



Treball Final de Grau

**Creation of a database of radium distribution coefficients in soils.
Creació d'una base de dades de coeficients de distribució de radi
en sòls.**

Guillem Sabaté Herrero

June 2021



UNIVERSITAT DE
BARCELONA

B:KC Barcelona
Knowledge
Campus
Campus d'Excel·lència Internacional

Aquesta obra esta subjecta a la llicència de:
Reconeixement–NoComercial–SenseObraDerivada



<http://creativecommons.org/licenses/by-nc-nd/3.0/es/>

I hope that posterity will judge me kindly, not only as to the things which I have explained, but also to those which I have intentionally omitted so as to leave to others the pleasure of discovery.

René Descartes

I would like to show my deepest appreciation for Prof. Miquel Vidal for his guidance and counseling provided during the extend of this project and beyond, his enthusiastic encouragement and critical judgement were of the outmost importance for the completion of this research work.

REPORT

CONTENTS

1. SUMMARY	3
2. RESUM	5
3. INTRODUCTION	7
3.1. Soil phases and properties affecting radionuclide interaction	7
3.2. The solid-liquid distribution coefficient (K_d)	9
3.3. Proposing K_d best estimates	11
3.4. Radium	12
3.4.1. Variables influencing the distribution coefficient of radium	13
4. OBJECTIVES	15
5. METHODS	16
5.1. Bibliographical research	16
5.2. Data collection and filtering	17
5.3. The database	18
5.4. Correlation analyses and data variability	20
5.5. Propose K_d (Ra) best estimates and explore data categorization	22
6. RESULTS AND DISCUSSION	24
6.1. Bibliographic search results	24
6.2. The database	25
6.3. Statistical analyses	27
6.3.1. Evaluation of the effect of the experimental approach on K_d (Ra) variability	27
6.3.2. Evaluation of the effect of soil properties on K_d (Ra) variability: correlation analysis	28
6.3.2.1. Univariate linear correlations	29
6.3.2.2. Multivariate linear correlations	29
6.3.3. Proposal of K_d (Ra) best estimates and description and reduction of K_d (Ra) variability for selected soil groups	31
6.3.3.1. K_d (Ra) best estimates and variability from soils grouped based on pH	31
6.3.3.2. K_d (Ra) best estimates and variability from soils grouped based on $[\text{CO}_3^{2-}]$	33
6.3.3.3. K_d (Ra) best estimates and variability from soils grouped based on $\text{CEC}/(\text{Ca}+\text{Mg})_{\text{ss}}$	34

7. CONCLUSIONS	36
8. REFERENCES AND NOTES	38
9. ACRONYMS	39
APPENDICES	41
Appendix 1: Database description	43
Appendix 2: List of publications in the database	46
Appendix 3: CDF <i>Matlab</i> program. Example for $CEC(Ca+Mg)_{ss}$	48

1. SUMMARY

Naturally occurring radionuclide materials (NORMs) are geological materials rich in indigenous radioactive elements. Some of these naturally occurring radionuclides (NOR) are present in many geological resources (e.g., mining ores and fossil fuels), whose exploitation might involve extracting NORs and exposing them into the surface. Then, they are more exposed to weathering and environmental factors that may increase their mobility. Consequently, understanding their mobility is important for risk assessment models involving radionuclides.

A possible approach to study the mobility of a radionuclide is using the distribution coefficient (K_d). It describes the relation between the concentration of the element in the solid phase and in the soil solution at equilibrium. So, a high K_d implies that the radionuclide is mostly sorbed into the solid fraction of a soil, i.e., it has a lower mobility.

Ideally, parametric models are to be proposed to predict the K_d as a function of soil properties. However, sorption involves many and complex mechanisms, which make it difficult to build such models. Alternatively, some variables may be used to propose K_d best estimates with a quantified variability for soil groups (which are to be made of intervals from those same variables). This was the case of this project, where a prospective analysis for radium (a NOR) was to be carried out.

In order to do so, a K_d (Ra) database of different soils was expanded and reviewed with literature research, both to add new entries and to improve the soil and experimental characterizations. The inclusion criteria and data organization had to be rearranged to fulfill this goal.

Once the dataset had been expanded, uni- and multivariant correlation analyses had to be carried out by creating new partial datasets for each variable, mainly to discern which soil properties were significant in reducing K_d (Ra) variability. Both pH and carbonate concentration (CO_3^{2-}) in the soil solution proved significant in the univariant correlation analyses. Additionally, the multivariant analyses showed many more significant correlations, but most importantly two of them: $\text{CEC}/(\text{Ca}+\text{Mg})_{\text{ss}}$ and $(\text{Ca}_{\text{exch}}+\text{Mg}_{\text{exch}})/(\text{Ca}+\text{Mg})_{\text{ss}}$. These last two combinations were so relevant to observe because they had been proven significant for other earth alkaline metals. Interestingly, some of the correlations found in this project are exceptionally good, which means that some of these variables could end up being used in establishing parametric models.

Finally, K_d (Ra) best estimates for grouped soils according to the value of certain variables were proposed. These variables were pH, CO_3^{2-} , and $\text{CEC}/(\text{Ca}+\text{Mg})_{\text{ss}}$. All these proposed best

estimates presented a smaller variability than the K_d (Ra) best estimated of the overall database. Thus, grouping soils according to their own properties helped to reduce the variability of the K_d (Ra) best estimate.

Keywords: radium, distribution coefficient, K_d variability.

2. RESUM

Els materials que contenen radionúclids d'origen natural (NORMs, sigles en anglès) són materials geològics rics en elements radioactius autòctons. Alguns d'aquests radionúclids d'origen natural (NOR, sigles en anglès) es troben a diversos recursos naturals (per exemple, materials miners i combustibles fòssils), l'explotació dels quals probablement implica l'extracció de NORs i la seva exposició a la superfície. Llavors, aquests estan més exposats a factors meteorològics i ambientals que poden incrementar la seva mobilitat. Conseqüentment, entendre la seva mobilitat és important per a models d'avaluació de riscos que incloguin radionúclids.

Una possible aproximació a l'estudi de la mobilitat d'un radionúclid és emprar el coeficient de distribució (K_d). Aquest descriu la relació entre la concentració de l'element a la fase sòlida i a la solució del sòl en equilibri. Llavors, una elevada K_d implica que el radionúclid està majoritàriament adsorbit a la fracció sòlida del sòl, és a dir, que presenta una menor mobilitat.

Idealment, s'han de proposar models paramètrics per a predir la K_d com una funció de propietats del sòl. No obstant, la sorció inclou múltiples i complexos mecanismes, els quals dificulten construir aquests models. Alternativament, algunes variables es poden emprar per a proposar *best estimates* de K_d s amb una variabilitat quantificada per a grups de sòls (els quals s'han de fer a partir d'interval·ls d'aquestes mateixes variables). Aquest va ser el cas per a aquest projecte, on es va dur a terme una anàlisi prospectiva per al radi, un NOR.

Per a dur-ho a terme, es va ampliar i revisar una base de dades de K_d (Ra) de diferents sòls amb una cerca bibliogràfica, tant per incloure noves entrades com per millorar les caracteritzacions experimentals i dels sòls. A més, es va haver de modificar criteris d'inclusió i l'organització de les dades.

Una vegada acabada l'ampliació de la base de dades, es va procedir a dur a terme anàlisis de correlacions uni- i multivariants amb la creació de bases de dades parcials per a cada variable, principalment per a discernir quines propietats del sòl van resultar significatives en reduir la variabilitat de la K_d (Ra). Ambdós, el pH i la concentració de carbonat (CO_3^{2-}) a la solució del sòl, es van mostrar significatius a les anàlisis de correlacions univariants. Addicionalment, les anàlisis multivariants van mostrar algunes correlacions significatives més, però les més importants essent: $\text{CEC}/(\text{Ca}+\text{Mg})_{\text{SS}}$ and $(\text{Ca}_{\text{exch}}+\text{Mg}_{\text{exch}})/(\text{Ca}+\text{Mg})_{\text{SS}}$. Aquestes dues combinacions van ser tant rellevants per a l'estudi per què ja s'havien mostrat significatives per a altres metalls alcalinoterris. De manera interessant, algunes d'aquestes correlacions van

resultar excepcionalment bones, fet que implica que aquestes variables podrien emprar-se per establir models paramètrics.

Finalment, es van proposar K_d (Ra) *best estimates* per a sòls agrupats d'acord amb el valor d'algunes variables. Aquestes variables eren pH, CO_3^{2-} i $\text{CEC}/(\text{Ca}+\text{Mg})_{\text{ss}}$. Totes aquestes estimacions van presentar una variabilitat menor que la K_d (Ra) estimades per a la base de dades sencera. Així, agrupar sòls d'acord amb les seves propietats ha ajudat a reduir la variabilitat de les K_d (Ra) *best estimates*.

Paraules clau: radi, coeficient de distribució, variabilitat de la K_d .

3. INTRODUCTION

Naturally occurring radionuclides (NOR) are a major radiation source, contained in many geological materials, that are a potential hazard to human beings. Mineral exploitations of naturally occurring radionuclide materials (NORM) increase human exposure to them. So, it is of great importance to treat any waste that may contain NORMs appropriately to avoid any kind of environmental contamination, which is not easy at all given the number of industries that use materials containing NORs, e.g., fossil fuels, mining ores, etc. (Vandenhove and Hees, 2007).

To control the disposal of NORMs is of great relevance due to the ability of most radionuclides to migrate. Most of them can incorporate into the trophic chain of an ecosystem because some plants tend to accumulate these elements (Al-Hamaneh *et al.*, 2016).

Consequently, knowledge regarding the mobility of the radionuclides is vital to predict human exposure to the radioactivity in any risk assessment model applied in the management of these materials. A more mobile radionuclide is to increase human radioactivity exposure once it spreads into the environment, as such compounds are to be more available to be incorporated by plants into the trophic chain, thus, increasing human internal dose (Vandenhove and Hees, 2007).

The mobility of a radionuclide in the soil-plant system is a complex process affected by a high number of mechanisms. This project will focus on the radionuclide sorption using the distribution coefficient, which in general terms estimates the capacity of a soil to retain a radionuclide, thus decreasing the possibility of a NOR to be incorporated by a plant (Al-Hamaneh *et al.*, 2016).

3.1. SOIL PHASES AND PROPERTIES AFFECTING RADIONUCLIDE INTERACTION

In a soil, three phases coexist in equilibrium: a liquid, a solid, and a gaseous phase. All components present in a soil are in equilibrium between its different phases. Yet, when a non-volatile substance is being studied, like most radionuclides, the role of the gaseous phase can be considered as negligible.

A radionuclide sorbed into the solid phase is not as readily available to disperse into other environments, whereas the fraction in the liquid phase is immediately ready to migrate. So, understanding soil-radionuclides interactions is key in predicting the expected sorbed fraction.

The solid phase is made of the mineral and the organic fractions. The former includes metallic oxides, carbonates, silicates, and other minerals. Depending on the particle size of these

components, it will present a different texture according to USDA classification: sand (from 50 μm to 2 mm), clay (less than 2 μm), and silt (from 2 to 50 μm). Mixtures of sands and clays are to produce loamy soils (Yolcubal *et al.* 2004). So, any element is expected to behave differently under different soil textures, due to each texture providing different number of sorption sites available. Thus, textural data can be used as factor to consider when studying the interaction of a radionuclide in a soil (IAEA, 2010).

On the other hand, the organic fraction is made both of organic compounds and biological residues that undergo a constant change and degradation, mostly done by microorganisms. Yet, organic matter tends to form humus, which are decomposition products and biological residues and tissues that tend to adopt a black-brownish color. So, the interactions of a radionuclide with a soil are expected to be highly sensitive to the presence of organic matter, due to the expected different interaction of the radionuclide between the mineral and the organic fractions (Brady, 1984).

The composition of the liquid phase of a soil (soil solution) affects radionuclide mobility. Some of its components may compete with the radionuclide for the sorption sites or may produce chelating or precipitation reactions that will affect the sorption degree. The main components of the soil solution to take into account are free cations (Ca^{2+} , Mg^{2+} , Na^+ , K^+), anions (CO_3^- , SO_4^{2-} , Cl^-), complexed metals; and the organic matter present in the liquid phase (DOC – Dissolved Organic Carbon).

Other parameters that will affect the soil-radionuclide interaction are pH and the CEC. The former is a measure of the concentrations of protons in which a soil is naturally found, and it is likely to influence the sorption process of a substance. The cationic exchange capacity (CEC) is a parameter that quantitatively defines the ability of a soil to exchange cations. Therefore, its units are $\text{cmol}_c \text{ kg}^{-1}$. Soils with a high organic and/or clay fractions contents are expected to present a higher CEC (Brady, 1984).

Nonetheless, there are even more parameters that may influence the soil-radionuclide interaction. Perhaps, one of the most relevant would be the soil mineralogy. The affinity of a radionuclide for a soil definitely depends on knowing the sort of minerals that it contains. But providing an exhaustive soil mineralogy is usually not easy and it is not as widely reported in literature as other parameters.

Likewise, another relevant soil property is the redox potential, as it may cause a change in the oxidation state of the element under study. If the radionuclide can form different species at different oxidations states, the redox potential is to be of paramount importance because these species may have very different chemical properties and produce different sorption behavior (Subdiaga *et al.*, 2019).

3.2. THE SOLID-LIQUID DISTRIBUTION COEFFICIENT (K_d)

The solid-liquid distribution or partition coefficient (K_d) can be understood as the ratio between the concentration of a target analyte (in this case, a radionuclide) in the solid phase (sorbed) and its concentration in the solution in equilibrium conditions:

$$K_d = \frac{\text{Sorbed concentration}}{\text{Concentration in solution}} \quad \text{Eq. 1}$$

The units of the distribution coefficient are usually expressed as $L \text{ kg}^{-1}$.

This coefficient provides a very general picture of the interaction of the substance with the soil. It describes the sorption of a substance, regardless of the mechanism. As a result of this, the K_d can be used to assess potential risks in case of an environmental contamination (Vandenhove and Hees, 2007).

Still, sorption is a process that depends on the target element and the soil-substance interactions. So, soil characterization is extremely important in establishing prediction models for proposing best estimate K_d in new soils. However, sorption measurements are also sensitive to the experimental method used in a determination (Vandenhove and Hees, 2007). Thence, it is important to identify all variability sources, as it is shown in Figure 1:

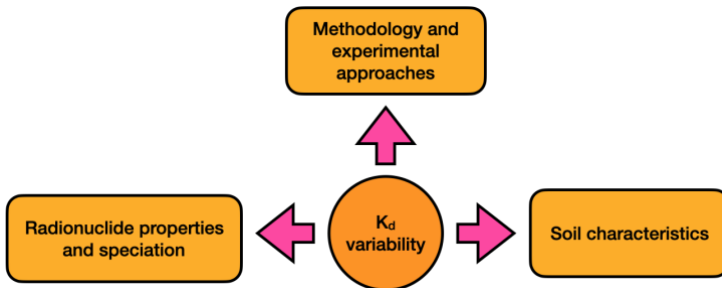


Figure 1. K_d variability sources.

However, the concept of K_d includes several assumptions and hypothesis that ought to be mentioned:

- Regardless of the method used to measure the K_d , it is assumed that the experiment is carried out under **equilibrium conditions**. So, all determinations that imply spiking or artificially contaminating a controlled soil are to be provided with a reasonable amount of time to achieve equilibrium. The term reasonable implies that in most cases it is going to be a pseudo-equilibrium (Vandenhove and Hees, 2007).
- Most experiments are to be planned to obtain a **reversible K_d** . Nonetheless, each radionuclide interacts differently with a soil, and some might irreversibly sorb a fraction if enough time is provided (Vandenhove and Hees, 2007).

Furthermore, the K_d can be obtained from long- and short-term experiments. The former are to be carried out to analyze an anthropogenic or indigenous elements, while the latter usually consist on spiking soils to measure the coefficient after a short equilibration period. For a few elements, like Cs, it has been proven that long-term experiments lead to the quantification of higher K_d values than short-term experiments (Ramírez-Guinart, 2020b), although this pattern has not been observed for other radionuclides, sometimes due to the lack of sufficient data (Ramírez-Guinart *et al.*, 2020a; Ramírez-Guinart *et al.*, 2020c). Table 1 summarizes the methodological approaches for the determination of K_d s.

Summarizing, short-term experiments are usually performed at a laboratory level, which implies the control over most of the soil and experimental parameters. Thus, they can better be used to study the effect of a limited set of variables, e.g., pH, OM, etc. Whereas long-term experiments are usually performed on contaminated aged samples, where the main focus is to obtain a K_d without going deeper in studying any properties in depth. (These are usually desorption experiments.) In contraposition, *in situ* experiments involve samples with little subsequent manipulation (the soil solution tends to be the pore water or river water, depending on the sample).

So, it is going to be of great relevance to relate any K_d not only to the soil properties, but also to this methodological approach, as it is crucial for risk assessment models to know whether the time of exposure of the contaminant to a soil will influence the K_d . If the time of exposure proves of significance, several K_d values are going to be needed to assess the effect of a radioactive deposit in the long-term.

Experimental approach	Description
Short-term	
<i>in situ</i>	K_d of an anthropogenic element in a recently contaminated soil, determined as the ratio between the concentration in the solid phase and in the pore water.
Sorption	K_d determined using batch experiments by means of spiking soil samples at a known concentration and waiting an equilibration time before doing the measurements. The K_d is obtained from the relation between the known initial concentration and the final concentration in the contact solution, after the equilibrium was reached.
Desorption	K_d of an anthropogenic element determined using extraction batch experiments in recently contaminated soils. The K_d is obtained from the relation between the known initial concentration and the final concentration in the contact solution, after the equilibrium was reached. K_d of an indigenous element in a contaminated area determined as the ratio between the concentration in the solid phase (obtained by performing a mild extraction) and in the contact solution.
Long-term	
<i>in situ</i>	K_d of an indigenous element, or an anthropogenic element in a soil contaminated at least a year before, determined between the ratio in the solid phase and in the pore water.
Desorption	K_d of an anthropogenic element determined using extraction batch experiments in soils contaminated at least a year before. The K_d is obtained from the relation between the known initial concentration and the final concentration in the contact solution, after the equilibrium was reached. K_d of an indigenous element determined as the ratio between the concentration in the solid phase (obtained by performing a strong extraction) and the contact solution.

Table 1. Experimental approaches for measuring distribution coefficients.

3.3. PROPOSING K_d BEST ESTIMATES

As a result of the assumptions that the K_d implies, modelist approaches to predict the sorption of a substance using the distribution coefficient are to have high inherent variability, as it is in itself a variable parameter. Especially, considering that soils are very complex matrixes and it is unrealistic to think that all its defining variables are to be available for any analysis. Thence, the building of these models is to be made by generating large K_d databases, measured in different soils with very detailed soil and methodological characterizations (Ramírez-Guinart *et al.*, 2020a).

Given the complexity of soils and the different experimental approaches, it is unrealistic to think that all variables can be simultaneously studied. Although, the more exhaustive the soil characterization is, more parameters can be studied. Thus, in absence of a complete soil characterization, parametric predictive models use whatever properties (previously identified as significant) are available to describe much of the variability of the K_d , which must satisfactorily

relate to experimental K_{ds} (high correlations and and low ordinate). However, this has been achieved for very few radionuclides before.

In case the previous scenario is not feasible, at least these variables can be used to group soils in intervals of these same variables, aiming at building probability function to describe variability associated with K_d best estimates for each group. These best estimates are to be calculated for each interval of a categorized variable and with quantified variability (Ramírez-Guinart *et al.*, 2020a).

Whether a parametric model is build or best estimates are proposed for intervals of different variables, it will greatly increase the predicting capabilities to infer a K_d in other soils, which may be of great importance in an environmental short- and long-term assessments, i.e., evaluating the impact of a recent accidental radionuclide spillage or the long-term management of a radionuclide deep geological repository (Vandenhove and Hees, 2007).

3.4. RADIUM

Radium (Ra) is target radionuclide that is studied in the current project. It is the heaviest of the alkaline earth metals and it is a natural radionuclide as radium-226 (half-life of 1600 years), although other isotopes can be artificially made. In addition, it is a decay product of uranium-238 (see Figure 2) and it is also an alpha particle producing radionuclide. Hence, it can be considered as a potential hazard due to its radiological impact on ecosystems (Bordelet *et al.*, 2013).

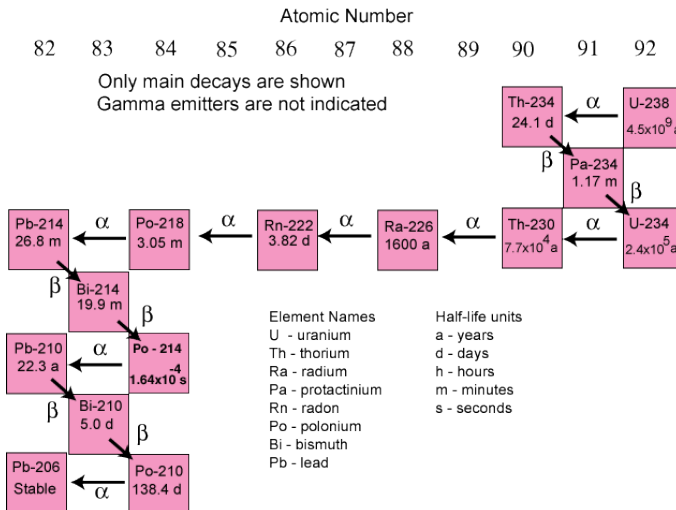


Figure 2. Uranium-238 decay chain and half-lives. From the U.S. Geological Survey.

On the other hand, Ra-226 ends up generating Rn-222, which is an even more dangerous radionuclide because it is a gas that may affect humans directly through the respiratory tract and accumulating itself in the lungs. According to the National Institutes of Health (NIH) indoor radon exposure is the second cause of lung cancer in the U.S., just after cigarette smoking (NIH, 2021).

Despite the danger posed by radium contamination, there is not much published literature regarding its distribution coefficient in different soils. The existing published K_d compilations are scarce and with not so many entries, as in much of the published literature radium is usually a more secondary element studied in parallel to others, like Sheppard *et al.*, 1984. Therefore, there is a special need to create an expanded database of K_d (Ra) to deal with all these deficiencies. Especially, given that some of these previous compendiums have presented K_d (Ra) differing in up to 5 orders of magnitude, which forecast an immense variability in any analyses of this radionuclide (Kumar *et al.*, 2020).

3.4.1. Variables influencing the distribution coefficient of radium

As an alkaline earth element, Ra is found in +2 oxidation state. Hence, it ought to be influenced by any parameter that would affect the sorption of any other cation, like the superficial charge of a soil or the presence of other competing cations (Kumar *et al.*, 2020; Nathwani and Philips, 1979b).

Given that a sandy soil usually provides less sorption sites than a clayey soil, sandy soils are expected to produce lower K_d (Ra) (Sheppard *et al.*, 1984). As marine sediments have a mostly sandy texture and they are very washed up, they are also expected to produce lower coefficients than soils. So, textural properties are expected to be of relevance in studying K_d (Ra) variability, although some contradicting information exist regarding on this matter: clay content may be expected to present a weak correlation with K_d (Ra) (Vandenhove and Hees, 2007; Simon and Ibrahim, 1990).

Besides texture, the following general parameters are expected to affect K_d (Ra):

- pH. A higher pH has been directly linked to higher K_d (Ra) (Kumar *et al.*, 2020).
- Organic matter content (OM). Higher OM has been linked to higher K_d (Ra) (Vandenhove and Hees, 2007).

- Sulphate. A co-precipitation reaction with BaSO_4 as Ba(Ra)SO_4 has been extensively described in literature. So, they ought to be directly correlated to the K_d (Ra) (Sheppard *et al.*, 1984; Grandia *et al.*, 2008).
- Carbonate. Radium is likely to present lower mobility in presence of carbonates due to precipitation reactions as RaCO_3 (Sheppard *et al.*, 1984; Grandia 2008).
- Cationic Exchange Capacity (CEC). It is expected to be directly correlated with the K_d (Ra) (Sheppard 1984). So, given that the organic matter is to produce a higher CEC, organic matter might also be expected to be directly correlated with the K_d (Ra) (IAEA 2014).
- Calcium and magnesium. Radium is expected to be free and competing with other divalent cations for sorption sites from pH 4 to 8. (Sheppard *et al.*, 1984). Ca has been proven to be inversely correlated to the K_d (Ra) by Nathwani and Philips, 1979b. An explanation for this fact would be that these cations are smaller than radium, so they can access the sorption sites more effectively than Ra.

More recent experiments have proven that the ratios $\frac{\text{CEC}}{(\text{Ca}^{2+} + \text{Mg}^{2+})_{SS}}$ and $\frac{\text{Ca}_{\text{exch}}^{2+} + \text{Mg}_{\text{exch}}^{2+}}{(\text{Ca}^{2+} + \text{Mg}^{2+})_{SS}}$ might be used to better explain variability of K_d (Sr) (Gil-García *et al.*, 2011). So, given that both Ra and Sr are earth alkaline metals, they are expected to present a similar sorption behavior.

Establishing prediction models with the previous parameters might have some success. However, before doing so, the evidence to back the previous statements ought to be reinforced, which is one of the goals of this project.

4. OBJECTIVES

The main goal of this project was to expand a database of distribution coefficients of radium (K_d (Ra)), and in its turn make a prospective study of which methodology and soil factors can help reduce the variability of K_d (Ra). To achieve this, a set of more enclosed goals was established:

- To perform an extensive bibliographic research to expand and critically review a database created by a fellow peer of radium distribution coefficients from published literature and *grey* literature.
 - o To add new entries to the database and describe important soil parameters, e.g., pH texture, etc.
 - o To review and adapt the previously included entries, with the purpose of checking their fulfillment of the inclusion criteria and to adapt the entries to the new criteria, if necessary.
- To study the main factors involved in radium sorption process in soils and, thus, in K_d variability.
- To perform a statistical analysis aiming to establish how each studied variable contributes to reduce the variability of K_d (Ra).
 - o To assess if the grouping of distribution coefficients of radium according with the experimental method by which they were obtained is a valid strategy to reduce the variability of a proposed K_d (Ra).
 - o To perform univariant and multivariant correlation analyses of the studied variables.
- To create partial datasets with the significant variables according to the correlation analyses, designed according to newly defined intervals for each variable. So, they can be used to produce and calculate best estimates (PCT50: 50th percentile and GM: Geometric Mean) and visualize the variability of the K_d (Ra) with probability functions, i.e., cumulative distribution functions (CDFs).

5. METHODS

This project was divided into 5 clear phases from the beginning. All of them are going to be detailed in this section. But the following diagram provides as a summary of them all, and provides a quick sum up of the processes involved in this project and the methodology:

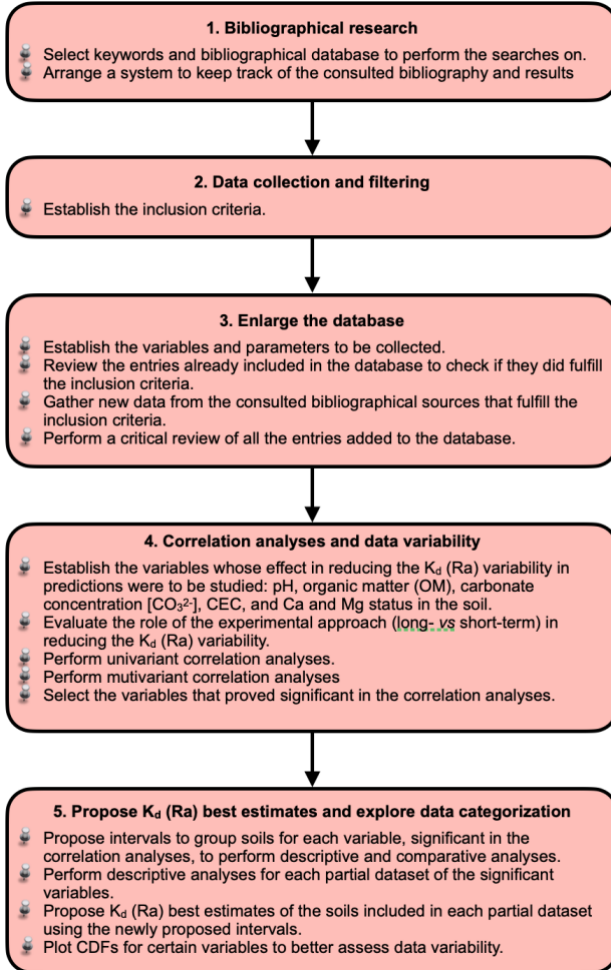


Figure 3. Diagram summarizing the methodology and steps followed in this project.

5.1. BIBLIOGRAPHICAL RESEARCH

A systematic bibliographical research was done regarding distribution coefficient data of radium in soils. The database used for the search was *Web of Science*, using only the *Core*

Collection. For the purpose of keeping a record of the consulted literature, an export of the results and keywords used to an *Excel* spreadsheet was made.

After making several testing searches, it was decided that the group of keywords to be used were: 'radium', 'soil', and 'distribution'; 'radium', 'soil', and 'kd'; and 'radium', 'soil', and 'partition'. Thus, after having all the results for a particular search in a spreadsheet, a consult through the *Web of Science* was made for each of the titles and abstracts. If after reading both, the article was deemed worthy, the original source was consulted. Otherwise, it was discarded before continuing down the list.

Additionally, after consulting the paper, a systematic classification was made. So, a paper could be tracked whether if it had been included into the database, or if relevant sources were obtained from it, or simply if it had been discarded. The tags used to keep the results ordered were: 'newly included into the database', 'already included', 'lead to other literature', 'consulted but discarded', 'discarded', and 'not found'. This last one is where all the papers that were considered worthy of consulting but could not be found were included; either because they were quite old and there was not an available online copy, or simply because it could not be found. The sources classified as 'discarded' were the ones for which only the abstract and titles were read and were directly dismissed.

All the papers classified as 'newly included into the database' were properly incorporated into the database and all available information was extracted. Afterwards, these sources were stored in a bibliographic manager: *Mendeley*.

5.2. DATA COLLECTION AND FILTERING

The main inclusion criterion for the database for any K_d (Ra) was that it had to emanate from an experimental source. Therefore, the database does not include neither K_d (Ra) derived from parametric equations nor averaged values from other data compilations; no data from mass-transport experiments was to be used either. Furthermore, pure phases distribution coefficients were neither accepted into the database, as this approach is out of the scope of the stated objectives of this project.

On the other hand, measurements under extreme soil conditions that would be unfeasible or unlikely in nature were also discarded. So, only K_d (Ra) between the pH range of 2 and 12 were

accepted. Yet, measurements of different isotopes were accepted because there ought not to be any differences between them in relation to their sorption degree.

Besides, entries providing K_d (Ra) by altering operational variables regarding the experimental approach were to be grouped as a single entry by means of the geometric mean (GM), e.g., solid-to-liquid ratio or equilibration time. All such factors were deemed as not relevant in the scope of this project. Thus, only individual K_d (Ra) values which differed in relevant soil properties were considered, like soil pH or organic matter content.

Summarizing, the database was only to include coefficients from soils, subsoils (>1m in depth), water sediments, gyttjas (which are a mud made of organic and mineral matter found close to certain lakes), and tills (which are a mixture of unsorted and non-stratified materials deposited by glacial ice). Nonetheless, all these geological materials were to be considered as soil analogues. Which was especially relevant for those coefficients which had to come from direct experimental determinations in the lab, under conditions in which they were expected to be found in nature.

So, all the reviews containing K_d (Ra) databases found during the forementioned systematic search were categorized as 'consulted but discarded'. They were consulted to get to the original sources to include them into the database if they ended up providing suitable information.

5.3. THE DATABASE

At the beginning of this project, a previous *Excel* database was provided that included 143 K_d (Ra) values of 17 different sources, both of published literature and technical reports. So, the task ahead was to expand it and review the present data to assert their inclusion into the database. At its initial state, there were only 36 columns to fill information in. However, modifications were made in order to improve the database and include further information. Some variables were renamed and readapted, others were newly added, and some simply removed. So, 26 new variables were added in addition to the original ones, making a total of 62 columns in the definitive database, each of them storing different kinds of information, e.g., bibliographical or soil characterization. Table 11 (see Annex 1 page 43-45) displays all the variables that the definitive database contained.

There were important differences between both the present and the original database, which are henceforth summed up. Firstly, despite not changing the bibliographical parameters, soil

parameters were greatly modified. The biggest change was the reorganization of soil parameters to include data from: the solid phase, the liquid phase or pore water, and the contact solution. The original database made only distinction between solid and liquid phases. Nonetheless, many sources provide both, data from the original pore water and from the contact solution of their experiment. Although the concentration in the contact solution is usually more relevant when provided, it was very important to include all the possible data for a future use of the database.

Secondly, the distinction of sorption and desorption coefficients was eliminated, as it was redundant considering the addition of the variable named as 'Experimental tag approach'. Here, all entries were classified according to their experimental approach, which they are explained in Table 1.

Thirdly, another soil classification strategy was used, which grouped them according to their organic matter content (OM) and textural data. So, for a soil to be included in the 'organic' group, it had to contain more than 20% of OM. Thus, all excluded soils were classified as 'mineral' soils; which could again be subdivided in 'sandy' soils (sand fraction $\geq 65\%$ and clay fraction $< 18\%$), 'clayey' soils (clay fraction $> 35\%$), and 'loamy' soils (those that were neither 'sandy' nor 'clayey') (Ramirez-Guinart *et al.*, 2020a). If one source did not provide enough information to establish to which category it should be included, it was classified as 'unspecified'. Although, if it was stated as a mineral soil, but it was not possible to subclassify the soil, the entries were just tagged as 'mineral' or 'mineral unspecified'.

Finally, an extra column named 'Observations' was also added. It was made for a personal purpose at first, so as to keep personal notes during the data analysis. After fulfilling its purpose, it was decided to keep it there for the sake of the future researcher who wants to study the database. Yet, it still holds some information that was considered relevant to keep, but it is not related directly to the study of the distribution coefficient.

There are several other columns, whose usefulness is rather doubtful due to the scarcity of such information in literature. Yet, it was considered that as the data of these variables had already been gathered, it should be left to decide to another researcher who might want to enlarge the database further if they should be removed or not. Some examples of them could be the 'ionic strength of the contact solution' or the 'surface area'. Additionally, it should be considered with chemical criteria if these parameters are expected to be relevant in predicting the sorption process of radium.

Besides, other columns were to simply provide a general picture of the database data and the type of soils that they contained. A fair example would be the 'sample tag'. Assigning a tag to a soil may in the future help stratify data further, but for now it just provides information about what sort of soils are included.

On the other hand, there is one very important remark left to detail. When considering the effect of a variable like the ionic concentrations, the concept of *real concentration* was always used. Which means that when two concentrations of the same substance are given for the soil pore water and for the contact solution, the one to be considered was usually the latter, as it was representative of the actual value when obtaining the K_d (R_a). Still, if both values were available, both were included in their respective variable for future studies because the first one provided the experimental conditions and the other further soil description. Nonetheless, if the contact solution data was not present, the soil concentrations were used instead.

Finally, it should be mentioned that the columns 'Contamination level' and ' ^{226}Ra in solution' are generally irrelevant. Yet, they were kept so as to discriminate any abnormality found in the partition coefficients that cannot be explained by any other variable. A huge addition of radium could affect its sorption degree, when compared with a sample with anthropogenic radium.

5.4. CORRELATION ANALYSES AND DATA VARIABILITY

Several statistical programs have been used to perform all the correlation and statistical analysis. An *Excel* spreadsheet was used to develop the database and most basic calculations. Yet, for more complex statistical tests *SPSS 26* and *Statsgraphics Centurion 18* were used. *SPSS* was used for the descriptive calculations for each variable and case, and the univariant and multivariant regression analysis. Nonetheless, *Statsgraphics* software was used to perform the Fisher's Least Significant Difference test (FLSD), due to *SPSS* only allowing to do the test as a *post-hoc* test of an ANOVA.

When doing all statistical analyses for a target variable, a previous data selection had to be done. The goal was to create subdatasets or partial datasets for each parameter, where there were no duplicated K_d (R_a). For example, when considering the pH variable, all K_d (R_a) obtained from the same soil under different pH had to be considered as separate entries for the subset. However, when analyzing a correlation with a soil parameter that was not modified (for instance the organic matter content), only one entry from the pH-varying experiments was considered

(specifically that entry closest to soil natural conditions when possible, or a geometric mean of all entries from the pH experiments). The reasoning behind such a laborious selection was to not overestimate a soil when studying a variable by eliminating repetitions.

In order to study the effect of a variable in the partition coefficient, *log-log* linear correlation analyses were done, both univariant and multivariant, under the assumption that the $\log K_d$ data array is a normally distributed variable. As a result, this project will show many correlations, which were checked with an *F* test to assess if the correlation was significant. Thus, discriminating which variables are relevant to study the K_d (Ra).

On the other hand, the FLSD test was used to compare means between the overall data, and the long- and short-term groups, simultaneously. So, it made possible to distinguish which groups individually had significant differences with each other. After performing the descriptive analyses, the *antilog* conversion had to be made to obtain desired statistical parameters.

Additionally, a strategy was devised to further reduce the variability of the different datasets and improve any correlations observed. As previous bibliography suggests for other earth-alkaline metals (like Cs), long-term coefficients ought to be generally greater than short-term ones (Ramírez-Guinart, 2020b). Consequently, all analyses were carried out for each variable for the overall data, and the long- and the short-term groups in parallel. Preliminary results of the comparison between long- and short-term, although not conclusive, also advised on performing separate analyses for both groups.

The variables selected to perform the univariant correlations tests on were pH, organic matter (OM), carbonate concentration, and CEC. So, a subdataset was to be created for each of these variables.

For the multivariant analyses, which involved creating additional subdatasets, several combinations between the significant variables (from the univariant correlation analyses) and Ca and Mg status in the soil were tested: the ration between the CEC and calcium and magnesium concentrations $\left(\frac{\text{CEC}}{(\text{Ca}^{2+} + \text{Mg}^{2+})_{ss}}\right)$, the ratio between exchangeable and non-exchangeable calcium and magnesium $\left(\frac{\text{Ca}_{\text{exch}}^{2+} + \text{Mg}_{\text{exch}}^{2+}}{(\text{Ca}^{2+} + \text{Mg}^{2+})_{ss}}\right)$, pH and OM, pH and $\frac{\text{CEC}}{(\text{Ca}^{2+} + \text{Mg}^{2+})_{ss}}$, pH and $\frac{\text{Ca}_{\text{exch}}^{2+} + \text{Mg}_{\text{exch}}^{2+}}{(\text{Ca}^{2+} + \text{Mg}^{2+})_{ss}}$, and carbonate and OM.

However, the multivariate correlation $\left(\frac{\text{Ca}_{\text{exch}}^{2+} + \text{Mg}_{\text{exch}}^{2+}}{(\text{Ca}^{2+} + \text{Mg}^{2+})_{\text{SS}}}\right)$ was the only one not done between significant factors. This correlation was done because previous bibliography of Sr, another earth-alkaline metal, shows a strong correlation between this parameter and the K_d (Sr) and a similar behavior should be expected for K_d (Ra) (Gil-García *et al.*, 2011).

5.5. PROPOSE K_d (Ra) BEST ESTIMATES AND EXPLORE DATA CATEGORIZATION

After performing the correlation analyses, the significant variables from the univariate correlation tests were selected to propose K_d (Ra) best estimates. Firstly, the data in each of the newly created subdatasets was treated to create intervals for each variable, that were intended to be as evenly distributed as possible, i.e., with an equilibrated number of entries for each of subgroups. So, a categorization process was made and it again was done in parallel for the overall data and the long- and short-term groups as a hypothetical strategy to reduce data variability.

Afterwards, a CDF was made for the variables pH and $\frac{\text{CEC}}{(\text{Ca}^{2+} + \text{Mg}^{2+})_{\text{SS}}}$, as a complement of the later performed FLSD tests to detect any significant differences between the new subdatasets. CDF stands for Cumulative Distribution Function (see Figure 4) and it describes the accumulated frequency against a data array in a plot. To do so, an array must be increasingly reordered to get assigned a probability in a normal distribution. Thus, the array can be used to do a non-linear fitting, according to Eq. 2:

$$f(x) = \frac{1}{2} \left(1 + \left(\text{erf} \left(\frac{\log K_d - \mu}{\sigma \sqrt{2}} \right) \right) \right) \quad \text{Eq.2. Fitting function to build a CDF.}$$

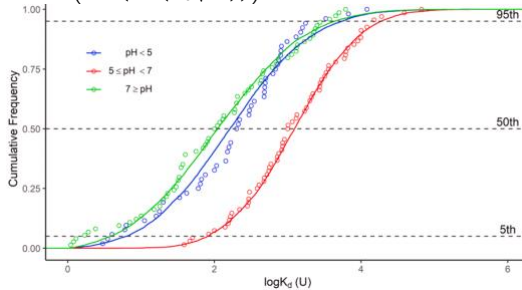


Figure 4. CDF of K_d (U) for pH groups. From Ramírez-Guinart *et al.* 2020.

The statistical parameters defined in Eq. 2 are a location parameter (μ), which is the central value of the distribution and is calculated as the arithmetic mean of the array; and the scale parameter of the normal distribution (σ), which express the data dispersion and corresponds to

the standard deviation of the array. Whereas the $\log K_d$ parameter is the variable used to iterate the array against the accumulated frequency. As Eq. 2 involves a complex fitting, the fitting package of *Matlab* had to be used to generate the plots, and to do so a short *Matlab* program for every CDF was written (see Annex 3 page 48 to view the program written to make Figure 7) (Ramírez-Guinart, 2020a).

From a CDF, a best estimate (PCTL50) can easily be deduced and a measure of the dispersion of data can be interpreted with just a glance. The more the experimental data deviates from its fitting, greater is the variability; while the proximity between fittings indicates how similar the groups are.

Afterwards, FLSD tests and descriptive parameters were performed on each subdataset for the most significant variables of the correlation analyses. In this case, the best estimate would be a geometric mean (GM) and data variability can be deduced from the geometric standard deviation (GSD). In addition, the descriptive analyses allowed for a comparison between the 5th and 95th percentiles, and the maximum and minimum ranges. All the best estimates and descriptive parameters that are to be provided in this project will be the *antilog* of the calculated value.

6. RESULTS AND DISCUSSION

The results of this project will be divided into three sections. Firstly, a summary for the bibliographic search results will be provided. Secondly, a sum up of the database shall also be provided with some descriptive statistics. And finally, the correlation results, the categorization analyses and best estimates proposal based on CDFs are going to be presented.

6.1. BIBLIOGRAPHIC SEARCH RESULTS

Table 2 provides an overall sum up of the results of the several queries done in the *Web of Science* database. After making several tests to check the availability of results in several different queries, it was decided to use the stated keywords for performing the searches. In addition, it is provided the date of the search and the classification of the provided results, regarding their contribution to this project.

Search	Date	Keywords	Results	Discarded	Included	CbD	LtoL	NF
1	07/03/21	"radium" AND "soil" AND "distribution"	224	79 %	4 %	13 %	2 %	2 %
2	13/03/21	"radium" AND "soil" AND "kd"	2				100 %	
3	12/04/21	"radium" AND "soil" AND "partition"	15	60 %	20 %	7 %	13 %	
Total			241	77 %	5 %	13 %	3 %	2 %
			241	186	12	30	8	4

CbD: Consulted but Discarded
 LtoL: Lead to other Literature
 NF: Not Found or not available

Table 2. Search parameters for *Web of Science* and classification of the results.

Although the keywords were quite specific, most of the sources provided in these queries were directly discarded due to not having anything to do with the distribution coefficient of radium. A 77% of the results were discarded; only a 3% of the results were used to access other documents; they consisted mostly of reviews, but also of a few previously published databases. Thus, just 5% of the results were suitable sources to be included into the present database.

Some of the sources that were classified as 'lead to other literature' were previous databases of radium Kds. And even when they were useful in obtaining new original sources, some of them were disappointing. Some of the sources they cited, like Sheppard *et al.* 2006, did not really provide the coefficients stated in their database. The only explanation that was found plausible is

that the authors of the reviews contacted to the researchers cited in their projects, in order to get raw data or some unpublished details. Yet, a similar contact attempt with the authors from Al-Masri *et al.* 2021 to get some unpublished data for this project was done, and it was unsuccessful.

On the other hand, a review was done from the original database entries. Originally, there were entries from 18 different sources, from which only 9 were kept in the present database. Such a reduction was made due to the deleted sources not fulfilling the inclusion criteria. Mainly, the erased sources consisted of published K_d (Ra) compilations, which did not contain any experimental data.

6.2. THE DATABASE

A digital copy of the overall updated database can be found at the link: https://ubarcelona-my.sharepoint.com/:x/g/personal/gsathe7_alumnes_ub_edu/EVD5kaeh_1ZDt3xcsFcUkr8Ba_dWjl00RPIFPn0XxegrqQQ?e=dVdlZS. Due to the extent of the database, a paper version could not be included in the annex; yet it has been made available for any consult. So, the purpose of this section is to provide some general descriptive information about the soils and experiment methodology of the included entries.

The final selection was made after several reviews, and it contains 185 entries from 29 sources. The disparity on the publishing formats and aims of the sources has made this database quite heterogeneous in its content. On the one hand, there are entries that provide an exhaustive soil characterization, whereas there are sources that barely provide any data but the coefficient and a vague description, e.g., Mishra *et al.*, 2012, and Krest and Harvey, 2003. This heterogeneity will manifest itself in the statistical treatment of the data, where it will limit the entries available for the analyses of each variable.

Firstly, a picture depicting the soil classification must be provided. On the 'sample tag' classification, the entries were found to belong to: sediments (33%), soils (57%), subsoils (9%), and tuffs (1%). This classification is not or should not be of much use to study the behavior of the K_d , because they do not imply inherent soil characteristics that can be quantified for any predictions.

Secondly, the soil classification overview can be found in Table 3. This classification is taken from Ramírez-Guinart *et al.*, 2020a and classifies soils according to their texture and the organic matter content as detailed in the previous section.

Classification	Entries	Entries [%]
Mineral	171	92 %
Sand	73	39 %
Loam	27	15 %
Clay	28	15 %
Unspecified	43	31 %
Organic	11	6 %
Unspecified	3	2 %

Table 3. Classification of the soils in the database according to their textural data and organic matter content.

Most of the distribution coefficients are from mineral soils. With just a 6% of the data, it can be seen that data from organic soils are quite scarce. On the other hand, the mineral soils, which can be classified as loamy, clayey, or sandy, are more evenly distributed with a higher proportion for sandy soils, whereas the mineral group without any textural information reported is also quite relevant (31%). The overall unspecified group is reserved for those sources which do not provide any textural characterization, neither qualitative nor quantitative, or organic matter content.

The isotope used in the sorption studies should not be a factor influencing K_d (Ra) values, although the radioisotope used in the experiments (mostly ^{226}Ra) is reported in the database.

Moreover, there is the experimental method used in obtaining the coefficient. As stated in the methodology, each entry had to be labelled in the variable "Experimental tag approach" as long-term, *in situ* or desorption; or as short-term, *in situ*, sorption, or desorption. So, Table 4 provides a description of the database, regarding the classification of the entries according to their experimental method. Most distribution coefficients were obtained with a short-term methodology, especially from sorption experiments, whereas almost a third were classified as long-term experiments. This designation was of great importance, because in the forthcoming statistical analyses the effect of the methodology was to be examined.

After taking perspective from the search and the review of the database, it can be stated that there is not much published data of radium distribution coefficients. A similar search for any other radionuclide, e.g., uranium, might produce a much bigger number of entries in the database. There is an important lack of available data.

Classification	Entries	Entries [%]
Long-term	58	31 %
<i>in situ</i>	19	10 %
desorption	39	21 %
Short-term	127	69 %
<i>in situ</i>	0	0 %
desorption	9	5 %
sorption	118	64 %

Table 4. Counting of the variable Experimental tag approach.

6.3. STATISTICAL ANALYSES

6.3.1. Evaluation of the effect of the experimental approach on K_d (Ra) variability

As a way of performing a statistical analysis, variability ought to be reduced to a minimal when possible. In order to do so, data was grouped according to the 'Experimental tag approach', as a means to improve any future data correlations. The reasoning for this segregation of data was that the expected K_d (Ra) is predicted to be greater for long-term experiments than for short-term ones, given that other radionuclides like Cs present such a behavior (Ramírez-Guinart *et al.*, 2020b). So, this section aims at elucidating the effect of the methodology to assess if such a grouping was appropriate; to do so, CDFs and best estimates were calculated and compared.

Subgroups	N	BE ^a	GSD ^b	PCTL5	PCTL95	Min.	Max.	FLSD ^c
Overall	185	977	12	13	29107	2	645654	A
Long-term	58	2042	12	5	21503	2	23988	A
Short-term	127	776	12	13	107647	3	645654	A

Units: K_d (L Kg⁻¹). PCTL: Percentile.

^a Best estimate (PCTL50).

^b Geometrical standard deviation

^c FLSD stands for Fisher's Least Significant Difference test. Different letters among the datasets for each variable indicate significant differences between the groups with a 95% confidence in *log* scale.

Table 5. Best estimate and group comparison between experimental approach groups.

An FLSD test was performed on the database to observe any potentially significant differences between the three groups: long- and short-term, and overall data. Table 5 shows the result of the test and the descriptive parameters of each group. The *Best Estimate* for the long-term group almost doubles the one for the short-term experiments. There are no significant

differences between them, although $PCTL50_{LT} > PCTL50_{ST}$. That does not mean that there are actually no differences, it just means that variability is too great to appreciate them.

One great factor to be considered after observing data from Table 5 is that some of the published data produce K_d (Ra) with great variability. Thus, confirming the initial expectations.

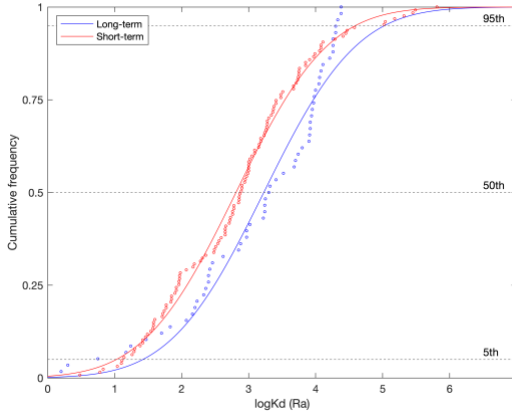


Figure 5. CDF for the long- and short-term groups.

Additionally, Figure 5 contains a plotted CDF to compare the long- and short-term groups. Both the CDF and Table 5 provide grounds to reinforce the hypothesis of long-term K_d (Ra) being greater than the short-term ones, despite not observing any statistically significant results. So, henceforth all the forthcoming analysis were performed for three main groups: overall, long-, and short-term.

6.3.2. Evaluation of the effect of soil properties on K_d (Ra) variability: correlation analysis

The goal of this section was to ascertain whether a soil factor can be used to group soils to reduce variability in the derived best estimate K_d (Ra). To do so, correlation analyses were performed between K_d (Ra) and selected soil variables from previous sorption studies, distinguishing between the overall, long-, and short-term groups. Hence, the studied variables were pH, organic matter (OM), carbonate concentration [CO_3^{2-}], Cation Exchange Capacity (CEC), and Ca and Mg status in the soil.

6.3.2.1. Univariate linear correlations

The univariate correlation analyses for OM (N=57) and CEC (N=68) proved to be not significant with the current number of entries in the database.

On the other hand, pH and carbonate concentration did show significant correlations. Table 6 displays the significant linear models and its correlation coefficient for both variables. Here, it can be seen that the strategy of using the experimental methodology as a stratifying element proved itself to be of great use. For instance, the carbonate correlation was only statistically significant in the short-term group. Neither the overall nor the long-term data showed any significant correlations. Nonetheless, it should not be inferred that the correlation does not equally affect long- and short-term experiments. In the partial database used to make these calculations, there were only 6 long-term entries that described a carbonate concentration. So, perhaps with more data, a correlation would have manifested. Yet, for now, the linear model for the short-term case proved an exceptionally high correlation ($R^2 = 0.56$). The whole partial dataset created to study the effects of carbonate contained 26 entries, though 20 of the were from short-term experiments.

Variable & Subgroup	Linear equation	N	R ²
pH			
Overall	$\log(K_d) = (0.17 \pm 0.11) \text{pH} + (1.87 \pm 0.73)$	138	0.07
Long-term*	$\log(K_d) = (0.37 \pm 0.13) \text{pH}$	43	0.45
[CO₃²⁻]			
Short-term	$\log(K_d) = (0.62 \pm 0.44) \log[\text{CO}_3^{2-}] + (1.99 \pm 0.57)$	20	0.56

* The slope of this equation was significant, though not the intercept.

Table 6. Significant univariate linear regression equations and their confidence intervals between K_d (Ra) and pH (overall and LT) and carbonate concentration (ST).

Likewise, pH did also show a significant correlation. Here, it can be observed that though the overall linear correlation is still significant, the inclusion of the short-term data into the array did increase the variability and the correlation coefficient was greatly reduced ($R^2 = 0.07$).

6.3.2.2. Multivariate linear correlations

Table 7 provides in detail all the multivariate analyses tested, leading to significant correlations.

Variable & Subgroup	Linear equation	N	R ²
[CO₃²⁻] + OM			
Overall	$\log(K_d) = (0.61 \pm 0.56) \log[\text{CO}_3^{2-}] + (1.59 \pm 1.21) \log(\text{OM})$	13	0.60
pH + OM			
Overall	$\log(K_d) = (0.26 \pm 0.12) \text{pH} - (0.52 \pm 0.39) \log(\text{OM}) + (1.63 \pm 0.81)$	69	0.26
Long-term*	$\log(K_d) = (0.41 \pm 0.35) \text{pH}$	12	0.51
Short-term	$\log(K_d) = (0.25 \pm 0.13) \text{pH} - (0.59 \pm 0.39) \log(\text{OM}) + (1.83 \pm 0.85)$	57	0.27
CEC/(Ca+Mg)_{ss}			
Overall	$\log(K_d) = (0.71 \pm 0.42) \log\left(\frac{\text{CEC}}{(\text{Ca}^{2+} + \text{Mg}^{2+})_{ss}}\right) + (1.90 \pm 0.48)$	24	0.33
Short-term	$\log(K_d) = (0.69 \pm 0.42) \log\left(\frac{\text{CEC}}{(\text{Ca}^{2+} + \text{Mg}^{2+})_{ss}}\right) + (1.96 \pm 0.49)$	23	0.33
(Ca_{exch}+Mg_{exch})/(Ca+Mg)_{ss}			
Short-term	$\log(K_d) = (0.64 \pm 0.31) \log\left(\frac{\text{Ca}_{exch}^{2+} + \text{Mg}_{exch}^{2+}}{(\text{Ca}^{2+} + \text{Mg}^{2+})_{ss}}\right) + (1.86 \pm 0.33)$	14	0.63
pH + CEC/(Ca+Mg)_{ss}			
Overall*	$\log(K_d) = (0.27 \pm 0.20) \text{pH} + (0.59 \pm 0.48) \log\left(\frac{\text{CEC}}{(\text{Ca}^{2+} + \text{Mg}^{2+})_{ss}}\right)$	25	0.43
Short-term*	$\log(K_d) = (0.64 \pm 0.41) \log\left(\frac{\text{CEC}}{(\text{Ca}^{2+} + \text{Mg}^{2+})_{ss}}\right)$	23	0.44

* Some coefficients of the regression model were not significant.

Table 7. Significant multivariate linear regression models.

The multivariate approaches led to a great improvement in the correlation coefficient, with respect to the univariate ones. Multivariate correlations were initially defined based on those variables showing significant univariate correlations, such as [CO₃²⁻] + OM and pH + OM. The former combination did also produce an especially increased correlation. Therefore, the organic matter content should not be disregarded when studying the distribution coefficient in future studies, although it is required that data on K_d (Ra) from soils with a high OM content become available.

The regression analysis for the relation between the CEC and calcium and magnesium concentrations (CEC/(Ca+Mg)_{ss}) was quite relevant. As according to literature, an estimate of the K_d (Sr) can be obtained with this ratio and it was decided to test if it would produce the same results with Ra (Gil-García *et al.*, 2011). However, as a result of the previous analysis, another

ratio was studied: the ratio between exchangeable and solved calcium and magnesium ($(Ca_{\text{exch}}+Mg_{\text{exch}})/(Ca+Mg)_{\text{ss}}$). Thus, a great increase in the coefficient correlation was observed, which confirms that Ra and Sr are chemical analogues and this ratio can be used to infer K_d (Ra). The correlation observed is so good, that it would allow for the construction of K_d (Ra) predicting models. But the problem of this ratio is that published data regarding the exchangeable cations in a soil is not as common as the CEC.

Additionally, $(Ca_{\text{exch}}+Mg_{\text{exch}})/(Ca+Mg)_{\text{ss}}$ and pH were combinedly studied and did not provide a much better correlation when combined.

6.3.3. Proposal of K_d (Ra) best estimates and description and reduction of K_d (Ra) variability for selected soil groups

In this section it was examined if grouping soils based on soil properties proved to affect K_d (Ra) values in order to derive K_d (Ra) best estimates with a lower related variability. Hence, partial datasets were created by grouping soils according to ranges of values of variables or combination of variables that showed significant correlations. Ranges were created to ensure partial datasets had a sufficient number of entries.

6.3.3.1. K_d (Ra) best estimates and variability from soils grouped based on pH

pH Subgroup	Intervals	N	BE ^a	GSD ^b	PCTL5	PCTL95	Min.	Max.	FLSD ^c
Overall	pH < 6	45	600	9	24	10872	3	28184	A
	6 ≤ pH < 8	57	800	8	13	26146	9	38019	A
	pH ≥ 8	36	4759	7	31	24033	28	23988	B
Long-term	pH < 6	10	1269	6	50	8318	50	8318	A
	6 ≤ pH < 8	12	1323	4	212	11220	214	11220	A
	pH ≥ 8	21	11410	2	1041	24182	933	23988	B
Short-term	pH < 6	35	540	10	14	13646	3	28184	A
	6 ≤ pH < 8	45	800	9	10	28314	9	38019	A
	pH ≥ 8	15	1132	7	28	4677	28	4677	A

Units: K_d (L kg⁻¹), PCTL: Percentile.

^a Best estimate: PCTL50.

^b Geometric standard deviation

^c FLSD stands for Fisher's Least Significant Difference test. Different letters among the datasets for each variable indicate significant differences between the groups with a 95% confidence in log scale.

Table 8. Proposed best estimates for soils at different pH ranges.

Table 8 involved grouping soils at different pH ranges: $\text{pH} < 6$, $6 \leq \text{pH} < 8$, and $\text{pH} \geq 8$. Although they were subjectively proposed, the goal was to produce ranges with as evenly distributed soils as possible. All the proposed best estimates of a K_d (Ra) for each soil group based on pH. The best estimate was not only proposed for the overall data but also for the LT and ST subdatasets.

Firstly, the influence of pH was confirmed: a higher pH related to a higher K_d (Ra). Secondly, above pH 8 the influence of pH is most notable, as surpassing this threshold might even produce an increase in the estimated K_d (Ra) of one order of magnitude, which statistically differs from the estimated K_d (Ra) of the other groups. Thirdly, at $\text{pH} \geq 8$ further FLSD tests were performed that showed that above this same threshold there are significant differences between the long- and short-term groups, and between the overall data. Nonetheless, if only data from Table 8 is considered, the K_d (Ra) best estimates for the long-term categories are higher than the short-term ones.

Finally, using soil grouping was proven to be useful in reducing the variability at proposing K_d (Ra) best estimates. If Table 8 is compared with Table 5 (which included the whole database), it can be inferred a much lesser variability for Table 8: the GSD is reduced, and the difference between the 5th-95th percentiles and the maximum and minimum ranges are greatly reduced.

Figure 6 provides, as an illustrating example the plotted CDFs of the overall data for the three pH soil groups.

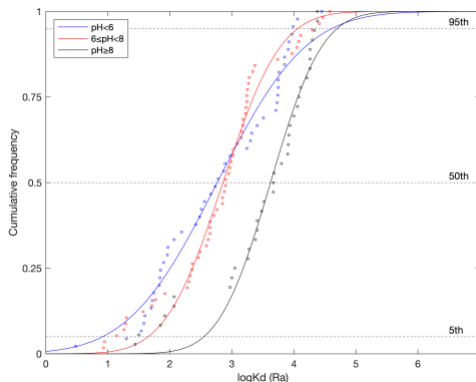


Figure 6. CDF of the pH intervals of the overall data.

Figure 6 confirms the observations made for Table 8, but also provides a better picture at presenting the variability of the data. In fact, the variability of the included K_d (Ra) has decreased to 3 orders of magnitude and has provided lesser variability to the best estimates for each soil group, even some of them being significantly different from the others, i.e., for $\text{pH} \geq 8$. The $\text{pH} \geq 8$ soil group is depicted as being significantly different from the other ranges, which overlap with each other in the 5th-95th percentile range.

As a sum up, the pH can be used to predict K_d (Ra) best estimates with a lower variability than the ones obtained from the overall database, some of which are even statistically different between each other.

6.3.3.2. K_d (Ra) best estimates and variability from soils grouped based on $[\text{CO}_3^{2-}]$

Table 9 involved grouping soils at different carbonate concentration (mg L^{-1}) ranges: $[\text{CO}_3^{2-}] < 10$, $10 \leq [\text{CO}_3^{2-}] < 50$, and $[\text{CO}_3^{2-}] \geq 50$.

$[\text{CO}_3^{2-}]$ - Subgroup	Intervals	N	BE ^a	GSD ^b	PCTL5	PCTL95	Min.	Max.	FLSD ^c
Overall	$[\text{CO}_3^{2-}] < 10$	13	150	6	2	1778	2	1778	A
	$10 \leq [\text{CO}_3^{2-}] < 50$	5	862	3	205	3020	204	3020	AB
	$[\text{CO}_3^{2-}] \geq 50$	8	946	8	50	12023	50	12023	B
Long-term	$[\text{CO}_3^{2-}] < 10$	3	60	23	2	933	2	933	A
	$10 \leq [\text{CO}_3^{2-}] < 50$	0	-	-	-	-	-	-	-
	$[\text{CO}_3^{2-}] \geq 50$	3	304	18	50	8318	50	8318	A
Short-term	$[\text{CO}_3^{2-}] < 10$	10	197	4	38	1778	38	1778	A
	$10 \leq [\text{CO}_3^{2-}] < 50$	5	862	3	205	3020	204	3020	B
	$[\text{CO}_3^{2-}] \geq 50$	5	1869	3	543	12023	537	12023	B

Units: K_d (L Kg^{-1}), $[\text{CO}_3^{2-}]$ (mg L^{-1}). PCTL: Percentile.

^a Best estimate: GM.

^b Geometric standard deviation

^c FLSD stands for Fisher's Least Significant Difference test. Different letters among the datasets for each variable indicate significant differences between the groups with a 95% confidence in log scale.

Table 9. $[\text{CO}_3^{2-}]$ best estimates for different intervals.

Table 9 shows all the proposed best estimates of a K_d (Ra) for each soil group base on $[\text{CO}_3^{2-}]$. The best estimate was not only proposed for the overall data but also for the LT and ST subdatasets.

Firstly, the influence of $[\text{CO}_3^{2-}]$ was confirmed: a higher $[\text{CO}_3^{2-}]$ related to a higher K_d (Ra). Secondly, at $[\text{CO}_3^{2-}] < 10$ the influence of $[\text{CO}_3^{2-}]$ is most notable, as surpassing this threshold might even produce an increase in the estimated K_d (Ra) of one order of magnitude. This interval statistically differed from the other ranges.

According to Table 9, data variability has also been reduced for this subdataset. In this case, K_d (Ra) variability was reduced to just three orders of magnitude and the 5th-95th percentile range is the same as the maximum-minimum range. This fact might be a consequence of studying categories with a reduced number of samples.

As a sum up, the variable carbonate concentration could be used to obtain K_d (Ra) best estimates with lower variability, some of which would be even statistically different between each other.

6.3.3.3. K_d (Ra) best estimates and variability from soils grouped based on $\text{CEC}/(\text{Ca}+\text{Mg})_{\text{ss}}$

Table 10 involved grouping soils at different $\text{CEC}/(\text{Ca}+\text{Mg})_{\text{ss}}$ ranges: $\text{CEC}/(\text{Ca}+\text{Mg})_{\text{ss}} < 5$, $5 \leq \text{CEC}/(\text{Ca}+\text{Mg})_{\text{ss}} < 50$, and $\text{CEC}/(\text{Ca}+\text{Mg})_{\text{ss}} \geq 50$.

CEC/(Ca+Mg)_{ss} - Subgroup	Intervals	N	BE^a	GSD^b	PCTL5	PCTL95	Min.	Max.	FLSD^c
Overall	X < 5	8	88	4	18	1072	18	1072	A
	5 ≤ X < 50	11	391	4	50	2291	50	2291	AB
	X ≥ 50	5	974	4	520	12023	525	12023	B
Short-term	X < 5	8	88	4	18	1072	18	1072	A
	5 ≤ X < 50	10	773	4	91	2291	91	2291	B
	X ≥ 50	5	974	4	520	12023	525	12023	B

X: $\text{CEC}/(\text{Ca}+\text{Mg})_{\text{ss}}$. Units: K_d (L Kg^{-1}), $\text{CEC}/(\text{Ca}+\text{Mg})_{\text{ss}}$ (L kg^{-1}). PCTL: Percentile.

^a Best estimate: PCTL50.

^b Geometric standard deviation

^c FLSD stands for Fisher's Least Significant Difference test. Different letters among the datasets for each variable indicate significant differences between the groups with a 95% confidence in log scale.

Table 10. $\text{CEC}/(\text{Ca}+\text{Mg})_{\text{ss}}$ best estimates for different intervals.

Table 10 shows all the proposed best estimates of a K_d (Ra) for each soil group based on $\text{CEC}/(\text{Ca}+\text{Mg})_{\text{ss}}$. The best estimates were again not only proposed for the overall data but also for the LT and ST subdatasets.

Firstly, the influence of $\text{CEC}/(\text{Ca}+\text{Mg})_{\text{ss}}$ was confirmed: a higher ratio related to a higher K_d (Ra). Secondly, significant differences were observed between the K_d (Ra) best estimates from

the extreme groups, although they did not significantly differ from the central soil group. The observed differences were in this case of one order of magnitude.

According to Table 10, data variability has also been reduced for this subdataset. In this case, K_d (Ra) variability was reduced to just two orders of magnitude and the 5th-95th percentile range was the same as the maximum-minimum range. This variability was considerably low when compared with the six orders of magnitude observed in Table 5 from the overall dataset.

Figure 7 provides, as an illustrating example the plotted CDFs of the overall data for the three $CEC/(Ca+Mg)_{ss}$ soil groups.

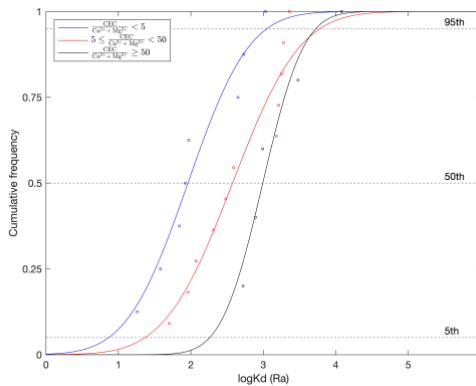


Figure 7. CDF of the $CEC/(Ca+Mg)_{ss}$ ratio for the overall data.

Figure 7 displays considerably good fittings, considering the small sample. Here, the results of the FLSD test can be better interpreted: significant differences exist between the highest and lowest soil ranges, but they do not manifest when they are respectively compared against the central range. That can be observed with the slight overlapping of two of the fittings.

As a sum up, the variable $CEC/(Ca+Mg)_{ss}$ could be used to propose K_d (Ra) best estimates with lower variability, some of which are even statistically different between each other.

7. CONCLUSIONS

The main goal of this project was achieved: a K_d (Ra) database was expanded to study which soil parameters can help reduce the variability in predicting a K_d (Ra) best estimate for a soil.

The database used in this project is an expanded version of a previous database, after adding a sizable number of entries and adjusting the parameters and variables stored within it. Additionally, the previously included entries were reviewed and some of them were discarded due to not fulfilling the main inclusion criterion: the original source had to provide an experimental K_d , not one from a data compilation. Despite the final number of entries being greatly increased, it is not as big and diverse as the ones existing for other radionuclides. Thus, it can be concluded after a thorough bibliographical research, that there is an important lack of published literature about K_d (Ra) in comparison to other radionuclides.

The number of factors involved in radium sorption are quite great and not all of them were studied in this project. A previous bibliographic research showed that parameters like the textural data or the soil mineralogy would be expected to play an important role in any sorption process, although not as much information for radium is available as there is for other radionuclides. Some of the parameters studied in this project were made under the assumption that radium would be expected to present a similar behavior as other alkaline earth elements, like strontium.

The univariant linear correlation analyses depicted considerably good correlations for pH and carbonate concentration in the soil solution with the K_d (Ra). No other univariant correlations were observed, although that may change in future studies when more data is being considered. Yet, the multivariant analyses did produce some exceptionally good correlations. Most interestingly, they did show that both the $CEC/(Ca+Mg)_{ss}$ and $(Ca_{esch}+Mg_{exch})/(Ca+Mg)_{ss}$ can quite significantly explain the K_d (Ra) variability, which is consistent with previous findings for other earth alkaline metals.

Other multivariant significant linear correlations were also described, being the combination of carbonate concentration and OM the one that produced the best correlation. Thus, highlighting a major role to the organic matter, previously undetected in the univariant correlation analyses.

On the other hand, using the experimental approach (long- or short-term) to further group data served its purpose: it allowed to propose smaller soil groups to study different parameters. This strategy reduced variability enough to observe some correlations that would otherwise have been missed.

Finally, K_d (Ra) best estimates were proposed for pH, carbonate concentration, and $CEC/(Ca+Mg)_{ss}$. Partial datasets were again used to group soils at different intervals of a variable or parameter, so as to produce a significant decrease in overall variability. Some of this categorized groups were different enough from each other to show significant differences, e.g., the case of $pH \geq 8$. Hence, it is important to remark that obtaining these best estimates at a reduced variability than those of the overall dataset is definitely good for building risk assessment models.

As a sum up, all the established goals were fulfilled within the scope of a prospective analysis of K_d (Ra). However, before proceeding to any further data analysis, additional K_d (Ra) entries should be included into the database inasmuch some soil groups, like organic soils, are underrepresented. So, all correlations (or the lack of them) observed within this project are slightly susceptible to change when more data is included.

11. REFERENCES AND NOTES

- Al-Hamameh, IF; Alkhomashi, N; Almasoud, FI. Study on the radioactivity and soil-to-plant transfer factor of ^{226}Ra , ^{234}U and ^{238}U radionuclides in irrigated farms from the northwestern Saudi Arabia. *J Environ Radioact*. **2016**, *160*, 1–7.
- Bordelet, G; Beaucaire, C; Phrommavanh, V; Descostes, M. Sorption Properties of Peat for U(VI) and ^{226}Ra in U Mining Areas. *Procedia Earth Planet. Sci.* **2013**, *7*, 85–88.
- Brady, NC. The Nature and properties of soils, 9th edition, McMillan Co., New York, USA (1984).
- Gil-García, CJ; Rigol, A; Vidal, M. The use of hard- and soft-modelling to predict radiostromtium solid-liquid distribution coefficients in soils. *Chemosphere*. **2011**, *85*, 1400-1405.
- Grandia, F; Merino, J; Bruno, J. Assessment of the Radium-Barium co-precipitation and Its Potential Influence on the Solubility of Ra in the Near-Field; TR-08-07. Svensk Kärnbränslehantering AB: Stockholm, 2008.
- Handbook of Parameter Values for the Prediction of Radionuclide Transfer in Terrestrial and Freshwater Environments; Vol. 472, International Atomic Energy Agency (IAEA): Vienna, 2010.
- Krest, JM; Harvey, JW. Using natural distributions of short-lived radium isotopes to quantify groundwater discharge and recharge. *Limnol. Oceanogr.* **2003**, *48*, 290–298.
- Krupka, KM; Kaplan, DI; Gene, W; Serne, RJ; Mattigod, SV. Understanding Variation in Partition Coefficient, Kd, Values. Volume I: the Kd Model, Methods of Measurement, and Application of Chemical Reaction Codes; EPA 402-R-99-004A. U. S. Environmental Protection Agency: Washington, 1999.
- Kumar, A; Rout, S; Pulhani, V; Kumar, AV. A review on distribution coefficient (Kd) of some selected radionuclides in soil/sediment over the last three decades. *J Radioanal Nucl Chem.* **2020**, *323*, 13–26.
- Nathwani, JS; Phillips, CR. Adsorption of ^{226}Ra by Soils (I). *Chemosphere*. **1979a**, *5*, 285–291.
- Nathwani, JS; Phillips, CR. Adsorption of ^{226}Ra by Soils in the Presence of Ca^{2+} Ions. Specific Adsorption (II). *Chemosphere*. **1979b**, *5*, 293–299.
- NIH. Radon [Online]. <https://www.niehs.nih.gov/health/topics/agents/radon> (accessed Jun 9, 2021).
- Ramírez-Guinart, O; Kaplan, D; Rigol, A; Vidal, M. Deriving Probabilistic Soil Distribution Coefficients (Kd). Part 1: General Approach to Decreasing and Describing Variability and Example Using Uranium Kd Values. *J Environ Radioact.* **2020a**, *222*, 106362.
- Ramírez-Guinart, O; Kaplan, D; Rigol, A; Vidal, M. Deriving Probabilistic Soil Distribution Coefficients (Kd). Part 2: Reducing Caesium Kd Uncertainty by Accounting for Experimental Approach and Soil Properties. *J Environ Radioact.* **2020b**, *223–224*, 106407.
- Ramírez-Guinart, O; Kaplan, D; Rigol, A; Vidal, M. Deriving Probabilistic Soil Distribution Coefficients (Kd). Part 3: Reducing Variability of Americium Kd Best Estimates Using Soil Properties and Chemical and Geological Material Analogues. *J. Environ. Radioact.* **2020c**, *223–224*, 106378.
- Sheppard, MI; Beals, DI; Thibault, DH; O'Connor, P. Soil Nuclide Distribution Coefficients and Their Statistical Distributions; AECL-8364. Atomic Energy of Canada Limited, AECL (Report). Pinawa; 1984.
- Simon, SL; Ibrahim, SA, 1990. Biological uptake of radium by terrestrial plants. In: The Environmental Behaviour of Radium, vol. 1, IAEA. Technical Report Series, No. 310., pp. 545–599.
- Subdiaga, E; Orsetti, S; Haderlein, SB. Effects of Sorption on Redox Properties of Natural Organic Matter. *Environ. Sci. Technol.* **2019**, *53* (24), 14319–14328.
- Vandenhove, H; Van Hees, M. Predicting Radium Availability and Uptake from Soil Properties. *Chemosphere*. **2007**, *69* (4), 664–674.
- Yolcubal, I.; Brusseau, M. L.; Artiola, J. F.; Wierenga, P. J.; Wilson, LG. Environmental Physical Properties and Processes. In *Environmental Monitoring and Characterization*; 2004; pp 207–239.

12. ACRONYMS

BE	Best Estimate
CDF	Cumulative Density Function
CEC	Cationic Exchange Capacity
DIC	Dissolved Inorganic Carbon
DOC	Dissolved Organic Carbon
FLSD	Fisher's Least Significant Difference test
GM	Geometric Mean
GSD	Geometric Standard Deviation
IAEA	International Atomic Energy Agency
LT	Long-term experiment
Max	Maximum
Min	Minimum
N	Sample population
NIH	National Institutes of Health
NOR	Naturally Occurring Radionuclide
NORM	Naturally Occurring Radionuclide Materials
K_d	Distribution Coefficient
OM	Organic Matter
DOC	Dissolved Organic Carbon
PCTL	Percentile
ST	Short-term experiment
TOC	Total Organic Carbon

APPENDICES

APPENDIX 1: DATABASE DESCRIPTION

	Variable	Specifications
Bibliographical data	Reference type	The kind of source, i.e., if it was either a published article or a technical report.
	Journal	The name of the publishing journal.
	Language	The language of the source.
	Authors	The authors of the publication.
	Volume	The volume number in the publishing journal.
	Pages	The pages in the publishing journal.
	Year	The year when it was published.
	Title	The title of the source.
General information	Element	The element under study.
	Isotope	The isotope with which the experiment was carried out.
	Climate zone	The climate zone wherein the sample was obtained.
	Country	The country wherein the sample was obtained.
	Sample tag	Basic classification as either 'subsoil', 'soil', 'sediment', 'glyttja', 'till', or 'tuff'; as it was stated in the inclusion criteria. New addition to the database.
	Soil code	The name or number assigned to the sample in the original source.
	Soil description	New name for the variable <i>Soil classification</i> . Brief description of the sample as given in the source.
Soil parameters	Handbook group	It specifies the soil type: 'organic' or 'mineral'; where mineral soils can be 'clays', 'sands', or 'loams.
	Soil pH	pH of the soil in its original or natural state.
	pH contact solution	Renaming of the original pH column. It states the actual pH under which the experiment was conducted. If the pH was not modified, the soil pH was the one considered as effective pH.
	OM (%)	Organic Matter content in the solid phase
	DOC (mg L ⁻¹)	Dissolved Organic Carbon.
	TOC (%)	Total Organic Carbon in the solid phase.
	Inorganic Carbon (%)	Inorganic Carbon content in the soil phase.
	DIC (mg L ⁻¹)	Dissolved Inorganic Carbon.
	CEC (cmol _c kg ⁻¹)	Cationic Exchange Capacity
	Sand (%)	Textural proportion of sand.
	Clay (%)	Textural proportion of clay.

Variable	Specifications
Ionic strength	Ionic strength of the contact solution or the liquid phase in a soil.
Exchangeable Ca ²⁺ (cmol _c kg ⁻¹)	Exchangeable calcium content in the solid phase
Exchangeable Mg ²⁺ (cmol _c Kg ⁻¹)	Exchangeable magnesium content in the solid phase
Exchangeable Na ⁺ (cmol _c kg ⁻¹)	Exchangeable Na ⁺ in the solid phase.
Exchangeable K ⁺ (cmol _c kg ⁻¹)	Exchangeable K ⁺ in the solid phase.
Exchangeable Al ³⁺ (cmol _c kg ⁻¹)	Exchangeable Al ³⁺ in the solid phase.
Exchangeable Mn (cmol _c kg ⁻¹)	Exchangeable Mn in the solid phase.
Exchangeable Fe (cmol _c kg ⁻¹)	Exchangeable Fe in the solid phase.
Na ⁺ contact solution (mEq L ⁻¹)	Actual Na ⁺ concentration in the contact solution in the sorption/desorption experiment. In a short-term experiment it must include the concentration in the contact solution, though in a long term one it is the concentration in the soil pore water or natural soil solution.
K ⁺ contact solution (mEq L ⁻¹)	Actual K ⁺ concentration in the contact solution in the sorption/desorption experiment. In a short-term experiment it must include the concentration in the contact solution, though in a long term one it is the concentration in the soil pore water or natural soil solution.
Ca ²⁺ contact solution (mEq L ⁻¹)	Actual Ca ²⁺ concentration in the contact solution in the sorption/desorption experiment. In a short-term experiment it must include the concentration in the contact solution, though in a long term one it is the concentration in the soil pore water or natural soil solution.
Mg ²⁺ contact solution (mEq L ⁻¹)	Actual Mg ⁺ concentration in the contact solution in the sorption/desorption experiment. In a short-term experiment it must include the concentration in the contact solution, though in a long term one it is the concentration in the soil pore water or natural soil solution.
Na ⁺ (mg L ⁻¹)	Na ⁺ concentration in the pore water or liquid phase. It was provided separately only when the concentration of this element in the contact solution was provided.
K ⁺ (mg L ⁻¹)	K ⁺ concentration in the pore water or liquid phase. It was provided separately only when the concentration of this element in the contact solution was provided.

Variable	Specifications
Ca ²⁺ (mg L ⁻¹)	Ca ²⁺ concentration in the pore water or liquid phase. It was provided separately only when the concentration of this element in the contact solution was provided.
Mg ²⁺ (mg L ⁻¹)	Mg ²⁺ concentration in the pore water or liquid phase. It was provided separately only when the concentration of this element in the contact solution was provided.
SO ₄ ²⁻ (mg L ⁻¹)	Sulphate concentration in solution.
CO ₃ ²⁻ (mg L ⁻¹)	Carbonate concentration in solution.
Oxalate extrac. Fe (mg L ⁻¹)	Iron concentration of the extractable fraction.
Fe (mg L ⁻¹)	Iron concentration in solution.
Total Fe (mg kg ⁻¹)	Total iron content in the solid phase
Total Na (g kg ⁻¹)	Total sodium in the solid phase.
Total K (g kg ⁻¹)	Total potassium in the solid phase.
Total Ca (g kg ⁻¹)	Total calcium in the solid phase.
Total Mg (g kg ⁻¹)	Total magnesium in the solid phase.
Total Al (g kg ⁻¹)	Total Al in the solid phase.
Total SO ₄ ²⁻ (mg kg ⁻¹)	Total sulphate in the solid phase.
CaCO ₃ (g kg ⁻¹)	Total calcium carbonate content in the solid phase
Surface area (m ² g ⁻¹)	Superficial area of the soil.
Sorption data	Kd (L kg ⁻¹) Sorption coefficient. The previous database made a distinction between sorption and desorption Kds and used different columns. Yet, this distinction was deleted due to the addition of the variable 'Experimental tag approach'.
Method	226Ra in solution Radium activity at an equilibrium phase in the liquid phase of a soil.
Contamination type	Specify if the experiment involved addition of radium or not, like an anthropogenic element.
Experimental tag approach	Tag to classify between long- and short-term methods. The tags used are ones used in Table 1.
Contamination level	Amount of radium in the sample, or the added amount. The units of the original source were the one used in each case.
Remarks sorption	Methodological specifications under which the sorption experiment was carried out, i.e., S/L ratio, equilibration time, etc.
Observations	Personal notes regarding an entry or the source.

Table 11. Variables included in the K_d (Ra) database.

APPENDIX 2: LIST OF PUBLICATIONS IN THE DATABASE

- Agüero, A. Performance Assessment Model Development and Parameter Acquisition for Analysis of the Transport of Natural Radionuclides in a Mediterranean Watershed. *Sci. Total Environ.* **2005**, *348*, 32–50.
- Al-Masri, MS; Amin, Y; Khalily, H; Al-Masri, W; Al-Khateeb, Y. Assessment of Natural Radionuclide Transfer Factors and Partition Coefficients in Some Syrian Soils. *J. Environ. Radioact.* **2021**, 229–230, 106538.
- Blanco Rodríguez, P; Lozano, JC; Vera Tomé, F; Prieto, C; Medeiros, A. Influence of Soil Conditions on the Distribution Coefficients of ²²⁶Ra in Natural Soils. *Chemosphere.* **2018**, *205*, 188–193.
- Bordelet, G; Beaucaire, C; Phrommavanh, V; Descostes, M. Sorption Properties of Peat for U(VI) and ²²⁶Ra in U Mining Areas. *Procedia Earth Planet. Sci.* **2013**, *7*, 85–88.
- Cochran, JK; Krishnaswami, S. Radium, Thorium, Uranium, and Pb210 in Deep-Sea Sediments and Sediment Pore Waters from the North Equatorial Pacific. *Am. J. Sci.* **1980**, *280*, 849–889.
- Colbert, SL; Hammond, DE. Shoreline and Seafloor Fluxes of Water and Short-Lived Ra Isotopes to Surface Water of San Pedro Bay, CA. *Mar. Chem.* **2008**, *108*, 1–17.
- Dalvi, AA; Kumar, SD; Reddy, AVR. A Site-Specific Study on the Measurement of Sorption Coefficients for Radionuclides. *Int. J. Environ. Sci. Technol.* **2014**, *11*, 617–622.
- Gillham, RW; Sharma, HD; Reddy, MR. Barium and Radium Migration in Unconsolidated Canadian Geological Materials; INFO-0048. Atomic Energy Control Board: Ottawa, 1981.
- Gonneea, ME.; Morris, P.J.; Dulajova, H; Charette, MA. New Perspectives on Radium Behavior within a Subterranean Estuary. *Mar. Chem.* **2008**, *109*, 250–267.
- Houmani, ZMM; Majid, AA; Radiman, S; Ahmad, Z. Influence of the Physico-Chemical Properties of Selangor Soil Series on the Distribution Coefficient (K_d-Value) of ²²⁶Ra. *J. Radioanal. Nucl. Chem.* **2010**, *285*, 271–277.
- Krest, JM; Harvey, JW. Using natural distributions of short-lived radium isotopes to quantify groundwater discharge and recharge. *Limnol Oceanogr.* **2003**, *48*, 290–298.
- Kumar, S; Kar, AS; Rawat, N; Maity, S; Mishra, S; Pandit, GG; Tomar, BS. Distribution Coefficients of Radionuclides around Uranium Mining Area and Effect of Different Analytical Parameters on Their Determination. *J. Radioanal. Nucl. Chem.* **2015**, *304*, 727–733.
- Lauria, DC; Almeida, RMR; Sracek, O. Behavior of Radium, Thorium and Uranium in Groundwater near the Buena Lagoon in the Coastal Zone of the State of Rio de Janeiro, Brazil. *Environ. Geol.* **2004**, *47*, 11–19.
- Legoux, Y; Blain, G; Guillaumont, R; Ouzounian, G; Brillard, L; Hussonnois, M. KdMeasurements of Activation, Fission and Heavy Elements in Water/Solid Phase Systems. *Radiochim. Acta.* **1992**, *58/59*, 211–218.
- Maity, S; Sahu, SK; Pandit, GG. Estimation of Distribution Coefficient of Radium around a Uranium Mining Site. *Radioprotection.* **2015**, *50*, 129–134.
- Meier, H; Zimmerhackl, E; Zeitler, G; Menge, P. Parameter Studies of Radionuclide Sorption in Site-Specific Sediment/Groundwater Systems. *Radiochim. Acta.* **2015**, 66–67, 277–284.
- Mishra S, Maity S, Pandit GG. Estimation of distribution coefficient of natural radionuclides in soil around uranium mines and its effect with ionic strength of water. *Radiat Prot Dosimetry.* **2012**, *152*, 229–233.
- Nathwani, JS; Phillips, CR. Adsorption of ²²⁶Ra by Soils (I). *Chemosphere.* **1979a**, *5*, 285–291.
- Nathwani, JS; Phillips, CR. Adsorption of ²²⁶Ra by Soils in the Presence of Ca²⁺ Ions. Specific Adsorption (II). *Chemosphere.* **1979b**, *5*, 293–299.
- Powell, BA; Miller, T; Kaplan, DI. On the Influence of Ionic Strength on Radium and Strontium Sorption to Sandy Loam Soils. *J. South Carolina Acad. Sci.* **2015**, *13*, 13–18.
- Rama; Moore, WS. Using the Radium Quartet for Evaluating Groundwater Input and Water Exchange in Salt Marshes. *Geochim. Cosmochim. Acta.* **1996**, *60*, 4645–4652.

- Sakamoto, Y; Takebe, S; Ogawa, H; Muraoka, S; Ishii, T; Inagawa, S; Gunji, Y. Sorption Behavior of Radium and Actinium on Soils. *Mater. Res. Soc. Symp. Proc.* **2001**, 663, 1207–1214.
- Salbu, B; Burkittbaev, M; Strømman, G; Shishkov, I; Kayukov, P; Uralbekov, B; Rosseland, BO. Environmental Impact Assessment of Radionuclides and Trace Elements at the Kurday U Mining Site, Kazakhstan. *J. Environ. Radioact.* **2013**, 123, 14–27.
- Sanchez, F; Rodriguez-Alvarez, MJ. Effect of PH, Temperature, Conductivity and Sediment Size on Thorium and Radium Activities along Jucar River (Spain). *J. Radioanal. Nucl. Chem.* **1999**, 24, 671–681.
- Sheppard, S; Sohlenius, G; Omberg, LG; Borgiel, M; Grolander, S; Nordén, S. Solid/Liquid Partition Coefficients (Kd) and Plant/Soil Concentration Ratios (CR) for Selected Soils, Tills and Sediments at Forsmark; R-11-24. Svensk Kärnbränslehantering AB: Stockholm, 2011.
- Sun, Y; Torgersen, T. Adsorption-Desorption Reactions and Bioturbation Transport of ²²⁴Ra in Marine Sediments: A One-Dimensional Model with Applications. *Mar. Chem.* **2001**, 74, 227–243.
- Thiry, Y; Van Hees, M. Evolution of PH, Organic Matter and ²²⁶Radium/Calcium Partitioning in U-Mining Debris Following Revegetation with Pine Trees. *Sci. Total Environ.* **2008**, 393, 111–117.
- Tripler, E; Haquin, G; Koch, J; Yehuda, Z; Shani, U. Sustainable Agricultural Use of Natural Water Sources Containing Elevated Radium Activity. *Chemosphere.* **2014**, 104, 205–211.
- Vandenhove, H; Van Hees, M. Predicting Radium Availability and Uptake from Soil Properties. *Chemosphere.* **2007**, 69, 664–674.
- Willett, IR; Bond, WJ. Sorption of Manganese, Uranium, and Radium by Highly Weathered Soils. *J. Environ. Qual.* **1995**, 24, 834–845.

APPENDIX 3: CDF MATLAB PROGRAM. EXAMPLE FOR CEC/(CA+MG)_{SS}

```
%*****
% Program: CEC_CaMg.m
% Description: CEC/Ca+Mg categories comparison
%*****
% Data and basic calculations for first Kd group
logKd = sort(load('CEC_CaMginf5.txt')); %paste data into this file
logKd_mean = mean(logKd);
logKd_sd = std(logKd);
logKd_freq = 1/length(logKd);
distr = normcdf(logKd,logKd_mean,logKd_sd);
logKd_freq_acum = [logKd_freq:logKd_freq:1]';
%-----
% Fitting: less than 5
[xData,yData] = prepareCurveData(logKd,logKd_freq_acum);
% Set up fitype and options.
ft = fitype( '1/2*(1+(erf((logKd-prom)./(desv*sqrt(2))))))', 'independent', 'logKd', 'dependent',
'logKd_freq_acum' );
opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
opts.Display = 'Off';
opts.StartPoint = [0.8002804688888 0.141886338627215];
% Fit model to data.
[fitresult, ~] = fit( xData, yData, ft, opts );
% The following expressions are used to extrapolate (change max X value)
Xfit = [0:0.05:6];
Yfit = fitresult(Xfit);
% Plot fit with data.
figure( 'Name', 'untitled fit 1' );
%h = plot( fitresult );
h = plot( Xfit, Yfit, 'b' );
hold on
scatter( xData, yData, 5, 'b' );
%*****
%Data and basic calculations for second Kd group
logKd = sort(load('CEC_CaMg5to50.txt')); %paste data into this file
logKd_mean = mean(logKd);
logKd_sd = std(logKd);
logKd_freq = 1/length(logKd);
distr = normcdf(logKd,logKd_mean,logKd_sd);
logKd_freq_acum = [logKd_freq:logKd_freq:1]';
%-----
% Fitting between 5 and 50
[xData,yData] = prepareCurveData(logKd,logKd_freq_acum);
% Set up fitype and options.
ft = fitype( '1/2*(1+(erf((logKd-prom)./(desv*sqrt(2))))))', 'independent', 'logKd', 'dependent',
'logKd_freq_acum' );
opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
opts.Display = 'Off';
```



```

opts.StartPoint = [0.8002804688888 0.141886338627215];
% Fit model to data.
[fitresult, gof] = fit( xData, yData, ft, opts );
% The following expressions are used to extrapolate (change max X value)
Xfit = [0:0.05:6];
Yfit = fitresult(Xfit);
% Plot fit with data.
%h = plot( fitresult );
h = plot( Xfit, Yfit, 'r' );
scatter( xData, yData, 5, 'r' );
%*****
%Data and basic calculations for third Kd group
logKd = sort(load('CEC_CaMgsup50.txt')); %paste data into this file
logKd_mean = mean(logKd);
logKd_sd = std(logKd);
logKd_freq = 1/length(logKd);
distr = normcdf(logKd,logKd_mean,logKd_sd);
logKd_freq_acum = [logKd_freq:logKd_freq:1];
%-----
% Fit: more than 50.
[xData,yData] = prepareCurveData(logKd,logKd_freq_acum);
% Set up fitype and options.
ft = fitype( '1/2*(1+(erf((logKd-prom)/(desv*sqrt(2))))))', 'independent', 'logKd', 'dependent',
'logKd_freq_acum' );
opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
opts.Display = 'Off';
opts.StartPoint = [0.8002804688888 0.141886338627215];
% Fit model to data.
[fitresult, gof] = fit( xData, yData, ft, opts );
% The following expressions are used to extrapolate (change max X value)
Xfit = [0:0.05:6];
Yfit = fitresult(Xfit);
% Plot fit with data.
%h = plot( fitresult );
h = plot( Xfit, Yfit, 'k' );
scatter( xData, yData, 5, 'k' );
%*****
%Legends and characterization for the plot
% Label axes
xlabel( 'logKd (Ra)', 'Interpreter', 'none' );
ylabel( 'Cumulative frequency', 'Interpreter', 'none' );
yline( 0.95, '--' ); yline( 0.5, '--' ); yline( 0.05, '--' );
text( 5.5, 0.97, '95th' ); text( 5.5, 0.52, '50th' ); text( 5.5, 0.07, '5th' );
yticks([0, 0.25, 0.5, 0.75, 1])
r = '$\frac{\mathrm{CEC}}{(\mathrm{Ca}^{2+}) + \mathrm{Mg}^{2+}} < 5$';
s = '$5 \leq \frac{\mathrm{CEC}}{(\mathrm{Ca}^{2+}) + \mathrm{Mg}^{2+}} < 50$';
t = '$\frac{\mathrm{CEC}}{(\mathrm{Ca}^{2+}) + \mathrm{Mg}^{2+}} \geq 50$';
legend( r, s, t, 'Location', 'NorthWest', 'Interpreter', 'latex' );
%*****

```