Machine Learning the Apparent Diffusion Coefficient of Se(IV) in Compacted Bentonite

Authors: XiaoqiongShi, JunleiTian, Jiacong Shen, Zhengye Feng, JiaxingFeng, TaoWu, QingfengLi, Tao Wu

Date: 2024-10-24T00:00:00+00:00

Abstract

Light Gradient Boosting Machine (LightGBM) and Random Forest (RF) algorithms were used to predict the apparent diffusion coefficient of Se(IV) in compacted bentonite. Seven instances of Se(IV) were measured using through-diffusion method. LightGBM (R2 = 0.98 and RMSE = 0.025) exhibited superior predictive accuracy with a training dataset consisting of 956 instances and eight input features from Japan Atomic Energy Agency (JAEA-DDB). Shapley Additive Explanation and Partial Dependence Plots analyses revealed valuable insights into the diffusion mechanism of adsorbed anion obtained by evaluating the relationships between the apparent diffusion coefficient and the dependency of each input feature.

Full Text

Preamble

Journal of Radioanalytical and Nuclear Chemistry https://doi.org/10.1007/s10967-024-09637-w

Application of Machine Learning in Predicting the Apparent Diffusion Coefficient of Se(IV) in Compacted Bentonite

Xiaoqiong Shi[1] · Junlei Tian[1] · Jiacong Shen[1] · Zhengye Feng[1] · Jiaxing Feng[2] · Tao Wu[1] · Qingfeng Li[2]

Received: 16 March 2024 / Accepted: 23 June 2024

Abstract

Light Gradient Boosting Machine (LightGBM) and Random Forest (RF) algorithms were employed to predict the apparent diffusion coefficient of Se(IV) in compacted bentonite. Seven instances of Se(IV) diffusion were measured using the through-diffusion method. LightGBM (R^2 = 0.98 and RMSE = 0.025) exhibited superior predictive accuracy when trained on a dataset comprising 956 instances and eight input features sourced from the Japan Atomic Energy Agency (JAEA-DDB) and relevant literature. Shapley Additive Explanation and Partial Dependence Plot analyses provided valuable insights into the diffusion mechanism of adsorbed anions by evaluating the relationships between the apparent diffusion coefficient and each input feature.

Keywords: Diffusion coefficient · Bentonite · Machine learning · Through-diffusion experiment

Introduction

Bentonite is widely utilized as an engineering barrier in high-level radioactive waste repositories to impede radionuclide release. Radionuclide transport in compacted bentonite is governed by diffusion-based mass transport following Fick's law [1–3]. The retardation mechanism is controlled by both diffusion and sorption processes [4]. The apparent diffusion coefficient (Da) and effective diffusion coefficient (De) are two crucial parameters for characterizing Fickian diffusion [5]. The effective diffusion coefficient is calculated using Fick's first law by analyzing radionuclide diffusion at steady-state. Since reaching steady-state requires considerable time, the through-diffusion method is often employed for weakly and non-sorbing radionuclides such as ^{75}Se(IV) [6], HTO [3, 7–9], and EuIII-EDTA$^-$ [10]. In contrast, the apparent diffusion coefficient measured at the transient stage accounts for the accessible porosity of bentonite in the accumulation term of Fick's second law for adsorbed radionuclides [4, 5, 11, 12]. The in-diffusion method is commonly used for strongly sorbing radionuclides such as ^{233}U(VI) [13], ^{237}Np(V) [14], and $^{134}Cs^+$ [15]. The relationship between the apparent diffusion coefficient (Da) and the effective diffusion coefficient (De) can be expressed as follows [3]:

$$D_a = \frac{D_e}{\varepsilon + \rho_d K_d} = \frac{D_e}{\alpha}$$

where , d, Kd, and α represent porosity, compacted dry density, distribution coefficient, and rock capacity factor, respectively.

Diffusion experiments are generally time-consuming and expensive for obtaining diffusion coefficients. The effective diffusion coefficient is frequently used in diffusion models, including empirical equations and numerical models, enabling rapid and cost-effective predictions. These models can quantitatively describe the dependence of the effective diffusion coefficient of radionuclide species on mineral characteristics (mineral composition, porosity, cation exchange capacity, and external surface area), solution properties (pH, Eh, and mixed ions), and experimental conditions (compacted dry density, temperature, and ionic strength) [6, 11, 12, 16, 17]. However, the predictive accuracy of these models remains unsatisfactory, likely due to approximations and assumptions made during modeling. For example, the effective diffusion coefficient predicted by Archie's equation deviated from experimental values by approximately 1–1.5 orders of magnitude, primarily because of these approximations [18]. Numerical models such as the integrated sorption and diffusion model [19], multi-porosity model [20, 21], and pore-scale model [22, 23] rarely provide predictive accuracy indices, suggesting unsatisfactory performance. Since in-diffusion experiments for strongly sorbing radionuclides are more complex than through-diffusion experiments [3, 14, 15], limited investigations have measured the apparent diffusion coefficient, which represents the coupled effect of radionuclide sorption and diffusion. Consequently, information regarding its relationship with the aforementioned influencing factors remains scarce.

Machine learning algorithms have been extensively developed for regression prediction [20, 24, 25]. These algorithms employ global analysis techniques such as Shapley Additive Explanation, Individual Conditional Expectation, and Partial Dependence Plots to visually analyze the nonlinear relationships between ion diffusion coefficients and influencing factors [24, 25]. Predictive accuracy depends on the dataset, including input features (influencing factors) and data size. Previous radionuclide diffusion experiments primarily focused on specific experimental conditions, such as radionuclide species, compacted dry density, ionic strength, pH, temperature, and bentonite origin [3, 7–10, 17, 26, 27]. However, most investigations considered only a limited number of influencing factors. The dataset size decreased as the number of input features increased, posing challenges for machine learning applications. This study examined the influence of input features and data size on predicting the apparent diffusion coefficient of Se(IV) in compacted bentonite using the through-diffusion method and two machine learning algorithms: Light Gradient Boosting Machine (LightGBM) and Random Forest (RF), based on a comprehensive diffusion dataset. Furthermore, the diffusion mechanism was investigated by analyzing the weight of influencing factors and the relationship between the apparent diffusion coefficient and each factor.

Materials and Methods

Materials

Gaomiaozi (GMZ) bentonite, obtained from the Beijing Research Institute of Uranium Geology, originates from the Gaomiaozi Mine in Xinghe County (Inner Mongolia, China). It was converted to Ba-bentonite by mixing $BaCl_2$ with GMZ powder at a mass ratio of 5%. The detailed preparation procedure for Ba-bentonite followed a previous study [28]. The bentonite was compacted into blocks measuring Ø 25.6 × 12.6 mm with compacted dry densities of 1300–1700 kg/m^3.

The Se(IV) stock solution was prepared by dissolving SeO_2 powder (Sinopharm

Reagent) in 0.1–0.5 mol/L NaCl solution. The initial Se(IV) concentration in diffusion experiments was $(12.5 \pm 1.0) \times 10^{-3}$ mol/L for Ba-bentonite experiments and $(17.9 \pm 0.8) \times 10^{-3}$ mol/L for GMZ experiments. All experimental solutions were prepared with ultrapure water from a Milli-Q system (Millipore, USA). Se concentrations were determined using an Optima 7000DV inductively coupled plasma optical emission spectrometer (PerkinElmer, USA).

Diffusion Experiments

Through-diffusion experiments were conducted under ambient conditions at room temperature (22 ± 3 °C). Bentonite blocks were fully saturated in 0.1–0.5 mol/L NaCl solution for 5 weeks. One side of the diffusion cell (x = 0) was connected to a source reservoir containing 200 mL of Se(IV) solution, while the other side (x = L) was connected to a target reservoir with 10 mL of NaCl solution. Se(IV) diffused through the 12.6 mm thick bentonite block and reached the target reservoir, which was regularly exchanged with fresh 10 mL NaCl solution without Se(IV) to maintain a constant concentration gradient.

Diffusion Data Processing

Two sets of experimental data were obtained: accumulated mass (Acum) of Se(IV) versus time, and flux (J(L,t)) versus time [3, 22]. The first dataset was used with the analytical solution of Fick's law to calculate the effective diffusion coefficient (De) and rock capacity factor (α) as follows [3]:

$$\text{Acum} = S \cdot L \cdot C_0 \left[\frac{D_e \cdot t}{L^2 \cdot \alpha} - \frac{2 \cdot \alpha}{L^2 \cdot \pi^2} \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \exp\left(-\frac{D_e \cdot n^2 \cdot \pi^2 \cdot t}{L^2 \cdot \alpha} \right) \right]$$

where S, L, and C_0 represent the cross-sectional area of the bentonite block, block thickness, and initial Se(IV) concentration in the source reservoir, respectively.

The second dataset was used to verify the effective diffusion coefficient and rock capacity factor as follows:

$$J(L,t) = \frac{\partial \text{Acum}}{\partial t}$$

Machine Learning Database Description and Analysis

The training dataset was sourced from the diffusion database system of the Japan Atomic Energy Agency (JAEA-DDB) (1982–2009) and relevant literature on radionuclide diffusion in bentonite (2010–2023) [8, 9, 21, 22, 28–32]. Insufficient descriptions of influencing factors in JAEA-DDB and the literature reduced the number of instances as the number of influencing factors increased, significantly affecting the predictive performance of ML models.

This study examined the number of instances and input features. Specifically, the input features numbered three, five, and eight, corresponding to 850, 820, and 739 instances in JAEA-DDB (designated J3, J5, and J8). Additionally, 106 instances were collected from published literature, resulting in total instances of 956, 926, and 845 for M3, M5, and M8, respectively. The dataset IM8 contained 956 instances that imputed M8 using the missForest imputation method [33].

The statistical information regarding the number of input features and instances is summarized in Table 1 . The input features included compacted dry density, ion diffusion coefficient in water, ionic strength, rock capacity factor, distribution coefficient, montmorillonite content, ionic charge, and temperature. Table S1 in the supporting information provides a summary of montmorillonite content for various bentonite types. Since no literature value for montmorillonite content of Kunipia-P was available, it was assumed equivalent to that of Na-montmorillonite [34]. The output variable was the apparent diffusion coefficient. Both the ion diffusion coefficient in water and apparent diffusion coefficient were transformed into logarithmic form to enhance predictive accuracy.

To assess the predictive performance of machine learning models, root mean square error (RMSE) and coefficient of determination (R^2) were utilized. These metrics were calculated using the following equations:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{Log}D_{a,i}^{\mathrm{exp}} - \mathrm{Log}D_{a,i}^{\mathrm{pred}})^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\mathrm{Log}D_{a,i}^{\mathrm{exp}} - \mathrm{Log}D_{a,i}^{\mathrm{pred}})^2}{\sum_{i=1}^{N}(\mathrm{Log}D_{a,i}^{\mathrm{exp}} - \mathrm{Log}D_{a}^{\mathrm{exp,ave}})^2}$$

where N represents the number of samples, LogD , and LogD , are the logarithms of the experimental and predicted apparent diffusion coefficients, respectively, and LogD , is the average experimental apparent diffusion coefficient. Models demonstrating high accuracy and performance are characterized by low RMSE and high R^2.

Feature Selection

Feature selection involves identifying the most significant features within a dataset to enhance machine learning model performance. Feature importance measured by the RF algorithm is illustrated in Fig. 1 [Figure 1: see original paper]. The montmorillonite content and ionic strength contributed approximately 5% and 1%, respectively, and were considered weak predictors for this dataset. Based on these results, only six features (rock capacity factor, compacted dry density, ionic charge, distribution coefficient, ion diffusion coefficient in water, and temperature) were selected as Da predictors for IM6 (containing 956 instances).

Results and Discussion

Measurement of the Apparent Diffusion Coefficient by Through-Diffusion Experiments

The through-diffusion method has been extensively used to study anionic radionuclide diffusion, attributed to high diffusivity resulting from the anionic exclusion effect [9, 10, 22]. ^{79}Se is a crucial radionuclide in repository safety assessment, with Se(IV) primarily present as $HSeO_3^-$ at pH 4–7 [35]. Figure 2 [Figure 2: see original paper] illustrates the breakthrough curves of $HSeO_3^-$ in Ba-bentonite and GMZ bentonite. The accumulated mass of $HSeO_3^-$ exhibited a linear increase during the steady-state stage, which is related to the effective diffusion coefficient. In contrast, the flux showed a significant increase during the transient stage, which is linked to the rock capacity factor. The time required to reach steady-state increased from approximately 10 days at a compacted dry density of 1300 kg/m^3 to about 20 days at 1700 kg/m^3.

Table 2 lists the diffusion parameters of $HSeO_3^-$ in compacted Ba-bentonite and GMZ bentonite. The apparent diffusion coefficient of Ba-bentonite decreased from 6.8×10^{-11} to 2.2×10^{-11} m^2/s as the compacted dry density increased from 1300 to 1700 kg/m^3. Additionally, the apparent diffusion coefficient of GMZ bentonite increased from 5.9×10^{-11} to 6.7×10^{-11} m^2/s as the ionic strength increased from 0.1 to 0.3 mol/L. This value was slightly lower than that reported in synthetic pore water (Da = 7.8×10^{-11} m^2/s) in a previous study [36], which could be attributed to higher pH and greater ionic complexity in the synthetic pore water.

Prediction of Machine Learning Algorithms

Both LightGBM and RF are popular and powerful tree-based machine learning algorithms for regression analysis. They have been applied in various diffusion studies, such as predicting the effective diffusion coefficient of Re(VII) in compacted bentonite [20] and the chloride diffusion coefficient in cement [25]. In this study, they were employed to predict the apparent diffusion coefficient of $HSeO_3^-$ in compacted bentonite. Hyperparameters are values specified before training machine learning models, and optimal performance can be achieved through successful hyperparameter optimization. Table 3 lists the optimized hyperparameters for LightGBM and RF algorithms.

The predictive results for the apparent diffusion coefficient using LightGBM and RF for different training sets are shown in Fig. 3 [Figure 3: see original paper] (J3, J5, J8, M3, M5, M8, IM8, and IM6). Figures 3A–H show LightGBM predictions, while Figs. 3I–P show RF predictions. Figures 3Q and S represent the application of IM8 in predicting Da values for ReO_4^-, $HCrO_4^-$, I^-, $CeEDTA^-$, HTO, and Cs^+ in compacted bentonite [37–40]. These species were selected to represent monovalent radionuclide anions, neutral molecules, and monovalent cations. Notably, the M-models exhibited lower RMSE and higher R^2 compared to the J-models, indicating that a larger number of instances enhanced predictive performance.

The RMSE ranked in the following order: J3 > J5 > J8 and M3 > M5 > M8, while R^2 ranked in the opposite order: J3 < J5 < J8 and M3 < M5 < M8. The predictive accuracy of IM8 was similar to M8 for the LightGBM algorithm, while IM8 exhibited higher predictive accuracy than M8 for the RF algorithm. These observations indicate that an increased number of input features enhanced predictive performance [41].

For J3 and M3, both LightGBM and RF produced relatively low prediction accuracy with R^2 below 0.8, indicating that three input features were insufficient for accurate predictions of the apparent diffusion coefficient of $HSeO_3^-$ in compacted bentonite. However, when the number of input features increased to five, the predictive accuracy of M5 using LightGBM significantly improved, with R^2 of 0.939 and RMSE of 0.042. IM8 demonstrated superior predictive accuracy for both LightGBM (R^2 = 0.98 and RMSE = 0.025) and RF (R^2 = 0.95 and RMSE = 0.039). The generalization of IM8 in predicting Da values of radionuclides also exhibited good performance, with R^2 of 0.79 and RMSE of 0.34 for LightGBM and R^2 of 0.63 and RMSE of 0.45 for RF (Fig. 3Q and S). This indicates that LightGBM has better generalization ability than RF. LightGBM utilizes gradient-based one-sided sampling, exclusive feature bundling, and histogram-based algorithm techniques, effectively preventing overfitting, accelerating training speed, and reducing memory consumption, thereby improving predictive accuracy. It is worth mentioning that eight input features may be insufficient to fully describe the complexities of radionuclide diffusion in compacted bentonite. Incorporating techniques such as recursive feature elimination or forward selection could further improve model performance and reduce overfitting when additional input features are considered.

The predictive accuracy (R^2 and RMSE) was similar for both IM8 and IM6 (Fig. 3G, H, O, and P). However, application tests showed that IM6 had lower predictive accuracy than IM8 (Fig. 3Q–T). Therefore, only the IM8 model is considered in this study.

Spearman and Shapley Additive Explanation Analyses

Figure 4 [Figure 4: see original paper] shows the correlation of each feature with Da in the training dataset using Spearman analysis and the weight of input features for IM8 using Shapley Additive Explanation (SHAP) analysis. Models J3–J8 and M3–IM8 showed similar Spearman correlation coefficients, indicating that the number of input features and instances had insignificant impact on the nonlinear relationship (Fig. 4A). Four parameters (compacted dry density, rock capacity factor, distribution coefficient, and ionic charge) were negatively correlated with the apparent diffusion coefficient, while three parameters (ion diffusion coefficient in water, montmorillonite content, and temperature) exhibited positive correlation. Ionic strength had an insignificant impact. The rock capacity factor, distribution coefficient, compacted dry density, and ionic charge had high absolute Spearman correlation coefficient values, indicating strong correlations with the apparent diffusion coefficient.

The Shapley Additive Explanations (SHAP) method is commonly used to analyze the weight or importance of input features on predictions. SHAP results for LightGBM and RF demonstrated approximately similar ranking orders (Fig. 4B). Specifically, LightGBM ranked the ion diffusion coefficient in water fifth, whereas RF ranked it fourth. The top two input features for predicting the apparent diffusion coefficient were the rock capacity factor and compacted dry density, contributing a total of 53.3% for LightGBM and 53.7% for RF compared to all features.

Montmorillonite content and ionic strength are two important parameters that have been extensively investigated, as they significantly influence the effective diffusion coefficient of anionic radionuclides [8, 22, 42]. However, SHAP analysis demonstrates that montmorillonite content and ionic strength made insignificant contributions to the prediction (Fig. 4B). The tendency for ionic strength is consistent with [9], who reported that its insignificant influence on the apparent diffusion coefficient of Cs^+ and Na^+ was attributed to the coupled effects of diffusion and sorption. Further studies are needed to verify these results, as the weight of input features is influenced by different types and numbers of instances, as well as machine learning algorithms.

Partial Dependence Plots Analysis

Partial Dependence Plots (PDP) can visualize the relationship between each input feature and the predicted apparent diffusion coefficient. Figure 5 [Figure 5: see original paper] shows the contribution of each input feature to the prediction for IM8 using PDP analysis. The gray column shows the distribution of data points for an input feature. PDP values for LightGBM are represented by solid lines, while dashed lines represent RF values. A flat curve suggests that the feature has an insignificant influence on the predicted outcome, whereas a steep curve indicates a stronger relationship. Both LightGBM and RF demonstrate similar trends for the dependency of the predicted apparent diffusion coefficient on each input feature.

The analysis of the relationship between the predicted apparent diffusion coefficient and input features related to experimental conditions was consistent with published experimental results [3, 13, 26]. Specifically, it showed a negative correlation with compacted dry density (Fig. 5A). This can be explained by the decrease in available pores for radionuclide diffusion with increasing compaction. At compacted dry densities above 1600 kg/m^3, the decreasing tendency changed slowly, attributed to the reduction of interlayer pores to a single water layer [21, 42, 43].

The apparent diffusion coefficient exhibited a positive correlation with temperature (Fig. 5B). In a repository, the temperature of clay near waste containers could exceed 100 °C [44]. The positive relationship between the effective diffusion coefficient and temperature was reported for HTO, $^{36}Cl^-$, and ReO_4^-, following the Arrhenius equation [3, 7, 26]. Figures 5C and D show that the apparent diffusion coefficient exhibited insignificant correlations with montmorillonite content and ionic strength. Slight negative and positive impacts were observed for montmorillonite content ranging from 0.5 to 0.8 and ionic strength below 1.0, respectively, which agrees with their correlations with the effective diffusion coefficient [20]. Numerous studies have shown that as ionic strength increases, the effective diffusion coefficient of radionuclides also increases [8, 22]. This tendency was explained by the decreased thickness of the electrical double layer in high-salinity solutions [9, 22]. In general, the consistency between PDP and experimental results implies that PDP analysis can provide insights into the diffusion laws and mechanisms of radionuclides.

The input features—rock capacity factor, distribution coefficient, ion diffusion coefficient in water, and ionic charge—are parameters related to radionuclide properties. Their dependencies on the apparent diffusion coefficient were unclear due to the coupled effects of solid, liquid, and radionuclide interactions. Figure 5E indicates that the rock capacity factor had a negative impact on the prediction. According to Eq. (3), the apparent diffusion coefficient is inversely proportional to the rock capacity factor. The negative relationship between the apparent diffusion coefficient and distribution coefficient can be attributed to the linear correlation between the rock capacity factor and distribution coefficient (Fig. 5F). The predicted apparent diffusion coefficient increased with increasing ion diffusion coefficient in water (Fig. 5G), indicating that radionuclides diffuse quickly in both liquid and solid phases. The RF algorithm shows that ionic charge had an insignificant effect on the apparent diffusion coefficient, whereas the LightGBM algorithm demonstrates that it decreased as the ionic charge reduced from 0 to +3 (Fig. 5H).

Conclusions

The effects of input features and instances on predicting the apparent diffusion coefficient were investigated using Light Gradient Boosting Machine (LightGBM) and Random Forest (RF) algorithms. $HSeO_3^-$ diffusion experiments in compacted bentonite were conducted using a through-diffusion method to validate predictive performance. Increasing the number of input features resulted in a decrease in instances. LightGBM ($R^2 = 0.98$ and RMSE = 0.025) and RF ($R^2 = 0.95$ and RMSE = 0.039) exhibited superior predictive accuracy for the training set of 956 instances and eight input features: compacted dry density, ion diffusion coefficient in water, ionic strength, rock capacity factor, distribution coefficient, montmorillonite content, ionic charge, and temperature.

Shapley Additive Explanations revealed that the top two input features for predicting the apparent diffusion coefficient were the rock capacity factor and compacted dry density. Partial Dependence Plots indicated that the apparent diffusion coefficient was negatively dependent on the rock capacity factor, compacted dry density, and distribution coefficient, while a positive relationship was observed between the apparent diffusion coefficient and the ion diffusion coefficient in water. This study presents a method for predicting the apparent diffusion coefficient, quantifying influencing factors, and understanding the effect of each input feature on the apparent diffusion coefficient. The insights into diffusion mechanisms are beneficial for repository safety assessment.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10967-024-09637-w.

Acknowledgements This work was supported by the Key Program of National Natural Science Foundation of China (12335008), Huzhou Science and Technology Planning Project (2021GZ60), and Scientific Research and Innovation Project for Postgraduate by School of Engineering, Huzhou University (2024GXYKYCX14).

References

1. Benning JL, Barnes DL (2009) Comparison of modeling methods for the determination of effective porosities and diffusion coefficients in through-diffusion tests. Water Resour Res 45(9):W09419. https://doi.org/10.1029/2008wr007236
2. Shackelford CD, Moore SM (2013) Fickian diffusion of radionuclides for engineered containment barriers: diffusion coefficients, porosities, and complicating issues. Eng Geol 152(1):133–147. https://doi.org/10.1016/j.enggeo.2012.10.014
3. Van Loon LR, Soler JM (2003) Diffusion of HTO, $^{36}Cl^-$, $^{125}I^-$ and $^{22}Na^+$ in opalinus clay: effect of confining pressure, sample orientation, sample depth and temperature. Paul Scherrer Institut (PSI):1015—2636
4. Liu Z, Emami-Meybodi H (2022) Apparent diffusion coefficient for adsorption-controlled gas transport in nanoporous media. Chem Eng J 450:138105. https://doi.org/10.1016/j.cej.2022.138105
5. Cormenzana JL, Garcia Gutiérrez M, Missana T, Junghanns Á (2003) Simultaneous estimation of effective and apparent diffusion coefficients in compacted bentonite. J Contam Hydrol 61(1–4):63–72. https://doi.org/10.1016/s0169-7722(02)00113-4
6. Effects of mineral compositions on matrix diffusion and sorption of ^{75}Se(IV) in Granite. Environ Sci Technol 52(3):1320–1329. https://doi.org/10.1021/acs.est.7b05795
7. Bestel M, Glaus MA, Frick S, Gimmi T, Juranyi F, Van Loon LR, Diamond LW (2018) Combined tracer through-diffusion of HTO and ^{22}Na through Na-montmorillonite with different bulk dry densities. Appl Geochem 93:158–166. https://doi.org/10.1016/j.apgeochem.2018.04.008

8. Fukatsu Y, Yotsuji K, Ohkubo T, Tachi Y (2021) Diffusion of tritiated water, $^{137}Cs^{+}$, and $^{125}I^{-}$ in compacted Ca-montmorillonite: experimental and modeling approaches. Appl Clay Sci 211:106176. https://doi.org/10.1016/j.clay.2021.106176

9. Tachi Y, Yotsuji K (2014) Diffusion and sorption of Cs^{+}, Na^{+}, I^{-} and HTO in compacted sodium montmorillonite as a function of porewater salinity: integrated sorption and diffusion model. Geochim Cosmochim Acta 132:75–93. https://doi.org/10.1016/j.gca.2014.02.004

10. Descostes M, Pointeau I, Radwan J, Poonoosamy J, Lacour JL, Menut D, Vercouter T, Dagnelie RVH (2017) Adsorption and retarded diffusion of EuIII-EDTA^{-} through hard clay rock. J Contam Hydrol 544:125–132. https://doi.org/10.1016/j.jhydrol.2016.10.014

11. Cho WJ, Oscarson DW, Hahn PS (1993) The measurement of apparent diffusion coefficients in compacted clays: an assessment of methods. Appl Clay Sci 8(4):283–294. https://doi.org/10.1016/0169-1317(93)90009-P

12. Gooddy DC, Kinniburgh DG, Barker JA (2007) A rapid method for determining apparent diffusion coefficients in chalk and other consolidated porous media. J Hydrol 343(1–2):97–103. https://doi.org/10.1016/j.jhydrol.2007.06.014

13. Joseph C, Van Loon LR, Jakob A, Steudtner R, Schmeide K, Sachs S, Bernhard G (2013) Diffusion of U(VI) in opalinus clay: influence of temperature and humic acid. Geochim Cosmochim Acta 109:74–89. https://doi.org/10.1016/j.gca.2013.01.027

14. Wu T, Amayri S, Drebert J, Van Loon LR, Reich T (2009) Neptunium(V) sorption and diffusion in opalinus clay. Environ Sci Technol 43:6567–6571. https://doi.org/10.1021/es9008568

15. Van Loon LR, Müller W (2014) A modified version of the combined in-diffusion/abrasive peeling technique for measuring diffusion of strongly sorbing radionuclides in argillaceous rocks: a test study on the diffusion of caesium in Opalinus Clay. Appl Radiat Isot 90:197–202. https://doi.org/10.1016/j.apradiso.2014.04.009

16. Cui LY, Masum SA, Ye WM, Thomas HR (2021) Investigation on gas migration behaviours in saturated compacted bentonite under rigid boundary conditions. Acta Geotech 17(6):2517–2531. https://doi.org/10.1007/s11440-021-01424-1

17. Kozaki T, Sawaguchi T, Fujishima A, Sato S (2010) Effect of exchangeable cations on apparent diffusion of Ca^{2+} ions in Na- and Ca-montmorillonite mixtures. Phys Chem Earth 35(6–8):254–258. https://doi.org/10.1016/j.pce.2010.04.006

18. Van Loon LR, Mibus J (2015) A modified version of Archie's law to estimate effective diffusion coefficients of radionuclides in argillaceous rocks

and its application in safety analysis studies. Appl Geochem 59:85–94. https://doi.org/10.1016/j.apgeochem.2015.05.003

19. Ochs M, Lothenbach B, Wanner H, Sato H, Yui M (2001) An integrated sorption-diffusion model for the calculation of consistent distribution and diffusion coefficients in compacted bentonite. J Contam Hydrol 47(2–4):283–296. https://doi.org/10.1016/S0169-7722(00)00157-1
20. Feng Z, Gao Z, Wang Y, Wu T, Li Q (2023) Application of machine learning to study the effective diffusion coefficient of Re(VII) in compacted bentonite. Appl Clay Sci 243:107076. https://doi.org/10.1016/j.clay.2023.107076
21. Wu T, Wang Z, Tong Y, Wang Y, Van Loon LR (2018) Investigation of Re(VII) diffusion in bentonite by through-diffusion and modeling techniques. Appl Clay Sci 166:223–229. https://doi.org/10.1016/j.clay.2018.08.023
22. Wu T, Yang Y, Wang Z, Shen Q, Tong Y, Wang M (2020) Anion diffusion in compacted clays by pore-scale simulation and experiments. Water Resour Res 56(11):e2019WR027037. https://doi.org/10.1029/2019wr027037
23. Yang Y, Churakov SV, Patel RA, Prasianakis N, Deissmann G, Bosbach D, Poonoosamy J (2024) Pore-scale modeling of water and ion diffusion in partially saturated clays. Water Resour Res 60(1):e2023WR035595. https://doi.org/10.1029/2023wr035595
24. Mohammadi Golafshani E, Kashani A, Kim T, Arashpour M (2022) Concrete chloride diffusion modelling using marine creatures-based metaheuristic artificial intelligence. J Cleaner Prod 374:134021. https://doi.org/10.1016/j.jclepro.2022.134021
25. Tran VQ (2022) Machine learning approach for investigating chloride diffusion coefficient of concrete containing supplementary cementitious materials. Constr Build Mater 328:127103. https://doi.org/10.1016/j.conbuildmat.2022.127103
26. Wu T, Wang Z, Li Q, Pan G, Li J, Van Loon LR (2016) Re(VII) diffusion in bentonite: effect of organic compounds, pH and temperature. Appl Clay Sci 127–128:10–16. https://doi.org/10.1016/j.clay.2016.03.039
27. Tian W, Li C, Liu X, Wang L, Zheng Z, Wang X, Liu C (2012) The effect of ionic strength on the diffusion of ^{125}I in Gaomiaozi bentonite. J Radioanal Nucl Chem 295(2):1423–1430. https://doi.org/10.1007/s10967-012-2284-y
28. Wu T, Feng Z, Geng Z, Xu M, Shen Q (2023) Restriction of Re(VII) and Se(IV) diffusion by barite precipitation in compacted bentonite. Appl Clay Sci 232:106803. https://doi.org/10.1016/j.clay.2022.106803
29. Geng Z, Feng Z, Li H, Wang Y, Wu T (2022) Porosity investigation of compacted bentonite using through-diffusion method and multi-porosity model. Appl Geochem 146:105480. https://doi.org/10.1016/j.apgeochem.2022.105480

30. Glaus MA, Frick S, Rossé R, Van Loon LR (2010) Comparative study of tracer diffusion of HTO, $^{22}Na^+$ and $^{36}Cl^-$ in compacted kaolinite, illite and montmorillonite. Geochim Cosmochim Acta 74(7):1999–2010. https://doi.org/10.1016/j.gca.2010.01.010

31. Tachi Y, Nakazawa T, Ochs M, Yotsuji K, Suyama T, Seida Y, Yamada N, Yui M (2010) Diffusion and sorption of neptunium(V) in compacted montmorillonite: effects of carbonate and salinity. Radiochim Acta 98(9–11):711–718. https://doi.org/10.1524/ract.2010.1768

32. Wu T, Wang Z, Wang H, Zhang Z, Van Loon LR (2017) Salt effects on Re(VII) and Se(IV) diffusion in bentonite. Appl Clay Sci 141:104–110. https://doi.org/10.1016/j.clay.2017.02.021

33. Joel LO, Doorsamy W, Paul BS (2024) On the performance of imputation techniques for missing values on healthcare datasets. ArXiv. https://doi.org/10.48550/arXiv.2403.14687

34. González Sánchez F, Van Loon LR, Gimmi T, Jakob A, Glaus MA, Diamond LW (2008) Self-diffusion of water and its dependence on temperature and ionic strength in highly compacted montmorillonite, illite and kaolinite. Appl Geochem 23(12):3840–3851. https://doi.org/10.1016/j.apgeochem.2008.08.008

35. Shi K, Ye Y, Guo N, Guo Z, Wu W (2013) Evaluation of Se(IV) removal from aqueous solution by GMZ Na-bentonite: batch experiment and modeling studies. J Radioanal Nucl Chem 299(1):583–589. https://doi.org/10.1007/s10967-013-2807-1

36. Wu T, Wang H, Zheng Q, Zhao YL, Van Loon LR (2014) Diffusion behavior of Se(IV) and Re(VII) in GMZ bentonite. Appl Clay Sci 101:136–140. https://doi.org/10.1016/j.clay.2014.07.028

37. Tachi Y, Yotsuji K, Seida Y, Yui M (2009) Diffusion of cesium and iodine in compacted sodium montmorillonite under different saline conditions. MRS Proc 1193:545–552. https://doi.org/10.1557/PROC-1193-545

38. Feng Z, Tian J, Wu T, Wei G, Li Z, Shi X, Wang Y, Li Q (2024) Unveiling the Re, Cr, and I diffusion in saturated compacted bentonite using machine-learning methods. Nucl Sci Tech 35(6):93. https://doi.org/10.1007/s41365-024-01456-8

39. Joseph C, Mibus J, Trepte P, Müller C, Brendler V, Park DM, Jiao Y, Kersting AB, Zavarin M (2017) Long-term diffusion of U(VI) in bentonite: dependence on density. Sci Total Environ 575:207–218. https://doi.org/10.1016/j.scitotenv.2016.10.005

40. Wu T, Hong Y, Shao D, Zhao J, Feng Z (2023) Experimental and modeling study of the diffusion path of Ce(III)-EDTA in compacted bentonite. Chem Geol 636:121639. https://doi.org/10.1016/j.chemgeo.2023.121639

41. Zhu T, Zhang Y, Tao C, Chen W, Cheng H (2023) Prediction of organic contaminant rejection by nanofiltration and reverse osmosis membranes using interpretable machine learning models. Sci Total Environ 857:159348. https://doi.org/10.1016/j.scitotenv.2022.159348

42. Van Loon LR, Glaus MA, Müller W (2007) Anion exclusion effects in compacted bentonites: towards a better understanding of anion diffusion. Appl Geochem 22(11):2536–2552. https://doi.org/10.1016/j.apgeochem.2007.07.008

43. Holmboe M, Wold S, Jonsson M (2012) Porosity investigation of compacted bentonite using XRD profile modeling. J Contam Hydrol 128(1–4):19–32. https://doi.org/10.1016/j.jconhyd.2011.10.004

44. Zheng L, Rutqvist J, Birkholzer JT, Liu HH (2015) On the impact of temperatures up to 200 °C in clay repositories with bentonite engineer barrier systems: a study with coupled thermal, hydrological, chemical, and mechanical modeling. Eng Geol 197:278–295. https://doi.org/10.1016/j.enggeo.2015.08.026