LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-852474

M4SF-23LL010301062-Surface Complexation/Ion Exchange Hybrid Model for Radionuclide Sorption to Clay Minerals

M. Zavarin, E. Chang, S. Han, H. Wainwright

July 31, 2023

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

July 21, 2023

M4SF-23LL010301062-Surface Complexation/Ion Exchange Hybrid Model for Radionuclide Sorption to Clay Minerals

M. Zavarin[1], Elliot Chang[1], Sol-Chan Han[1], Haruko Wainwright[2]

[1] Glenn T. Seaborg Institute, Physical & Life Sciences, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA.
[2] Massachusetts Institute of Technology, Cambridge, MA

DISCLAIMER

Contents

1. Introduction....1
2. Status of L-SCIE Sorption Database....2
3. A Chemistry-Informed Hybrid Machine Learning Approach to Predict Metal Adsorption onto Mineral Surfaces....3
4. Aqueous Chemistry and Physics Constrained Autonomous-Closed-Loop Modeling for Quantifying Adsorption Processes as Applied to Metal-Mineral Interface Geochemistry....5
5. Planned FY24 Efforts....6
6. Acknowledgments....7
7. References....7

1. Introduction

This progress report (Level 4 Milestone Number M4SF-23LL010301062) summarizes research conducted at Lawrence Livermore National Laboratory (LLNL) within the Argillite International Collaborations Activity Number SF-23LL01030106. The activity is focused on our long-term commitment to engaging our partners in international nuclear waste repository research. The focus of this milestone is the establishment of international collaborations for surface complexation modeling and the associated impacts of unlocking larger, community-based datasets. More specifically, we are developing a database framework for Spent Fuel and Waste and Science Technology (SFWST) that is aligned with the Helmholtz Zentrum Dresden Rossendorf (HZDR) sorption database development group in support of the database needs of the SFWST program.

In our FY22 effort, we described a detailed analysis of U(VI) sorption to quartz through both traditional surface complexation modeling and through a hybrid ML framework. In FY23, effort was placed on publication of these results and expansion of the LLNL surface complexation and ion exchange database (L-SCIE) in order to assess mineral-based radionuclide retardation under a wider variety of geochemical conditions (e.g., ionic strength, varying electrolyte compositions). Efforts were initiated to expand L-SCIE to include radionuclide surface complexation and ion exchange to clays that are relevant to subsurface geochemical processes occurring at nuclear waste repositories. In particular, a large source of sorption data for clays resides at the Paul Scherrer Institute (PSI) (work primarily by Bradbury and Baeyens) and we initiated discussions on how to retrieve those data and apply FAIR principles to those datasets.

In addition to L-SCIE development, two hybrid models that incorporate AI/ML were investigated and compared to discern the most promising approaches for accurate and precise estimations of radionuclide retardation. Key considerations for future model development include (1) the ability to reduce computational burden on determining retardation coefficients for PA and (2) the ability to quantify and predict radionuclide-mineral partitioning at a more efficient, rapid pace due to automated workflows. Upon the careful consideration of the most effective modeling approaches, we are identifying ways to implement these approaches into PA.

Ultimately, the data science-based workflows will provide a major incentive for other institutions to adopt a FAIR-formatted, interoperable database. LLNL will play a key role in disseminating sorption data and acting as good data stewards by updating the database in a consistent format and assessing the quality of the newly assimilated data in an organized fashion. To this end, all data and workflows are open access and made available on the LLNL Seaborg research website (https://seaborg.llnl.gov/resources/geochemical-databases-modeling-codes).

2. Status of L-SCIE Sorption Database

To develop a comprehensive surface complexation database in support of the SFWST program, we continued to build a digital sorption database to support surface complexation database development. The LLNL Surface Complexation/Ion Exchange (L-SCIE) digital sorption database and workflow was developed in Microsoft Access with a series of linked tables as reported previously. The structure of the database was recently reported in Zavarin et al. (2022) and will not be reported here. The efforts in FY23 focused on applying the L-SCIE database to a number of problems rather than expanding the content of the database. In this capacity, various debugging efforts and database enhancements were performed. While the longterm goal is to provide all database and associated workflows in a web-accessible format, all materials associated with this project are presently made available by request to the authors. The materials made available include the following:

- Database: Contains the full access database and all linked files (zipped) that are necessary to fully implement the L-SCIE database. To work the full database, you should unzip the linked files and place them in the same directory as the access database.

- How-to videos: These videos explain the process of digitizing data from the literature and loading them into the L-SCIE database. They provide step-by-step instructions for selecting literature references, extracting figures and metadata, extracting data from figures using the DataThief software, and importing data into the L-SCIE database.

- L-SCIE unification code: This folder contains the R code (two files: Server.r and ui.r) that are used to run the unification code. The code is run using R Studio. The code takes the exported .csv files from the L-SCIE database (dataset.csv and data.csv) and performs numerous unit conversions, formatting, and other processes to "unify" all the data in the database (e.g. takes Kd data and converts it all to aqueous and sorbed concentrations). This yields a file called sc.dataset.csv: the file that contains all the sorption data in the database in a uniform format. You can also use the L-SCIE code to produce a subset file (e.g. extract all Sr-goethite sorption data) called sc.subset.csv. The L-SCIE code can also be used to produce input files for use in phreeqc/pest parameter estimation processes.

Our L-SCIE digital sorption database **includes 246 references, 2331 datasets, and 27,000 individual data points.** The database is also linked to a large Kd database available from the JAEA. A python code automates the import of JAEA Kd data into the L-SCIE workflow increased our total data holdings to 44,000 data points. Details regarding the L-SCIE database and associated workflows were reported in the FY22 annual reports and will not be repeated here.

In support of the Argillite work package, we began focusing on clay sorption data available in the literature. At present, the L-SCIE database, when combined with the

JAEA database, contains approximately 16,000 data for radionuclide sorption to clay minerals. Discussions with Maria Marquez and Dan Miron (PSI) have been continuing since January, 2023, to establish a mechanism to share the vast clay data available at PSI. However, these data are not readily available at this time and require reworking into a digital format. We are assisting PSI in this process to ensure compatibility between the L-SCIE and PSI databases. Meanwhile, we are moving forward with digitizing other clay sorption data that are identified in the RES3T database (Helmholtz-Zentrum Dresden-Rossendorf, 2022). In total, this database contains 187 unique references to radionuclide sorption to bentonite, biotite, illite, kaolinite, montmorillonite, muscovite, smectite, nontronite, and beidellite clay. This is the starting point for our clay data digitization effort that will be initiated in FY24. We anticipate that a significant fraction of these data are from PSI and access to that digital resource will significantly reduce the burden of digitizing these datasets.

3. A Chemistry-Informed Hybrid Machine Learning Approach to Predict Metal Adsorption onto Mineral Surfaces

Chang, E., Zavarin, M., Beverly, L., and Wainwright, H. 2023. A Chemistry-Informed Hybrid Machine Learning Approach to Predict Metal Adsorption onto Mineral Surfaces, Applied Geochemistry, 155: 105731, https://doi.org/10.1016/j.apgeochem.2023.105731.

Zavarin, M., Chang, E., Beverly, L., Wainwright, H. 2023. A Data Mining and Hybrid Machine Learning Approach to Assess Mineral Retardation of Radionuclides, WM2023, February 26-March 2, 2023, Phoenix, AZ.

The information presented below is a summary from a manuscript published in June 2023 in Applied Geochemistry (Chang et al., 2023). The publication describes our L-SCIE sorption database (Zavarin et al., 2022) and the hybrid workflow used to fit sorption behavior using these data. Importantly, we show that the hybrid approach is sufficiently flexible to be applied to surface complexation and ion exchange dominated systems. This indicates that the approach, initially developed using a U(VI)-quartz test case, can be expanded to simulation of ion exchange on clays. This is tested for the case of U(VI) sorption to montmorillonite. The results were also published in the proceedings of the 2023 Waste Management conference (Zavarin et al., 2023).

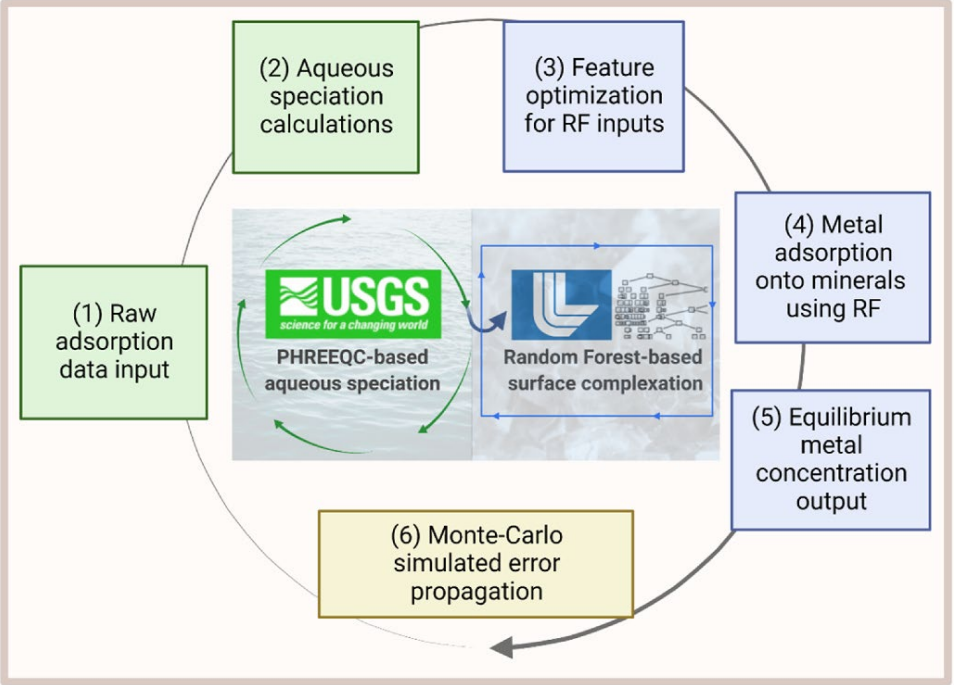

Figure1. L-SURF workflow chart with chronological steps: (1) Adsorption data and selected thermodynamic database are imported into L-SURF module, (2) Aqueous speciation calculations are conducted and all geochemical features are output, (3) Choice of most impactful geochemical features and hyperparameters are optimized, (4) Optimal features are used to train and test a random forest adsorption model, (5) Equilibrium aqueous metal sorbate concentrations are output, and (6) Steps 1–5 are repeated using Monte-Carlo simulations with randomly sampled input data ± experimentally-determined measurement uncertainty.

Historically, surface complexation model (SCM) constants and distribution coefficients (Kd) have been employed to quantify mineral-based retardation effects controlling the fate of metals in subsurface geologic systems. Our recent SCM development workflow, based on the L-SCIE database, illustrated a community FAIR data approach to SCM development by predicting uranium(VI)-quartz adsorption for a large number of literature-mined data. In this manuscript, we present an alternative hybrid machine learning (ML) approach that shows promise in achieving equivalent high-quality predictions compared to traditional surface complexation models. At its core, the hybrid random forest (RF) ML approach is motivated by the proliferation of incongruent SCMs in the literature that limit their applicability in reactive transport models. Our hybrid ML approach implements PHREEQC-based aqueous speciation calculations; values from these simulations are automatically used as input features for a random forest (RF) algorithm to quantify adsorption and avoid SCM modeling constraints entirely. Named the LLNL Speciation Updated Random Forest (L-SURF) model, this hybrid approach is shown to have applicability to U(VI) sorption cases driven by both ion-exchange and surface complexation, as is shown for quartz and montmorillonite cases. The approach can be applied to reactive transport modeling and may provide an alternative to the costly development of self-consistent SCM reaction databases.

The L-SURF code is available at https://ipo.llnl.gov/technologies/software/l-surf

The U(VI)-quartz dataset used in testing the workflow can be found at: https://seaborg.llnl.gov/resources/geochemical-databases-modeling-codes

4. Aqueous Chemistry and Physics Constrained Autonomous-Closed-Loop Modeling for Quantifying Adsorption Processes as Applied to Metal-Mineral Interface Geochemistry

Chang, E. Beverly, L., and Wainwright, H. Submitted. Constrained Autonomous Modeling of Metal-Mineral Adsorption, *In* Methods and Applications of Autonomous Experimentation, Marcus Michael Noack and Daniela Ushizima, Eds.

This manuscript was submitted as a book chapter for inclusion in Methods and Applications of Autonomous Experimentation, edited by Marcus Noack and Daniela Ushizima. The book chapter was prepared to illustrate the use of high-throughput, "big data" modeling approaches to investigate fluid-mineral interactions in subsurface geochemical processes. To this extent, working data in a FAIR formatted structure have been compiled using the Lawrence Livermore National Laboratory Surface Complexation/Ion Exchange (L-SCIE) database (Zavarin et al., 2022). Utilizing this consistently formatted set of data, the authors pose a test case using the mobility of the U(VI) exposed to quartz. The chapter illustrates the application of Gaussian process (GP) modeling that is informed by chemical aqueous speciation to develop predictive capabilities. A workflow is outlined based on autonomous optimization of the GP regressor, constrained by physics information - in this case mass conservation and statistical knowledge of high-variance regions. This approach demonstrates how chemistry and physics can be incorporated in the generation of a fine-tuned adsorption model through an autonomous, closed-loop process.

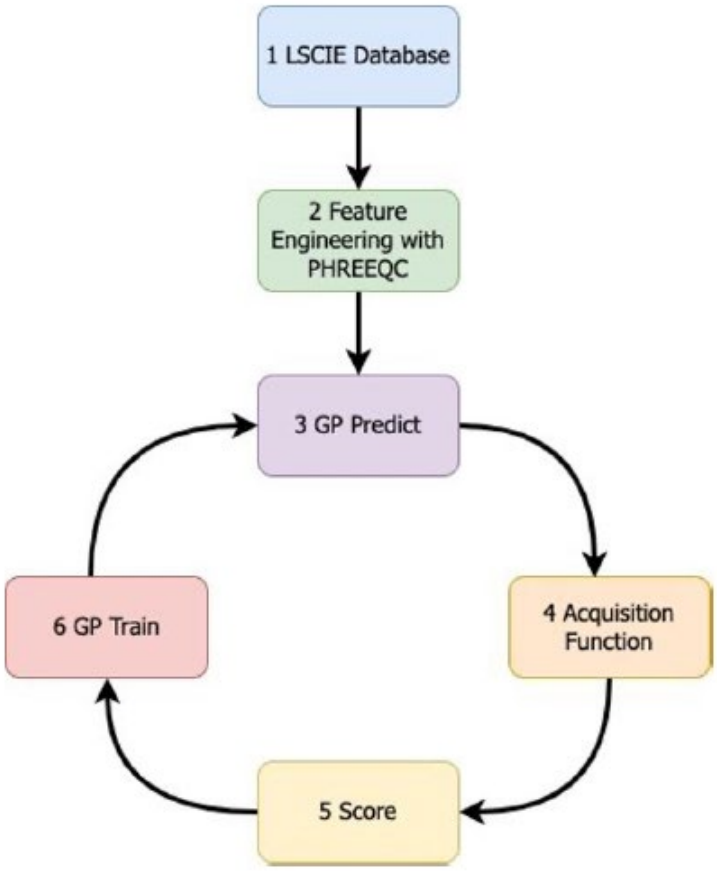

Figure 2. Autonomous experimentation loop for virtual instrument optimization

5. Planned FY24 Efforts

In FY24, effort will be placed on the growth of the surface complexation and ion exchange database (L-SCIE) with a particular focus on clay minerals relevant to argillite rock. First, we will assist our collaborators at PSI to digitize their clay sorption data in a format consistent with our L-SCIE database. Second, these data will be supplemented by sorption reference contained in the RES^{3}T database and the data already captured from the JAEA Kd database. Once these datasets are digitized and unified, we will begin developing high throughput automated workflows to model sorption of a series of radionuclides and clays. This effort will move us from workflows that focus on single radionuclide-mineral pairs to more comprehensive modeling workflows. This effort will include the testing of L-SURF to quantify K_d values for numerous different radionuclide-clay pairs under varying geochemical conditions. We will conduct direct comparisons of the L-SURF approach with various surface complexation (Non-electrostatic, diffuse layer, etc.) and ion exchange (Vanselow, Gapon, etc.) models. These important model inter-comparisons will provide a clearer path forward in incorporating traditional SCM or/and modern hybrid ML approaches into repository flow and transport models.

Key considerations for future modeling development will include (1) the ability to reduce computational burden on determining retardation coefficients and (2) a new capability to quantify and predict radionuclide-mineral partitioning at a more efficient, rapid pace due to automated workflows. We will begin discussing efforts for the integration of our new surface complexation/ion exchange constants or hybrid-ML models into PFLOTRAN.

In addition to the modeling effort described above, we plan to upgrade our L-SCIE database to make it web-accessible. This will require converting the database that is presently supported as a Microsoft Access database to a more portable SQL format that is open access and readily shared with the scientific community, following the FAIR data principles (Wilkinson et al., 2016).

6. Acknowledgments

This work was supported by the Spent Fuel and Waste Science and Technology campaign of the Department of Energy's Nuclear Energy Program. Prepared by LLNL under Contract DE-AC52-07NA27344.

7. References

Chang, E., Zavarin, M., Beverly, L., Wainwright, H., 2023. A chemistry-informed hybrid machine learning approach to predict metal adsorption onto mineral surfaces. Applied Geochemistry 155, 105731. doi.org/https://doi.org/10.1016/j.apgeochem.2023.105731.

Helmholtz-Zentrum Dresden-Rossendorf, 2022. RES³T-Rossendorf Expert System for Surface and Sorption Thermodynamics. www.hzdr.de/res3t. 2022).

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3(1), 160018. doi.org/10.1038/sdata.2016.18.

Zavarin, M., Chang, E., Beverly, L., Wainwright, H., 2023. A Data Mining and Hybrid Machine Learning Approach to Assess Mineral Retardation of Radionuclides, WM2023. Phoenix, AZ.

Zavarin, M., Chang, E., Wainwright, H., Parham, N., Kaukuntla, R., Zouabe, J., Deinhart, A., Genetti, V., Shipman, S., Bok, F., Brendler, V., 2022. Community Data Mining Approach for Surface Complexation Database Development. Environmental Science & Technology 56(4), 2827-2838. doi.org/10.1021/acs.est.1c07109.