Efficient models for predicting the non-compliance of food crops with regulation limits for metallic contaminants

Christophe Nguyen, Laurence Denaix, Emma Vivien, Jean-Yves Cornu, Agathe Roucou, Benoît Méléard

Efficient models for predicting the non-compliance of food crops with regulation limits for metallic contaminants

Nguyen[1] C., Denaix[1] L., Vivien[1] E., Cornu[1] JY., Roucou[2] A., Méléard[2] B.

1: UMR 1391 Ispa, Inrae, Bordeaux France INRAE

2: Arvalis Institut, Boigneville France ARVALIS

christophe.nguyen@inrae.fr

Will this future harvest comply with the regulation limits for toxic metals in food products?

Background

- Increasing concerns about the dietary exposure of humans to toxic metals, especially cadmium (Cd)
- Regulation for toxic metals in food products becomes more strict (EC1323/2021)
- The rates of non-compliance of harvests become more frequently critical (ex: Cd in Durum wheat, lineseed)
- Prediction of non-compliance of harvests with regulation limits for metals is needed for anticipating corrective actions

State of art & Hypothesis

- Bioavailability of contaminants in soils is generally the main factor governing the contamination of a given plant species (McLaughlin *et al.*, 2021)
- Bioavailability of many cationic metals in soils is strongly governed by the metal content, the pH and by the organic matter (Antoniadis *et al.*, 2017)

Hypothesis
The compliance of a future harvest with regulation limit for toxic metals can be reliably predicted by binary classification models by using the soil metal content, the pH and the organic carbon content

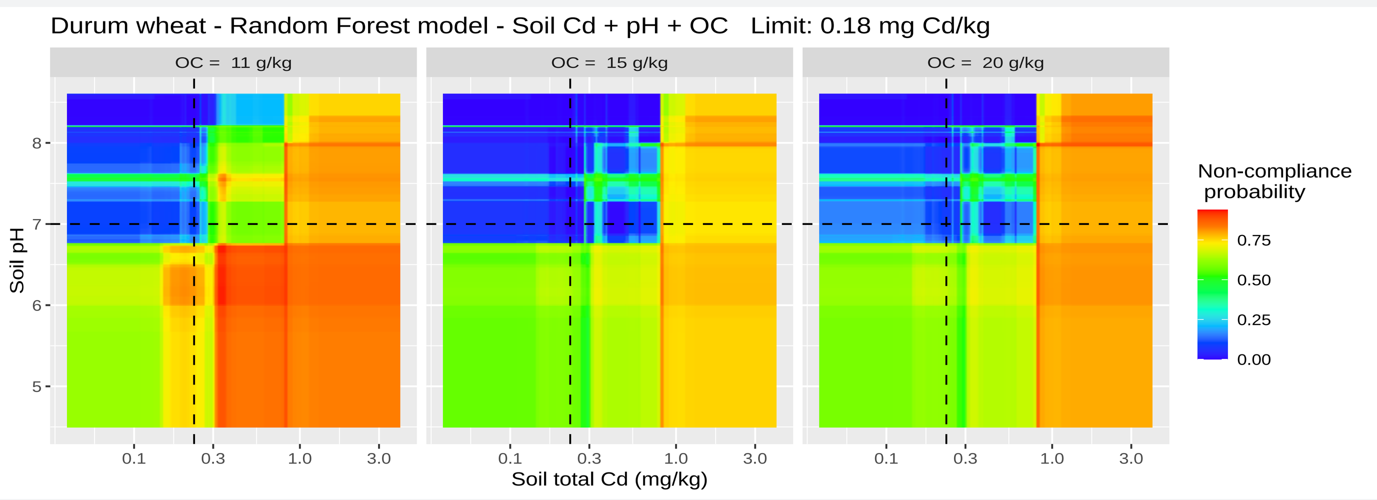
Methods

- Modelling of the probability of compliance from datasets of paired soil + plant samples (Durum and bread wheat, lettuce, potato) from research projects and from literature
- Soil predictors are: the total (pseudo-total) metal content, pH(water), organic carbon content (OC), clay+loam when available
- Use of machine learning classifiers: Logisitic regression, random forest
- Cross-validation (K=5 folds, repetitions=100) for estimating the model performances

Proof of concept for Cadmium

Main Results

Random Forest models



Dashed line: median for French cropland soils

- Random Forest models simulate non-realistic discontinuous effects of soil Cd, pH and OC due to over-parameterization

Logistic models: Performances

$$log\left(\frac{p}{(1-p)}\right) = a_0 + a_1.log_{10}(Cd_{soil}) + a_2.log_{10}(H^+_{soil}) + a_3.log_{10}(OC_{soil})$$

p: probability of non-compliance

Model Performances estimated by 5-folds cross validation (100 repetitions)

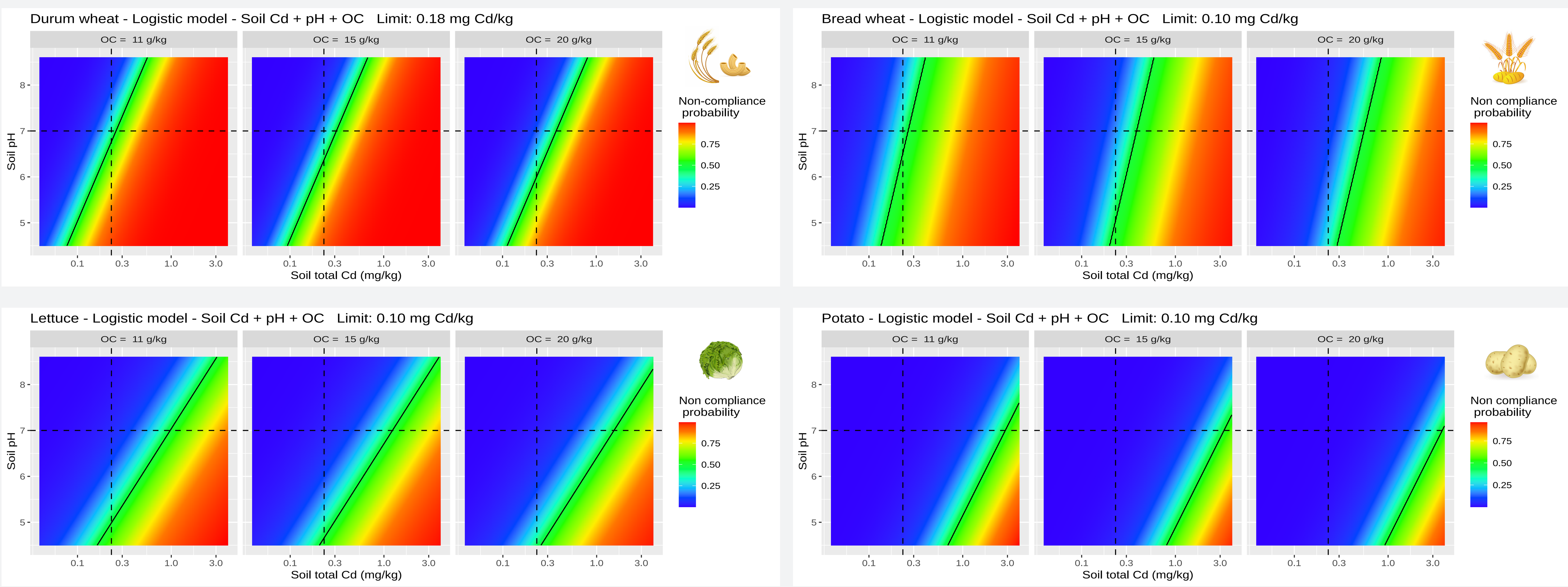
	Durum wheat (n=476)	Bread wheat (n=441)	Lettuce (n=106)	Potato (n=106)
Correct predictions (Accuracy, %)	84.9	85.9	92.5	78.3
Detected non compliance (% of actuals)	75.8	37.3	72.2	75.0
Detected compliance (% of actuals)	87.0	93.5	96.6	79.5
Correct predicted non-compliance (% of predicted)	58.0	46.8	81.2	56.8
Correct predicted compliance (% of predicted)	93.8	90.6	94.4	89.9

- Logistic models with soil Cd, pH and OC gives good and valuable performances for detecting compliant and non-compliant cases for Durum wheat, Lettuce and Potato. Detection of non compliance for Bread wheat is less good. More predictors are likely required.
- For Durum wheat adding clay or loam did not improve the models
- Models can surestimate the risk of non-compliance (low robustness for predicted non-compliance), which is protective.

Logistic models: Predictions

- Soil Cd and the pH are more influential than the OC content
- Ranking of the risk of non-compliance: Durum wheat > Bread wheat > Lettuce > Potato
- The effect of soil Cd, pH and soil OC depends on the regulation limit and on the species potential for Cd accumulation
- For the 4 crops, **simple guidelines** would be:
  Safe conditions :
  - soil Cd <0.1 mg/kg
  - Soil Cd<0.3 mg/kg AND soil pH >7

  Risky conditions :
  Soil Cd>2 mg/kg AND pH<6
- The Durum wheat model is freely available at: https://ispa.bordeaux.inra.fr/services/blesur/



Dashed line: median for French cropland soils *Solid line: Non compliance probability p=0.5 boundary*

Conclusions

- The good performances of the models predicting the compliance with Cd limit allow to have tools to anticipate problems with risky soils by choosing varieties with a lower accumulation potential
- The approach could be extended to other crops and other metals such as Ni to anticipate the future regulations
- Open access databases of paired soil+crop analyses are highly desirable to build and test the models

References:
Antoniadis *et al.*, 2017. Earth-Science Reviews 171, 621–645. doi:10.1016/j.earscirev.2017.06.005
McLaughlin *et al.*, 2021. Advances in Agronomy. pp. 1–129. doi:10.1016/bs.agron.2020.10.004

Fundings:
- Arvalis, Cadur CR127518
- French National Research Agency (ANR): Cadon, ANR-15-CE21-0001-04, Blésur LabEx COTE ANR-10-LABX-45
- French Ministry of Agriculture: Multicontamination of crops CASDAR n°10054, Quasagro CASDAR n°5423

anr agence nationale de la recherche