Proceedings of the 7th International Conference on Materials Chemistry and Environmental Engineering
DOI: 10.54254/2755-2721/2027.38839

Environmental Behavior of Heavy Metal Chemical Speciation and Migration Regulated by Dissolved Organic Matter in Soil

Shuya Teng

School of Environmental Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, China
wellington589125@gmail.com

Abstract. Heavy metal soil pollution is still a serious environmental issue as the risk assessment relies not only on total metal concentration but also on chemical speciation and mobility. DOM impacts the metal transport via complexation, competitive adsorption, and colloid-facilitation; however, its effect varies depending on molecular structure and concentrations. The column experiments of soil for Cd, Cu, and Pb were carried out with the application of control, low- and high-concentration straw-derived DOM, and low- and high-concentration humic-rich DOM. The BCR sequential extraction, ICP-MS, UV-Vis spectroscopy, EEM-PARAFAC, and XGBoost-SHAP techniques were combined to associate DOM properties with the speciation and mobility of metals. Simulation results which fit the experimental setting demonstrated that high straw-derived DOM increased the Cd acid-soluble fraction up to 61.3% and cumulative migration up to 31.4%; however, for high humic-rich DOM, the parameters equal 39.5% and 11.6%. The R^2 of XGBoost models was higher than that of Random Forest and SVR models and equal to 0.93, 0.90, and 0.88 for Cd, Cu, and Pb migration, correspondingly. According to SHAP analysis, acid-soluble fraction and DOC played the dominant role for the migration of Cd; however, humic-like fluorescence and SUVA254 had a greater impact on Cu and Pb.

Keywords: dissolved organic matter, heavy metal speciation, soil pollution, migration behavior, machine learning

1. Introduction

Metal species bound to agricultural soils can travel into groundwater systems or into food chains depending on changes in the chemical nature of the soil that can cause an increase in the mobile forms of these species. Hence, metal speciation is a better means to assess environmental risks than the total concentrations, especially for Cd, Cu, and Pb [1]. DOM has been shown to be highly reactive since the carboxyl, phenolic, and aromatic functionalities are highly interactive with metal ions and surfaces of minerals. Thus, the interactions cause changes in the adsorption, complexation, and transportation processes [2]. Newer studies have revealed that the role of DOM in facilitating metal mobility by forming soluble complexes or the role it plays in immobilization of metal ions on

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2027.38839

soil by humic materials is unpredictable due to the complex interplay of DOM composition, pH, metal-specific affinity, and DOM concentration. A system combining soil column study with BCR fractionation, optical DOM analysis, and ICP-MS measurement along with predictive data mining was employed in this work.

2. Literature review

2.1. Chemical reactivity of DOM

DOM is not homogeneous but a combination of several substances where reactivity will depend on composition rather than carbon content alone. DOM extracted from straw will generally be degradable and protein-like, while humified DOM will comprise mainly aromatic compounds interacting with minerals and metals [3]. EEM-PARAFAC allows for decomposing the overlapping fluorescence bands to humic-like, fulvic-like, and protein-like compositions, thus detecting variations not visible in bulk DOC analysis [4]. The UV-Vis indicators give additional insight: SUVA254 shows aromatic content, HIX is an indicator of humification, and BIX represents biological DOM recently formed. The indicators may vary in soils as a result of preferential sorption of specific DOM compositions. Therefore, even with equal DOC content, there may be different metal binding conditions, hence the need for both concentration and optical indicators.

2.2. Heavy metal speciation transformation

The degree of metal speciation regulates how much can be released or transported through pore space. The modified BCR approach allows the separation into four groups of elements: acid-soluble, reducible, oxidizable, and residual. This is the way of assessment of potential mobility [5]. The mobility of cadmium is regulated by pH changes and competition, while for copper retention, there is an impact from ionic strength, DOM adsorption, and competition [6]. The DOM impacts metals differently: while the soluble forms cause the increase of desorption, the humified molecules enhance the retention or even get immobilized with metals in complexes. Lead binding is influenced by the differences between DOM components and their transformations, proving that the fluorescence composition is not a passive characteristic of the matter [7].

2.3. Data-driven analysis of metal migration

The column tests yield repetitive leaching results, optical characteristics, soil chemistry, and operational metal fractions, yielding non-linear dependencies that cannot be described using simple one-dimensional correlations. Machine-learning techniques allow taking into account all these factors and possible interactions between the pH, organic matter characteristics, and metal availability; XGBoost has already proven its ability to predict the availability of Cd from soil and the environment [8]. However, even accurate predictions do not provide any insight regarding whether the presence of a variable is conducive to or inhibits the transport of the contaminant at certain concentrations. The Explainable machine learning resolves this issue by breaking down the predictions into individual contributions of the features used. XGBoost-SHAP has already been employed for the prediction of Cd, Cu, and Pb contamination to reveal environmental controls and interaction patterns [9].

DOI: 10.54254/2755-2721/2027.38839

3. Experimental methods

3.1. DOM gradient column leaching experiment

Topsoil samples of agricultural land were collected from the 0-20 cm depth, air dried, homogenized, and sieved through a 2mm sieve. Soil was spiked with 5 mg kg^{-1} Cd, 100 mg kg^{-1} Cu, and 200 mg kg^{-1} Pb and equilibrated at 60% water-holding capacity, followed by aging of soil for 30 days. Five sets of treatments were created including control without the addition of any DOM, low and high concentrations of straw-based DOM, and low and high concentrations of humic-rich DOM. Six columns of each treatment were used, making 30 columns in total. Columns made up of PVC were packed to same bulk density using 20 cm depth of soil bed and pre-equilibrated with background electrolyte solution before conducting leaching experiment. Solutions of DOM were passed through 0.45 μm membrane filters and standardized based on DOC content before adding to columns. Each column was subjected to leaching of 12 pore volumes at a constant flow rate with sampling of effluent after each pore volume. Sampling was done for ICP-MS, DOC, pH, EC, UV-Vis, and EEM analysis.

3.2. Chemical characterization of DOM and metals

DOM and metal chemistry were determined at each pore volume in order to relate optical composition to the distribution in solid phase. The concentrations of DOC were determined by TOC Analyzer, while the pH and EC measurements were done right away after sampling. DOM-specific UV absorbance (an operational measure of DOM aromaticity) was calculated from UV-Vis absorbance at 254 nm according to Equation (1).

$$SUVA_{254}=\frac{100A_{254}}{DOC\times l} \tag{1}$$

Where A254 is absorbance at 254 nm, DOC is dissolved organic carbon concentration in mg L^{-1}, and l is optical path length in cm. EEM spectra were corrected for blanks and inner-filter effects before PARAFAC decomposition to quantify humic-like and protein-like components [10]. Modified BCR extraction separated acid-soluble, reducible, oxidizable, and residual metal pools before and after column leaching. The proportion of each fraction was calculated using Equation (2).

$$F_i=\frac{M_i}{\sum_{j=1}^{4} M_j}\times 100\% \tag{2}$$

Where M_i is the metal mass assigned to fraction i and the denominator is the summed mass of all four fractions. Cd, Cu, and Pb in leachates and BCR extracts were quantified by ICP-MS using matrix-matched calibration and procedural blanks. These measurements produced synchronized DOM descriptors, aqueous metal concentrations, and solid-phase fractions for subsequent migration modeling.

DOI: 10.54254/2755-2721/2027.38839

3.3. XGBoost migration prediction model

Individual prediction models were built for each metal, with DOC, SUVA254, HIX, BIX, PARAFAC intensity, pH, EC, pore volume, and BCR fractions serving as the predictors. Two response variables were kept: leachate concentration and cumulative migration percentage. The XGBoost algorithm served as the main nonlinear prediction model, with Random Forest and support vector regression as baseline models. The objective function in Equation (3) was minimized during the training phase.

$$\mathcal{L}^{(t)}=\sum_{i=1}^{n} l\left(y_i,\hat{y}_i^{(t-1)}+f_t\left(x_i\right)\right)+\Omega\left(f_t\right) \tag{3}$$

Where l is the loss function, f_t is the tree added at iteration t, and Ω penalizes model complexity. Hyperparameters were tuned within the training folds, while the five-fold group-wise cross-validation was applied with full column identity as the group, which ensures that no data point from one column gets into both the training and testing sets [11]. Prediction results were analyzed using R^2, RMSE, and MAE. Afterward, SHAP values were determined for every XGBoost prediction using Equation (4).

$$f(x)=\phi_0+\sum_{j=1}^{p}\phi_j \tag{4}$$

Where f(x) is the predicted migration response, ϕ_0 is the baseline prediction, and ϕ_j is the contribution of predictor j. Mean absolute SHAP values ranked global importance, while dependence patterns were examined for thresholds associated with increasing or decreasing migration.

4. Results

4.1. DOM-induced speciation and migration responses

Contrasting effects of high concentrations of DOM source were seen in operational speciation and cumulative leaching of metals, as reported in Table 1. The concentration of Cd acid-soluble fraction went up to 61.3% from 48.2% of the control, with cumulative Cd migration increasing to 31.4% from 18.7%. The concentration of oxidizable Cu fraction was reduced from 37.6% to 31.8% in the former, with that of oxidizable Pb fraction decreasing from 28.4% to 23.7%. Migration of the two metals increased to 12.8% and 7.1%, respectively. Under high concentration of humic-rich DOM, Cd acid-soluble fraction was reduced to 39.5%, with oxidizable Cu and Pb fractions increasing to 46.8% and 38.1%, respectively. Their cumulative migration decreased to 4.1% for Cu and 2.0% for Pb, while Cd migration decreased to 11.6%. This shows that mobilization of Cd fraction and leaching were simultaneously increased by straw-derived DOM, while oxidizable fractions were increased under humic-rich DOM.

DOI: 10.54254/2755-2721/2027.38839

Table 1. Heavy-metal speciation and cumulative migration under DOM treatments

Treatment	Cd acid-soluble fraction (%)	Cu oxidizable fraction (%)	Pb oxidizable fraction (%)	Cd migration (%)	Cu migration (%)	Pb migration (%)
Control	48.2	37.6	28.4	18.7	7.2	3.8
Low straw-derived DOM	54.6	34.1	25.9	24.9	9.6	5.3
High straw-derived DOM	61.3	31.8	23.7	31.4	12.8	7.1
Low humic-rich DOM	44.7	41.2	32.6	15.2	5.8	2.9
High humic-rich DOM	39.5	46.8	38.1	11.6	4.1	2.0

4.2. Migration prediction and key variable identification

XGBoost produced the strongest grouped-validation performance, with R^2 values of 0.93 for Cd, 0.90 for Cu, and 0.88 for Pb cumulative migration; corresponding RMSE values were 1.62, 0.73, and 0.41 percentage points. Random Forest achieved R^2 values of 0.86, 0.83, and 0.79, while SVR reached 0.78, 0.74, and 0.71. Figure 1 shows the mean absolute SHAP contributions for the principal predictors. Cd migration was dominated by the acid-soluble fraction at 4.10 percentage points and DOC at 3.62, followed by pH at 2.75. Cu was most sensitive to the humic-like PARAFAC component at 2.48 and SUVA254 at 2.16, whereas Pb showed the largest contributions from the humic-like component at 1.54 and SUVA254 at 1.31. These patterns indicate that Cd transport depended more strongly on labile speciation, while Cu and Pb migration was governed primarily by DOM aromaticity and humified binding capacity.

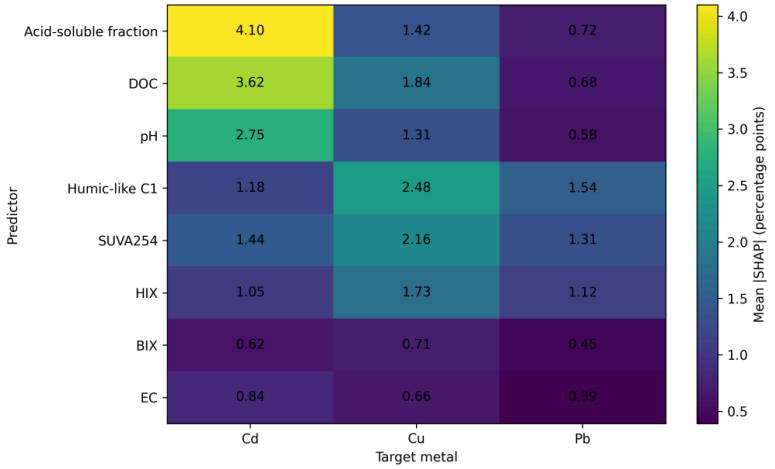


Figure 1. Mean absolute SHAP contributions to cumulative metal migration

DOI: 10.54254/2755-2721/2027.38839

5. Discussion

The opposing influences of straw-derived DOM and humic-dominated DOM on the fate of metals result from competition between mobilization facilitated by soluble ligands and sorption by strongly adsorbing organic complexes. The larger share of acid-soluble Cd fraction in the presence of high straw-derived DOM is related to its lower sorption capacity compared with Cu and Pb. Therefore, mobilization can be facilitated by soluble ligands and competing surface processes. Cu and Pb showed higher responses to SUVA254 and humic-like PARAFAC groups. Thus, less migration in the presence of humic-rich DOM points to the higher sorption capacity of the complexes of DOM with metals. SHAP thresholds must be understood as transition ranges supported by modeling since pH, DOC, and fluorescence composition change together in the course of elution through the column. XGBoost better accounted for non-linear dependencies, and grouped validation restricted the overestimation due to repeated observations. Migration of metals is defined not only by the quantity of DOM but also by the quality of DOM and initial metal fractions.

6. Conclusion

The source and concentration of DOM determined migration of cadmium, copper, and lead by altering their dissolution and operational speciation. DOM extracted from straw favored migration by increasing their mobile fractions and leaching potential, wherein high DOM treatment resulted in 31.4% of Cd migration, while DOM rich in humic composition increased their oxidizable fractions and limited migration. XGBoost outperformed Random Forest and SVR in predicting the migration behavior of the heavy metals, and SHAP indicated that Cd is mainly associated with acid-soluble fraction and DOC, while Cu and Pb correlated with humic-like fluorescence and aromaticity. The laboratory setup involved single soil type and controlled DOM additions, which limits direct application in field conditions.

References

[1] Wang, Z., Tang, W., Ding, X., Dong, Q., Guo, Y., Liu, G., Liu, Y., Liang, Y., Yin, Y., Cai, Y., & Jiang, G. (2025). Different extractable pools of Cd and Pb in agricultural soil under amendments: Water-soluble concentration sensitively indicates metal availability. *Journal of Environmental Sciences, 150*, 297–308.

[2] Tao, M., Ke, X., Ma, J., Liu, L., Qiu, Y., Hu, Z., & Liu, F. (2025). Dissolved organic matter (DOM)-driven variations of cadmium mobility and bioavailability in waterlogged paddy soil. *Journal of Hazardous Materials, 492*, 138065.

[3] Kölbl, A., Mosley, L., Fitzpatrick, R. W., & Kaiser, K. (2024). Sorption retards remediation of clayey sulfuric soils with straw-derived dissolved organic matter. *European Journal of Soil Science, 75*(5), e13574.

[4] Li, W., Lu, L., & Du, H. (2024). Deciphering DOM-metal binding using EEM-PARAFAC: Mechanisms, challenges, and perspectives. *Environmental Science and Pollution Research, 31*(10), 14388–14405.

[5] Firmino, F. H. T., Ker, J. C., Fontes, M. P. F., de Mello, J. W. V., de Melo, H. F., & Mesquita, L. F. (2025). Efficiency of sequential extraction schemes in partitioning toxic elements in acid sulfate soils with a thionic horizon. *European Journal of Soil Science, 76*(2), e70090.

[6] Zhou, X., & Cao, H. (2024). Effect of common ions in agricultural additives on the retention of Cd, Cu, and Cr in farmland soils. *Sustainability, 16*(11), 4870.

[7] Ren, H., Shen, X., Shen, D., Wang, K., Jiang, X., & Qadeer, A. (2024). Regional differences in lead (Pb) and tetracycline (TC) binding behavior of sediment dissolved organic matter (SDOM): Effects of DOM heterogeneity and microbial degradation. *Journal of Hazardous Materials, 474*, 134785.

[8] Yang, Z., Xia, H., Guo, Z., Xie, Y., Liao, Q., Yang, W., Li, Q., Dong, C., & Si, M. (2024). Development and application of machine learning models for prediction of soil available cadmium based on soil properties and climate features. *Environmental Pollution, 355*, 124148.

DOI: 10.54254/2755-2721/2027.38839

[9] Yan, Y., & Yang, Y. (2025). Revealing the synergistic spatial effects in soil heavy metal pollution with explainable machine learning models. *Journal of Hazardous Materials, 482*, 136578.
[10] Cui, S., Liu, L., Zhang, F., Fu, Q., Ma, C., & Ding, Y. (2025). Compositional evolution of dissolved organic matter mobilized by straw incorporation and its climate-driven interactions with lead in cold-region black soil: Decoding mechanisms through PARAFAC and complexation modeling. *Carbon Research, 4*, 56.
[11] Proshad, R., Asha, S. M. A. A., Tan, R., Lu, Y., Abedin, M. A., Ding, Z., Zhang, S., Li, Z., Chen, G., & Zhao, Z. (2025). Machine learning models with innovative outlier detection techniques for predicting heavy metal contamination in soils. *Journal of Hazardous Materials, 481*, 136536.