



Comparison of mechanistic and PLS-based regression models to predict radiocaesium distribution coefficients in soils

C.J. Gil-García, A. Rigol, M. Vidal*

Departament de Química Analítica, Universitat de Barcelona, Martí i Franqués 1-11, 3^a Planta, 08028 Barcelona, Spain

ARTICLE INFO

Article history:

Received 10 April 2011

Received in revised form

12 September 2011

Accepted 13 September 2011

Available online 17 September 2011

Keywords:

Soil

Sorption

Radiocaesium interception potential

Distribution coefficient

Modelling

ABSTRACT

Radiocaesium sorption interaction descriptors were examined in 30 soil samples from Spain. Mechanistic and regression models were used to predict the solid–liquid distribution coefficients of radiocaesium ($K_d(\text{Cs})$) based on soil properties, and the obtained values were compared with the experimental ones, which were derived from batch experiments. The batch experiments used two contact solutions: one simulated the composition of the soil solution, and the other was the wash-off from the soil. Several mechanistic models of different complexity were tested based on the Radiocaesium Interception Potential (RIP), with satisfactory agreement between experimental and predicted values. A simplified model based on either the RIP, or the clay content and K status of the soil was proposed. Various multivariate regression models, which were constructed using the Partial Least Square Regression (PLS), were also evaluated. The RIP, clay content, and the K and NH_4^+ contents were also identified by the regression models as the most relevant soil parameters to predict the K_d . As seen for the mechanistic models, the goodness of fit of the regression models was demonstrated by an excellent agreement between experimental and predicted values.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Understanding the behaviour of radionuclides in the environment is necessary to predict the impact of radioactive contamination and to assess its derived risk. The mobility of radionuclides such as radiocaesium within the environment is mainly controlled by their sorption in soils and by the reversibility of this process. Radiocaesium has been the focus of a large number of environmental studies because it was the most relevant radionuclide that affected European ecosystems after the Chernobyl accident [1]. Numerous studies in the literature have focused on describing the mechanisms that govern radiocaesium sorption in soils and their subsequent mobility in the soil–plant system [2–4]. It has been shown that radiocaesium sorption is controlled by ionic exchange at two types of sites with varying sorption affinities: the high-affinity frayed edge sites (FES), which are inter-lattice sites that are found at the end of expanded clay layers, and the low-affinity regular exchange sites (RES), which are found in organic matter phases and at external positions in clay minerals. With the exception of soils with a high organic matter content or a negligible 2:1 phyllosilicate content, the FES govern radiocaesium sorption, which also ensures a high sorption irreversibility [2–7].

The solid–liquid distribution coefficient (K_d) is a useful parameter to characterise the radionuclide behaviour at solid–liquid interfaces. Although it is not a thermodynamic constant, the K_d integrates simultaneous physicochemical processes, and it is highly dependent on liquid and solid phase characteristics [8–10]. The quantification of K_d values can be made using various experimental approaches [11–13], although batch experiments are the most widely used approach [14,15]. There is a general agreement that K_d values can be used as input data for environmental decision support models [16]. However, there is still a major concern about the apparently high variability of the K_d data because the values depend not only on soil characteristics, but also on experimental conditions. These variables include the cationic composition of the contact solution, the contact time, and the concentration of the stable element related to the radioisotope. Recent studies have succeeded in reducing the variability of the best estimate of the radiocaesium distribution coefficient ($K_d(\text{Cs})$) for a given type of soil [17,18]. In these approaches, soils were grouped with respect to textural classes, organic matter content, or on the basis of specific parameters that govern radiocaesium interaction in soils. After data that were considered outliers or were generated from non-representative experimental conditions were discarded, an estimation of the $K_d(\text{Cs})$ could be made [17]. However, for a given soil group, the best estimate of the $K_d(\text{Cs})$ still varied by several orders of magnitude [17], and a univariate correlation between $K_d(\text{Cs})$ values and a given soil

* Corresponding author. Tel.: +34 934039276; fax: +34 934021233.

E-mail address: miquel.vidal@ub.edu (M. Vidal).

property could not be properly established. Thus, the possibility of calculating an individual $K_d(\text{Cs})$ value for a fully characterised soil was reduced.

Here, we tested two modelling approaches to describe the range of $K_d(\text{Cs})$ values that have been experimentally obtained for a set of soils [19] and to predict an individual $K_d(\text{Cs})$ value based on soil properties. A mechanistic model was constructed based on the knowledge of radiocaesium interaction in soils, which relied on the concept of the Radiocaesium Interception Potential (RIP), the cationic composition of the contact solution, and the clay content in the soil samples. This approach was compared with a regression model based on the use of the Partial Least Square Regression (PLS), which has been successfully applied to predict other radiological parameters, such as soil–plant transfer [20] and the distribution coefficient of other radionuclides [21].

2. Materials and methods

2.1. Soil samples

Thirty agricultural soils were sampled in locations near radioactive facilities and nuclear power plants in Spain. The samples were taken from portions of the surface layer (0–10 cm depth). Prior to the analyses and experiments, the samples were air-dried, passed through 2-mm sieves and homogenised in 5–10 L cylinder bottles for 90 h with a roller table.

2.2. General soil characteristics

Table 1 summarises the ranges of values for general and specific quantified soil characteristics. Individual values are available elsewhere [19]. Soils were mineral-based, with a maximum C_{org} content lower than 10%. Most soils had a loamy texture, and a few had clay (5 soils) and sandy (1 soil) textures. Unlike soils from temperate areas, many of the examined soils had a clay content over 20%, a pH value over 8, and a CaCO_3 content over 25%.

2.3. X-ray diffraction (XRD) analyses of the soil samples

Soil samples were also analysed by XRD to quantify the total content of phyllosilicates and 2:1 phyllosilicates, in addition to other crystalline phases. The determination of the soil mineralogy was carried out with powder samples, whereas for the determination of the 2:1 phyllosilicates the preparation of oriented aggregates was required. The fraction smaller than 2 μm was obtained by sedimentation in aqueous media, and the characterisation of phyllosilicates in oriented aggregates was performed after they were subjected to several treatments (air dried, solvation with ethyleneglycol, thermal treatment at 550 °C). The XRD patterns were obtained on a Bruker D8-Advance diffractometer, with $\text{CuK}\alpha$ radiation, from 3° to 70° 2θ with a step of 0.03° and a counting time of 1 s, whereas for the oriented aggregates they were obtained from 3° to 40° 2θ with a step of 0.02° and a counting time of 1.5 s. The quantitative analyses were carried out following the procedure outlined elsewhere [22].

Table 2 lists the main crystalline phases that were quantified in the examined soils. Phyllosilicate content ranged from 6.8% at ALM to 71.9% at CABRIL, whereas the 2:1 phyllosilicate content ranged from 4.3% at ALM soil to 64% at CABRIL.

2.4. Determination of radiocaesium solid–liquid distribution coefficients

The $K_d(\text{Cs})$ values were determined using two batch sorption tests, which differed in the composition of the contact solution. The first set of data was obtained in a medium that simulated the

cationic composition of the soil solution. Soil samples (1 g) were equilibrated for 16 h with 50 mL of a solution that represented the cationic composition of the soil solution of each soil. After four pre-equilibrations, soils were equilibrated for 24 h with the same solution, but were labelled with ^{137}Cs (CS137ELSB45; LEA FRAMATONE). The $K_d(\text{Cs})$ values were determined by measuring the ^{137}Cs level in the supernatant before and after equilibration.

A second dataset was obtained with a contact solution derived from the soil wash-off. This solution was obtained by equilibrating soil samples (1 g) in water (30 mL g^{-1}) with end-over-end shaking for 16 h. Subsequently, the wash-off solution in contact with the soil was spiked with ^{137}Cs and shaken for 24 h. The $K_d(\text{Cs})$ values were calculated from ^{137}Cs levels in the supernatant before and after equilibration.

The ^{137}Cs activity was measured in samples derived from the sorption experiments in 20 mL capacity polyethylene vials using a solid scintillation detector (PACKARD MINAXI 5000 Series) equipped with a 3-inch NaI (TI activated) crystal. The measurement time was set so that the RSD < 0.5%.

Table 1 also includes the ranges of $K_d(\text{Cs})$ values obtained in the two contact media. From the statistical descriptors, it could be observed that the $K_d(\text{Cs})$ values obtained in the batch experiments with the contact solution that simulated the soil solution composition were systematically lower.

2.5. Radiocaesium-specific soil parameters

2.5.1. Determination of Radiocaesium Interception Potential

In a soil homoionically saturated with K or NH_4^+ , which are competitive species for Cs, the sorption of radiocaesium at the FES depends on the total capacity of these sites [FES], the sorption selectivity of Cs compared with that of K or NH_4^+ ($K_C^{\text{FES}}(\text{Cs}/\text{X})$) and the concentration of K or NH_4^+ in the contact aqueous phase (m_X), as shown in the following equation:

$$K_d^{\text{FES}}(\text{Cs}) = \frac{[\text{FES}]}{m_X} \cdot K_C^{\text{FES}}(\text{Cs}/\text{X}) \quad (1)$$

Because the FES capacity is difficult to measure, one of the approaches to predict the $K_d^{\text{FES}}(\text{Cs})$ is based on the Radiocaesium Interception Potential (RIP) concept, which represents the $K_d^{\text{FES}}(\text{Cs}) \cdot m_X$ product and estimates the soil ability for specific radiocaesium sorption because it also gives a direct measure of $K_C^{\text{FES}}(\text{Cs}/\text{X}) \cdot [\text{FES}]$.

The RIP was determined in K and NH_4^+ scenarios (RIP_K and RIP_N). RIP_K was determined after preequilibrating the soil sample (1 g) with 50 mL of a solution containing 100 mmol L^{-1} of Ca and 0.5 mmol L^{-1} of K ($K_{\text{SS}} = 0.5$). After four preequilibrations, soils were equilibrated for 24 h with the same solution but labelled with ^{137}Cs , and the related $K_d(\text{Cs})$ was determined. The calculated product, $K_d(\text{Cs}) \cdot K_{\text{SS}}$, was associated with the RIP_K value [23]. RIP_N was determined using the same procedure but using a 0.5 mmol L^{-1} NH_4^+ solution.

2.5.2. Changes in the $K_d(\text{Cs})$ due to changes in NH_4^+ concentration in the contact solution

The $K_d(\text{Cs})$ values were obtained in a solution with Ca and K (100 and 10 mmol L^{-1} , respectively). The $K_d(\text{Ca–K})$ values were obtained in an additional five solutions with the same concentration of Ca and K but with increasing NH_4^+ concentrations (0.50, 1.0, 2.5, 4, 5 mmol L^{-1}). The $K_d(\text{Ca–K–NH}_4)$ values were obtained to evaluate how the system responded to increasing concentrations of NH_4^+ in the mixed Ca–K– NH_4^+ scenarios. Soil samples (1 g) were preequilibrated with 50 mL of these solutions, and after 3 preequilibrations, soils were equilibrated for 24 h with the same solution but labelled with ^{137}Cs .

Table 1
Range of values for the soil characteristics.

Variable	Code	Unit	Min	Max	AM	GM
pH	pH	–	4.3	9.2	7.0	6.9
Organic carbon	C _{org}	%	0.2	9.4	3.3	2.1
Carbonate content	CaCO ₃	%	1	51	13.9	6.6
Cationic exchange capacity	CEC	cmol _c kg ⁻¹	21.3	89.3	42.2	39.0
Exchangeable Na	Na _{exch}	cmol _c kg ⁻¹	0.12	19.6	1.4	0.69
Exchangeable K	K _{exch}	cmol _c kg ⁻¹	0.19	10.2	1.3	0.81
Exchangeable Ca	Ca _{exch}	cmol _c kg ⁻¹	3.5	52.0	13.1	10.8
Exchangeable Mg	Mg _{exch}	cmol _c kg ⁻¹	0.45	14.1	2.7	1.8
Exchangeable NH ₄ ⁺	NH _{4exch}	cmol _c kg ⁻¹	0.06	2.6	0.36	0.22
Clay fraction (wrt. initial soil weight)	Clay	%	6.3	52.4	22.9	20.7
Sand fraction (wrt. initial soil weight)	Sand	%	4.5	86.8	43.1	36.8
Field capacity	FC	%	6.2	56.1	27.8	26.1
Electrical conductivity	EC	μS cm ⁻¹	89	3390	738	481
Wash-off experiments						
Na in the contact solution	Na _{ss}	mmol L ⁻¹	0.01	3.6	0.35	0.11
K in the contact solution	K _{ss}	mmol L ⁻¹	0.02	0.35	0.095	0.075
Ca in the contact solution	Ca _{ss}	mmol L ⁻¹	0.01	6.6	0.70	0.29
Mg in the contact solution	Mg _{ss}	mmol L ⁻¹	0.01	0.76	0.12	0.065
NH ₄ in the contact solution	NH _{4ss}	mmol L ⁻¹	0.02	0.96	0.091	0.047
Soil solution experiments						
Na in the contact solution	Na _{ss}	mmol L ⁻¹	0.17	128	11.8	2.0
K in the contact solution	K _{ss}	mmol L ⁻¹	0.13	23.7	1.8	0.79
Ca in the contact solution	Ca _{ss}	mmol L ⁻¹	0.44	52.1	14.3	8.3
Mg in the contact solution	Mg _{ss}	mmol L ⁻¹	0.17	72.5	6.8	2.4
NH ₄ in the contact solution	NH _{4ss}	mmol L ⁻¹	0.016	11.4	1.1	0.24
Radiocaesium distribution coefficient						
Wash-off experiments	K _d	L kg ⁻¹	153	34945	8203	5239
Soil solution experiments	K _d	L kg ⁻¹	10	19437	3111	1185

AM, arithmetic mean; GM, geometric mean.

If radiocaesium sorption were governed by the RES, the K_d (Cs) should not change, because NH₄⁺ and K are equally competitive at these sites and, consequently, the ratio K_d (Ca–K)/ K_d (Ca–K–NH₄) would tend to approach unity. If radiocaesium sorption were associated with the FES, the K_d (Cs) should be sensitive to the changes

in the NH₄⁺ concentrations in the contact solution because NH₄⁺ is more competitive than K at these sites. This latter case is described by the following equation [24,25]:

$$\frac{K_d(\text{Ca} - \text{K})}{K_d(\text{Ca} - \text{K} - \text{NH}_4)} = 1 + K_C^{\text{FES}}(\text{NH}_4/\text{K}) \frac{\text{NH}_{4,ss}}{K_{ss}} \quad (2)$$

Table 2
Main crystalline phases quantified by XRD analyses (% wrt. initial soil weight).

Soil sample	Quartz	Calcite + dolomite	Phyllosilicates	Illite	Smectite	Vermiculite	2:1 phyllosilicates ^a
ALM	86	nd ^b	6.8	4.2	nd	0.1	4.3
ANDCOR	<5	<3	12.2	7.2	nd	5.0	12.2
ASCO	21	50	13.9	10.7	nd	0.3	11.0
AYUD	26	16	50.2	44.2	0.5	nd	44.7
BAD1	67	nd	26.5	11.4	10.3	nd	21.7
BAD2	57	<5	34.6	10.4	19.3	nd	29.7
BILBAO	29	nd	30.3	25.1	0.6	nd	25.7
CABRIL	25	nd	71.9	64.0	nd	nd	64.0
DELTA1	nd	<5	30.0	24.0	nd	0.3	24.3
DELTA2	10	59	14.4	9.8	nd	nd	9.8
ENUSA	59	nd	15.9	9.1	0.2	0.6	9.9
FONCOR	26	nd	13.3	6.0	nd	0.5	6.5
FROCOR	31	nd	18.4	10.5	0.2	0.7	11.4
GAROÑA	31	44	19.1	12.2	1.0	nd	13.2
GOLOSO	23	nd	7.4	3.6	1.4	nd	5.0
GRACOR	46	nd	37.4	4.1	nd	1.5	5.6
LEON	52	nd	34.4	17.5	nd	nd	17.5
MALAGA	18	18	63.8	34.5	10.9	nd	45.3
OVI01	38	nd	43.0	9.9	nd	nd	9.9
OVI03	55	nd	30.7	4.9	nd	0.9	5.8
TENF1	<5	nd	49.4	28.2	4.4	1.5	34.1
TENF2	5	nd	42.9	0.9	28.8	0.4	30.1
TRILLO	19	56	19.2	14.2	nd	nd	14.2
UIB	25	30	32.8	26.5	0.7	nd	27.2
UPV	<5	70	16.7	16.7	nd	nd	16.7
USAL	39	<3	22.9	9.4	10.3	nd	19.7
VAN1	33	33	10.9	8.4	0.5	0.1	9.1
VAN2	27	38	30.9	22.3	nd	0.6	22.9
VILCOR	77	nd	18.6	7.8	nd	1.1	8.9
ZORITA	21	50	22.0	16.4	nd	0.3	16.7

^a 2:1 phyllosilicates: sum of illite, smectite and vermiculite contents.^b nd, not detected.

where K_{ss} and $NH_{4,ss}$ are the K and NH_4^+ concentrations in the contact solution in $mmol\ L^{-1}$, $K_d(Ca-K)$ and $K_d(Ca-K-NH_4)$ are the K_d in the absence or presence of NH_4^+ , respectively, and $K_c^{FES}(NH_4/K)$ is the NH_4 -to-K selectivity coefficient at FES.

2.6. Quality control

Intermediate activity solutions of ^{137}Cs were prepared by diluting weighed amounts of commercial solutions of each radionuclide with deionised water and used as internal controls for measurements with a PACKARD MINAXI 5000 Series instrument. Additionally, parallel blank experiments were carried out for each analysis without a soil sample but with the same procedures for the sorption tests.

2.7. Data treatment

The application of regression models was carried out in consideration of three $K_d(Cs)$ datasets, which were dependent on what the models examined: $K_d(Cs)$ from the wash-off experiments, $K_d(Cs)$ from the soil solution experiments, or the global dataset. The application of the mechanistic models was only carried out in consideration of the global dataset. Single and multiple linear correlations related to the mechanistic models were performed using STATGRAPHICS Plus (StatPoint Technologies, US) software, whereas Unscrambler® 6.11a software (CAMO ASA, Norway) was used to carry out the regression modelling based on the Partial Least Squares regression (PLS).

3. Results and discussion

3.1. Use of mechanistic models

3.1.1. Construction of the global mechanistic model and estimation of the $K_c^{FES}(NH_4/K)$

As radiocaesium sorption in soils is controlled by FES and RES, two contributions are expected for the total $K_d(Cs)$ in soils: the $K_d^{FES}(Cs)$ and the $K_d^{RES}(Cs)$. The relative weight of each contribution depends on the soil type [4,7,26].

The $K_d^{FES}(Cs)$ can be predicted by dividing the RIP_K value (in $mmol\ kg^{-1}$) by the sum of the K, NH_4^+ and Na concentrations in the soil solution (in $mmol\ L^{-1}$) and by multiplying the NH_4^+ and Na contributions by their respective selectivity coefficients in FES ($K_c^{FES}(NH_4/K)$ and $K_c^{FES}(Na/K)$) because at the FES, the relative selectivity of these monovalent cations differs from 1 [23,27]. The $K_d^{RES}(Cs)$ can be calculated by dividing the sum of the exchangeable K, NH_4^+ and Na (in $mmol\ kg^{-1}$) by the sum of the K, NH_4^+ and Na concentrations in the soil solution [17,23]. The derived equation may be written as follows:

$$K_d(Cs) = K_d^{FES}(Cs) + K_d^{RES}(Cs) = \frac{RIP_K}{K_{ss} + K_c^{FES}(NH_4/K) \cdot NH_{4,ss} + K_c^{FES}(Na/K) \cdot Na_{ss}} + \frac{K_{exch} + NH_{4,exch} + Na_{exch}}{K_{ss} + NH_{4,ss} + Na_{ss}} \quad (3)$$

Sodium has little significance in Cs sorption, with the exception of Cs sorption in saline soils; a good estimation of $K_c^{FES}(Na/K)$ is 0.02. By including NH_4^+ in the prediction of $K_d^{FES}(Cs)$, a proper quantification of the $K_c^{FES}(NH_4/K)$ value is required. $K_c^{FES}(NH_4/K)$ can be determined using two approaches. In the first approach, the ratio between the RIP obtained in the K and NH_4^+ scenarios [28] can be used, and in the second, the changes in the $K_d(Cs)$ after changing the NH_4^+ concentration in the contact solution in mixed Ca–K– NH_4^+ scenarios (as described in Section 2.5.2) can be used.

Table 3 presents the $K_c^{FES}(NH_4/K)$ values obtained by the two approaches, and Fig. 1 displays four examples of the changes in $K_d(Cs)$ that occurs with increasing NH_4^+ concentration. In this case, the $K_c^{FES}(NH_4/K)$ is derived from the slope of the

Table 3

RIP_K ($mmol\ kg^{-1}$), RIP_N ($mmol\ kg^{-1}$), and $K_c^{FES}(NH_4/K)$ values obtained by two methods.

Soil sample	RIP_K	RIP_N	$K_c^{FES}(NH_4/K)$	
			RIP_K/RIP_N	Derived from Eq. (2)
ALM	1002	316	3.2	8.7
ANDCOR	366	140	2.6	6.2
ASCO	2034	512	4.0	6.5
AYUD	4929	1368	3.6	7.5
BAD1	1523	376	4.1	8.3
BAD2	2594	613	4.2	6.2
BILBAO	4419	2052	2.2	9.4
CABRIL	4852	1234	3.9	10.2
DELTA1	1186	298	4.0	7.2
DELTA2	1463	337	4.3	7.4
ENUSA	744	249	3.0	7.3
FONCOR	340	128	2.7	4.4
FROCOR	179	74	2.4	7.7
GAROÑA	2080	749	2.8	5.3
GOLOSO	672	204	3.3	5.2
GRACOR	1017	303	3.4	8.3
LEON	2314	910	2.5	10.0
MALAGA	5958	1824	3.3	6.9
OVI01	3221	1009	3.2	8.3
OVI03	1718	825	2.1	7.6
TENF1	6669	4914	1.4	4.4
TENF2	6758	2128	3.2	5.0
TRILLO	2897	838	3.5	7.4
UIB	7000	1595	4.4	7.1
UPV	3213	714	4.5	7.6
USAL	1874	753	2.5	5.7
VAN1	660	205	3.2	6.2
VAN2	4488	803	5.6	8.0
VILCOR	225	103	2.2	5.1
ZORITA	2371	687	3.5	6.8
Min	179	74	1.4	4.4
Max	7000	4914	5.6	10.2
AM	2626	875	3.3	7.1
GM	1760	557	3.2	6.9

AM, arithmetic mean; GM, geometric mean.

$K_d(Ca-K)/K_d(Ca-K-NH_4)$ vs. $NH_{4,ss}/K_{ss}$ correlation. The quantification of the $K_c^{FES}(NH_4/K)$ coefficient depends on the approach followed, because the values from the mixed scenario experiments were consistently higher (around 2.3-fold higher, as a mean value) than those derived from the RIP ratio. This is the first time that a systematic comparison of $K_c^{FES}(NH_4/K)$ coefficients derived from the two approaches has been performed. An explanation for the higher values of the $K_c^{FES}(NH_4/K)$ coefficient derived from the mixed scenario experiments may be that the lower K and NH_4^+ concentrations used for the RIP calculations might not ensure homoionic FES in soils with high clay content. When one focuses on the range of values derived from both approaches, the fact that the $K_c^{FES}(NH_4/K)$

values were consistently higher than 1 confirms that FES govern the radiocaesium sorption in these soils.

3.1.2. Prediction of the $K_d(Cs)$ on the basis of the global and simplified equations

The use of the $K_c^{FES}(NH_4/K)$ values derived from either the RIP ratio or the mixed scenario experiments did not affect the mechanistic model, because similar experimental vs. predicted $K_d(Cs)$ were obtained ($\log K_d(Cs)_{exp} = 0.98 (\pm 0.02) \log K_d(Cs)_{pred}$, $n=60$, $r=0.94$; and $\log K_d(Cs)_{exp} = 1.04 (\pm 0.02) \log K_d(Cs)_{pred}$, $n=60$, $r=0.90$, respectively). Both of these cases show that the multiple correlation represented by Eq. (3) succeeded in predicting the

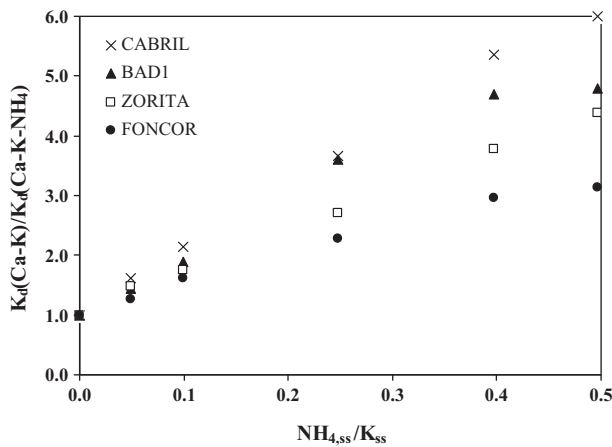


Fig. 1. Response of the $K_d(\text{Cs})$ to increasing NH_4^+ concentrations in a Ca–K– NH_4 scenario in four of the soils of the data set: (●) Foncor, (□) Zorita, (▲) Bad1 and (×) Cabril soils.

$K_d(\text{Cs})$ in soils. As an example, Fig. 2 shows the results of the prediction exercise from Eq. (3) by using the $K_c^{\text{FES}}(\text{NH}_4/\text{K})$ value derived from the mixed scenario experiments.

The use of Eq. (3) requires a full characterisation of the soils. Therefore, simpler models could be necessary in case there is a limited soil characterisation. As has been suggested previously [4], a value of 5 can be taken for $K_c^{\text{FES}}(\text{NH}_4/\text{K})$ if its value is unknown for a given soil. This value is also consistent with the arithmetic mean value of all the $K_c^{\text{FES}}(\text{NH}_4/\text{K})$ values obtained in the present work. Additionally, the role of Na can be disregarded due to the low value of the $K_c^{\text{FES}}(\text{Na}/\text{K})$. Taking these two assumptions into account, the model still predicted the experimental values of the $K_d(\text{Cs})$: the resulting correlation was similar to the correlation obtained before the assumptions were made ($\log K_d(\text{Cs})_{\text{exp}} = 1.01(\pm 0.02) \log K_d(\text{Cs})_{\text{pred}}$, $n = 60$, $r = 0.92$).

Further assumptions to simplify the model might have a more significant effect on the correlations. One option is to focus on the $K_d^{\text{FES}}(\text{Cs})$ and to neglect the $K_d^{\text{RES}}(\text{Cs})$. A rapid calculation confirmed that for many soils, the $K_d^{\text{FES}}(\text{Cs})$ accounted for more than 90% of the global $K_d(\text{Cs})$, but in a few soils, the percentage decreased to 53% (e.g., the FROCOR soil). As the additional information needed

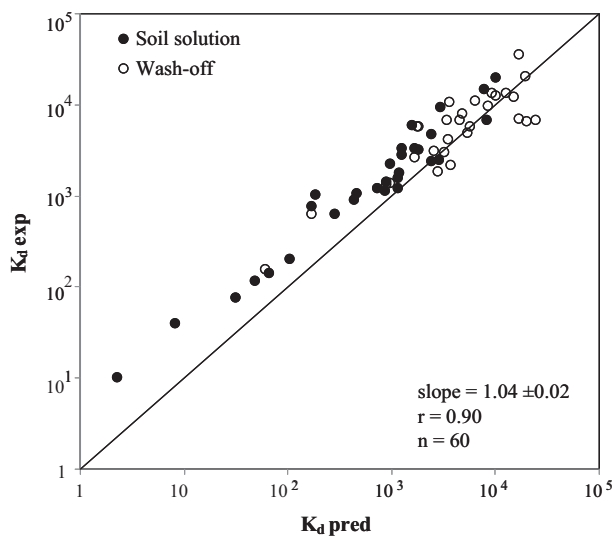


Fig. 2. Prediction of $K_d(\text{Cs})$ values with Eq. (3), using $K_c^{\text{FES}}(\text{NH}_4/\text{K})$ values derived from mixed scenario experiments. The solid line represents the ideal $K_d(\text{Cs})_{\text{exp}} = K_d(\text{Cs})_{\text{pred}}$ relationship.

to calculate the $K_d^{\text{RES}}(\text{Cs})$ is the concentration of cations in the exchangeable complex, this simplification would not add significant value, because this information is often available. Another option is not to include the NH_4^+ data. The NH_4^+ concentration, and thus its role in the prediction of $K_d(\text{Cs})$, could be high in wet and peat soils, whereas it could be lower than the value of K in mineral soils [19]. The equation that results from this additional simplification would be as follows:

$$K_d(\text{Cs}) = K_d^{\text{FES}}(\text{Cs}) + K_d^{\text{RES}}(\text{Cs}) = \frac{\text{RIP}_K}{K_{\text{SS}}} + \frac{K_{\text{exch}}}{K_{\text{SS}}} \quad (4)$$

The correlation obtained with the simplified equation was slightly worse than the previous correlations ($\log K_d(\text{Cs})_{\text{exp}} = 0.88(\pm 0.02) \log K_d(\text{Cs})_{\text{pred}}$, $r = 0.89$), but it still ensured a proper prediction of the $K_d(\text{Cs})$ values.

A major limitation of these mechanistic models' ability to predict the $K_d(\text{Cs})$ values is the fact that a $K_d(\text{Cs})$ value must be obtained to quantify the RIP_K value. To date, attempts to predict the RIP_K value based on soil properties have only been partially successful, because the RIP_K value is multivariately dependent not only on the clay content but also on the type of clay and geological origin of the soil [29]. Both the phyllosilicate and 2:1 phyllosilicate contents had similar correlations with respect to the RIP_K values (for instance, $\text{RIP}_K = 92 (\pm 30)$ phyllosilicate, $n = 30$, $r = 0.76$) in the examined soils; both variables were strongly correlated ($r = 0.85$). Another parameter that may correlate with the RIP_K is the clay content of the corresponding textural class, which is a parameter that is easier to quantify because it does not require additional XRD analyses. In this case, the correlation was slightly worse ($\text{RIP}_K = 130 (\pm 60)$ clay, $n = 30$, $r = 0.65$). The inclusion of the silt content, which was also derived from textural analyses, only improved the percentage of the RIP_K variability in the case of a multiple regression with the phyllosilicate content ($\text{RIP}_K = 93 (\pm 26)$ phyllosilicate + 61 (± 28) silt – 1700 (± 1140), $r = 0.86$). Any of these correlations could be used to predict the $K_d(\text{Cs})$ when a previous prediction of the RIP_K values exists. For instance, when the RIP_K value is substituted by its correlation with the clay content, a worse correlation was obtained ($\log K_d(\text{Cs})_{\text{exp}} = 0.84(\pm 0.02) \log K_d(\text{Cs})_{\text{pred}}$, $r = 0.74$). With this model, the $K_d(\text{Cs})$ is predicted based only on general soil properties, such as the texture and K status in the exchangeable complex and the soil solution. However, the correlation is still good enough to generate input data for environmental and decision support models.

3.2. PLS-based regression-modelling

An alternative to using mechanistic-modelling to predict the $K_d(\text{Cs})$ from soil properties is to use regression-modelling, which requires a large database of samples and variables and takes into account soil properties that either are directly associated with the radiocaesium sorption in soils or are properties that affect the sorption process. A data matrix was constructed with rows related to the soil samples and columns related to the soil variables, with the aim of evaluating the capacity of multivariate regression-modelling in the prediction of $K_d(\text{Cs})$ values. In order to assume normal distributions of variables, the data were transformed to logarithmic units, with the exception of pH. Data were also autoscaled prior to analysis – i.e., mean centred (each element was subtracted by its mean column) and scaled to unit variance (each element was divided by the standard deviation of its column) – to give the same weight to all variables in the analyses.

PLS builds a model between a sample variable (y) and a set of other sample descriptors (X) in a low-dimensional space that is formed by PLS components. It finds the model with the maximum covariance of the relationship between the X - and y -space [30,31], as $y = XB$, where B is the matrix of regression coefficients

calculated in the PLS space. Full cross-validation was used to determine the number of PLS components that provided a minimal value of the Root Mean Square Error of Cross Validation (RMSECV):

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (y_{\text{predicted}}^i - y_{\text{experimental}}^i)^2}{n}} \quad (5)$$

Further details on the theoretical background of the PLS algorithm and on the applied software can be found elsewhere [32].

3.2.1. Identification of the relevant variables in the regression models

PLS was first used to determine which soil variables were relevant to the prediction of the $K_d(\text{Cs})$. This was achieved by evaluating the quality parameters of the models and the value and sign of the regression coefficients. The soil solution, the wash-off, and the global datasets were modelled separately to reveal which variables better explained the $K_d(\text{Cs})$ in each dataset. The role of the clay fraction, which is one of the key soil properties that affects radiocaesium interaction in soils, was evaluated by either the phyllosilicate content, determined by mineralogical analysis, or the clay content, determined by textural analysis. The weight of RIP_K in the model was also assessed. Table 4 summarises the statistical description of the cross-validated PLS-based regression model obtained in each case. Four PLS components were considered for comparison purposes because the explained cumulative variance of y was higher than 80% in many cases. Additional PLS components slightly improved the quality parameters of the models, but the relevance of variables did not vary.

When the RIP_K variable was used to construct the model, almost identical results were obtained for the three data sets, regardless the variable selected to account for the role of clays, because the $K_d(\text{Cs})$, and thus the regression model, was strongly affected by the RIP variable. Satisfactory regressions ($r > 0.9$) with slopes close to 1 were obtained, confirming the absence of bias in the models and, therefore, of systematic overestimations or underestimations of $K_d(\text{Cs})$. However, the y explained cumulative variance was slightly better for the global dataset due to the wider range of values. The RMSECV values, calculated after back-transforming the logarithmic values of the predicted $K_d(\text{Cs})$ to their original units, were also very similar for the three models (from 3300 to 4200). The regression coefficients of the regression models were also compared to identify the most relevant variables that describe the $K_d(\text{Cs})$ variability in each case. The direction and magnitude of the influence of variables can generally be associated with the sign and the magnitude of its regression coefficient. However, as soil variables may be partially correlated, the related regression coefficients cannot be considered in an individual or completely independent way. Only the most salient trends in the regression coefficients can safely be used for chemical interpretation. For example, Fig. 3a shows the regression coefficients obtained for the global dataset, including RIP_K and clay texture variables. Non-significant variations in the relative relevance of variables were observed when phyllosilicates were included instead of clay texture; similar conclusions could also be drawn from the regression coefficients of the three data sets. Among the clearest trends was the high negative correlation of the concentration of competitive species, K and NH_4^+ , in the contact solution; this trend was evident in all three data sets. Another general trend was the relevance of RIP_K and clay texture, which presented high positive regression coefficients. These findings were consistent with the mechanisms that dictate the $K_d(\text{Cs})$ [2,25,33,34].

As shown in Table 4, when RIP_K was excluded from the databases, the quality parameters of the models were still satisfactory, but they were worse than when the RIP_K variable was

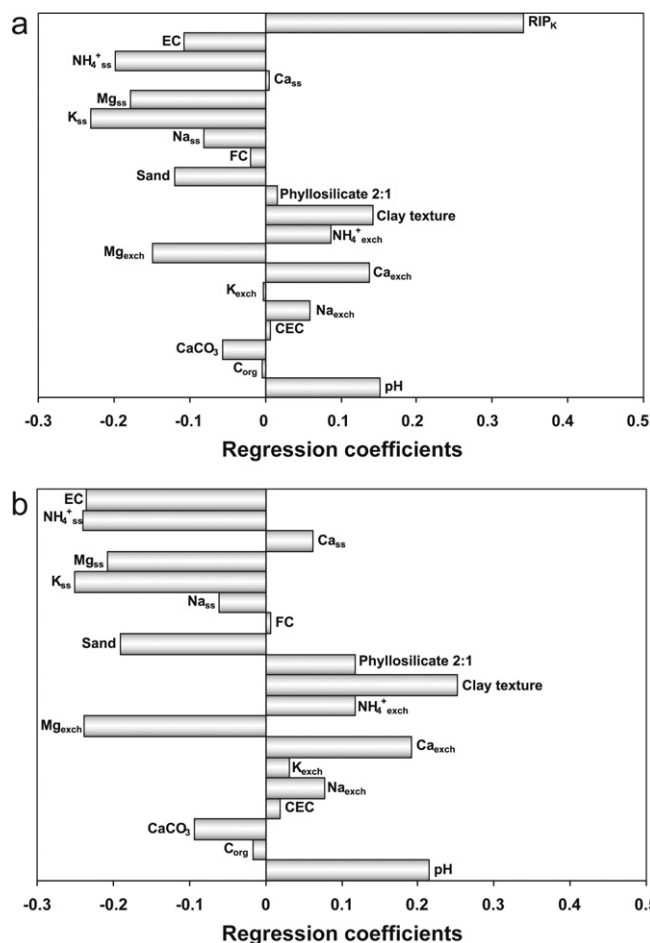


Fig. 3. Regression coefficients of the PLS models using the global dataset and the clay texture to describe the clay content, (a) including RIP_K and (b) excluding RIP_K .

retained, especially in the datasets with a smaller amount of data. Again, the models that considered the clay texture were similar to those that considered phyllosilicates. Based on the regression coefficients (Fig. 3b), the relative relevance of the soil parameters, except for clay texture, is similar to the results from the model constructed with the RIP_K variable. The results are especially similar with respect to the 2:1 phyllosilicate variable and the concentration of cationic species in the contact solution, which became significant in the prediction of $K_d(\text{Cs})$.

3.2.2. External validation of the PLS-based regression models

In contrast to mechanistic-modelling, which employs a fixed equation based on soil parameters to predict the $K_d(\text{Cs})$, the PLS-based regression models require an additional step to predict $K_d(\text{Cs})$ values for samples that have not been used to calibrate the model. For the external validation step, two-thirds of the samples of the global dataset were selected for calibration and to construct the regression model using the Kennard–Stone algorithm [35], and the remaining samples were used as the prediction set. Models were constructed with or without RIP_K , and in each case, the total phyllosilicates or clay texture variables were considered.

The regression models obtained with the reduced calibration sets had a pattern of regression coefficients and quality parameters similar to those obtained in the previous section. For instance, the y explained cumulative variance was 85% when RIP_K was not considered and 90% when RIP_K was considered. The slopes were 1, the regression coefficients were greater than 90, and the RMSECV values were of the same order of magnitude as

Table 4
Description of the cross-validated PLS models (4 PLS components) to predict K_d (Cs).

Dataset	n	y explained cumulative variance (%)	RMSECV ^a	Experimental vs. predicted regression	
				Slope	R
<i>With RIP and total phyllosilicates</i>					
Soil solution	30	85	3700	0.99 ± 0.04	0.90
Wash-off	30	83	3300	1.00 ± 0.02	0.90
Soil solution + wash-off	60	89	4100	1.00 ± 0.02	0.94
<i>With RIP and clay texture</i>					
Soil solution	30	86	3400	0.99 ± 0.04	0.91
Wash-off	30	86	3300	1.00 ± 0.02	0.91
Soil solution + wash-off	60	89	4200	1.00 ± 0.02	0.94
<i>Without RIP and with total phyllosilicates</i>					
Soil solution	30	79	5300	0.99 ± 0.04	0.86
Wash-off	30	67	4100	0.99 ± 0.03	0.79
Soil solution + wash-off	60	84	4100	0.99 ± 0.02	0.91
<i>Without RIP and with clay texture</i>					
Soil solution	30	80	4300	0.99 ± 0.04	0.87
Wash-off	30	70	3600	1.00 ± 0.03	0.81
Soil solution + wash-off	60	85	4100	0.99 ± 0.02	0.92

^a Value calculated after back-transforming the logarithmic predicted values to original y units.

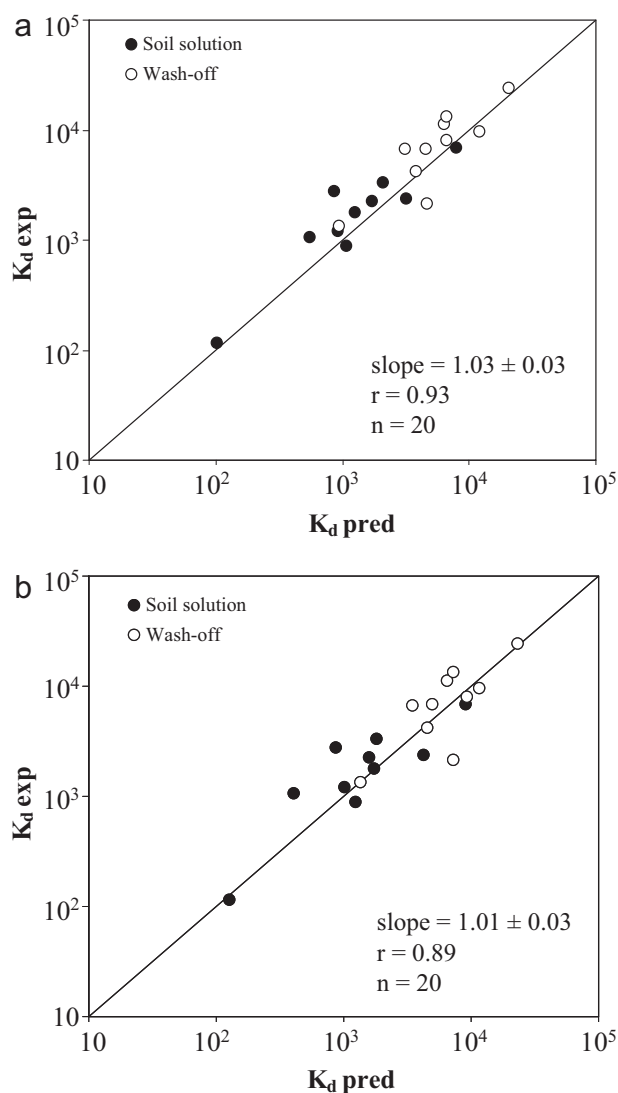


Fig. 4. Validation of the PLS models with the prediction set. Experimental vs. predicted K_d (Cs) correlations obtained with the PLS models using the global dataset and the clay texture to describe the clay content, (a) including RIP_K and (b) excluding RIP_K , in which the solid line represents the ideal $K_d(Cs)_{exp} = K_d(Cs)_{pred}$ relationship.

previously obtained. These statistics confirmed that the data used to construct the new regression model was representative data.

In each case, the model was externally validated with the corresponding test set. For example, Fig. 4a and b shows the correlations between experimental and predicted K_d (Cs) values with or without RIP_K , respectively, for data sets with the clay texture variable. The two correlations were satisfactory, with slopes close to unity and regression coefficients close to or higher than 0.9. These results confirmed the robustness and the absence of bias in the PLS-based regression models and indicated that the K_d (Cs) could also be predicted with the regression models without the RIP_K variable.

4. Conclusions

Both mechanistic and PLS-based regression models succeeded in predicting K_d (Cs) with satisfactory correlations between experimental and predicted values. Among mechanistic models, the model with the highest requirements in terms of soil data provided the best prediction. However, it was possible to propose simpler, similarly satisfactory models based on soil properties that are available from routine analyses, such as the clay content and the K status in the soil. The PLS-based regression models confirmed that soil parameters such as the RIP_K , the clay content, and the NH_4^+ and K statuses explained the variability in the K_d (Cs) values. Only slight differences in the models were observed when the clay content, obtained by textural analysis, or the phyllosilicate content, obtained by mineralogical analysis, was used. In all cases, the regressions had quality parameters similar to those of the mechanistic models. Although better correlations between experimental and predicted K_d (Cs) were obtained when the RIP_K was included as a soil variable, the prediction of K_d (Cs) was also satisfactory when this specific soil parameter was not considered in the regression model. As the mechanisms of the interaction of a contaminant in soils are not always well known, PLS-based regression models are recommended as a powerful approach to examine and to predict the interaction parameters of any contaminant in soils.

Acknowledgements

This research was funded by the Spanish Government (CICYT, contracts PPQ2002-00264 and CTM2005-03847/TECNO) and the Generalitat de Catalunya (contract 2009SGR1188). C. J. Gil-García acknowledges the Ministerio de Educación, Cultura y Deporte for a PhD fellowship. The authors thank Universidad de Salamanca, Universidad Politécnica de Valencia, Universidad de Oviedo,

Universidad de Extremadura, Universidad de León, Universitat de les Illes Balears, ENRESA, ENUSA, SCAR (Generalitat de Catalunya), Universidade da Coruña and Universidad de Málaga for the soil samples supplied. The authors also thank Dra. A. Sahuquillo for the use of MATCONTROL facilities to prepare the samples.

References

- [1] Radioactive contamination of the environment, in: Environmental Consequences of the Chernobyl Accident and their Remediation: Twenty Years of Experience. Radiological Assessment Reports Series, International Atomic Energy Agency, Vienna, 2006, pp. 18–68.
- [2] A. Cremers, A. Elsen, P. De Preter, A. Maes, Quantitative analysis of radiocaesium retention in soils, *Nature* 335 (1988) 247–249.
- [3] J. Wauters, L. Sweeck, E. Valcke, A. Elsen, A. Cremers, Availability of radiocaesium in soils: a new methodology, *Sci. Total Environ.* 157 (1994) 239–248.
- [4] M. Vidal, M. Roig, A. Rigol, M. Llauro, G. Rauret, J. Wauters, A. Elsen, A. Cremers, Two approaches to the study of radiocaesium partitioning and mobility in agricultural soils from the Chernobyl area, *Analyst* 120 (1995) 1785–1791.
- [5] R.N.J. Comans, M. Haller, P. De Preter, Sorption of caesium on illite: non-equilibrium behaviour and reversibility, *Geochim. Cosmochim. Acta* 55 (1991) 433–440.
- [6] A.B. Hird, D.L. Rimmer, Total caesium-fixing potentials of acid organic soils, *J. Environ. Radioact.* 26 (1995) 103–118.
- [7] G. Rauret, S. Firsakova, The transfer of radionuclides through the terrestrial environment to agricultural products, including the evaluation of agrochemical practices, in: EUR 16528 EN, European Commission, Luxembourg, 1996.
- [8] J. Hilton, R.N.J. Comans, Chemical forms of radionuclides and their quantification in environmental samples, in: E. Van der Stricht, R. Kirchmann (Eds.), *Radioecology, Radioactivity and Ecosystems*, Fortemps, Liège, 2001, pp. 99–111.
- [9] Environmental Protection Agency, Understanding Variation in Partitioning Coefficients, K_d , values. Volume II: Review of Geochemistry and Available K_d values for Cadmium, Caesium, Chromium, Lead, Plutonium, Radon, Strontium, Thorium, Tritium and Uranium, EPA 402-R-99-004B, US-EPA Office of Air and Radiation, Washington, USA, 1999.
- [10] S. Staunton, Sensitivity of the distribution coefficient, K_d , of nickel to changing soil chemical properties, *Geoderma* 122 (2004) 281–290.
- [11] H. Bachhuber, K. Bunzl, W. Schimmack, The migration of ^{137}Cs and ^{90}Sr in multilayered soils: results from batch column and fallout investigations, *Nucl. Technol.* 59 (1992) 291–301.
- [12] A.L. Sanchez, E. Smolders, K. Van den Brande, R. Merckx, S.M. Wright, C. Naylor, Predictions of in situ solid–liquid distribution of radiocaesium in soils, *J. Environ. Radioact.* 63 (2002) 35–47.
- [13] L. Dewiere, D. Bugai, C. Grenier, V. Kashparov, N. Ahamdach, ^{90}Sr migration to the geo-sphere from a waste burial in the Chernobyl exclusion zone, *J. Environ. Radioact.* 74 (2004) 139–150.
- [14] C. Elejalde, M. Herranz, F. Legarda, F. Romero, Determination and analysis of distribution coefficients of ^{137}Cs in soils from Biscay (Spain), *Environ. Pollut.* 110 (2000) 157–164.
- [15] Q. Tianwei, T. Hongxiao, C. Jiajun, W.J. Sheng, L. Chunli, W. Guibin, Simulation of the migration of ^{85}Sr in Chinese loess under artificial rainfall condition, *Radiochim. Acta* 89 (2001) 403–406.
- [16] J. Gutiérrez, C. Vázquez, R. Meckbach, B. Wilkins, B. Rafferty, E. Holm, M. Badie, O. Burton, TEMAS Project. Techniques and management strategies for environmental restoration and their ecological consequences, *Colección Documentos Ciemat*, pp. 16–27.
- [17] C.J. Gil-García, A. Rigol, M. Vidal, New best estimates for radionuclide solid–liquid distribution coefficients in soils. Part 1: radiostromium and radio-caesium, *J. Environ. Radioact.* 100 (2009) 690–696.
- [18] M. Vidal, A. Rigol, C. Gil-García, Soil-radionuclide interactions, in: Quantification of radionuclide transfer in terrestrial and freshwater environments for radiological assessments, IAEA-TECDOC-1616, IAEA, Vienna, 2009, pp. 71–102.
- [19] C.J. Gil-García, A. Rigol, G.G. Rauret, M. Vidal, Radionuclide sorption–desorption pattern in soils from Spain, *Appl. Radiat. Isot.* 66 (2008) 126–138.
- [20] A. Rigol, M. Camps, A. de Juan, G. Rauret, M. Vidal, Multivariate soft-modelling to predict radiocaesium soil-to-plant transfer, *Environ. Sci. Technol.* 42 (2008) 4029–4036.
- [21] C.J. Gil-García, A. Rigol, M. Vidal, The use of hard and soft modelling to predict radiostromium solid–liquid distribution coefficient in soils, *Chemosphere*, 10.1016/j.chemosphere.2011.08.015.
- [22] P. Aparicio, R.E. Ferrell, An application of profile fitting and CLAY++ for the quantitative representation (QR) of mixed-layer clay minerals, *Clay Min.* 36 (2001) 501–514.
- [23] J. Wauters, A. Elsen, A. Cremers, A.V. Konoplev, A.A. Bulgakov, R.N.J. Comans, Prediction of solid liquid distribution coefficients of radiocaesium in soils and sediments. Part 1: A simplified procedure for the solid characterization, *Appl. Geochem.* 11 (1996) 589–594.
- [24] J. Wauters, M. Vidal, A. Elsen, A. Cremers, Prediction of solid liquid distribution coefficients of radiocaesium in soils and sediments. Part 2: A new procedure for the solid characterization, *Appl. Geochem.* 11 (1996) 595–599.
- [25] A. Rigol, M. Vidal, G. Rauret, C.A. Shand, M.V. Cheshire, Competition of organic and mineral phases in radiocaesium partitioning in organic soils of Scotland and the area near Chernobyl, *Environ. Sci. Technol.* 32 (1998) 663–669.
- [26] L. Sweeck, Soil–chemical availability of radiocaesium in mineral soils, Ph.D. Thesis, Katholieke Universiteit Leuven, 1996.
- [27] J. Wauters, M.J. Madruga, M. Vidal, A. Cremers, Solid-phase speciation of radiocaesium in bottom sediments, *Sci. Total Environ.* 187 (1996) 121–130.
- [28] L. Sweeck, J. Wauters, E. Valcke, A. Cremers, The specific interception potential of soils for radiocaesium, in: G. Desmet, P. Nassimbeni, M. Belli (Eds.), *Transfer of Radionuclides in Natural and Seminatural Environments*, EUR 12448, Elsevier Applied Science, 1990, pp. 249–258.
- [29] N. Waegeneers, E. Smolders, R. Merckx, A statistical approach for estimating the radiocaesium interception potential of soils, *J. Environ. Qual.* 28 (1999) 1005–1011.
- [30] A.E. Boardman, B.S. Hui, H. Wold, The partial least squares-fix point method of estimating interdependent systems with latent variables, *Comm. Stat. Theory Methods* 10 (1981) 613–639.
- [31] S. Wold, H. Martens, H. Wold, The multivariate calibration problem in chemistry solved by the PLS method, *Lect. Notes Math.* 973 (1983) 286–293.
- [32] K. Esbensen, S. Schönkopf, T. Midtgaard, *Multivariate Analysis in Practice Computer-Aided Modelling*, AS Norway, 1994.
- [33] S. Staunton, Adsorption of radiocaesium on various soils: interpretation and consequences of the effects of soil:solution ratio and solution composition on the distribution coefficient, *Eur. J. Soil Sci.* 45 (1994) 409–418.
- [34] J.P. Absalom, S.D. Young, N.M.J. Crout, Radiocaesium fixation dynamics: measurement in six Cumbrian soils, *Eur. J. Soil Sci.* 46 (1995) 461–469.
- [35] R. Kennard, L.A. Stone, Computer aided design of experiments, *Technometrics* 11 (1969) 137–148.