Radiocaesium soil-to-plant transfer: a meta-analysis of key variables and data gaps on a global scale

Margot Vanheukelom [a,b,*] ⓘ, Mark Mng'ong'o [c], Floris Abrams [a,c], Surya Gupta [d], Talal Almahayni [a], Lieve Sweeck (deceased)[a], Jos Van Orshoven [c], Erik Smolders [b]

[a] *Biosphere Impact Studies, Belgian Nuclear Research Centre (SCK CEN), Boeretang 200, 2400, Mol, Belgium*
[b] *Division of Soil and Water Management, KU Leuven, Kasteelpark Arenberg 20, 3001, Leuven, Belgium*
[c] *Division of Forest, Nature and Landscape, KU Leuven, Celestijnenlaan 200e, box 2411, 3001, Leuven, Belgium*
[d] *RadioPharma Research, Belgian Nuclear Research Centre (SCK CEN), Boeretang 200, 2400, Mol, Belgium*

ARTICLE INFO

ABSTRACT

A harmonized, publicly accessible database of worldwide observations and experiments on radiocaesium transfer from soil to plants is lacking. Such a database is needed for evaluating and establishing transfer models, especially for regions with limited research but operational or planned nuclear reactors. Therefore, we systematically screened the literature for radiocaesium soil-to-plant transfer factors (CR, i.e., concentration ratios), extracted data that met the criteria for experimental soundness, relevance, and traceability, and compiled a harmonized database. The database included 7,182 CR data points and associated variables from 139 source documents. The CRs ranged from 0.000028 to 380 kg kg^{-1}, with the highest CR observed with soils from tropical climates and the lowest with soils from temperate climates. However, data from tropical ($N = 411$) and arid climates ($N = 335$) remained limited. Univariate and multivariate analyses revealed that CRs were most influenced by the specific study (methods and designs) in which the data were obtained, followed by soil properties and plant species-based categories. On a subset ($N = 199$) that contained all variables required for semi-mechanistic models, it was found that these models fitted the CR data rather well ($R^2 = 0.42$–0.50). Slightly better predictions with the same data were found with a random forest model ($R^2 = 0.51$) or a statistical mixed-effects model ($R^2 = 0.58$). More adequate machine learning models could not yet be created due to insufficient reliable data. The harmonized database in this study can be further completed and analyzed to support machine learning applications and improve impact assessments of food chain contamination following accidental radiocaesium deposition on agricultural land.

1. Introduction

The expansion of nuclear reactors from the global north to the south calls for a revision of radiological impact assessments so that remedial strategies are ready in case of a nuclear accident anywhere on earth. Release of radioactive caesium (or radiocaesium) is a concern because it has a long half-life (2 years for [134]Cs and 30 years for ^{137}Cs) and because it is similar to potassium (K), so that it can be taken up by plants and hence enter the food chain. While the foliar pathway dominates immediately after radiocaesium deposition, especially during the growing season, transfer of radiocaesium through the root system in soil becomes the main pathway of radiocaesium entering the food chain in the medium to long term (Strebl et al., 2007). The potential transfer of radiocaesium from soil to plants is commonly quantified using a concentration ratio (CR, or transfer factor), expressed as:

$$CR = \frac{\text{concentration of caesium isotope in plant } \left(\frac{Bq}{\text{kg dry plant part}}\right)}{\text{concentration of caesium isotope in soil } \left(\frac{Bq}{\text{kg oven dry soil}}\right)}$$

The CR mainly depends on soil and plant properties and on the time after the contamination. Many data and models have been published on the fate of radiocaesium in the terrestrial environment, especially after the Chornobyl and Fukushima accidents (e.g., reviewed by Almahayni et al., 2019; IAEA, 2020, 2009, 2006; IUR, 1992; Li et al., 2024; Nisbet et al., 1999). The International Atomic Energy Agency (IAEA) published

---

* Corresponding author. Biosphere Impact Studies, Belgian Nuclear Research Centre (SCK CEN), Boeretang 200, 2400, Mol, Belgium.
E-mail address: margot.vanheukelom@kuleuven.be (M. Vanheukelom).

reference values for radiological impact assessments based on these research findings (IAEA, 2021, 2020, 2009). In the IAEA reports, CRs were compiled into categories for various crop types and soils with different mineral and organic matter contents so that CR estimates based on geometric means could be made. However, it is well established that there is a large variability of the CR within these soil-plant categories (Almahayni et al., 2019). While using generic CR values for screening environmental radiological impact assessments might be adequate in some cases, more detailed assessments likely require site-specific CR values to reflect environmental conditions at the site and to reduce assessment uncertainty. For example, CRs typically vary over six orders of magnitude, and within the soil and plant categories, at least three orders of magnitude variation have been reported, such as for leafy vegetables grown on loamy soils (CR = 0.0003–0.7 kg kg^{-1}; IAEA, 2009).

Alternatively, semi-mechanistic models have been developed to estimate CRs based on soil variables such as *clay content, exchangeable K content, ammonium (NH_4) concentration in solution, organic matter content* and *pH* and the *soil-radiocaesium contact time* (Absalom et al., 2001; Almahayni et al., 2019). These models have been constructed based on the concept that the CR is explained by *soil solution ^{137}Cs and K concentrations* as the measure of bioavailable ^{137}Cs (Absalom et al., 1999; Smolders et al., 1997). These models perform reasonably well (Vanheukelom et al., 2024). However, it is clear that considerable variation in CR data remains unexplained (Vanheukelom et al., 2024) and that the calibration and validation have been mainly limited to the soil-plant conditions found in the areas affected by the Chornobyl (e.g., Brimo et al., 2021) and Fukushima accidents (e.g., Uematsu et al., 2016).

Established (e.g., Multivariate Linear Regression) and newer data-driven modeling approaches (e.g., machine learning) are promising for CR estimations on a larger, i.e., a global scale. Moreover, integrating approaches such as Digital Soil Mapping (McBratney et al., 2003) could enable spatially explicit estimations of CRs because it uses spatially available predictors (e.g., soil properties, climate data, land use) and can exploit spatial autocorrelation, i.e., the tendency of nearby locations to have similar values, to improve estimates (e.g., exchangeable K content) when there are sufficient observations. Recent studies have demonstrated the potential of machine learning (random forest models) to estimate CRs in Germany (Urso et al., 2023) and Fukushima Prefecture, Japan (Shuryak, 2022). These studies highlight the power of random forest models in handling both numerical (e.g., *soil-radiocaesium contact time*) and categorical (e.g., *plant category*) variables, as well as complex interactions between variables (e.g., Cs^+-K^+ competition in soil). However, their effectiveness depends on large, diverse datasets covering a wide range of soil-plant conditions, as random forest models cannot extrapolate beyond the range of the calibration data. Despite this potential, a comprehensive database with proper georeferencing of radiocaesium CRs and information on the variables affecting radiocaesium transfer from soil to plant is not openly accessible. For example, the International Union of Radioecology (IUR) database (IUR, 1992) is available on demand but lacks essential data such as *exchangeable K content* in soil. The IAEA reports often include summarized values for soil and plant types, but the underlying data are not released for research purposes by third parties (IAEA, 2020, 2009). Additionally, national databases, such as those of the Ministry of Agriculture, Forestry and Fisheries (MAFF) of Japan, are reported in Japanese (MAFF, 2017, 2015, 2014) making the exploitation challenging for non-Japanese researchers.

This study aims to improve estimates of radiocaesium transfer in agricultural soils on a global scale so that these estimates can be used with confidence in the event of future nuclear accidents or for the analysis of hypothetical scenarios. By building a comprehensive database, the study seeks to reduce uncertainty in impact assessments and support food safety measures. Data were collected from multiple sources, screened for quality and harmonized, and a database was constructed. The number of soil, plant, climate, and other variables associated with each CR data point varied among studies as expected. With these data points, a multivariate analysis was performed, and different models were compared to evaluate their performance. A comparison was made between the semi-mechanistic models (Absalom et al., 2001; Tarsitano et al., 2011), log-linear multiple regression and random forest models (Breiman, 2001), using a subset of data points for which all relevant variables of the semi-mechanistic model were available.

2. Materials and methods

2.1. Data collection

A total number of 7,182 CR data points, derived from 139 different source documents were collated. Data points were in- or excluded based on a systematical screening. The data collection and selection procedure is summarized in Fig. 1. The data sources included pot experimental datasets, IAEA technical reports, and articles obtained by systematically searching in Web of Science™ (© Clarivate, 2022) and Scopus® (Elsevier) literature databases.

Data from pot experiments (Gilis and Sweeck, 2019; Sanchez et al., 1999; Smolders et al., 1997; Uematsu et al., 2018; Vanheukelom et al., 2023, 2024) with similar setup were included in the database because of data reliability and high quality due to the standardized conditions in which the experiments were performed.

Most data originated from IAEA technical reports (IAEA, 2021, 2020, 2009, 2006). Data in these reports were carefully compiled by experts. However, these CRs have been reported as descriptive statistics per plant-soil categories and not as raw data. Therefore, that dataset was reconstructed by revisiting the source documents. In total, 271 source documents had been cited in the IAEA reports, of which 163 were accessible online (most non-accessible were sources in Russian and Japanese), of which 123 met all the selection criteria (Table A.2).

Relevant publications indexed by the Web of Science and Scopus were obtained by means of a search string of keywords covering four concepts: *caesium* as the element of interest, *soil* as substrate, *plant* as receiving organism, and *transfer* as a process (Table A.3). The search string was validated using the source documents cited in the IAEA reports relevant to this study. Of the 102 sources of IAEA reports in the Web of Science database, 96 were obtained using that string. The Web of Science and Scopus platforms (Table A.4) were simultaneously searched on December 19, 2022 at 15:00 (CEST) using the search string. After deduplication with EndNote™ (version 20.4.1) software, fewer sources remained. To keep the work feasible, we selected a subset of 100 sources using a random number generator. The subset was screened using Rayyan software (Ouzzani et al., 2016). Relevant sources were selected first based on the titles and abstracts and then on the full articles using the selection criteria (Table A.2) if they were accessible online. Only 10 out of 100 sources from the random subset were retained.

Data were extracted from tables from source documents by copying and data were extracted from graphs by reconstructing them with an online extraction tool (https://plotdigitizer.com/app).

2.2. Selection criteria

Selection criteria were used to systematically filter data during the screening process, details are given in the Supplementary information (Table A.2) and are briefly summarized here. Data were rejected if:

- Radiocaesium in or on the plant originated from foliar interception of fallout, as this database referred to root uptake of radiocaesium from soil, and/or
- Plants were grown in hydroponics, agar, or other substrates not representative of agricultural soils, and/or
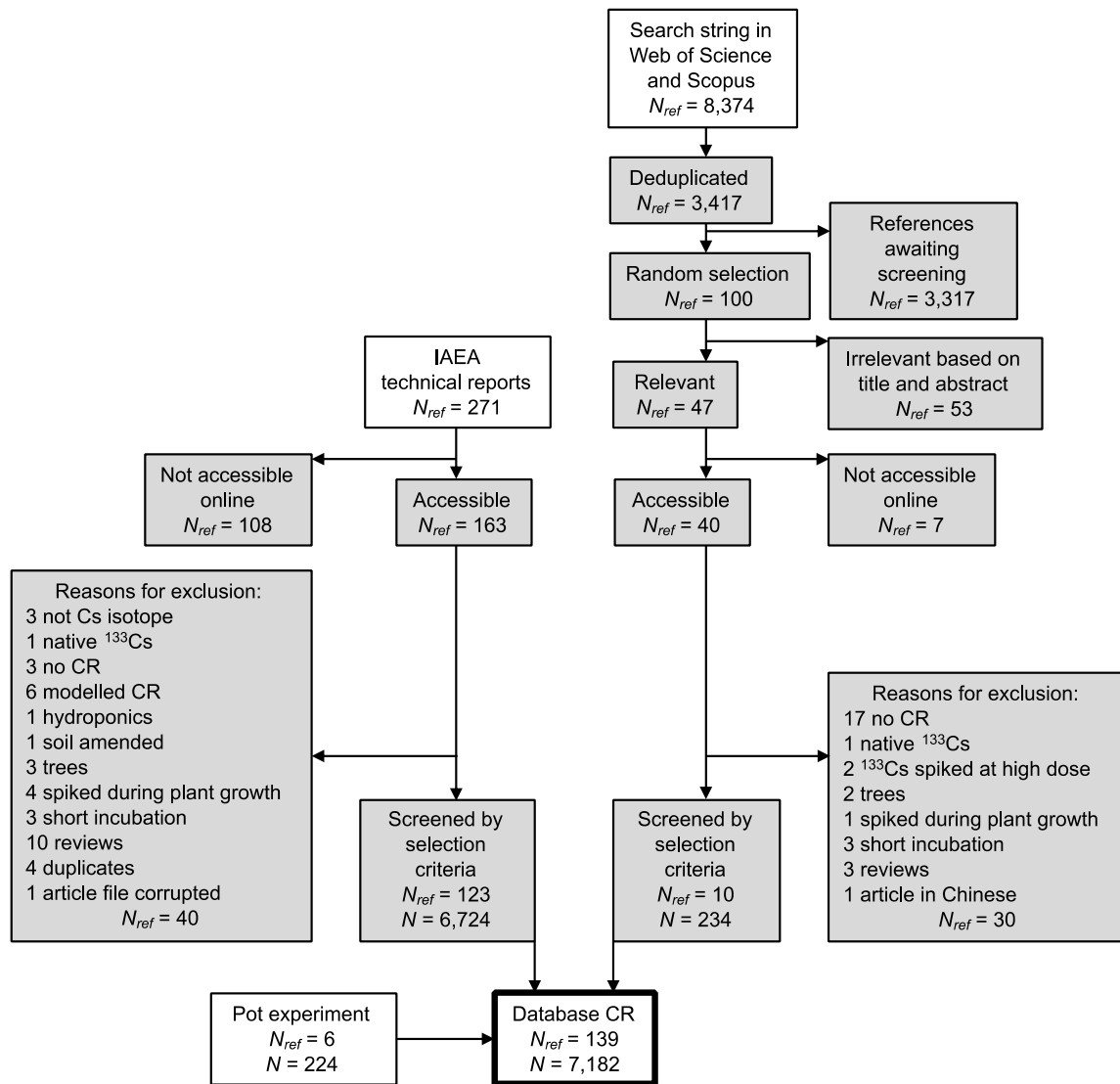
Fig. 1. Process of collecting and selecting data for the CR database. The selection criteria are detailed in Supplementary Information (Table A.2). N_{ref} = number of source documents; N = number of data points. Grey boxes are data selection steps.

- The contact time between radiocaesium and soil was less than 2 months at the time of plant collection to ensure that a pseudo-equilibrium had been established, as shorter times may reflect surface contamination rather than true soil-to-plant transfer (IAEA, 2006), and/or
- The CR was based on stable Cs added to soil unless the dose was less than 0.001 μmol kg^{-1} substrate, and/or
- The soil was amended with minerals (e.g., zeolite, bentonites), and/or
- The nutrients in the soil were experimentally depleted before plant growth started, and/or
- The plants were trees or non-terrestrial (marine) plants, and/or
- Background information on the CR calculation method was absent, and/or
- The CR was based on modeled or inferred rather than measured radiocaesium concentrations in the soil and plant, and/or
- The CR was reported in a source (e.g., reviews) citing other sources documents, and/or
- The CR was non-paired (derived CR), i.e., radiocaesium concentrations in the plant were measured in a different substrate than the one in which the plant was growing, and/or

- Radiocaesium concentrations were not corrected for radioactive decay.

2.3. Database structure

Detailed background information behind each CR was collected from source documents to reflect possible site-specific effects of soil, plant, climate and experimental variables. This collation followed recommendations from reports (IAEA, 2021; IUR, 1982; FAO/IAEA, 2010).

The data was collected with a total of 100 attributes (or columns) that contained categorical and numerical information. These attributes were further organized and grouped into nine main categories (Table A.1): reference (i.e., source document), experiment site condition (i.e., experiment setup), caesium (isotope type), soil origin (sampling location), soil class, plant info, quality control, sampling, measurement (caesium concentrations and soil and plant properties), and plant growth condition (methodology). These nine designed main categories provide future scalability by allowing easy integration of novel datasets or expansion of existing datasets without compromising the structure of the collected data. The current database is not georelational because the paired soil-plant samples are not georeferenced.

For each CR value, all 100 attributes were documented. These attributes are further denoted as variables when used in models. The variables were classified into soil variables (e.g., *soil pH, clay content*), plant variables (e.g., *plant species, plant part*) or miscellaneous variables (e.g., *plant growth method, CR calculation method*). Each CR data point (in a row) in the database was completed with all available information from the original source documents, with missing data inferred from related information, such as *climate zone of soil* derived from soil sampling location using climate zone maps (section 2.4). Unique sample codes and source information were assigned to each row for traceability. Key variables affecting radiocaesium transfer were selected and used in all studied statistical and machine learning models.

2.4. Data harmonization

Data from different source documents were harmonized to ensure compatibility and comparability. This process included:

- Converting units (e.g., Ci to Bq, g to mol, mL to L);
- Converting plant material fresh weights to dry weights using default dry-to-fresh ratios (FAO/IAEA, 2010);
- Converting aggregated radiocaesium CR (CR_{agg}) expressed in m^2 soil per kg dry weight (DW) plant to CR expressed in kg DW soil per kg DW plant units by using the *soil depth* and *soil density* given in the source document or using an assumed 10 cm soil depth for grass and 20 cm for other plants (IAEA, 2009) and soil density of 1300 kg per m^3 (common for loamy soils; USDA, 2001) as follows:

$$CR\left(\frac{\frac{Bq}{kg\,DW\,plant}}{\frac{Bq}{kg\,DW\,soil}}\right) = CR_{agg}\left(\frac{\frac{Bq}{kg\,DW\,plant}}{\frac{Bq}{m^2\,soil}}\right) \cdot \frac{soil\,depth\,(cm)}{100} \cdot soil\,density\left(\frac{kg}{m^3}\right)$$

(1)

The variable *original CR expression method* indicates whether the CR was originally given as CR or whether it was converted as given above with given or assumed soil density;

- Converting soil humus content or soil organic matter to *soil organic carbon* by using a conversion factor of 0.58 (Pribyl, 2010);
- Converting exchangeable cations reported in terms of cation oxides such as mg K_2O per 100 g or mg CaO per 100 g to $cmol_c$ per kg;
- Soil pH values measured in 0.01 mol L^{-1} $CaCl_2$ or 1 mol L^{-1} KCl solution were used as equivalent numbers (with preference to soil pH in 0.01 mol L^{-1} $CaCl_2$) while pH measured in water (pH_{water}) was converted to *soil pH* using the calibrated line (Ahern et al., 1995):

Soil pH = $0.999 \cdot pH_{water} - 0.877$ 　　(2)

- Converting Soil Taxonomy classes (Soil Survey Staff, 1999) and soil classes from national soil classification systems (e.g., the Russian soil classification system) to the WRB (IUSS Working Group WRB FAO, 2022);
- If only minimum and maximum values were reported for the same plant species on similar soils, the average was calculated and entered into the database. If only mean values were reported for the same plant species on similar soils, the reported mean was entered into the database;
- Missing data on the climate zone (Köppen Climate Classification System; Beck et al., 2018) where the soil was sampled (*climate zone of soil*) was entered into the database based on the available, often approximate location where the soil was sampled, such as the city;
- Missing data on the contact time of fallout radiocaesium with substrate and time of plant harvest was entered in the database based on the most recent large-scale nuclear deposition event relative to the date of the study. Studies on radiocaesium deposition of fallout barely report contact time. Therefore, three reference dates were

chosen corresponding to the major accidental radiocaesium deposition events. On 05/08/1963, the Partial Test Ban Treaty was signed, greatly reducing nuclear weapons testing. This date was selected for fallout from nuclear weapons testing because radiocaesium deposition occurred over decades, peaked in the 1960s, and thus could not be traced to a single date. More recent events later overshadowed radiocaesium deposition of fallout from nuclear weapons testing. On 26/04/1986, the Chornobyl nuclear accident occurred, leading to the deposition of radiocaesium around the world. On 11/03/2011, the Fukushima accident caused radiocaesium deposition in east Asia;

- If the time of the experiment was missing, an average of 90 days between sowing and harvesting was assumed and entered into the database. This is a rough estimate because the growing period depends on the plant species and growing conditions. If only the year of harvest was given but not the exact date, it was assumed that harvesting took place on July 1 of the year if harvesting took place one year after radiocaesium contamination in the substrate.

2.5. Data analysis: models, regressions and machine learning

Statistical analyses were performed using JMP® Pro software (Version 17.0.0 SAS Institute Inc., Cary, NC, 1989–2023) or R Statistical Software. The data analysis steps are shown schematically for clarification (Fig. 2). Univariate analyses examined the relationships between single variables and the CR. Geometric means (GM) and their standard errors (SE) were calculated for CRs grouped by *climate zone of soil, plant category* (e.g., cereals, rice, grass, leafy vegetables), *soil category* (clay, loam, sand, organic, coral sand), and *dominant clay mineral*. Significant differences in CR within categories were tested using Dunn's non-parametric comparison for all pairs, at a 5 % significance level ($p < 0.05$). Variables were [10]log-transformed if skewness was >1 (except *pH*). Single linear regressions were applied to evaluate the effects of numeric variables on [10]log-transformed CR, while categorical variables were analyzed using one-way analysis of variance (ANOVA). Multivariate analysis with a random effect was performed using standard least squares multiple regression, accounting for interdependencies among variables. Variance components were estimated using the Restricted Maximum Likelihood (REML) method to improve the unbiased estimation of random effects.

Three semi-mechanistic models were evaluated: Absalom et al., 2001, Absalom 2011; Tarsitano et al., 2011 (Table 3). The Absalom et al., 2001 model (Absalom et al., 2001) estimates the CR based on Cs^+ concentration in the soil solution within the rooting zone, considering processes such as Cs^+-K^+-NH_4^+ competition, selective Cs adsorption in soil and fixation over time. That model was parameterized using pot trial data of grass grown on mineral and organic soils. The Absalom 2011 model (Tarsitano et al., 2011) retained the same structure but was recalibrated with a larger dataset including grass, barley, and wheat grown in laboratory and field conditions. The Tarsitano et al., 2011 model (Tarsitano et al., 2011) was a reduced version of Absalom et al. (2001), omitting soil pH as an input variable and recalibrated on the same larger dataset.

To compare model performance, subsets of the full database were selected that contained only paired, measured CR data where both plant and soil samples were collected from the same location, and radiocaesium concentrations were measured in both. These paired samples ensured that the soil sample was representative of the rooting zone relevant to the corresponding plant sample. Data points were rejected if a single soil sample was used to represent the entire field plot from which plant samples were taken. The three semi-mechanistic models (Absalom et al., 2001; Tarsitano et al., 2011) were compared with statistical and machine learning models using the same set of variables: *plant category, soil clay content* (g 100 g^{-1}), *soil organic matter content* (g 100 g^{-1}), *soil pH, exchangeable K content* ($cmol_c$ kg^{-1}), *NH₄ concentration in soil solution* (mmol L^{-1}), and *soil-radiocaesium contact time* (days). A subset of data points was selected for which all these variables were
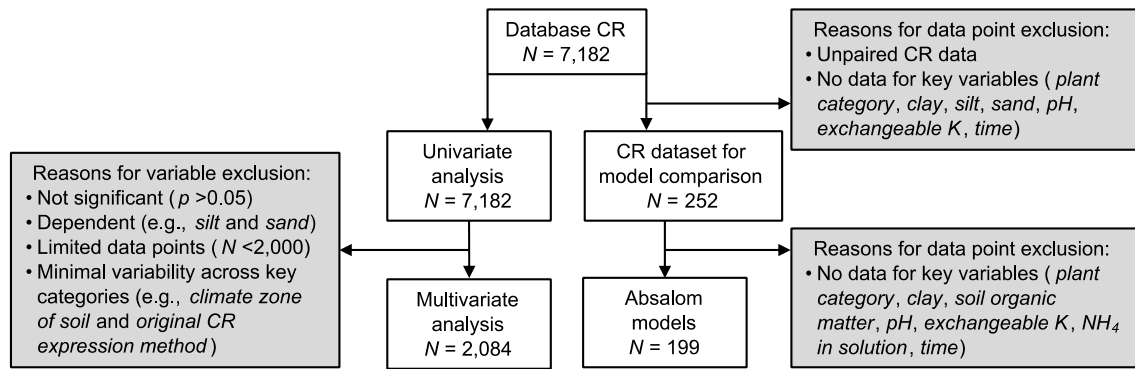
Fig. 2. Subsets of the CR database with justification, used for univariate and multivariate analyses (section 3.1) and modeling (comparison between Absalom and other models and CR estimation with mixed-effects and random forest models; section 3.2). N = number of data points. Grey boxes are variable and data selection steps.

known ($N = 199$).

The semi-mechanistic models were available for a single plant category (i.e., grass), so modeled CR needed to be converted for other plant categories, whereas this was not necessary for statistical and machine learning models that considered *plant category* as a variable. Plant conversion factors for plant species (Beresford and Willey, 2019) were grouped into plant categories, and their geometric means were calculated (Table A.5), because too few CRs were reported at the plant species level. These conversion factors (f_{plant}) were then used to convert CR modeled for grass ($CR_{model\ for\ grass}$) to other plant categories (CR_{plant}):

$$CR_{plant} = CR_{model\ for\ grass} \cdot f_{plant} \quad (3)$$

A more comprehensive data analysis was attempted, where data availability was less restrictive, using statistical and machine learning models. A subset of data was selected ($N = 252$) for which the following variables were known: *plant category*, soil texture (*clay, silt, sand*), *soil pH*, *exchangeable K content*, and *soil-radiocaesium contact time*.

The statistical models were linear mixed-effects regressions created using the package "lme4" (Bates et al., 2015) in R version 4.4.1 (R Core Team, 2024), including fixed effects (all variables except *source document*) and a random effect (*source document*) estimated using REML. The random effects covariance structure was modeled with constrained variance estimates using Cholesky decomposition and optimized via penalized quasi-likelihood (PQL). Dummy variables were used for *plant categories* with grass as the reference level. For model comparison, marginal R^2, representing the variance explained solely by the fixed effects, and conditional R^2, representing the variance explained by both the fixed and random effects (not shown), were calculated (Nakagawa and Schielzeth, 2013). These mixed-effects models were chosen to estimate CR as accurately as possible, modeling random effects explicitly to improve prediction, while the previously discussed multivariate models were appropriate to assess the significance of variables.

The machine learning models were created using random forest regression algorithms (Breiman, 2001) using the R package "random-Forest" (Liaw and Wiener, 2002). In total, 500 trees were used, and the dataset was split into a 70 % training set and a 30 % validation set. The models were set up to estimate ^{10}log-transformed CRs without transforming the variables. For this exercise, models were simplistic with no missing data, i.e., based on a limited number of data points and no cross-validation. Variable importance was checked using variable importance plots (Figure A.8). R^2, marginal R^2 in the case of mixed-effects models, and root mean square error (RMSE) values were calculated to compare model performances.

3. Results

3.1. Radiocaesium soil-to-plant transfer database

The database created contained 7,182 data points with information extracted from 139 source documents. Overall, the CRs ranged 7 orders of magnitude, i.e., from 0.000028 kg kg^{-1} to 380 kg kg^{-1} (Figure A.1). Even within the categorized climate zones, plant types, soil types, or dominant clay minerals, variability is high (5–7 orders of magnitude differences in CRs within the categories with $N > 100$).

3.2. Univariate analysis

The univariate analysis showed the most important variables that explained the ^{10}log-transformed CRs (Table 1). The *source document of the data points*, i.e., the study in which data were obtained, explained most of the CR. This may imply that methodological effects, despite efforts to harmonize data, and experimental design effects, such as *plant growth methods*, introduce considerable variability in CRs in addition to the expected soil- and plant-related effects.

The *NH$_4$ concentration in soil solutions* and *radiocaesium interception*

Table 1
Single variable effect sizes on ^{10}log-transformed CRs ($p < 0.05$), ranked by R^2. n. s. = not significant ($p > 0.05$). N = number of data points. Note that N differs largely among variables because these variables were often not present for different data points.

variable class	variable	N	R^2
miscellaneous	source document (i.e., reference)	7,182	0.62
soil	soil solution NH$_4$	98	0.46
soil	radiocaesium interception potential	356	0.40
soil	soil classification (WRB)	1,722	0.34
soil	climate zone of soil	5,870	0.27
plant	plant category (based on plant species)	7,182	0.25
soil	soil category (particle size class, organic soil, coral sand)	4,295	0.25
plant	plant part (analyzed compartment of plant)	7,182	0.24
soil	exchangeable K content	3,471	0.19
soil	mineralogy	378	0.15
miscellaneous	original CR expression method	7,182	0.11
soil	cation exchange capacity	3,186	0.08
miscellaneous	plant growth method	7,182	0.08
miscellaneous	land use where soil was sampled	2,476	0.08
soil	clay content	3,304	0.04
soil	sand content	3,068	0.04
soil	soil organic carbon content	3,257	0.02
miscellaneous	soil-radiocaesium contact time	4,923	0.01
soil	silt content	2,962	<0.01
soil	soil pH	3,983	n.s.

potential (RIP) in soils, both related to radiocaesium adsorption processes in soils, had significant and expected effects on CRs, but data were limited ($N < 500$). *Particle size category* (including the category organic soil), *exchangeable K content* and *clay content* of the soil also affected the CR due to their known role in radiocaesium adsorption in soil. *Soil classification* and *climates of which soils originated* had larger effects than *plant category* and *plant part*, suggesting soil variables had a larger effect than plant variables on the CR. The *original CR expression method*, either as CR (kg kg^{-1}) or CR$_{agg}$ (m^2 kg^{-1}), also played a significant role, suggesting that the conversion method used in this study, assuming a homogeneous contamination in the rooting zone, was inadequate.

The CRs measured in soils from different climates, grouped as tropical, arid, temperate or continental, showed significant differences (Fig. 3). The CRs measured in soils from tropical climates (grouped Af, Am, etc.) were significantly higher ($p < 0.05$, Dunn's test) than those from other climate groups (Figure A.2). While the CRs in temperate climates (grouped Cfa, Cfb, etc.) were significantly lower. Most CRs from temperate climates were measured on Japanese soils ($N = 1,663$ out of 2,814), which were significantly lower compared to CRs measured on soils from other locations (Figure A.5).

The CRs grouped by *plant category* showed no significant differences except rice that had significantly lower CR (Fig. 4). Similarly, CRs grouped by *plant part* did not exhibit a consistent trend except that shoots and leaves had significantly higher CRs than grains (Figure A.6). The CRs measured in organic and coral sandy soils (dominated by carbonates and not silicate clay, e.g., island and atoll soils) were significantly higher than in other soil categories, while clay and loamy soils had the lowest CR (Fig. 5). The CRs measured in soils dominated by allophane/imogolite and kaolinite were significantly larger compared to illite (Fig. 6). However, limited data were available on soil mineralogy ($N = 378$ out of 7,182).

The highest CRs (>50 kg kg^{-1}) were measured in coral soils from the Marshall Islands (Robison et al., 2006), in sandy soils from the south-west coast of Norway (Selnæs and Strand, 1992), and in peat soils from Northern Ireland (McGee et al., 1992). The CR$_{agg}$ of the sandy and peat soils, expressed in m^2 kg^{-1}, needed to be converted to CR, expressed in kg kg^{-1}, using Equation (1). A soil density of 1300 kg m^{-3} was assumed and was likely too low for sandy soil and too high for peat soils.

Indeed, the CR$_{agg}$ values measured in peat soils (0.038–1.3 m^2 kg^{-1}, GM = 0.20 m^2 kg^{-1}, $N = 35$, McGee et al., 1992) were the largest compared to all CR$_{agg}$ values in this database (0.000010–1.3 m^2 kg^{-1}, GM = 0.0018 m^2 kg^{-1}, $N = 861$), so the method by which the CR was obtained and the assumption by which data were harmonized affected the data analyses. CRs calculated from radiocaesium deposition data ($N = 1,045$) were significantly higher than CRs based on data from measured radiocaesium in soil ($N = 6,115$).

In addition, the plant growth methods under which the CR was obtained significantly affected the CR, with a ranking of field monitoring < field experiment < lysimeter and pot experiment (Figure A.3). The CRs from reconstituted substrates did not significantly differ from those in lysimeter and pot experiments. However, the CRs were not corrected for time (i.e., related to the duration of the experiment), which likely differed for each plant growth method, as clearly shown in Figure A.4, i. e., the contact time and plant growth method effects might be confounded (see below: multivariate analysis).

3.3. Multivariate analysis

In the first step, a standard least squares regression model was constructed using the Restricted Maximum Likelihood method. The *source document* (or reference, $N_{ref} = 139$ values) was used as a random effect. Indeed, a high multicollinearity (variance inflation factor = 5) was found between *source document* (linked to the experiment) and *contact time*, as experiments from the same setup were typically conducted during the same time period (Figure A.7). All significant variables that were identified with univariate analyses (Table 1) were entered as fixed categorical or numerical variables (Fig. 2). However, *silt* and *sand content* were removed, and the same was true by removing the variable *soil category* (clay, loam, sand, organic soil, coral sand) as these variables covary with *clay content* and *organic matter content*. The number of variables differed (Table 1), and hence, a final model with all these variables yielded a model that was based on a small number of data ($N = 46$) that had all these variables. Hence, a compromise was made by manually entering or removing variables to obtain at least 25 % of the data ($N = 2,084$). The variable *original CR expression method* was not entered as it had only one expression in several climate classes. The result of that first
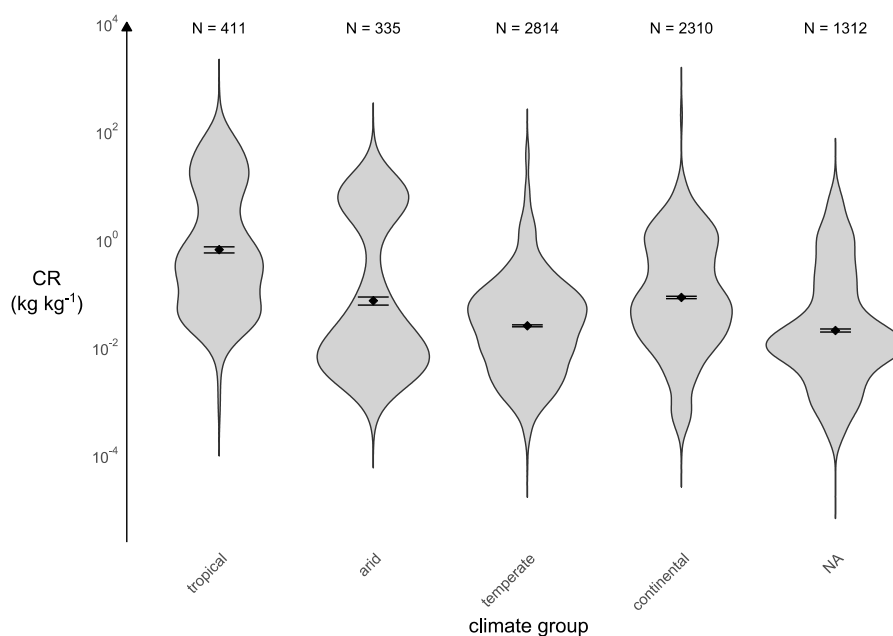

Fig. 3. Violin plots of CR on log-scale categorized by soils sampled in these climate zones show a significantly higher CR in tropical climates than in other climates. Diamonds show geometric means, and error bars are their standard errors. NA = data not available. Note that the number of data points per climate group is largely variable.
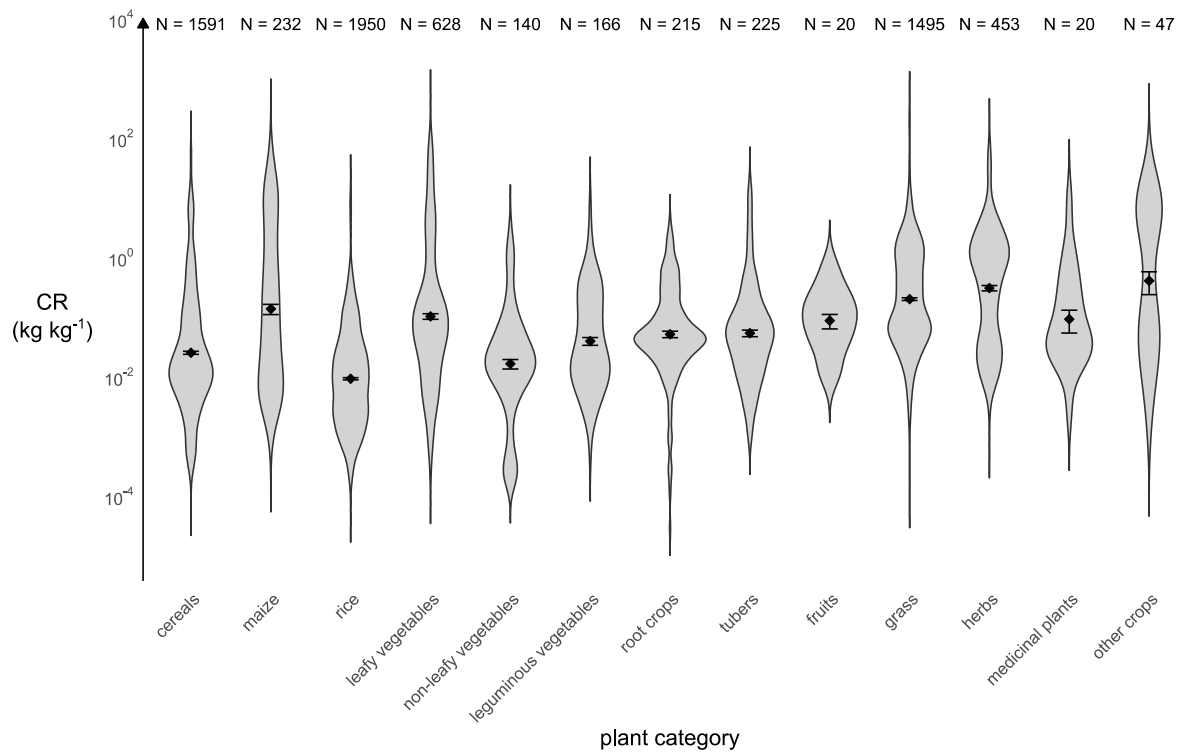
Fig. 4. Violin plots of CR on log-scale categorized by plant category. Diamonds show geometric means, and error bars are their standard errors.
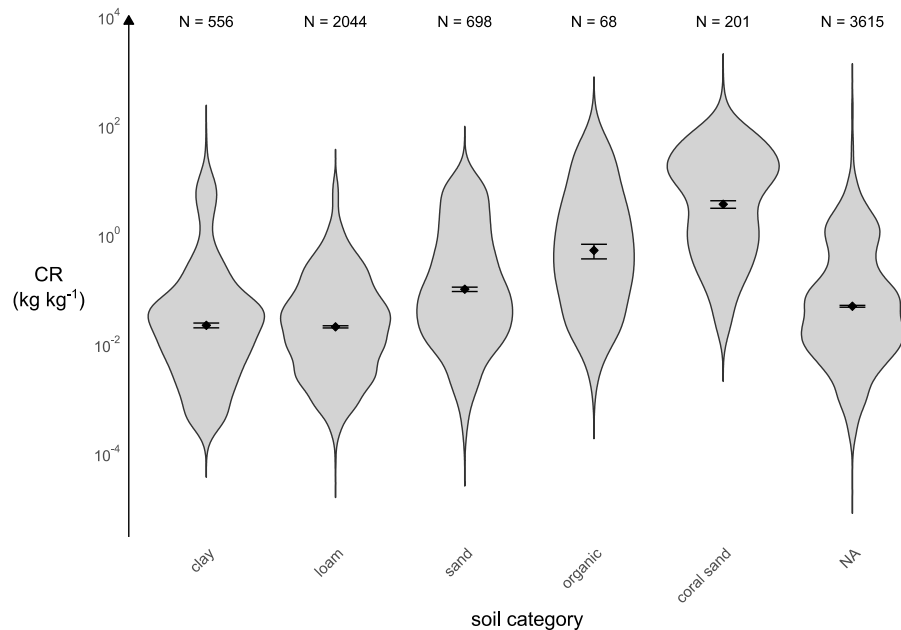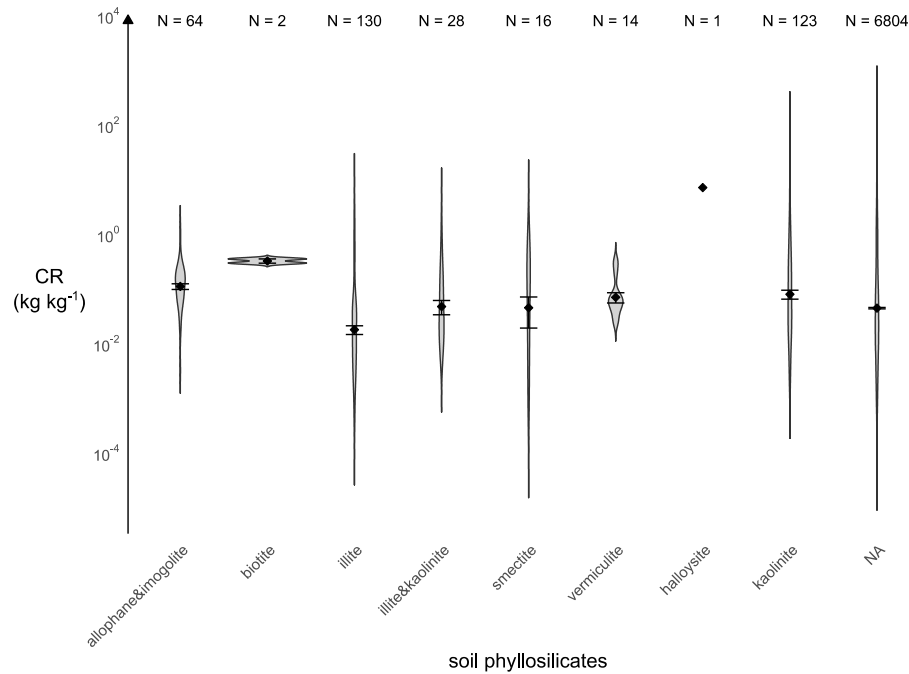

Fig. 5. Violin plots of CR on log-scale categorized by soil (texture) category show a significantly higher CR for organic and coral sandy soils (coral sands are dominated by carbonates and not silicate clay, e.g., island and atoll soils). Diamonds show geometric means, and error bars are their standard errors. NA = data not available.

analysis (Table 2) yielded a model that explained 66 % of the variation ($N = 2,084$) and showed that variables affecting the ^{10}log CR ranked *plant part > plant category > soil pH > climate zone of soil > contact time > exchangeable K content*. The effects of plant variables *plant part* and *plant category* were more important than soil variables in explaining the ^{10}log CR in the multivariate analysis. *Soil pH* remarkably ranked high here, whereas univariate analyses showed only poor effects of pH. Hence, *soil*

pH is likely a proxy for other factors affecting radiocaesium bioavailability, e.g., more weathered soils, often devoid of clay minerals that adsorb radiocaesium and are richer in oxides, typically have lower pH. The *climate zone of the soil* is significant, likely due to the leverage effect of the outliers described above (tropical soils, such as coral sands). The effect of the *plant growth methods* (pot trials, field monitoring; Figure A.3) was less significant than the effect of *contact time*, corroborating the

Fig. 6. Violin plots of CR on log-scale categorized by dominant phyllosilicate in soil show a significantly lower CR for illite compared to allophane/imogolite and kaolinite. Diamonds show geometric means (except for halloysite, arithmetic mean), and error bars are their standard errors. NA = data not available. Note that the number of data points on mineralogy is very limited.

Table 2
Comparison of the variable effect sizes on ^{10}log-transformed CRs in a multivariate regression model with *source document* (the reference) as random effect. The adjusted R^2, i.e., adjusting for the number of variables in the model, the slope of the effect (positive or negative relation), and the log-transformation of *p*-values (logworth) are given. Larger values of the logworth indicate higher statistical significance in explaining the CR. n.s. = not significant ($p > 0.05$).

N	R^2_{adj}	variable	slope	logworth	p-value
2,084	0.66	plant part (analyzed compartment of plant)		12	<0.0001
		plant category (based on plant species)		12	<0.0001
		soil pH	–	10	<0.0001
		climate zone of soil		9.1	<0.0001
		^{10}log(soil-radiocaesium contact time)	–	5.3	<0.0001
		^{10}log(exchangeable K content)	–	3.8	0.0002
		^{10}log(cation exchange capacity)	–	2.1	0.0072
		^{10}log(soil organic carbon content)	+	1.8	0.017
		plant growth method		1.4	0.035
		clay content			n.s.

speculation above. Most strikingly, the *clay content* was not significant.

3.4. Model comparison

In a second step, different models were compared (Table 3). For a small subset of data ($N = 199$) where all soil variables required for the semi-mechanistic models (Absalom et al., 2001, Absalom 2011; Tarsitano et al., 2011) were available, the semi-mechanistic models performed similarly well as random forest models but slightly less good as the a statistical mixed-effects model (*source document* as random effect, other variables as fixed effects).

In a subsequent step, we attempted to identify better-performing models using a slightly larger dataset ($N = 252$ instead of 199) that included key variables such as *plant category*, *clay*, *silt*, *sand*, *soil pH*, *exchangeable K*, and *contact time*. While the mixed-effects model and random forest model based on this slightly larger dataset showed improved performance (Figure A.8) compared to the previously tested semi-mechanistic models, their effectiveness was constrained by the

limited size and coverage of datasets across *climate regions* (tropical $N = 16$, arid $N = 2$, continental $N = 8$) and *plant categories* (grass $N = 139$, leafy vegetables $N = 26$, non-leafy vegetables $N = 3$, rice $N = 41$, root crops $N = 40$, tubers $N = 3$). Additionally, data scarcity for variables, such as soil organic carbon ($N = 214$), further limited model performance.

4. Discussion

4.1. Radiocaesium soil-to-plant transfer database

Harmonizing radiocaesium soil-to-plant transfer data is challenging due to substantial variability introduced by methodological and experimental design factors, which can confound the expected soil- and plant-related effects. In radiocaesium transfer studies, the objectives and experimental setups vary (e.g., field vs. pot experiments), leading to differences in measurement units, compartments studied, available information, etc. Moreover, harmonizing these datasets requires certain

Table 3

Comparison of semi-mechanistic, log-linear multiple regression, and random forest ($p < 0.05$) to estimate the ^{10}log-transformed CR. The R^2 and RMSE (^{10}log(kg kg^{-1})) of the best performing model are given in bold. The reference level for dummy variables for *plant category* was grass in regression equations. Standard errors of estimates are in brackets and the normal distribution of random intercepts (i.e., 'N(x,y)') are given in regression equations.

type	model[a]	N	R^2	RMSE
semi-mechanistic	Absalom et al., 2001 based on plant category, clay, orgC, pH, exchK, solutionNH$_4$, time.	199	0.50	0.78
semi-mechanistic	Absalom 2011 based on plant category, clay, orgC, pH, exchK, solutionNH$_4$, time.	199	0.46	0.76
semi-mechanistic	Tarsitano et al., 2011 based on plant category, clay, orgC, exchK, solutionNH$_4$, time.	199	0.42	0.79
multiple regression	^{10}log(CR) = 0.77($\pm$0.66) + 0.00026($\pm$0.0035)$\cdot$clay + 0.26($\pm$0.12)$\cdot$^{10}log(orgC) − 0.32($\pm$0.05)$\cdot$pH − 1.0($\pm$0.1)$\cdot$^{10}log(exchK) + 0.060($\pm$0.027)$\cdot$solutionNH$_4$ − 0.21($\pm$0.28)$\cdot$^{10}log(time) + 0.11($\pm$0.32)plant$_{\text{leafy vegetables}}$ + 0.19($\pm$0.28)plant$_{\text{rice}}$ − 0.24($\pm$0.31)plant$_{\text{root crops}}$ + N(0, 0.059)	199	**0.58**[b]	**0.61**
machine learning	Random forest based on plant category, clay, orgC, pH, exchK, solutionNH$_4$, time.	199	0.51[c]	0.67

[a] clay = clay content, orgC = soil organic carbon content, pH = soil pH, exchK = exchangeable potassium content, solutionNH$_4$ = NH$_4$ concentration in soil solution, time = soil-radiocaesium contact time;

[b] marginal R^2.

[c] R^2 was calculated on the validation set that contained 30 % of the data points (i.e., $N = 60$).

assumptions that may inadvertently introduce uncertainties. By considering the variability introduced by the study and harmonization process, the effects of soil and plant characteristics on CR can be revealed, as shown in the multivariate analysis (Table 2). Furthermore, extracting and filtering relevant data from these studies is time-consuming and prone to errors. With many studies found but not reviewed ($N > 3,000$), ongoing efforts to gather and harmonize data are important for improving the reliability of radiocaesium transfer models.

A rather sensitive assumption is required to convert the area-based CR (m^2 per kg), i.e., aggregated CR to the mass-based CR where depth of soil sampling and bulk density data are often missing. For example, peat soils, high in organic matter, typically have bulk densities much lower than the assumed 1,300 kg per m^3, whereas highly weathered soils, dominated by iron oxides, often exceed this density. Such conversions can substantially overestimate CRs for organic soils or underestimate them for highly weathered soils.

In this study, the database reveals a wide range of CRs of more than seven orders of magnitude (Figure A.1). Most of these data are primarily post-accidental and predominantly biased towards temperate climate regions in Europe and Japan (Figure A.5). Our results confirm that there is a data gap for other regions such as tropical and arid climates, where radiocaesium bioavailability can be markedly different due to environmental conditions (Fig. 3). Radioecological studies in these regions have historically been deprioritized due to lower radiocaesium deposition rates, yet the variability in CRs in these areas can be substantial. High CRs are often observed in tropical regions, particularly coral sandy soils common on islands and atolls (Fig. 5), with limited capacity to adsorb radiocaesium. Only limited information is available on soil mineralogy (Fig. 6), even though it is crucial in determining the extent of radiocaesium adsorption in soils. For this database, the significance of differences in CRs among plant categories is limited (Fig. 4).

4.2. Univariate and multivariate analyses

There are notable differences in CRs between plant growth methods, with pot trials yielding higher CRs than field trials (Figure A.3). This difference was previously attributed to factors such as more K depletion in limited-volume pots, resulting in less Cs$^+$-K$^+$ competition for plant uptake and to the limited radiocaesium contact time with soil after spiking (Frissel et al., 1990; IAEA, 2006). However, when corrected for random variation of the source document (or study), for time and other soil and plant variables, the *plant growth method* is of less importance (Table 2), as previously disputed (Gerzabek et al., 1998; Horak and Gerzabek, 1989; Strebl et al., 2007).

The bioavailability of radiocaesium in soils decreases over time,

which is confirmed here by the negative relation of the *soil-radiocaesium contact time* and CR (Table 2). The aging of radiocaesium in soils depends on both the clay mineralogy and structure of soils. Radiocaesium is strongly adsorbed in micaceous clay minerals (high RIP) (Cremers et al., 1988) where it gradually migrates to clay interlayers (Fuller et al., 2015; Krouglov et al., 1997) and becomes less bioavailable. In contrast, organic-rich soils in cooler climates, such as peat soils, often lack clay minerals (low RIP) and have low K content, so that radiocaesium remains more bioavailable (Sanchez et al., 1999). Soil structure also plays a crucial role (Fernagut and Merckx, 1997). In field conditions, radiocaesium is deposited on undisturbed soils where Cs$^+$ may initially remain mobile because clay minerals are cemented in aggregates that limit access to selective adsorption sites, delaying fixation in the soil.

The multivariate regression model indicates that *soil pH* is the most important soil variable explaining CRs (Table 2), but not in explaining the variation of CR as a single variable (Table 1). This difference highlights how multivariate models can capture interactions between variables and account for indirect effects that univariate analyses may overlook. Some studies report a negative effect of soil pH on radiocaesium plant uptake (e.g., Ogura et al., 2014), while others report no effect (e.g., Keum et al., 2007; Van Bergeijk et al., 1992). It is possible that this statistical effect of pH is not a causal effect, but that *soil pH* may be a proxy for other properties that have an expected (mechanisms known) effect on CR. For example, pH correlates positively with RIP and is explained by the parent material and its weathering products (Waegeneers et al., 1999).

A striking outcome of the multivariate regression model is the lack of an effect of *soil clay content* on CRs after correcting for other factors. On the one hand, this might be related to the correction for *cation exchange capacity* and *organic carbon content* which are covarying with clay content. On the other hand, it is still striking as it contrasts the paradigm in many models and studies proposing *clay content* as the most important soil variable explaining radiocaesium bioavailability (e.g., Skarlou et al., 1996; Squire and Middleton, 1966). Yet, this striking observation also corroborates the findings of comparative studies showing no effect of *clay content* on radiocaesium bioavailability (Keum et al., 2007; Smolders et al., 1997; Vanheukelom et al., 2023, 2024). This failure is likely due to the various types of clay minerals and Cs$^+$ adsorption sites on those clay minerals that are not differentiated by mere particle size (Sandalls and Bennett, 1992; Vanheukelom et al., 2024). Moreover, assuming the same mineralogy, an increase in clay content has two counteracting effects (Smolders et al., 1997): a higher RIP due to more selective adsorption sites for radiocaesium, which decreases bioavailability, but also a lower K concentration in soil solution due to the higher cation exchange capacity, which increases radiocaesium uptake. These

counteracting effects may explain why clay content alone does not consistently estimate radiocaesium bioavailability.

4.3. Radiocaesium soil-to-plant transfer estimation

Grouping plants into categories generalizes the effect of plant type on the CR. In this database, the significance of differences in CRs among plant categories is limited (Fig. 4). While investigating plant species effects on CR would be valuable, sufficient data points are currently lacking. The semi-mechanistic model requires plant categories as input variables, so we derived conversion factors (Beresford and Willey, 2019) that can convert the standard estimated CR for grass to other plant categories, such as rice. These conversion factors were compared with previously reported factors (Table A.5), but they do not agree. It is questionable whether plant conversion factors representing average values at the category level capture variability in the CR. Still, plant variables, such as *plant part* and *plant categories*, significantly affected CRs (Tables 1 and 2). So, plant properties have an important effect on CRs, but so do other variables. This highlights the importance of comparing plants grown in similar conditions on the same substrate (Beresford and Willey, 2019; Frissel et al., 2002), particularly regarding nutrient (K) availability to clarify the effect of plant properties on radiocaesium uptake.

The performance of radiocaesium soil-to-plant transfer models depends on the dataset size (Table 3). Semi-mechanistic models (Absalom et al., 2001; Tarsitano et al., 2011), although not calibrated to this dataset, performed reasonably well with a small dataset ($N = 199$). Mixed-effects regression and, to a smaller extent, random forest models calibrated to the small dataset performed slightly better. Random forest models generally benefit from larger datasets, with a rule of thumb suggesting a dataset size of approximately 200 times the number of variables (Van Der Ploeg et al., 2014), i.e., for 7 variables $N = 1,400$, well above the number of high quality data here. This allows random forest models to reduce overfitting and capture complex interactions, leading to better generalization across diverse conditions.

Recent studies using random forest models for CR estimations in Germany ($N = 1,205$; 1980–2004; Urso et al., 2023) and the Fukushima Prefecture, Japan ($N = 132$; 05/2011–11/2012; Shuryak, 2022), confirm the importance of key variables such as *plant species*, *plant part*, *exchangeable K*, *soil pH*, and *time*. For Germany, random forest models outperformed semi-mechanistic models (Absalom et al., 2001; Tarsitano et al., 2011), even though in the latter the authors had not included time, NH_4 concentration in soil solution or plant category effects (Urso et al., 2023).

Model uncertainties remain substantial and vary with dataset coverage and size. Assuming normally distributed errors (degrees of freedom is > 100), the 95 % confidence interval of CR predictions is $\pm 1.96 \times$ RMSE. For instance, the random forest model in this study has an RMSE of 0.67 for ^{10}log-transformed CR (Table 3). This implies that, for a predicted CR value of CR $= 0.1$ kg kg^{-1} (or ^{10}log(CR) $= -1$ ^{10}log(kg kg^{-1})), we expect that about 95 % of the ^{10}log(CR) predictions will fall within the interval $[-2.3$ ^{10}log(kg kg^{-1}); 0.31 ^{10}log(kg kg^{-1})] or for CR [0.0049 kg kg^{-1}; 2.1 kg kg^{-1}]. Thus, the range of possible CR values reflects a large uncertainty in the random forest model predictions, which in this study spans over a factor of 423, compared with factor 35 (Germany; Urso et al., 2023) and factor 6.6 (Fukushima; Shuryak, 2022) in other recent studies. These lower uncertainties likely reflect the reduced variability in soil and crop types within these regional datasets compared to the global dataset used in this study.

We suggest collecting and harmonizing data in undersampled regions, including information on soil K content and clay mineralogy, and critically validating radiocaesium CR models to make them suitable for impact assessments. Although coordinate information of sampling locations has not been collected in the database because it was not available in many studies, georeferenced data points would enable spatially explicit modeling, such as Digital Soil Mapping, which would improve the applicability of these models for regional and global impact assessments.

5. Conclusion

This study aims to support impact assessments with estimates of radiocaesium transfer from soil to crops on a global scale, moving beyond generic transfer factors. A database was constructed encompassing a large share of published radiocaesium soil-to-plant transfer data and associated variables which allows to develop new and evaluate existing radiocaesium transfer models. The findings confirm that radiocaesium transfer is governed by well-known mechanisms controlled by plant and soil properties, competing cation concentrations (K^+ and NH_4^+) in soil and soil-radiocaesium contact time. While soil clay content appears to have limited influence on radiocaesium transfer, clay mineralogy (linked to weathering processes) and exchangeable K are likely more critical. The latter is especially important in soils with barely any clay minerals (e.g., sandy and peaty soils). The role of the climate zone (where soil was sampled) underscores the need for more data points from diverse climatic regions to improve model applications on a global scale.

Radiocaesium transfer models benefit from larger datasets that include a wide range of variables and better geographical coverage. Data gaps remain, including limited data points from tropical and arid climates, limited information on soil exchangeable K content and clay mineralogy, and limited georeferenced data points suitable for spatially explicit modeling, such as Digital Soil Mapping. Future research should focus on the influence of soil structure and mineralogy on radiocaesium bioavailability in field conditions, particularly in tropical climates, where little is known about the long-term behavior of radiocaesium.

CRediT authorship contribution statement

Margot Vanheukelom: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mark Mng'ong'o:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation. **Floris Abrams:** Writing – review & editing, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Surya Gupta:** Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Talal Almahayni:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Lieve Sweeck:** Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Jos Van Orshoven:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Erik Smolders:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

librarians of KU Leuven Libraries – 2Bergen (Leuven, Belgium), for their help conducting the systematic literature search. The authors thank Neil Willey of Centre for Research In Bioscience (UWE Bristol), for sharing insights regarding plant conversion factors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvrad.2025.107704.

Data availability

I have shared the link to my data https://data.mendeley.com/datasets/767syw9xsd/2.

References

Absalom, J.P., Young, S.D., Crout, N.M.J., Nisbet, A.F., Woodman, R.F.M., Smolders, E., Gillett, A.G., 1999. Predicting soil to plant transfer of radiocesium using soil characteristics. Environ. Sci. Technol. 33, 1218–1223. https://doi.org/10.1021/es9808853.

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. https://doi.org/10.1016/S0265-931X(00)00098-9.

Ahern, C.R., Baker, D.E., Aitken, R.L., 1995. Models for relating pH measurements in water and calcium chloride for a wide range of pH, soil types and depths. Plant Soil 171, 47–52. https://doi.org/10.1007/BF00009563.

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. https://doi.org/10.1016/j.jenvrad.2019.01.006.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J Stat Softw 67, 1–48. https://doi.org/10.18637/jss.v067.i01.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future köppen-geiger climate classification maps at 1-km resolution. Sci. Data 5, 1–12. https://doi.org/10.1038/sdata.2018.214.

Beresford, N.A., Willey, N., 2019. Moving radiation protection on from the limitations of empirical concentration ratios. J. Environ. Radioact. 208–209, 106020. https://doi.org/10.1016/j.jenvrad.2019.106020.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brimo, K., Pourcelot, L., Métivier, J.M., Gonze, M.A., 2021. Evaluation of semi-mechanistic models to predict soil to grass transfer factor of 137Cs based on long term observations in French pastures. J. Environ. Radioact. 227. https://doi.org/10.1016/j.jenvrad.2020.106467.

Cremers, A., Elsen, A., Preter, P. De, Maes, A., 1988. Quantitative analysis of radiocaesium retention in soils. Nature 335, 247–249. https://doi.org/10.1038/335247a0.

Fernagut, B., Merckx, R., 1997. De Beschikbaarheid Van Radiocesium in Graslandbodems.

Frissel, M.J., Deb, D.L., Fathony, M., Lin, Y.M., Mollah, A.S., Ngo, N.T., Othman, I., Robison, W.L., Skarlou-Alexiou, V., Topcuoğlu, S., Twining, J.R., Uchida, S., Wasserman, M.A., 2002. Generic values for soil-to-plant transfer factors of radiocesium. J. Environ. Radioact. 58, 113–128. https://doi.org/10.1016/S0265-931X(01)00061-3.

Frissel, M.J., Noordijk, H., Bergeijk, K.E. van, 1990. The Impact of Extreme Environmental Conditions, as Occurring in Natural Ecosystems, on the Soil-To-Plant Transfer of Radionuclides. Elsevier Applied Science, United Kingdom.

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. https://doi.org/10.1016/j.clay.2015.02.008.

Gerzabek, M.H., Strebl, F., Temmel, B., 1998. Plant uptake of radionuclides in lysimeter experiments. Environmental Pollution 99, 93–103. https://doi.org/10.1016/S0269-7491(97)00167-X.

Gilis, B., Sweeck, L., 2019. Bodem-plant transfer van radioactief cesium-137 Transfer naar spinazie en radijzen. Thomas More, Geel, p. 58.

Horak, O., Gerzabek, M.H., 1989. Österreichisches Forschungszentrum 4515.

IAEA, 2021. Soil–plant transfer of radionuclides in non-temperate environments. TECDOC Series. INTERNATIONAL ATOMIC ENERGY AGENCY, Vienna.

IAEA, 2020. Environmental transfer of radionuclides in Japan following the accident at the Fukushima daiichi nuclear power plant. TECDOC Series. INTERNATIONAL ATOMIC ENERGY AGENCY, Vienna.

IAEA, 2009. Quantification of Radionuclide Transfer in Terrestrial and Freshwater Environments for Radiological Assessments, TECDOC Series. INTERNATIONAL ATOMIC ENERGY AGENCY, Vienna.

IAEA, 2006. Classification of soil systems on the basis of transfer factors of radionuclides from soil to reference plants. TECDOC Series. INTERNATIONAL ATOMIC ENERGY AGENCY, Vienna. https://doi.org/10.1051/radiopro/2002097.

IUR, 1992. VIIIth Report of the Working Group Soil-To-Plant Transfer Factors. Balen, Belgium.

IUR, 1982. Report IUR Working Group Soil-To-Plant Transfer. Part I. Wageningen.

IUSS Working Group WRB FAO, 2022. World Reference Base for Soil Resources. International Soil Classification System for Naming Soils and Creating Legends for Soil Maps, fourth ed. International Union of Soil Sciences (IUSS)21, Vienna.

Joint FAO/IAEA Division of Nuclear Techniques in Food and Agriculture, V. (Austria), Values, P., Transfer, R., Environments, F., 2010. Handbook of Parameter Values for the Prediction of Radionuclide Transfer in Terrestrial and Freshwater Environments. IAEA, International Atomic Energy Agency (IAEA).

Keum, D.K., Lee, H., Kang, H.S., Jun, I., Choi, Y.H., Lee, C.W., 2007. Predicting the transfer of 137Cs to rice plants by a dynamic compartment model with a consideration of the soil properties. J. Environ. Radioact. 92, 1–15. https://doi.org/10.1016/j.jenvrad.2006.08.007.

Krouglov, S.V., Filipas, A.S., Alexakhin, R.M., Arkhipov, N.P., 1997. Long-term study on the transfer of 137Cs and 90Sr from Chernobyl-contaminated soils to grain crops. J. Environ. Radioact. 34, 267–286. https://doi.org/10.1016/0265-931X(96)00043-4.

Li, P., Gong, Y., Tanaka, T., Thiry, Y., Huang, Q., Komatsuzaki, M., 2024. Modeling long-term transfers of radiocesium in farmland under different tillage and cover crop treatments. Sci. Total Environ. 907, 167849. https://doi.org/10.1016/j.scitotenv.2023.167849.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R. News 2, 18–22.

MAFF, 2017. Annual Report of Ministry of Agriculture, Forestry and Fisheries Affiliated Radioactivity Research: H26 Report. Tokyo.

MAFF, 2015. Annual Report of Ministry of Agriculture, Forestry and Fisheries Affiliated Radioactivity Research: H25 Report. Tokyo.

MAFF, 2014. Annual Report of Ministry of Agriculture, Forestry and Fisheries Affiliated Radioactivity Research: H24 Report. Tokyo.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma. https://doi.org/10.1016/S0016-7061(03)00223-4.

McGee, E.J., Colgan, P.A., Dawson, D.E., Rafferty, B., O'Keeffe, C., 1992. Effects of topography on caesium-137 in montane peat soils and vegetation. Analyst 117, 461–464. https://doi.org/10.1039/an9921700461.

Nakagawa, S., Schielzeth, H., 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol. Evol. 4, 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x.

Nisbet, A.F., Woodman, R.F.M., Haylock, R.G.E., 1999. Recommended Soil-To-Plant Transfer Factors for Radiocaesium and Radiostrontium for Use in Arable Systems. National Radiological Protection Board.

Ogura, S.I., Suzuki, T., Saito, M., 2014. Distribution of radioactive cesium in soil and its uptake by herbaceous plants in temperate pastures with different management after the Fukushima Dai-Ichi Nuclear Power Station accident. Soil Sci. Plant Nutr. 60, 790–800. https://doi.org/10.1080/00380768.2014.954269.

Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A., 2016. Rayyan-a web and mobile app for systematic reviews. Syst. Rev. 5, 1–10. https://doi.org/10.1186/s13643-016-0384-4.

Pribyl, D.W., 2010. A critical review of the conventional SOC to SOM conversion factor. Geoderma. https://doi.org/10.1016/j.geoderma.2010.02.003.

R Core Team, 2024. R: A Language and Environment for Statistical Computing.

Robison, W.L., Hamilton, T.F., Conrado, C.L., Kehl, S., 2006. Uptake of Caesium-137 by Leafy Vegetables and Grains from Calcareous Soils. International Atomic Energy Agency (IAEA).

Sanchez, A.L., Wright, S.M., Smolders, E., Naylor, C., Stevens, P.A., Kennedy, V.H., Dodd, B.A., Singleton, D.L., Barnett, C.L., 1999. High plant uptake of radiocesium from organic soils due to Cs mobility and low soil K content. Environ. Sci. Technol. 33, 2752–2757. https://doi.org/10.1021/es990058h.

Sandalls, J., Bennett, L., 1992. Radiocaesium in upland herbage in Cumbria, UK: a three year field study. J. Environ. Radioact. 16, 147–165. https://doi.org/10.1016/0265-931X(92)90013-J.

Selnæs, T.D., Strand, P., 1992. Comparison of the uptake of radiocaesium from soil to grass after nuclear weapons tests and the chernobyl accident. Analyst 117, 493–496. https://doi.org/10.1039/an9921700493.

Shuryak, I., 2022. Machine learning analysis of 137Cs contamination of terrestrial plants after the Fukushima accident using the random forest algorithm. J. Environ. Radioact. 241, 106772. https://doi.org/10.1016/j.jenvrad.2021.106772.

Skarlou, V., Papanicolaou, E.P., Nobeli, C., 1996. Soil to plant transfer of radioactive cesium and its relation to soil and plant properties. Geoderma 72, 53–63. https://doi.org/10.1016/0016-7061(96)00011-0.

Smolders, E., Van den Brande, K., Merckx, R., 1997. Concentrations of 137 Cs and K in soil solution predict the plant availability of 137 Cs in soils. Environ. Sci. Technol. 31, 3432–3438. https://doi.org/10.1021/es970113r.

Soil Survey Staff, 1999. Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys, second ed. Natural Resources Conservation Service, Washington DC, USA. https://doi.org/10.1007/BF01574372. U.S. Department of Agriculture Handbook 436.

Squire, H.M., Middleton, L.J., 1966. Behaviour of cs137 in soils and pastures a long term experiment. Radiat. Bot. 6, 413–423. https://doi.org/10.1016/s0033-7560(66)80074-1.

Strebl, F., Ehlken, S., Gerzabek, M.H., Kirchner, G., 2007. Behaviour of radionuclides in soil/crop systems following contamination. Radioact. Environ. 10, 19–42. https://doi.org/10.1016/S1569-4860(06)10002-9.

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. https://doi.org/10.1016/j.jenvrad.2010.11.017.

Uematsu, S., Vandenhove, H., Sweeck, L., Van Hees, M., Smolders, E., 2018. Radiocaesium bioavailability to flooded paddy rice is related to soil solution

radiocaesium and potassium concentrations. Plant Soil 428, 415–426. https://doi.org/10.1007/s11104-018-3686-6.

Uematsu, S., Vandenhove, H., Sweeck, L., Van Hees, M., Wannijn, J., Smolders, E., 2016. Variability of the soil-to-plant radiocaesium transfer factor for Japanese soils predicted with soil and plant properties. J. Environ. Radioact. 153, 51–60. https://doi.org/10.1016/j.jenvrad.2015.12.012.

Urso, L., Petermann, E., Gnädinger, F., Hartmann, P., 2023. Use of random forest algorithm for predictive modelling of transfer factor soil-plant for radiocaesium: a feasibility study. J. Environ. Radioact. 270. https://doi.org/10.1016/j.jenvrad.2023.107309.

USDA, 2001. Soil quality test kit guide. Soil Quality Institute 82. https://doi.org/10.1037/t15144-000.

Van Bergeijk, K.E., Noordijk, H., Lembrechts, J., Frissel, M.J., 1992. Influence of pH, soil type and soil organic matter content on soil-to-plant transfer of radiocesium and -strontium as analyzed by a nonparametric method. J. Environ. Radioact. 15, 265–276. https://doi.org/10.1016/0265-931X(92)90062-X.

Van Der Ploeg, T., Austin, P.C., Steyerberg, E.W., 2014. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med. Res. Methodol. 14, 1–13. https://doi.org/10.1186/1471-2288-14-137.

Vanheukelom, M., Sweeck, L., Almahayni, T., De Bruyn, M., Steegmans, P., Fondu, L., Van Gompel, A., Van Hees, M., Wannijn, J., Smolders, E., 2024. Highly weathered mineral soils have highest transfer risk of radiocaesium contamination after a nuclear accident: a global soil-plant study. Sci. Total Environ., 173583 https://doi.org/10.1016/j.scitotenv.2024.173583.

Vanheukelom, M., Sweeck, L., Van Hees, M., Weyns, N., Van Orshoven, J., Smolders, E., 2023. Quantitative clay mineralogy predicts radiocesium bioavailability to ryegrass grown on reconstituted soils. Sci. Total Environ. 873, 162372. https://doi.org/10.1016/J.SCITOTENV.2023.162372.

Waegeneers, N., Smolders, E., Merckx, R., 1999. A statistical approach for estimating the radiocesium interception potential of soils. J. Environ. Qual. 28, 1005–1011. https://doi.org/10.2134/jeq1999.00472425002800030034x.