RadoNorm
Managing risks from radon and NORM

Deliverable 2.17

Report on the capability of models to predict site-specific NOR K_d

Work Package 2

Frank Bok, Viktor Dück (HZDR), Miquel Vidal, Anna Rigol, Dídac Serraïma-López and Joan Serra-Ventura (University of Barcelona)

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Document information

Project Acronym	**RadoNorm**
Project Title	**Towards effective radiation protection based on improved scientific evidence and social considerations - focus on radon and NORM**
Project Type	**RIA**
EC grant agreement No.	**900009**
Project starting / end date	**1st September 2020 – 31st August 2025**
Work Package No.	**2**
Work Package Title	**Exposure**
Deliverable No.	**2.17**
Deliverable Title	**Report on the capability of models to predict site-specific NOR K_d**
Lead Beneficiary	**HZDR**
Contractual Delivery Date	**31/05/2025**
Actual Delivery Date	**31/05/2025**
Type	**Report**
Dissemination level	**PU**
Authors	**Frank Bok, Viktor Dück (HZDR), Miquel Vidal, Anna Rigol, Dídac Serraïma-López and Joan Serra-Ventura (University of Barcelona)**

To be cited as:

Bok F., Dück V., Vidal M. Rigol A., Serraïma-López D., Serra-Ventura J. (2025). Report on the capability of models to predict site-specific NOR K_d. Final version as of 31.05.2025 of deliverable D2.17 of project RadoNorm.

Disclaimer

This document reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

Acknowledgement

This document is a deliverable of the RadoNorm project. This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Status of deliverable		
	By	Date
Delivered (Lead Beneficiary)	Frank Bok, Viktor Dück (HZDR), Miquel Vidal, Anna Rigol, Dídac Serraïma-López and Joan Serra-Ventura (University of Barcelona)	22/05/2025
Verified (WP Leader)	L. Février (IRSN)	23/05/2025
Reviewed (Reviewers)	T. Arnold / S. Sachs (HZDR)	22/05/2025
Approved (PC)	PC	28/05/2025
Submitted to EC (PC)	PC	28/05/2025

Executive Summary

The dispersion of naturally occurring radionuclides (NOR), such as uranium (U), thorium (Th), radium (Ra), polonium (Po), radon (Rn), and lead (Pb), poses a risk to the environment and population, and must therefore be avoided. To better assess this potential risk, improved models are urgently needed to predict the sorption and subsequent mobility of NOR in soils. One objective of the RadoNorm project is to contribute to the improvement of these models.

This requires detailed knowledge of the chemical behaviour of NOR in the environment. To achieve this, complementary laboratory and in situ studies on a limited number of NORs were performed in Task 2.7, with the aim of generating new knowledge on the interaction behaviour of NOR in the environment and improving the models used to predict NOR mobility. The NORs of interest in our studies are U, Ra, Po and Pb.

Model improvement is partially based on expanding datasets with sorption and desorption parameters, which will eventually enable the development of thermodynamic, probabilistic, or parametric equations. These equations can be used to predict various parameters, including the solid-liquid distribution coefficient (K_d) of NOR. Within the radio-ecological community, the K_d value of a contaminant between the immobilised phase (mineral precipitation or sorption) and the aqueous phase is used as an indicator of mobility. It is possible to model the saturation of NOR mineral phases, as well as the sorption of NOR on the surfaces of naturally occurring mineral phases, using mechanistic models or probabilistic distribution functions. This allows the K_d value to be calculated using the model parameters.

This work aims to apply the smart-K_d method to predict the K_d values of some NORs in selected natural scenarios, and to compare the results obtained with measured field data. The wetland at the Rophin site was used as test site. For the whitish layer of the Rophin site, the predicted K_d value for U (K_d^{calc} = 66 to 98 L/kg) was in very good agreement with an experimentally determined value (K_d^{exp} = ~50 L/kg).

This report also presents an integrated approach to generating K_d (NOR) prediction models, particularly for Ra, Po and U. A dataset containing new experimental data is created that is supplemented by literature data and contains a large number of characteristic soil properties. In the next step, the most important properties determining the interaction of a particular NOR in soils are identified using chemometric tools such as principal component analyses and partial least squares regression analysis. These properties are then used to construct parametric equations derived from univariate or multivariate linear regressions. This allows as much of the K_d (NOR) variability as possible to be explained. Where detailed soil data is lacking, probabilistic distribution functions help estimate Kd(NOR) for similar soil groups. This is exemplified with Ra. The approach successfully predicts K_d(NOR), often explaining over 80% of its variance, and is suitable for use in radioecological risk assessment codes.

Furthermore, expected K_d values were predicted as a function of variable geochemical parameters in a blind prediction exercise. Despite the incomplete thermodynamic surface complexation data and input parameters required for thermodynamic modelling, partition coefficients for U, Ra, and Pb could be calculated using the smart-K_d approach, and compared with the available field data. There was good agreement between the calculated and measured K_d values, particularly for U. Similar good results were obtained using the approach of constructing parametric models and probabilistic distribution functions. The reasons for the deviations observed in the case of Ra lie in the incomplete surface complexation modelling (SCM) data available for Ra in many mineral phases. This situation is improved by Viktor Dück's PhD thesis, in which the chemical analogues of Ra, barium (Ba) and strontium (Sr), are investigated for their sorption behaviour on gibbsite and quartz. The aim is to develop a generic sorption model for alkaline earth metal ions on feldspars. In his work a mechanistic understanding of Sr(II) and Ba(II) sorption on gibbsite, a representative mineral for aluminol surface sites, was provided and it was found that sorption is strongly pH-dependent and sensitive to ionic strength. Outer-sphere complexation that was confirmed by extended X-ray absorption spectroscopy was the dominant sorption process. Furthermore, an unexpected and pronounced influence of the background electrolyte (NaCl) on the sorption

behaviour was observed, shifting sorption edges significantly. This is in contrast with previous studies and suggests electrolyte-specific interactions at the gibbsite surface.

In conclusion, in addition to a complete set of SCM parameters, the reliable prediction of K_d values requires comprehensive characterisation of the soil, which is rarely available. One critical parameter is the specific surface area of the mineral phases present in the soil. This is often given as a total surface area at best and is not broken down into the individual minerals. The inclusion of organic substances also poses a problem for prediction as they are often poorly characterised (or not characterised at all), and for many of them, especially those with a high molecular weight, e.g. humic and fulvic acids, there is often no thermodynamic data available for interaction with NOR.

Despite the above-mentioned uncertainties, there is a very good agreement between the predicted and measured K_d values for both approaches. Using the smart-K_d approach and the probabilistic distribution functions allows large parameter ranges to be captured quickly and helps to identify relevant sorption processes. Smart-K_d predictions can be further improved by extending the SCM dataset to fill existing data gaps.

Table of content

List of figures

List of Tables

Glossary

AICc	Corrected Akaike Information Criterion
CDF	Cumulative distribution functions
CEC	Cation exchange capacity
DDL	Diffuse double layer sorption model
DIC	Dissolved inorganic carbon, the sum of inorganic carbon species in a solution
DOC	Dissolved organic carbon, the sum of organic carbon species in a solution
CPU	Central processing unit (in terms of CPU time, computing time for numerical task)
E_H	Redox potential is a measure of the tendency of a chemical species to acquire electrons from or lose electrons to an electrode and thereby be reduced or oxidised respectively. Redox potential is expressed in volts (V).
FLSD	Fisher's Least Significant Difference
Gbs	Gibbsite $(AlOH)_3$
HZDR	Helmholtz-Zentrum Dresden-Rossendorf e.V.
I	Ionic strength of a solution is a function of the concentration of all ions present in that solution.
ICP-MS	Inductively coupled plasma mass spectrometry
IEx	Ion exchange, immobilization process where ions are incorporated in the interlayer of some minerals, esp. sheet silicates (montmorillonite, illite).
K_d	Distribution coefficient of a contaminant between the immobile phase and the mobile, aqueous phase.
log*K*	Decadic logarithm of the reaction constant of a chemical reaction derived by the mass-law-equation
LQ	Limit of quantification
LT	Long-term
NOR	Naturally occurring radionuclide
PCA	Principal component analyses
PCTL n	n^{th} percentile
pH	pH value is the negative decimal logarithm of the activity of hydrogen ions in the solution.
PLS	Partial least squares regression analysis
Qtz	Quartz (SiO_2)
RMSE	Root-mean-square
RMSECV	Root-mean-square of cross-validation
SCM	Surface Complexation Model, quasi-thermodynamic model to describe sorption as a chemical reaction of a dissolved species on a mineral's surface.

Smart-K_d	Automated calculation of K_d values based on mechanistic models (SCM, IEx) with respect to varible geochemical parameters (pH, E_H, DIC, Ca, Mg and SO_4 concentration) resulting in a multidimensional matrix of parameter and corresponding K_d.
ST-D	Short-term desorption
ST-S	Short-term sorption
TENORM	Technologically enhanced naturally occurring radioactive material
TOC	Total organic carbon, the sum of organic carbon in a sample
UB	University of Barcelona
VIP	Variable importance in projection score

1. The smart-K_d concept

The K_d value is a very important parameter describing the distribution of an element between the liquid and solid phases. It is required as an essential input in geochemical models describing the dispersion of contaminants in the environment. The so-called smart-K_d concept is a theoretical framework that addresses the processes of radionuclide sorption in natural systems. The smart-K_d values are derived primarily from mechanistic surface complexation models (SCM) and ion exchange (IEx) reactions. The distribution coefficients can thus reflect temporal and spatial variability, depending on the prevailing geochemical conditions. Given the extensive parameter space inherent in real-world applications, the identification and quantification of the most relevant parameters poses a significant challenge. Sorption is described by the "component additivity approach" (bottom-up strategy), in which the contribution of sorption of one element is considered on every single mineral fraction of a sediment, also including respective competitive effects. A variety of possible combinations of geochemical parameters are generated by a numerical tool as individual parameter vectors and then passed to a geochemical speciation code (often using an appropriate ASCII data processing tool as an intermediate layer), which calculates the smart-K_d value associated with that parameter vector. This process is then repeated for all parameter combinations. The fundamental strategy for the generation and utilisation of the smart-K_d matrices is described in detail in Stockmann et al. (2017).

The calculation code used to create the Smart-K_d matrices was the geochemical software “Geochemist's Workbench” (React, Vers. 17.0.1) (Bethke, 2022) and a Python framework was used to vary the geochemical parameters (pH, E_H, Ca, Mg, SO_4, DIC).

1.1 Application examples

1.1.1 Rophin site

The Rophin site is one of the uranium (U) mines exploited in France between 1948 and 1957 in the department of Puy-de-Dôme in the Auvergne-Rhône-Alpes region. During the exploitation of U ores, several discharges of processed U ore waters were carried out, resulting in the accumulation of mining deposits, including namely U and radium (Ra) in a wetland located downstream of the site of Rophin. (Grangeon et al., 2023).

The “whitish layer” in the wetland downstream of the Rophin old U mine was used in this project as an example for the application of the smart-K_d concept to predict the K_d value in a natural and complex geochemical scenario. This layer, at a depth from ~−13 cm to ~−23 cm below surface, has shown to hold a high concentration of naturally occurring radionuclides (NOR) (Martin et al., 2020).

The field data (mineral and pore water composition, NOR concentration in the dried solid) were kindly provided by Gilles Montavon (SUBATECH) (Montavon, 2021). The obtained data used in the smart-K_d calculation are shown in *Table 1* (mineral composition), *Table 2* (pore water composition) and *Table 3* (NOR concentration). More information is given in Nivesse et al. (2023).

1.1.1.1 Chemical and mineralogical data

Table 1 – Mineral composition (semi-quantitative analysis) of the "whitish layer" at Rophin site (Montavon, 2021)

Mineral	Formula	Percentage
Quartz	SiO_2	76.0 %
Albite	$NaAlSi_3O_8$	5.7 %
Orthoclase	$KAlSi_3O_8$	3.4 %
Microcline	$KAlSi_3O_8$	6.6 %
Kaolinite	$Al_2Si_2O_5(OH)_4$	2.6 %
Montmorillonite	$Na_{0.3}Mg_{0.36}Al_{1.67}Si_4O_{10}(OH)_2$	3.1 %
Illite	$K_{0.6}Mg_{0.25}Al_{1.8}Al_{0.5}Si_{3.5}°_{10}(OH)_2$	2.6 %

Table 2 – Chemical parameters of pore water in the "whitish layer" at Rophin site (Montavon, 2021)

Chemical parameter	Value	Unit
pH	5.25 ± 0.05	-
E_H	261 ± 4	mV
Na^+	9.33 ± 0.3	$mg \cdot L^{-1}$
K^+	5.73 ± 0.15	$mg \cdot L^{-1}$
Mg^{2+}	0.87 ± 0.02	$mg \cdot L^{-1}$
Ca^{2+}	4.10 ± 0.03	$mg \cdot L^{-1}$
Cl^-	16.28 ± 0.12	$mg \cdot L^{-1}$
SO_4^{2-}	10.58 ± 0.28	$mg \cdot L^{-1}$

Table 3 – NOR concentration in the "whitish layer" at Rophin site (Montavon, 2021)

NORM	Value	Unit
^{226}Ra	16.23 ± 1.66	$Bq \cdot g^{-1}$
U	804.7 ± 27.11	$\mu g \cdot g^{-1}$
Pb	307.0 ± 10.6	$\mu g \cdot g^{-1}$

1.2 Model development

1.2.1 Specific surface areas of the minerals

For the calculation of K_d values using thermodynamic and surface complexation modelling, for each mineral in the soil composition one specific surface area (SSA) value has to be used. For the Rophin site, no SSA value is available in the description of the "whitish layer" so far. Since the given example for the Rophin site did not contain specific information on the surface areas of the individual minerals in the soil composition, an SSA value was estimated for each mineral phase. For this purpose, the SSA values from the RES³T sorption database (RES³T, 2025 & Brendler et al., 2003) were used. Values that were falsified by improper sample preparation (e.g. by grinding) were removed and the remaining values were averaged. The mean SSA value was then used for each individual mineral in the modelling (see *Table 4*).

Table 4 – Mean specific surface area values for the mineral phases used in the Rophin site's K_d calculation, mean values from RES³T database (RES³T, 2025 & Brendler et al., 2003)

Mineral	Formula	Mean SSA [m²/g]	No of averaged datasets
Quartz	SiO_2	3.5	73
Albite	$NaAlSi_3O_8$	3.8	9
Orthoclase	$KAlSi_3O_8$	2.4	13
Microcline	$KAlSi_3O_8$	5.4	4
Kaolinite	$Al_2Si_2O_5(OH)_4$	17.0	151
Muscovite	$KAl_3Si_3O_{10}(OH)_2$	6.5	15
Chlorite	$Mg_5Al_2Si_3O_{10}(OH)_8$	2.8	21
Illite	$K_{0.6}Mg_{0.25}Al_{1.8}Al_{0.5}Si_{3.5}°_{10}(OH)_2$	63.3	66
Montmorillonite	$Na_{0.3}Mg_{0.36}Al_{1.67}Si_4O_{10}(OH)_2$	230.7	159
Goethite	FeOOH	59.9	295
Ferrihydrite	$Fe(OH)_3$	272.6	81

1.2.2 Surface Complexation Models

For the sorption modelling, classical thermodynamic data for the reactions in the aqueous phase and for mineral precipitation/dissolution are necessary as well as SCM data for the actual sorption reactions. Additionally, for some minerals like sheet silicates (montmorillonite, illite), ion-exchange parameters for the cation exchange reactions occurring in the interlayer have to be included.

For the thermodynamic database, the PSI Chemical Thermodynamic Database was chosen (Thoenen et al., 2014). This database is comprehensive and consistent in terms of radionuclide/NOR (U, Ra, Pb) aqueous chemistry as well as for the soil minerals needed and it is actively maintained.

1.2.2.1 Collection of SCM data

An SCM dataset consists of the SSA of the sorbing mineral and the surface site density (SSD) being the number of identical binding sites per unit area, *e.g.* 1 site nm^{-2}, for each type of binding site. Furthermore, this also includes the protonation and deprotonation reactions constants ($pK°_{a1,2}$) of the binding site and the sorption reaction constants ($\log K°$) of the NOR and the main competing matrix ions (Mg^{2+}, Ca^{2+}, Al^{3+} & SO_4^{2-}) in solution.

$$\equiv OH_2^+ \rightleftharpoons \equiv OH + H^+ \qquad pK°_{a1}$$

$$\equiv OH \rightleftharpoons \equiv O^- + H^+ \qquad pK°_{a2}$$

$$m \equiv OH + M^{n+} \rightleftharpoons (\equiv O)_m M^{(n-m)+} + m\ H^+ \qquad \log K°$$

A literature search was conducted for SCM data for the required combinations of mineral and sorbing ions, and the data found were critically evaluated. SCM datasets have been compiled for the NORs U, Ra and Pb. For polonium (Po) and bismuth (Bi), no appropriate SCM datasets were available in literature to set up a consistent and complete model that would reflect the sorption on all given minerals (see *Table 1*). Thus, those two NORs were not included in this work. An SCM database for the diffuse double layer (DDL) sorption model was created and formatted according to the "Geochemist's Workbench" (Bethke, 2022) geochemical software requirements.

For Ra sorption, there are still many gaps in the SCM data in the literature, which leads to greater uncertainty in the K_d values obtained. By using chemical analogies both for the sorbing ion (use of barium (Ba) or strontium (Sr) data) and for the sorbing mineral phases (e.g. use of SCM data of other feldspars than orthoclase/albite), attempts were made to close data gaps.

No proper SCM datasets were available for Pb, Ca and Mg onto the feldspar surfaces and for Al onto the goethite surface. The sources of the selected data sets and information on the conversions/normalisations performed are given in Appendix A.

1.2.2.2 Conversion from SCM data file to zero ionic strength

Reaction constants ($\log K$) of surface complexation reactions are usually determined at a constant ionic strength (I) of a certain background electrolyte. To use the collected SCM data in a speciation code (e.g. Geochemist's Workbench), the $\log K$ value had to be converted to the corresponding value at infinite dilution (zero ionic strength). The Davies equation (Eq. 1, 2) was chosen as a simple algorithm to recalculate $\log K$ values (Davies, 1962). This equation is valid up to an ionic strength of I = 0.5 $mol \cdot L^{-1}$ and consists only of the charge (z) and the stoichiometric factor (v) of the reactants (i), ionic strength (I) as well as two fixed Davies parameters (A and b) to calculate the activity coefficient (γ_i).

$$\log \gamma_i = -A \cdot z_i^2 \cdot \left(\frac{\sqrt{I}}{1+\sqrt{I}} - b \cdot I \right) \qquad \text{Eq. 1}$$

$$\log K_0^\circ = \log K_I + \sum_{i=1}^{n} \nu_i \cdot z_i^2 \cdot \left[-A \cdot \left(\frac{\sqrt{I}}{1+\sqrt{I}} - b \cdot I \right) \right] \qquad \text{Eq. 2}$$

Note that the surface species are *not* considered within the $\Sigma v_i \cdot z_i^2$ term. The Davies parameters A and b are set to:

- $A_{298.15\ K}$ = 0.5093 $kg^{½} \cdot mol^{-½}$
- $b_{298.15\ K}$ = 0.3 $kg^{½} \cdot mol^{-½}$

1.2.2.3 Standardization of SCM data with regard to the site-density used

The SSD value corresponds to the number of binding sites on a mineral surface and is usually given in sites per nm². Different SSDs are often used for one mineral in the literature on the sorption of different NORs. Furthermore, different types of binding sites are assumed in the literature – some use only one type of binding site (*e.g.* ≡SiOH), others use several types of binding sites (strong and weak site concept, *e.g.* ≡(s)SiOH & ≡(w)SiOH) with different site density values each. Within one geochemical calculation, however, all sorption reactions for one single mineral must refer to a mineral-specific SSD value in order to achieve consistency. The reference site density was set to 2.31 nm^{-2} (Davis et al., 1990) for all minerals. In case of models including strong and weak binding sites, the sum of the site densities is set to this value. According to Kulik (2006) surface complexation constants are normalized with equation (3), where *n* is the stoichiometric number of surface binding sites in the reaction equation of the surface reaction and Γ_{exp} the binding site density as reported in the original reference.

$$\log K_{normalized} = \log K_{exp} + n \cdot \log \frac{\Gamma_{exp}}{2.31} \qquad \text{Eq. 3}$$

1.2.2.4 Collection of cation exchange data

Cation exchange (IEx) is a process that only occurs in a few of the mineral phases analysed: montmorillonites and illites. Therefore, parameters for calculating the ion exchange only had to be collected for these two phases. The references of the compiled IEx parameters are given in Appendix B. No proper IEx datasets were available for Mg on montmorillonite.

1.3 Results

1.3.1 Rophin site

A small smart-K_d matrix was created for the Rophin site, in which most of the geochemical parameters were varied within the scope of the measurement uncertainty. Only the E_H range and the carbonate concentration were varied over larger ranges (see *Table 5*), as these parameters have the greatest influence on the chemistry of U and thus the retention through sorption or precipitation. This also applies to the elements Ra and Pb, albeit to a lesser extent. The ranges of parameter variation are sufficient for a consideration of the geochemical conditions in the “whitish layer” in the sediment of the wetlands downstream of the tailings of the former uranium mine.

Table 5 – Ranges of geochemical parameter variation in the smart-K_d calculation for the Rophin site

Parameter	Range	Step-width
pH	5.2 – 5.3	0.05
E_H	−200 – 300 mV	10 mV
Ca^{2+}	101 – 103 µmol/L	1 µmol/L
Mg^{2+}	35 – 37 µmol/L	1 µmol/L
SO_4^{2-}	107 – 113 µmol/L	1 µmol/L
DIC	1 – 101 µmol/L	10 µmol/L

All combinations of the parameters in these ranges result in a number of 106,029 individual K_d calculations with a central processing unit (CPU) time of approx. 13 hours.

1.3.1.1 Uranium

The K_d values for U vary over a range of ~5 orders of magnitude ($\log_{10}K_d$ = −2 – 3) but with an asymmetric shape in the histogram. This effect is caused by two processes: the change in the U redox state (+IV/+VI) during the E_H variation and the strong pH and DIC concentration's dependency of the U(VI) sorption. Tetravalent U is immobilized by the formation of sparsely soluble hydroxide phases ($U(OH)_4$(am) or UO_2(cr)) while the hexavalent U is much more mobile in the environment under the given geochemical conditions. At the actual Rophin site conditions, the slightly acidic pH value prevents stronger sorption of dissolved U on mineral surfaces, but also the formation of highly negatively charged U(IV) and U(VI) carbonato complexes, which would increase U mobility at higher pH values.

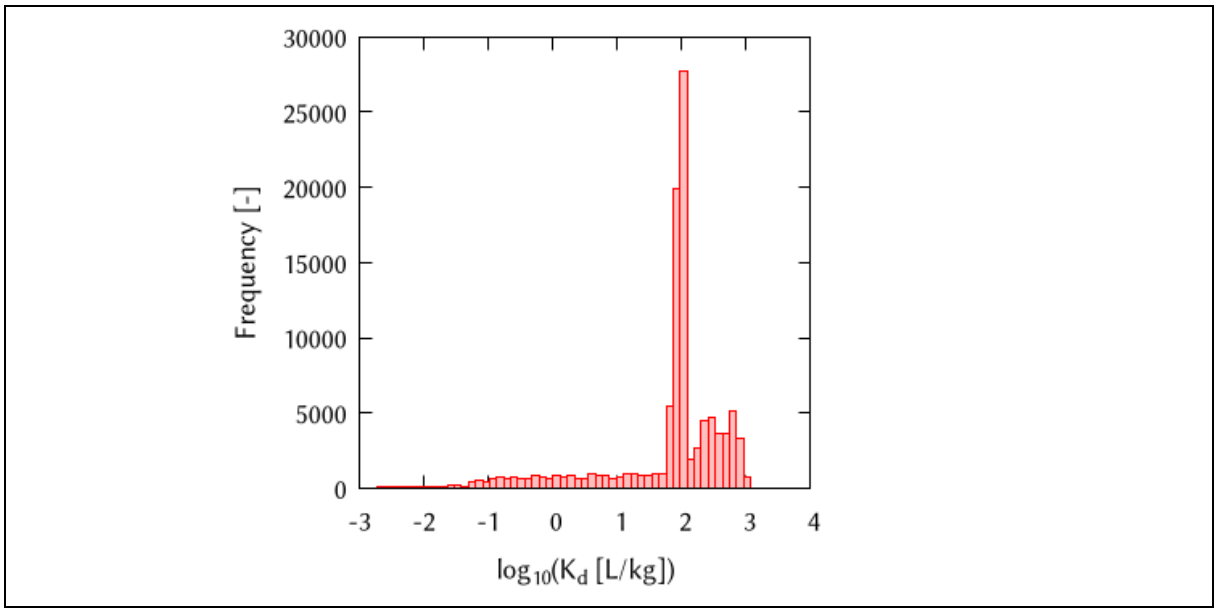


Figure 1 – Distribution of K_d values for uranium calculated for the "Whitish Layer" of the wetland downstream of the tailings of Rophin site

The K_d value calculated for the exact geochemical conditions (E_H = 261 mV) in the "Whitish Layer" of the soil downstream of the tailings of the Rophin site was K_d^{calc} = 66 to 98 L/kg ($\log_{10}K_d^{calc}$ = 1.82 to 1.99), which is in very good agreement with the experimentally determined value of K_d^{exp} = ~50 L/kg ($\log_{10}K_d^{exp}$ = 1.70). The results are currently under review in the publication by Montavon et al. (2025).

1.3.1.2 Radium

The K_d values for Ra vary within a small range ($\log_{10}K_d$ = 1.85 – 2.2), see *Figure 2*. This is caused by the uniform chemistry of Ra that neither undergo redox transformation nor changes in the aqueous speciation within the varied geochemical conditions. The slightly acidic pH conditions also prevent the formation of Ra-containing carbonate phases. Immobilisation therefore occurs mainly by the formation of insoluble sulphate mineral phases and to a lesser extent, due to the acidic pH, by sorption of Ra ions on mineral surfaces.

A major source of uncertainty in the calculated K_d values for Ra is the lack of Sr and Ba concentration data in the Rophin site field data. As a result, the incorporation of Ra in poorly soluble solid phases such as baryte ($BaSO_4$) or celestine ($SrSO_4$) or the formation of a solid solution from the sulphates of the three alkaline earth metal ions (Klinkenberg et al., 2018 & Brandt et al., 2020) could not be included in the calculation. This would reduce the Ra concentration in solution by ~3 orders of magnitude compared

to the calculated values in equilibrium with $RaCO_3$(cr), which will not form in nature (Grandia et al., 2008).

The same applies to the incorporation of Ra in whiterite ($BaCO_3$) or strontianite ($SrCO_3$), but in this case, it is of minor importance as the low pH value prevents the formation of carbonate solid phases.

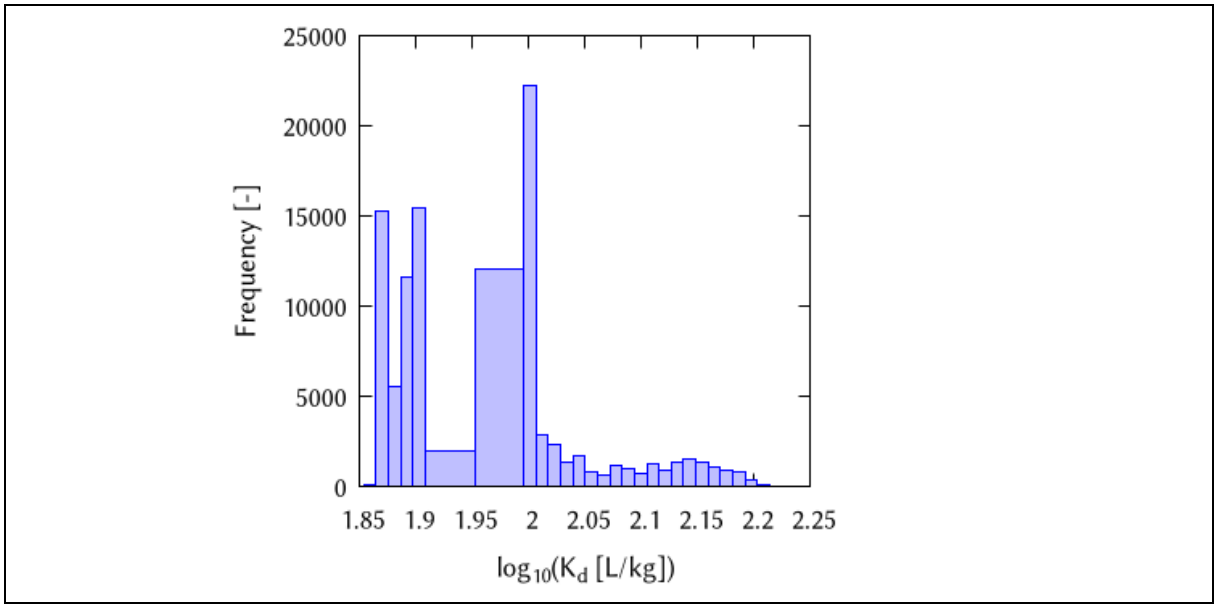


Figure 2 – Distribution of K_d values for radium calculated for the "Whitish Layer" of the wetland downstream of the tailings of Rophin site

1.3.1.3 Lead

The K_d values for Pb vary within a small range ($\log_{10}K_d$ = 1.55 – 1.9), see *Figure 3*.

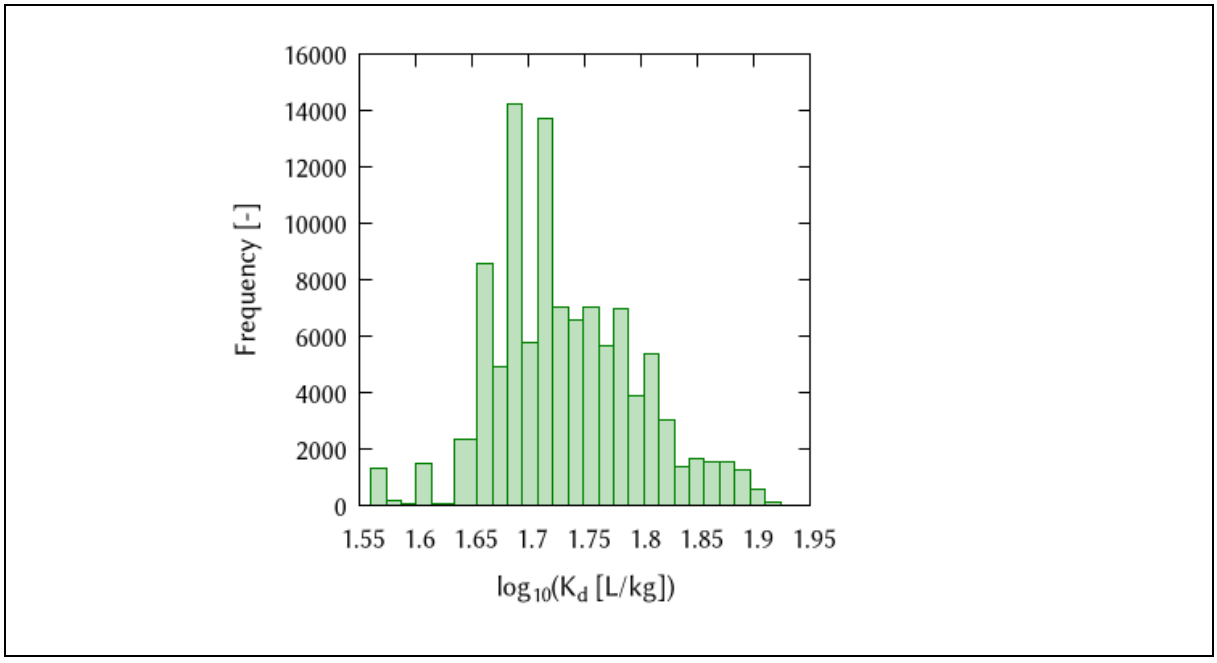


Figure 3 – Distribution of K_d values for lead calculated for the "Whitish Layer" of the wetland downstream of the tailings of Rophin site

Like in the Ra case, the sorption of Pb on the mineral phases is rather weak due to the acidic pH value.

Unfortunately, the authors could not be provided with measured field K_d values for Ra or Pb to evaluate the K_d values calculated using the smart-K_d approach.

1.3.1.4 Sources of uncertainties

For U(VI), the effect of inorganic carbonate in the presence of alkaline earth metal ions can be estimated as follows. With increasing carbonate content, the formation of easily soluble and less sorbing alkaline earth uranyl(VI) tricarbonato complexes significantly increases the mobility of U(VI). The formation of negatively charged U(IV) carbonato complexes also increases the solubility of U(IV) and reduces sorption in slightly alkaline media (pH > 8). Carbonate can lead to the formation of poorly soluble alkaline earth carbonate phases, which reduce mobility in alkaline media by incorporating radium as a solid solution. The formation of sparingly soluble Pb carbonate significantly reduces the mobility of Pb at pH values >7.

Another source of uncertainty in the K_d calculation is the lack of specific surface area for each mineral in the soil. The SSAs were therefore estimated from mean values of the surface area measurements available in the literature. A complicating factor is that it is not known which part of the mineral surfaces are accessible to the aqueous solution or are in principle inaccessible for sorption reactions due to intergrowth of the particles or coatings with secondary phases e.g. ferrihydrite or goethite. This impact cannot be determined and therefore cannot be taken into account in the calculation of the K_d values. Further research is required here with regard to sampling and SSA determination.

1.4 Earth alkaline metal ion sorption on gibbsite (Ph.D. work)

1.4.1 Introduction

Radium-226 (^{226}Ra), a naturally occurring radionuclide from the U decay chain, is of considerable environmental concern due to its potential mobility in soils and groundwater systems. Understanding its interaction with mineral surfaces is critical for assessing its environmental behaviour and for improving predictive models used in risk assessment and waste disposal strategies. However, due to its radioactivity and low environmental concentrations, direct experimental studies with Ra are complex and limited. Therefore, chemically similar elements such as strontium (Sr) and barium (Ba) are commonly employed as analogues in sorption studies, as they share comparable geochemical characteristics, including charge (+2), ionic radii, and hydration behaviour.

A central challenge in modelling Ra retention in natural systems lies in the mineralogical complexity of environmental matrices. At the Rophin site in France, for instance, over 90% of the soil composition consists of quartz and feldspars (Grangeon et al., 2023 & Montavon, 2021, see chapter 1.1.1.1). Yet, experimental sorption data for these mineral phases – particularly for quartz and feldspar (albite and microcline) – remain sparse. *Table 6* highlights the existing data gaps for Ra sorption on these mineral phases. While some information exists for clays such as smectite and kaolinite, limited data are available for Ra sorption on quartz and feldspar (microcline, albite) and muscovite.

Since quartz and feldspar together account for over 90% of the soil composition (54% quartz, 38% feldspar), understanding their sorption behaviour is crucial for reliable predictive models.

Table 6 – Quantitative mineralogy of the Rophin ICPE site (average value of samples from Grangeon et al. (2023) with sorption data of ^{226}Ra (or chemical analogues) onto the respective mineral (data from RES³T database (RES³T,2025)).

Mineral	wt-%	^{226}Ra sorption data	Available data for chemical analogues	Reference
Quartz	54.26	None	Ca onto quartz	Sverjensky, 2006
Microcline	22.58	None	Sr onto orthoclase	Britz, 2018
Albite	16.74	None	Sr onto orthoclase	Britz, 2018
Smectite	2.42	✓	Ra onto smectite	Tachi et al., 2001
Kaolinite	1.24	✓	Ra onto kaolinite	Riese, 1982
Muscovite	2.76	None	Sr onto muscovite	Hu et al., 2017

Feldspars and aluminosilicates possess two dominant types of surface functional groups: silanol (≡SiOH) and aluminol (≡AlOH). To deconvolute their individual contributions to metal ion retention, a systematic and mechanistic stepwise investigation is pursued. In this framework, gibbsite ($Al(OH)_3$) is studied as a model for aluminol sites, quartz (SiO_2) for silanol sites, and albite ($NaAlSi_3O_8$) as a representative feldspar containing both types of functional groups. This mechanistic breakdown provides the foundation for future integration into surface complexation models, ultimately supporting more accurate predictions of Ra mobility.

This study builds upon prior work, such as Katz et al. (2013), which explored the sorption of alkaline earth metals (including Sr(II) and Ba(II)) on gibbsite. However, their study focused on temperature effects and inner vs. outer-sphere complexation, while this work extends the understanding of pH-dependent sorption and the impact of ionic strength.

1.4.2 Experimental approach

1.4.2.1 Batch sorption experiments

Batch sorption experiments were performed to investigate the uptake of Sr(II) and Ba(II) on gibbsite (Gbs) and quartz under varying pH conditions (see *Table 7* for specific details), different background electrolyte concentrations, and at various initial sorptive concentrations. For gibbsite, both concentration-dependent (isotherm) and pH-dependent sorption experiments were conducted, alongside solid-liquid ratio variation. Quartz was investigated under similar conditions (see *Table 7*) for initial pH-dependent sorption screening at 0.01 mol·L^{-1} NaCl. All experiments were performed in a N_2-filled glove box to maintain CO_2-free conditions and prevent carbonate formation. Pre-equilibration of the solid phase with electrolyte was performed for 72 h before the addition of Sr(II) or Ba(II), followed by an additional 72 h for sorption equilibration. Supernatants were analyzed by ICP-MS to quantify residual metal concentrations.

Table 7 – Experimental conditions for the batch sorption experiments for the sorption of Ba and Sr onto gibbsite, quartz and albite (all experiments were performed at 22 ± 2 °C).

Experiments	m/V ratio	Pre-equilibration time	Equilibration time (sorption)	pH	$BaCl_2$	$SrCl_2$	Ionic strength
	[g/L]	[h]	[h]	[-]	[mol/L]	[mol/L]	[mol/L]
Time dependent (pre-equilibration)	0.5	72	-	11	-	-	0.01 and 0.1
Time dependent (sorption)	0.5	72	72	11	10^{-6}	10^{-6}	0.01 and 0.1
pH dependent	0.5 and 4.0	72	72	8 to 12	10^{-6}; $2.5{\cdot}10^{-5}$ and $5{\cdot}10^{-5}$	10^{-6}; $2.5{\cdot}10^{-5}$ and $5{\cdot}10^{-5}$	0.01 and 0.1
Isotherm	0.5	72	72	9.5 and 11	10^{-8} to 10^{-4}	10^{-8} to 10^{-4}	0.01
EXAFS	4.0	72	72	9 to 12	$5{\cdot}10^{-5}$	$5{\cdot}10^{-5}$	0.01 and 0.1
Zeta potential	0.5	72	72	3 to 12	0	0	0.005 and 0.01
Zeta potential with sorptive	0.5	72	72	7 to 12	$2.5{\cdot}10^{-5}$	$2.5{\cdot}10^{-5}$	0.01
Zeta potential with sorptive	0.5	72	72	7 to 12	$2.5{\cdot}10^{-5}$	$2.5{\cdot}10^{-5}$	0.01
Quartz – SiO_2							
pH dependent	30	72	72	8 to 12	10^{-5}; 10^{-4} and 10^{-3}	-	0.01
Albite – $NaAlSi_3O_8$							
Time dependent (pre-equilibration)	3	72	-	9 and 10	-	-	0.01 and 0.1
Time dependent (sorption)	3	72	72	9 and 10	10^{-6}	10^{-6}	0.01 and 0.1
Zeta potential	0.5	72	72	2 to 12	0	0	0.01

1.4.2.1 Surface charge measurements (Zeta potential)

Surface charge behaviour was assessed by zeta potential measurements to determine the isoelectric point (IEP) and to evaluate the impact of sorption on surface electrostatics. For gibbsite, these measurements were performed across a broad pH range (pH 3–12) at two ionic strengths (0.005 and 0.01 mol·L^{-1} NaCl) and with or without sorptive present. Albite was investigated analogously at 0.01 mol·L^{-1} NaCl and is currently undergoing evaluation at 0.005 mol·L^{-1} NaCl.

1.4.2.2 EXAFS spectroscopy

EXAFS spectroscopy was used to probe the local coordination environment of Sr(II) and Ba(II) sorbed onto gibbsite. This method helped to distinguish between inner-sphere (direct bonding) and outer-sphere (electrostatic) complexation mechanisms. X-ray absorption spectroscopy (XAS) was performed at the Rossendorf Beamline II (ROBL-II) ((Scheinost et al., 2021) of the European Synchrotron Radiation Facility (ESRF) in Grenoble (France). The experimental conditions are listed in *Table 7*.

1.4.3 Results and Discussion

1.4.3.1 pH-dependent sorption behaviour

The batch sorption experiments were conducted to investigate the sorption behaviour of Sr(II) and Ba(II) on gibbsite under controlled conditions. The equilibrium time for sorption was determined through time-dependent batch experiments, where Sr(II) or Ba(II) sorption reached 80% after approximately 24 h. To ensure complete equilibration, a conservative equilibration time of 72 h was selected for all further experiments. The same observation was made for pre-equilibration of gibbsite, using time-dependent experiments. So 72 h was set as a pre-equilibration time for gibbsite.

The pH-dependent sorption experiments were performed at different initial concentrations of Sr(II) and Ba(II), varying the mass-to-volume ratio and ionic strength (*Figure 4A-4D*). The sorption edges are shown in *Figure 4B* and *4D*. In both cases (Sr(II) and Ba(II)) the sorption starts between pH 8 and 9 and is not reaching a maximum at pH 12.The sorption shows strong pH dependency. The sorption of $1{\cdot}10^{-6}$ mol·L^{-1} Sr(II) or Ba(II) with 0.01 mol·L^{-1} NaCl as background electrolyte exhibited a negligible sorption below pH 8 and a rapid increase above pH 9. For Sr(II) (black points, *Figure 4A*), a steep sorption edge was observed between pH 9 and 11, reaching nearly complete sorption at pH 12. A similar trend was observed for Ba(II) (black points, *Figure 4C*), although sorption initiated at slightly lower pH values compared to Sr(II), indicating differences in affinity for gibbsite surfaces. For Sr(II) (yellow points in *Figure 4A*), the sorption starts below pH 8 and reaches a maximum at pH 10. The yellow points (mass-to-volume ratio of 4 g·L$^{-1)}$ are observed to run parallel to the black points (mass-to-volume ratio of 0.5 g·L$^{-1)}$. A comparable trend was observed for the Ba(II) data (yellow points, *Figure 4C*), though the sorption maximum appears to be reached at a slightly higher pH (10.5) than for Sr(II).

The influence of ionic strength on Sr(II) and Ba(II) was studied with batch sorption samples containing $1{\cdot}10^{-6}$ mol·L^{-1} Sr(II) or Ba(II), 0.5 g·L^{-1} gibbsite and 0.1 or 0.01 mol·L^{-1} NaCl. The effect of the ionic strength can be seen in *Figure 4A* and *4C* by comparing the red and black points. This effect of ionic strength cannot be seen in the results of the sorption experiments for Sr(II) or Ba(II) of Katz et al. (2013). They worked, especially in the case of sorption experiments with Ba(II), with similar conditions, but used $NaNO_3$ instead of NaCl. The source of the influence must therefore be in the different electrolyte.

Sorption isotherms were measured at pH 9.5 to determine the relationship between the equilibrium concentration of Ba(II) in solution and the amount sorbed on gibbsite (*Figure 4E*). At low concentrations (<$1{\cdot}10^{-6}$ mol·L^{-1} Ba(II)) sorption increases linearly. An explanation for that could be that Ba(II) as a sorptive occupies available adsorption sites without reaching a saturation. The data followed a linear trend, suggesting no weak or strong sorption sites on the gibbsite surface. At higher concentrations (>$1{\cdot}10^{-6}$ mol·L^{-1} Ba(II)) the sorption increase slows slightly, because a saturation or equilibrium between sorbed and free Ba(II) was reached.

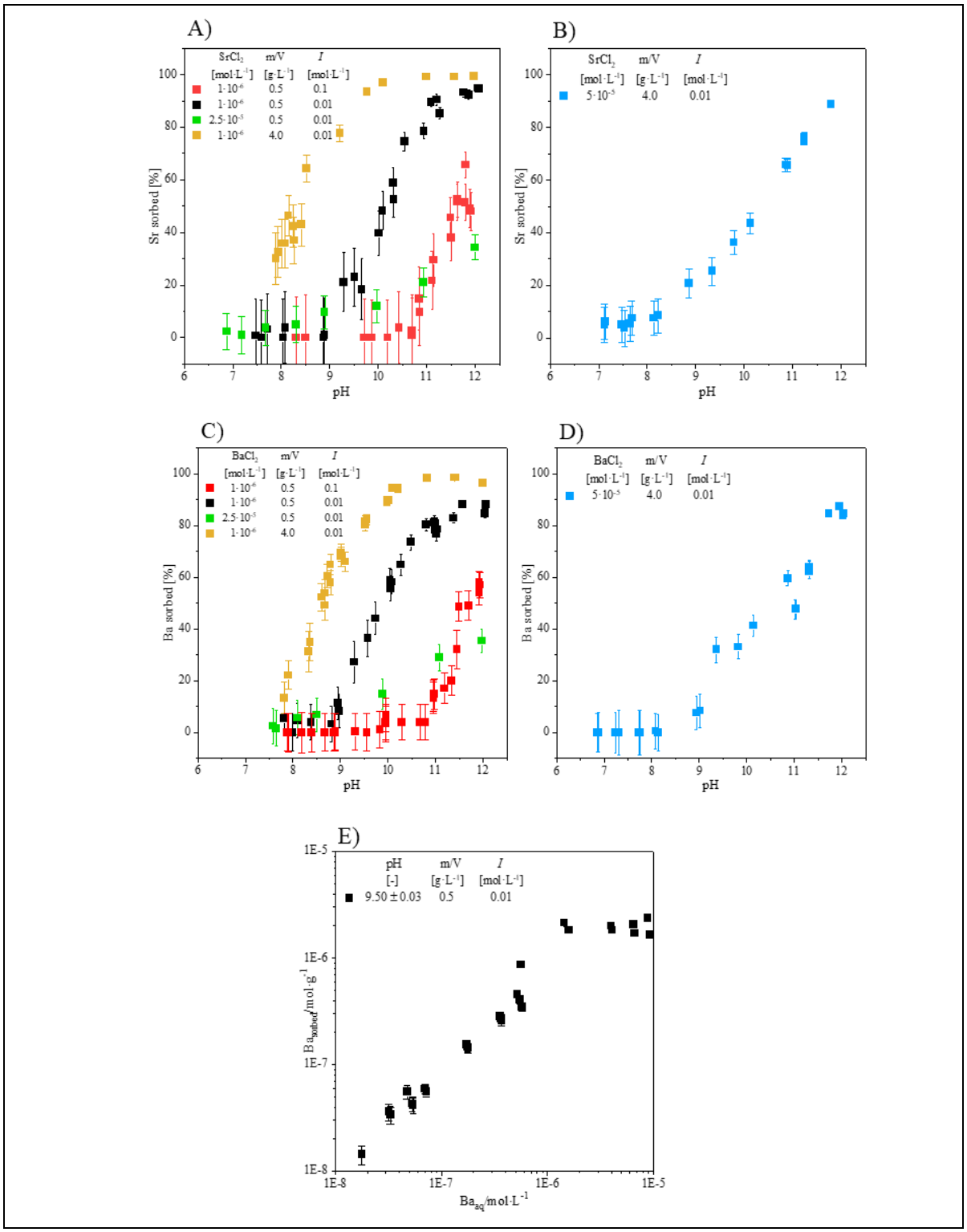


Figure 4 – A) Sorption curve of Sr(II) and C) Ba(II) on gibbsite under N_2(g) at different concentrations, mass to volume (m/V) ratio and ionic strength. B) Sr(II) and D) Ba(II) sorption edges of EXAFS samples. E) Sorption isotherm of Ba(II) at pH 9.5 ± 0.03. All concentrations are initial ones of the sorptive (error bars represent standard deviations at each point).

In addition to experiments on gibbsite, first batch sorption data for Ba(II) on quartz were collected (*Figure 5*). The experiments were carried out at 30 g·L^{-1} quartz with 0.01 mol·L^{-1} NaCl and Ba(II) concentrations of 1·10^{-5}, 1·10^{-4}, and 1·10^{-3} mol·L^{-1}. All data show pH-dependent sorption behavior, with the sorption

edge shifting to higher pH values as Ba(II) concentration increased. This observation is consistent with limited sorption site density and increasing competition at higher Ba(II) loads. Despite the high solid content, maximum sorption remained below 100%, supporting the assumption of weak, electrostatically driven interactions on quartz surfaces and potentially limited inner-sphere complexation.

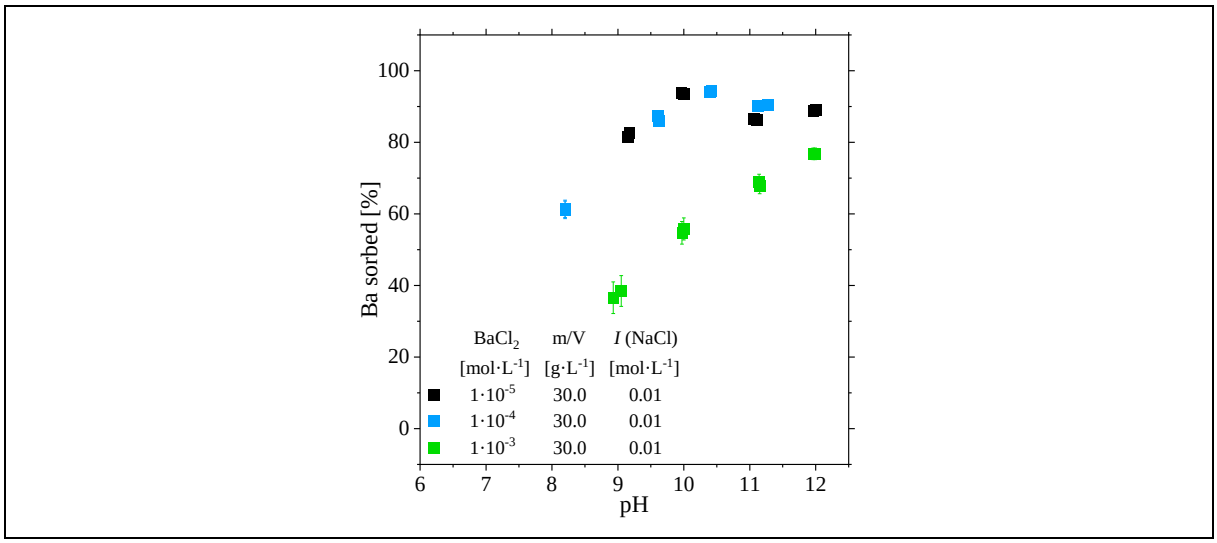


Figure 5 – Sorption curve of Ba(II) on quartz under N_2(g) at different concentrations (error bars represent standard deviations at each point).

1.4.3.2 Zeta Potential Measurements

Zeta potential measurements of gibbsite revealed an isoelectric point (IEP) at pH 8.1, which remained unchanged across different ionic strengths (0.005 and 0.01 mol·L⁻¹ NaCl). Upon the addition of $2.5 \cdot 10^{-5}$ mol·L⁻¹ Sr(II) or Ba(II), the surface potential of gibbsite shifted toward more positive values, particularly at pH > 8, indicating a reduction in surface charge. This behavior is consistent with outer-sphere complexation dominated by electrostatic interactions, rather than inner-sphere complex formation. Outer-spheric complexation of earth alkaline metal ions is known from other (hydr-)oxide mineral phases as well (Axe et al., 1998, 2000, Nie et al., 2017).

In comparison, zeta potential measurements of albite in 0.01 mol·L⁻¹ NaCl show no distinct isoelectric point within the investigated pH range of 2–12 (*Figure 6*). The surface remains negatively charged across all pH values and becomes increasingly negative with rising pH, reaching values below −70 mV at pH 11.

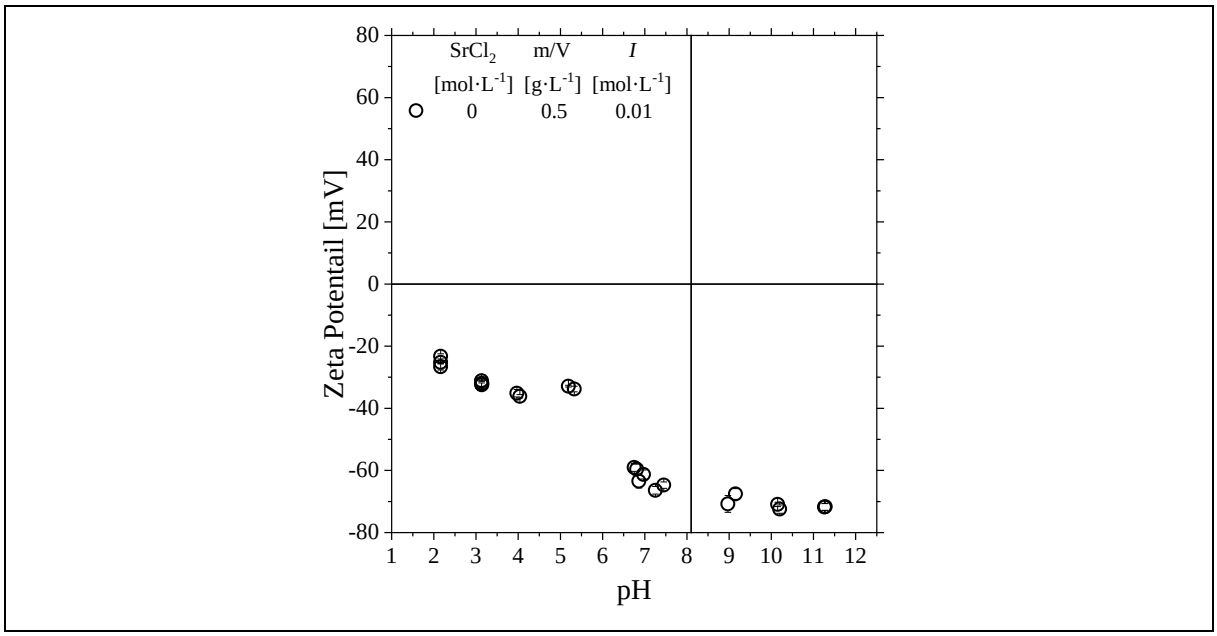


Figure 6 – Zeta potential of albite with 0.01 mol·L⁻¹ NaCl

1.4.3.3 EXAFS Analysis

EXAFS spectroscopy was used to investigate the local coordination environment of Sr(II) and Ba(II) sorbed on gibbsite at different pH values and NaCl concentrations (*Figure 7*, *Table 8*). The EXAFS spectra of Sr-Gbs do not show any detectable backscattering from Al atoms. For Sr(II) sorption, the coordination number (CN) was determined to be 7.1 at pH 9 and 6.7 at pH 12, with Sr-O bond distances of 2.58 Å and 2.57 Å, respectively. These values are in agreement with the findings of Pappalardo et al. (2021), who reported a CN of 8 and a bond length of 2.57 Å for hydrated Sr(II).

For Ba-Gbs, the EXAFS results are less distinct due to the limited k-range available for analysis (data truncated at k = 7 Å^{-1}). Despite this limitation, the shell-fitting analysis suggests that Ba(II) retains a CN of 6.8 at pH 9 and 7.3 at pH 12, with Ba-O bond distances of 2.90 Å and 2.78 Å, respectively. These values are in reasonable agreement with the hydration parameters of Ba(II) in solution, where a CN between 9 and 10 and a bond length of 2.81 Å have been reported (Pappalardo et al., 2021). The slight discrepancy in CN values could be attributed to experimental limitations and the reduced resolution of the Fourier-transformed EXAFS spectra. Due to the absence of backscattering with Al atoms, the formation of outer-sphere complexes can be concluded. This is in good agreement with the results of the zeta potential measurements.

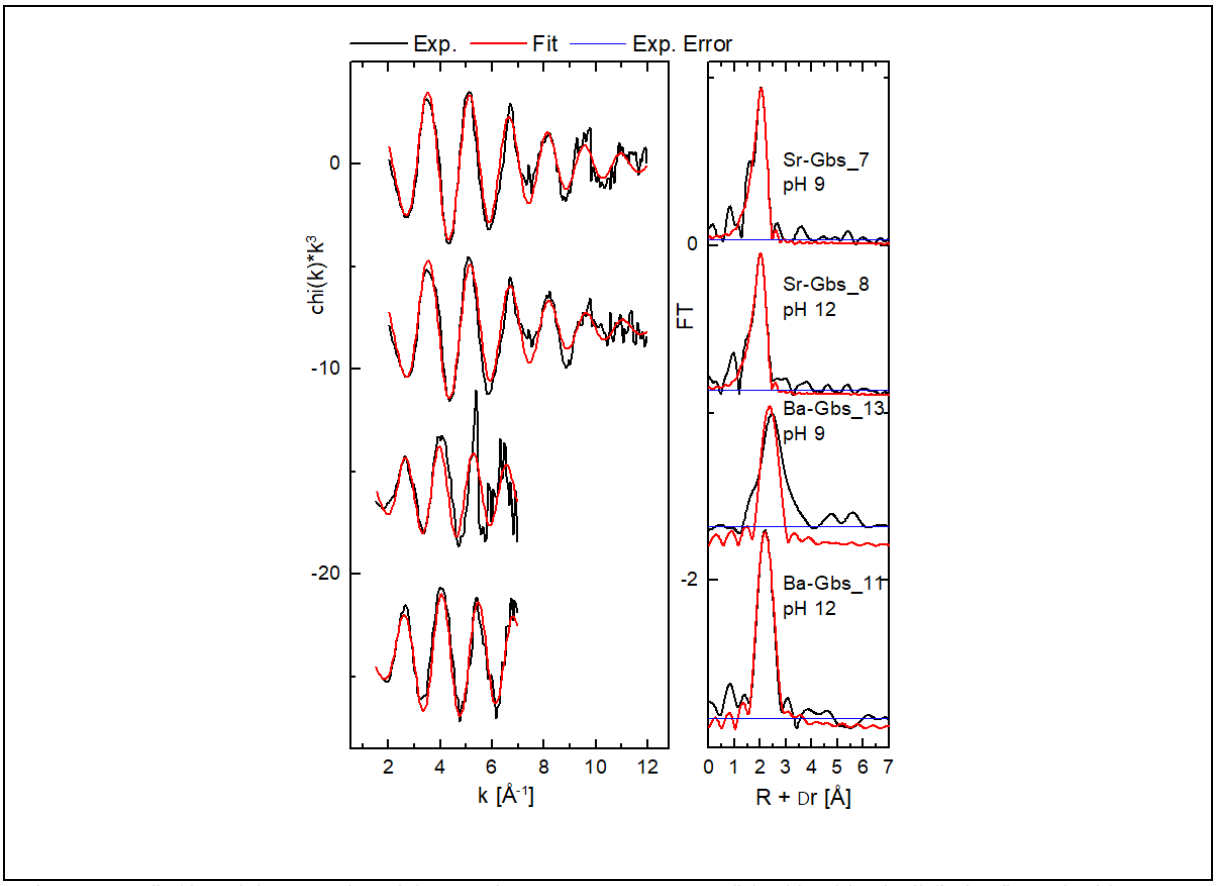


Figure 7 – (left) Sr(II) K- and Ba(II) L_{III}-edge EXAFS spectra (black) with shell fit (red) and with corresponding Fourier-Transform (FT) (right). Estimated level of experimental error (blue). The corresponding sorption curve of Sr(II) and Ba(II) on gibbsite can be seen in Figure 4. The initial concentration for each sorptive is $5 \cdot 10^{-5}$ mol·L^{-1} with an ionic strength of 0.1 or 0.01 mol·L^{-1} NaCl

Table 8 – Sr-O and Ba-O bond lengths. Gbs – gibbsite, CN - coordination number, R - radial distance, DW - Debye-Waller factor, dE_0 - shift in energy threshold. Estimated standard deviations of the variable parameter in parenthesis. $S_0^2 = 0.9$.

Sample	I [mol·L^{-1}]	pH	Path	CN	R [Å]	DW [Å²]	dE_0 [eV]	Sorptive sorbed [ppm]
Sr-Gbs_7	0.01	9	Sr-O	7.1(3)	2.584(3)	0.0102(5)	4.2(3)	356.9
Sr-Gbs_8	0.1	12	Sr-O	6.7(3)	2.569(4)	0.0109(6)	3.8(3)	1343.7
Ba-Gbs_13	0.01	9	Ba-O	6.0(8)	2.90(1)	0.012(3)	9.0(3)	401.5
Ba-Gbs_11	0.1	12	Ba-O	7.3(4)	2.778(5)	0.010(1)	5.6(2)	1053.9

1.4.4 Conclusion

This study provides new insights into the sorption behaviour of Sr(II) and Ba(II) onto gibbsite, demonstrating sorption dominated by electrostatic interactions, confirming outer-sphere complexation. The sorption processes are strongly pH-dependent and show sensitivity to ionic strength, particularly when NaCl is used as the background electrolyte. EXAFS results confirm that Sr(II) retain their hydration shells upon sorption onto gibbsite, further supporting the hypothesis of electrostatic sorption mechanisms. These findings contribute to a mechanistic understanding of Ra analogues' interactions with aluminol sites and will be extended to silanol sites in future research.

The absence of an IEP indicates that albite surfaces remain consistently electronegative in NaCl solution, which may affect sorption behavior differently than in the case of gibbsite and quartz.

1.4.5 Outlook

Future research should focus on sorption experiments of Ra on gibbsite, on performing titration experiments with gibbsite to determine a pK_a value to enable thermodynamic modelling, and on comparative studies with other minerals (such as quartz or orthoclase) to better predict the behaviour of Ra in complex natural environments. These insights contribute to a deeper understanding of radionuclide mobility, with significant implications for remediation of sites contaminated with NOR and radioactive waste management.

Future efforts will focus on extending sorption experiments to quartz, where pH-dependent sorption studies under different ionic strengths will be conducted to better understand sorption onto silanol sites. For albite, zeta potential measurements will be completed at higher ionic strength, followed by batch sorption experiments under similar conditions as for gibbsite and quartz. These studies will contribute to a more comprehensive understanding of how mineral surface chemistry controls the retention of Sr(II), Ba(II), and Ra(II), ultimately enhancing our ability to predict radionuclide mobility in complex environmental matrices.

2. Parametric models and probabilistic distribution functions

2.1 Construction of parametric models and probabilistic distribution functions to predict and derive soil K_d (NOR)

2.1.1 Parametric models: statistical analysis of the relationships between K_d (NOR) and soil properties

2.1.1.1 Setting up the K_d (NOR) datasets

The construction of parametric models was based on datasets generated from sorption and desorption experiments carried out within the RadoNorm project and, when necessary, on extended datasets that also included data obtained from a thorough review of published studies. For the extended datasets, new K_d (NOR) data were accepted if they came from similar experimental approaches (i.e., batch experiments with similar soil-solution ratios and equilibration times). The entries were carefully analysed to meet conditions similar to those used in our own experiments to determine the K_d (NOR) parameter. First, the entries were filtered by experiment type, excluding column experiments and diffusion tests. Second, only studies in soils were considered, excluding sorption experiments on pure mineral phases. Finally, the data were filtered to maintain a liquid-solid ratio similar to that used in our experiments.

The experimental approach used to obtain the K_d (NOR) values is in this context a broad term that allowed the data to be classified according to the source term and time elapsed since NOR incorporation, thus defining short-term sorption, short-term desorption, and long-term datasets. Short-term sorption K_d (NOR) values (ST-S) corresponded to batch experiments performed by adding aqueous solutions with known NOR concentrations and equilibrating them with the solid sample for a given time. Short-term desorption K_d (NOR) values (ST-D) corresponded to batch desorption experiments of freshly contaminated samples. Long-term K_d (NOR) values (LT) were generally obtained by the ratio of the concentration of native NOR in the solid phase to that of the aqueous solution in contact with it.

Regarding the soil collection used in the frame of the RadoNorm project, the contents of sand, silt, clay and calcium carbonate ($CaCO_3$) were analysed in the solid phase of the soil samples, as well as the amorphous iron and manganese content (Fe_{am} and Mn_{am}, respectively), organic carbon content, the SSA, the exchangeable sodium, potassium, calcium and magnesium ($(Na+K)_{exch}$ and $(Ca+Mg)_{exch}$) contents, and the cation exchange capacity (CEC). The contact solution was also characterised in terms of pH, dissolved inorganic and organic carbon (DIC and DOC, respectively) content, water-soluble sodium, potassium, calcium and magnesium ($(Na+K)_{ws}$ and $(Ca+Mg)_{ws}$) contents, and the contents of water-soluble anions (Cl^-_{ws}, Br^-_{ws}, $NO_2^-_{ws}$, $NO_3^-_{ws}$, $SO_4^{2-}_{ws}$ and $PO_4^{3-}_{ws}$). Further details on soil characteristics can be found in section 3 of the deliverable 2.16 (Vidal et al., 2024).

2.1.1.2 Principal component analyses (PCA) and partial least squares (PLS) regression analysis

PCA was applied to reduce the dimensionality of the dataset while preserving the most significant variance. PCA transforms the original correlated variables into a set of uncorrelated components, known as principal components, which explain the maximum variance in the dataset. This allows for a clearer visualization of relationships and patterns that may not be immediately apparent. For exploratory and classificatory analysis, a score plot was constructed using an X matrix that included all soil properties, ensuring no redundant information, along with the response variable, here K_d (NOR). When the value of a given soil property was lower than the limit of quantification (LQ), the value inserted was the LQ divided by four. Relationship and correlation degrees between variables was further analysed through a loading plot, which helped to screen out less relevant properties by examining their correlation and orthogonality with K_d (NOR), thereby reducing the number of variables for subsequent PLS analyses. Data were log10-transformed (except for pH) and autoscaled prior to PCA to standardize the variables

and mitigate any discrepancies in their magnitude and amplitude, ensuring that all variables contributed equally to the analysis.

A partial least squares (PLS) regression was performed to explore the relevant soil properties affecting K_d (NOR) by building a low-dimensional space composed of latent variables (LVs). LVs are constructed from linear combinations of the predictor variables (here, soil properties) to maximise the covariance explained by the predictor and response variables (here, K_d (NOR)). The PLS regression analysis was performed with an X data matrix of samples and soil properties (thus derived from the previous PCA) and a Y data matrix of the K_d (NOR) values. The original variables were first log10-transformed (except for pH) and autoscaled to ensure they presented the same weight in the model.

The statistics and indexes provided by the PLS regression, such as Hotelling's T2, Q residuals, and the variable importance in projection (VIP) scores, allowed us to identify the key soil properties correlating with K_d (NOR) that were going to be considered in further statistical analyses. Hotelling's T2 is a generalisation of the Student's t-test statistic and indicates the similarity of the soil samples across the multivariate space (Hotelling, 1931). The Q residual indicates the difference between the observed K_d (NOR) and the value predicted with the soil properties forming the PLS regression model. The VIP score is the index for each soil variable resulting from the sum, over LV, of the PLS values weighted by the percentage of the explained K_d (NOR) variance by each LV (Cocchi et al., 2018).

The reliability of the derived statistics and the optimum LV from the PLS regression analysis was checked with a cross-validation step of the k-fold venetian blinds method, evaluating the root-mean-square (RMSE) and root-mean-square of cross-validation (RMSECV) errors. The PCA and PLS regression analysis were performed with the PCA and PLS Toolbox 652 of MATLAB R2009a (MathWorks, Inc., USA).

2.1.1.3 Linear regression analysis

The K_d (NOR) data were subjected to univariate linear regression (ULR) using the least-squares method to assess their mathematical relationship with the soil properties and identify the properties that best described the K_d (NOR) data variance for the development of the K_d (NOR) prediction models. Multivariate linear regression (MLR) analyses were then performed by combining the most relevant soil properties identified by the ULR, following the principle of explaining the maximum K_d (NOR) data variance with the minimal number of parameters. The empirical equations obtained that best explained the K_d (NOR) data variance (highest regression coefficient) and had a p-value < 0.05 were used as prediction models to estimate the K_d (NOR) values in soils based on the soil properties selected.

In some cases, model selection was performed using the *glmulti* package in R (version 4.4.1), where models were evaluated according to an information criterion approach, such as the corrected Akaike Information Criterion (AICc). This allowed for the comparison and validation of different regression models build from the relevant soil properties identified through the PLS analysis, as all candidate models are ranked based on their AICc values. This criterion assesses the relative quality of statistical models, balancing model fit with complexity by penalizing models with more parameters, besides correcting the bias that can occur when the number of observations is low relative to the number of model parameters, which allows for the determination of the free parameters of a statistical model (Akaike, 1974 & Aldrich, 1997). The AICc is defined as follows:

$$AICc = 2k - 2\ln(L) + \frac{2k \cdot (k+1)}{n-k-1}$$

where *k* is the number of parameters in the model, *ln(L)* is the logarithm of the maximum likelihood, and *n* is the number of observations. A lower AICc value indicates a better fitting model that also balances simplicity. Proposed models were restricted to a maximum of four predictor variables and pairwise interaction terms between variables were included only when searching for more complex models.

All statistical analyses and treatments were performed at a 95% confidence level.

2.1.2 Construction of cumulative distribution functions of K_d (NOR) values

As the probability that K_d (NOR) takes a given value can be represented by a continuous function, the construction of cumulative distribution functions (CDFs) allows the description of K_d (NOR) populations and the assessment of their variability. Fitting the CDF to a specific population distribution (e.g., normal, log-normal) allowed the population to be described by statistical parameters and can be a satisfactory approach to suggest a best estimate of the K_d (NOR) value when a parametric equation cannot be applied due to the lack of data on the specific soil properties governing NOR interactions.

The K_d (NOR) coefficients followed log-normal distributions, which was confirmed by the application of the Kolmogorov-Smirnov test to the log K_d (NOR) population. Thus, a log10-transformed K_d (NOR) population followed a normal distribution, with an associated location parameter, which is considered the most probable best estimate value (namely, the 50th percentile, PCTL 50), and a scale parameter, which informs on the dispersion of the population. These statistical parameters were obtained by fitting the cumulative distribution of log10-transformed K_d (NOR) values of a population to the theoretical CDF of a normal distribution (Ramírez-Guinart et al., 2020). Briefly, the log K_d (NOR) values were ordered by increasing value and given the same empirical frequency of 1/n, where n was the total number of entries. The experimental cumulative distribution was obtained through the sum of the preceding frequencies (up to 1) of the previously ordered log K_d (NOR) population. The fitting to the theoretical CDF was performed using the least-squares method with the Curve Fitting Toolbox of MATLAB R2022a.

When necessary, Fisher's Least Significant Difference (FLSD) test was used for the pairwise comparison of the partial K_d (NOR) populations.

2.1.3 The case of radium

2.1.3.1 Influence of soil properties on Ra sorption

PLS regression was performed with the dataset of the 31 soil samples to identify the soil properties that were related to the sorption K_d (Ra) values. The X dataset comprised a matrix of 31 soil samples x 16 soil properties (deduced from a previous PCA), which were pH, DIC, DOC, Cl^-_{ws}, $NO_3^-{}_{ws}$, $SO_4^{2-}{}_{ws}$, clay, sand, $CaCO_3$, Fe_{am}, Mn_{am}, C_{org}, $(Ca+Mg)_{exch}$, CEC, K_d (Ca+Mg), and SSA. The Y dataset corresponded to a 31 x 1 matrix of the sorption K_d (Ra) values.

The use of PLS regression led to a model that explained around 80% of the K_d (Ra) data variance. The model contained two LVs, the lowest RMSECV value and the lowest difference between the RMSE and RMSECV parameters (see *Figure 8A*). We did not observe samples with a higher Q residual than the 95% confidence threshold, while only one clay soil sample had a Hotelling's T2 value beyond the threshold determined by the 95% confidence level. As the exclusion of this soil sample did not affect the predicting ability of the PLS model, no samples were excluded.

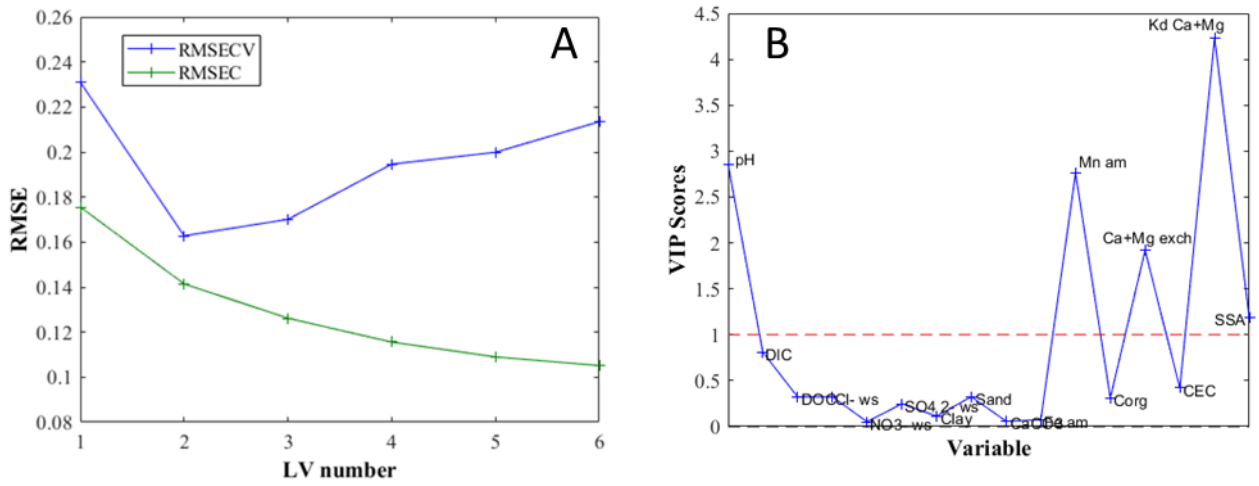


Figure 8 – PLS regression analysis: (A) Changes in RMSE and RMSECV with increasing LV; (B) VIP loading plot of the PLS-based model for predicting the K_d (Ra) value.

The plot of the VIP values confirmed the variables that could be considered good descriptors of the K_d (Ra) data variance (VIP score > 1). As can be seen in *Figure 8B*, the pH of the contact solution, the content of Mn_{am} in the soil, the levels of exchangeable Ca and Mg, K_d (Ca+Mg), and the SSA were the best descriptors to be included in a model predicting the sorption K_d (Ra) of a soil sample. These soil properties are related to the availability of exchangeable sorption sites in the soil, which is partially in agreement with the sorption pattern reported for radiostrontium, whose sorption was related to the CEC and to the levels of Ca and Mg in the soil solution (Gil-García et al., 2011).

2.1.3.2 Construction of linear regression models with the Ra expanded dataset

Univariant linear regression (ULR) equations were explored to assess quantitatively the correlations between sorption K_d (Ra) and the soil properties. The most statistically significant ULR was found between K_d (Ra) and K_d (Ca+Mg) values, which explained more than 75% of K_d (Ra) variance (see also *Figure 9*):

$$\log K_d \text{ (Ra)} = 0.7 \log K_d \text{ (Ca+Mg)} + 1.8$$

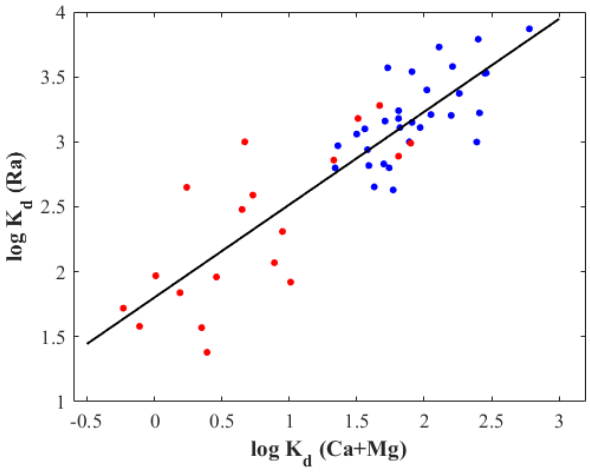


Figure 9 – ULR model with the expanded dataset between K_d (Ra) and K_d (Ca+Mg). The blue dots indicate our own data, while the red dots indicate the literature data.

This correlation was considered a reliable starting point for developing a K_d (Ra) prediction model based only on characterizing the levels of water-soluble and exchangeable Ca and Mg in the soil.

Multivariate linear regressions (MLR) were also tested. *Table 9* summarizes the MLR equations that described K_d (Ra) data variance better than the previous ULR equation.

Table 9 – MLR models for predicting K_d (Ra).

Correlation	N	Variance explained (%)
log K_d (Ra) = 0.7 log K_d (Ca+Mg) + 0.08 pH + 1.3	49	80
log K_d (Ra) = 0.5 log K_d (Ca+Mg) + 0.2 log SSA + 0.07 pH + 0.2 log Mn_{am} + 1.3	31	82

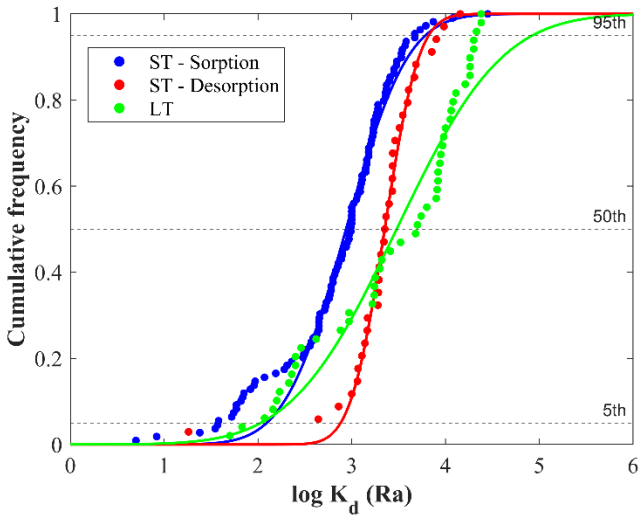
Two MLR models succeeded in explaining more than 80% of the K_d (Ra) data variance. In agreement with the PLS regression and ULR analyses, the soil parameters playing a key role in Ra sorption in soils were K_d (Ca+Mg) and pH, as well as the SSA and Mn_{am} for the 4-term MLR correlation. The choice on using one or the other equation should be based on the availability of information on the solid and liquid phases of the soil sample. For instance, the four soil parameters used in the first equation of *Table 9* are not currently reported in any of the literature studies reviewed, although it is the parametric equation that explains the highest data variance.

2.1.3.3 Use of CDF to derive K_d (Ra) best estimates

The construction of CDFs allowed for a rapid, comparative and global analysis of K_d (Ra) populations and for the suggestion of the best K_d (Ra) estimates based on the available characterization information if it would be insufficient to apply parametric models. In this context, the creation of partial populations based on ranges of property values with available information can allow for the derivation of best estimates with lower associated variability.

2.1.3.3.1 Kd (Ra) partial datasets according to the experimental approach

The construction of CDFs from partial datasets according to the experimental approach allowed us to identify possible K_d (Ra) differences between long-term incorporated Ra (largely, native Ra) and recently incorporated Ra. Differences among ST-S and ST-D datasets with respect to the LT dataset was examined. The CDF and derived parameters for the comparison of the three K_d (Ra) populations are shown in *Figure 10*.



Dataset	PCTL 50	PCTL 5	PCTL 95
ST-S	870	120	6100
ST-D	2240	760	6610
LT	3010	110	84500

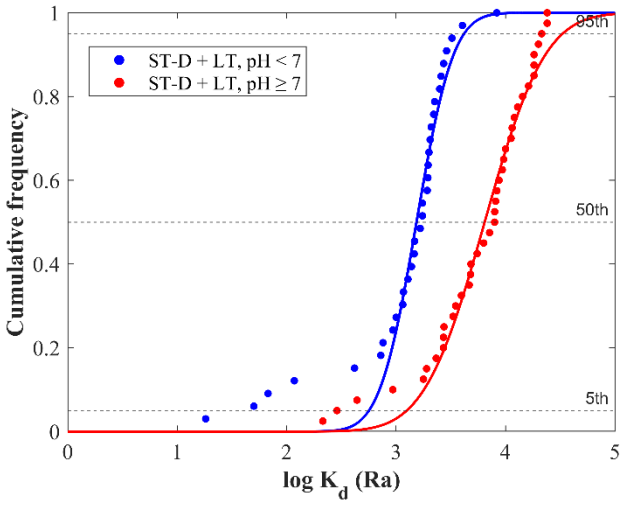
Figure 10 – CDFs and derived parameters (PCTL 50, PCTL 5, and PCTL 95: L kg^{-1}) of the K_d (Ra) ST-S, ST-D and LT populations.

The CDFs constructed from ST-S and ST-D populations were statistically different, the latter presenting a most probable value (PCTL 50) that was more than two-fold greater than the sorption dataset, thus demonstrating a certain sorption irreversibility. Moreover, this meant that the information on the desorption behaviour of recently incorporated Ra in a solid sample could not be directly extracted from sorption tests, unless a correction factor related to the hysteresis (around 2.5) was applied. On the other hand, the application of the FLSD test revealed no differences between the ST-D and LT datasets, partly due to the large variability of the LT dataset. Therefore, when predicting Ra remobilisation, the elapsed time is not a key factor and other factors (mainly soil properties) probably play a major role. In terms of radioecological risk assessments, these findings confirm that sorption (i.e., incorporation of Ra from effluents into solid compartments) and desorption scenarios (i.e., Ra remobilisation) should use different input data, as sorption irreversibility should be considered. Furthermore, if Ra sorption data are used to predict Ra remobilisation, an overestimation of Ra mobility may occur.

Considering the lack of statistical differences between the ST-D and LT populations, the integration of the two datasets (defining an overall desorption dataset) allowed us to derive the parameters of the resulting population (PCTL 50 = 2760 L kg^{-1}; PCTL 5 = 255 L kg^{-1}; PCTL 95 = 29830 L kg^{-1}). Consequently, the hysteresis factor was recalculated as 3.2.

2.1.3.3.2 Effect of pH on the proposal of the best estimates of desorption K_d (Ra)

To further examine the desorption K_d (Ra) dataset, the potential role of the soil properties in K_d (Ra) variability was examined. The analysis could only be performed by grouping the samples according to the pH, as it was one of the few properties available in a relevant number of entries. A pH grouping criterion with the threshold at pH 7 was chosen because of the chemical significance of differentiating between acidic and alkaline samples, as well as to ensure large enough partial datasets that could be used to derive reliable CDFs and related parameters. *Figure 11* displays the CDFs and derived parameters obtained from using pH to examine the desorption K_d (Ra) population.



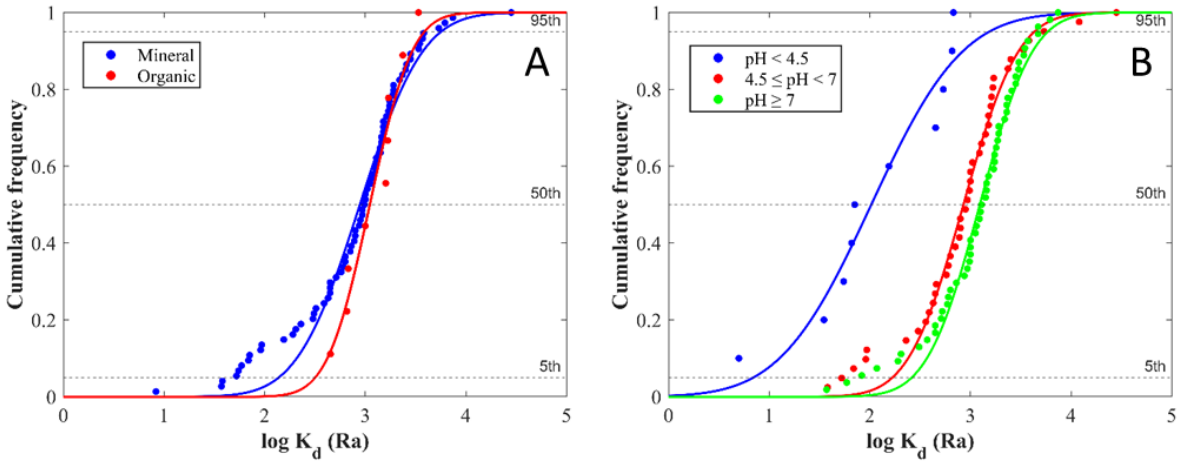
Dataset	PCTL 50	PCTL 5	PCTL 95
pH < 7	1540	560	4180
pH ≥ 7	6440	1250	33210

Figure 11 – The CDFs and derived parameters (PCTL 50, PCTL 5, and PCTL 95: L kg^{-1}) of the combined desorption K_d (Ra) populations for soils grouped according to the pH.

Each partial desorption dataset presented lower variability than the overall ST-D + LT K_d (Ra) dataset. Additionally, the two pH datasets were statistically different. Thus, the best estimate PCTL 50 value for the pH ≥ 7 partial dataset was significantly higher than that for the pH < 7 partial dataset (more than four-fold greater), clearly indicating that Ra remobilisation decreases with increasing pH, as Ra tended to solubilise more in acidic media. This behaviour is in agreement with the expected increase in Ra sorption because of the negative charges generated on the surface of the solid phase in the alkaline pH range, since Ra occurs mainly in the form of the free ion Ra^{2+}. Furthermore, the lower desorption K_d (Ra) in the acidic pH range is also related to an increased rate of dissolution of the mineral phases retaining Ra, as previously reported (Thiry et al., 2008).

2.1.3.3.3 Effect of organic matter content and pH on the proposal of best estimates of ST-S K_d (Ra)

The ST-S K_d (Ra) dataset was divided into partial datasets according to the textural information and OM content. The CDFs for the ST-S mineral and organic datasets and the parameters derived are shown in *Figure 12A*.



Dataset	PCTL 50	PCTL 5	PCTL 95
Mineral	880	140	5660
Organic	1110	320	3870

Dataset	PCTL 50	PCTL 5	PCTL 95
pH < 4.5	100	7.0	1540
4.5 ≤ pH < 7	850	170	4330
pH ≥ 7	1240	260	5870

Figure 12 – The CDFs and derived parameters (PCTL 50, PCTL 5, and PCTL 95: L kg^{-1}) of the ST-S K_d (Ra) populations for mineral and organic soils (A) and according to the defined pH ranges (B).

The partial ST-S mineral and organic datasets exhibited lower variability than the overall ST-S K_d (Ra) dataset. However, the mineral and organic distributions were statistically equal, also when compared to the overall dataset. This outcome agreed with the fact that textural and OM-related properties (e.g., clay and sand contents, DOC and C_{org}) were not relevant for directly predicting K_d (Ra) (Serra-Ventura et al., 2024), as evidenced by the PCA and ULR analyses. It was also confirmed that textural and OM-related properties alone were not suitable for decreasing K_d (Ra) variability and/or suggesting more representative K_d (Ra) best estimates compared to the overall ST-S dataset. This is in contrast to that observed in previous studies with smaller datasets (Vandenhove et al., 2009).

Regarding pH, ranges for defining the partial datasets were selected according to the protolysis data of representative minerals from the most abundant phases in the soils selected (Brendler et al., 2003). Three partial datasets could be thus analysed as shown in *Figure 12B*. The derived K_d (Ra) PCTL 50 ranged within one order of magnitude, from 100 to 1240 L kg^{-1} for the most acidic and the most alkaline ranges, respectively. Below pH 4.5, the silanol groups and ancillary groups from argillaceous minerals such as kaolinite and illite (with pK_{a2} values of around 6.5) and quartz (pK_{a1} values of around −1.5) were expected to be mostly positively charged (Mahoney et al., 1991 & Sverjensky, 2006), resulting in low K_d (Ra) values. Conversely, the alkaline pH led to a higher K_d (Ra) PCTL 50, since commonly occurring minerals such as (oxy)hydroxides and feldspars were expected to be deprotonated (Arnold et al., 2001), thus enhancing the sorption of Ra at the mineral surface. This outcome agrees with previous studies confirming that Mn minerals such as manganite and pyrolusite show significant positive correlations with Ra sorption in soils, presenting pK_{a2} values of around 9 and, hence, generating negative charges at an alkaline pH to immobilise Ra. Moreover, this aligns with the fact Mn_{am} is a significant descriptor of sorption K_d (Ra) variability. In addition to this, other commonly occurring minerals in the soil from the feldspar group, such as orthoclase and albite that also have pK_{a2} values of around 8, probably contributed to the higher K_d (Ra) PCTL 50 at alkaline pH values.

Statistically significant differences between the pH < 4.5 dataset and the two other partial datasets were observed. Thus, a unique K_d (Ra) best estimate could be proposed for short-term risk assessments of scenarios involving acidic contact solutions (below pH 4.5) of 100 L kg^{-1}. Furthermore, the combination

of the pH ≥ 4.5 datasets resulting from their statistical equality led to the obtention of a differentiated K_d (Ra) best estimate for the cases with near neutral and alkaline waters of 1070 L kg^{-1}, presenting lower variability than the overall ST-S dataset (PCTL 5 = 210 L kg^{-1}; PCTL 95 = 5540 L kg^{-1}).

2.1.4 The case of polonium

2.1.4.1 Influence of soil properties on Po sorption

An initial exploratory analysis of the role of pH on sorption and desorption K_d (Po) variability was carried out. As seen in *Figure 13*, despite the fact that the highest K_d (Po) values were observed within the pH range 7–8, no clear effect of the soil pH on Kd (Po) variability pattern could be drawn.

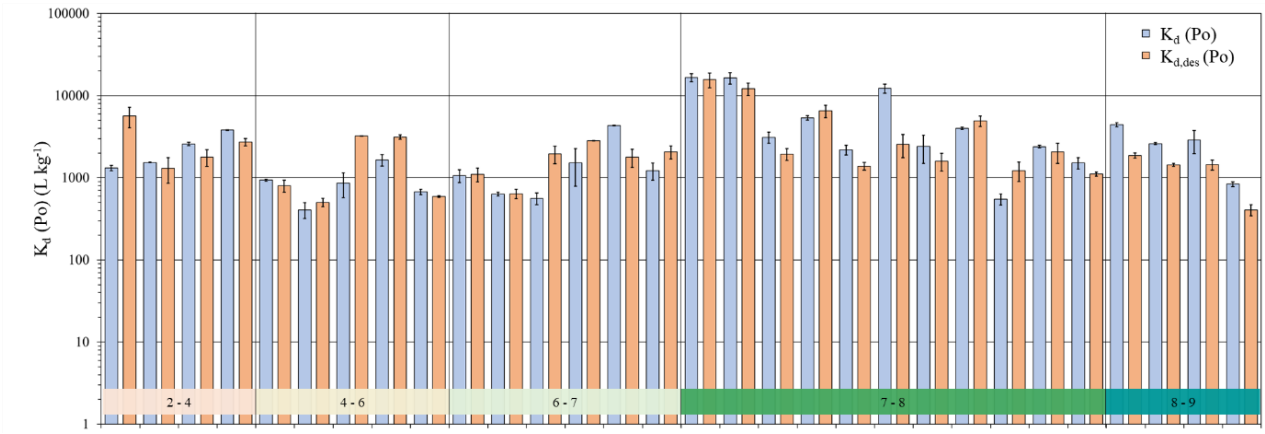


Figure 13 – Dependence of K_d (Po) vs. pH for the soil collection.

In a similar way to what was done with Ra, PLS regression was performed with a dataset of 30 soil samples to identify the soil properties that were related to the sorption K_d (Po) values. The X dataset comprised a matrix of 29 soil samples x 15 soil properties (deduced from a previous PCA), which for the case of Po were pH, DIC, DOC, $(Na+K)_{ws}$, $(Ca+Mg)_{ws}$, $(Na+K)_{exch}$, $(Ca+Mg)_{exch}$, silt+clay, sand, SOM, $CaCO_3$, Fe_{am}, Mn_{am}, and SSA. The PLS analysis permitted to observe a soil sample (TENF1) with a Q residual higher than the 95% confidence threshold (see *Figure 14 top*). Thus, this soil was excluded for further analyses.

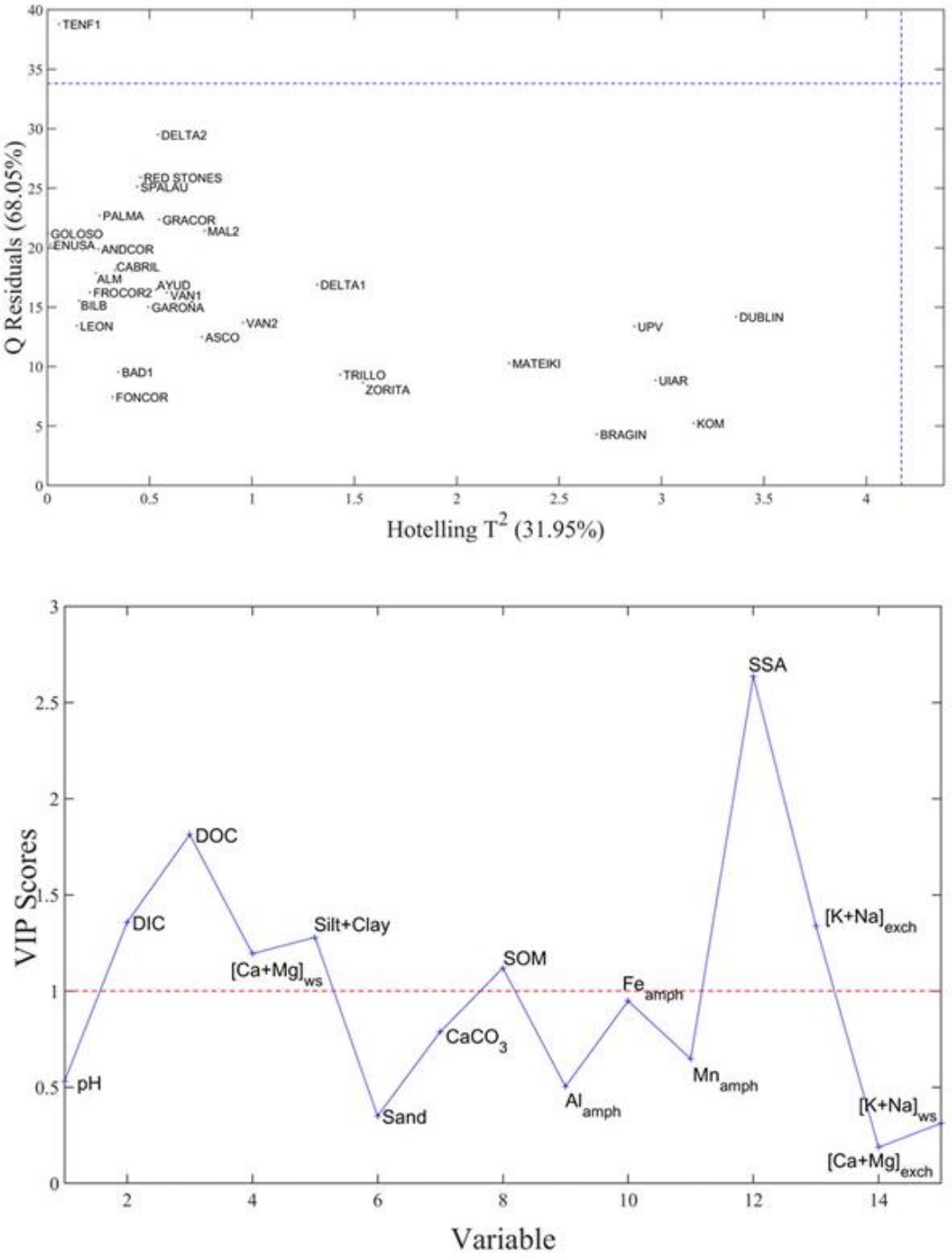


Figure 14 – PLS regression analysis: Hotelling's T2 and Q residuals of the soil collection regarding K_d (Po) (top); VIP loading plot of the PLS-based model for predicting the K_d (Po) value (bottom).

Regarding the VIP plot (see *Figure 14 bottom*), it confirmed the variables that could be considered good descriptors of the K_d (Po) data variance (VIP score > 1), from which SSA stands out. Besides this variable, a role of DOC (which may form soluble complexes with Po, thus preventing Po sorption in the solid phase), $(K+Na)_{exch}$ (an increase of this variable would lead to a decrease in the Ca+Mg status in the exchangeable complex, thus enhancing Po sorption), silt+clay. $(Ca+Mg)_{ws}$ and SOM could be expected. In addition to this, the pH variable did not get a VIP value higher than 1, in agreement with the lack of K_d (Po) dependence over pH previously observed.

2.1.4.2 Construction of linear regression models with the Po own dataset

A first attempt to obtain parametric models to predict K_d (Po) was carried out with the K_d (Po) dataset obtained within the RadoNorm project, which had 29 entries once the TENF1 soil was excluded.

Table 10 – ULR and MLR models for predicting K_d (Po) (N=29).

Model equation	Variance explained (%)
$\log K_d\,(Po) = 0.5 \log SSA + 3.0$	68
$\log K_d\,(Po) = 0.6 \log SSA + 0.2 \log [K+Na]_{exch} + 2.9$	73
$\log K_d\,(Po) = 0.3 \log SSA + 0.4 \log [K+Na]_{exch} - 0.3 \log DOC + 3.4$	81
$\log K_d\,(Po) = 0.3 \log SSA + 0.4 \log [K+Na]_{exch} - 0.4 \log DOC - 0.3 \log [Ca+Mg]_{ws} + 4.0$	87

As expected from the PLS-VIP results, the use of SSA permitted to derive the best ULR equation, with an associated AICc value of 5.4, which described near to 70% of the K_d (Po) variability. Regarding MLR equations, the application of the *glmulti* analysis permitted to suggest the most promising MLR parametric models. The inclusion of other relevant variables observed in the VIP plot (such as $(K+Na)_{exch}$, DOC and $(Ca+Mg)_{ws}$, increased the prediction capacity of the derived equations, with negative AICc values (as low as −12.5 for the 4-variables equation), which were able to describe up to 87% of the K_d (Po). The external validation of these models with data gathered from the literature remains as a future exercise to be performed.

2.1.5 The case of uranium

2.1.5.1 Influence of soil properties on U sorption

A PLS regression was performed using a collection of 35 soils, with an X matrix consisting of the 16 predictor variables derived from a previous PCA analyses (pH, DIC, DOC, silt+clay, sand, $CaCO_3$, SOM, Al_{amph}, Fe_{amph}, Mn_{amph}, K_{exch}, SSA, water soluble native U ($^{238}U_{ws}$)) and a Y matrix containing the corresponding K_d (U) values as the response variable. Note that experiments were performed with 233U, which makes the distinction with soil-native U possible. The LV space explained up to 78% of the variance of K_d (U) using two LVs, which ensured a minimum RMSE while maintaining minimal differences between this error and the RMSECV. Additionally, outliers or poorly explained samples were explored via analysis of Q residuals and Hotelling's T^2. None of the samples exceeded the limits in either case with 95% confidence, so the initial set of 35 soils was retained.

The VIP scores plot (*Figure 15*) identified the variables with the most significant influence on predicting K_d (U). As seen, variables affecting U sorption already identified in previous reported works, such as pH, silt+clay content, $CaCO_3$, and SOM, as well as less commonly considered variables, such as SSA, $^{238}U_{ws}$, or K_{exch}, were associated with VIP > 1.

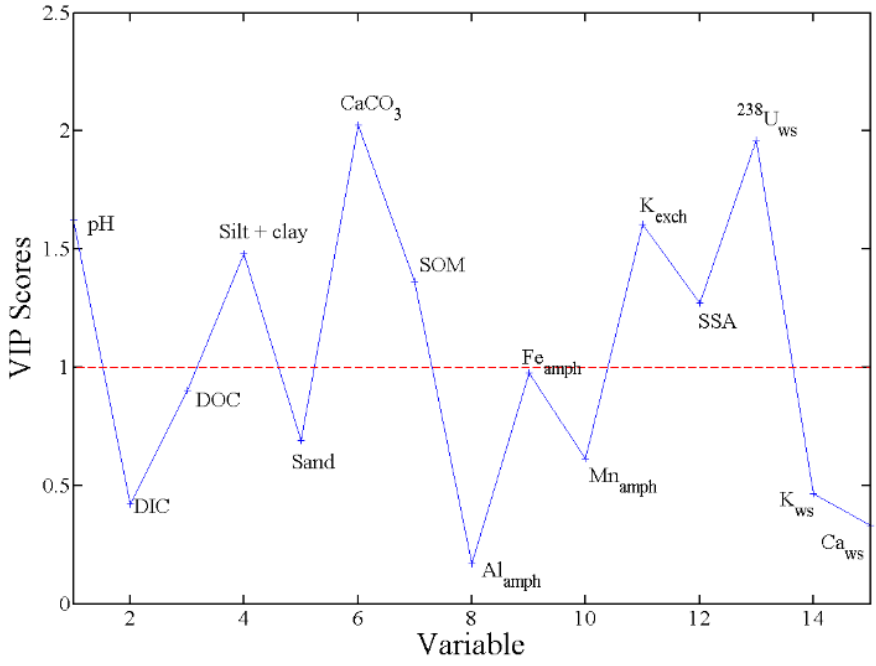


Figure 15 – VIP scores plot derived from PLS regression for K_d (U) prediction.

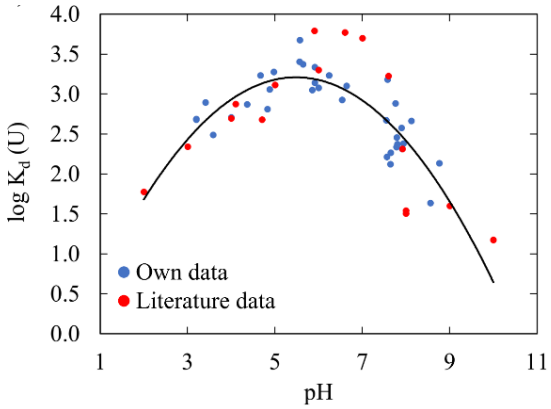
After reducing the set of relevant predictor variables to 8 (those with VIP values >1), the remaining variables were used to construct prediction K_d (U) models, enabling computationally intensive searches that would have been highly time-consuming and incurred significant computational costs, if conducted with the initial set of 26 variables.

2.1.5.2 Construction of linear regression models with the U extended dataset

The application of the *glmulti* analysis to predict K_d (U) from the relevant soil properties identified in the PLS analysis started focusing on simple univariate regressions and progressed towards more complex multivariate models which eventually incorporate pairwise interaction terms. A K_d (U) extended dataset, comprising own and literature data, was used to derive the parametric equations.

2.1.5.2.1 Simple regression models: ULR with pH

Considering the key role of pH on U speciation, this was the variable to be tested first to construct ULR models. The relationship between K_d (U) and pH followed a bell-shaped trend as previously observed (Vandenhove et al., 2009), then suggesting a quadratic equation as the initial equation to be tested. Therefore, the pH^2 term was added as an additional predictor variable, since *glmulti* package does not account for quadratic terms by default. As shown in the table associated to *Figure 16*, the best model found was based, as expected, on a second-order equation.



Regression equation	N	Variance explained (%)
$\log K_d\,(U) = -0.13\ pH^2 + 1.4\ pH - 0.6$	53	66

Figure 16 – ULR pH-based regression. Plot built from expanded dataset is represented together with the corresponding regression equation, number of entries (N) and variance explained in the associated table.

2.1.5.2.2 MLR regression models

To improve the percentage of K_d (U) variance explained by the model based solely on pH, the construction of MLR models that combine a greater number of variables, including interaction terms, was explored through *glmulti* package. This approach yielded a total of more than 450 and 25,000 models, without and with interaction terms, respectively.

Table 11 summarizes the most relevant models according to the AICc criterion (with values ranging from 5.9 to −5.4) and K_d (U) variance explained. Besides pH, they included terms of $^{238}U_{ws}$, SSA and $CaCO_3$ content. Model 2, based solely on pH and $CaCO_3$ content, increased the explained K_d (U) variance near to 80%. The high percentage of explained K_d (U) variance and the fact that it is based on two easily obtainable properties makes it a promising model for predicting soil sorption K_d (U). Other notable models were model 3, which incorporates a variable that had previously appeared in model 1 ($^{238}U_{ws}$), and model 4, which introduces SSA. In both cases, more than 80% of explained variance percentage is achieved, although they were based in variables less available in the literature.

Table 11 – MLR models for predicting K_d (U).

Model No.	Model equation	N	Variance explained (%)
	Interactions excluded		
1	$\log K_d\,(U) = -0.11\ pH^2 + 1.2\ pH - 0.2 \log {}^{238}U_{ws} - 0.3$	35	75
2	$\log K_d\,(U) = -0.13\ pH^2 + 1.5\ pH - 0.3 \log CaCO_3 - 1.0$	38	78
3	$\log K_d\,(U) = -0.12\ pH^2 + 1.4\ pH - 0.2 \log CaCO_3 - 0.2 \log {}^{238}U_{ws} - 1$	35	81
4	$\log K_d\,(U) = -0.16\ pH^2 + 1.9\ pH - 0.3 \log CaCO_3 + 0.2\ \log SSA - 2$	35	81
	Interactions included		
5	$\log K_d\,(U) = -0.15\ pH^2 + 1.8\ pH - 0.3\ \log CaCO_3 + 0.3\ \log SSA \cdot \log SOM - 2$	35	84

The model 5, with an interaction term, resulted in describing the highest K_d (U) variance percentage. Despite the apparent complexity of the model, it is well-suited to predict U behaviour in a sorption scenario for recently contaminated soils with a wide range of edaphic properties, as it makes use of soil properties that are easy to obtain.

2.1.6 Conclusion

In this report, an integrated approach for the construction of K_d (NOR) prediction models is also presented, specifically for Ra, Po, and U. In a first phase, a dataset is created with new experimental data, supplemented with data collected from the literature when necessary, and including a large number of characterised soil properties. Next, the main properties governing the interaction of a given NOR in soils are identified through the combined use of chemometric tools such as principal component analyses (PCA) and partial least squares regression analysis (PLS). Subsequently, the relevant properties are used in the construction of parametric equations derived from uni- or multivariate linear regressions, with the aim of explaining as much of the K_d (NOR) variability as possible. In cases where the lack of a complete characterisation of soils makes the application of parametric models difficult, the use of probabilistic distribution functions, with data grouped according to intervals of relevant property values, allows the derivation of K_d (NOR) best estimates for groups of soils with close values of certain properties, which is illustrated for the case of Ra. This overall approach to predict K_d (NOR) to be used as input data in radioecological risk assessment codes proved to be successful, as derived parametric equations were able to predict K_d (NOR), with a percentage of described K_d (NOR) variance often higher than 80%.

3. Blind prediction exercise

3.1 Examples

The approaches developed by the University of Barcelona and the Helmholtz-Zentrum Dresden-Rossendorf e.V. (HZDR) and presented above to predict the K_d values have been used in the framework of a blind prediction exercise within the ALLIANCE Topical Roadmap working group NORM in spring 2023. The exercise, which focused on the prediction of K_d values for U and Ra, was initiated to evaluate the performance of these two approaches under real conditions. As part of the blind prediction exercise, both groups were only given field information on soil or solid samples and neither group was allowed to carry out any experimental work for their modelling. In preparation for this exercise, it was necessary to obtain suitable site-specific (field) data sets, including the associated K_d values for U or Ra, ideally for both elements. To this end, numerous colleagues were contacted and asked to check the availability of such data sets and to make them available for the blind prediction exercise. In addition, a thorough review of the existing literature was undertaken to determine the availability of relevant data sets. It was imperative to ensure that the participating groups remained blind to the origin of the datasets.

To facilitate the survey, a spreadsheet was circulated to colleagues outlining the different types of soil information required. Colleagues were asked to complete this table by simply ticking the information that was available. For the blind prediction exercise, only datasets that already contained experimentally determined K_d values (for U and/or Ra) and as much as possible of the soil information listed in the distributed table were considered.

After evaluating the datasets received, it turned out to be much more difficult than expected to find suitable datasets for the blind prediction exercise. There were many comprehensive datasets, but some important information, such as mineralogical composition, was very often missing. In the end, two datasets were selected for the blind prediction exercise. The information given for the chemical scenarios for the K_d blind prediction exercise are compiled in Appendix C.

3.2 Results

In the blind prediction exercise, the K_d values of five chemical scenarios (1A–C & 2) were to be determined using only the geochemical parameters and without knowledge of the measured field K_d values. Within the blind prediction exercise, in the scenarios 1A–C (see page 58), an overall SSA of 52.3 m^2/g was given, for the scenarios 2A–B (see page 59) no SSA value was available. As the description of the geochemical scenarios did not provide uncertainty data for the concentrations of dissolved ions or mineral contents in the soil, the geochemical parameters were not varied and only one K_d value was calculated for each. The results obtained are compared with the measured field K_d values in *Table 12*, in each case as K_d or, for better comparability, as sorbed percentage values.

3.2.1 Dataset 1A–C

In examples 1A–C, the predicted retention of U agrees remarkably well with the measured field data. The strikingly high deviation of the K_d value in example 1A is clearly relativized when converted into sorbed percentages and is a numerical effect of the K_d value calculation with almost complete sorption, where each decimal place of sorbed NOR percentage increases the K_d value by a factor of 10. In the case of complete sorption, large numbers (>10 000 L/kg) are therefore no longer meaningful.

The slightly larger deviation in example 1B is due to the significantly more acidic pH value (pH = 4.96 compared to pH = 7.23 in examples 1A and 1C). In this range, sorption is much more sensitive to changes in pH and therefore shows much larger uncertainty ranges in the experimental sorption data required to generate the SCM data. Nevertheless, the accuracy achieved in predicting the retention of the NOR can be considered good.

Table 12 – Comparison of the field data and the predicted K_d values of uranium and radium in the blind prediction exercise

Field data	**1A (U)**	**1B (U)**	**1C (U)**	**2 (U)**	**2 (Ra)**
K_d	21 680 L/kg	3 465 L/kg	5 113.64 L/kg	62 L/kg	413 L/kg
Log K_d	4.336 L/kg	2.540 L/kg	3.709 L/kg	1.792 L/kg	2.616 L/kg
Sorbed percentage	98.83 %	93.27 %	95.34 %	91.61 %	98.64 %
Prediction	**1A (U)**	**1B (U)**	**1C (U)**	**2 (U)**	**2 (Ra)**
Helmholtz-Zentrum Dresden-Rossendorf e.V. (HZDR)					
K_d	786 314 L/kg	7 793 L/kg	5 310 L/kg	2443 L/kg	64 L/kg
Log K_d	5.896 L/kg	3.892 L/kg	3.725 L/kg	3.388 L/kg	1.812 L/kg
Sorbed percentage	99.97 %	96.89 %	95.50 %	99.77 %	91.96 %
Δ sorbed percentage	1.14 %	3.62 %	0.16 %	8.16 %	−6.68 %
University of Barcelona (UB)					
K_d	1680 – 7230 L/kg (mean: 4455 L/kg)	2780 – 3545 L/kg (mean: 3162 L/kg)	1680 – 7230 L/kg (mean: 4455 L/kg)	8545 – 18030 L/kg (mean: 13288 L/kg)	1275 – 1475 L/kg (mean: 1375 L/kg)
Log K_d (from mean K_d)	3.649 L/kg	3.500 L/kg	3.649 L/kg	4.123 L/kg	3.138 L/kg
Sorbed percentage	87.05 – 96.66 % (mean: 91.86 %)	91.75 – 93.41 % (mean: 92.85 %)	87.05 – 96.66 % (mean: 91.86 %)	99.93 – 99.97 % (mean: 99.95 %)	99.56 – 99.61 % (mean: 99.59 %)
Δ sorbed percentage	7.0 %	0.69 %	3.48 %	−8.34 %	−0.95 %

Of particular note is the good agreement of the predicted K_d values using the Smart-K_d approach with the measured field data of example 1C. Example 1C represents a K_d value measured in a natural system (natural pore water with increased CO_2 partial pressure due to borehole depth), in contrast to examples 1A & B, which were obtained under simplified laboratory conditions. The thermodynamic data for the dissolved U-carbonato complexes play an important role in combination with the SCM data, which could be demonstrated in this case. In general, smart-K_d works in principle, allowing rapid coverage of large parameter ranges and helping to identify relevant processes. There is potential for further improvement of smart-K_d predictions by improving the SCM dataset by filling existing data gaps.

3.2.2 Dataset 2

The predicted K_d values for example 2 deviate slightly more from the measured field data for both U and Ra compared to example 1. The main reason for this is most likely the amount of unspecified organic substances (33 % in the dry sample). Without information in the nature of the organic substances (or at least the substance class), it is not possible to set up a thermodynamic model for their interaction with NOR. Therefore, the influence of these substances could not be taken into account when predicting the NOR's K_d values. More specific site information is required to increase the quality of the K_d prediction. The lack of Ba and Sr concentrations prevented the correct calculation of the Ra concentration, as only the formation of solubility determining solid solutions could be taken into account, see also Chapter 1.3.1.2. These parameters should also be recorded in future measurement campaigns.

Together with the results from the collaborating group from Barcelona University (methodology given in Vidal et al., 2024), the results have been presented at the Workshop of the ALLIANCE Topical Roadmap WG NORM (Bok, 2023). Both datasets were not perfect for the two participating groups, but they started the prediction exercise willing to fill the gaps with their respective assumptions. In any case, the survey showed that such field datasets hardly exist and that it would be a nice task to compile useful and complete datasets for both groups for another blind prediction exercise.

4. Conclusions

The Smart-K_d concept has demonstrated its suitability for predicting the sorption behaviour of NOR. However, further research is necessary to improve the reliability and robustness of the modelling. The role of organic matter in soils, and its influence on sorption processes, are still not well understood and require further investigation. Moreover, to reduce uncertainties in K_d predictions, more accurate characterization of the specific surface areas of soil minerals is essential. The study also revealed significant gaps in the availability of thermodynamic sorption data for Ra on naturally occurring soil minerals. Although efforts are currently underway to address these deficiencies studying the sorption behaviour of Sr and Ba as chemical analogues for Ra, but also of Ra, data gaps for elements such as bismuth and Po are even more pronounced. This highlights an urgent need for additional research to support accurate modelling of their environmental behaviour. The results could be significantly improved if the description of the geochemical factors (field measurement data) were more complete. In particular, the lack of information on barium and strontium concentrations in the aqueous phase has a significant influence on the thermodynamic modelling of Ra behaviour, since solid solutions containing Ra, Ba and Sr sulphate or carbonate — which are highly influential in Ra chemistry — cannot be included. This alters the solubility-limited concentration of radium in an aqueous solution by several orders of magnitude.

A source of uncertainty was that the complexation of the NOR to naturally occurring organics - especially humic substances - could not be taken into account, as an indication of the dissolved organic carbon/total organic carbon (DOC/TOC) content was not part of the sample characterization obtained. Even then, complexation would not be calculable as it would require knowledge of the exact nature of the TOC, which would require an extensive analysis of the organic substances. Depending on the nature of the organic matter in soil, the effect on NOR mobility can be different and hard to predict. Small dissolved organic molecules can form aqueous complexes with NOR, slightly increasing solubility and reducing NOR sorption. However, these complexation reactions are often weak compared to inorganic ligands. The effect of larger organic molecules is difficult to predict as the sorption of NOR on mineral phases can be decreased or increased with different concentrations of e.g. humic acid (Křepelová et al., 2006).

The work described in this report represents a relevant step forward in the prediction of sorption K_d (NOR) by combining new experimental data in well-controlled conditions and chemometric tools able to elucidate the relevant properties governing NOR sorption, and subsequently propose best derived model equations. Regarding Ra, the PCA and PLS analyses permitted to identify those soil properties relevant to describe Ra interaction, such as K_d (Ca+Mg), Mn_{am}, SSA and pH. A simple URL model based solely on K_d (Ca+Mg) described around 75% of K_d (Ra) variability, whereas a MLR parametric model based on K_d (Ca+Mg), SSA, pH and Mn_{am} the K_d (Ra) succeeded in explaining 82% of K_d (Ra) variance. Furthermore, the strategy of constructing CDFs for partial K_d (NOR) datasets grouped according to different criteria, here tested for the case of Ra, succeeded in deriving robust K_d (Ra) best estimates that can be derived according to the information available for the scenario to be assessed. Regarding Po, SSA was confirmed to be a relevant soil property governing Po interaction in soils. Besides this variable, the additional role of DOC, $(K+Na)_{exch}$, silt+clay and $(Ca+Mg)_{ws}$ were pointed out. Based on these variables, the application of the *glmulti* tool allows to derive several K_d (Po) prediction models of different complexity. The single use of SSA allowed to describe nearly a 70% of K_d (Po) variance, whereas more complex models explained up to 87% of K_d (Po) variability, using SSA, DOC, $(Na+K)_{exch}$ and $(Ca+Mg)_{ws}$ data. For the case of U, the relevant soil properties governing U sorption were pH, $CaCO_3$ and SOM, as well as additional soil properties not previously considered, such as SSA and water-soluble native U. The most relevant models to predict K_d (U) were a pH second-order equation, which explained around 70% of K_d (U) variance, and MLR models including pH, $CaCO_3$, SOM and SSA (up to 84% of the K_d (U) variance). Therefore, it can be concluded that for the first time, end-users have an array of models to

predict K_d (NOR), to be selected according to the available information on those key soil properties governing NOR interaction in soils.

The blind prediction test further underscored the lack of consistent and comprehensive databases for surface complexation modelling (SCM) data. Currently, individual modellers rely on independently compiled datasets, which vary considerably in terms of documentation quality and quality assurance practices. As a result, comparability between modelling results from different research groups is severely limited. Additionally, the test revealed that the characterization of soil samples is often incomplete. Moving forward, the application of a standardized formalism for sample characterization is recommended to ensure the completeness and comparability of field data, including the documentation of parameter variability ranges.

Further blind prediction exercises are strongly encouraged—ideally with expanded participation from international partners—to enhance data harmonization, methodological transparency, and to improve the predictive capabilities for K_d values urgently needed in reactive transport models.

5. Acknowledgements

Viktor Dück and Frank Bok would like to thank Sabrina Beutner, Salim Shams, and Paul Chekhonin (HZDR) for analytical support, the ROBL team at the ESRF beamline for the opportunity to perform EXAFS measurements and their support during the measurements and the data analysis, and Norbert Jordan and Vinzenz Brendler for the Ph.D. work supervision and fruitful discussions.

6. References

Akaike, H. (1974) A new look at the statistical model identification, *IEEE Trans. Automat. Contr.*, **19**(6), 716–723.

Aldrich, J. (1997). R. A. Fisher and the making of maximum likelihood 1912–1922, *Stat. Sci.*, **12**(3), 162–176.

Ali, M.A. et al. (1996) Effects of simple organic acids on sorption of Cu^{2+} and Ca^{2+} on goethite, *Geochim. Cosmochim. Acta*, **60**(2), 291–304.

Arnold, T. et al. (2001) Sorption behavior of U(VI) on phyllite: experiments and modelling, *J. Contam. Hydrol.*, **47**(2–4), 219–231.

Axe, L. et al. (1998) An XAFS Analysis of Strontium at the Hydrous Ferric Oxide Surface, *J. Colloid Interface Sci.*, **199**(1), 44–52.

Axe, L. et al. (2000) Local Structure Analysis of Strontium Sorption to Hydrous Manganese Oxide, *J. Colloid Interface Sci.*, **224**(2), 408–416.

Bethke, C.M. (2022) *Geochemical and Biogeochemical Reaction Modeling*, 3rd Ed., Cambridge University Press, Cambridge (UK), ISBN-13: 978-1108790864, 524 p.

Berka, M. et al. (2001) Surface complexation modeling of K^+, NO_3^-, SO_4^{2-}, Ca^{2+}, F^-, Co^{2+}, and Cr^{3+} ion adsorption on silica gel, *J. Colloid Interface Sci.*, **233**(1), 131–135.

Bok, F. (2023) *Blind prediction of K_d values from NORM sites using the smart-K_d approach*, Workshop of the ALLIANCE Topical Roadmap WG NORM, 06.-7.6.2023, Granada (Spain).

Bradbury, M.H. et al. (2005a) *Experimental and modelling investigations on Na-illite: Acid-base behaviour and the sorption of strontium, nickel, europium and uranyl*, NTB Technical Report 04-02, Nagra, Wettingen (Switzerland), 87 p.

Bradbury, M.H. et al. (2005b) Modeling the sorption of Mn(II), Co(II), Ni(II), Zn(II), Cd(II), Eu(III), Am(III), Sn(IV), Th(IV), Np(V) and U(VI) on montmorillonite: Linear free energy relationships and estimates of surface binding constants for some selected heavy metals and actinides, *Geochim. Cosmochim Acta*, **69**(4), 875–892.

Brandt, F. et al. (2020) Recrystallization and Uptake of ^{226}Ra into Ba-Rich $(Ba,Sr)SO_4$ Solid Solutions, *Minerals*, **10**(9), 812.

Brendler, V. et al. (2003), RES^{3}T-Rossendorf expert system for surface and sorption thermodynamics, *J. Contam. Hydrol.*, **61**(1-4), 281–291.

Britz, S.M. (2018) *Europium sorption experiments with muscovite, orthoclase, and quartz: Modeling of surface complexation and reactive transport*, Ph.D. thesis, Technische Universität Carolo-Wilhelmina zu Braunschweig, Braunschweig (Germany), 284 p.

Bruggeman, C. et al. (2010) Influence of Boom Clay organic matter on the adsorption of Eu(3+) by illite - geochemical modelling using the component additivity approach, *Radiochim. Acta*, **98**, 597–605.

Charlet, L. et al. (1993) Cation adsorption on oxides and clays: The aluminum case, *Aquat. Sci.*, **55**(4), 291–303.

Chen, M.A. et al. (2018) Radium Sorption to Iron (Hydr)oxides, Pyrite, and Montmorillonite: Implications for Mobility, *Environ. Sci. Technol.*, **52**(7), 4023–4030.

Cocchi, M. et al. (2018) Chapter Ten - Chemometric Methods for Classification and Feature Selection, *Compr. Anal. Chem.*, 82, 265–299.

Davies, C.W. (1962) *Ion association*, Butterworths, London (UK), 190 p.

Davis, J.A. et al. (1990) *Surface complexation modeling in aqueous geochemistry* in Book: *Mineral-Water Interface Geochemistry*, Hochella, M. F., White, A. F. (Eds.), Mineralogical Society of America, Washington, DC (USA), ISBN: 0-939950-28-6, 177–258.

Dzombak, D.A. et al. (1990) *Surface complexation modeling. Hydrous ferric oxide*, Wiley, New York (USA), ISBN: 0-471-63731-9, 393 p.

Fletcher, P. et al. (1989) The chemical modeling of clay/electrolyte interactions for montmorillonite. *Clay Miner.*, **24**(2), 375–391.

Gil-García, C.J. et al. (2011) The use of hard- and soft-modelling to predict radiostrontium solid-liquid distribution coefficients in soils. *Chemosphere*, **85**(8), 1400–1405.

Gilbert, M. et al. (1965) Exchange properties of hydrogen ions in clays, *Soil Sci.*, **100**(3), 157–162.

Grandia, F. et al. (2008) *Assessment of the radium-barium co-precipitation and its potential influence on the solubility of Ra in the near-field.* Swedish Nuclear Fuel and Waste Management Co., SKB Technical Report No. TR-08-07, ISSN: 1404-0344, 48 p.

Grangeon, S. et al. (2023) Geochemical and mineralogical characterization of streams and wetlands downstream a former uranium mine (Rophin, France), *App. Geochem.*, **150**, 105586.

Gu, X. et al. (2010) Modeling the adsorption of Cd (II), Cu (II), Ni (II), Pb (II) and Zn (II) onto montmorillonite, *Geochim. Cosmochim. Acta*, **74**(20), 5718–5728.

Hotelling, H. (1931) The generalization of Student's ratio. *Ann. Math. Stat.*, **2**, 360–378.

Hu, B. et al. (2017) Macroscopic and microscopic investigation on adsorption of Sr(II) on sericite, *J. Mol. Liq.*, **225**, 563–568.

Ikhsan, J. et al. (2005), Surface complexation modeling of the sorption of Zn(II) by montmorillonite, *Colloids Surf. A*, **252**(1), 33–41.

Katz, L.E., et al. (2013) Temperature effects on alkaline earth metal ions adsorption on gibbsite: Approaches from macroscopic sorption experiments and molecular dynamics simulations, *J. Colloid Interface Sci.*, **399**, 68–76.

Klinkenberg, M. et al. (2018) The solid solution–aqueous solution system $(Sr,Ba,Ra)SO_4 + H_2O$: A combined experimental and theoretical study of phase equilibria at Sr-rich compositions, *Chem. Geol.* **497**, 1–17.

Klinkenberg, M. et al. (2021) Adsorption of barium and radium on montmorillonite: A comparative experimental and modelling study, *Appl. Geochem.*, **135**, 105117.

Křepelová, A et al. (2006) Uranium(VI) sorption onto kaolinite in the presence and absence of humic acid, *Radiochim. Acta*, **94**(12), 825–833.

Kuan, W.H. et al. (2004) Modeling and electrokinetic evidences on the processes of the Al(III) sorption continuum in SiO_2(s) suspension, *J. Colloid Interface Sci.*, **272**(2), 489–497.

Kulik, D.A. (2006) Standard molar Gibbs energies and activity coefficients of surface complexes on mineral-water interfaces (thermodynamic insights) in Book: Surface Complexation Modelling, Lützenkirchen, J. (Ed.), Academic Press, Amsterdam (Netherlands) ISBN: 0123725720, 171–250.

Lützenkirchen, J. et al. (2014) Adsorption of dissolved aluminum on sapphire-c and kaolinite: implications for points of zero charge of clay minerals, *Geochem. Trans.*, **15**(9), 1–14.

Mahoney, J.J. et al. (1991) Adsorption of Sr on Kaolinite, Illite and Montmorillonite at High Ionic Strengths, *Radiochim. Acta*, **54**(3), 139–144.

Marques Fernandes, M. et al. (2019) Cation exchange and surface complexation of lead on montmorillonite and illite including competitive adsorption effects, *Appl. Geochem.*, **100**, 190–202.

Martin, A. et al. (2020) An integrated approach combining soil profile, records and tree ring analysis to identify the origin of environmental contamination in a former uranium mine (Rophin, France), *Sci. Total Environ.*, **747**, 141295.

Mathur, S.S. et al. (2006) Surface complexation modeling: Goethite in Book: Surface Complexation Modelling, Lützenkirchen, J. (Ed.), Academic Press, Amsterdam (Netherlands) ISBN: 0123725720, 443–468.

Missana, T. et al. (2003) Experimental and modeling study of the uranium(VI) sorption on goethite, *J. Colloid Interface Sci.*, **260**(2), 291–301.

Montavon, G. (2021) personal communication on the description of the U-rich, white clayey horizon at Rophin site, Email as of 2021-05-03.

Montavon, G. et al. (2025) 70 years later... Uranium speciation in a contaminated wetland and impact on its mobility, *Environ. Sci. Technol.*, in review.

Nie, Z. et al. (2017) Adsorption of selenium and strontium on goethite: EXAFS study and surface complexation modeling of the ternary systems, *Environ. Sci. Technol.*, **51**(7), 3751–3758.

Nivesse, A.-L. et al. (2023) *Determination of site-specific K_d values,* RadoNorm, Milestone 31, 2023.

Olin, M. et al. (1997) *Application of surface complexation modelling: Nickel sorption on quartz, manganese oxide, kaolinite and goethite, and thorium on silica*, POSIVA97-10, Posiva Oy, Helsinki (Finland), 32 p.

Pappalardo, R.R. et al. (2021) Hydration of heavy alkaline-earth cations studied by molecular dynamics simulations and X-ray absorption spectroscopy. *Inorg. Chem.*, **60**(17), 13578-13587.

Ramírez-Guinart, O. et al. (2020) Deriving probabilistic soil distribution coefficients (K_d). Part 1: General approach to decreasing and describing variability and example using uranium K_d values, *J. Environ. Radioact.*, **222**, 106362.

Reich, T.J. et al. (2010) Surface complexation modeling of Pb(II) adsorption on mixtures of hydrous ferric oxide, quartz and kaolinite, *Chem. Geol.*, **275**(3–4), 262–271.

RES³T (2025) Rossendorf Expert System for Surface and Sorption Thermodynamics, Helmholtz-Zentrum Dresden-Rossendorf, https://www.hzdr.de/res3t.

Richter, C. et al. (2016) Macroscopic and spectroscopic characterization of uranium(VI) sorption onto orthoclase and muscovite and the influence of competing Ca^{2+}, *Geochim. Cosmochim. Acta*, **189**, 143–157.

Riese, A.C. (1982) Adsorption of radium and thorium onto quartz and kaolinite: A comparison of solution/surface equilibria models, Ph.D. thesis, Colorado School of Mines, Golden, CO (USA), 292 p.

Scheinost, A.C. et al. (2021) ROBL-II at ESRF: A synchrotron toolbox for actinide research, *J. Synchrotron Radiat.*, **28**(1), 333–349.

Serra-Ventura, J. et al. (2024) Construction and validation of parametric models to predict radium sorption in soils, *Sci. Total Environ.*, **944**, 173953.

Stockmann, M. et al. (2017) Smart Kd-values, their uncertainties and sensitivities - Applying a new approach for realistic distribution coefficients in geochemical modeling of complex systems. *Chemosphere*, **187**, 277–285.

Stockmann, M. et al. (2022) New insights into U(VI) sorption onto montmorillonite from batch sorption and spectroscopic studies at increased ionic strength, *Sci. Total Environ.*, **806**(2), 150653.

Sverjensky, D.A. (2006) Prediction of the speciation of alkaline earths adsorbed on mineral surfaces in salt solutions, *Geochim. Cosmochim. Acta*, **70**(10), 2427–2453.

Tachi, Y. et al. (2001) Experimental and modeling studies on sorption and diffusion of radium in bentonite, *J. Contam. Hydrol.*, **47**(2–4), 171–186.

Thiry, Y. et al. (2008) Evolution of pH, organic matter and ^{226}radium/calcium partitioning in U-mining debris following revegetation with pine trees, *Sci. Total Environ.*, **393**(1), 111–117.

Thoenen, T. et al. (2014) *The PSI/Nagra Chemical Thermodynamic Database 12/07*, PSI Report No. 14-04, ISSN: 1019-0643, Pauf-Scherrer-Institute, Villigen, Switzerland. 445 p.

Tournassat, C. et al. (2007) Two cation exchange models for direct and inverse modelling of solution major cation composition in equilibrium with illite surfaces, *Geochim. Cosmochim. Acta*, **71**(5), 1098–1114.

Turner, D.R. et al. (1996) Approaches to sorption modeling for high-level waste performance assessment, *J. Contam. Hydrol.*, **21**(1–4), 311–332.

Vandenhove, H. et al. (2009) New best estimates for radionuclide solid-liquid distribution coefficients in soils. Part 2. Naturally occurring radionuclides, *J. Environ. Radioact.*, **100**(9), 697–703.

Vidal, M. et al. (2024) *Laboratory sorption-desorption experiments and updated databases on soil sorption and desorption parameters for NOR.* Final version as of 29.02.2024 of deliverable D2.16 of project RadoNorm, 50 p.

Waite, T.D. et al. (1994) Uranium(VI) adsorption to ferrihydrite: Application of a surface complexation model, *Geochim. Cosmochim. Acta*, **58**(24), 5465–5478.

Zachara, J.M. (1990) Sorption of aminophthalene and quinoline on amorphous silica, *Environ. Sci. Technol.*, **24**(1), 118–126.

Zavarin, M. et al. (2022) Community Data Mining Approach for Surface Complexation Database Development, *Environ. Sci. Technol.*, **56**(4), 2827–2838.

Appendix A. SCM data compilation

Table 13 – Compilation of literature SCM data for NOR K_d prediction

Quartz		
Ion	***Reference***	***Comment***
H^+	Zavarin et al., 2022	-
UO_2^{2+}	Zavarin et al., 2022	-
Ra^{2+}	Berka et al., 2001 Olin et al., 1997 Riese, 1982 Zachara, 1990	Ca as chemical analogue for Ra Mean of log*K* from different references, values normalized to reference binding site 2.31 sites/nm² and zero ionic strength Transformed to strong & weak site concept
Pb^{2+}	Reich et al., 2010	log*K* normalized to reference binding site 2.31 sites/nm² and zero ionic strength Transformed to strong & weak site concept
Mg^{2+}	Olin et al., 1997	log*K* normalized to reference binding site 2.31 sites/nm² and zero ionic strength Transformed to strong & weak site concept
Ca^{2+}	Berka et al., 2001 Olin et al., 1997 Riese, 1982 Zachara, 1990	Mean of log*K* from different references, values normalized to reference binding site 2.31 sites/nm² and zero ionic strength Transformed to strong & weak site concept
Al^{3+}	Kuan et al., 2004	Re-Fit of exp. data from KLW04 log*K* normalized to reference binding site 2.31 sites/nm² and zero ionic strength Transformed to strong & weak site concept
Albite		
Ion	**Reference**	**Comment**
H^+	Arnold et al., 2001	-
UO_2^{2+}	Richter et al., 2016	Orthoclase as chemical analogue for albite
Ra^{2+}	Britz, 2018	Sr as chemical analogue for Ra
Pb^{2+}	-	No SCM data available
Mg^{2+}	-	No SCM data available
Ca^{2+}	-	No SCM data available

Al^{3+}	Lützenkirchen et al., 2014	Kaolinite as chemical analogue for albite
Orthoclase		
Ion	**Reference**	**Comment**
H^+	Arnold et al., 2001	Albite as chemical analogue for orthoclase
UO_2^{2+}	Richter et al., 2016	-
Ra^{2+}	Britz, 2018	Sr as chemical analogue for Ra
Pb^{2+}	-	No SCM data available
Mg^{2+}	-	No SCM data available
Ca^{2+}	-	No SCM data available
Al^{3+}	Lützenkirchen et al., 2014	Kaolinite as chemical analogue for orthoclase
Microcline		
Ion	**Reference**	**Comment**
H^+	Arnold et al., 2001	Albite as chemical analogue for microcline
UO_2^{2+}	Richter et al., 2016	Orthoclase as chemical analogue for microcline
Ra^{2+}	Britz, 2018	Sr as chemical analogue for Ra
Pb^{2+}	-	No SCM data available
Mg^{2+}	-	No SCM data available
Ca^{2+}	-	No SCM data available
Al^{3+}	Lützenkirchen et al., 2014	Kaolinite as chemical analogue for microcline
Kaolinite		
Ion	**Reference**	**Comment**
H^+	Turner et al., 1996	-
UO_2^{2+}	Turner et al., 1996	-
Ra^{2+}	Riese, 1982	Sorption on ≡AlOH groups assumed
Pb^{2+}	Reich et al., 2010	Normalized to reference binding site 2.31 sites/nm² and zero ionic strength
Mg^{2+}	Riese, 1982	Ca as chemical analogue for Mg Sorption on ≡AlOH groups assumed

Ca^{2+}	Riese, 1982	Sorption on ≡AlOH groups assumed
Al^{3+}	Lützenkirchen et al., 2014	Sorption on ≡AlOH groups assumed Original measurement in 1.0 mol/L $NaClO_4$
Montmorillonite		
Ion	**Reference**	**Comment**
H^{+}	Stockmann et al., 2022	-
UO_2^{2+}	Stockmann et al., 2022	-
Ra^{2+}	Klinkenberg et al., 2021	-
Pb^{2+}	Bradbury et al., 2005b	Normalized to reference binding site 2.31 sites/nm² and zero ionic strength Transformed to strong/weak1/weak2-site concept
Mg^{2+}	Bruggeman et al., 2010	Normalized to reference binding site 2.31 sites/nm² and zero ionic strength Transformed to strong/weak1/weak2-site concept
Ca^{2+}	Tachi et al., 2001 Ikhsan et al., 2005	Mean of Log*K*-values from different sources, normalized to reference binding site 2.31 sites/nm² and zero ionic strength Transformed to strong/weak1/weak2-site concept
Al^{3+}	Charlet et al., 1993	Mean of Log*K*-values from different sources, normalized to reference binding site 2.31 sites/nm² and zero ionic strength Transformed to strong/weak1/weak2-site concept
Illite		
Ion	**Reference**	**Comment**
H^{+}	Stockmann et al., 2022	Montmorillonite as chemical analogue for illite
UO_2^{2+}	Stockmann et al., 2022	Montmorillonite as chemical analogue for Illite
Ra^{2+}	Klinkenberg et al., 2021	Montmorillonite as chemical analogue for illite
Pb^{2+}	Bradbury et al., 2005b	Montmorillonite as chemical analogue for illite Normalized to reference binding site 2.31 sites/nm² and zero ionic strength Transformed to strong/weak1/weak2-site concept
Mg^{2+}	Bruggeman et al., 2010	Normalized to reference binding site 2.31 sites/nm² and zero ionic strength Transformed to strong/weak1/weak2-site concept

Ca^{2+}	Tachi et al., 2001 Ikhsan et al., 2005	Mean of Log*K*-values from different sources, normalized to reference binding site 2.31 sites/nm² and zero ionic strength Transformed to strong/weak1/weak2-site concept
Al^{3+}	Charlet et al., 1993	Mean of Log*K*-values from different sources, normalized to reference binding site 2.31 sites/nm² and zero ionic strength Transformed to strong/weak1/weak2-site concept
Ferrihydrite		
Ion	**Reference**	**Comment**
H^{+}	Waite et al., 1994	-
UO_2^{2+}	Arnold et al., 2001	-
Ra^{2+}	Chen et al., 2018	Normalized to reference binding site 2.31 sites/nm² and zero ionic strength
Pb^{2+}	Dzombak et al., 1990	-
Mg^{2+}	Dzombak et al., 1990	-
Ca^{2+}	Dzombak et al., 1990	-
Al^{3+}	-	No SCM data available
Goethite		
Ion	**Reference**	**Comment**
H^{+}	Missana et al., 2003	-
UO_2^{2+}	Missana et al., 2003	-
Ra^{2+}	Chen et al., 2018	Normalized to reference binding site 2.31 sites/nm² and zero ionic strength
Pb^{2+}	Mathur et al., 2006	Normalized to reference binding site 2.31 sites/nm² and zero ionic strength
Mg^{2+}	Mathur et al., 2006	Normalized to reference binding site 2.31 sites/nm² and zero ionic strength
Ca^{2+}	Ali et al., 1996	Normalized to reference binding site 2.31 sites/nm² and zero ionic strength
Al^{3+}	-	No SCM data available

Appendix B. IEx data compilation

Table 14 – Compilation of literature IEx data for NOR K_d prediction

Montmorillonite		
Ion	**Reference**	**Comment**
H^+	Fletcher et al., 1989	-
UO_2^{2+}	Stockmann et al., 2022	-
Ra^{2+}	Klinkenberg et al., 2021	-
Pb^{2+}	Gu et al., 2010	-
Mg^{2+}	-	No IEx data available
Ca^{2+}	Stockmann et al., 2022	-
Illite		
Ion	**Reference**	**Comment**
H^+	Gilbert et al., 1965	-
UO_2^{2+}	Bradbury et al., 2005a	-
Ra^{2+}	Bradbury et al., 2005a	Sr as chemical analogue for Ra
Pb^{2+}	Marques Fernandes et al., 2019	-
Mg^{2+}	Tournassat et al., 2007	-
Ca^{2+}	Tournassat et al., 2007	-

Appendix C. Scenarios for the K_d blind prediction exercise

Table 15 – Original information for scenarios no. 1A–C in the K_d blind prediction exercise

Required soil/solid information			
Dataset 1 for blind prediction of K_d values	1A	1B	1C
pH (in contact solution)	7.23	4.96	7.20
The mineralogical composition of the soil (e.g. 20 wt% quartz, 30 wt% feldspar, …)	Kaolinite 60 % (mass %) Quartz 12.9 % Muscovite (mica) 3.5 % Chlorite 6.6 % Goethite 16.7 % Ferrihydrite 0.3 %	Kaolinite 60 % (mass %) Quartz 12.9 % Muscovite (mica) 3.5 % Chlorite 6.6 % Goethite 16.7 % Ferrihydrite 0.3 %	Kaolinite 60 % (wt %) Quartz 12.9 % Muscovite (mica) 3.5 % Chlorite 6.6 % Goethite 16.7 % Ferrihydrite 0.3 %
Cation exchange capacity (CEC)	-	-	-
$CaCO_3$ content (%)*	-	-	-
Texture (textural class) or grain size distribution (clay, silt, sand, …	Grain size < 10 mm	Grain size < 10 mm	Grain size < 10 mm
Specific surface area (SSA)	52.3 m^2/g	52.3 m^2/g	52.3 m^2/g
The composition of the aqueous phase (Na, K, Mg, Ca, Sr, Ba, SO_4^{2-}, Cl^-, dissolved inorganic carbon (DIC)).	0.1 M $NaNO_3$	0.1 M $NaNO_3$	Mg^{2+} 11.3 mg/L Ca^{2+} - Na^+ 1.64 mg/L K^+ 0.57 mg/L HCO_3^- 102 mg/L Cl^- 3.3 mg/L SO_4^{2-} 3.4 mg/L Electrical conductivity: 166 µS/cm
Redox potential E_H	Air atmosphere	Air atmosphere	145 mV
Organic carbon content and/or loss on ignition (LOI)	Below detection limit	Below detection limit	Below detection limit
Solid-liquid ratio and/or porosity	200 mg/50mL	200 mg/50mL	200 mg/50mL
Partial pressure CO_2	air	air	$10^{-2.2}$
Concentration of U in solid [mg U/g solid]	232 µg/g	232 µg/g	225 µg/g
Concentration of Ra in solid [mg Ra/g solid]	-	-	-
Helpful further information			
Dissolved organic carbon (DOC)	-	-	-
Amorphous Fe and Al content (extractable)**	Fe: 1.69 mg/g Al: 1.82 mg/g (mg extracted / g of solid)	Fe: 1.69 mg/g Al: 1.82 mg/g (mg extracted / g of solid)	Fe: 1.69 mg/g Al: 1.82 mg/g (mg extracted / g of solid)

* It means the value derived from the carbonate analyses in the solid phase (following any recommended procedure for the analyses of soils and/or solid environmental matrices).

** There are common procedures for their quantification (for instance, based on oxalate extraction)

Table 16 – Original information for scenario no. 2 in the K_d blind prediction exercise

Required soil/solid information	
Dataset 2 for blind prediction of K_d values	2
pH (in contact solution)	5.6
The mineralogical composition of the soil: Mineral fraction (%) (dry sample)	Quartz 37.93 % Kaolinite 9.93 % Orthoclase 5.98 % Albite 6.12 % Muscovite 9.36 %
Cation exchange capacity (CEC)	-
$CaCO_3$ content (%)*	-
Texture (textural class) or grain size distribution (clay, silt, sand, …	Soil material unsieved
Specific surface area (SSA)	-
The composition of the aqueous phase (Na, K, Mg, Ca, Sr, Ba, SO_4^{2-}, Cl^-, dissolved inorganic carbon (DIC)).	$[Na^+]$ = 4.1 mg·L^{-1}; $[K^+]$ = 70.2 mg·L^{-1}; $[Mg^{2+}]$ = 10.8 mg·L^{-1}; $[Ca^{2+}]$ = 1.82 mg·L^{-1}; $[Cl^-]$ 90.4 mg·L^{-1}; $[SO_4^{2-}]$ = 16.8 mg·L^{-1}; $[Ba^{2+}]$ = 38 µg·L^{-1}
Redox potential E_H	210 mV
Organic carbon content and/or loss on ignition (LOI) Organic matter (%) dry sample	33 %
Solid-liquid ratio and/or porosity	176 g/L (S/L)
Partial pressure CO_2	-
Concentration of U in solid [mg U/kg dry matter]	1206 mg/kg
Accessible uranium in %	0.9
Concentration of Ra in solid [ng Ra/kg dry matter]	518
Accessible radium in %	3.2
Helpful further information	
Dissolved organic carbon (DOC)	-
Amorphous Fe and Al content (extractable)**	-

* It means the value derived from the carbonate analyses in the solid phase (following any recommended procedure for the analyses of soils and/or solid environmental matrices).

** There are common procedures for their quantification (for instance, based on oxalate extraction)