Bioaccessibility of Cd and Pb as affected by soil parameters in contaminated agricultural soils: A statistical modeling approach

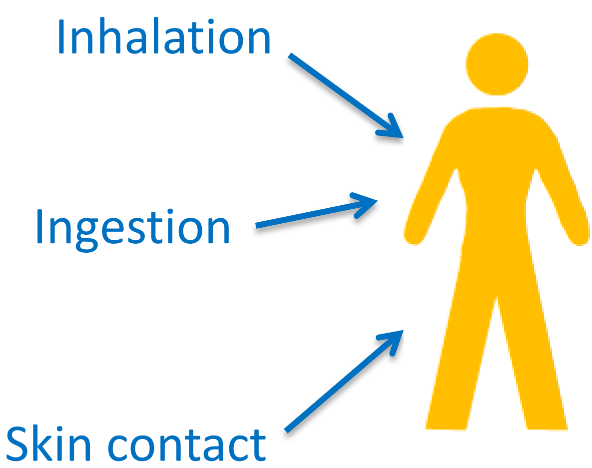
A. Pelfrêne, C. Waterlot, M. Mazzuca, C. Nisse, G. Bidar, F. Douay

Intersol 2012 – March, 27 to 30, 2012 - Paris

Context

- Metal mining and smelting activities are the major sources of metals in the environment
  - Soil contamination
  - Human health risks

Soil exposure pathways



Soil ingestion

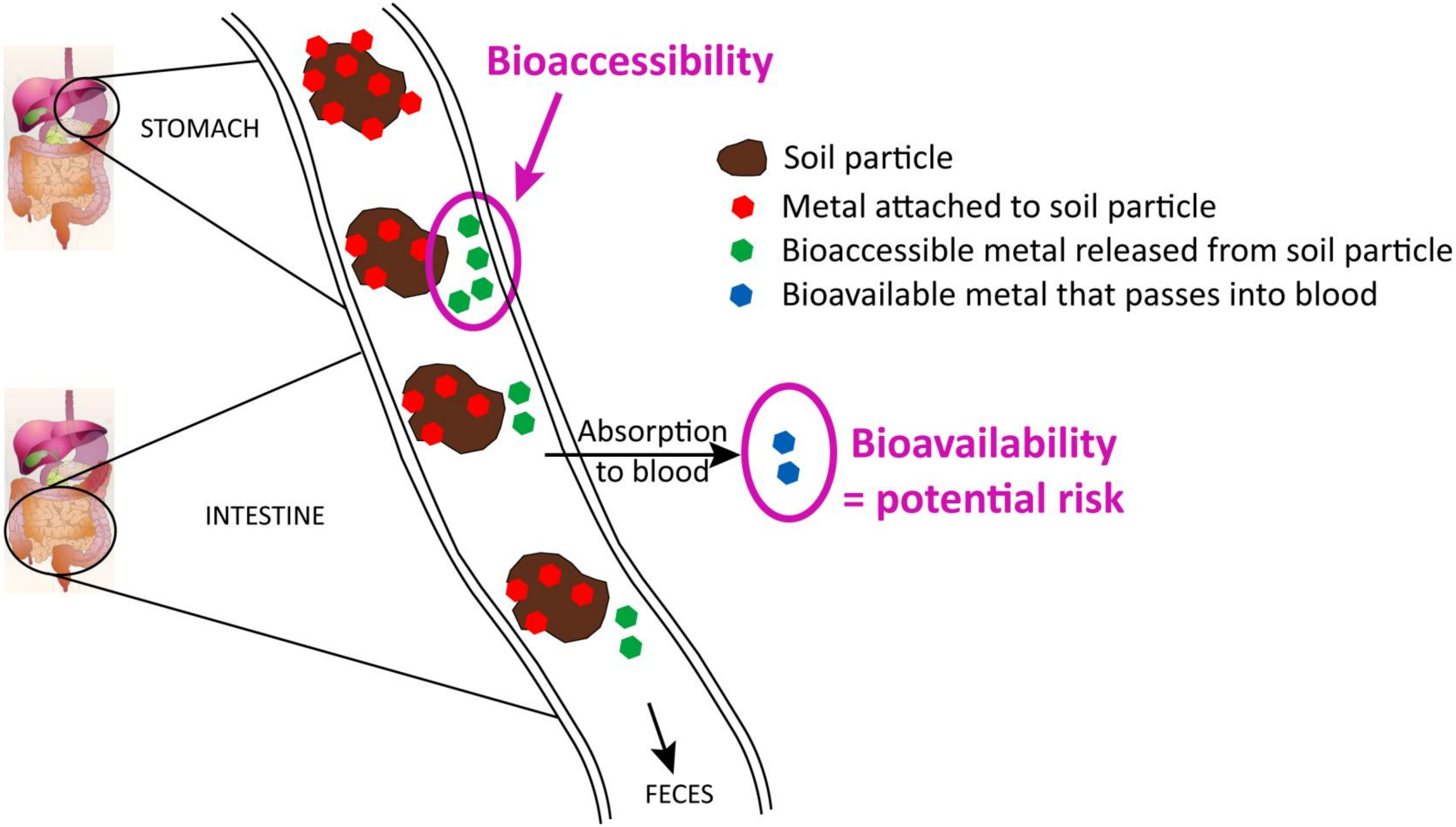
Adults: 20-100 mg day^{-1}

Children: 80-400 mg day^{-1}

Current risk models assume that 100% of the ingested contaminants pass into the bloodstream

Context

Fate of metals after ingestion of soil particles

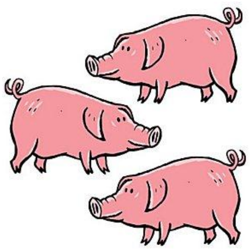


Total metal concentrations ➡ Overestimation of the exposure to soil contaminants

Context

How do we measure metal bioavailability?

In vivo tests: animal models

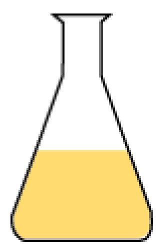


Expensive
Time-consuming
Ethical issues

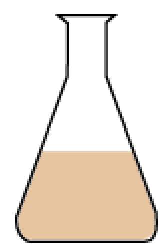
In vitro tests: bioaccessibility tests

Typically use a two-stage method that simulates the chemical and physical environment of the gastrointestinal tract

Soil sample



Gastric conditions



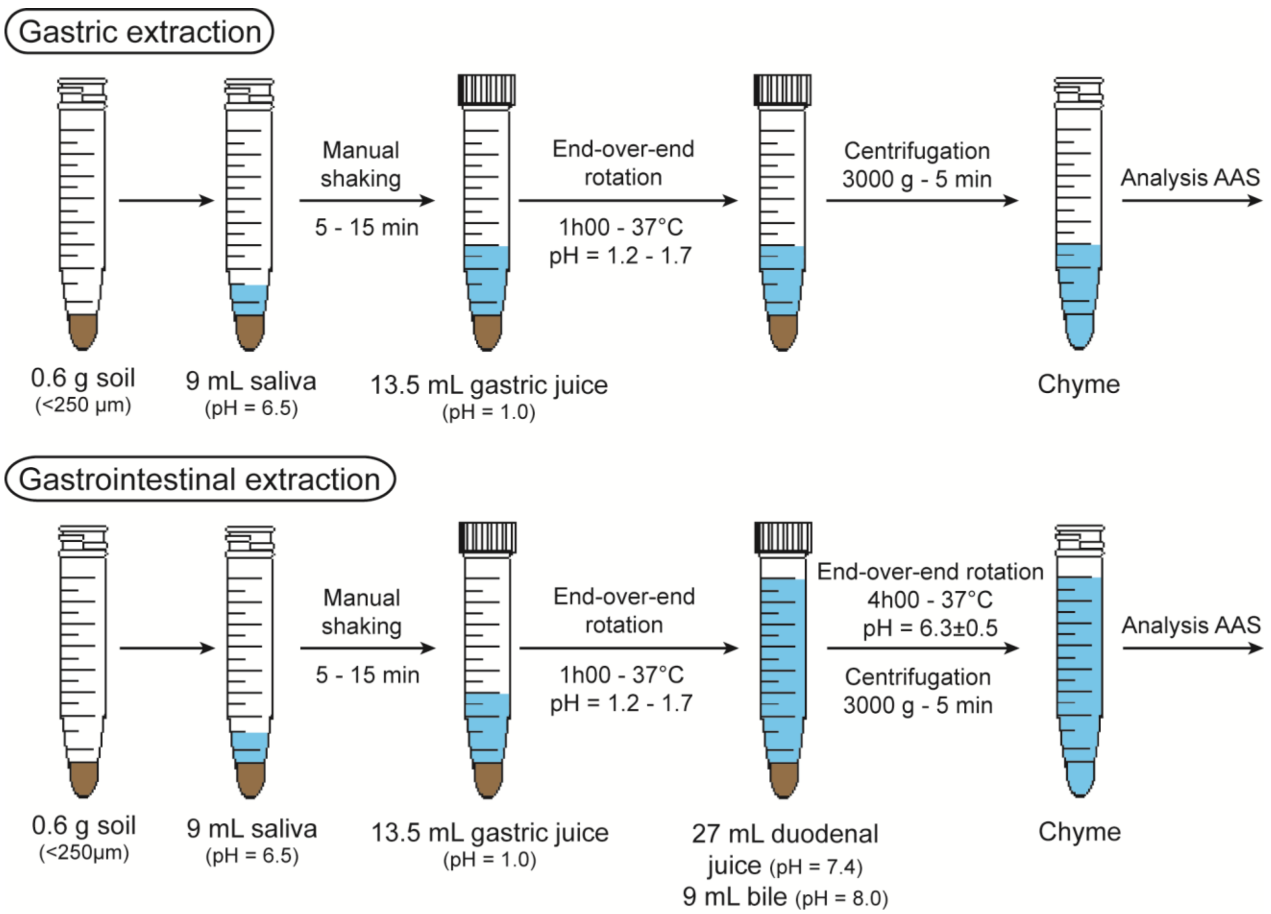
Intestinal conditions

Analysis

Several *in vitro* tests developed: PBET, IVG, SHIME, TIM, RIVM, … and UBM

Context

UBM - Unified Bioaccessibility Method



Validated against an *in vivo* model for Cd, Pb and As

At a given site: information on metal bioaccessibility can promote a more proportionate and cost-effective remediation of contaminated land

Study area

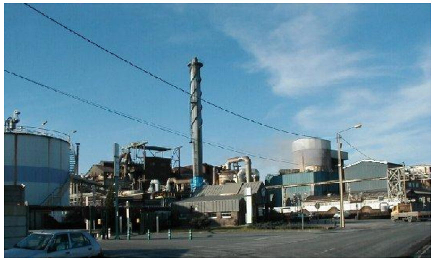
Located around two lead and zinc smelters in northern France

soils highly contaminated by the past smelter activities

Metaleurop Nord

(1894-2003)

Nyrstar



(since 1869)

Contribute to an assessment of the health risks
for inhabitants living on this contaminated area

- Estimation of the metal bioaccessibility (UBM test) through ingestion of soil particles
- **Integrate the most significant soil parameters and their interactions to obtain a model predicting the bioaccessibility of Cd and Pb in the topsoils**

Methods

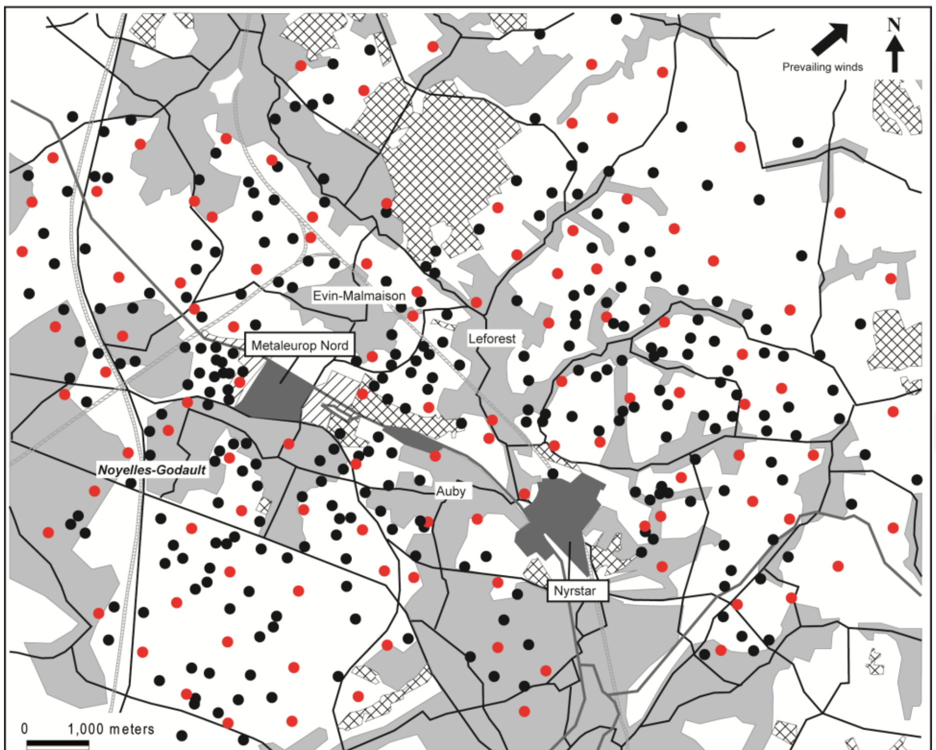
Elaboration of a robust model from an extended database of 390 agricultural topsoil samples

Methods

Elaboration of a robust model from an extended database of 390 agricultural topsoil samples

- *Analytical determinations*
  - ✓ Soil parameters: clay, silt, sand, pH water, total $CaCO_3$, organic matter, assimilated P, free Fe-, Mn-, Al-oxides, total Fe, Mn, Al, Cd, Pb
  - ✓ Bioaccessibility of Cd and Pb in the G and GI phases

- *A statistical modeling approach*



Training set

● **280 samples**

Model calibration

Test set

● **110 samples**

(selected using a regular 1000 x 1000-m grid laid over the survey area)

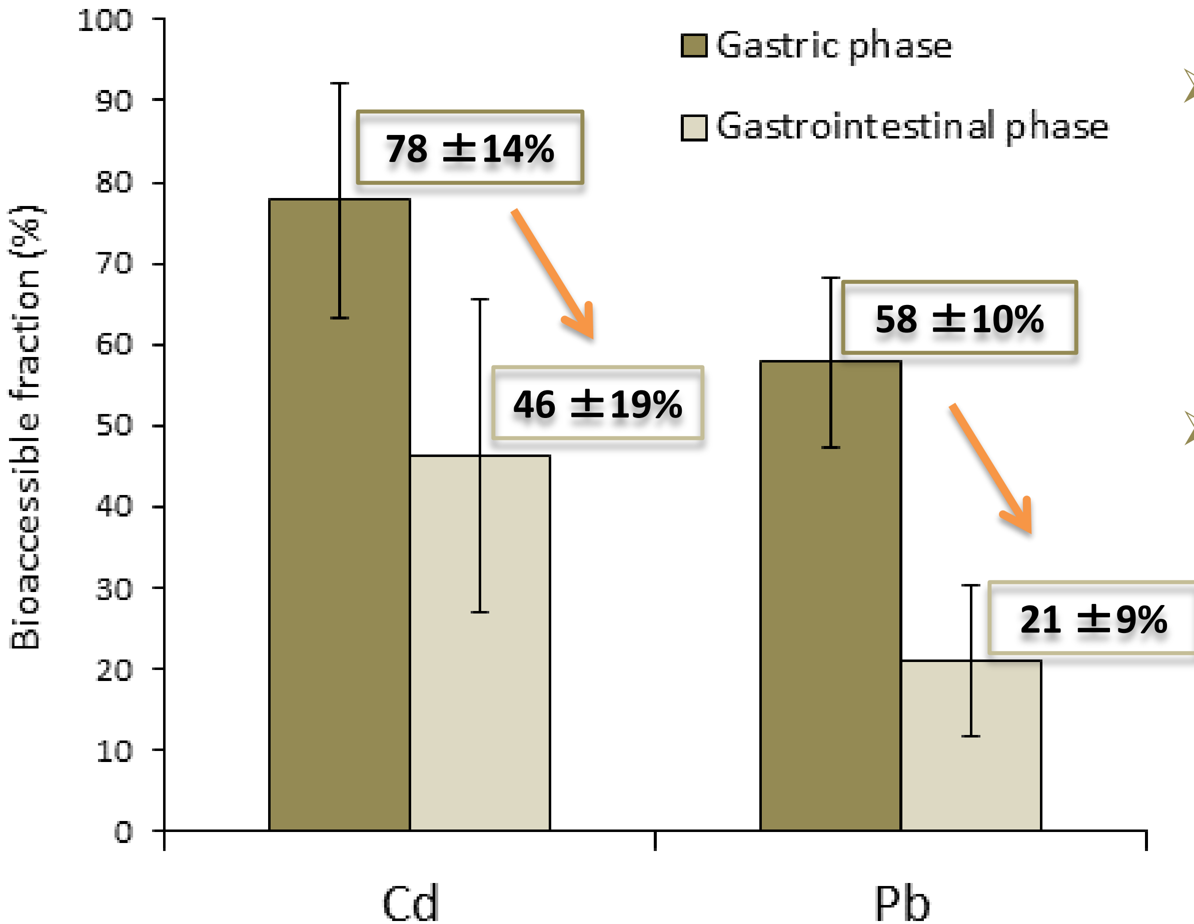
Model validation

Physico-chemical parameters of the 390 agricultural topsoils

		Mean	Min	Max
Clay	g kg^{-1}	198	84	681
Silt	g kg^{-1}	537	94	777
Sand	g kg^{-1}	266	63	782
pH water		7.6	5.0	8.6
Total $CaCO_3$	g kg^{-1}	13.8	1.0	280
Organic matter	g kg^{-1}	37.9	14.9	177
Assimilated P	g kg^{-1}	0.425	0.017	2.659
Free Fe oxide	g kg^{-1}	3.26	0.16	37.3
Free Mn oxide	g kg^{-1}	0.24	0.06	0.56
Free Al oxide	g kg^{-1}	0.32	0.09	1.41
Total Fe	g kg^{-1}	22.1	3.2	71.3
Total Mn	g kg^{-1}	0.37	0.04	0.91
Total Al	g kg^{-1}	9.08	0.56	64.5
Total Cd	mg kg^{-1}	3.5	0.9	15.0
Total Pb	mg kg^{-1}	153	42	854

- Soils around Metaleurop Nord developed mainly on loessic materials
- Soils around Nyrstar originated from alluvial deposits
- Large variability in terms of contamination degree and physico-chemical parameters

Bioaccessibility of Cd and Pb in the 390 agricultural soils



- A sharp decrease in extracted Cd and Pb as they moved from the G phase to the GI phase
- The mean values of the bioaccessible fractions of Cd and Pb in the GI phase < 50%

A significant portion of the total soil-borne metal concentration may not be available following incidental soil ingestion

$$\text{Bioaccessibility (\%)} = \frac{[\text{metal}]_{in\ vitro}\ (\text{mg kg}^{-1})}{[\text{metal}]_{tot}\ (\text{mg kg}^{-1})} \times 100$$

Model calibration from the training set (280 samples)

To study the dependence of the Cd and Pb bioaccessibility on the corresponding total metal concentrations and soil physico-chemical parameters

1 Select the relevant physico-chemical parameters to describe metal bioaccessibility

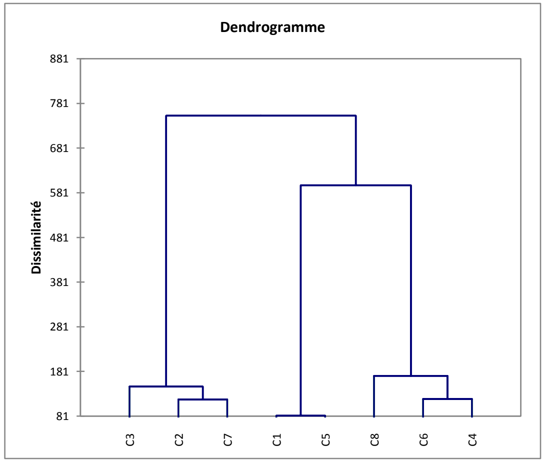
Statistical analysis
(Pearson's correlation, multicollinearity effects)

Sand
Total $CaCO_3$
Organic matter
Assimilated P
Free Fe-, Mn-oxides
Total Fe, Al

Soil parameters included in the multiple regression analysis

2 Define the best population to be taken into account among the 280 soil samples and construct multivariate regression models

Statistical analysis
(stratified sampling, ascending hierarchical clustering)



8 well-differentiated groups

Model calibration from the training set (280 samples)

Several steps

AHC

Group	*n*	Size (%)
1	26	9.3
2	17	6.1
3	86	30.7
4	16	5.7
5	90	32.1
6	16	5.7
7	23	8.2
8	6	2.1
Total	280	100

Step 1: Use the smallest sample population by taking:

- ✓ one random soil sample in the smallest group (group 8)
- ✓ n random samples in the other groups in relation to the size of the group

Elaboration of models (multiple regression analysis) with the selected samples (n = 47)

Step 6: Elaboration of models with the full set of 280 samples

At each step

Use of the full set of 280 samples to test and calibrate the models elaborated

Model calibration from the training set (280 samples)

How do we choose the best predictive models among the 6 steps?

Use of statistical indicators

$$\mathrm{RMSE} = \left[\frac{1}{n}\sum_{i=1}^{n}(B_i - B_i^*)^2\right]^{1/2}$$

Root mean square error

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}(B_i - B_i^*)$$

Mean error

$$r_p^2 = 1 - \frac{\sum_{i=1}^{n}(B_i^* - B_i)^2}{\sum_{i=1}^{n}(B_i - \overline{B})^2}$$

Coefficient of determination

B_i	Measured bioaccessibility at location *i*
B_i^*	Predicted bioaccessibility at location *i*
$\overline{B}$	Mean of the measured values
n	Sample size (*280*)

The best predictive models

Cd G et GI → **Step 4 (*n* = 186)**

Pb G et GI → **Step 5 (*n* = 233)**

Model calibration from the training set (280 samples)

Results (significance level: $p < 0.0001$)

Cd G = $-0.053 + 0.012\ Al_{tot} + 0.75\ Cd_{tot}$ $R^2 = 0.90$

Cd GI = $0.56 - 0.0024\ CaCO_3 + 0.035$ Free Fe oxide $+ 0.24\ Cd_{tot}$ $R^2 = 0.46$

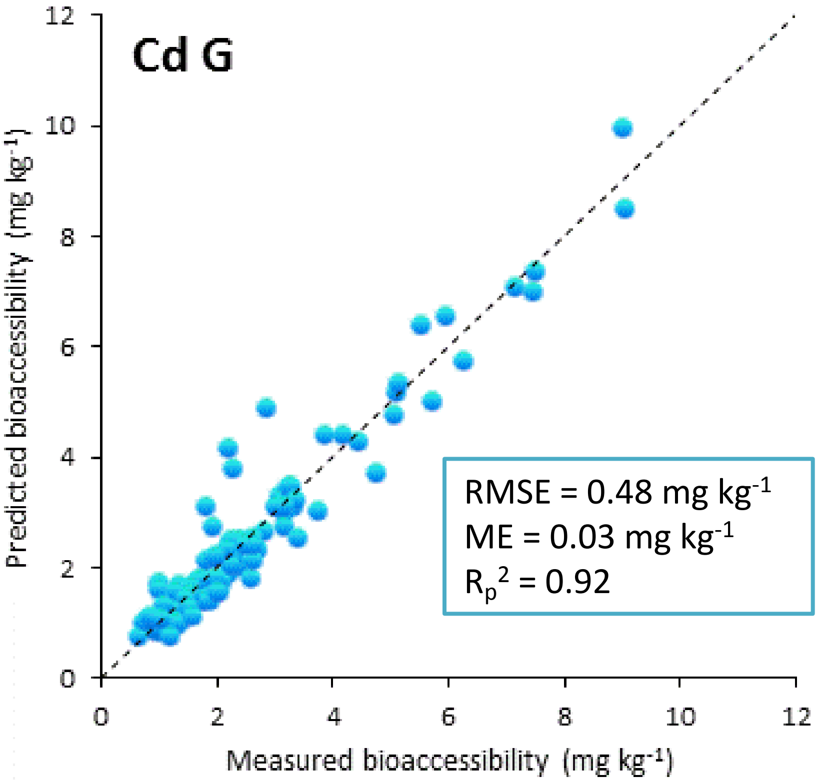
Pb G = $6.06 - 0.25\ CaCO_3 - 0.42$ OM $+ 16.08$ Assimilated P $- 0.88\ Al_{tot} + 0.67\ Pb_{tot}$ $R^2 = 0.90$

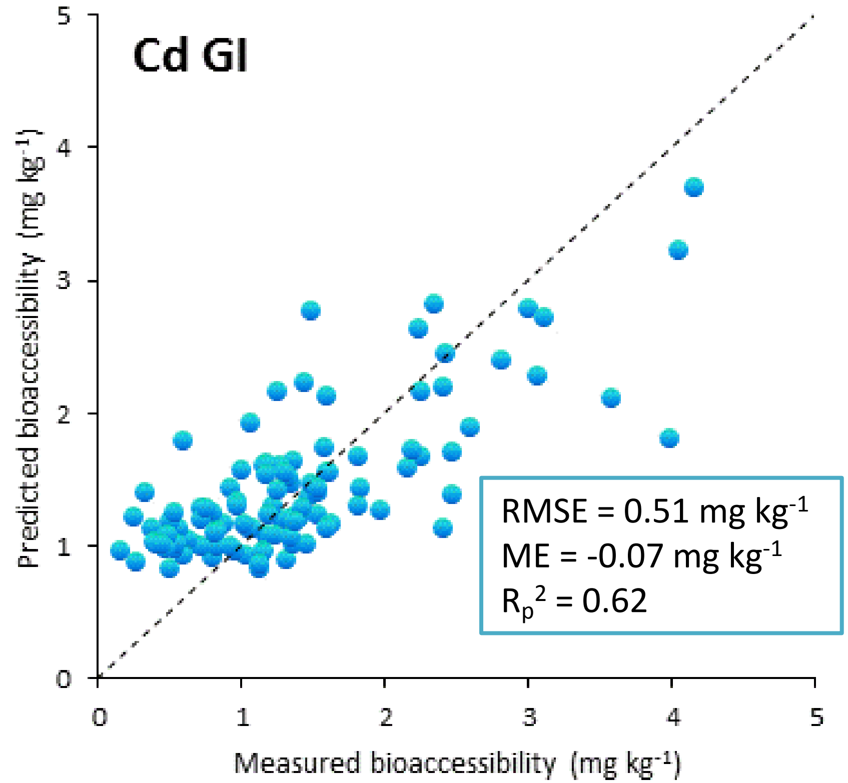
Pb GI = $-5.12 + 17.07$ Assimilated P $+ 0.19\ Pb_{tot}$ $R^2 = 0.65$

Significant correlations between concentrations extracted with UBM, total metal concentrations, and selected physico-chemical parameters

Model validation from the test set (110 samples)

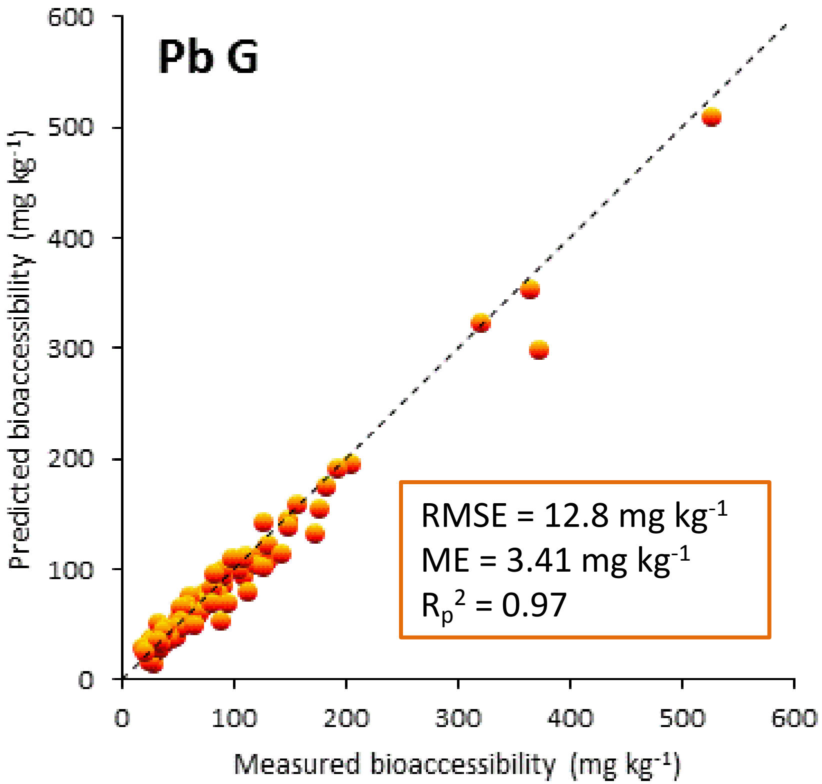
Use of the test set to testify the ability of the regression models

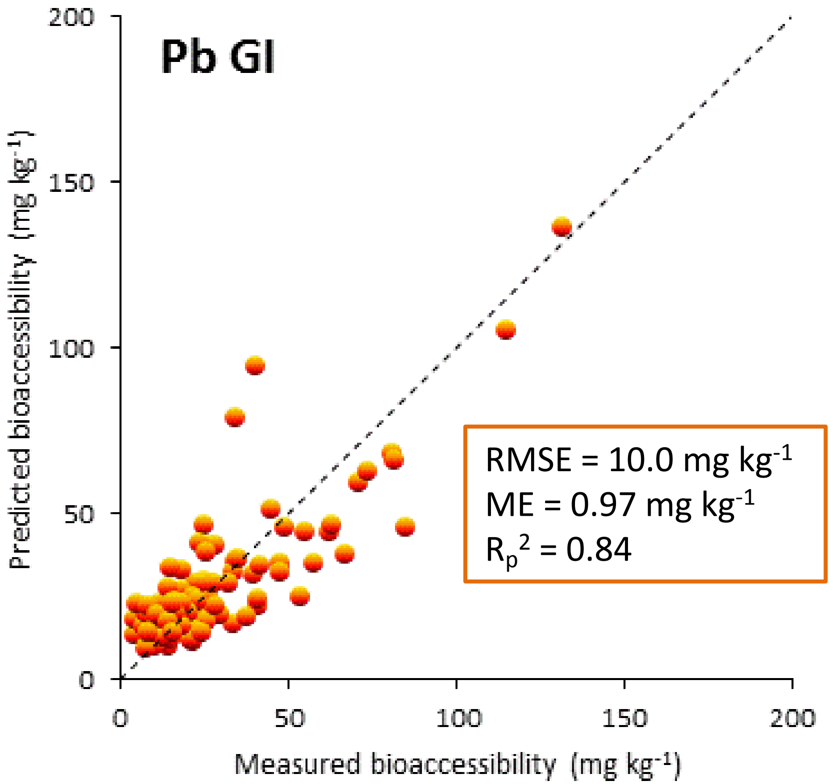




- In the G phase: the model can predict Cd bioaccessibility quite successfully (r_p^2 = 0.92)
- In the GI phase: the model shows a lower predictability (r_p^2 = 0.62)

Model validation from the test set (110 samples)





- In both phases: the models can predict Pb bioaccessibility quite successfully ($r_p^2 > 0.84$)
- In the G phase: the highest prediction accuracy

For all models, RMSE and ME values are low	→	A satisfactory and effective calibration of the models

Conclusions and perspectives

- Elaboration and validation of robust models on agricultural soils contaminated by the past atmospheric emissions of two smelters

- The main variables governing metal bioaccessibility:
  - total carbonates
  - organic matter
  - assimilated P
  - free Fe-oxide
  - total Al, Cd and Pb

- Next stage: Testify the ability of models on data with a larger variability in according to uses (residential, woody habitats) and contamination ways of soils

- These models are based on soil parameters easily measured in laboratory

  ➡ Fast and cost-effective assessment of bioaccessible fractions of Cd and Pb in soils

- Contribution to the management of these contaminated soils

Thank you very much

for your attention!