Deriving probabilistic soil distribution coefficients (K_d). Part 2: Reducing caesium K_d uncertainty by accounting for experimental approach and soil properties

Oriol Ramírez-Guinart[a], Daniel Kaplan[b], Anna Rigol*[,a], and Miquel Vidal[a]

[a] *Chemical Engineering and Analytical Chemistry Department, Faculty of Chemistry, University of Barcelona, Martí i Franqués 1-11, 08028, Barcelona, Spain.*

[b] *Savannah River National Laboratory, Aiken, South Carolina, USA*

*Corresponding author: Telephone: (+34) 93 403 9274; Fax: (+34) 93 402 1233

Email address: annarigol@ub.edu (A. Rigol)

Acknowledgements

This work was carried out in the frame of the IAEA MODARIA and MODARIA II programmes. It was supported by the Ministerio de Ciencia e Innovación de España (CTM2014-55191 and CTM2017-87107-R), the Generalitat de Catalunya (AGAUR 2014SGR1277), and the Department of Energy's Office of Sciences, Subsurface Biogeochemistry Research Program (DE-AD09-96SR18500), and the Savannah River National Laboratory's LDRD program (LDRD-2017-00005). Oriol Ramírez-Guinart would like to thank the support of an APIF pre-doctoral fellowship from the University of Barcelona. Authors would also like to thank Dr. Brenda Howard for their fruitful discussions and Oriol Toll for his contribution in data treatment.

Deriving probabilistic soil distribution coefficients (K_d). Part 2: Reducing caesium K_d uncertainty by accounting for experimental approach and soil properties

Abstract

The solid-liquid distribution coefficient (K_d) is a key input parameter in radioecological models. However, its large variability hampers its usefulness in modelling transport processes as well as its accuracy in representing soil-radionuclide interactions. For the specific case of radiocaesium, the analyses of a Cs K_d soil dataset (769 entries) showed that values varied over a five order of magnitude range, and the resulting Cs K_d best estimate (calculated as a geometric mean = 2.5×10^3 L kg^{-1}) lacked reliability and representativity. Grouping data and creation of partial datasets based on the experimental approach (short-term (< ~1 yr) *vs.* long-term experiments (> ~1 yr)) and soil factors affecting Cs interaction (i.e., the ratio of the radiocaesium interception potential (RIP) to the potassium content in soil solution (K_{ss}); organic matter content (OM) and soil texture) succeeded in reducing variability a few orders of magnitude, with Cs K_d best estimates also differing by one-two orders of magnitude depending on the type of soil and experimental approach. The statistical comparison of the Cs K_d best estimates and related cumulative distribution functions of the partial datasets revealed a relevant effect of the sorption dynamics on Cs K_d values (with long-term values systematically higher than short-term ones), and that the RIP/K_{ss} ratio was an excellent predictor of Cs K_d for short-term scenarios, whereas the RIP parameter could be predicted on the basis of texture information. The OM threshold to distinguish between OM threshold to distinguish between Mineral and Organic soils subclasses, regarding Cs interaction was determined to be 50% and 90% OM for short- and long-term scenarios, respectively. It was then recommended to select the Cs K_d input data depending on the soils and scenarios to be assessed (e.g., short- vs. long-term; OM %) to improve the reliability and decrease the uncertainty of the radioecological models.

Keywords

Distribution coefficient; soil; cumulative distribution function; radiocaesium; radiocaesium interception potential; probabilistic modeling

1. Introduction

Radiocaesium (Cs) interaction with soils and other materials like clays has been thoroughly studied in the last couple decades, which has led to a strong understanding of the sorption mechanisms governing Cs sorption in soils and to the identification of related soil physicochemical properties (Comans et al., 1989; Cremers et al., 1990; Vidal et al., 1995; Rigol et al., 1998; Okumura et al., 2018). Cs speciation in solution is very simple. In the natural environment, the predominant Cs species in solution are hydrolysed cations, and thus Cs-soil interaction is based on cation exchange reactions involving two main types of sorption sites with contrasting affinity for Cs and selectivity for monovalent and divalent cations. In those soils with high organic matter content and absence of 2:1 phyllosilicates, Cs sorption is mainly caused by the interaction with regular exchange sites (RES) (Rigol et al., 1998; Rigol et al., 2002). RES are present in organic matter and clay minerals as a result of deprotonation of certain functional groups and isomorphic substitutions, respectively; and they can be roughly estimated with soil cation exchange capacity (CEC). Since RES have low affinity and low selectivity coefficients for monovalent cations (*e.g.*, Cs/K or Cs/NH_4^+), Cs sorption in RES is considered as a weak and non-specific interaction (Vidal et al., 1995) that can be highly inhibited by the presence in solution of other monovalent cations and, specially, divalent cations presenting a higher electrostatic affinity, due to sorption competition processes (Comans et al., 1989; Cremers et al., 1990). When only trace levels of 2:1 phyllosilicates are present in soils, Cs sorption becomes controlled by its interaction with sorption sites located in the interlayer space of these minerals, which include illite, vermiculite or smectite and, particularly, in the frayed edge sites (FES). FES have an extremely high-affinity for monovalent cations (described in below) (Cremers et al., 1988; Okumura et al., 2018). In this context, a prediction of the Cs K_d may be attempted on the basis of the following equation (Gil-Garcia et al., 2011):

$$K_d = K_d^{FES} + K_d^{RES} = \frac{RIP_K}{K_{ss} + K_C^{FES}(NH_4/K) \cdot NH_{4,ss} + K_C^{FES}(Na/K) \cdot Na_{ss}} + \frac{K_{exch} + NH_{4,exch} + Na_{exch}}{K_{ss} + NH_{4,ss} + Na_{ss}} \quad [1]$$

in which RIP_K is the Radiocaesium Interception Potential (RIP) obtained by measuring the amount of caesium sorbed in a medium containing 100 mmol L^{-1} of Ca and 0.5 mmol L^{-1} of K (Wauters et al., 1996), K_{ss}, $NH_{4,ss}$ and Na_{ss} refer to K, NH_4^+ and Na concentrations in soil solution, K_{exch}, $NH_{4,exch}$ and Na_{exch} stand for K, NH_4^+ and Na concentrations in the exchangeable complex, and K_C^{FES} is the monovalent trace selectivity coefficients at FES.

K_C^{FES} (Na/K) takes a value around 0.02, whereas K_C^{FES} (NH_4/K) roughly varies within a 4-8 range. This equation can be simplified by only considering the RIP_K, that accounts for the soil capacity to specifically sorb Cs (Sweeck et al., 1990; Wauters et al., 1996), and the K concentration in the soil solution, as key soil properties governing Cs K_d values, specifically when sorption at RES can be disregarded, as illustrated in the Equation 2 (Gil-Garcia et al., 2011):

$$K_d^{FES} = RIP_K / K_{ss} \qquad [2]$$

The specific Cs sorption, K_d^{FES}, is crucial for the understanding the long-term Cs-soil interaction, which is the result of complex Cs-clays multiple reactions, especially with illite, montmorillonite and mordenite clays (de Koning and Comans, 2004; Ohnuki and Kozai, 2013; Durrant et al., 2018; Okumura et al., 2018). Sorbed Cs species undergo a progressive dehydration reaction with time that causes the so-called clay interlayer spacing collapse and implies that the dehydrated sorbed Cs may become trapped in the clay bulk (Wampler et al. 2012; Fuller et al., 2015). Because of this, Cs sorption to FES is a slow dynamic process, resulting in the fraction of sorbed Cs that virtually is irreversibly sorbed and no longer participates in the partition between the solid and liquid phases to increase with time. This process is known as sorption aging and enhances the fraction of irreversibly bound Cs (Absalom et al., 1995; Roig et al., 2007; Wang and Staunton, 2010; Wampler et al. 2012 Söderlund et al., 2016). Therefore, important considerations affecting the risk posed by Cs contaminated soils is whether there has been a short or long Cs contact time with the soil and whether 2:1 clays are present.

The K_d parameter is used in many radioecological risk assessment models for multiple purposes, including estimating radionuclide transport, plant-soil partitioning, and desorption from a source term (Krupka et al. 1999; Almahayni et al, 2019). K_d values are used to determine radionuclide partitioning between the dissolved and solid phases and when combined with the bulk density and porosity of the soil it can be used to calculate the retardation factor (Krupka et al. 1999), which in turn can be used to estimate the mobile radionuclide fraction (Krupka et al. 1999). In an identical calculation but with a different intent, the K_d can be used to estimate the release of radionuclides from a contaminated source. K_d values may also be used to estimate the plant to soil concentration ratio based on the assumption that plants take up primarily radionuclides from the porewater solution phase, as estimated by the denominator of the K_d. Users can estimate external doses to organisms by inputting either soil or water radionuclide activity concentrations, and then the model estimates the associated soil or water radionuclide activity concentration through the use of K_d values (Beaugelin-Seiller et al., 2002; Brown et al., 2016). K_d values can be used to estimate leaching of a radionuclide from a surface soil to an underlying zone (e.g., vadose zone or aquifer) by accounting for both plant uptake (plant:soil partitioning) and sorption during transport (retardation factor). Finally, in semi-mechanistic models focusing in Cs, specific parameters such as RIP can also be used to internally predict Cs K_d values (Equation 2) (Absalom et al, 2001; Tarsitano et al., 2011). In each of these uses of the K_d parameter, data can be entered as a single value or as a probability density function (Simon-Cornu et al., 2015).

This work is the second in an initial series of three publications (Ramírez-Guinart et al., 2020a; Ramírez-Guinart et al., 2020b) aiming at deriving sorption data suitable for risk assessment from soil K_d datasets, as well as to develop a strategy to reduce and describe K_d variability based on probabilistic models, including the construction of distribution functions to statistically describe the K_d values of a target radionuclide (RN), in this case radiocaesium. As explained in part 1 (Ramírez-Guinart et al., 2020a), the International Atomic Energy Agency (IAEA) K_d dataset described in Technical Reports Series Number 472 (TRS-472; IAEA, 2010) was the starting point of the work. Under the auspices of the IAEA-MODARIA (Modelling and Data for Radiological Impact Assessment) project, the TRS-472 dataset was updated and critically reviewed following agreed acceptance criteria by the MODARIA Working Group 4, including: 1) rejecting any K_d value not directly quantified as the ratio between concentrations of the target element measured in a liquid and a solid phase (i.e., reject data from parametric equations, mass-transport experiments, or K_d reference values were excluded); 2) rejecting K_d values created by pooling values originating from varying non-relevant operational or soil variables; 3) excluding values obtained from experiments not representative for environmental conditions (such as extremely low or high pH); 4) accepting data from stable isotopes obtained at the lowest concentration range; and 5) rejecting data from pure (soil) mineral phases, such as clay minerals or metal (hydro)oxides. The resulting critically reviewed dataset contains >7000 soil K_d entries for 83 elements (Ramírez-Guinart et al., 2020a), of which 769 entries describe soil Cs K_d values.

Previous work has derived Cs K_d best estimate values from large datasets, as well from partial datasets created based on the RIP parameter and texture and organic matter content (Gil-Garcia et al., 2009). The objectives of this study are to evaluate and quantify new potential sources of variability of Cs K_d values. More specifically, the objectives of this study were: 1) to evaluate if Cs contact time with soil (i.e., short-term vs. long-term; < ~1 yr and > ~1 yr, respectively) affects Cs K_d values and therefore be a source for Cs K_d variability; 2) to evaluate several groupings aligned with various soil properties, including RIP/K_{ss}, and soil OM+Texture; 3) because RIP is not commonly measured, evaluate whether it can be estimated by common soil texture properties; and 4) to determine the organic matter (OM) content threshold to optimally distinguish between mineral and organic soils to permit reducing Cs K_d uncertainty. The intent of this study was not only to identify significant differences between these various soil categories, but also to improve present approaches to selecting Cs K_d values to minimize uncertainty, thereby improving input data of radioecological models.

2. Data collection and treatment

2.1. Current status of the Cs K_d compilation

The updated MODARIA Cs dataset contains 769 entries with related soil characteristics and details regarding the experimental approach (Ramírez-Guinart et al., 2020a). With respect to the TRS-472 compilation, a significant set of data gathered from desorption experiments of indigenous Cs has been integrated because of the changes introduced in data acceptance criteria (POSIVA, 2014; SKB, 2014). The soil Cs K_d overall dataset contained values ranging within up to five orders of magnitude (Min-Max range of 4×10^0 - 4.5×10^5 L kg^{-1}). The large Cs K_d variability of the dataset denotes the presence of data from soils with contrasting key properties (*e.g.*, RIP, OM, clay content, K in soil solution, etc.) and from different experimental approaches.

Besides fields within the dataset related to the sources of information, radioisotopes, soil characteristics and ancillary information (such as pH, organic matter content, cationic exchange capacity (CEC), clay and sand contents referred to mineral matter; exchangeable K and NH_4; concentration of K and NH_4 in soil solution; and RIP), additional fields were included related to the experimental approach (either short or long-term experiments).

2.2. Soil factors and developed criteria to group Cs K_d data

Cs K_d values have to be grouped according to soil factors specific to the Cs sorption mechanisms. As discussed above, the Cs K_d^{FES}, which is equal to the RIP/K_{ss} (Eq. 2), was used as a criterion for reducing Cs K_d variability (Gil-García et al., 2009). Four RIP/K_{ss} ranges were created: $RIP/K_{ss} < 10^2$; $10^2 \leq RIP/K_{ss} < 10^3$; $10^3 \leq RIP/K_{ss} < 10^4$; and $RIP/K_{ss} > 10^4$, as previously agreed in past analyses (IAEA, 2009).

Ideally the soil factor used to categorize soils for predicting Cs K_d values would be based directly on the concentration of the 2:1 clay minerals that provide FES, rather than RIP. However, such measurements are costly (involving multiple X-ray diffraction analyses of a single soil sample) and are typically not conducted in routine soil analyses. Consequently, few mineralogy characterisation data are available for soils, thus making any mineralogy-based grouping criteria of limited practical use. Thus, a K_d grouping criterion based on the soil texture and OM content, the so-called OM+Texture criterion previously defined and agreed upon (IAEA, 2010), was also applied to group Cs K_d data. In short, a Cs K_d value was included in the rganic group if the soil had an OM content ≥ 20%, whereas it was included in the Mineral group if OM was lower than 20%. Secondly, the K_d data contained in the Mineral group were split in three textural groups (Sand, Loam, and Clay) when textural data were available. Based on percentage of the mineral fraction, the Sand group was defined by a sand fraction ≥ 65%, and a clay fraction < 18%; Clay group: clay fraction > 35%; and Loam group, rest of cases. The suitability of the OM+Texture K_d grouping criterion to propose soil-type Cs K_d data lies on the fact that even though the OM+Texture criterion is not based on the fundamental description of the underlying sorption processes of Cs in soils, it partially captures some of the soil properties that can play a key role in the Cs-soil interaction.

The OM threshold traditionally used to discern between an organic and a mineral soil may not be the most appropriate for Cs K_d. A 20% OM threshold to distinguish between soils with a high Cs sorption capacity due to the presence of the mineral fraction and those with a low sorption capacity due to the absence of FES could not be appropriate, as a minor content of mineral fraction can govern Cs sorption even in organic soils such as histosols (Vidal et al., 1995; Rigol et al.,1998). Therefore, the effect of varying the OM concentration threshold was also analysed to redefine the OM+Texture criterion.

2.3. Analysis of the influence of the experimental approach on Cs K_d data variability

As with other papers in this series, the influence of the experimental approach was simultaneously evaluated along with relevant soil factors for reducing Cs K_d variability. From the three experimental approach categories (short-term sorption (ST-S), short-term desorption (ST-D), and long-term desorption (LT-D)) (Ramirez-Guinart et al., 2020a), the greatest number of K_d entries were in the "short-term sorption" category (that is, Cs K_d derived from applying a sorption batch test based on putting in contact for short times a non-contaminated soil with a solution spiked with radiocaesium or with low concentrations of stable Cs), and "long-term desorption" (that is, K_d of anthropogenic Cs derived from applying an extraction test to long-term contaminated solid materials with radiocaesium, or K_d derived from indigenous Cs from performing an extraction test when the total content of the indigenous Cs at the solid matrix is quantified). There were no entries that could be considered as "short-term desorption data" (Cs K_d derived from anthropogenic radiocaesium from applying an extraction batch test to soils recently contaminated with radiocaesium). Therefore, according to the entries available for the experimental approach categories, only the differences between the overall short-term versus long-term datasets were finally examined.

The data treatment was based on group mean centering (GMC) to minimize the effect of soil factors on the interaction terms and to examine better the individual role of the experimental approaches on Cs K_d variability (Bell et al., 2018). Regarding soil factors and based on the previous experience of similar grouping exercises (Gil et al., 2009), these analyses could be carried out by either considering the RIP/K_{ss} or the OM+Texture criteria. For the GMC treatment, the use of the RIP/K_{ss} factor was dismissed as the RIP concept accounts for short-term and reversible sorption scenario. Therefore, the GMC was only addressed to minimize the effect of OM and texture on elucidating the role of the experimental approach factor. Firstly, the overall Cs dataset was log-transformed, the log Cs K_d data was then grouped according to the OM+Texture criterion. The arithmetic mean (AM) of the log Cs K_d values of each soil-type group was calculated and each log Cs K_d value within a given group was corrected by subtracting the AM log Cs K_d value of the respective soil-type group. Subsequently, the GMC-corrected log Cs K_d datasets were divided according to the type of the experimental approach and statistical tests (Fisher's least significant differences (FLSD) test for multiple means; 95% confidence level; StatGraphics 18) were performed to check whether the Cs K_d means for each experimental approach significantly differed.

2.4. Construction of cumulative distribution functions to describe Cs K_d variability

Cumulative Distribution Functions (CDF) of Cs K_d data were constructed to describe the population and variability of each groupings' datasets. Since the K_d parameter is a ratio of concentrations, K_d data are expected to follow a lognormal distribution (Sheppard et al., 2011). Thus, lognormal was the first function distribution of the Cs K_d data tested. For the construction of CDFs, Cs K_d data were log-transformed and the presence of possible outlier values in the datasets was examined by performing an exploratory analysis based on box-and-whisker plots. The log Cs K_d data within every dataset were sorted by increasing value and an empirical frequency ($f_{exp,i}$) equal to 1/N (where N is the total number of Cs K_d entries in the respective dataset) was assigned to each entry. Experimental group cumulative frequency distributions were constructed by assigning to each sorted log Cs K_d value their corresponding cumulative frequency ($F_{exp,i}$), *i.e.*, the sum of the preceding frequencies $\left(F(K_{d,j}) = \sum_{i=0}^{j} f(K_{d,i})\right)$. The Kolmogorov-Smirnov test was then applied to ascertain whether the underlying frequency distribution in each Cs K_d dataset was significantly different from the lognormal distribution. As expected, it was confirmed that overall and partial Cs K_d datasets followed a lognormal distribution. Consequently, the experimental cumulative frequency distributions constructed with the log Cs K_d data were fitted to the theoretical normal CDF equation, and the related geometric mean (GM) and percentile ranges were derived. Further details can be found elsewhere (Ramírez-Guinart et al., 2020a).

To derive properly a reliable CDF from a given K_d dataset it is necessary that it contains a minimum number of entries ($N \approx 10$). However, there were a few partial datasets containing less than 10 entries for which CDFs were evaluated. Those partial datasets that provided a good fit to the lognormal distribution were reported and for the rest of the cases only GM values were calculated.

3. Analyses of Cs K_d distributions

3.1. Influence of the experimental approach on Cs K_d data

The overall Cs K_d dataset contained K_d data gathered by applying sorption experiments in a short-term scenario (ST-S), and desorption experiments in a long-term scenario (LT-D). When the statistical analysis was performed after applying the GMC to the partial datasets created from the application of the OM+Texture criterion, significant differences were observed between ST-S and LT-D (data not shown). Thus, short-term and long-term Cs K_d data must be distinguished before testing any further grouping criteria, which agrees with the fact that long-term incorporated Cs may undergo an aging process leading to an increase in the Cs sorption irreversibility.

Figure 1 shows the Cs K_d descriptors of the distributions of the short-term and long-term partial datasets from the overall dataset, as well as the related CDFs. The Cs K_d data and CDFs for the long-term incorporated Cs (GM and 5^{th} – 95^{th} percentile range values) were an order of magnitude or more greater than those of short-term. An implication of this finding is that the use of long-term Cs K_d data should be avoided to predict soil Cs sorption behaviour in recently contamination systems, and *vice versa,* values from short-term measurements of Cs K_d should be avoided to predict long-term Cs sorption behaviour.

3.2. Cs K_d best estimates and CDFs based on the RIP/K_{ss} criterion

Of the 769 entries in the Cs K_d dataset, 328 contained sufficient ancillary information for calculating the RIP/$K_{ss.}$ The log-log correlation between experimental Cs K_d and the respective RIP/K_{ss} values could explain 64% of the total Cs K_d variance, representing a profound result considering the large variability of the dataset (see equation 3 and Figure S1 in the Supplementary Material):

$$\log K_d = 0.76\ (\pm\ 0.18) + 0.73\ (\pm\ 0.06) \times \log(\mathrm{RIP}/K_{ss}) \quad (N = 328;\ r = 0.80;\ p = 2.6\ x10^{-74}) \quad [3]$$

In this context, the construction of the CDFs helps to describe the variability associated with the use of the RIP/K_{ss} ratio, to propose Cs K_d best estimates derived from the use of the RIP/K_{ss} criterion, and to quantify their uncertainty.

Figure 2 depicts the graphical representation of the CDFs constructed from each partial dataset created by applying the RIP/K_{ss} criterion as well as the main quantitative outcomes from the CDF construction. The Cs K_d GMs and related 5^{th}-95^{th} percentile ranges increased when increasing each RIP/K_{ss} group. Whereas the examination of the Cs K_d GM is of a lesser importance in this case (as an end-user with available RIP and K_{ss} data can straightforwardly calculate an approximate, related Cs K_d value), the 5^{th}-95^{th} percentiles permit to quantify and describe the Cs K_d variability within each dataset. The variability in the partial datasets was only of one to two orders of magnitude. Besides, the constructed CDFs curves did not overlap among them. Therefore, the calculation of the RIP/K_{ss} ratios permits a rapid estimation of the Cs K_d, with an associated uncertainty calculated from the corresponding CDF.

A major limitation of this approach lies on the fact that it is necessary to have the soil RIP value, a parameter that albeit being more and more frequently determined it is not yet characterised on routine analyses. Therefore, end-users may find useful an equation enabling the prediction of soil RIP values from soil properties often available or, at least, much easier to determine than the RIP parameter. Previous studies demonstrated that RIP values can be roughly predicted from soil clay and silt contents (Waegeneers et al. 1999; Gil-García et al., 2011). Here, a multiple linear regression is created from data from the current compilation as well as from data from other works in which RIP was measured along with other soil properties (Vandebroek et al., 2012; Uematsu et al., 2015). The correlation captured around 70% of total RIP variability and reliably correlated RIP values of soils with their clay and silt contents:

$$\log \text{RIP} = 1.24\ (0.09) + 0.76\ (0.06)\times\log \text{Clay} + 0.68\ (0.06)\times\log \text{Silt} \quad (N = 225;\ r = 0.82;\ p = 1.4\,x10^{-54}) \quad [4]$$

This model could be improved if the 2:1 phyllosilicates content would be quantified (Nakao et al., 2015; Uematsu et al., 2015), although this information is expensive and not available in routine analyses. Thus, the development of an equation to predict RIP values not only from clay content but from clay mineralogy remains a future challenge.

A secondary limitation of this approach is that the K concentration in the soil solution is not routinely analysed by researchers. However, there have been a few attempts to estimate this parameter from other soil properties, such as from exchangeable K and K K_d estimates reported for mineral and organic soils; from the total K content, the CEC in clay minerals and from the percentage of the exchange sites on soil clay minerals occupied by K. In the case of organic soils, the gravimetric humus content of the soil is also required as well as to distinguish between CEC of humus and clay sites (Absalom et al., 2001; Gil-García et al., 2009).

3.3. Cs K_d best estimates and CDFs based on the OM+Texture criterion

3.3.1. Cs K_d best estimates and CDFs based on the initial OM+Texture criterion

The Cs K_d dataset, refined to include those entries with the information required for the OM+Texture criterion, contained 573 entries, an additional 100 entries compared to the TRS-472 dataset (IAEA, 2010). Table 1 summarises the Cs K_d data obtained from the CDFs constructed by applying the OM+Texture criterion, distinguishing between short- and long-term partial datasets, as well as by mineral and organic soils, and when statistically significant, textural classes within the mineral soils.

With the short-term data, the GMs derived from the CDFs created with the short-term data evidenced that Cs K_d values for the Mineral dataset were statistically greater than that of the Organic dataset. Within the Mineral dataset, the K_d GMs increased along with the soil clay content ($GM_{sand} < GM_{loam} < GM_{clay}$). However, and although kept separately in the table, Clay and Loam datasets were not significantly different. This pattern is generally consistent with reported Cs sorption mechanisms, Cs partitioning between FES and RES sites and the relatively weak sorption of Cs to natural OM sites (Rigol et al., 2002). For the long-term data, the GM derived for the Mineral group was also one order of magnitude higher than that of the Organic group, although no statistical differences were observed among the textural groups among them and with respect to the Mineral dataset.

The GM values (Table 1) derived from the Mineral soil group created from the short-term dataset was statistically lower than that of the long-term dataset (more than one order of magnitude) and the 5^{th}-95^{th} intervals of the long-term Mineral dataset were shifted to higher Cs K_d values, which corroborates the influence of the sorption dynamics on the Cs K_d values as observed in the Section 3.1. Moreover, the same pattern was also observed for the Organic datasets, as the GM values of the Organic dataset in the long-term scenario was one order of magnitude higher than that of the short-term. This may indicate that the 20% OM content threshold to distinguish between mineral and organic soils regarding Cs interaction is too low for this radionuclide. This observation may be attributed to FES out competing the RES that account for Cs sorption to OM (Rigol et al., 1998; Roig et al, 2007).

The comparison of the 5^{th}-95^{th} percentile ranges of Cs K_d values obtained from the OM+Texture partial datasets with that of the overall Cs K_d dataset indicates that the application of the OM+Texture criterion to short-term data allowed us to create Cs K_d mineral textural groups with a much lower variability (lower GSD and more narrow percentile ranges) than that of the overall data set (down to one-to-two orders of magnitude in a few cases) (Table 1). Conversely, organic soils datasets contained Cs K_d values still varying within ranges similar to those of overall short- and long-term datasets. Thus, these results suggest that the initial OM+Texture criterion could be improved by better establishing the OM% threshold for Cs to distinguish between mineral and organic soils.

3.3.2. Cs K_d best estimates and CDFs based on the redefined OM+Texture criterion

The effect of changing the OM thresholds to better distinguish between mineral from organic soils were tested for short- and long-term partial Cs K_d datasets. Criteria to select the new OM thresholds were based on 1) to analyse significant changes in the derived Cs K_d GMs; 2) to obtain organic and mineral soil distributions with enough entries and minimum variability; and 3) to obtain similar Cs K_d best estimates for short- and long-term organic soil datasets.

A few entries were excluded if the OM content was not reported. As summarized in Table S1 in the Supplementary Material, the GMs of the Mineral datasets were statistically the same regardless the OM threshold, whereas the GMs from the Organic datasets progressively decreased by increasing the OM threshold up to 50%, and remained statistically constant for the 60% OM threshold. Besides this, the variability of the Mineral datasets was roughly the same for all the OM thresholds tested, whereas variability decreased significantly for the Organic datasets when increasing the OM threshold up to 50%, with no further improvement for the 60% threshold. Therefore, a 50% OM threshold is suggested for short-term datasets.

Regarding long-term data, data limitation required that we evaluate greater threshold values, up to 90%. The GM Cs K_d and related variability of the Mineral datasets remained constant regardless the OM threshold, whereas for the Organic datasets GMs and related variability were gradually decreased as the OM threshold value was increased. The decrease was statistically significant when the 90% OM threshold was applied. From these results, it is suggested a 90% OM content as a threshold to distinguish between organic and mineral soils for long-term data.

The third criterion for establishing the new thresholds was only partially achieved. Whereas the GM of the organic short-term and long-term datasets approached with the new thresholds (from one-order of magnitude difference to a lower 4 fold), the datasets were still statistically different at 95% confidence level, although with a *p* value of 0.043. Therefore, the two datasets could be considered as comparable. However, greater organic datasets according to the redefined thresholds with more entries (especially for the long-term scenario) are needed to be able to statistically fulfil this criterion. If an overall short-term plus long-term organic dataset is built up with the redefined OM thresholds, with 45 entries, a best-estimate of 1.1×10^2 L kg^{-1} (GSD = 4.2) can be derived from the overall organic dataset.

Figure 3 summarises the Cs K_d data obtained from the CDFs constructed by applying the redefined OM+Texture criteria to the short-term and long-term partial datasets and the CDF graphical representations. The changes introduced concerning the OM thresholds resulted in partial datasets with much lower variability than when using the initial OM+Texture criterion. The derived Cs K_d GM values increased with increasing clay content leading to well-defined CDFs among textural groups, also for the long-term datasets. Besides, for a given textural soil group, the long-term Cs K_d data were systematically much higher (around one order of magnitude) than those corresponding to the short-term data.

4. Conclusions and recommendations

From the analyses performed to the Cs K_d dataset, it was found that sorption dynamics effects (i.e., long-term vs. short-term scenarios) had a strong impact on the Cs K_d values and related variability. This fact should be taken into consideration when dealing with risk assessment exercises in which Cs K_d data are required. Besides, it was also evidenced that soil properties either directly related to the mechanisms governing Cs sorption in soils, like soil RIP and K concentration in soil solution, or indirectly related, such as the soil OM and to a lesser extent the soil texture, dramatically affected the Cs K_d values and their variability. These soil properties should be available for a proper estimation and selection of Cs K_d.

A single Cs K_d best estimate and/or CDF has little practical value for modelling because it is fraught with high variability and it is not assured that it is representative of the target scenario. Alternatively, it is highly recommended to end-users to select the Cs K_d best estimates and CDF that corresponds to their scenario of interest. First, it is crucial to identify if the assessment is made for a recent contamination episode such as right after a radioactive accidental release (short-term scenario, $< \sim 1$ yr) or for a post-contamination episode that occurred a long time ago or for predictions extended to the future (e.g., in the context of safety and performance assessment of deep geological repositories or long-term impact assessment of contamination episodes). If the radiological assessment exercise is done for a short-term scenario and the RIP and K_{ss} data are available for the studied soil, it is recommended to use a Cs K_d value based on the direct calculation of the RIP/K_{ss} ratio, associated with the uncertainty derived from the constructed CDF for the corresponding RIP/K_{ss} group, in which, if required, RIP can be predicted directly from the clay and silt contents of the soil. K_{ss} could also be derived from other soil properties, such as total K, and CEC of clay and humus fractions. Both from short- and long-term scenarios, if soil organic matter content of the target soil is known, it is suggested to use the CDF that also suits the soil type (Organic or Mineral), whereas if soil texture data is also available it is suggested to refine the CDF election also according to the textural group.

References

Absalom, J. P., Young, S. D., Crout, N. M., 1995. Radio-caesium fixation dynamics: measurement in six Cumbrian soils. Eur. J. Soil Sci. 46, 461-469.

Absalom, J.P., Young, S.D., Crout, N.M.J., Sanchez, A., Wright, S.M., Smolders, E., Nisbet, A.F., Gillett, A.G., 2001. Predicting the transfer of radiocaesium from organic soils to plants using soil characteristics. J. Environ. Radioact. 52, 31–43.

Almahayni, T., Beresford, N.A., Crout, N.M.J., Sweeck, L. 2019. Fit-for-purpose modelling of radiocaesium soil-to-plant transfer for nuclear emergencies: a review. J. Environ. Radioact. 201, 58-66.

Beaugelin-Seiller, K., Boyer, P., Garnier-Laplace, J., Adam, C. 2002. Casteaur: a simple tool to assess the transfer of radionuclides in waterways, Health. Phys. 84, 539–542.

Bell, A.; Jones. K., Fairbrother, M. 2018. Understanding and misunderstanding group mean centering: a commentary on Kelley et al.'s dangerous practice. Quality & Quantity 52, 2031–2036.

Brown, J.E., Alfons, B., Avila, R., Beresford, N.A., Copplestone, D., Hosseini, A. 2016. A new version of the ERICA tool to facilitate impact assessments of radioactivity on wild plants and animals, J. Environ. Radioact. 153, 141–148.

Comans, R. N., Middelburg, J. J., Zonderhuis, J., Woittiez, J. R., De Lange, G. J., Das, H. A., Van der Weijden, C. H., 1989. Mobilization of radiocesium in pore water of lake sediments. Nature 339, 367-369.

Cremers, A., Elsen, A., De Preter, P., Maes, A., 1988. Quantitative analysis of radiocesium retention in soils. Nature 335, 247-249.

Cremers, A., Elsen, A., Valcke, E., Wauters, J., Sandalls, F. J., Gaudern, S. L., 1990. In: Desmet, G., Nassibeni, P., Belli, M. (Eds.) *Transfer of radionuclides in natural and semi-natural environments*, New York, USA: Elsevier Applied Science Publishers.

de Koning, A., Comans, R. 2004. Reversibility of radiocaesium sorption on illite. Geochem. Cosmochim. Acta, 68, 2815-2823.

Durrant, C.B., Begg, J.D., Kersting, A.B., Zavarin, M. 2018. Cesium sorption reversibility and kinetics on illite, montmorillonite, and kaolinite. 2018. Sci. Total Environ. 610–611, 511-520.

Fuller, A.J., Shaw, S., Ward, M.B., Haigh, S.J., Mosselmans, J.F.W., Peacock, C.L., Stackhouse, S., Dent, A.J., Trivedi, D., Burke, I.T. 2015. Caesium incorporation and retention in illite interlayers. Appl. Clay Sci. 108, 128-134.

Gil-García, C., Rigol, A., Vidal, M., 2009. New best estimates for radionuclide solid–liquid distribution coefficient in soils. Part 1: radiostrontium and radiocaesium. J. Environ. Radioact. 100, 690–696.

Gil-García, C., Rigol, A., Vidal, M., 2011. Comparison of mechanistic and PLS-based regression models to predict radiocaesium distribution coefficients in soils. J. Hazard. Mat. 197, 11– 18.

IAEA, 2009. TECDOC-1616 - Quantification of radionuclide transfer in terrestrial and freshwater environments for radiological assessments. International Atomic Energy Agency, Austria.

IAEA, 2010. TRS-472 - Handbook of Parameter Values for the Prediction of Radionuclide Transfer in Terrestrial and Freshwater Environments. International Atomic Energy Agency, Austria.

Krupka, K.M., Kaplan, D.I., Whelan, G., Serne, R.J., Mattigod, S.V. 1999. Understanding variation in partition coefficient, K_d, values. Volume 1: The K_d model, methods of measurement, and application of chemical reaction codes. Rep. No. EPA 402-R-99-004A. US Environmental Protection Agency, Washington, DC.

Nakao, A., Takeda, A., Ogasawara, S., Yanai, J., Ito, T. 2015. Relationship between paddy soil Radiocesium Interception Potentials and physicochemical properties in Fukushima, Japan. J. Environ. Qual. 44, 780-788.

Okumura, M., Kerisit, S., Bourg, I.C., Lammers, L.N., Ikeda, T., Sassi, M., Rosso, K.M., Machida, M. 2018. Radiocesium interaction with clay minerals: Theory and simulation advances Post–Fukushima. J. Environ. Radioact. 189, 135-145.

Ohnuki, T., Kozai, N. 2013. Adsorption behavior of radioactive cesium by non-mica minerals. J. Nucl. Sci. Technol. 50, 369-375.

POSIVA, 2014. Working Report 2013-66 - Geochemical and physical properties, distribution coefficients of soils and sediments at the Olkiluoto Island and in the reference area in 2010-2011. POSIVA, Finland.

Ramírez-Guinart, O., Kaplan, D., Rigol, A., Vidal, M. 2020a. Deriving probabilistic soil distribution coefficients (K_d). Part 1: General approach to decreasing and describing variability and example using uranium K_d values. Submitted to J. Environ. Radioactiv.

Ramírez-Guinart, O., Kaplan, D., Rigol, A., Vidal, M. 2020b. Deriving probabilistic soil distribution coefficients (K_d). Part 3: Reducing variability of americium K_d best estimates using soil properties and chemical and geological material analogues. Submitted to J. Environ. Radioactiv.

Rigol, A., Vidal, M., Rauret, G., 1998. Competition of organic and mineral phases in radiocesium partitioning in organic soils of Scotland and the area near Chernobyl. Environ. Sci. Technol. 32, 663-669.

Rigol, A., Vidal., M., Rauret, G., 2002. An overview of the effect of organic matter on soil-radiocaesium interaction: implications in root uptake. J. Environ. Radioactiv. 58, 191-216.

Roig, M., Vidal, M., Rauret, G., Rigol, A., 2007. Prediction of radionuclide aging in soils from the Chernobyl and Mediterranean areas. J. Environ. Qual. 36, 943-952.

Sheppard, S. C., 2011. Robust Prediction of K_d from Soil Properties for Environmental Assessment. Hum. Ecol. Risk Assess. 17(1), 263-279.

Simon-Cornu, M., Beaugelin-Seiller, K., Boyer, P., Calmon, P., Garcia-Sanchez, L., Mourlon, C., Nicoulaud, V., Sy, M., Gonze. M.A. 2015. Evaluating Variability and Uncertainty in Radiological Impact Assessment Using SYMBIOSE. J. Environ. Radioact. 139, 91-102.

SKB, 2014. TR-14-002 - Initial state report for the safety assessment SR-PSU. Svensk Kämbränslehantering AB-Swedish Nuclear Fuel and Waste Management Co, Sweden.

Söderlund, M., Virtanen, S., Välimaa, I., Lempinen, J., Hakanen, M., Lehto, J. 2016. Sorption of cesium on boreal forest soil II. The effect of time, incubation conditions, pH and competing cations. J. Radioanal. Nucl. Chem. 309, 647–657.

Sweeck, L., Wauters, J., Valcke, E., Cremers, A., 1990. In: Desmet, G., Nassibeni, P., Belli, M. (Eds.), *Transfer of radionuclides in natural and semi-natural environments*, New York, USA: Elsevier Applied Science Publishers.

Tarsitano, D., Young, S.D., Crout, N.M.J., 2011. Evaluating and reducing a model of radiocaesium soil–plant uptake. J. Environ. Radioact. 102, 262–269.

Uematsu, S., Smolders, E., Sweeck, L., Wannijn, J., Van Hees, M., Vandenhove, H., 2015. Predicting radiocaesium sorption characteristics with soil chemical properties for Japanese soils, Sci. Total Environ. 524-525, 148-156.

Vandebroek, L., van Hees, M., Delvaux, B., Spaargaren, O., Thiry, Y., 2012. Relevance of Radiocaesium Interception Potential (RIP) on a worldwide scale to assess soil vulnerability to 137Cs contamination. J. Environ. Radioactiv. 102, 87-93.

Vidal, M., Roig, M., Rigol, A., Llauradó, M., Rauret, G., Wauters, J., Elsen, A., Cremers, A., 1995. Two approaches to the study of radiocaesium partitioning and mobility in agricultural soils from the Chernobyl area. Analyst 120, 1785-1791.

Waegeneers, N., Smolders, E., Merckx, R., 1999. A statistical approach for estimating the radiocaesium interception potential of soils. J. Environ. Qual. 28, 1005–1011.

Wampler, J. M., Krogstad E. J., Elliott W. C., Kahn B., Kaplan D.H., 2012. Long-term selective retention of natural Cs and Rb by highly weathered Coastal Plain soils. Environ. Sci. Technol. 46, 3837-3843.

Wang, G., Staunton, S. 2010. Dynamics of caesium in aerated and flooded soils: Experimental assessment of ongoing adsorption and fixation. Eur. J. Soil Sci. 61(6), 1005-1013.

Wauters, J., Elsen, A., Cremers, A., Konoplev, A. V., Bulgakov, A. A., Comans, R. N. J., 1996. Prediction of solid liquid distribution coefficients of radiocaesium in soils and sediments. Part 1: A simplified procedure for the solid characterisation, Appl. Geochem. 11, 589–594.

Table 1

Descriptors of Cs K_d ($L\ kg^{-1}$) distributions after applying the initial OM+Texture criterion.

Partial dataset	N	GM	GSD	FLSD[a]	5th	95th
Short-term						
Overall	405	2.2×10^{3}	5.7	A[b]	5.2×10^{1}	2.2×10^{4}
Organic	60	1.8×10^{2}	6.1	B[b]	2.0×10^{1}	4.1×10^{3}
Mineral	345	2.7×10^{3}	4.3	C[b]	1.3×10^{2}	2.4×10^{4}
Clay	32	5.9×10^{3}	3.4	A[c]	6.8×10^{2}	3.5×10^{4}
Loam	190	3.7×10^{3}	3.2	A[c]	5.9×10^{2}	2.6×10^{4}
Sand	110	1.3×10^{3}	5.0	B[c]	5.6×10^{1}	1.0×10^{4}
Long-term						
Overall	168	2.4×10^{4}	4.2	A[d]	1.6×10^{3}	1.5×10^{5}
Organic	20	2.0×10^{3}	7.4	B[d]	1.1×10^{2}	9.2×10^{4}
Mineral	148	2.8×10^{4}	2.6	C[d]	6.8×10^{3}	1.5×10^{5}

N = number of observations, GM = geometric mean, GSD = geometric standard deviation

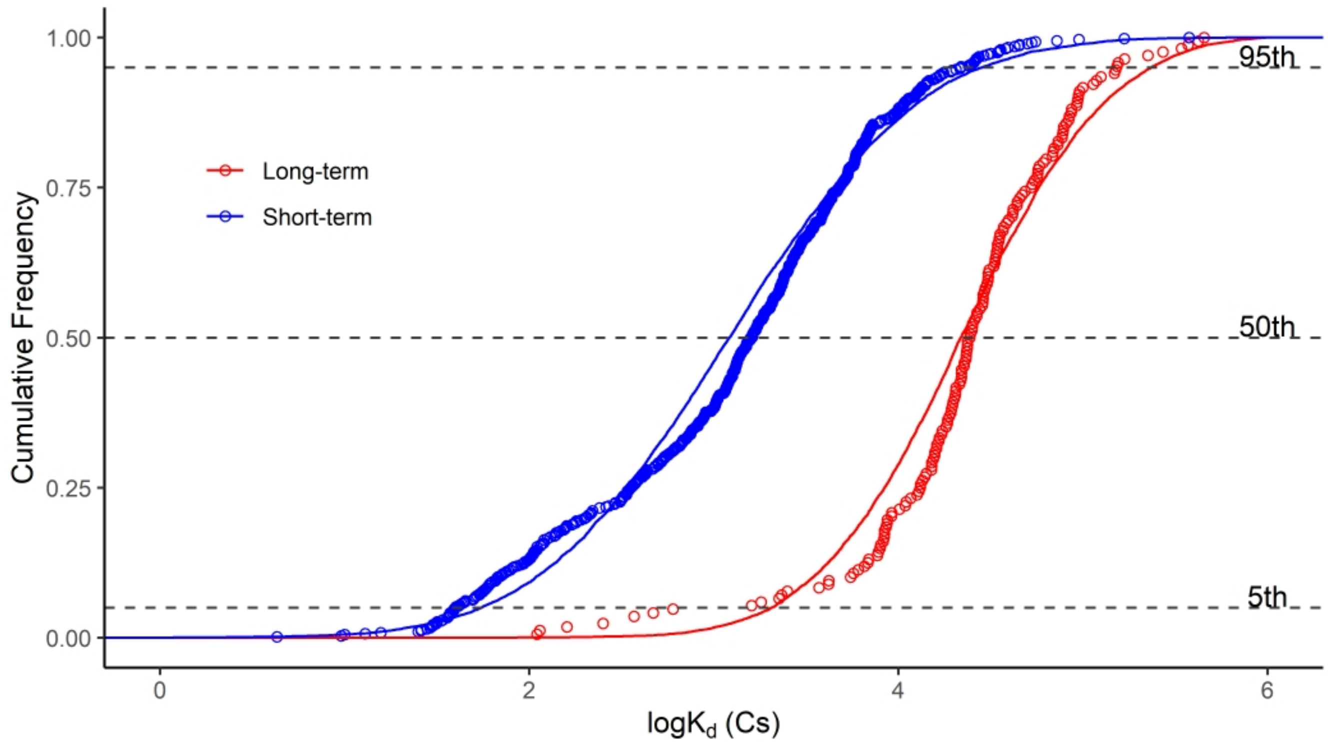
[a] Different letters among the datasets compared indicate statistically significant differences between GMs according to the Fisher's Least Significant Differences test. Dataset comparisons shown here are: [b] short-term overall, mineral, and organic datasets; [c] short-term textural datasets; [d] long-term overall, mineral, and organic datasets.

Figure captions

Fig. 1. CDFs and descriptors of Cs K_d (L kg^{-1}) distributions for soils grouped according to the Experimental Approach. Data for the Overall dataset are included for comparison. Points indicate individual dataset values whereas lines indicate the fitted distributions.

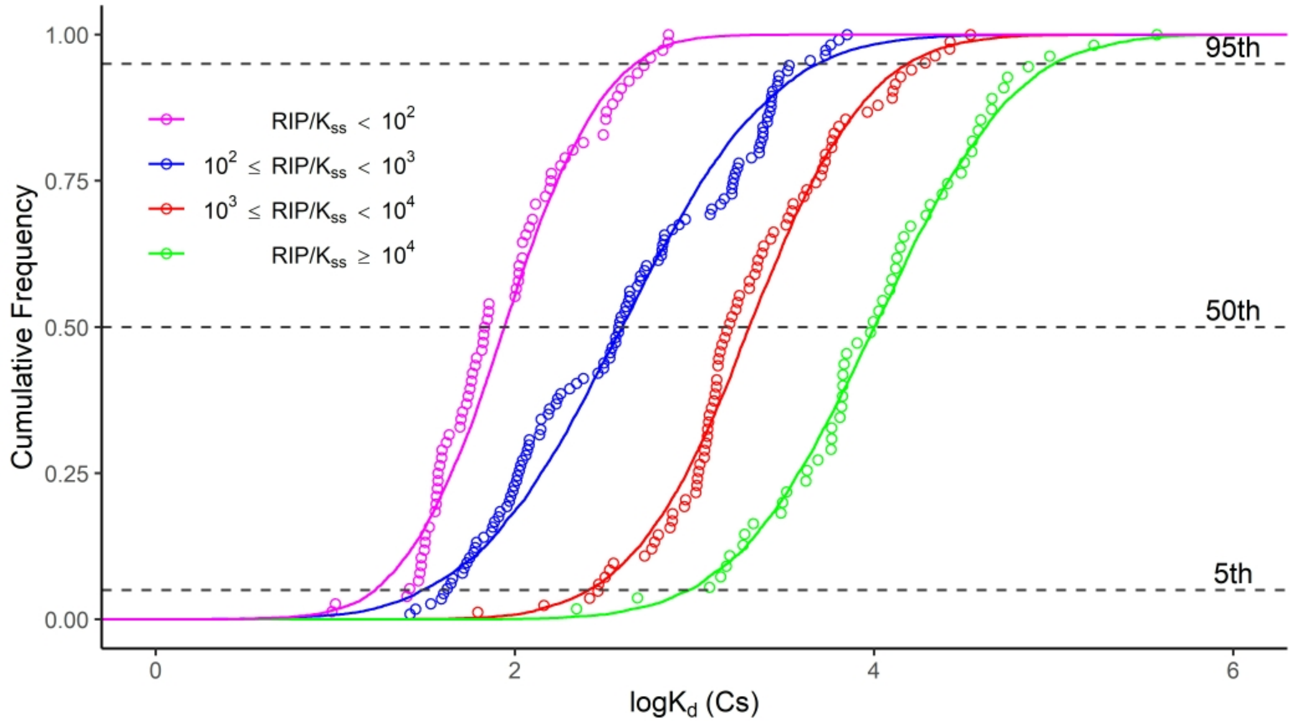
Fig. 2. CDFs and descriptors of Cs K_d (L kg^{-1}) distributions for soils grouped according to RIP/K_{ss} criterion. Points indicate individual dataset values whereas lines indicate the fitted distributions.

Fig. 3. CDFs and descriptors of Cs K_d (L kg^{-1}) distributions derived from the redefined OM+Texture criterion for short-term (A) and long-term (B) partial datasets. Points indicate individual dataset values whereas lines indicate the fitted distributions.



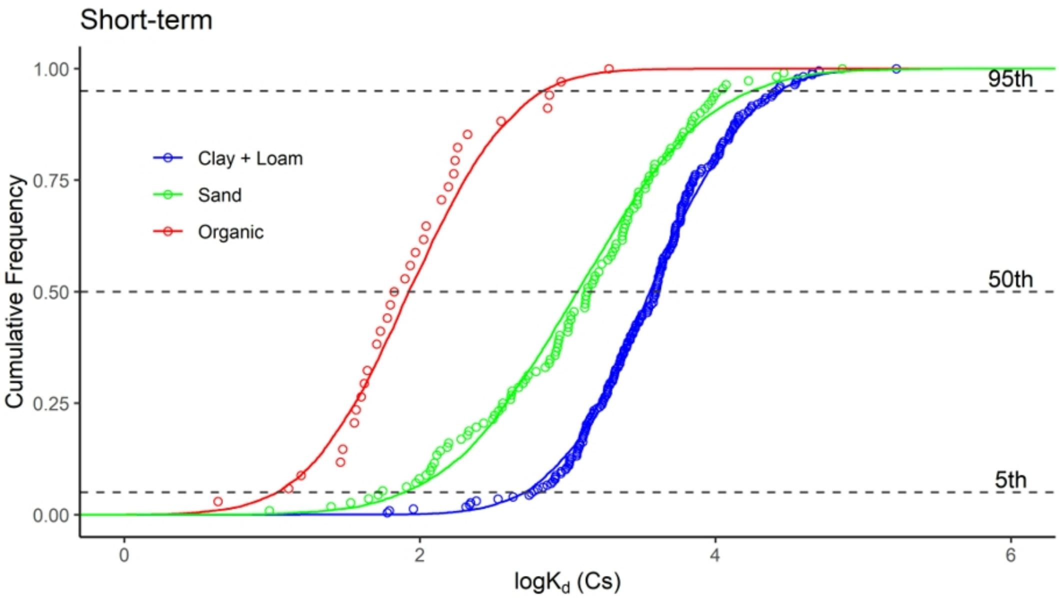
Partial dataset	N	GM	GSD	FSLD[a]	5th	95th
Overall	769	2.5×10^3	8.6		5.0×10^1	6.3×10^4
Short-term	601	1.6×10^3	6.6	A	4.0×10^1	2.2×10^4
Long-term	168	2.4×10^4	4.2	B	1.6×10^3	1.5×10^5

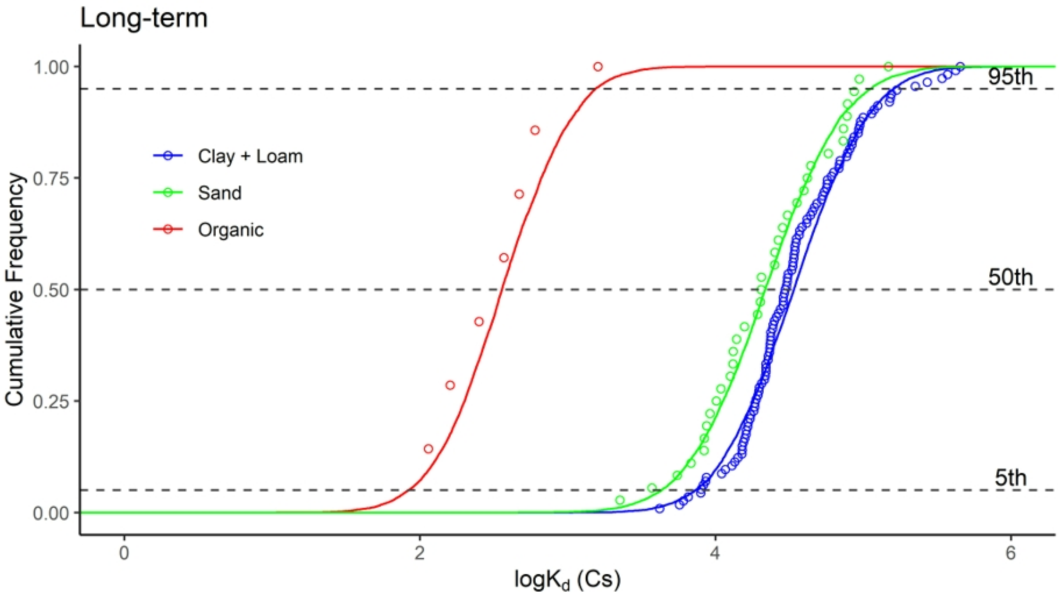
N = number of observations, GM = geometric mean, GSD = geometric standard deviation
[a] Different letters among the datasets compared indicate statistically significant differences between GMs according to the Fisher's Least Significant Differences test.



Partial dataset	N	GM	GSD	FLSD[a]	5th	95th
$RIP/K_{ss} < 10^2$	74	6.9×10^1	2.7	A	2.6×10^1	5.7×10^2
$10^2 \leq RIP/K_{ss} < 10^3$	116	3.8×10^2	4.5	B	4.2×10^1	4.4×10^3
$10^3 \leq RIP/K_{ss} < 10^4$	83	1.6×10^3	3.4	C	2.9×10^2	1.9×10^4
$RIP/K_{ss} \geq 10^4$	55	1.0×10^4	4.1	D	1.2×10^3	9.5×10^4

N = number of observations, GM = geometric mean, GSD = geometric standard deviation
[a] Different letters among the datasets compared indicate statistically significant differences between GMs according to the Fisher's Least Significant Differences test.





Partial dataset	N	GM	GSD	FLSD[a]	5th	95th
Short-term						
Mineral (OM < 50%)	367	2.5×10^3	4.5	A[b]	1.2×10^2	2.4×10^4
Clay+Loam	227	3.9×10^3	3.3	B[b]	5.9×10^2	2.6×10^4
Sand	112	1.4×10^3	5.2	C[b]	5.6×10^1	1.1×10^4
Organic (OM ≥ 50%)	38	8.9×10^1	4.2	D[b]	1.3×10^1	1.9×10^3
Long-term						
Mineral (OM < 90%)	161	2.5×10^4	3.2	A[c]	4.2×10^3	1.5×10^5
Clay+Loam	114	3.0×10^4	2.6	B[c]	8.0×10^3	2.2×10^5
Sand	36	2.0×10^4	2.7	A[c]	3.7×10^3	9.4×10^4
Organic (OM ≥ 90%)	7	3.7×10^2	2.4	C[c]	1.1×10^2	1.6×10^3

N = number of observations, GM = geometric mean, GSD = geometric standard deviation
[a] Different letters among the datasets compared indicate statistically significant differences between GMs according to the Fisher's Least Significant Differences test. Dataset comparisons shown here are: [b] short-term mineral, textural and organic datasets; [c] long-term mineral, textural and organic datasets.

SUPPLEMENTARY MATERIAL

Deriving probabilistic soil distribution coefficients (K_d). Part 2: Reducing caesium K_d uncertainty by accounting for experimental approach and soil properties

Table S1

Examination of the effect of the OM threshold in deriving Mineral and Organic datasets.

OM Threshold (%)	Mineral				Organic			
	N	GM	GSD	FLSD[a]	N	GM	GSD	FSLD[a]
Short term								
20	345	2.7×10^3	4.3	A[b]	60	1.8×10^2	6.1	A[d]
30	355	2.4×10^3	4.3	A[b]	50	1.5×10^2	5.8	AB[d]
40	360	2.3×10^3	4.5	A[b]	45	1.2×10^2	4.6	AB[d]
50	367	2.3×10^3	4.5	A[b]	38	8.9×10^1	4.2	B[d]
Long term								
20	148	2.8×10^4	2.6	A[c]	20	2.0×10^3	7.4	A[e]
60	150	3.1×10^4	2.6	A[c]	18	1.2×10^3	4.9	AB[e]
70	153	2.9×10^4	3.0	A[c]	15	1.2×10^3	4.6	AB[e]
80	154	2.9×10^4	3.0	A[c]	14	1.1×10^3	4.5	AB[e]
90	161	2.5×10^4	3.2	A[c]	7	3.7×10^2	2.4	B[e]

N = number of observations, GM = geometric mean, GSD = geometric standard deviation

[a] Different letter among the datasets compared indicate statistically significant differences according to the Fisher's Least Significant Differences test. Dataset comparisons shown here are: [b] short-term mineral datasets; [c] long-term mineral datasets; [d] short-term organic datasets; [e] long-term organic datasets.

Figure S1

Plots of $\log K_d$ vs. $\log (RIP/K_{ss})$ and $\log K_d$ vs. $\log K_{d,pred}$ ($K_{d,pred}$: calculated from equation [3]).