential for estimating the environmental risk. In this study partial least-square (PLS) regression was tested for its capability to deconvolute Cd L_3 -edge X-ray absorption near-edge structure (XANES) spectra of multi-compound mixtures. For this, a library of Cd reference compound spectra and a spectrum of a soil sample were acquired. A good coefficient of determination (R^2) of Cd compounds in mixtures was obtained for the PLS model using binary and ternary mixtures of various Cd reference compounds proving the validity of this approach. In order to describe complex systems like soil, multi-compound mixtures of a variety of Cd compounds must be included in the PLS model. The obtained PLS regression model was then applied to a highly Cd-contaminated soil revealing $\text{Cd}_3(\text{PO}_4)_2$ (36.1%), $\text{Cd}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$ (24.5%), $\text{Cd}(\text{OH})_2$ (21.7%), CdCO_3 (17.1%) and CdCl_2 (0.4%). These preliminary results proved that PLS regression is a promising approach for a direct determination of Cd speciation in the solid phase of a soil sample.

Keywords: X-ray absorption; PLS regression; mixtures; Cd; speciation.

1. Introduction

High cadmium (Cd) concentrations are becoming a more and more serious environmental problem in various agro-ecosystems and urban areas, as the increasing use of Cd-bearing phosphorus (P) fertilizer and industrial pollution leads to Cd enrichments in soil (McLaughlin & Singh, 1999). The heavy metal Cd is known for its high toxicity and can, especially in its mobile forms, easily be accumulated in the food chain by plant and animal uptake (Adriano, 2001) with the risk of serious health (e.g. consumers) and economic (e.g. farmers) problems. A first step for risk assessment is the determination of total Cd concentrations in soil. Knowledge about the abundance and behavior of Cd species is fundamental to estimating their fate in the environment, associated risks to human health, and to develop strategies for Cd remediation or immobilization. Sequential extractions have been developed to characterize Cd pools in environmental samples (e.g. Tessier *et al.*, 1979; Zeien & Brümmer, 1989). However, these extractions provide assignments to Cd pools of different solubility rather than chemically defined Cd speciation *sensu strictu*. Furthermore, involved successive extraction procedures can alter the Cd species in the sample (Bertsch *et al.*, 1997). To overcome such

analytical limitations, X-ray absorption spectroscopy (XAS) can be applied, which enables an element-specific and non-invasive direct speciation of Cd in the solid phase (Fendorf & Sparks, 1996). In general, a Cd XANES spectrum is the weighted sum of signals from all Cd species within the sample. Various approaches were applied to deconvolute the spectrum of an unknown sample into proportions of Cd species. The majority of the previously published studies used extended X-ray absorption fine-structure (EXAFS) spectroscopy at the Cd K -edge to investigate Cd complexes in various types of environmental samples by fitting the $\chi(k)$ model function to the EXAFS oscillation by using least-square methods (Sery *et al.*, 1996; Pickering *et al.*, 1999; Mishra *et al.*, 2009; Jalilehvand *et al.*, 2009; Mah & Jalilehvand, 2010; Demchenko *et al.*, 2010). However, there are only a limited number of beamlines available that are capable of providing a stable flux at the necessary energy range of the Cd K -edge (~ 27 keV). Thus, access to those beamlines is restricted by the limited number of possible users. For this reason, attempts were made to use other edges for Cd speciation. The three Cd L -edges are in the energy range between 4018 and 3538 eV, which is available at a great number of beamlines. The capability of Cd L -edge XANES for Cd speciation has been reported in several studies

(e.g. Pickering *et al.*, 1999; Isaure *et al.*, 2006; Jalilehvand *et al.*, 2009), although the sensitivity of the Cd L -edges is lower compared with the Cd K -edge being amenable at concentrations as low as 1 mg kg^{-1} (Mishra *et al.*, 2010). By contrast, the energy resolution at the Cd L -edges is higher than for the Cd K -edge. Most studies used the XANES region of the Cd L_3 -edge, as the EXAFS region is distorted by the feature of the potassium (K) K -edge superimposing the EXAFS oscillation of the Cd L_3 -edge. Therefore, Cd L_3 -edge EXAFS is not suitable for those environmental samples that contain K.

A common approach for XANES data evaluation is the so-called ‘fingerprint approach’. In this, spectra of known reference compounds are visually compared with the spectrum of unknown composition to assign major constituents. Most of the previously published studies evaluated Cd L_3 -edge spectra using this approach as a first step before using more sophisticated methods (e.g. Pickering *et al.*, 1999; Isaure *et al.*, 2006; Jalilehvand *et al.*, 2009). This method gives a first indication of the qualitative composition of the sample. A more sophisticated method, which also enables a quantitative estimation of the Cd composition, is linear combination fitting (LCF). This approach involves finding the best combination of Cd reference compounds spectra which fit the sample spectra when weighted and summed. For example, Isaure *et al.* (2006) used LCF to discriminate Cd bound to O and N ligands from S ligands in plant samples. In complex matrices such as sediment and soil samples many different Cd species may be abundant. Therefore, some researchers have applied principal components analysis (PCA) and an adjacent target transformation in order to preselect major reference compounds for subsequent LCF analysis. This approach was successfully applied to analyze the effect of calcium (Ca) addition on the Cd speciation in tobacco plants (Isaure *et al.*, 2010) and to investigate the interactions and sorption mechanisms of Cd to red mud (Luo *et al.*, 2011).

Partial least-square (PLS) regression is another approach to evaluate spectral data. By using this multivariate calibration method latent variables are computed to model independent variables. By correlating them with dependent variables, linear calibration models are constructed enabling prediction of proportions of individual compounds in spectra (Lindberg *et al.*, 1983). So far, PLS regression has been extensively applied to evaluate infrared spectroscopic data for analyzing the composition of multi-compound samples (e.g. Brown *et al.*, 2004; Janik *et al.*, 2009). However, this method has been applied only rarely to XANES data (Kuno *et al.*, 1999; Kuno & Matsuo, 2000; Yasoshima *et al.*, 2001); these authors tested the capability of PLS regression to estimate the proportions of iron (Fe)-bearing minerals in mixtures and then they successfully applied the obtained PLS model to estuarine sediment samples.

The objectives of the present study were (i) to provide a comprehensive library of Cd L_3 -edge XANES spectra for subsequent studies, (ii) to test the capability of PLS regression to estimate the proportions of known multi-compound Cd mixtures, and (iii) to apply this approach to qualitatively and quantitatively estimate the Cd forms in a soil sample.

2. Experimental

2.1. Cd reference compounds

For XANES analysis 12 reagent-grade Cd reference standards were purchased from different chemical suppliers: Cd(OH)₂, CdO, CdCO₃, Cd(CH₃COCHCOCH₃)₂ (Cd-Org), and Cd₃(PO₄)₂ (Cd-P 1) were purchased from Alfa Aesar (76057 Karlsruhe, Germany), Cd(NO₃)₂·4H₂O and Cd(CH₃COO)₂·2H₂O (Cd-Ac) from J. T. Baker (72793 Pfuldingen, Germany), CdCl₂ was obtained from Fluka (30926 Seelze, Germany), CdSO₄ and 3CdSO₄·8H₂O were purchased from Merck (64293 Darmstadt, Germany). The Cd phosphate standard Cd₅H₂(PO₄)₄·4H₂O (Cd-P 2) was synthesized by applying a slightly modified procedure as previously published by Matusik *et al.* (2008). Briefly, CdCl₂ and K₂HPO₄ were added to 65 ml deionized water to yield a concentration of [Cd_{aq}] = 4.80 mM and [PO₄] = 3.52 mM, respectively. Both solutions were dropwise mixed (~ 5 ml min⁻¹) under continuous stirring to a 500 ml beaker containing 370 ml deionized water. The pH of the solution was kept constant at 5.00 throughout the reaction using potassium hydroxide and nitric acid. The resulting precipitate was filtered and washed with deionized water and subsequently dried at 313 K. All reference compounds were used without further purification and ground in an agate stone mortar and pestle. Solid binary and ternary mixtures of various Cd standards were prepared in various proportions on a Cd atomic mass basis. Furthermore, artificial binary to senary mixtures were mathematically calculated using the normalized spectra of all reference compounds. For this, spectra were summed in various proportions and the resulting spectra were normalized as described below.

2.2. XANES data collection and data treatment

The Cd L_3 -edge XANES measurements were performed at the Canadian Light Source, Canada, a 2.9 GeV third-generation synchrotron source, on the soft X-ray micro-characterization beamline (SXRMB) covering an energy range of 1.7 to 10 keV (Hu *et al.*, 2010). The air-dried and homogenized samples were spread as thin film onto a double-sided carbon tape and mounted onto a copper sample holder before being placed in the vacuum chamber. A Si(111) double-crystal monochromator ($\Delta E/E \simeq 10^{-4}$) was used for all measurements. All XANES spectra of reference compounds were recorded in total electron yield mode to avoid self-absorption effects, and the soil sample was measured in fluorescence yield mode. All fluorescence yield spectra were recorded within the energy range between 3500 and 3650 eV using a silicon drift detector. The chosen step size was of 0.7 eV (3500 to 3530 eV), 0.2 eV (3530 to 3580 eV) and 0.8 (3580 to 3650 eV). Furthermore, a dwell time of 1 s for reference compounds and 10 s for the soil sample were chosen. Three to four scans were recorded, and the beam was moved to a ‘fresh’ sample spot after each scan to avoid X-ray-induced changes in the sample. Afterwards, spectra of these individual scans were averaged to improve the signal-to-noise ratio and then background corrected by subtracting a linear

regression fit through the pre-edge region. Finally, all spectra were normalized to the mean of the last 28 data points. This data treatment was performed using *OriginPro8* (OriginLab, MA, USA).

2.3. PLS regression

For PLS analysis the PLS package (Wehrens & Mevik, 2007) *R* 2.13.0 (R Development Core Team, 2011) was used. All PLS models were built using 264 independent variables representing the data points of the Cd spectra in the energy range between 3530 and 3590.4 eV. Depending on the composition of the mixtures up to six dependent variables representing the relative abundance (0 to 1) of the reference compounds were used. The leave-one-out cross-validation method was used as only a limited number of spectra of reference compounds and mixtures were available to build a PLS calibration model. The spectral data were separated into a calibration subset used for building the PLS calibration model and a test subset based on individual scans. The PLS calibration model was then applied to predict and evaluate the composition of the mixtures in the test subset and in the soil sample.

2.4. PCA

A principal component analysis of the normalized spectra in the range 3530–3590.4 eV (264 variables) was calculated with *R* 2.13.0 using a singular value decomposition of the correlation matrix (command `prcomp`) after scaling to a mean of 0 and a standard deviation of 1.

2.5. Soil sample analyses

The soil particle size distribution was determined by the standard pipette method. Total Cd and P concentrations of the soil sample were determined after microwave-assisted digestion with HNO_3 and HCl (CEN, 2000) using an inductively coupled plasma-optical emission spectroscope (JY 238, Jobin Yvon, France). The pH of the soil was measured in 0.1 M CaCl_2 solution at a soil:solution ratio of 1:2.5 (Blume *et al.*, 2011). The soil had a particle size distribution of 26.5% clay, 55.3% silt and 18.2% sand, a pH of 6.2 (CaCl_2 method), and exhibited a total Cd concentration of 1028 mg kg^{-1} .

3. Results and discussion

3.1. Cadmium reference compounds

A first direct comparison between Cd L_2 -edge and L_3 -edge XANES spectra of two selected compounds is shown in Fig. 1. The spectra of the Cd L_2 -edge were similar in shape to those of the Cd L_3 -edge, but exhibited beamline features between 3590 and 3605 eV that are due to the Si(111) crystal glitches and difficult to normalize. Additionally, the L_3 -edge spectra exhibited more intensive and slightly better resolved spectral features. Therefore, the Cd L_3 -edge was chosen for further analysis.

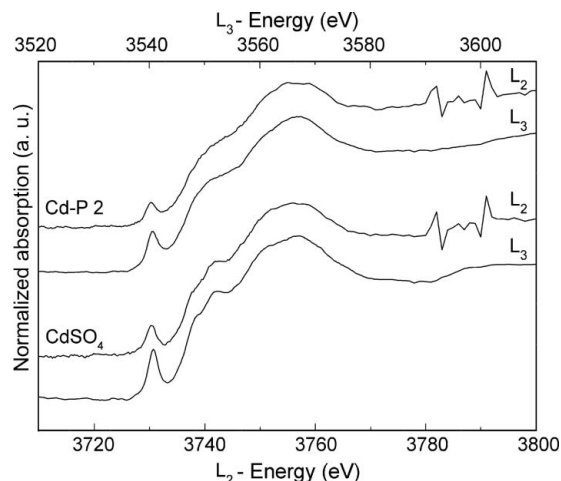


Figure 1
Comparison of the Cd L_2 - and L_3 -edge spectra for $\text{Cd}_3\text{H}_2(\text{PO}_4)_4 \cdot 4\text{H}_2\text{O}$ (Cd-P 2) and CdSO_4 .

The Cd L_3 -edge of the different Cd(II) reference compounds exhibited several spectral features, which varied in energy position and shape (Fig. 2). The exact energy positions of the characteristic features are listed in Table 1. The majority of the spectra exhibited a sharp feature in the energy region between 3540.2 and 3541.2 eV. This feature can be assigned to the transitions from the Cd $2p_{3/2}$ state to unoccupied states to $5s$ orbitals, which are formed in the presence of O and N ligands (Pickering *et al.*, 1999; Isaure *et al.*, 2006). In contrast, this feature was absent in the spectra of CdS, since Cd in this compound is not associated with O or N ligands (Fig. 2). Organic Cd compounds exhibited less resolved spectral features most probably due to the covalent character of the Cd bonding, resulting in less defined orbitals and transitions.

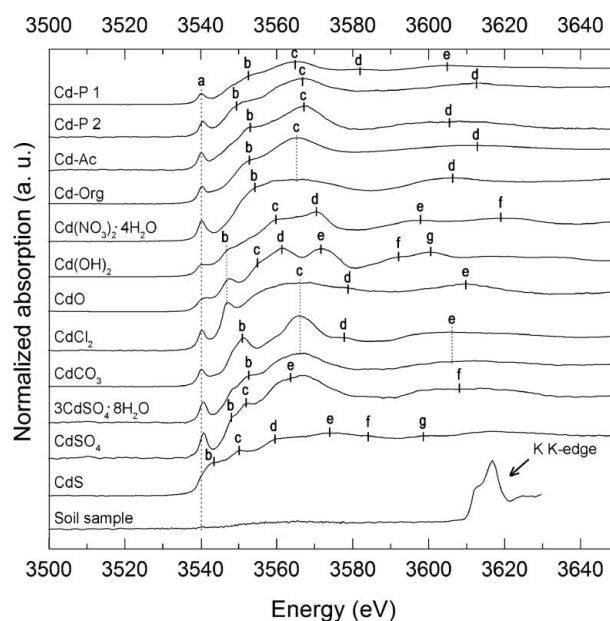


Figure 2
Cd L_3 -edge XANES spectra of different reference compounds. Note that the Cd features of the soil spectra are scaled down owing to the high intensity of the K peak.

Table 1

Energy positions of the main XANES features (in eV) for all Cd reference compounds.

b: broad; sh: shoulder; w: weak; CAS refers to the CAS registry number.

Reference compound	CAS	Feature						
		<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>
Cd-P 1	13477-17-2	3540.2	3553.8 ^{sh}	3565.2	3581.6 ^b	3604.8		
Cd-P 2	–	3540.6	3551.4 ^{sh}	3566.6	3612.0 ^b			
Cd-Org	14689-45-3	3540.2	3553.0 ^{sh}	3567.4	3606.4 ^b			
Cd-Ac	5743-04-4	3540.2	3552.8 ^{sh}	3565.2	3611.2 ^b			
Cd(NO ₃) ₂ ·4H ₂ O	10022-68-1	3540.2	3555.4 ^{sh}	3563.8 ^b	3605.6 ^b			
Cd(OH) ₂	21041-95-2	3540.2	3547.4 ^{sh}	3560.4 ^{sh}	3570.6	3596.8 ^b	3619.2 ^b	
CdO	1306-19-0	3541.2	3547.6	3555.8 ^{sh}	3561.4	3571.6	3592.0 ^{sh}	3600.8
CdCl ₂	10108-64-2	3540.2	3547.0	3565.8 ^b	3578.0 ^{sh}	3609.6 ^b		
CdCO ₃	513-78-0	3540.2	3551.0	3565.6	3606.4 ^b			
3CdSO ₄ ·8H ₂ O	15244-35-9	3540.6	3553.0 ^{sh}	3566.2	3612 ^b			
CdSO ₄	10124-36-4	3540.6	3548.8 ^{sh}	3552.0 ^w	3566.6	3610.4 ^b		
CdS	1306-23-6	3543.4 ^{sh}	3545.4 ^{sh}	3550.2	3562.8 ^{sh}	3575.2 ^b	3584 ^{sh}	3600 ^b

Table 2

Results of PLS regression of binary and ternary mixtures of various standard Cd compounds using the PLS regression model based on real measured mixtures.

Numbers in parentheses are the percentage of actual amount of the compounds within the mixture.

Test subset	Cd(NO ₃) ₂ ·4H ₂ O	CdCl ₂	CdSO ₄
1	−0.8 (0)	2.5 (0)	98.3 (100)
2	3.9 (0)	1.3 (0)	94.8 (100)
3	55.7 (50)	−4.1 (0)	48.3 (50)
4	75.4 (80)	0.6 (0)	24.0 (20)
5	25.6 (20)	−1.0 (0)	75.4 (80)
6	38.6 (33.3)	30.9 (33.3)	30.5 (33.3)
7	65.2 (60)	18.6 (20)	16.2 (20)
8	24.2 (20)	57.9 (60)	17.8 (20)
9	26.3 (20)	16.3 (20)	57.4 (60)
Standard error (%)	4.9	2.4	3.4
R ²	0.964	0.985	0.988

Generally, the shape of the presented spectra agreed with Cd L₃-edge spectra previously published in the literature (e.g. Pickering *et al.*, 1999; Isaure *et al.*, 2006) but had better resolved features for some compounds, e.g. CdO. Furthermore, for all spectra in Fig. 2 the Cd L₃-edge EXAFS oscillations were also recorded for a longer post-edge in the energy range than previously published spectra, which were recorded only to ~3590 eV. This longer post-edge range enables a more precise edge step normalization of the spectra for subsequent quantitative data evaluation, since the EXAFS features observed, for example, for CdO and Cd(OH)₂ can be better averaged out.

Fig. 2 also shows the spectrum of a Cd-polluted soil sample. This spectrum was characterized by a strong feature originating from the potassium (K) 1s → π transition. Note that any Cd features in the soil spectrum are scarcely visible owing to the high intensity of the K peak in Fig. 2. However, a scaled version of this spectrum showing only Cd-relevant features is shown in Fig. 4. Because of the abundance of K in soils the speciation and quantification of Cd using Cd L₃-edge XANES spectra is possible only in the energy range below 3605 eV.

Therefore, the energy range 3530–3590.4 eV was used for subsequent spectral data evaluation.

3.2. PLS regression applied to spectra of mixtures of Cd compound

To test the ability of the PLS regression to estimate the relative abundance of Cd compounds in multi-compound samples, spectra from solid mixtures containing Cd(NO₃)₂·4H₂O, CdCl₂ and two Cd sulfates were subjected to PLS regression analysis. Since the spectra of CdSO₄ and 3CdSO₄·8H₂O were very similar in their spectral features, they were treated as one compound. The

number of latent variables was determined by constructing the PLS models and varying the number of latent variables (one to ten). Cross-validation and comparison of the error of prediction indicated that six latent variables were sufficient to estimate the proportion of the different compounds within the mixtures. The calibration model achieved good accuracy with high R² values for each Cd compound used in the test subset (Table 2). The calculated standard errors for individual mixtures varied from 2.4 to 4.9%, proving that PLS regression is suitable to deconvolute the mixtures (Table 2).

However, PLS regression can only predict proportions of compounds that were used for building and calibrating the PLS regression model. Therefore, to apply the PLS regression to more complex systems like soil and soil-related samples containing various Cd forms, data sets used for PLS model construction must contain spectra of multi-compound mixtures of all expected soil-relevant Cd compounds. Thus, a large data set of spectra from various binary to *n*-nary mixtures of Cd compounds in varying proportions must be included in the PLS regression model. Since the number of necessary mixtures can easily exceed the feasibility of measurements at a synchrotron facility (limited beam time), it was not possible to record Cd L₃-edge spectra of such a large sample set. Thus, we tested whether PLS regression models built using artificial mixtures will yield comparable results as ‘real’ measured mixtures. For this, mixtures with the same composition and proportions of the above three Cd compounds were constructed mathematically using the spectra of Cd reference compounds (Fig. 2) and a PLS regression model was constructed. Using the same test subset for both models better R² values were obtained for the mathematical model (Table 3).

The obtained proportions of the mixture constituents were comparable and even more precise, as indicated by the reduced standard error being in the range 1.3–2.4%. This can be explained by the inaccuracy of the ‘real’ mixtures by, for example, inhomogeneity of the dry powder mixtures (Chen & Yu, 2004) and other analytical and chemical bias, which can be ruled out when using mathematical artificial mixtures. To avoid possible influences of such mixture inaccuracies on the

Table 3

Results of PLS regression of binary and ternary mixtures of various standard Cd compounds using the PLS regression model based on mathematical mixtures.

Numbers in parentheses are the percentage of actual amount of the compounds within the mixture.

Test subset	Cd(NO ₃) ₂ ·4H ₂ O	CdCl ₂	CdSO ₄
1	-0.7 (0)	2.2 (0)	98.4 (100)
2	-3.2 (0)	1.9 (0)	101.3 (100)
3	54.0 (50)	-2.3 (0)	48.3 (50)
4	75.6 (80)	0.1 (0)	24.3 (20)
5	22.3 (20)	-0.2 (0)	77.9 (80)
6	33.3 (33.3)	33.3 (33.3)	33.4 (33.3)
7	59.5 (60)	20.5 (20)	19.9 (20)
8	20.0 (20)	60.0 (60)	20.0 (20)
9	20.4 (20)	19.3 (20)	60.3 (60)
Standard error (%)	2.4	1.3	1.8
R ²	0.991	0.996	0.997

data evaluation and owing to the above-mentioned beam time constraints a mathematical model was created to quantify all relevant Cd compounds in soils. A PCA was applied to define the number and type of Cd compound that should be included in the mathematical PLS regression model.

3.3. PCA of a soil sample

Fig. 3 shows the score plot of the PCA including all Cd reference compounds and the soil sample. The obvious clustering of the two Cd sulfates also supported the above consideration of these two compounds as one (Cd-SO₄). Furthermore, Cd(NO₃)₂·4H₂O, Cd-P 1 and Cd-Ac were clustering. This is in agreement with visual inspection of the spectra of Cd-P 1 and Cd-Ac, as they were very similar in spectral features. This is different for Cd K-edge EXAFS, where the spectra of Cd-Ac and Cd-P are distinguishable from each other (Mishra *et al.*, 2009). Since pretests on Cd L₃-edge

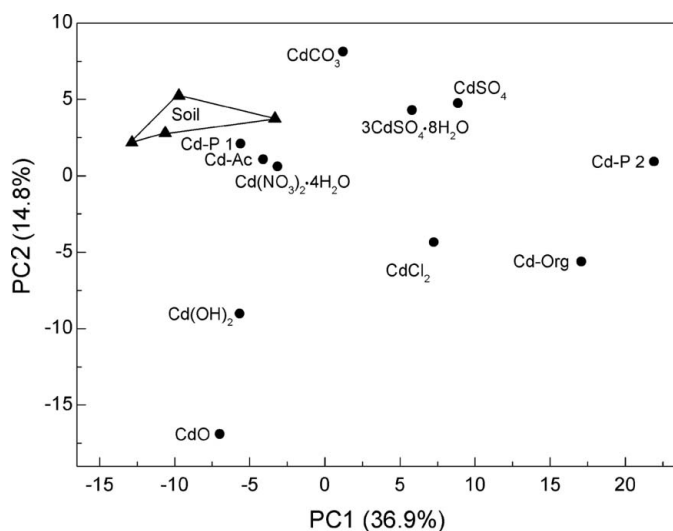


Figure 3
PCA score plot of Cd reference compounds and the soil sample; Cd(CH₃COCHCOCH₃)₂ (Cd-Org), Cd(CH₃COO)₂·2H₂O (Cd-Ac), Cd₃(PO₄)₂ (Cd-P 1) and Cd₅H₂(PO₄)₄·4H₂O (Cd-P 2).

mixtures containing Cd-P 1 and Cd-Ac revealed that PLS regression was not able to differentiate these two compounds (data not shown), only Cd-P 1 was used for further analysis representing both compounds (Cd-P 1/Cd-Ac).

As individual scans of the soil sample were used in the PCA, they showed less clustering owing to the heterogeneity of the soil (Fig. 3). Furthermore, PCA revealed that six components were adequate for describing the data, accounting for 84.1% of total variance of the soil spectra. Thus, the following six Cd compounds were selected for PLS regression grouping closest to the soil cluster: Cd-P 1/Cd-Ac, Cd(NO₃)₂·4H₂O, CdCO₃, Cd-SO₄, Cd(OH)₂ and CdCl₂.

3.4. PLS regression applied to spectra of a soil sample

Based on the PCA results the following Cd standards were used to build a PLS model out of the Cd reference compound spectra and 45 artificial (mathematical) mixtures: Cd(NO₃)₂·4H₂O, CdCl₂, CdSO₄, Cd-P 1/Cd-Ac, CdCO₃ and Cd(OH)₂. For building the PLS regression model mathematically, mixtures containing up to six Cd compounds were calculated at different ratios in the proportions of 0 to 1. The lowest species amount used for creating the mathematical model was 14.3% of each compound. Thus, concentrations were chosen to incorporate the critical amount of a species that are difficult to assess using XANES spectroscopy. The model was applied in the energy range between 3530 and 3590.4 eV (264 independent variables) and six latent variables were used. To validate this model the PLS regression was applied to the same test subset as above. Standard errors for individual mixtures are presented in Table 4 and ranged between 0.9 and 7.2%, which is comparable with and even better than the above tested models built out of only three Cd reference compounds.

Furthermore, the PLS regression model was tested with another set of spectra from the above 45 mathematical mixtures (up to six Cd compounds, separate scans not used for model construction, data not shown). R² ranged from 0.989 to 0.999 for each Cd compound used in the test subset, indicating a good predictability of Cd compounds with that model, which means that PLS regression can resolve XANES spectra and quantify the Cd species. Because the reference compounds used for creating the model are the main Cd forms in soil solution (Sposito, 1998), and Cd sulfates and Cd carbonates are the main anthropogenic Cd contaminants in soil (Bataillard *et al.*, 2003), the created model appears applicable for many types of Cd contamination.

As a first application the PLS regression model was used to analyze the Cd composition of a highly Cd-contaminated soil sample. The PLS regression revealed the following composition: Cd-P 1/Cd-Ac (36.3%), Cd(NO₃)₂·4H₂O (24.5%), Cd(OH)₂ (21.7%), CdCO₃ (17.1%) and CdCl₂ (0.4%). The individual spectra of these compounds, weighted by their proportions in soil, are shown in Fig. 4. The presence of CdCl₂ is precarious in the soil sample, because it is unlikely that such small proportions can be quantified by PLS regression even though CdCl₂ exhibited a strong feature at around 3547.0 eV.

Table 4

Results of PLS regression of binary and ternary mixtures of various standard Cd compounds using the PLS regression model based on mathematical mixtures.

Numbers in parentheses are the percentage of actual amount of the compounds within the mixture.

Test subset	Cd(NO ₃) ₂ ·4H ₂ O	CdCl ₂	CdSO ₄	Cd-P 1	CdCO ₃	Cd(OH) ₂
1	10.1 (0)	6.2 (0)	77.4 (100)	2.6 (0)	3.8 (0)	0 (0)
2	0.9 (0)	3.0 (0)	94.5 (100)	0.0 (0)	1.5 (0)	0 (0)
3	47.3 (50)	0.0 (0)	39.9 (50)	12.8 (0)	0.0 (0)	0 (0)
4	63.9 (80)	0.0 (0)	24.1 (20)	5.8 (0)	0.0 (0)	0 (0)
5	21.3 (20)	0.0 (0)	60.0 (80)	18.7 (0)	0.0 (0)	0 (0)
6	34.5 (33.3)	31.2 (33.3)	34.3 (33.3)	0.0 (0)	0.0 (0)	0 (0)
7	61.9 (60)	17.6 (20)	20.5 (20)	0.0 (0)	0.0 (0)	0 (0)
8	19.3 (20)	54.6 (60)	19.2 (20)	2.1 (0)	0.0 (0)	0.5 (0)
9	20.3 (20)	17.6 (20)	60.2 (60)	4.3 (0)	0.0 (0)	0 (0)
Soil sample	24.5	0.4	0	36.3	17.1	21.7
R ²	0.992	0.991	0.991	0.999	0.999	0.989

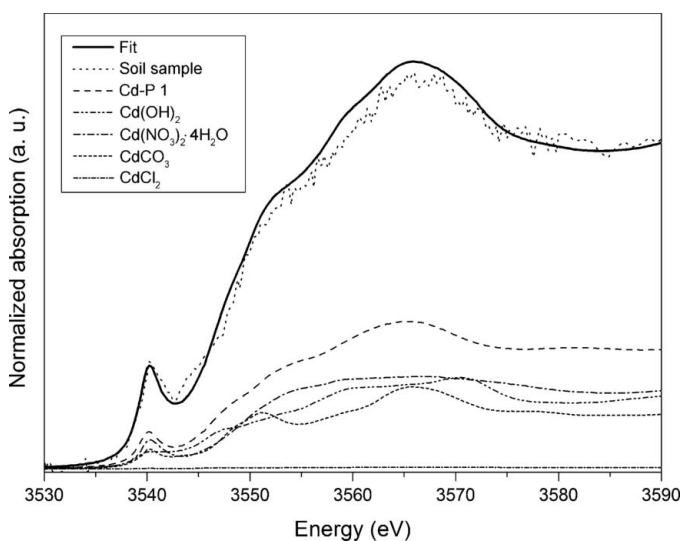


Figure 4 Normalized Cd XANES spectra of model compounds and the soil sample (the solid curve is the fit for the soil sample).

Generally, it is assumed that a species can be quantified by XANES if it comprises more than 10% within the mixture (Beauchemin *et al.*, 2003). Therefore, CdCl₂ should be considered as not present in the soil sample, but for the sake of completeness it was not set to zero. The fit obtained from the summed spectra of individual compounds sufficiently agreed with the measured spectrum of the soil sample. Previous attempts to speciate Cd soils often relied on sequential extractions that yield the following operationally defined fractions: (1) soluble/exchangeable Cd representing the most easily available metals, (2) carbonate Cd, (3) oxidizable Cd (including Cd oxides bound to organics), (4) reducible and (5) residual Cd (Tessier *et al.*, 1979). Publications have shown that, in soils having diverse physicochemical properties and receiving Cd contamination from various sources, the majority of Cd is bound in the first three fractions (*e.g.* Ahnstrom & Parker, 1999; Zimmer *et al.*, 2011). This is supported by the high proportion (36.3%) of Cd as Cd-P 1/Cd-Ac [representing Cd₃(PO₄)₂ belonging to fraction (3) and Cd bound to organics] as derived from the PLS regression of the L₃-edge XANES

spectra (Fig. 4). Furthermore, it was shown that Cd can precipitate as CdCO₃ at high Cd loadings (McBride, 1980; Prieto *et al.*, 2003), agreeing with the substantial contribution of the CdCO₃ (17.1%) to the fit of the soil sample. However, these comparisons are restricted by the fact that no single Cd species has been identified by these operationally defined fractions yet. So far only one study directly determined the Cd speciation in the solid phase of a contaminated soil (Khaokaew *et al.*, 2011). The Cd K-edge EXAFS spectra revealed that the majority of Cd was associated with humic acid in the highly contaminated alkaline paddy soil (142 mg kg⁻¹ Cd). Despite quite different soils (terrestrial *versus* paddy) and methods (XANES *versus* EXAFS), speciation obtained by these synchrotron-radiation-based studies agreed to some extent, *i.e.* the importance of organically bound Cd and proportions of CdCO₃.

4. Conclusions

The present study showed for the first time that PLS regression applied to Cd L₃-edge XANES spectra has great potential to directly determine Cd speciation in solid multi-compound mixtures and highly contaminated soils. Since soil-relevant Cd compounds were used to build the PLS regression model, the method described here can be applied to various soil conditions and Cd contaminations. These results may encourage the application of PLS regression for Cd XAS data evaluation. However, a direct comparison between PLS regression and the more common LCF approach is desirable to test the sensitivity of both approaches. Owing to the lower sensitivity of the Cd L-edge compared with the K K-edge, the data evaluation is still challenging. Therefore, more work has to be done to enhance the sensitivity by, for example, improving the signal-to-noise ratio by inhibiting the signals from the background and elastic scattering *via* filters (Mishra *et al.*, 2009).

Since Cd contamination is a serious environmental problem world-wide, its direct speciation in contaminated samples is the first crucial step towards the estimation of its environmental risks and the development of remediation strategies and to evaluate their success. In a forthcoming study we will use this approach to track changes in the Cd speciation of soils that were subjected to the common sequential Cd extractions. This will contribute to a validation or modification of the operationally defined Cd fractions as carried out previously for phosphorus (*e.g.* Beauchemin *et al.*, 2003; Kruse *et al.*, 2010).

The Cd L₃-edge XANES measurements were performed at the Canadian Light Source, which is supported by the Natural Sciences and Engineering Research Council of Canada, the

National Research Council Canada, the Canadian Institutes of Health Research, the Province of Saskatchewan, Western Economic Diversification Canada, and the University of Saskatchewan. We thank Dr Volkmar König, Thuringian State Research Centre for Agriculture (TLL), for providing the soil used in this study. NS also acknowledges a PhD grant from the Federal State of Mecklenburg-Western Pomerania in Germany.

References

- Adriano, D. C. (2001). *Trace Elements in Terrestrial Environments: Biogeochemistry, Bioavailability and Risk of Metals*. Vol. 2, pp. 540–800. New York: Springer.
- Ahnstrom, Z. S. & Parker, D. R. (1999). *Soil Sci. Soc. Am. J.* **63**, 1650–1658.
- Bataillard, P., Cambier, P. & Picot, C. (2003). *Eur. J. Soil Sci.* **54**, 365–376.
- Beauchemin, S., Hesterberg, D., Chou, J., Beauchemin, M., Simard, R. R. & Sayers, D. E. (2003). *J. Environ. Qual.* **32**, 1809–1819.
- Bertsch, P. M., Hunter, D. B., Nuessle, P. R. & Clark, S. B. (1997). *J. Phys. IV*, **7**, 817–818.
- Blume, H. P., Stahr, K. & Leinweber, P. (2011). *Bodenkundliches Praktikum*, Vol. 3. p. 255. Heidelberg: Spektrum Akademischer Verlag.
- Brown, D. J., Brickley, R. S. & Miller, P. R. (2004). *Geoderma*, **129**, 251–267.
- CEN (2000). *EN 13346, Bestimmung von Spurenelementen und Phosphor, Euräisches Komitee für Normung*. Brussels, Belgium.
- Chen, C. & Yu, C. (2004). *Mater. Chem. Phys.* **85**, 227–237.
- Demchenko, I. N., Denlinger, J. D., Chernyshova, M., Yu, K. M., Speaks, D. T., Olalde-Velasco, P., Hemmers, O., Walukiewicz, W., Derkachova, A. & Lawniczak-Jablonska, K. (2010). *Phys. Rev. B*, **82**, 075107.
- Fendorf, S. E. & Sparks, D. L. (1996). *Methods of Soil Analysis: Chemical Methods, SSSA Book Series 5*, edited by D. L. Sparks, pp. 377–416. Madison: Soil Science Society of America.
- Hu, Y. F., Coulthard, I., Chevrier, D., Wright, G., Igarashi, R., Sitnikov, A., Yates, E. L., Hallin, E. L., Sham, T. K. & Reiniger, R. (2010). *AIP Conf. Proc.* **1234**, 343–346.
- Isaure, M. P., Fayard, B., Sarret, G., Pairis, S. & Bourguignon, J. (2006). *Spectrochim. Acta*, **12**, 1242–1252.
- Isaure, M. P., Sarret, G., Harada, E., Choi, Y. E., Marcus, M. A., Fakra, S. C., Geoffroy, N., Pairis, S., Susini, J., Clemens, S. & Manceau, A. (2010). *Geochim. Cosmochim. Acta*, **74**, 5817–5834.
- Jalilehvand, F., Leung, B. O. & Mah, V. (2009). *Inorg. Chem.* **48**, 5758–5771.
- Janik, L. J., Forrester, S. T. & Rawson, A. (2009). *Chemom. Intell. Lab. Syst.* **97**, 179–188.
- Khaokaew, S., Chaney, R. L., Landrot, G., Ginder-Vogel, M. & Sparks, D. L. (2011). *Environ. Sci. Technol.* **45**, 4249–4255.
- Kruse, J., Negassa, W., Appathurai, N., Zuin, L. & Leinweber, P. (2010). *J. Environm. Qual.* **39**, 2179–2184.
- Kuno, A. & Matsuo, M. (2000). *Anal. Sci.* **16**, 597–602.
- Kuno, A., Matsuo, M. & Numako, C. (1999). *J. Synchrotron Rad.* **6**, 667–669.
- Lindberg, J. A., Persson, J. A. & Wold, S. (1983). *Anal. Chem.* **55**, 643–648.
- Luo, L., Ma, C., Ma, Y., Zhang, S., Lv, J. & Cui, M. (2011). *Environ. Pollut.* **159**, 1108–1113.
- Mah, V. & Jalilehvand, F. (2010). *J. Biol. Inorg. Chem.* **15**, 441–458.
- Matusik, J., Bajda, T. & Manecki, M. (2008). *J. Hazard. Mater.* **152**, 1332–1339.
- McBride, M. B. (1980). *Soil Sci. Soc. Am. J.* **44**, 26–28.
- McLaughlin, M. J. & Singh, B. R. (1999). *Cadmium in Soils and Plants*, pp. 128–135. Dordrecht: Kluwer Academic.
- Mishra, B., Boyanov, M. I., Bunker, B. A., Kelly, S. D., Kemner, K. M. & Fein, J. B. (2010). *Geochim. Cosmochim. Acta*, **74**, 4219–4233.
- Mishra, B., Boyanov, M. I., Bunker, B. A., Kelly, S. D., Kemner, K. M., Nerenberg, R., Read-Daily, B. L. & Fein, J. B. (2009). *Geochim. Cosmochim. Acta*, **73**, 4311–4325.
- Pickering, I. J., Prince, R. C., George, G. N., Rauser, W. E., Wickramasinghe, W. A., Watson, A. A., Dameron, C. T., Dance, I. G., Fairlie, D. P. & Salt, D. E. (1999). *Biochim. Biophys. Acta*, **1429**, 351–364.
- Prieto, M., Cubillas, P. & Fernandez-Gonzales, A. (2003). *Geochim. Cosmochim. Acta*, **67**, 3859–3869.
- R Development Core Team (2011). *The R Project for Statistical Computing*, <http://www.R-project.org/>.
- Sery, A., Manceau, A. & Greaves, G. N. (1996). *Am. Mineral.* **81**, 864–873.
- Sposito, G. (1998). *Bodenchemie*, p. 230. Stuttgart: Ferdinand Enke Verlag.
- Tessier, A., Campbell, P. G. C. & Bisson, M. (1979). *Anal. Chem.* **51**, 844–851.
- Wehrens, R. & Mevik, B. H. (2007). *ppls*, <http://mevik.net/work/software/ppls.html>.
- Yasoshima, M., Matsuo, M., Kuno, A. & Takano, B. (2001). *J. Synchrotron Rad.* **8**, 969–971.
- Zeien, H. & Brümmer, G. W. (1989). *Boden. Mitt. Deutsch. Bodenk. Gesellsch.* **59**, 505–510.
- Zimmer, D., Kiersch, K., Baum, C., Meissner, R., Müller, R., Jandl, G. & Leinweber, P. (2011). *Soil Air Water*, **39**, 328–337.