

Lawrence Berkeley National Laboratory

LBL Publications

Title

Machine learning surrogates for surface complexation model of uranium sorption to oxides

Permalink

<https://escholarship.org/uc/item/2s01979v>

Journal

Scientific Reports, 14(1)

ISSN

2045-2322

Authors

Li, Chunhui
Adeniyi, Elijah O
Zarzycki, Piotr

Publication Date

2024

DOI

10.1038/s41598-024-57026-w

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



OPEN

Machine learning surrogates for surface complexation model of uranium sorption to oxides

Chunhui Li¹✉, Elijah O. Adeniyi²✉ & Piotr Zarzycki¹

The safety assessments of the geological storage of spent nuclear fuel require understanding the underground radionuclide mobility in case of a leakage from multi-barrier canisters. Uranium, the most common radionuclide in non-reprocessed spent nuclear fuels, is immobile in reduced form (U(IV)) and highly mobile in an oxidized state (U(VI)). The latter form is considered one of the most dangerous environmental threats in the safety assessments of spent nuclear fuel repositories. The sorption of uranium to mineral surfaces surrounding the repository limits their mobility. We quantify uranium sorption using surface complexation models (SCMs). Unfortunately, numerical SCM solvers often encounter convergence problems due to the complex nature of convoluted equations and correlations between model parameters. This study explored two machine learning surrogates for the 2-pK Triple Layer Model of uranium retention by oxide surfaces if released as U(IV) in the oxidizing conditions: random forest regressor and deep neural networks. Our surrogate models, particularly DNN, accurately reproduce SCM model predictions at a fraction of the computational cost without any convergence issues. The safety assessment of spent fuel repositories, specifically the migration of leaked radioactive waste, will benefit from having ultrafast AI/ML surrogates for the computationally expensive sorption models that can be easily incorporated into larger-scale contaminant migration models. One such model is presented here.

Development of sustainable, low-carbon energy production technologies is of paramount importance due to growing energy demand in the realm of the global warming^{1,2}. Nuclear power emerges as a promising solution, offering stable and clean energy while minimizing CO₂ emissions to mitigate climate change^{3,4}. However, the safe treatment and disposal of nuclear waste remain a challenge^{5,6}.

Geological repositories are considered to store spent nuclear fuel^{7,8}. This approach involves multi-barrier isolation of the radioactive waste, with the host rock formation being the last barrier^{9,10}. The safety assessment of such repositories requires knowledge of the mobility of radionuclides that could be released to the surrounding environment (e.g., nuclear waste leakage, Fig. 1a) and upward transport to groundwater and soils remain a concern^{11–13}.

To assess the safety of the spent nuclear waste disposal sites in case of leakage, one must quantify the strength and extent of radionuclides retention by mineral/soil matrix. This study focused on uranium because it is the most abundant radionuclide in non-reprocessed spent nuclear fuel and is highly mobile if present in the oxidized form. Most of the repositories are planned deep underground, where reducing conditions prevail. In these conditions, uranium is present in a reduced U(IV) form and is strongly attracted to negatively charged organic and inorganic surfaces; thus, it is immobile. In the event of a leakage, released U(IV) could be retained within the leakage zone via sorption to mineral surfaces^{14,15}. Particles with adsorbed U(IV) could be transported upward to oxic conditions^{16–18}. Oxidized uranium, U(VI), remains stable in solution and is considered an environmental threat if released into groundwater¹⁹.

Our work refers to a hypothetical leakage of spent nuclear fuel from multibarrier containers, followed by its upward transport to the subsurface, where oxidation reactions mobilize uranium transport by transforming immobile uranium (IV) into mobile uranium (VI) (Fig. 1a). We consider a redox equilibrium between reduced and oxidized uranium but assume a positive redox potential ($pe = 0.5$). Consequently, uranium (VI) is a dominant oxidation state, and U is present as UO_2^{2+} ion. However, if U(IV) is strongly adsorbed to the mineral surfaces, it may remain as U(IV), protected from oxidation in the solution to U(VI). In this study, we consider a broad range of U(IV) and U(VI) sorption affinities to account for the fact that uranium may remain as U(IV) adsorbed to the mineral surfaces despite oxidative conditions.

¹Energy Geosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ²Department of Earth Sciences, Montana State University, Bozeman, MT, USA. ✉email: chunhuili@lbl.gov; elijahadeniyi@montana.edu

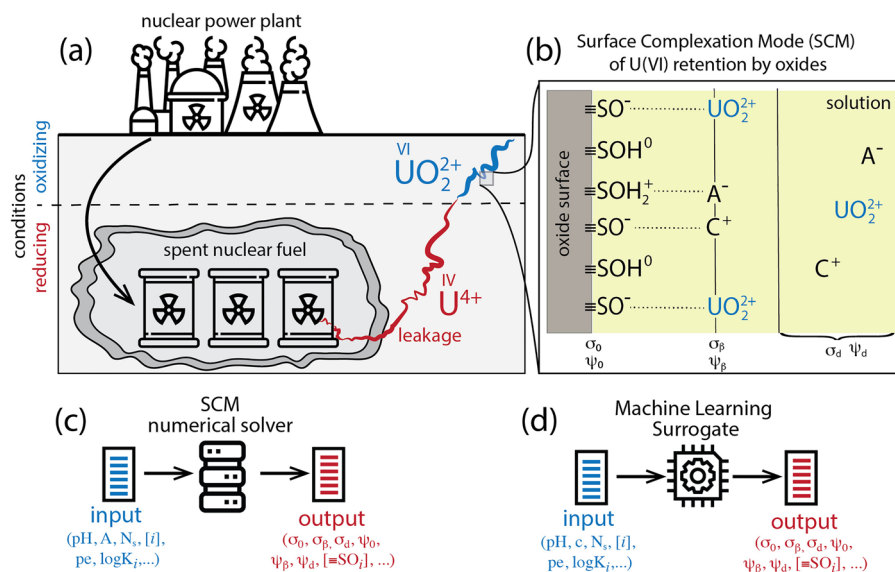


Figure 1. Schematic illustration of the hypothetical leakage of spent nuclear fuel from geological storage, mobilization of uranium (U) in oxidizing subsurface conditions, and retention by sorption to oxides minerals (a). In panel (b), we show a schematic diagram of the 2-pK Triple Layer Model (TLM) of the oxide/U-containing solution interface: 0-plane at the mineral surface, d-plane at diffuse layer and β -plane is the layer where electrolyte ions adsorb non-specifically. C^+ and A^- represent the electrolyte cations and anions (i.e., Na^+ , Cl^-). In panels (c, d), we show the numerical approach to predict thermodynamic properties of the mineral/electrolyte interface and the extent of ion sorption (c) and its Machine Learning surrogate - described in this report (d). The input contains a description of system conditions such as pH, ions concentrations ($[i]$, $i = \text{C}^+$, U), surface area (A), site density (N_s), and ion/proton affinity constants ($\log K_i$). The output contains charge densities and electrostatic potentials at the surface, β and diffuse layers ($\sigma_0, \sigma_\beta, \sigma_d, \psi_0, \psi_\beta, \psi_d$) and concentration of various types of surface species ($[\equiv\text{SO}_i]$). Despite oxidizing conditions, we also consider a special case in which U(IV) is adsorbed strongly to the surface, and thus, it is protected against oxidation in the solution.

Uranium mobility and environmental toxicity in the event of a compromised nuclear waste repository depend on its interactions with the surrounding soil/rock and their constituents. The key environmental descriptors controlling uranium sorption are pH, redox conditions, salinity, and presence of complexing ligands/mineral surfaces^{14,15,20}. The retention by mineral surfaces can be calculated using the interfacial speciation models known as surface complexation models (SCMs)^{20–23}. One of the challenges of applying SCM models to estimate uranium retention by mineral surfaces is the complexity of convoluted and coupled equations that need to be numerically solved^{20,24}. This translates to convergence problems and computational costs²⁵, requiring inputs from domain expertise to resolve the issue.

Recently, we demonstrated the effectiveness of applying traditional machine learning methods while developing surface composite models (SCMs)²⁶. In particular, we employed an ensemble of random forest regressors to accurately predict SCM parameters, simplifying and enhancing the SCM development process. In this study, we curated extensive datasets, comprising more than one million data points, covering diverse conditions of uranium sorption on oxides using advanced numerical SCM solvers. Recognizing the advantages of deep learning for large datasets, in addition to traditional machine learning methods, we explored a deep neural network model. We aim to create a cost-effective yet accurate surrogate model with minimal human input for predicting uranium sorption to oxides across a broad range of surface conditions. This effort contributes to the safety assessment of spent fuel repositories by introducing a class of artificial intelligence/machine learning (AI/ML) surrogates capable of predicting uranium retention for a broader range of subsurface conditions at a fraction of the computational cost compared to analytical models.

Results

We generated a comprehensive dataset comprising 1,589,411 entries via a traditional SCM solver²⁷. Each entry consists of 12 system descriptors: pH, ions concentrations ($[i]$, where $i = \text{C}^+$, U), surface area (A), site density (N_s), and ion/proton affinity constants ($\log K_i$), which will be used as inputs to the surrogate model (see Table S3, Supporting Information). The information of charge densities and electrostatic potentials at the surface, β and diffuse layers, and concentration of various surface species ($[\equiv\text{SO}_i]$) were then extracted and used as target values when developing uranium sorption surrogate models. Subsequently, we developed two uranium sorption surrogate models: a multioutput regressor composed of an ensemble of random forests (RF-surrogate) and a deep neural network model constructed from six fully connected layers (DNN-surrogate). Both surrogate models have the same input parameter format with 12 system descriptors as the numerical SCM solver in generating the dataset.

We then trained both models on 1,271,528 data. The performance of RF-surrogate on predicting uranium sorption to oxide surfaces was evaluated by simultaneously testing the remaining 317883 conditions (input entries). For the Deep Neural Network surrogate model (DNN-surrogate), 158,941 data points from the remaining data were allocated for hyperparameter tuning and to mitigate overfitting. The model performance was then assessed on the remaining 158,942 conditions. A more detailed description of the model construction and dataset can be found in the “Methods” section.

Figure 2 is the parity plot of DNN-surrogate, which compares the charge densities, electrostatic potentials at different layers, and surface adsorption from numerical SCM solver to machine-learned results under diverse conditions. The overall R^2 score of RF-surrogate consistently exceeds 0.999 for all targets (see Supplementary Information, Fig. S1), whereas DNN-surrogate consistently achieves R^2 scores higher than 0.9999. Although RF-surrogate also received a remarkable value of R^2 . The DNN surrogate improves upon the RF surrogate, with fewer outliers and more minor prediction errors, especially for surface sorption and charge density/electrostatic potential at the diffuse layer. The maximum absolute error of normalized data between the machine-learned prediction and the numerical SCM solver’s solution is approximately 0.015 for DNN-surrogate and around 0.04 for RF-surrogate. Utilizing DNN-surrogate has resulted in a 62.5% improvement in accuracy. We also calculated each target’s mean squared error (MSE) for each target variable while preserving the original units of the data,

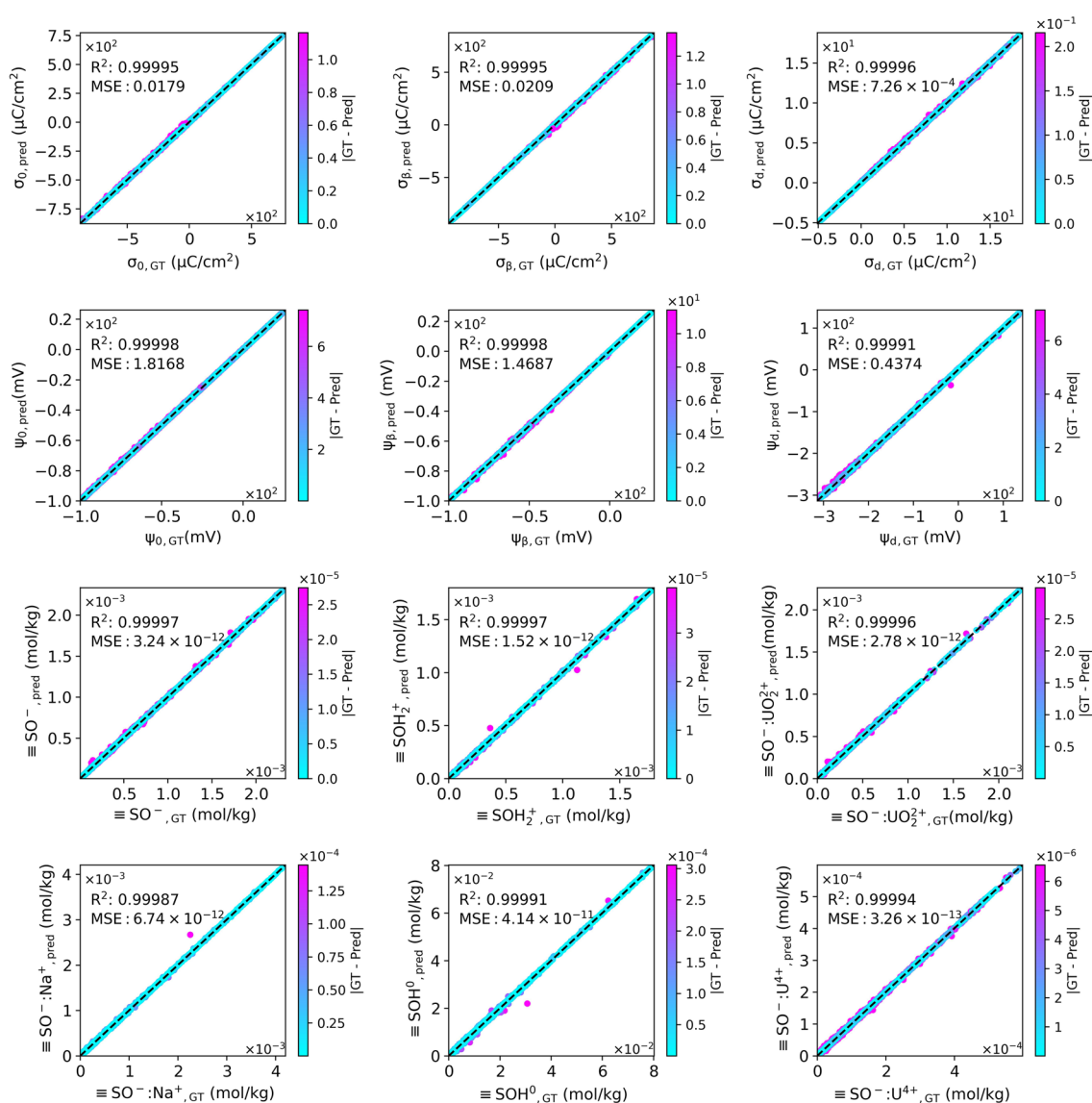


Figure 2. Performance of DNN-surrogate for uranium sorption on oxides. Parity plot for DNN-surrogate model prediction vs. numerically calculated ground truth using GWB. R^2 score represents the goodness of fit of surrogate predictions to corresponding ground truth values. MSE represents the mean squared error between prediction and target. The absolute error ($|GT - Pred|$) between DNN-surrogate prediction and the corresponding ground truth value is shown as the color bar.

as displayed in Fig. 2. Compared to RF-surrogate, DNN-surrogate decreases the mean squared error (MSE) by 56.64–89.2% for each target.

We conducted two additional tests to assess the surrogate model's performance further. We randomly generated two titration sets of input entries (see Table S3, Supporting Information) to compare the surrogate model's and numerical solver's predictions regarding pH-dependent charge densities and electrostatic potentials.

DNN-surrogate can produce identical results as numerical SCM solvers for charge densities and electrostatic potentials regardless of locations (Fig. 3, first set). On the other hand, RF-surrogate generates similar results for β layer charge density and diffuse layer potential (see Supplementary Information, Fig. S2), but fails to predict other targets accurately.

In the second test, we choose a set of parameters for which the numerical solver encounters convergence problems for pH above 5 (Fig. 4). In contrast, both surrogate models can produce the results for the whole pH range. Specifically, DNN-surrogate (but not RF-surrogate) consistently delivers reliable predictions for targets below and above the threshold of pH equal to 5. What is more, RF-surrogate cannot extrapolate when pH is beyond 5 (see Supplementary Information, Fig. S3).

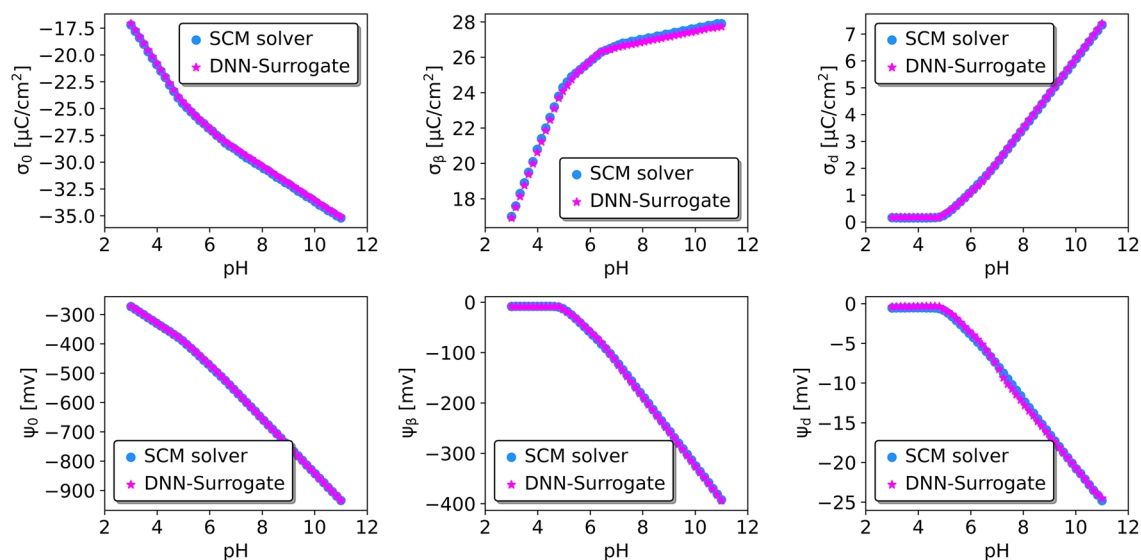


Figure 3. Comparison of charge densities and potentials at the surface, β , and diffuse layers predicted by DNN-surrogate's predictions and those obtained by fully converged numerical SCMs.

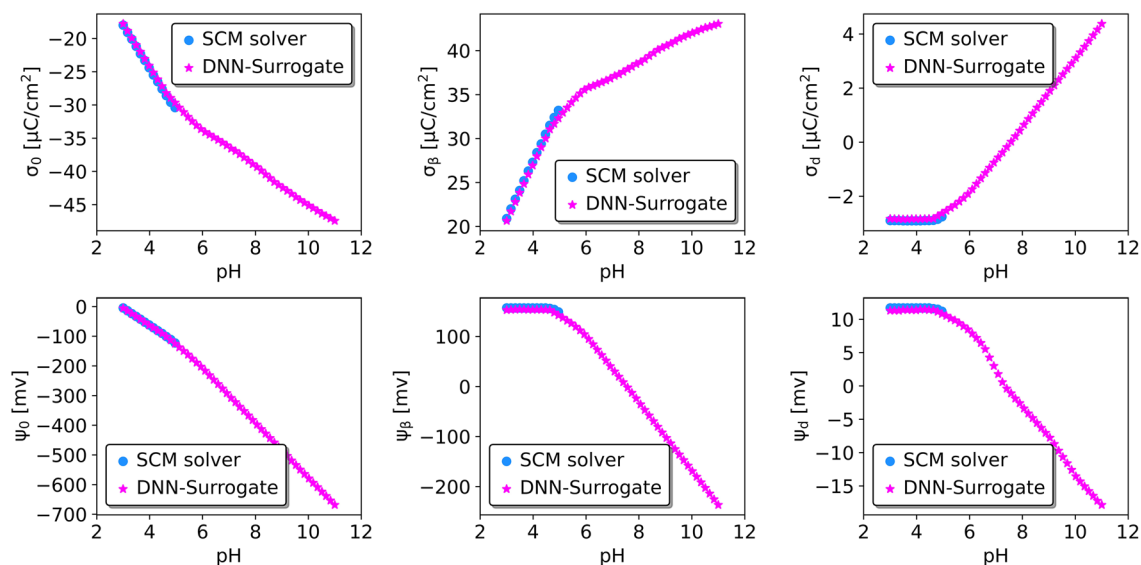


Figure 4. Comparison of charge densities and potentials at the surface, β , and diffuse layers predicted by DNN surrogate with those obtained by the numerical solver that encounters convergence issues. Note that DNN-surrogate can provide predictions for all pH values in contrast to numerical SCM.

To better understand how the DNN-surrogate model learned to map an input to output, we performed feature analysis by measuring how each input feature impacts each output value. Here, we implemented the integrated gradient algorithm²⁸ to calculate the feature attribution by assigning an importance score to each input feature via approximating the integral of the gradients of the model's output concerning the inputs.

For surface, β -layer, and diffuse layer charge densities, the top five crucial input features consistently include the uranium initial concentration in the solution, anions binding constant $\log K_a$, pH, protonation equilibrium constant $\log K_1$, and inner-layer capacitances c_1 . By definition, the surface charge is the sum of positive and negative charges at the surface, the charge at β -plane is determined by accumulated charges in β -plane, and charge density at the diffuse layer is highly correlated to the charge density at the surface and β -plane, can be calculated via $\sigma_0 + \sigma_\beta + \sigma_d = 0$ ^{29–33}. Therefore, it is reasonable for the DNN surrogate to discover that pH conditions, ion concentrations, and affinity constants that control the uptake and release of ions and protons are essential parameters for all charge density predictions.

The electrostatic potentials are coupled to the charge densities in the SCM. First, c_1 directly links the charge at the surface with the potential drop at a distance β with the equation: $\sigma_0 = c_1(\psi_0 - \psi_\beta)$ ³¹. Regarding feature importance, we anticipate similar patterns for ψ_0, ψ_β with surface charge density σ_0 . Analysis results reveal that c_1 is always the fifth important feature for predicting $\sigma_0, \psi_0, \psi_\beta$, and the top five most important features of potential near the surface (0 and β layers) are in agreement with surface charge density, having the most relevant features being pH, $[U^{4+}]$, affinity constants, and c_1 . For diffuse layer potential, it follows the relation: $\sigma_d = c_2(\psi_d - \psi_\beta)$. Therefore, it is not surprising that the importance of c_1 decreased. In literature²⁹, c_2 is often set to a constant value; hence, it is not considered in our dataset. Thus, an inconsistent feature importance between ψ_d and ψ_β, σ_d is expected.

The extent of accessible binding sites for surface sorption is directly governed by surface area. The analysis underscores the essential role of surface area as a critical feature for predicting surface reactions using DNN-surrogate. The analysis confirms the importance of affinity constants in accurately predicting surface reactions. For instance, equilibrium constants are key features in predicting proton/deprotonation processes and ion complexation reactions, and $\log K_{UO_2^{2+}}$ and $\log K_{U^{4+}}$ play crucial roles in predicting UO_2^{2+} and U^{4+} sorption to mineral surfaces, respectively (as expected based on assumed chemistry in the SCM construction, see “Methods” section).

SCM equations are highly non-linear, and often, solutions show a convoluted interdependence between parameter values and predicted properties. Here, we identify several such correlations (refer to Fig. S6). The SCM construction imposes coupling between the electrostatic properties at various layers. Here, by analyzing training datasets, we noticed strong correlations between diffusion layer charge (σ_d) and the β layer potential (ψ_β); between the surface charge (σ_0) and the β layer charge (σ_β). One can expect similar correlations in the results of the feature importance analysis of these targets. As illustrated in Fig. 5, σ_0 and σ_β exhibit nearly identical feature importance trends, differing only in the rank reversal of $\log K_1$ and c_1 . In the case of σ_d and ψ_β , they share the same top eight features and trends, while the remaining input features, $N_s, \log K_c, Na^+$, and $\log K_{U^{4+}}$, display distinct importance trends and are all less significant in accurately predicting the targets. However, the feature analysis of RF-surrogate (Fig. S6) reveals that only the initial four important features align for σ_d and ψ_β , and the fourth and last three input features are consistent for σ_0 and σ_β . Merely 4 of 12 input features display analogous importance in strongly correlated pairs. In conclusion, the DNN-surrogate successfully recognized the underlying physicochemical relationships from training data, whereas the RF-surrogate only partially captured these relationships.

Table S1 summarizes the computational efficiency and utilized computing resources for three approaches in estimating uranium sorption under over 300,000 conditions. AI surrogates exhibit exceptional speed, making simultaneous predictions for all 300,000 conditions in seconds: 27 s for DNN-surrogate and 86 s for RF-surrogate. In contrast, traditional numerical SCM solvers need to solve for these conditions sequentially, taking over 5 h.

Conclusion

The safety assessments of geological repositories for spent nuclear fuel require knowledge of the radionuclide transport and retention by surrounding rocks in case of a catastrophic event of leakage. If U(IV), a dominant form in reducing conditions, is transported upwards, it can be oxidized to a more mobile U(VI).

Here, we investigated two distinct AI surrogates for the SCM of uranium retention by oxides. We modeled various minerals varying in acid-base properties, site densities, surface area, and ion affinities for different ion concentrations (including uranium). We used the 2-pK TLM model of uranium sorption on the oxide/electrolyte interface, which was solved numerically by the GWB solver, to generate over a million entries for training AI models.

The DNN surrogate does not experience any convergence issues as compared to the numerical solver and RF surrogate, and it can make predictions at a fraction of the computational time required for the SCM solver. Moreover, DNN-surrogate outperformed RF-surrogate regarding the range of applicability and accuracy. The feature analysis showed that DNN-surrogate captured how uranium sorption depends on the system descriptors and model parameters—demonstrating that DNN learned the physicochemistry of SCM construct to some extent directly from the data.

Here, we showed how to construct a synthetic dataset and trained AI to mimic the 2-pK TLM model of uranium sorption. Our surrogates are the first step toward constructing an AI-based framework for modeling radionuclide subsurface migration that is essential in the safety assessment of geological nuclear waste repositories. More extensive and complex chemistries will be examined in the future.

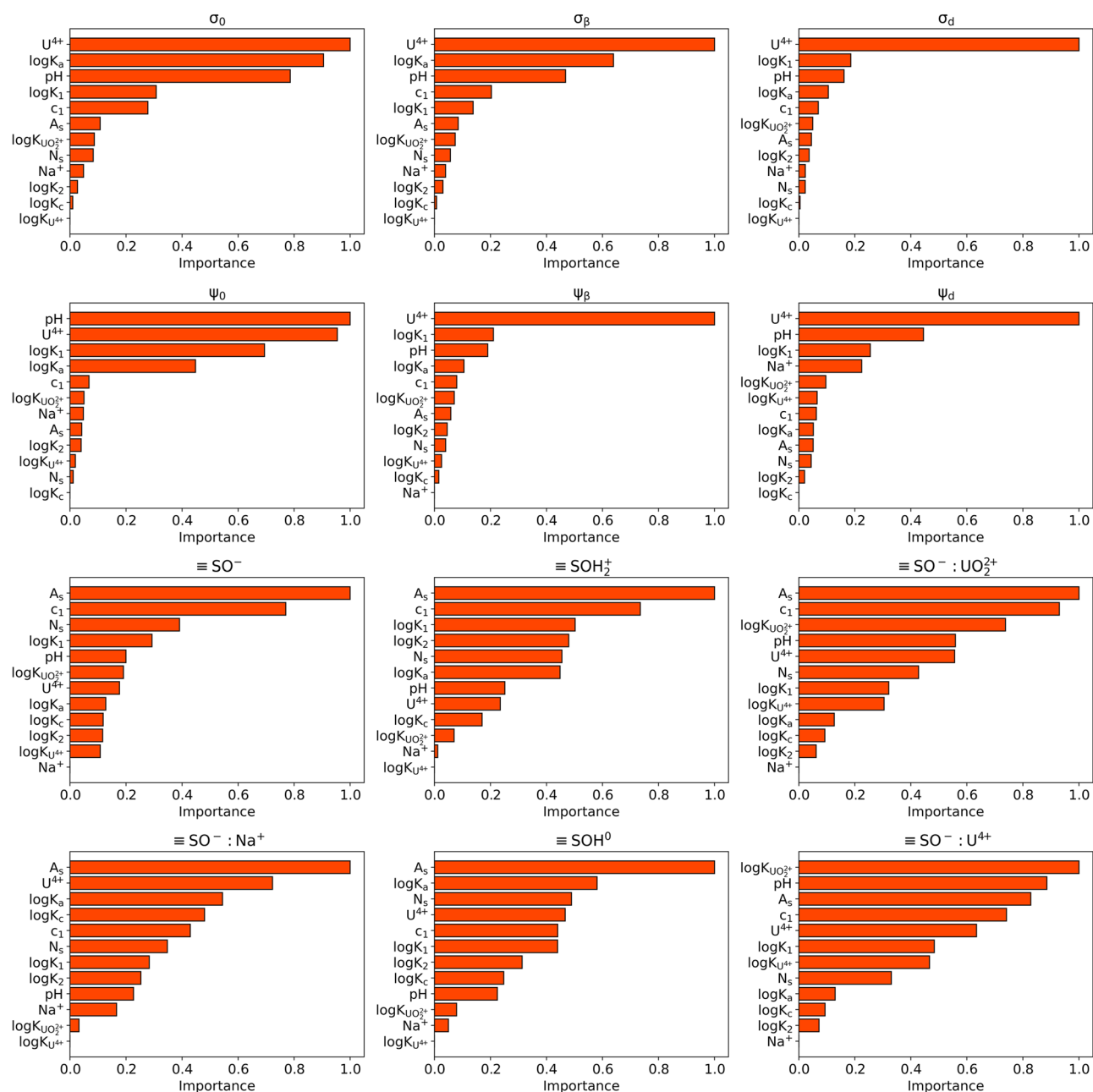


Figure 5. Analysis of input feature attribution in predicting target values for DNN-surrogate.

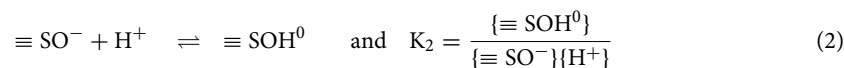
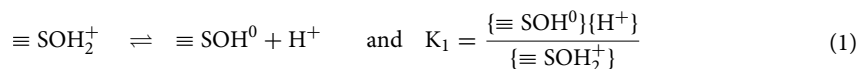
Methods

U(VI) SCM surrogate model development

The first step towards developing our U(VI) SCM surrogate is constructing the database of input-output pairs using an existing SCM solver. Here, we used the robust SCM solver provided by Geochemist's Workbench (GWB) software²⁷ because it is one of the most frequently used software in geochemical modeling. We also used the random forest algorithm (RF) and deep neural networks (DNN) for the surrogate model development, because RF is simple and efficient in classification and regression tasks³⁴, and DNN is effective for making data-driven decisions³⁵.

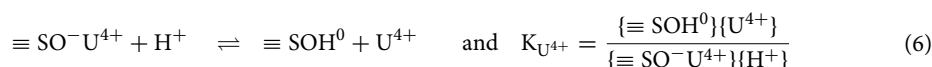
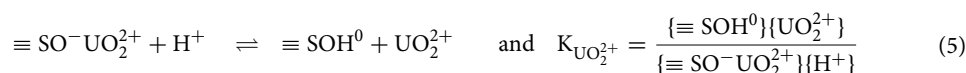
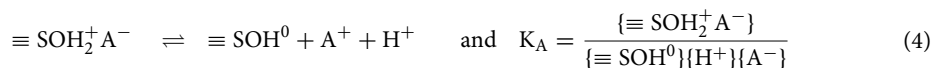
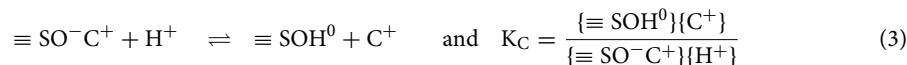
Data set generation

SCMs consist of chemical reactions characterizing the mineral surface charging process and ion complexation combined with geometric models of spatial charge distribution at the interface. To model the mineral surfaces, we used a generic oxide surface covered by reactive amphoteric surface sites ($\equiv \text{SOH}$). These sites can act as an acid or a base by either releasing a proton or binding an additional proton. The acid-base chemistry of the surface group is described using the 2-pK protonation model^{36–38}:



where $\equiv \text{SOH}^0$, $\equiv \text{SOH}_2^+$ and $\equiv \text{SO}^-$ represent charge-neutral, positively and negatively charged sites, respectively.

The electrolyte and uranium ions accumulate near the surface as outer-sphere complexes in the β layer (Fig. 1b). The ions accumulation is described as surface complexation. Here, we used notation adopted by Sverjensky²⁹ and describe ions accumulation via the following desorption reactions:



The electric field generated by a charged mineral surface affects ions ion concentrations near the surface. SCMs account for this effect by weighting bulk concentrations with electric-field energy terms as follows:^{21,26,36–38}

$$\{X^{z_X}\} = [X^{z_X}] \exp\left(\frac{-z_X e \psi_j}{k_B T}\right) \quad (7)$$

where $\{X\}$, $[X]$ represents the interfacial and bulk concentrations of ion X, the electrostatic potential ψ_j is equal to the surface potential for $X = \text{H}^+$ or β -layer potential for $X = \text{C}^+$, A^- , UO_2^{2+} , U^{4+} (see Fig. 1b), z_X is the formal ion charge, k_B is the Boltzmann constant and T is the temperature.

By solving the SCM model described above, one can determine concentrations of all types of surface complexes, charge densities (σ_0 , σ_β , σ_d), and potentials (ψ_0 , ψ_β , ψ_d) at the surface, β - and diffuse layers (see Fig. 1b). A detailed discussion of the SCM equations and numerical methodology can be found elsewhere^{21–23,26}.

The uptake of U(VI) by mineral surface depends on the surface charge of oxide at a given pH condition²⁰. The non-specific adsorption is driven mainly by electrostatic interactions; in the case of uranium ions, they are attracted to positively charged cations by negatively charged surfaces. Consequently, uranium sorption occurs if pH exceeds the Point of Zero Charge (PZC) and generally increases with increasing pH. However, in the case of specific sorption, or sorption of UO_2^{2+} complexed with Cl^- , U(VI) accumulation can occur even at pH below PZC (see Supplementary Information, Fig. S1). To generate an extensive dataset that covers most situations as much as possible for the ML training, we developed a Python script that modifies the uranium thermodynamic database by varying the environmental conditions, such as pH and salt concentrations. We also modified model parameters describing acid-base properties of the surface, surface site density, surface area, capacitance, logK values, and electrolyte/uranium affinity to the mineral surface. The ranges of explored inputs parameter space and environmental conditions are listed in Supplementary Information, Table S2. Consequently, we generated over 1.5 million data by fully exploring the range of uranium sorption conditions. To prepare the generated dataset for the random forest algorithm, we extracted the modified inputs (12) in the SCM solver and the uranium thermodynamic database and their subsequent outputs (12) (Fig. 6).

We split the database of extracted input parameters and output values described above into training and test sets with a ratio of 80:20 % for RF-surrogate and training, validation, and test sets of 80:10:10 % for DNN-surrogate. The min-max normalization strategy is applied to the data to improve training efficiency. We trained the model on the training dataset and evaluated its performance on the test set, which was never used during the training process.

Surrogate model construction and training strategy

We used Scikit-learn³⁹ library for RF-surrogate's model development, training, and testing. As previously²⁶, we used the MultiOutputRegressor module from Scikit-learn to estimate multiple output values. A Random Forest Regressor was used as the base estimator for predicting each target, and a multioutput regressor consisting of 12 random forest regressors was implemented to predict 12 target values. Each random forest consists of 100 decision trees (Fig. 6). The RF-surrogate was trained on a single CPU for 1.5 h.

DNN-surrogate consists of 7 layers of interconnected nodes, including input, output, and 5 hidden layers. The number of neurons in each hidden layer is 512/512/512/512/512 respectively. Input and output dimensions are both 12 and stored in the same format as the one used in RF-surrogate (Fig. 6). During the training, we adopted Adam optimizer⁴⁰ with default parameters for minimizing the loss function. A batch size of 128 was chosen; the

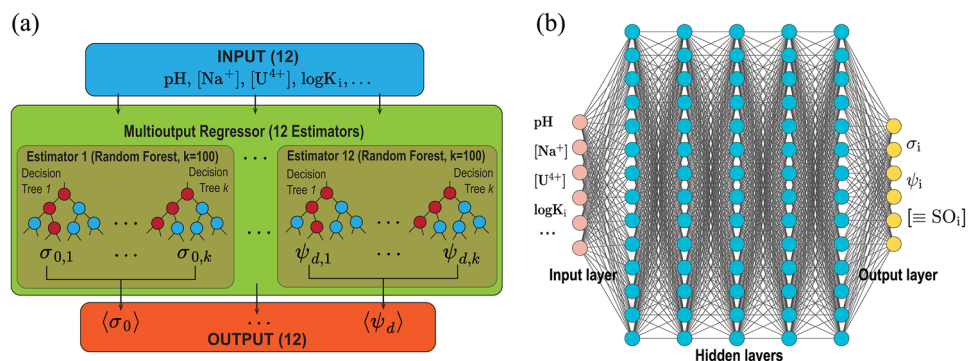


Figure 6. Illustration of AI Surrogates for Uranium Sorption on Mineral Surfaces: **(a)** RF-surrogate, a multioutput regressor as a SCM surrogate, comprising of 12 random forests (estimators), each trained to predict a single target value (output) based on a complete input vector. **(b)** DNN-surrogate, a deep neural network used as a surrogate model for SCM, consists of 5 hidden layers with 512 neurons per layer. Both surrogate models use the same input/output format. The input consists of a description of the oxide surface (surface area, site density), the SCM parameter values, such as affinity constants ($\log K_i$) capacitances, and a description of solution composition (12 values). The output consists of charge density and electrostatic potential in each layer and concentrations of surface complexes (12 values).

initial value of the learning rate was set to 0.001 and was dynamically adjusted based on validation performance. Additionally, an early stopping strategy was implemented to reduce overfitting. The above selections of the number of hidden layers, optimization algorithms, and batch size are determined by RayTune framework⁴¹, which is a library for tuning the hyperparameters of deep learning models. The DNN-surrogate model was implemented in PyTorch⁴² and trained on a single NVIDIA A100 GPU for 2.3 h.

For both surrogate models, mean square error (MSE) was used to minimize the loss between the ground truth and the surrogate prediction during the training procedure.

The source code for our surrogate model is provided (see the “Data availability” section).

Data availability

The data used in this work is available for download from <https://zenodo.org/records/10815543>. Jupyter Notebook used to train the RF- and DNN-SCM models is available at: <https://github.com/nodameCL/Uranium-Sorption-Surrogate>.

Received: 15 May 2023; Accepted: 13 March 2024

Published online: 19 March 2024

References

- Dincer, I. & Acar, C. A review on clean energy solutions for better sustainability. *Int. J. Energy Res.* **39**, 585–606. <https://doi.org/10.1002/er.3329> (2015).
- Goldemberg, J. The promise of clean energy. *Energy Policy* **34**, 2185–2190. <https://doi.org/10.1016/j.enpol.2005.03.009> (2006).
- Jaiswal, K. K. *et al.* Renewable and sustainable clean energy development and impact on social, economic, and environmental health. *Energy Nexus* **7**, 100118. <https://doi.org/10.1016/j.nexus.2022.100118> (2022).
- Rehm, T. E. Advanced nuclear energy: The safest and most renewable clean energy. *Curr. Opin. Chem. Eng.* **39**, 100878. <https://doi.org/10.1016/j.coche.2022.100878> (2023).
- Wainwright, H. M. *et al.* Nuclear waste educator’s workshop: What and how do we teach about nuclear waste?. *J. Environ. Radioact.* **270**, 107288. <https://doi.org/10.1016/j.jenvrad.2023.107288> (2023).
- Alwaeli, M. & Mannheim, V. Investigation into the current state of nuclear energy and nuclear waste management—A state-of-the-art review. *Energies* <https://doi.org/10.3390/en15124275> (2022).
- Jacoby, M. As nuclear waste piles up, scientists seek the best long-term storage solutions. *Chem. Eng. News* (2020).
- Frankel, G. S. *et al.* Recent advances in corrosion science applicable to disposal of high-level nuclear waste. *Chem. Rev.* **121**, 12327–12383 (2021).
- Duquette, D. J., Latanision, R. M., Di Bella, C. A. W. & Kirstein, B. E. Corrosion issues related to disposal of high-level nuclear waste in the Yucca mountain repository—peer reviewers’ perspective. *Corrosion* **65**, 272–280. <https://doi.org/10.5006/1.3319133> (2009).
- Butler, D. France digs deep for nuclear waste: Geological storage of long-lived radioactive material is moving closer to reality in Europe. *Nature* **466**, 804–806 (2010).
- King, F. Container materials for the storage and disposal of nuclear waste. *Corrosion* **69**, 986–1011 (2013).
- Féron, D., Crusset, D. & Gras, J.-M. Corrosion issues in nuclear waste disposal. *J. Nucl. Mater.* **379**, 16–23 (2008).
- Hall, D. S., Behazin, M., Binns, W. J. & Keech, P. G. An evaluation of corrosion processes affecting copper-coated nuclear waste containers in a deep geological repository. *Prog. Mater. Sci.* **118**, 100766 (2021).
- Tan, X., Fang, M. & Wang, X. Sorption speciation of lanthanides/actinides on minerals by TRILFS, EXAFS and DFT studies: A review. *Molecules* **15**, 8431–8468 (2010).
- Maher, K., Bargar, J. R. & Brown, G. E. Jr. Environmental speciation of actinides. *Inorg. Chem.* **52**, 3510–3532 (2013).
- Gronowski, A. Effects of uranium mining on groundwater-geochemical modeling of aqueous uranium speciation due to changing redox conditions (2013).
- McMahon, P. B., Chapelle, F. H. & Bradley, P. M. Evolution of redox processes in groundwater. in *Aquatic Redox Chemistry, Vol. 1071 of ACS Symposium Series*, 581–597 (American Chemical Society, 2011). <https://doi.org/10.1021/bk-2011-1071.ch026>.

18. Bone, S. E. *et al.* Complexation by organic matter controls uranium mobility in anoxic sediments. *Environ. Sci. Technol.* **54**, 1493–1502 (2019).
19. Langmuir, D. Uranium solution-mineral equilibria at low temperatures with applications to sedimentary ore deposits. *Geochim. Cosmochim. Acta* **42**, 547–569 (1978).
20. Geckeis, H., Lützenkirchen, J., Polly, R., Rabung, T. & Schmidt, M. Mineral–water interface reactions of actinides. *Chem. Rev.* **113**, 1016–1062. <https://doi.org/10.1021/cr300370h> (2013).
21. Dzombak, D. A. & Morel, F. M. M. *Surface Complexation Modeling: Hydrous Ferric Oxide* (Wiley, 1990).
22. Lützenkirchen, J. (ed.) *Surface Complexation Modelling* (Elsevier, 2006).
23. Karamalidis, A. K. & Dzombak, D. A. *Surface Complexation Modeling: Gibbsite* (Wiley, 2010).
24. Davis, J. A., Meece, D. E., Kohler, M. & Curtis, G. P. Approaches to surface complexation modeling of uranium (vi) adsorption on aquifer sediments. *Geochim. Cosmochim. Acta* **68**, 3621–3641 (2004).
25. Allison, J. D., Brown, D. S. & Novo-Gradac, K. J. *MINTEQA2/PRODEFA2, a geochemical assessment model for environmental systems: Version 3.0 user's manual* (Environmental Research Laboratory, Office of Research and Development, US Environmental Protection Agency, 1991).
26. Li, C. & Zarzycki, P. A computational pipeline to generate a synthetic dataset of metal ion sorption to oxides for AI/ML exploration. *Front. Nucl. Eng.* **1**, 977743. <https://doi.org/10.3389/fnucn.2022.977743> (2022).
27. Bethke, Craig M. *Geochemical and Biogeochemical Reaction Modeling* 2nd edn. (Cambridge University Press, 2007).
28. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks (2017). [arXiv:1703.01365](https://arxiv.org/abs/1703.01365).
29. Sverjensky, D. A. Prediction of surface charge on oxides in salt solutions: Revisions for 1:1 (M+L-) electrolytes. *Geochim. Cosmochim. Acta* **69**, 225–257 (2005).
30. Robertson, A. & Leckie, J. Cation binding predictions of surface complexation models: Effects of pH, ionic strength, cation loading, surface complex, and model fit. *J. Colloid Interface Sci.* **188**, 444–472. <https://doi.org/10.1006/jcis.1996.4752> (1997).
31. Piasecki, W. Determination of the parameters for the 1-pK triple-layer model of ion adsorption onto oxides from known parameter values for the 2-pK TLM. *J. Colloid Interface Sci.* **302**, 389–395. <https://doi.org/10.1016/j.jcis.2006.06.039> (2006).
32. Charms, R. *et al.* Influence of electrolyte ion adsorption on the derivative of potentiometric titration curve of oxide suspension— theoretical analysis. *Colloids Surf., A* **244**, 9–17. <https://doi.org/10.1016/j.colsurfa.2004.06.001> (2004).
33. Zarzycki, P., Charms, R. & Piasecki, W. Formal mathematical analysis of the existence of the common intersection point in relation to determining the parameters describing ion adsorption at the oxide/electrolyte interface: Comparison of the triple and four-layer models. *Adsorption* **10**, 139–149 (2004).
34. Liu, Y., Wang, Y. & Zhang, J. New machine learning algorithm: Random forest. in *International Conference on Information Computing and Applications*, 246–252 (Springer, 2012).
35. Liu, W. *et al.* A survey of deep neural network architectures and their applications. *Neurocomputing* **234**, 11–26. <https://doi.org/10.1016/j.neucom.2016.12.038> (2017).
36. Yates, D. E., Levine, S. & Healy, T. W. Site-binding model of the electrical double layer at the oxide/water interface. *J. Chem. Soc., Faraday Trans. 1* **70**, 1807–1818. <https://doi.org/10.1039/F19747001807> (1974).
37. Hohlf, H. & Stumm, W. Interaction of Pb²⁺ with hydrous γ -Al₂O₃. *J. Colloid Interface Sci.* **55**, 281–288. [https://doi.org/10.1016/0021-9797\(76\)90035-7](https://doi.org/10.1016/0021-9797(76)90035-7) (1976).
38. Davis, J. A., James, R. O. & Leckie, J. O. Surface ionization and complexation at the oxide/water interface: I. Computation of electrical double layer properties in simple electrolytes. *J. Colloid Interface Sci.* **63**, 480–499. [https://doi.org/10.1016/S0021-9797\(78\)80009-5](https://doi.org/10.1016/S0021-9797(78)80009-5) (1978).
39. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
40. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization (2017). [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
41. Liaw, R. *et al.* Tune: A research platform for distributed model selection and training (2018). [arXiv:1807.05118](https://arxiv.org/abs/1807.05118).
42. Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library (2019). [arXiv:1912.01703](https://arxiv.org/abs/1912.01703).

Acknowledgements

This work was supported by the U.S. Department of Energy, Office of Nuclear Energy, Used Fuel Disposition Campaign, under Award Number DE-AC02-05CH11231. E.A. was supported by the Ingenuity Internship (Next-Generation Nuclear Waste Disposal Internship), provided by LBNL via the U.S. Department of Energy, Office of Nuclear Energy, Used Fuel Disposition Campaign, under Award Number DE-AC02-05CH11231. This research used the National Energy Research Scientific Computing Center (NERSC) resources, a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using NERSC award BES-ERCAP-0024871 and BES-ERCAP-0024868.

Author contributions

C.L. and P.Z. conceived the project. P.Z. and E.A. generated and prepared the data set. E.A. and C.L. developed the RF-surrogate model, and C.L. developed the DNN-surrogate model. C.L. trained, tested, and validated the AI/ML model, visualized and analyzed data, and provided the feature importance analysis. All authors wrote and reviewed the manuscript.

Competing interests

The authors declare that they have no competing interests as defined by Nature Research or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-57026-w>.

Correspondence and requests for materials should be addressed to C.L. or E.O.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024