1 **Deriving probabilistic soil distribution coefficients (K_d). Part 1: General**

2 **approach to decreasing and describing variability and example using**

3 **uranium K_d values**

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

5 *[a] Chemical Engineering and Analytical Chemistry Department, Faculty of Chemistry,*

6 *University of Barcelona, Martí i Franqués 1-11, 08028, Barcelona, Spain.*

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

8

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

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

11

12

13

14

15

16

1   **Deriving probabilistic soil distribution coefficients (K_d). Part 1: General**

2   **approach to decreasing and describing variability and example using**

3   **uranium K_d values**

4

5   **Abstract**

6   A general approach is presented to derive probabilistic radionuclide distribution coefficients

7   (K_d) in soils from a K_d dataset. The main aim was to derive informed estimates with a low

8   inherent uncertainty by restricting the K_d value data to subsets based on key soil factors and

9   the experimental approach used to calculate the K_d value (e.g., sorption and desorption tests).

10   As an example, the general approach was applied to uranium (U) K_d values that are part of a

11   critically reviewed dataset containing more than 5000 soil K_d entries for 83 elements and an

12   additional 2000 entries of K_d data for 75 elements gathered from a selection of other, non-soil,

13   geological materials. The overall soil U K_d dataset included 196 values spanning a range of

14   four orders of magnitude (1 to 67,000 L kg^{-1}), with additional 50 entries for other geological

15   materials. Whereas the effect of the experimental approach could be disregarded, major

16   factors in decreasing U K_d variability were pH and organic matter concentration (OM).

17   Limitation in the number of entries made it difficult to use texture information (sand, silt,

18   clay) to further decrease U K_d variability. The integrated combination of pH+OM permitted

19   some soil groups to have U K_d confidence intervals as narrow as two orders of magnitude.

20   Specifically for U K_d, data in the Mineral (<20% OM) and Organic (≥20% OM) partial

21   datasets were significantly different. Analogue data from geological materials other than soils,

22   such as subsoil, till and gyttja (a lacustrine mud containing elevated organic matter (OM)

23   concentrations), were also statistically evaluated to determine whether they could be used to

24   fill U K_d data gaps. It was shown that U K_d from subsoils and tills, but not gyttjas, could be

25   used to enhance soil U K_d datasets. Selection of probabilistic K_d values for risk modelling can

26   be made more reliably and with less uncertainty by using appropriate geochemical data

27   representative of the study site to narrow the wide range of potential K_d values.

28

1. Introduction

The partitioning of radionuclides (RN) between solid and liquid phases largely affects the RN mobility and bioavailability in soils. RN partitioning is often quantified by the solid-liquid distribution coefficient (K_d), defined as the ratio of activity concentration of RN sorbed on a specified solid phase (C_{solid}) to the RN activity concentration in a specified liquid phase (C_{liquid}) in equilibrium with the former (ICRU, 2001). Despite that the K_d model is based on several theoretical assumptions, which are not always fulfilled (EPA, 1999a), its use constitutes the simplest sorption model available. Risk assessment models use soil K_d values to derive RN concentration in solution from the concentrations of RNs in contaminated soils to estimate: the transfer of the bioavailable pool of RNs between soil layers and to plants, the leaching of RNs from the surface of contaminated soils and further transport in the unsaturated zone (vadose zone) and saturated zone (groundwater), and the diffusion-driven transport of RN through soils (Simon-Cornu et al., 2015; Sy et al., 2016).

The K_d parameter alone does not give information about the sorption mechanisms governing the phase partitioning of the target species. Reported K_d values of RNs can vary greatly due to the experimental methods used for their quantification. Among the methods used to measure K_d values are: 1) in situ measurements of RN concentrations in paired samples of soil and porewater collected from the field; 2) batch uptake tests with dispersed samples and RN-spiked aqueous phase; 3) desorption tests with RN-contaminated soils; and 4) dynamic column flow experiments or dynamic tank-leaching tests with compacted samples (OECD, 2000; ASTM, 2003; Gil-García et al., 2008; ASTM, 2010; Aldaba et al., 2010; EPA, 2017). To complicate matters further, several experimental parameters also influence the measured K_d values, including contact time of the RN and soil, the soil:solution ratio, and solution chemical characteristics that is in contact with the soil during the uptake or desorption experiments.

To make the K_d data derived from compilations suitable for risk assessment purposes, it must be ensured that they were derived from experiments using similar solids, liquids, and RN species as exist at the contamination scenario under study. This requires identifying the factors controlling the RN-soil interaction to explain, and when possible to reduce, the K_d variability. Ideally, risk assessments models use site-specific K_d values, however such data are often not available. To overcome this limitation, probabilistic models are a useful option (Degryse et al., 2009). Probabilistic models use K_d compilations to derive, among other parameters, the most likely/best-estimate K_d values, statistical functions describing the overall variability of K_d values, and to derive confidence intervals at specified significance levels

66   (Sheppard and Thibault, 1990; Stenhouse, 1994; EPA, 2001; Simon-Cornu et al., 2015).

67   However, the derived statistics from these K_d compilations are often suitable only for

68   screening purposes because the variability is high. One approach to reducing this variability is

69   to create K_d data subsets for different soil types based on the soil characteristics governing the

70   target RN sorption in soils, if this information is available (IAEA, 2010) and describing K_d

71   variability by using distribution functions, so the input data in risk assessment models are

72   entered with their variability thus contributing to the final uncertainty derived from the

73   application of the model.

74   The International Atomic Energy Agency (IAEA) launched the Environmental

75   Modelling for Radiation Safety Programme (EMRAS) (2003-2007), whose main purpose was

76   to identify and reduce uncertainties in the predictive capability of environmental models. The

77   EMRAS Programme produced the Technical Report Series 472 'Handbook of Parameter

78   Values for the Prediction of Radionuclide Transfer in Terrestrial and Freshwater

79   Environments' (IAEA, 2010), and a related TECDOC 1616 document entitled, 'Quantification

80   of radionuclide transfer in terrestrial and freshwater environments for radiological

81   assessments' (IAEA, 2009). These documents summarized a new soil K_d data compilation

82   that included approximately 2900 records for 67 elements and were a major update of the

83   previous K_d compilation (IAEA, 1994).

84   One of the main outcomes resulting from the analyses of this compilation was the

85   proposal of K_d best estimate values (as a geometric mean value) and their related variability

86   (expressed as a geometric standard deviation value, and with the minimum-maximum K_d

87   values range) (Gil-García et al., 2009a; Gil-García et al., 2009b; Vandenhove et al., 2009).

88   Partial datasets were constructed based mainly on soil groups according to the organic matter

89   (OM), and clay and sand (textural) contents of the mineral soil samples, the so-called

90   OM+Texture criterion. This approach was uniformly applied to all RNs with sufficient data to

91   permit such analysis. But even after following this approach, it was shown that the K_d data

92   could still vary more than 4 orders of magnitude within one OM+Texture soil group, thus

93   compromising the usefulness and reliability of the derived best-estimate values.

94   The Modelling and Data for Radiological Impact Assessments (MODARIA)

95   Programme was auspicied by IAEA during the period 2012 and 2015, which was followed by

96   the MODARIA II Programme (Development, Testing and Harmonization of Models and Data

97   for Radiological Impact Assessment; 2016-2019). Working groups were established whose

98   main goal was to analyse radiological data aiming at, among others, identifying the data gaps

99   of those parameters relevant for the radiological assessments; filling the priority data gaps by

100   updating previous parameter data compilations with recently available data and exploring the

101   viability of applying approaches based on the extrapolation of the current existing data to

102   foresee other scenarios lacking of data, *i.e.*, analogue approach; and developing strategies for

103   handling of variability of parameter values to provide end-users with less uncertain data for

104   radiological assessment purposes. Specifically, an aim of the working groups was to update

105   the available soil RN K_d dataset, as well as to develop a strategy to reduce and describe K_d

106   variability based on probabilistic models, including the construction of distribution functions

107   to statistically describe the K_d values of a target RN.

108   This is the first in a series of three papers, in which we describe and illustrate the

109   approach followed to derive best-estimate K_d values and K_d statistical distributions from an

110   updated and critically reviewed compilations of RN K_d values (Ramírez-Guinart et al. 2020a;

111   Ramírez-Guinart et al. 2020b). In this first paper the approach to reducing the distribution of

112   K_d values is described and illustrated for uranium K_d values. Parts two and three in this series

113   of papers show the application of the approach to radiocaesium and americium K_d values,

114   respectively. In all three papers, the most suitable criteria for reducing uncertainty, such as

115   soil characteristics (*e.g.*, pH, texture, organic matter content, etc.) and factors related to the

116   experimental approach applied (*e.g.*, sorption *vs.* desorption test, or effect of long-term

117   interaction), are evaluated; the cumulative distribution functions of K_d values for the

118   optimized groupings are proposed; and the viability of using K_d data gathered from geological

119   materials other than soils (material analogue approach) and/or for elements chemically similar

120   to the target element (chemical analogue approach) to enhance datasets containing insufficient

121   entries of the target element are assessed.

122

123

124   **2. Data collection and treatment**

125   *2.1. Soil RN K_d compilation update: revision of data acceptance criteria*

126   The former TRS-472 soil K_d data compilation was critically reviewed and updated, by

127   examining more than 100 new documents (reviewed papers and grey literature). The K_d

128   values available, along with ancillary information regarding the method used and soil

129   characteristics, such as soil pH, organic matter content, clay and silt contents in the mineral

130   fraction and cationinc exchange capacity, were incorporated into the K_d compilation.

131   K_d values not directly quantified as the ratio between concentrations of the target

132   element measured in a liquid and a solid phase were rejected. Consequently, the current

133   compilation does not include data: 1) indirectly derived from parametric equations, 2)

134 indirectly deduced from mass-transport experiments, such as column tests or diffusion

135 experiments, or 3) pooled or averaged K_d values from former K_d data compilations.

136 Furthermore, K_d data originated from the same element-sample combination by applying a set

137 of tests at varying operational variables related to the experimental approach, but not

138 considered as a relevant factor for grouping K_d values (*e.g.*, contact time or solid-to-liquid

139 ratio in batch experiments) were pooled to become a single entry, calculated as a geometric

140 mean (GM). Instead, they were entered into the data base as individual entries if the variable

141 tested could be considered as a relevant factor (*e.g.*, experiments at varying pH). K_d data were

142 only accepted if they were obtained under experimental conditions representative of typically

143 environmental conditions. For example, K_d measurements measured under extreme chemical

144 conditions (< pH 2 or > pH 12) were not included. K_d data for stable isotopes were accepted

145 so long as they were obtained using isotope concentrations within the linear sorption range. In

146 agreement with the previous compilation (IAEA, 2009), K_d data from pure mineral phases,

147 such as clay minerals or metal (hydro)oxides, were not accepted. The K_d compilation also

148 included data from subsoils, surface sediments (formed by deposition of transported organic

149 matter or mineral soils by the action of wind, water, or ice, and/or by the force of gravity

150 acting on the particles), gyttja (which refer to a mud formed from organic and mineral matter,

151 which can be found at the bottom or near the shore of certain lakes) and till materials (which

152 are unsorted and non-stratified materials deposited directly by glacial ice, which consist of a

153 mixture of clay, silt, sand, gravel, stones, and boulders in any proportion).

154

155 *2.2. Structure of the updated soil RN K_d compilation: creation of partial datasets*

156   The current soil RN K_d compilation is a spreadsheet-type document in which K_d and

157 ancillary data are organized based on elements in independent sheets. It contains more than

158 5000 entries of soil K_d for 83 elements, and additional more than 2000 entries of K_d data for

159 75 elements gathered from a selection of other solid materials. Besides fields related to the

160 sources of information, radioisotopes, soil characteristics and ancillary information, new

161 fields related to the experimental approach and type of material were added. The category of

162 'type of material' is used to distinguish between K_d data for soils (the unconsolidated

163 geological material comprising the terrestrial root zone) versus those from other types of

164 geological materials, such as sediment, subsoils, till and gyttja. Various categories were

165 created to classify the experimental approaches (see Table 1: short or long-term experiments;

166 in situ, batch sorption or batch desorption tests) that permit selecting K_d values for specific

167 transport scenarios. Either for the overall dataset, or from partial datasets based on the

168   experimental approach, grouping criteria based on soil factors were used to create further

169   partial datasets from which were derived K_d best estimate values and related distribution

170   function with a lower variability obtained.

171   A strategy was then developed and applied to create partial datasets with less

172   variability, based on grouping the K_d data of a given element according to optimized criteria

173   that consider factors related to the soil characteristics and the experimental approach applied,

174   relevant for the RN-soil interaction. The aim was to identify the soil properties that reduced

175   RN K_d variability, especially those routinely analysed (and, thus, more likely to have

176   available information); to establish a series of criteria to properly create K_d partial datasets;

177   and to construct cumulative distribution functions (CDFs) or, at least, to derive best-estimate

178   values for each group.

179   The OM content was the first soil property to be tested (the OM criterion), and the

180   effect of the soil texture for mineral soils was also tested when relevant (the OM+Texture

181   criterion). Thus, four soil groups were created based on the texture (Sand, Loam, and Clay

182   soils) and organic matter content (Organic soils). Initially, a soil was included in the Organic

183   group if its OM content was >20%. For the mineral soils, three groups were created according

184   to the following criteria (% refers to mineral matter): Sand group: sand fraction $\geq$ 65%; clay

185   fraction < 18%; Clay group: clay fraction > 35%; and Loam group (IAEA, 2010). Additional

186   soil factors were also checked for specific radionuclides. For instance, the pH and hierarchical

187   application of various soil factors, such as pH and OM, were tested as described later in

188   Section 3. Finally, statistical tests (Fisher's least significant differences (FLSD) test for

189   multiple samples; 95% confidence level; StatGraphics 18) were performed to check whether

190   the geometric mean K_d values for the partial datasets were significantly ($p \leq 0.05$) different.

191

192   *2.3. Analysis of the influence of the experimental approach on RN K_d data variability*

193   The influence of the experimental approach was simultaneously evaluated along with

194   relevant soil factors. Before applying the grouping criteria based on the experimental

195   approach factor, a data treatment based on group mean centering (GMC) was carried out to

196   minimize the effect of soil factors identified as relevant to the interaction of the target

197   element. Figure 1 illustrates this approach for the pH+OM criterion as the key soil-factor for

198   the GMC treatment for U K_d values. Other soil factors may be applicable to other

199   radionuclides. The GMC consisted in log-transforming the overall dataset of the target

200   element, grouping the K_d data according to the soil criteria previously established based on a

201   key soil factor, calculating the arithmetic mean (AM) of log K_d values of each soil-type group

202   created and correcting each single log K_d value within a given group by subtracting the AM

203   log K_d value of the respective soil-type group. Secondly, the GMC-corrected log K_d datasets

204   were splitted in subsets according to the type of the experimental approach, that is, firstly

205   long- versus short-term experiments, and then a further splitting was carried out based on in

206   situ, sorption and/or desorption experiments. No in situ data were available in the current

207   datasets and, thus, it was only necessary to split the short-term partial datasets into sorption

208   and desorption categories. Finally, statistical tests (Fisher's Least Significant Differences

209   (FLSD) test for multiple samples; 95% confidence level; StatGraphics 18) were performed to

210   check whether the K_d geometric means for each partial datasets were significantly different.

211

212   *2.4. Construction of Cumulative Distribution Functions (CDF) to describe K_d variability*

213   A suitable method to describe the population and variability of K_d values for a certain RN-

214   material combination is the construction of Cumulative Distribution Functions (CDF). In

215   short, CDFs are equations describing the different values of a real-valued variable, in this case

216   K_d, and their related accumulated frequency, which means that the probability that K_d takes a

217   value less than or equal to a certain value can be explained in a continuous form with a

218   function. CDFs are built with the statistical parameters (*e.g.*, arithmetical mean, geometric

219   mean, mode, variance, etc.) of the underlying frequency distribution (*e.g.*, normal, lognormal,

220   exponential, etc.) (Ciffroy et al., 2009). Thus, in order to construct reliable CDFs it is

221   necessary to unequivocally ascertain the statistical distribution describing the data population,

222   as well as to properly derive the corresponding statistical parameters.

223   Since the K_d parameter is a ratio of concentrations, K_d data are expected to follow a

224   lognormal distribution (Sheppard et al., 2011). If a K_d dataset is lognormally distributed, then

225   it inherently implies that the log-transformed K_d values of this dataset (log K_d) follows a

226   normal distribution. The statistical parameters describing a symmetrical log K_d distribution

227   are: the location parameter (μ), which can be considered as the most probable log K_d value

228   since it corresponds to the 50th percentile of the log K_d distribution, and the scale parameter

229   (σ), which gives an estimation of the dispersion among log K_d values. Both statistical

230   parameters can be determined either by calculating the arithmetic mean (AM) and standard

231   deviation (SD), respectively, of the experimental log K_d values comprising a given dataset or

232   by fitting the experimental cumulative distribution of the log K_d dataset to the theoretical CDF

233   equation of the normal distribution. Figure S1 in the Supplementary Material illustrates the

234   graphical representation of the CDF of an ideal lognormal dataset that contains K_d values

235   within the 1 to 10^6 L kg^{-1} range. CDFs also informs of all the potential values and their

236   probability of occurrence (Ciffroy et al., 2009). Thus, confidence intervals of K_d values can

237   also be established from a CDF constructed with a K_d dataset by calculating the

238   corresponding percentile ranges (*e.g.*, the 90% and 95% confidence intervals correspond to

239   the 5th - 95th and 2.5th - 97.5th percentile ranges, respectively).

240   For the construction of CDFs (see Figure S2 of the Supplementary Material), K_d data

241   were log-transformed and the presence of possible outlier values in the datasets was examined

242   by performing an exploratory analysis based on box-and-whisker plots. A threshold of three

243   times the interquartile range was established to identify if a log K_d values was an outlier. The

244   log K_d data within every dataset were sorted by increasing value and an empirical frequency

245   ($f_{exp,i}$) equal to 1/N (where N is the total number of K_d entries in the respective dataset) was

246   assigned to each entry. Experimental cumulative frequency distribution were constructed by

247   assigning to each sorted log K_d value their corresponding cumulative frequency ($F_{exp,i}$), *i.e.*,

248   the sum of the preceding frequencies $\left(F(K_{d,j}) = \sum_{i=0}^{j} f(K_{d,i})\right)$. Following this, the

249   Kolmogorov-Smirnov test was applied to test if the experimental cumulative distribution

250   function did not statistically differ from the theoretical cumulative frequency distribution

251   constructed by assuming a lognormal distribution. As expected, it was found that all the K_d

252   datasets analysed followed a lognormal distribution. Consequently, the experimental

253   cumulative frequency distributions constructed with the log K_d data were fitted to the

254   theoretical normal CDF equation (Eq. 1) and the location and scale parameters (μ and σ,

255   respectively) were derived to construct CDFs.

256

257    $P\left(\log K_{d,i} \leq \log K_{d,j}\right) = \sum_{\log K_{d,i} \leq \log K_{d,j}} p\left(\log K_{d,i}\right) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{\log(K_{d,i}) - \mu}{\sigma\sqrt{2}}\right); K_{d,i} > 0.$ [1]

258

259   where P is the cumulative probability, erf is the error function, and subscripts *i* and *j* represent

260   two different K_d values in the ranked grouping.

261

262   Since the statistical parameters determined are in log-scale, the derived best-estimate value

263   and confidence range are also in log-scale. Thus, the corresponding antilog parameters were

264   calculated, that is, the geometric mean (GM) and geometric standard deviation (GSD), as

265   shown in Eq. (2) and Eq. (3):

266    $\mathrm{GM}\,(K_d) = \left(\prod_{i=1}^{N} K_{di}\right)^{\frac{1}{N}} = 10^{\left(\frac{1}{N}\sum_{i=1}^{N}\log(K_{di})\right)} = 10^{\mu} = \mathrm{antilog}\,(\mu)$ [2]

267 $$\text{GSD } (K_d) = 10^{\left(\sqrt{\dfrac{\Sigma_{i=1}^{N}(\log \frac{K_{d_i}}{GM})^2}{n}}\right)} = 10^{\sigma} = \text{antilog }(\sigma) \qquad [3]$$

268

269 To properly derive a reliable CDF from a given K_d dataset it is necessary that it contains a

270 minimum number of entries (around $N > 10$). Exceptionally, from those datasets containing a

271 lower number of entries, CDFs were constructed as far as the datasets presented clear

272 experimental cumulative distributions and a good fitting of data could be done. For the rest of

273 cases only GM values were calculated directly from the dataset.

274

275 **3. Example of the application of the general approach for the data analyses using**

276 **uranium K_d values**

277 *3.1. Soil factors governing U sorption: derived soil criteria to group U K_d data*

278 Uranium sorption in soils is known to be very complex and affected by several soil properties

279 and phases, mainly pH, soil texture, specific surface area (SSA), cation exchange capacity

280 (CEC), dissolved carbonate, amorphous iron oxides (Fe_{amorph}) and OM matter contents (EPA,

281 1999b; Payne et al., 2011). Uranium oxidation state or the redox status of the experimental

282 systems are also extremely important factors influencing U K_d values. U(IV) binds

283 appreciable more strongly and has a much lower solubility than U(VI) (Langmuir, 1978;

284 Vandenhove et al., 2010), which is expected to be the dominant U species in most topsoils. At

285 pH below 5, U(VI) is present as the uncomplexed uranyl ion, UO_2^{2+}. At a higher pH, the

286 uranyl ion hydrolyses, forming aqueous hydroxide complexes, which dominate U(VI)

287 speciation in the absence of dissolved inorganic ligands (*e.g.*, carbonate, sulphate or

288 phosphate). At the pH range of 5–10, highly soluble carbonate complexes dominate the U

289 speciation (Vandenhove et al., 2010; Li et al. 2014). Few studies included in the U K_d dataset

290 provided information about the specific U species in their system and thus this information

291 could not be incorporated as a significant field of the dataset or be used for data treatment.

292 In general, the sorption of U by soils is low at pH values less than 3, increases rapidly with

293 increasing pH from 3 to 5, reaches a maximum in the pH range from 5 to 7, and then

294 decreases with increasing pH values greater than 7 (EPA, 1999b). Thus, U sorption in soils

295 frequently shows an inverted-U shaped trend in relation to the pH. The increase in sorption

296 between pH 3 and 5 has been attributed to increased number of available sorption sites at clay

297 minerals, while the decrease sorption above pH 7 has been attributed to the dominance of

298 U(VI)-carbonate complexes that have weak affinities for mineral surfaces (EPA, 1999b).

299   Organic matter and clay minerals provide exchange sites that are expected to increase

300   sorption of UO_2^{2+} and other positively-charged U species (Kaplan and Serkiz 2001; Li et al.

301   2014; Li et al. 2015). The influence of organic matter on U interaction is twofold: an

302   increased sorption through exchange mechanisms on soil-bound OM and a decreased sorption

303   due to formation of soluble organic complexes, for those samples having large amounts of

304   dissolved organic matter content and colloids. U(VI) was found to be strongly retained by

305   organic aggregates and organic coatings on quartz minerals (Crançon and van der Lee, 2003),

306   whereas a large fraction of U(VI) was also found to be associated to humic colloids in soil,

307   thus forming a potential mobile uranium phase. Other studies also highlight the importance of

308   iron oxides/hydroxides for the sorption of U (Hsi and Langmuir, 1985; Waite et al., 1994;

309   Duff and Amrhein, 1996; Payne et al., 1996; Kaplan et al. 2016). The positively charged U-

310   species are sorbed to the negatively charged surfaces of the sesquioxides or U-species become

311   structurally incorporated in the iron-oxides (coatings) over the many dissolution–precipitation

312   cycles of these amorphous or poorly crystalline iron oxides (Sowder et al., 2003). The soil

313   specific surface area (SSA; units of m^2 kg^{-1}) is also known to play a significant role in

314   uranium sorption, as shown for pure mineral phases (Payne et al., 2011). Soils containing

315   large fractions of amorphous or clay minerals are expected to have very large specific surface

316   area values. Unfortunately, there were an insufficient number of U K_d experiments in the

317   literature that included specific surface area soil measurements, consequently it was not

318   possible to assess soil surface area as an independent variable for predicting U K_d values.

319   Due to the varying soil properties affecting U sorption, it is difficult to use

320   geochemical modelling or parametric equations to predict U K_d based on soil characteristics,

321   although there have been recent advances to notice. The use of the smart-K_d concept offers

322   new possibilities for the modelling of K_d for elements with complex speciation system as U,

323   as the coupling of a surface complexation model with a geochemical speciation code permits

324   to calculate more realistic distribution coefficients as a function of varying environmental

325   conditions (Stochmann et al., 2017). Using regression statistics, Echevarria et al. (2001)

326   explored the effect of soil characteristics on U sorption for a reduced soil dataset and deduced

327   a linear relationship between soil U K_d and pH that evidenced that U K_d values decreased

328   when increasing pH in the 5.5–8.8 pH range. A similar pattern was also observed by

329   Vandenhove et al. (2007) for a controlled dataset comprising soils with pH $\geq$ 6, which was

330   explained by the increased amount of soluble uranyl–carbonate complexes at increasing basic

331   pH. However, Sheppard et al. (2006) found that, when considering heterogeneous dataset

332   composed by soils with pH ranging from 5.5 to 8.8 and data from varying sources, only

333   relative low percentage of U K_d variance could be explained by the U K_d vs. pH correlations

334   obtained, indicating that the U K_d cannot be univariantly predicted from pH variation.

335   Sheppard (2011) recently suggested updated K_d *vs.* pH and clay correlations, based on a

336   dataset enriched with a significant contribution of field data with indigenous uranium.

337   The use of soil carbonate content, or more specifically the CO_3^{2-} and HCO_3^-

338   concentration in soil solution, and the iron hydro(oxides) content or soil specific surface area

339   to group U K_d values, was not considered as this information usually is not available in the

340   current compilation. Instead, U K_d data was initially grouped according to the OM criterion to

341   distinguish between mineral and organic soils, as this approach indirectly includes the role of

342   not only organic matter but also specific surface area. To complete the examination of the

343   effect of specific surface area, the mineral dataset was splitted when possible into the different

344   textural groups. As pH strongly affects U species present in the soil solution, the U K_d overall

345   dataset was also split into three pH groups (pH < 5; $5 \leq$ pH < 7; and $7 \leq$ pH < 9) defined

346   according to the U speciation (Vandenhove et al., 2009), which defines the pH criterion.

347   Finally, it was evaluated for the first time the hierarchical application of soil factors to group

348   U K_d data (pH, pH+OM; pH+OM+Texture). Thus, the U K_d groups created when the U

349   overall dataset was firstly split according to the pH criterion were further split according to

350   the OM content to create Mineral and Organic groups (pH+OM criterion). In a further step, U

351   K_d data in the pH-Mineral groups were also split according to the soil texture into the three

352   textural groups (pH+OM+Texture criterion).

353

354   *3.2. Influence of experimental approach on U K_d data*

355   The overall U K_d dataset contained 196 entries with U K_d values varying more than 4 orders

356   of magnitude. The overall GM value was 320 L kg^{-1} (see Figure 2), in general agreement with

357   previous reported values (Sheppard et al., 2006; Vandenhove et al., 2009). However, the GM

358   or CDF that could be derived from the overall dataset are affected by a more than 4-orders of

359   magnitude uncertainty (the 5th-95th percentile range was 3.5×10^0-9.9×10^3 L kg^{-1}) and they

360   may be not suitable for risk assessment in specific scenarios.

361   The overall U dataset contained entries of short-term sorption (ST-S), short-term

362   desorption (ST-D) and long-term desorption (LT-D) experimental approaches, as defined in

363   Table 1. The GMC data treatment to examine the effect of the experimental approach on U K_d

364   variability considered the combination of soil factors most relevant to U sorption, that is, pH

365   and OM. The FLSD statistical analyses performed are summarised in Table S1 of the

366   Supplementary Material. When splitting the datasets according to the partial pH+OM groups,

367 no effect of the experimental approach was observed for U K_d for data obtained neither at pH

368 < 5 nor at pH ≥ 7, and only a few non-systematic discrepancies were noticed for the 5 ≤ pH <

369 7 dataset. Therefore, as the possible effect of the experimental methology was not

370 unequivocally proven by the statistical analyses of the partial datasets, U K_d data were not

371 seggregated based on experimental approach in the subsequent analyses.

372

373 *3.3. U K_d best estimates and CDFs based on the OM and OM+Texture criteria*

374 The current U dataset when refined for applying the OM criterion contained 153 entries

375 varying in the same range of U K_d values as the overall dataset. Figure 2 displays the data

376 obtained when applying the OM criterion as well as the graphical representation of the CDFs

377 constructed from the overall and partial datasets.

378   The U K_d data in the Mineral and Organic partial datasets were significantly different

379 (Figure 2). The GM of the Organic group was around one order of magnitude greater than that

380 of the Mineral group, which confirmed the role observed in previous studies regarding the

381 enhancement of the uranium sorption due to its interaction with the soil organic matter

382 (Crançon and van der Lee, 2003). The application of the OM criterion allowed proposing

383 CDFs for the Mineral and Organic groups, which comprised K_d values varying within a

384 narrower range than in the case of the CDF constructed from the overall dataset, especially for

385 the Organic soils group. Despite this improvement, the usefulness of these CDFs for

386 performing radiological assessments may still be limited since they are constructed from

387 datasets with large variability.

388   The Mineral soils were splitted according to the soil texture and no significant

389 differences were observed in U K_d data among textural groups (Clay: GM=96 L kg^{-1}, N=8;

390 Loam: GM=258 L kg^{-1}, N=62; and Sand=204 L kg^{-1}, N=28), which did not follow a sequence

391 consistent with expected effects based on specific surface area. Thus, the OM+Texture

392 criterion to group soils did not add further relevant information than the simpler OM criterion.

393 Such results indicate that soil texture alone is not a soil factor relevant enough to control the

394 U-soil interaction and, thus, it is unnecessary to distinguish between textural soil-types for

395 mineral soils without further factors being preliminarily considered.

396   Since it was corroborated that soil OM content has a strong influence on the U

397 sorption behaviour, an exercise to optimise the threshold of OM content to consider a given

398 soil as "organic" was done. The aim of this exercise was to explore the possibility of

399 obtaining partial datasets with less variability, allowing a better distinction of those soils in

400 which the OM fraction really controls the U-soil interaction and, in turn, to succeed in

401 distinguishing U K_d data of mineral soils among textural groups. The possibility was explored

402 to decrease the initial OM threshold from 20% down to 5%, by examining the new partial

403 datasets based on grouping values according to 5% and 20% OM thresholds. Results,

404 displayed in Table S2 of the Supplementary Material, showed that by lowering the threshold

405 to 5% OM there was no improvement neither in terms of decreasing the U K_d variability nor

406 deriving statistically different GMs among the new mineral and organic datasets. Therefore,

407 concerning the interaction of U in soils, the initial 20% OM threshold was maintained to

408 consider whether a soil is grouped as organic or as mineral.

409

410 *3.4. U K_d best estimates and CDFs based on soil factors related to U sorption mechanisms.*

411 *3.4.1. The pH criterion*

412 U K_d values were initially segregated into three pH categories (pH < 5; 5 $\leq$ pH < 7; and pH $\geq$

413 7) based on U speciation. Figure 3 presents the U K_d data and CDF derived by applying the

414 pH criterion, including the CDF of the overall dataset. The parameters derived from the pH

415 partial datasets were significantly different among them and contained U K_d values varying

416 less than in the overall dataset, as the 5th-95th interval of U K_d values decreased down to less

417 than three orders of magnitude. The derived U K_d best estimates and related confidence

418 ranges indicated a significant increase in U K_d values when increasing pH, reaching a

419 maximum within the 5-7 pH range, whereas U K_d values decreased at higher pH values. This

420 pH-dependent U sorption agreed with previous reports (Dalvi et al., 2014; Payne et al., 2011;

421 Vandenhove et al, 2009). At low pH values (pH < 5) the U sorption in soils is decreased due

422 to the competition of the uranyl cation with protonated sites whereas at pH > 7 the decrease in

423 U sorption is caused by the formation of stable, weakly sorbing uranyl carbonate complexes.

424 The 5th-95th percentile CDF region of the 5 $\leq$ pH < 7 and pH $\geq$ 7 partial datasets partially

425 overlapped. This fact, in conjunction with the large U K_d variability still existing within the

426 pH groups from which the CDF were constructed, suggested that it was still necessary to seek

427 a more meaningful U K_d grouping by means of taking into a simulateneous consideration of

428 several soil factors.

429

430 *3.4.2. Hierarchical application of pH and OM soil factors*

431 The pH+OM criterion

432 Figure 4 summarizes the U K_d data obtained from the datasets created based on the

433 hierarchical application of the pH and OM soil factors (the pH+OM criterion) and depicts the

434 related CDFs. The simultaneous use of pH and OM factors led to the creation of partial

435   datasets with a much lower variability than the pH partial datasets. Besides, they followed the

436   same pH dependence as the pH-based partial datasets. Thus, for mineral soils the U K_d values

437   for the intermediate pH mineral group ($5 \leq pH < 7$) were statistically the highest, whereas for

438   organic soils, as the $pH \geq 7$ data set has only 2 entries, the only significant comparison

439   showed that the GM value of the intermediate group was also higher than that of the $pH < 5$

440   organic dataset. For a given pH group, the FLSD test confirmed that organic soils always had

441   greater GMs and the 5th-95th percentile ranges were shifted to greater values than those of the

442   respective mineral group. These trends denote the suitability of using two soil factors for U K_d

443   grouping, as both are relevant and govern uranium soil interactions. The lack of U K_d values

444   for organic soils at $pH \geq 7$ prevented the construction of the related CDF. The scarcity of such

445   data is not surprising as most organic soils have acid-to-neutral pH values. Therefore, this soil

446   group has a reduced representativity of natural environmental conditions.

447

448   <u>The pH+OM+Texture criterion</u>

449   A final strategy to refine the U K_d data of the pH-Mineral groups was tested by splitting them

450   into textural groups according to the clay and sand content of the soils, in order to check

451   whether the absence of texture effect on U K_d variability observed in section 3.3 was due to

452   not consider a previous grouping of soils by pH. As can be seen in Table S3 in the

453   Supplementary Material, most of the derived textural groups had scarce data ($N < 10$), which

454   hindered a proper evaluation of the role of introducing textural information to further decrease

455   U K_d variability. As an improvement from the previous analyses, the GM of the textural

456   classes now followed the sequence based on specific surface area ($GM_{Clay} > GM_{Loam} >$

457   GM_{Sand}). Thus, the possibility of using U K_d data from solid environmental materials other

458   than soils to enhance the textural datasets was explored.

459   The analogy between soils and other solid environmental materials (till, subsoil and

460   gyttja) with respect to their capacity to sorb U was evaluated. These materials were mostly

461   mineral, although a few entries of organic gyttja materials were also available. For each

462   material, all U K_d data available were pooled in an overall dataset and each material and

463   partial dataset were initially created according to the pH criterion. The U K_d statistics derived

464   from each dataset are summarised in Table 2. U K_d data for gyttja materials had higher GMs

465   or were significantly higher than for soils (as the FLSD test revealed at $pH < 5$). Instead, U K_d

466   data for till and subsoils were not significantly different to the soil U K_d data. From this

467   analysis it can be concluded that U K_d data from tills and subsoils can be considered

468   analogous to U K_d values from soils. Conversely, U K_d values measured in gyttja were

469 significantly different from those of soils. Thus, available till and subsoil data, but not gyttja,

470 were used to enhance the soil subsets.

471 Table 3 summarises the data derived from the datasets in which soil and analogous materials

472 U K_d values were pooled. The inclusion of subsoil and till U K_d data on the pH+OM+Texture

473 partial datasets allowed constructing CDFs with a higher number of entries, although the pH-

474 Clay groups still lacked enough data. Although the GMs and 5th-95th percentile ranges that

475 could be derived for the pH+OM+texture groups were again consistent with the U sorption

476 mechanisms (and the expected lower specific surface area values for sandy soils), no

477 statistically significant differences were observed among the textural groups at the 95% of

478 confidence level. However, low-p values, in some cases < 0.1, suggest that it may be

479 warrented to treat clay+loam and sand as separate textural classes, as their GM values differed

480 by more than one order of magnitude. Accordingly, the proposal of U K_d data for soil groups

481 based on pH+OM criteria, and further distinguishing clay+loam from sandy soils, is the more

482 detailed and statistically-based approach for radiological assessments from the available

483 entries in the current dataset.

484

485 **4. Conclusions and recommendations**

486 An integrated data treatment approach was demonstrated to decrease and describe K_d

487 variability by creating partial datasets based on the experimental approach and relevant soil

488 factors governing RN-soil interaction. When soil properties governing radionuclide

489 interactions are identified and datasets with a sufficient number of entries are available, this

490 approach can successfully derive best-estimate K_d values and related distribution functions

491 with a lower inherent variability. K_d values with lower variability are more useful for risk

492 assessment exercise.

493 As tested here for uranium, the analyses performed to the overall U K_d dataset

494 demonstrated that a single U K_d best-estimate value should be avoided as input data for

495 radiological risk assessment models. Instead, it is recommended that modellers make use of

496 informed data, more related to the scenario to be assessed, with as much information as

497 possible of the soil system, focusing on refining the input data according to the values of key

498 soil properties that govern uranium-soil interaction. This approach would lead to chose among

499 K_d best estimates that may differ in various orders of magnitude, as was observed here for the

500 case for U K_d (*e.g.*, sandy soils at pH < 5 (see Table 3; GM=20 L kg^{-1}) compared to organic

501 soils within the $5 \leq$ pH < 7 range (see Figure 4; GM=8000 L kg^{-1})).

502   In principle, it is not required to refine U K_d data based on RN-soil contact time

503   (short-term vs. long-term). However, soil factors had a greater effect on U K_d values and on

504   their variability than aging processes due to interaction dynamics. These conclusions cannot

505   be extrapolated to other radionuclides. In fact, it is recommended to examine the effect of the

506   experimental approach on K_d variability for each radionuclide under study. In the absence of

507   sufficient redox and U speciation data accompanying the reported U K_d values in the

508   literature, it is suggested to refine U K_d data based on both the pH and the OM content of the

509   soils of the scenarios to be assessed, and for every pH range, to distinguish between

510   clay+loam and sandy soils. However, if pH information is not available, it is recommended to

511   at least distinguish mineral and organic soils, as the latter has higher U K_d values.

512   It was evidenced that despite the efforts made to update the U K_d compilation there are

513   still evident U K_d gaps in certain soil groups, such is the case of organic soils at basic pH, or

514   clay soils, although these two scenarios are uncommon in the environment. It was

515   demonstrated that there were no statistical differences between the U K_d data gathered from

516   soils and other geological materials such as tills and subsoils. Therefore, data from these

517   analogue materials can be used to enhance the number of entries of the U K_d soil datasets and

518   contribute to their continuous update.

519 **References**

520 Aldaba, D., Rigol, A., Vidal, M., 2010. Diffusion experiments for estimating radiocesium and

521   radiostrontium sorption in unsaturated soils from Spain: Comparison with batch

522   sorption data. J. Hazard. Mater. 181, 1072.

523 ASTM, 2003. ASTM D4646-03 - Standard Test Method for 24-h Batch-type Measurement of

524   Contaminant Sorption by Soils and Sediments. American Society for Testing and

525   Materials, USA.

526 ASTM, 2010. ASTM C1733–10 - Standard Test Method for Distribution Coefficients of

527   Inorganic Species by the Batch Method. American Society for Testing and Materials,

528   USA.

529 Ciffroy, P., Durrieu, G., Garnier, J. M., 2009. Probabilistic distribution coefficients (Kds) in

530   freshwater for radioisotopes of Ag, Am, Ba, Be, Ce, Co, Cs, I, Mn, Pu Ra, Ru, Sb, Sr

531   and Th. Implications for uncertainty analysis of models simulating the transport of

532   radionuclides in rivers, J. Environ. Radioactiv. 100, 785-794.

533 Crançon, P., van der Lee, J., 2003. Speciation and mobility of uranium (VI) in humic- contain

534   soils. Radiochim. Acta, 91 (11), 673–679.

535 Dalvi A. A., Kumar A. V., Reddy V. R. 2014. A site specific study on the measurement of

536   sorption coefficients for radionuclides. Int. J. Environ. Sci. Technol. 11, 617-622.

537 Degryse F., Smolders, E., Parker, D. R., 2009. Partitioning of metals (Cd, Co, Cu, Ni, Pb, Zn)

538   in soils: concepts, methodologies, prediction and applications – a review. European

539   Journal of Soil Science 60: 590-612.

540 Duff, M. C., Amrhein, C., 1996. U(VI) sorption on goethite and soil in carbonate solutions.

541   Soil Sci. Soc. Am. J. 60, 1393–1400.

542 Echevarria, G., Sheppard, M., Morel, J. L., 2001. Effect of pH on the sorption of uranium in

543   soils. J. Environ. Radioactiv. 53, 257–264.

544 EPA, 1999a. EPA 402-R-99-004A - Understanding variation in partition coefficient, K_d,

545   values. Volume I: The Kd model, methods of measurement, and application of chemical

546   reaction codes, U. S. Environmental Protection Agency, USA.

547 EPA, 1999b. EPA 402-R-99-004B - Understanding Variation in Partition Coefficient, Kd,

548   Values. Volume II: Review of Geochemistry and Available Kd Values for Cadmium,

549   Cesium, Chromium, Lead, Plutonium, Radon, Strontium, Thorium, Tritium (3H), and

550   Uranium, U. S. Environmental Protection Agency, USA.

551 EPA, 2001. EPA 540-R-02-002 - Risk assessment guidance for superfund: volume III - Part
552   A, Process for conducting probabilistic risk assessment, U. S. Environmental Protection
553   Agency, USA.

554 EPA, 2017. SW-846 Test Method 1315: Mass Transfer Rates of Constituents in Monolithic or
555   Compacted Granular Materials Using a Semi-Dynamic Tank Leaching Procedure, 1st
556   revision. 37 pp, U. S. Environmental Protection Agency, USA.

557 Gil-García, C., Rigol, A., Rauret, G., Vidal, M., 2008. Radionuclide sorption-desorption
558   pattern in soils from Spain. Appl. Radiat. Isot. 66, 126–138.

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

562 Gil-García, C., Tagami, K., Uchida, S., Rigol, A., Vidal, M., 2009b. New best estimates for
563   radionuclide solid-liquid distribution coefficients in soils. Part 3: miscellany of
564   radionuclides (Cd, Co, Ni, Zn, I, Se, Sb, Pu, Am, and others). J. Environ. Radioact. 100,
565   704-715.

566 Hsi, C. K. D., Langmuir, D., 1985. Adsorption of uranyl onto ferric oxy-hydroxides:
567   application of the surface complexation site-binding model. Geochim. Cosmochim.
568   Acta 49, 1931–1941.

569 IAEA, 1994. TRS-364 - Handbook of Parameter Values for the Prediction of Radionuclide
570   Transfer in Temperate Environments, International Atomic Energy Agency, Austria .

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

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

577 ICRU, 2001. ICRU Report 65 - Quantities, Units and Terms in radioecology. International
578   Commission on Radiation Units and Measurements, Journal of the ICRU 1.

579 Kaplan, D. I., Serkiz, S. M. 2001. Quantification of thorium and uranium sorption to
580   contaminated sediments. J. Radioanal. Nuclear Chem. 248, 529-535.

581 Kaplan, D. I., Kukkadapu, R., Seaman, J. C., Arey, B. W., Dohnalkova, A. C., Buettner, S.,
582   Li, D., Varga, T., Scheckel, K. G., Jaffé, P. R. 2016. Iron mineralogy and uranium-
583   binding environment in the rhizosphere of a wetland soil. Sci. Total Environ. 569, 53-
584   64.

585   Langmuir, D., 1978. Uranium solution–mineral equilibria at low temperature with

586   applications to sedimentary ore deposits. Geochim. Cosmochim. Acta 42, 547–569.

587   Li, D., Seaman, J. C., Chang, H. S., Jaffé, P., Groos, P. K. v., Jiang, D. T., Chen, N., Lin, J.,

588   Authur, Z., Pan, Y., Scheckel, K., Newville, M., Lanzirotti, A., Kukkadapu, R. 2014.

589   Retention and chemical speciation of uranium in an oxidized wetand sediment from the

590   Savannah River Site. J. Environ. Radioactiv. 131, 40-46.

591   Li, D., Kaplan, D. I., Chang, H.-S., Seaman, J. C., Jaffé, P. R., Koster van Groos, P.,

592   Scheckel, K. G., Segre, C. U., Chen, N., Jiang, D.-T. 2015. Spectroscopic evidence of

593   uranium immobilization in acidic wetlands by natural organic matter and plant roots.

594   Environ. Sci. Technol. 49, 2823-2832.

595   OECD, 2000. Guideline 106 - Guideline for the testing of chemicals: Adsorption-desorption

596   using a batch equilibrium method. Organisation for Economic Cooperation and

597   Development, France.

598   Payne, T. E., Brendler, V., Comarmond, M. J., Nebelung, C., 2011. Assessment of surface

599   area normalisation for interpreting distribution coefficients (K_d) for uranium sorption. J.

600   Environ. Radioactiv. 102, 888-895.

601   Payne, T. E., Davis, J. A., Waite, T. D., 1996. U sorption on ferrihydrite – effects of

602   phosphate and humic acid. Radiochim. Acta 74, 239–243.

603   Ramírez-Guinart, O., Kaplan, D., Rigol, A., Vidal, M. 2020a. Deriving probabilistic soil

604   distribution coefficients (K_d). Part 2: Reducing caesium K_d uncertainty by accounting

605   for experimental approach and soil type. Submitted to J. Environ. Radioactiv.

606   Ramírez-Guinart, O., Kaplan, D., Rigol, A., Vidal, M. 2020b. Deriving probabilistic soil

607   distribution coefficients (K_d). Part 3: Reducing variability of americium K_d best

608   estimates using soil properties and chemical and geological material analogues.

609   Submitted to J. Environ. Radioactiv.

610   Sheppard, M. I., Thibault, D. H., 1990. Default soil solid/liquid partition coefficients, Kds, for

611   four major soil types: a compendium. Health Phys. 59, 471.

612   Sheppard, S. C., 2011. Robust Prediction of K_d from Soil Properties for Environmental

613   Assessment. Hum. Ecol. Risk Assess. 17(1), 263-279.

614   Sheppard, S. C., Sheppard, M. I., Tait, J. C., Sanipelli, B. L., 2006. Revision and metaanalysis

615   of selected biosphere parameter values for chlorine, iodine, neptunium, radium, radon

616   and uranium. J. Environ. Radioactiv. 89, 115–137.

617 Simon-Cornu, M., Beaugelin-Seiller, K., Boyer, P., Calmon, P., Garcia-Sanchez, L., Mourlon,
618   C., Nicoulaud, V., Sy, M., Gonze, M. A., 2015. Evaluating variability and uncertainty in
619   radiological impact assessment using SYMBIOSE. J. Environ. Radioactiv. 139, 91-102.

620 Sowder, A.G., Bertsch, P.M., Morris, P.J., 2003. Partitioning and availability of uranium and
621   nickel in contaminated riparian sediments. J. Environ. Qual. 32, 885–898.

622 Stenhouse, M. J., 1994. Technical Report NTB 93-06 - Sorption Databases for Crystalline,
623   Marl and Bentonite for Performance Assessment. NAGRA, Switzerland.

624 Stockmann, M., Schikora, J., Becker, D.-A., Flügge, J., Noseck, U., Brendler, V. 2017. Smart
625   K_d-values, their uncertainties and sensitivities - Applying a new approach for realistic
626   distribution coefficients in geochemical modeling of complex systems. Chemosphere
627   187, 277-285.

628 Sy, M., Gonze, M. A., Métivier, J. M., Nicoulaud-Gouin, V., Simon-Cornu, M., 2016.
629   Uncertainty analysis in post-accident risk assessment models: application to the
630   Fukushima accident. Annals of Nuclear Energy 93, 94-106.

631 Vandenhove, H., van Hees, M., Wouters, K., Wannijn, J., 2007. Can we predict uranium
632   bioavailability based on soil parameters? Part 1: Effect of soil parameters on soil
633   solution uranium concentration. Environ. Pollut. 145, 587–595.

634 Vandenhove, H., Gil-García, C., Rigol, A., Vidal, M. 2009. New best estimates for
635   radionuclide solid-liquid distribution coefficients in soils. Part 2. Naturally occurring
636   radionuclides. J. Environ. Radioactiv. 100, 697-703.

637 Vandenhove, H., Hurtgen, C., Payne, T.E. 2010. "Uranium", Radionuclides in the
638   Environment (Atwood, D.A., Ed.). John Wiley & Sons.

639 Waite, T. D., Davis, J. A., Payne, T. E., Waychunas, G. A., Xu, N., 1994. U(VI) sorption to
640   ferrihydrite: application of the surface complexation model. Geochim. Cosmochim.
641   Acta 58, 5465–5478.

Table 1

Description of the experimental approach categories followed for K_d quantification.

Experimental approach categories		Examples of experiments
Short-term	In-situ	K_d of anthropogenic elements determined from their concentration in paired water-and-soil samples collected in areas recently (<~1 yr) contaminated with radioisotopes or with low concentrations of stable isotopes of the target element.
		K_d of indigenous elements (radioisotopes or stable isotopes) derived from determining their concentration in paired water-and-soil samples collected in non-contaminated areas. The target element concentration in the soil-porewater and in the reversible fraction of the sorbed indigenous element (*e.g.* using mild extractant reagents, such as acetic acid 0.43M) are quantified.
	Sorption	K_d of elements derived from applying a sorption batch test based on putting in contact for short times (<~1 yr) a clean soil (initially non-contaminated) with a solution spiked with radioisotopes or with low concentrations of stable isotopes of the target element. Soil bound radionuclide concentrations are derived by difference between aqueous radioisotope concentration before and after contact with the soil.
	Desorption	K_d of anthropogenic elements derived from estimating the solid material concentration by applying an extraction batch test to soils recently contaminated with radioisotopes or with low concentrations of stable isotopes of the target element (*e.g.*, soil residues coming from a previous sorption test or soils sampled from recently contaminated areas). The aqueous concentration used in the K_d calculation is the concentration of the target radionuclide in the contact solution.
		K_d of indigenous elements (radioisotopes or stable isotopes) derived from estimating the solid phase concentration by applying an extraction test to contaminated solid materials when only the reversible fraction of the sorbed indigenous element is quantified (*e.g.*, using mild extractant reagents, such as ammonium acetate 1M or acetic acid 0.43M). The aqueous concentration is the concentration of the target radionuclide in the contact solution.
Long-term	In-situ	Similar to "Short-term/In-situ K_d Experimental Approach" except conducted with soils aged with target element or radionuclide for >1 yr.
	Desorption	Similar to "Short-term/Desorption K_d Experimental Approach" except conducted with soils aged with target indigenous element or radionuclide for >1 yr.

Table 2

Descriptors of U K_d (L kg^{-1}) distributions derived from soils and other environmental materials after applying the pH criterion.

Dataset	Material	N	GM	GSD	FLSD[1]	5th	95th
pH < 5	Soils	53	1.5×10^2	7.8	a	3.5×10^0	1.6×10^3
	Till	8	3.5×10^2	8.8	a	2.6×10^0	3.6×10^3
	Subsoil	9	1.0×10^2	5.0	a	2.6×10^1	5.0×10^3
	Gyttja	11	3.4×10^3	3.7	b	8.0×10^2	4.4×10^4
$5 \leq$ pH < 7	Soils	68	1.0×10^3	5.2	a	8.3×10^1	1.5×10^4
	Till*	5	1.5×10^3	2.6	a	n.a.	n.a
	Subsoil	13	2.2×10^3	8.7	a	7.9×10^0	1.0×10^4
	Gyttja*	2	1.6×10^3	n.a.	-	n.a.	n.a.
pH $\geq$ 7	Soils	75	1.0×10^2	9.5	a	1.3×10^0	3.7×10^3
	Till	7	8.0×10^1	12	a	8.1×10^{-1}	3.7×10^2
	Subsoil	24	4.7×10^1	11	a	6.3×10^{-1}	1.7×10^3
	Gyttja*	2	6.6×10^2	n.a.	-	n.a.	n.a.

N = number of observations, GM = geometric mean, GSD = geometric standard deviation
[1] Different letters among the datasets compared indicate statistically significant differences between GMs according to the Fisher's Least Significant Differences test.
*CDF not constructed due to lack of K_d data (N << 10); n.a.: not applicable

Table 3

Descriptors of U K_d (L kg^{-1}) distributions based on the pH+OM+Texture criterion, including data gathered from soils and from solid materials analogous to soils.

Partial dataset		N	GM	GSD	FLSD[1]	FLSD[1]	5th	95th
pH < 5	Mineral	58	1.6×10^2	7.3	a		3.0×10^0	3.6×10^3
	Clay*	1	4.8×10^2	n.a.		-	n.a.	n.a.
	Loam	15	2.2×10^2	3.6		a	1.6×10^1	1.6×10^3
	Clay+Loam	16	2.6×10^2	3.5		a	1.6×10^1	1.6×10^3
	Sand	17	2.0×10^1	16		a	7.0×10^{-1}	6.7×10^3
5 ≤ pH < 7	Mineral	73	1.0×10^3	5.3	b		4.6×10^1	1.0×10^4
	Loam	42	1.2×10^3	3.9		a	1.5×10^2	1.0×10^4
	Sand	15	7.2×10^2	4.6		a	8.3×10^1	6.7×10^4
pH ≥ 7	Mineral	95	8.0×10^1	9.1	c		9.4×10^{-1}	2.0×10^3
	Clay	7	2.1×10^2	7.7		a	5.0×10^0	4.7×10^2
	Loam	42	1.0×10^2	9.0		a	1.2×10^0	2.0×10^3
	Clay+Loam	49	1.0×10^2	8.7		a	1.2×10^0	2.0×10^3
	Sand	13	3.5×10^1	4.6		a	7.0×10^0	1.3×10^3

N = number of observations, GM = geometric mean, GSD = geometric standard deviation
[1] Different letters among the datasets compared indicate statistically significant differences between GMs according to the Fisher's Least Significant Differences test.
*CDF not constructed due to lack of K_d data (N << 10); n.a.: not applicable

Figure captions

Fig. 1. Illustration of the Group Mean Centering (GMC) statistical approach used to minimize the effect of soil factors identified as relevant to the interaction of the target element. In this example, the effects of the pH+OM criterion on experimental approach datasets are corrected. AM(#) is the arithmetic mean for grouping #; FLSD is the Fisher's least significant differences test.

Fig. 2. CDFs and descriptors of U K_d (L kg^{-1}) distributions for Mineral and Organic soil-types (data for the Overall dataset are included for comparison). Points indicate individual dataset values whereas lines indicate the fitted distributions.
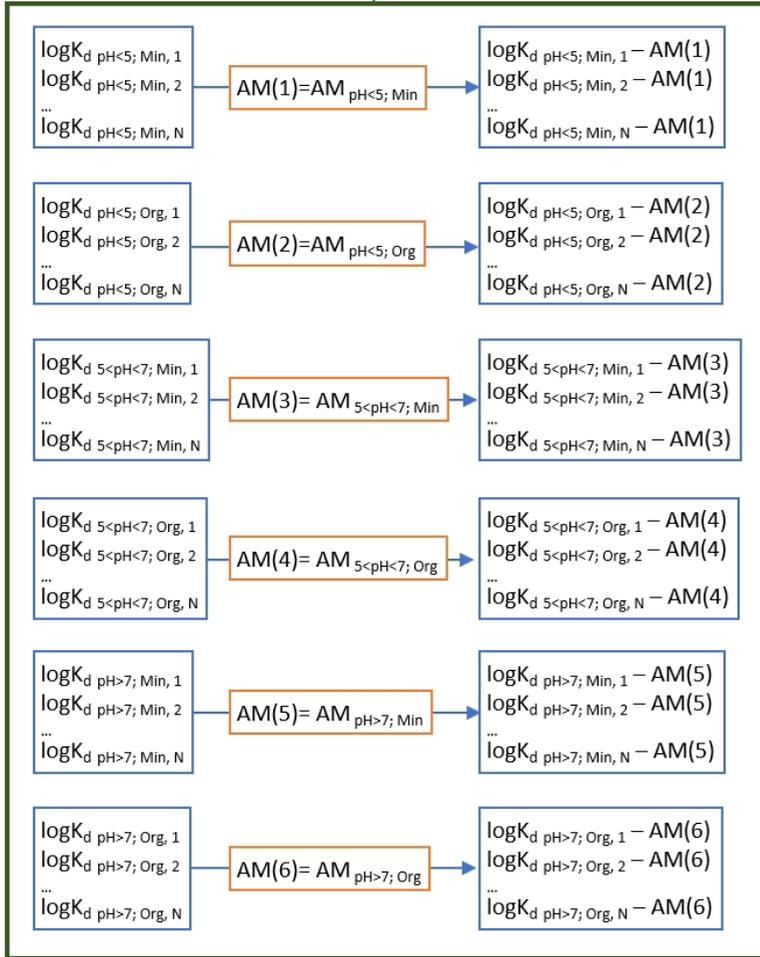
Fig. 3. CDFs and descriptors of U K_d (L kg^{-1}) distributions for pH groups.

Fig. 4. CDFs and descriptors of U K_d (L kg^{-1}) distributions for soil-types according to the pH + OM criterion.
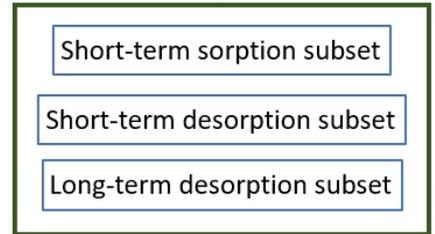
```
K_d dataset
        │
        ▼
   logK_d dataset
        │
        ▼
logK_d grouping according to pH+OM criterion
```

logK_d pH<5; Min, 1
logK_d pH<5; Min, 2
…
logK_d pH<5; Min, N

AM(1)=AM_{pH<5; Min}

logK_d pH<5; Min, 1 − AM(1)
logK_d pH<5; Min, 2 − AM(1)
…
logK_d pH<5; Min, N − AM(1)

logK_d pH<5; Org, 1
logK_d pH<5; Org, 2
…
logK_d pH<5; Org, N

AM(2)=AM_{pH<5; Org}

logK_d pH<5; Org, 1 − AM(2)
logK_d pH<5; Org, 2 − AM(2)
…
logK_d pH<5; Org, N − AM(2)

logK_d 5<pH<7; Min, 1
logK_d 5<pH<7; Min, 2
…
logK_d 5<pH<7; Min, N

AM(3)= AM_{5<pH<7; Min}

logK_d 5<pH<7; Min, 1 − AM(3)
logK_d 5<pH<7; Min, 2 − AM(3)
…
logK_d 5<pH<7; Min, N − AM(3)

logK_d 5<pH<7; Org, 1
logK_d 5<pH<7; Org, 2
…
logK_d 5<pH<7; Org, N

AM(4)= AM_{5<pH<7; Org}

logK_d 5<pH<7; Org, 1 − AM(4)
logK_d 5<pH<7; Org, 2 − AM(4)
…
logK_d 5<pH<7; Org, N − AM(4)

logK_d pH>7; Min, 1
logK_d pH>7; Min, 2
…
logK_d pH>7; Min, N

AM(5)= AM_{pH>7; Min}

logK_d pH>7; Min, 1 − AM(5)
logK_d pH>7; Min, 2 − AM(5)
…
logK_d pH>7; Min, N − AM(5)

logK_d pH>7; Org, 1
logK_d pH>7; Org, 2
…
logK_d pH>7; Org, N

AM(6)= AM_{pH>7; Org}

logK_d pH>7; Org, 1 − AM(6)
logK_d pH>7; Org, 2 − AM(6)
…
logK_d pH>7; Org, N − AM(6)

GMC logK_d dataset

GMC logK_d grouping according to the experimental approach

Short-term sorption subset

Short-term desorption subset

Long-term desorption subset

FLSD test

Statistical differences?
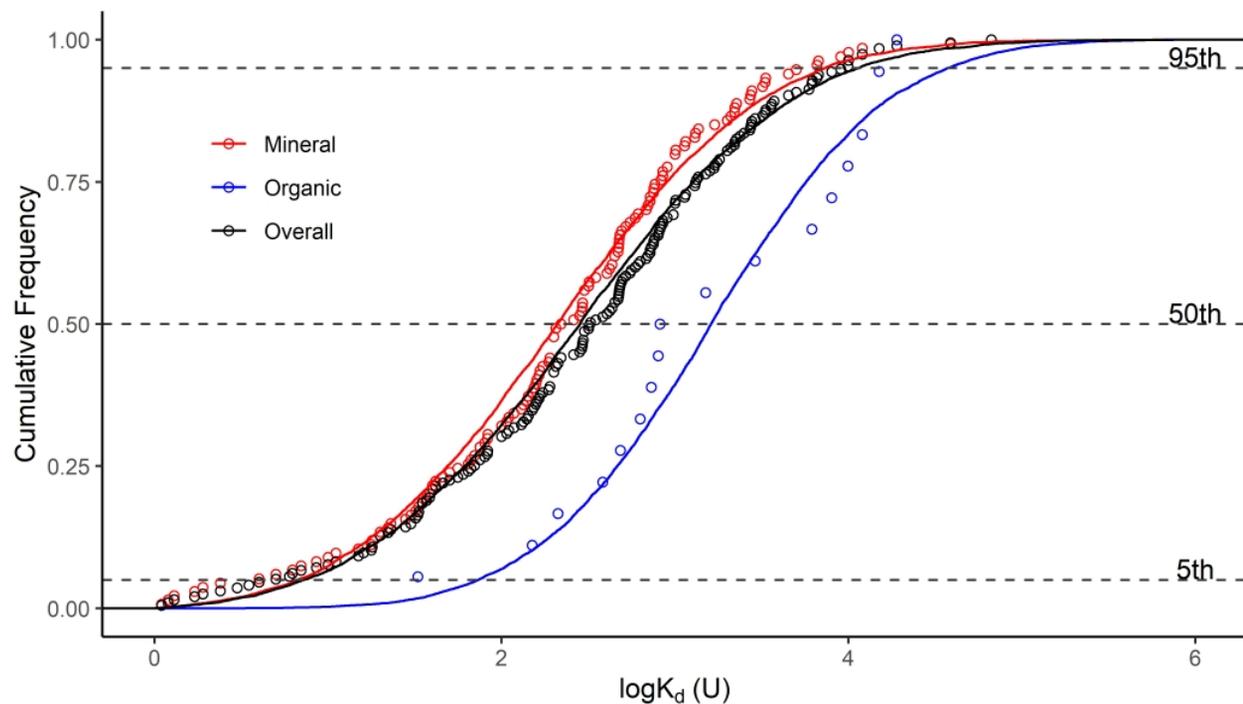
Partial dataset	N	GM	GSD	FLSD[1]	5th	95th
Overall	196	3.2×10^2	9.8	a	3.5×10^0	9.9×10^3
Mineral+Organic	153	3.0×10^2	9.3	a	4.0×10^0	1.0×10^4
Mineral (OM < 20%)	135	2.2×10^2	8.7	a	2.4×10^0	6.6×10^3
Organic (OM ≥ 20%)	18	1.1×10^3	6.4	b	3.3×10^1	1.9×10^4

N = number of observations, GM = geometric mean, GSD = geometric standard deviation
[1] Different letters among the datasets compared indicate statistically significant differences according to the Fisher's Least Significant Differences test.
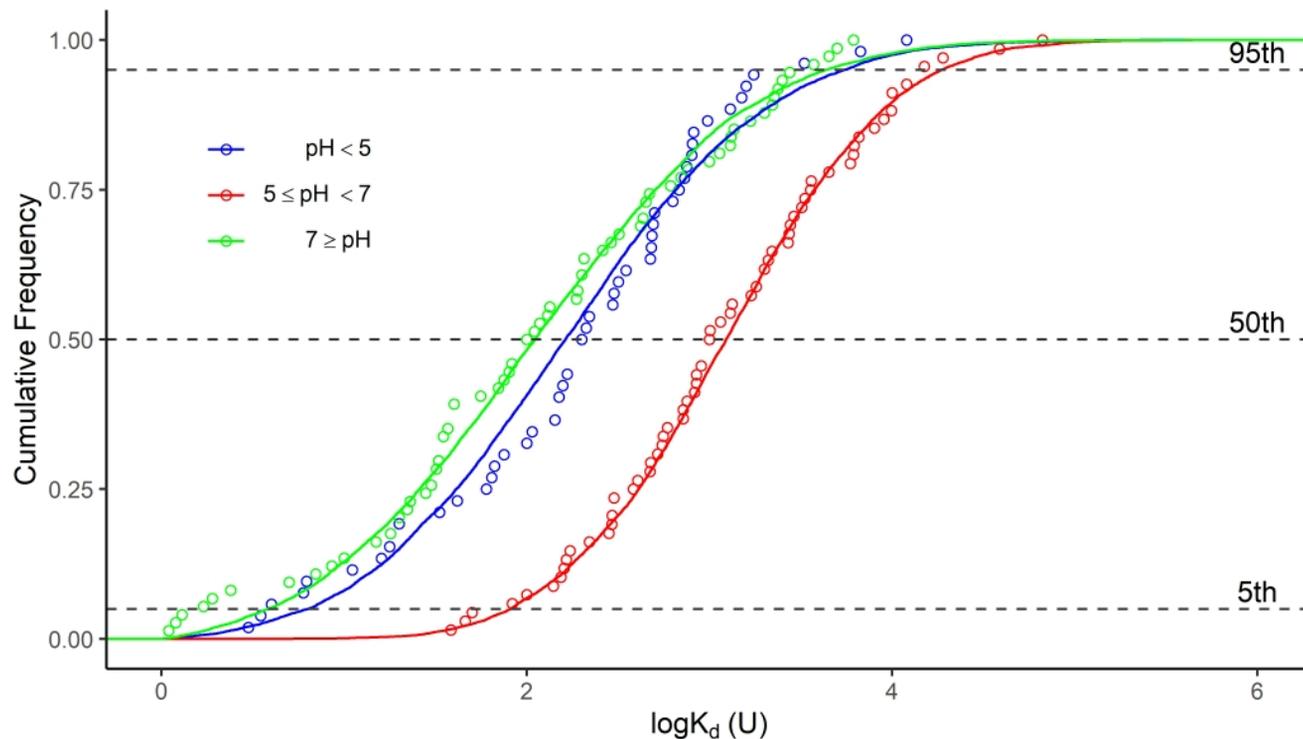
Partial dataset	N	GM	GSD	FLSD[1]	5th	95th
pH < 5	53	2.0×10^2	7.8	a	3.5×10^0	3.3×10^3
5≤ pH <7	68	1.0×10^3	5.2	b	8.3×10^1	1.5×10^4
pH ≥ 7	75	1.0×10^2	9.5	a	1.3×10^0	3.7×10^3

N = number of observations, GM = geometric mean, GSD = geometric standard deviation
[1] Different letters among the datasets compared indicate statistically significant differences according to the Fisher's Least Significant Differences test.
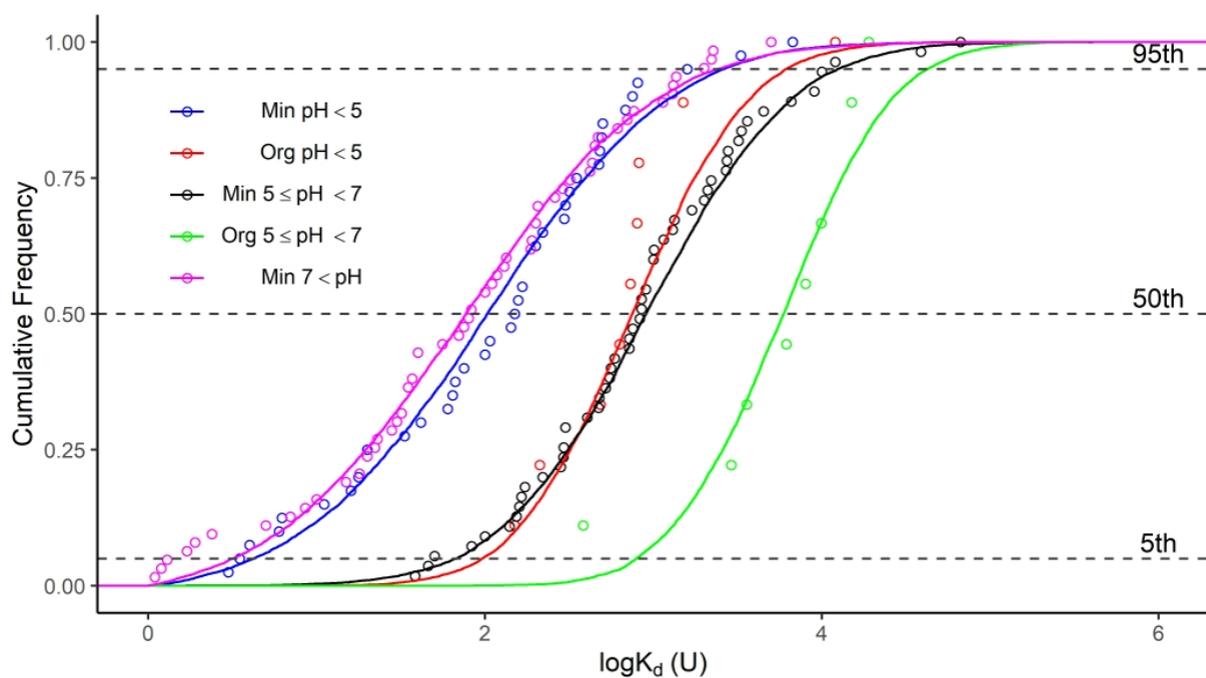
Partial dataset		N	GM	GSD	FLSD[1]	5th	95th
pH < 5	Mineral	41	1.5×10^2	7.4	a	3.5×10^0	1.6×10^3
	Organic	9	7.3×10^2	3.5	b	1.5×10^2	1.2×10^4
$5 \leq pH < 7$	Mineral	55	8.5×10^2	5.0	a	5.0×10^1	1.2×10^4
	Organic	9	8.0×10^3	3.4	b	3.8×10^2	1.9×10^4
$pH \geq 7$	Mineral	64	8.1×10^1	8.2	n.a.	1.3×10^0	2.0×10^3
	Organic*	2	4.5×10^2	n.a.	n.a.	n.a.	n.a.

N = number of observations, GM = geometric mean, GSD = geometric standard deviation
[1] Different letters among the datasets compared indicate statistically significant differences according to the Fisher's Least Significant Differences test.
*CDF not constructed due to lack of K_d data (N << 10); n.a.: not applicable

SUPPLEMENTARY MATERIAL

Deriving probabilistic soil distribution coefficients (K_d). Part 1: General approach to decreasing and describing variability and example using uranium K_d values

Table S1

Effect of the experimental approach and pH and mineral grouping on U K_d values.

pH Group	Organic + Mineral Group	Experimental Approach[1]	N	FLSD[2]
pH < 5	Mineral	ST-S	8	a
		ST-D	17	a
		LT-D	4	a
	Organic	ST-S	-	-
		ST-D	2	-
		LT-D	7	-
	Min+Org	ST-S	8	a
		ST-D	19	a
		LT-D	11	a
5 ≤ pH < 7	Mineral	ST-S	18	a
		ST-D	20	b
		LT-D	11	a
	Organic	ST-S	2	-
		ST-D	-	-
		LT-D	5	-
	Min+Org	ST-S	20	ab
		ST-D	20	b
		LT-D	16	a
pH ≥ 7	Mineral	ST-S	32	a
		ST-D	1	-
		LT-D	24	a
	Organic	ST-S	2	-
		ST-D	-	-
		LT-D	-	-
	Min+Org	ST-S	34	a
		ST-D	1	-
		LT-D	24	a

[1] Experimental approach: ST-S = Short-term sorption; ST-D = Short-term desorption; LT-D = Long-term desorption.

[2] Fisher's Least Significant Difference. Different letters among the datasets compared indicate statistically significant differences according to the Fisher's Least Significant Differences test.

Table S2

Optimisation of the OM% threshold to define organic and mineral soils for U K_d.

Partial data set	N	GM	GSD	FLSD[1]
Mineral (OM < 5%)	82	1.6×10^2	7.0	a
Mineral (OM < 20%)	119	2.2×10^2	7.9	a
Organic (OM ≥ 5%)	52	6.8×10^2	9.0	b
Organic (OM ≥ 20%)	15	1.8×10^3	7.4	b

[1] Different letter among the datasets compared indicate statistically significant differences according to the Fisher's Least Significant Differences test.

Table S3

Descriptors of U K_d (L kg^{-1}) distributions based on the pH+OM+Texture criterion for mineral soils.

Partial dataset		N	GM	GSD	FLSD[1]	5th	95th
pH < 5	Clay*	1	4.8×10^2	n.a.	-	n.a.	n.a.
	Loam	10	2.6×10^2	3.7	a	1.6×10^1	1.6×10^3
	Sand	14	1.4×10^1	18	a	7.0×10^1	6.7×10^3
5 ≤ pH < 7	Loam	34	1.0×10^3	4.2	a	1.4×10^2	1.2×10^4
	Sand	13	5.6×10^2	5.0	a	8.3×10^1	6.7×10^4
pH ≥ 7	Clay	7	2.1×10^2	7.7	a	5.0×10^0	4.7×10^2
	Loam	33	7.5×10^1	10	a	1.1×10^0	2.2×10^3
	Sand	11	3.5×10^1	5.0	a	7.0×10^0	1.3×10^3

N = number of observations, GM = geometric mean, GSD = geometric standard deviation
[1] Different letter among the datasets compared indicate statistically significant differences according to the Fisher's Least Significant Differences test.
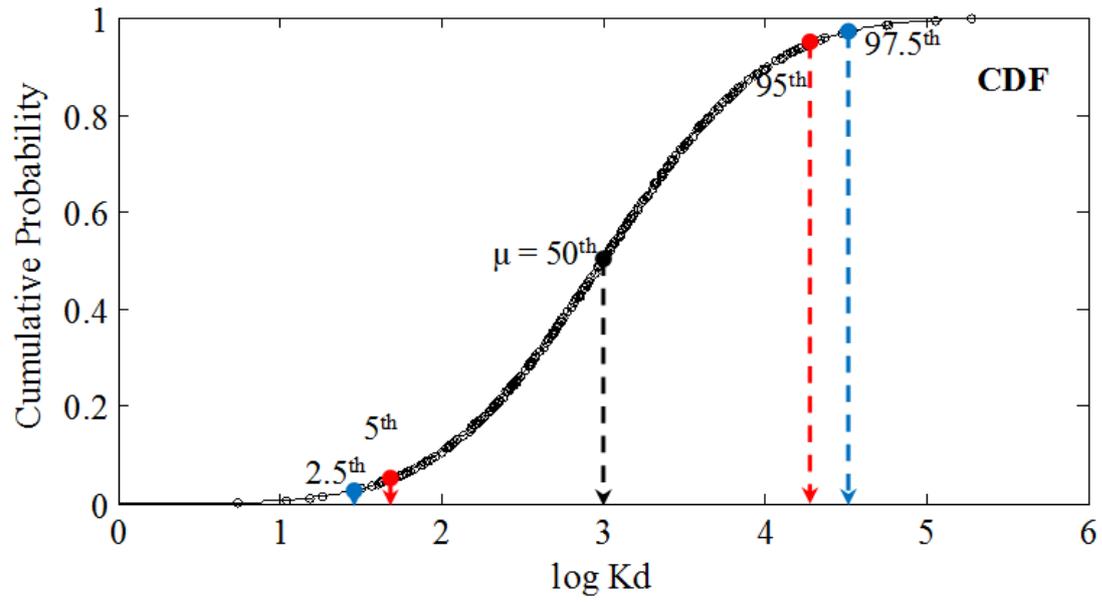*CDF not constructed due to lack of K_d data (N << 10); n.a.: not applicable

Fig. S1. Graphical representation of the CDF of an ideal lognormally distributed K_d dataset.
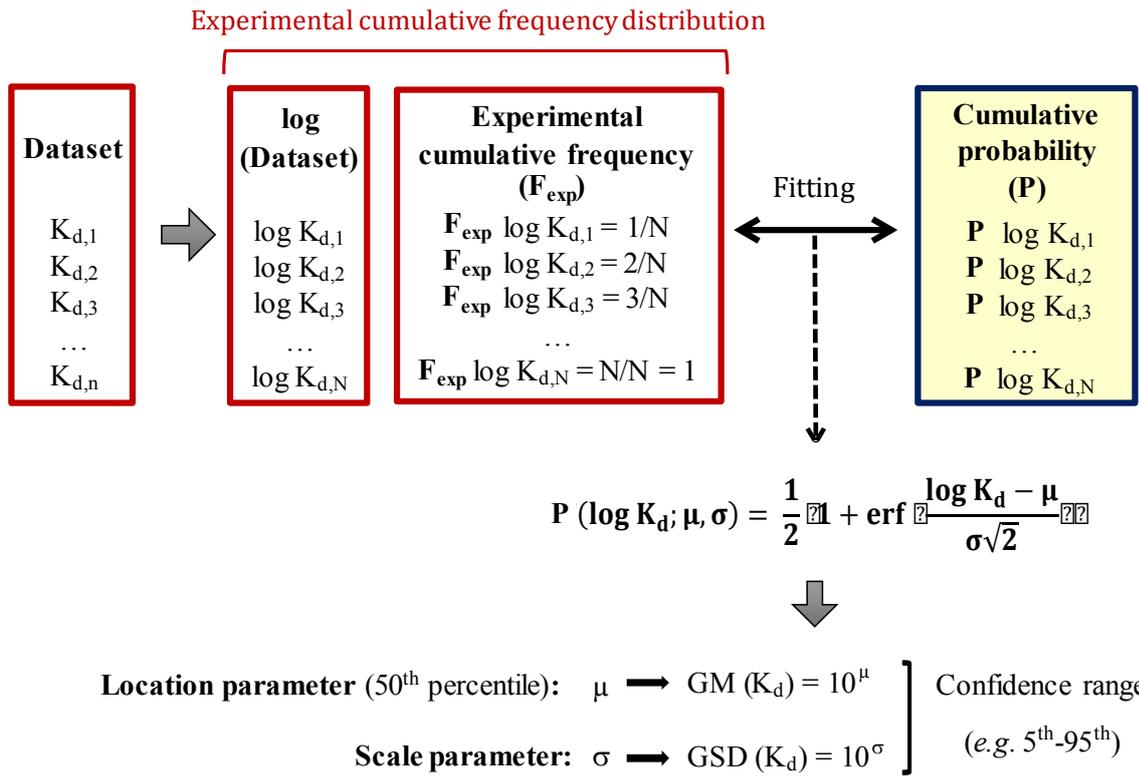
Fig. S2. Schematic process to construct CDFs and to derive best estimates values and related confidence intervals from K_d datasets.