AI4EO CHALLENGE

SEEING BEYOND THE VISIBLE

06/07/2023

INTRODUCTION

AI4EO Challenge : Seeing Beyond the Visible

Organizers and partner

Autonomous space solutions company (Poland)

European Space Agency

Precision farming software company (Poland)

Goal

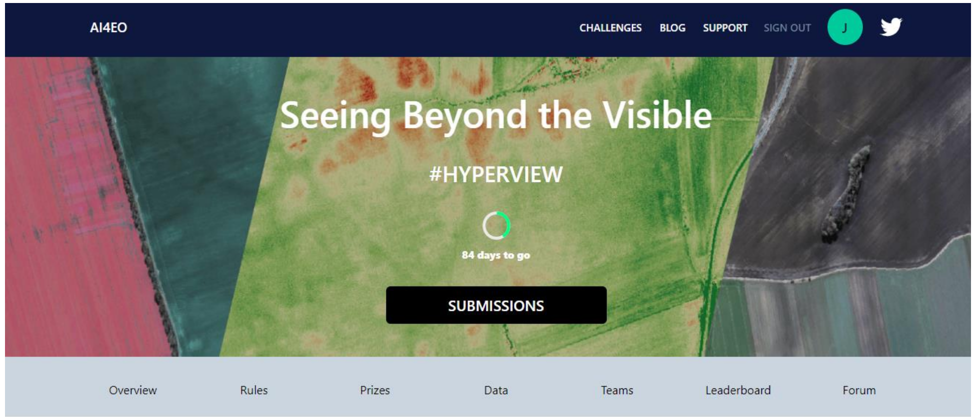
Generate a model for prediction of **soil parameters** from **hyperspectral imaging** data of agricultural areas

Timeline

From the 9^{th} of February to the 1^{st} of July 2022

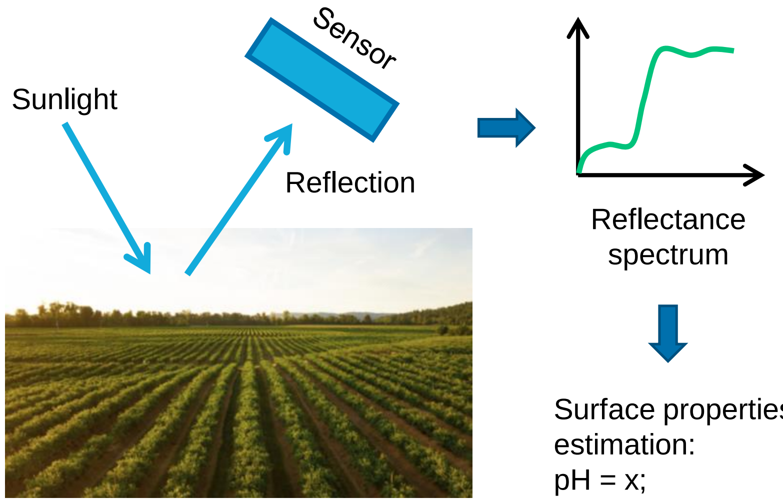
Prizes

Best predictive model will be deployed on Intuition-1 satellite set to launch in Q1 2023
Top 5 teams invited for joint-paper publication



Soil parameter estimation for sustainable farming

- As the global population is increasing, there is a growing need for more **efficient** methods of **food production**
- **Macro-nutrients** (N, K, P, …) have a direct impact on agriculture yields, and are commonly supplied in the form of **fertilizers**
- Excessive use of fertilizers has a **detrimental** impact on **biodiversity** and **public health**
- Methods of soil property **estimation** from **reflectance spectra** will help minimize resource input and make farming more sustainable (precision farming)



Surface properties estimation:
pH = x;
N = y mg/km;
K = z mg/km;
…

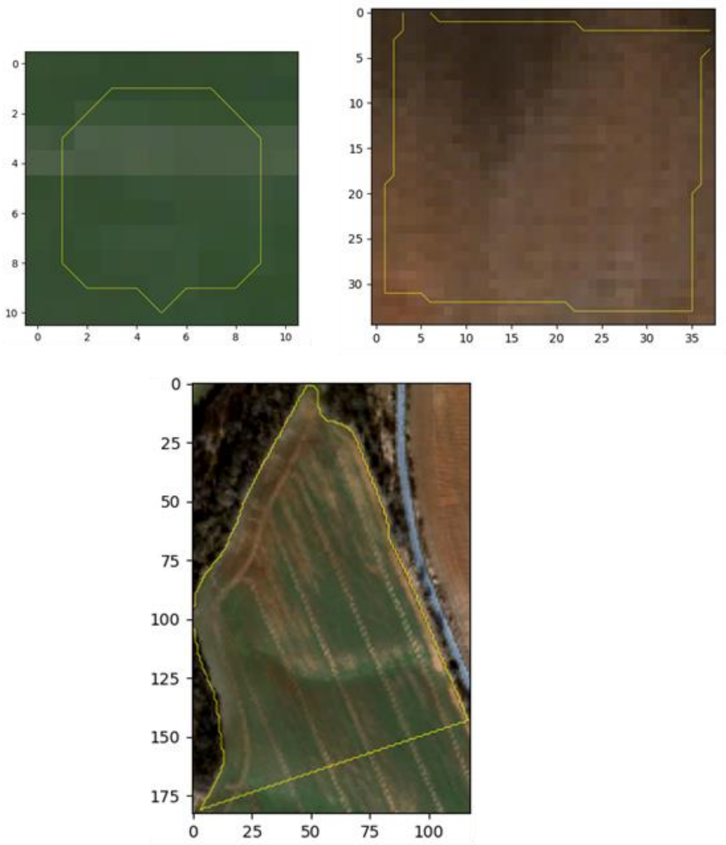
Resources :
Chabrillat, et al. Imaging Spectroscopy for Soil Mapping and Monitoring. Surv Geophys 40, 361–399 (2019). https://doi.org/10.1007/s10712-019-09524-0
Burton, et al. 2020, J. Electrochem. Soc. https://doi.org/10.1149/1945-7111/ab6f5d

DATA SET ACQUISITION

Challenge dataset

Training set
1732 patches
Known soil parameters

Testing set
1154 patches
Unknown soil parameters



RGB representation of patches :

Yellow outline : masks
delimiting parcels

Soil Parameters :

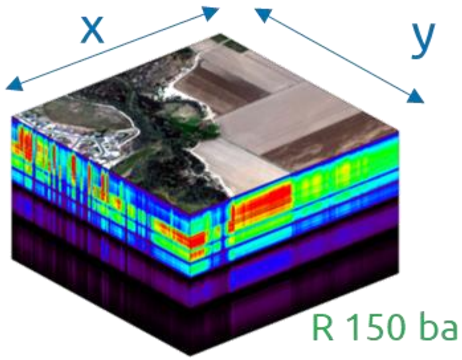
P2O5 : Phosphorus content of fertilizers
K : Potassium
Mg : Magnesium
pH : measure of soil acidity

Spectral range :

150 bands : 462-942nm, 3.2nm resolution
Visible range and early Near-Infrared

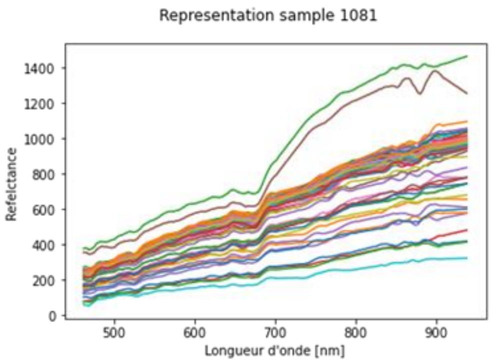
Spatial resolution :

1px = 4m²



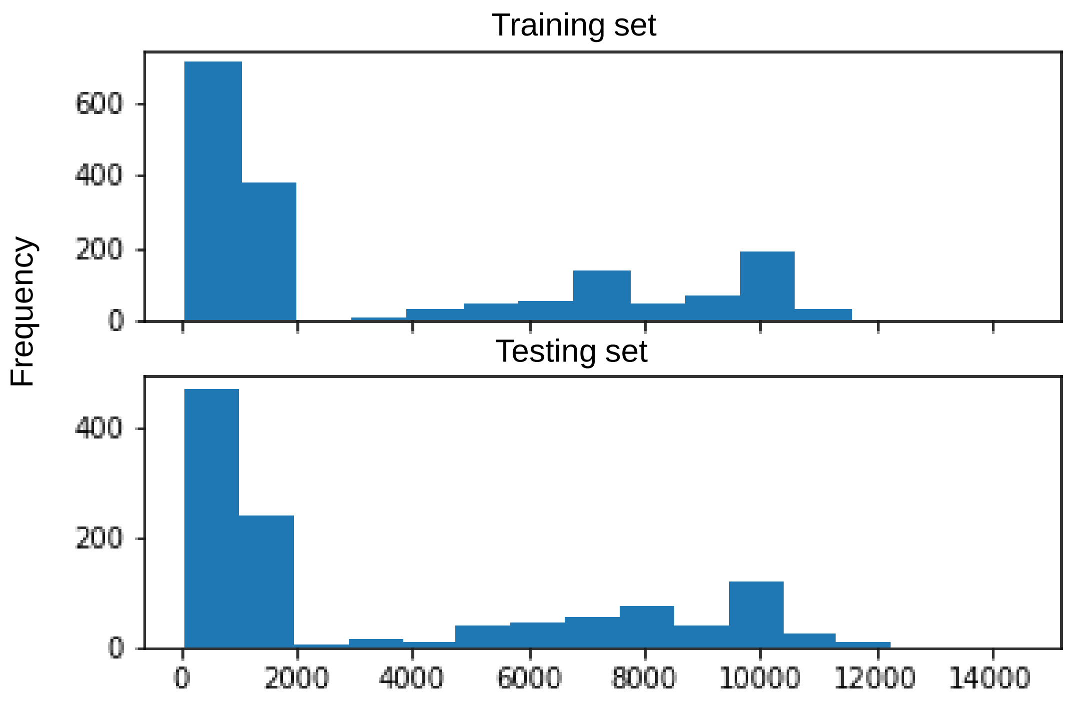
R 150 bandes

3D

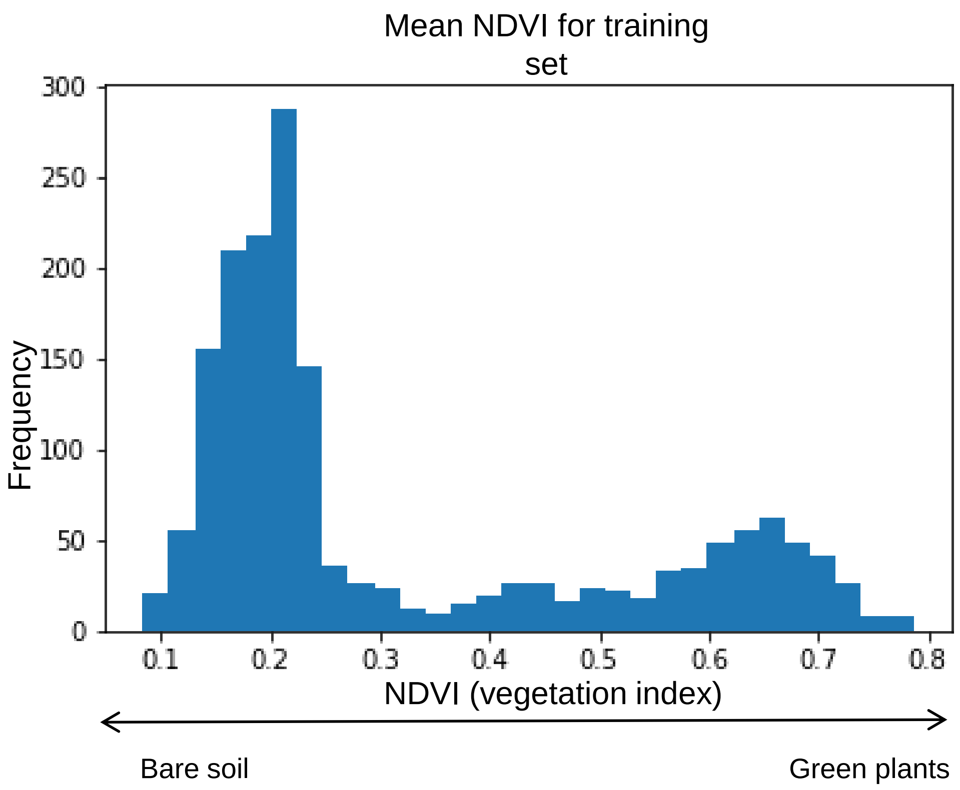


2D

Limited,high dimension and heterogeneous



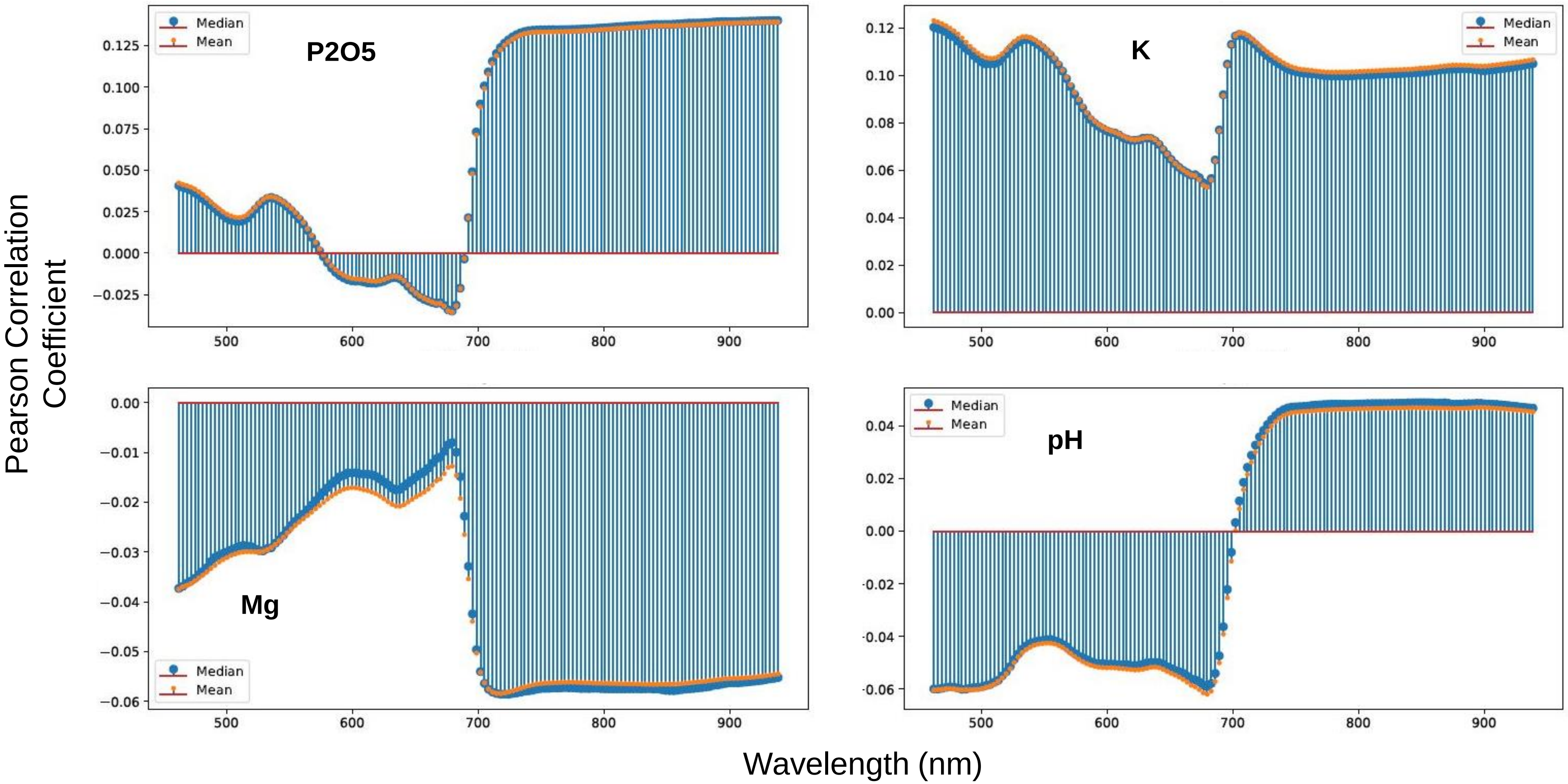
Répartition des images par taille
(Nb de pixels)



NDVI = (ρNIR - ρRed) / (ρNIR + ρRed)

Reflective content is **heterogeneous**

Correlation between band reflectance and soil parameters



Very **Weak correlation** (< 0.12) between reflectance per wavelength and soil parameters
Note : Correlation between Vegetation Index and soil parameters also low : ~0 R-squared value

How to develop a predictive model for soil parameters from hyperspectral imaging data ?

Challenge scoring of testing set parameters predictions :

$$Score = \sum_{i=1}^{4} \left(\frac{MSE_i}{MSE_i^{base}}\right) /4 \qquad with:\ MSE_i = \frac{\sum_{j=1}^{|\psi|} \left(p_j - \hat{p}_j\right)^2}{|\psi|}$$

Challenge score : Ratio of MSE (mean square error) of predictions over MSE of base solution, averaged over the four parameters

Lower score = **Better** predictions

How to develop a predictive model for soil parameters from hyperspectral imaging data ?

Challenge scoring of testing set parameters predictions :

$$Score = \sum_{i=1}^{4} \left(\frac{MSE_i}{MSE_i^{base}}\right) /4 \qquad with:\ MSE_i = \frac{\sum_{j=1}^{|\psi|} \left(p_j - \hat{p}_j\right)^2}{|\psi|}$$

Challenge score : Ratio of MSE (mean square error) of predictions over MSE of base solution, averaged over the four parameters

Lower score = **Better** predictions

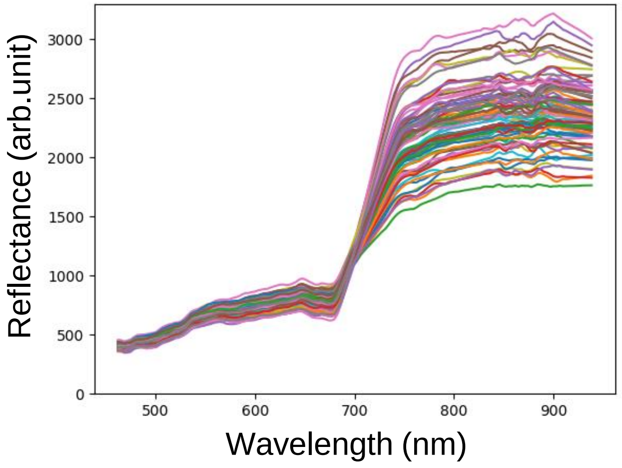
Main obstacles :

- Low **sample size** (1732 for training set), high **heterogeneity**
- **High dimensionality**
- Those soil nutrients usually do **not** display clear **absorption features**

DATA
PRE-PROCESSING

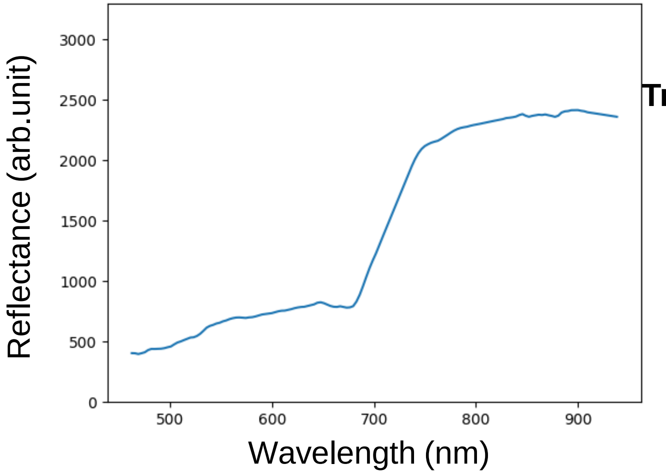
Pre-processing

Raw spectra of all pixels of an image

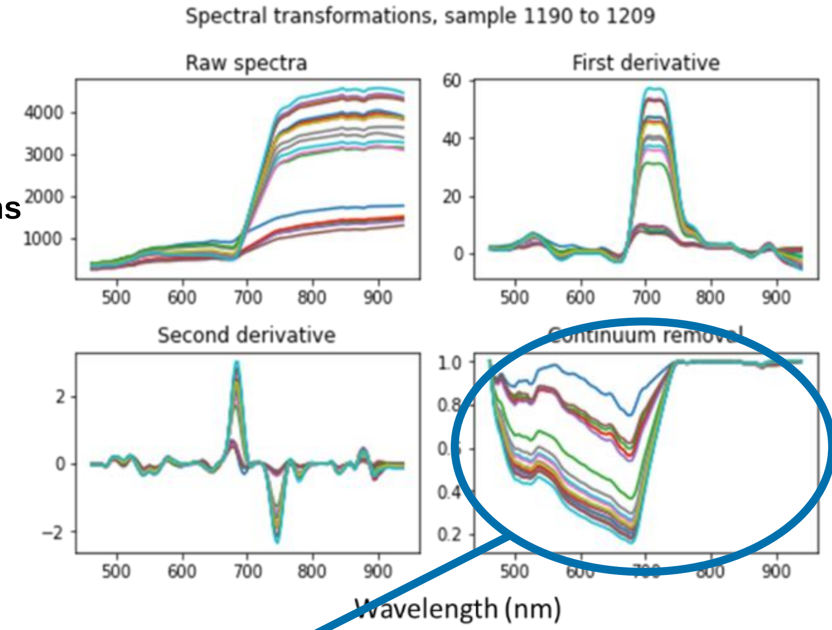


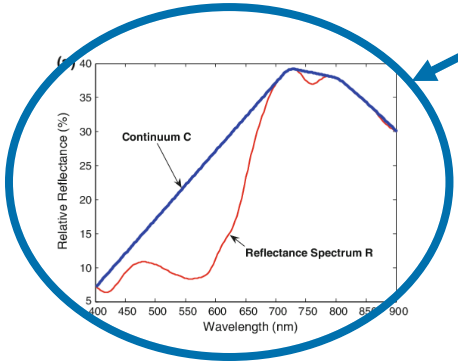
Average spectrum

Average reflectance spectrum of an image



Transformations



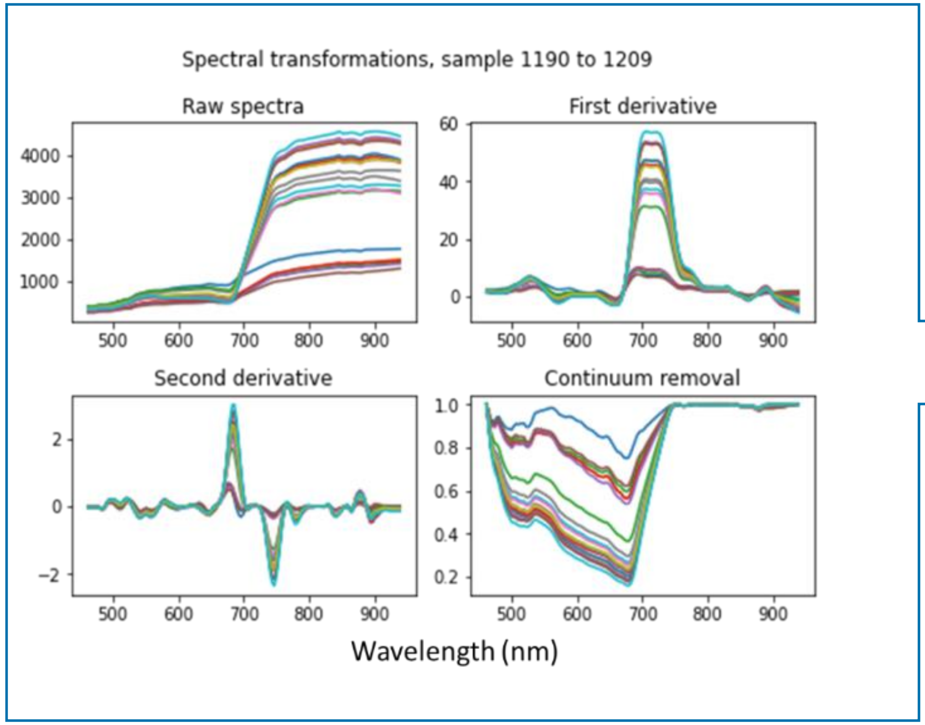


- An image is represented by a vector of shape (x, 150) where x is the number of pixels

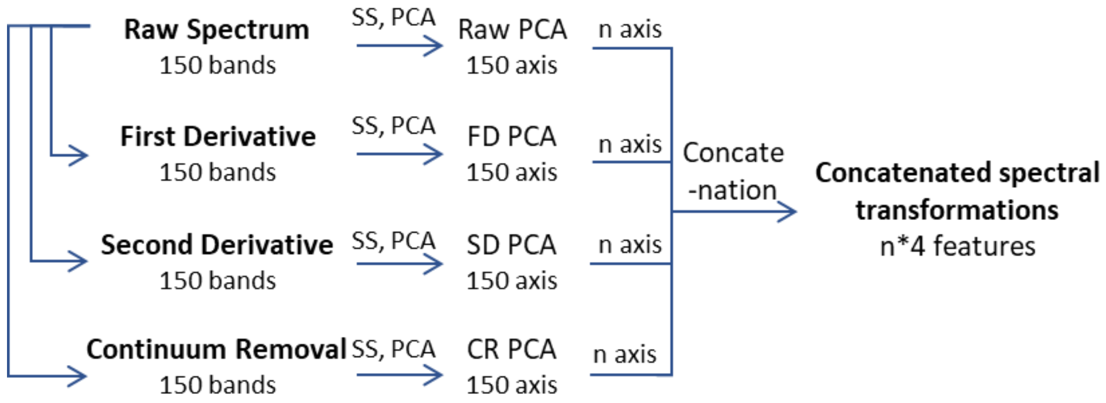
- It is now represented by a (1, 150) vector

- Finally it is represented by a (1, 600) vector (150 dims x 4 transformations)

Pre-processing



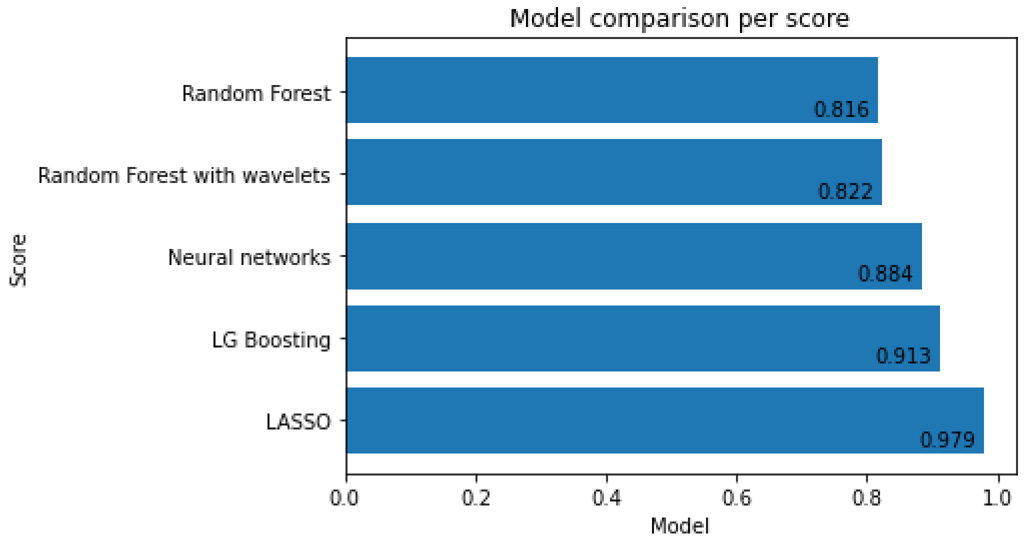
PCA reduction des dimensions



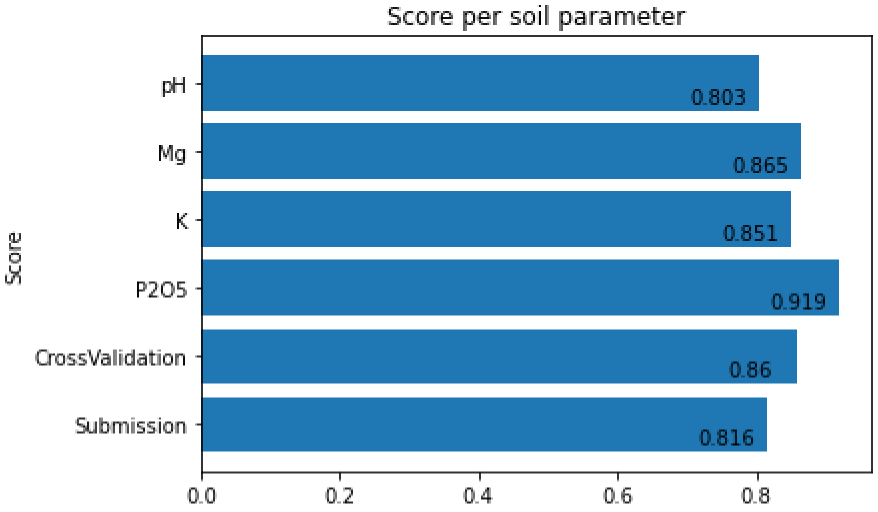
RESULTS

RESULTS

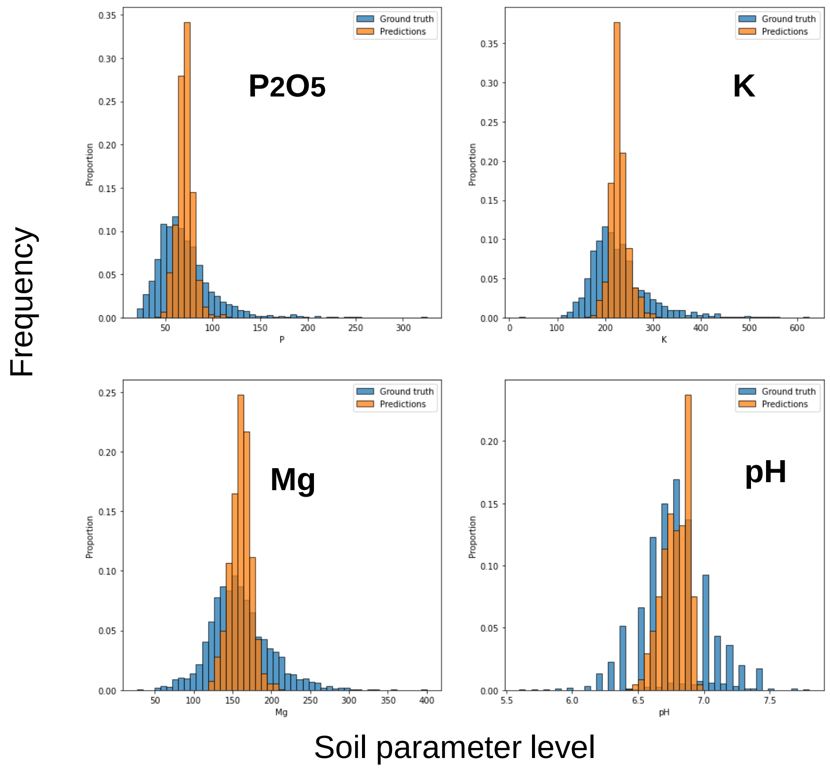
- ✓ Several methods of Machine Learning have been implemented, in order to choose the best prediction model.
- ✓ Using cross validation to choose the best hyperparameters of ML models and analyse the contribution of spectrum transformations and dimension reduction
- ✓ Best submission scores :



Random Forest predictions



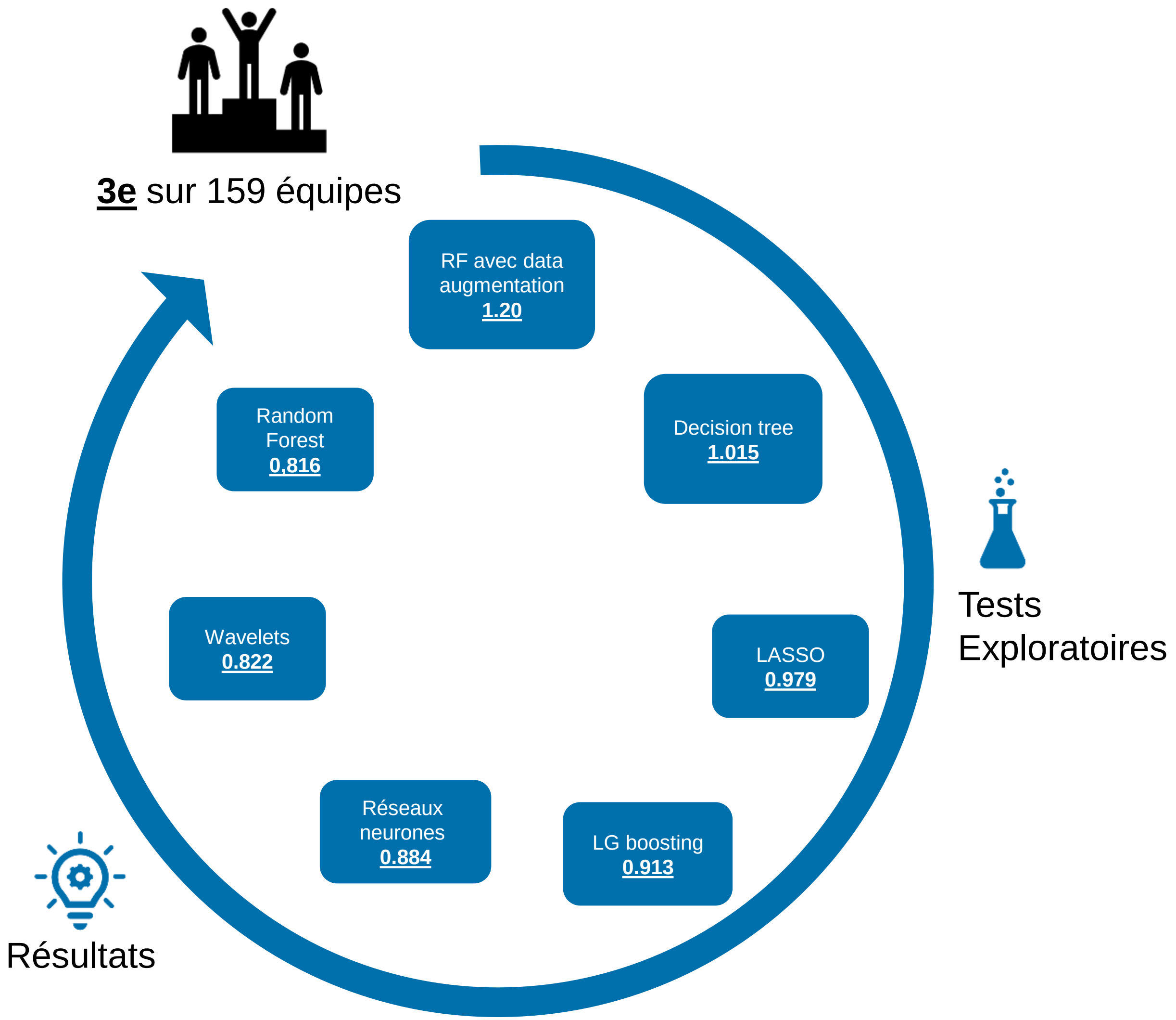
Training truth
RF prediction output



- ✓ **Prediction** distribution centred around **parameter mean**
- ✓ Model performs better for **pH**, worse for **P2O5**

RESULTS

LEGENDE	Objectif initial



Model	Date	Score officiel	Cross-validation	P2O5	K	Mg	pH	Pre-processing	Parameters	Comments
RandomForest		0,816	0,860	0,919	0,851	0,865	0,803	Mean of sample > (Transfo : FD/SD/CR) > StandardScaler > PCA > Concatenation of n first PCA axis for each transfo (n*4 feature total)	RF : bootstrap = False, max_features = log2 (P) / sqrt (K, Mg, pH), n_estimator = 10000, min_samples_leaf = 2/4	40 pca axis for P (160 total), 35 for K, Mg (140 total), 50 for pH (200 total) 2ème soumission (19/06) : 0,817

CONCLUSION

CONCLUSION

- Ø First step for sustainability agriculture
- ➢ Asset to France 2030 program support by Science&IA, C&CA offer
- ➢ Prospects : Space Agencies, Env & Agriculture Ministry, Carbon Farming

MERCI DE VOTRE ATTENTION