Sensitivity of soil property prediction obtained from VNIR/SWIR Lab data to spectral configurations

K.R.M. Adeline[1], C. Gomez[1], N. Gorretta[2], J.-M. Roger[2]

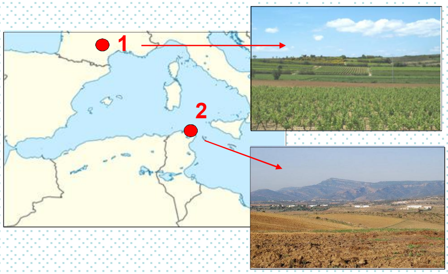
[1] IRD, UMR LISAH (INRA-IRD-SupAgro), 34060 Montpellier, France
[2] IRSTEA, UMR ITAP, 34196 Montpellier, France
Contact: Karine.adeline@supagro.inra.fr

1) CONTEXT AND OBJECTIVES:

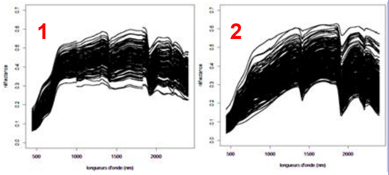
Research in agriculture precision and environmental monitoring leads to the observation of several physical and biophysical processes in the soil conditions that requires at least to know the soil structure and composition. Thereafter, the use of **VNIR-SWIR Laboratory spectroscopy (350-2500nm)** has proven to be a good alternative to costly physical and chemical laboratory soil analysis. As well, the number of studies using VNIR-SWIR hyperspectral airborne imaging in soil property mapping has also increased (e.g. [1]). The main issue is now to achieve to transfer these promising results to future satellite and Unmanned Aerial Vehicle (UVA) data. As such, the objective of this study is to assess the **sensitivity** of soil property prediction results to different **spectral configurations** (including the spectral resampling [2], the spectral resolution and the number of spectral bands); which may offer a first insight of the **potential of future hyperspectral UAV and satellite sensors (i.e. HYPXIM, PRISMA, Shalom, ENMAP and HyspIRI) for soil applications and mapping**.

2) MATERIALS:

- Study area:
  2 Mediterranean sites with different soil environments
  - Site n°1: Peyne, in France (0.91km2)
  - Site n°2: Lebna, in Tunisia (300km2)
- Collection of **in-situ soil samples** (site, number, year):
  - Peyne, M = **148**, 2010
  - Lebna, M = **262**, 2008-2009-2010



- **Soil reflectance spectra** measurements:
  - ASD (Analytical Spectral Devices Inc.)
  - spectral range: 350-2500nm
  - number of spectral band: 1961 (= $\mathbf{N_{init}}$)
  - spectral sampling: 1nm



- Physico-chemical analysis for **4 soil properties**:
  - Clay content (granulometric fraction < 2µm)
  - Calcium Carbonate (CaCO3)
  - Iron oxides
  - pH
- **Correlation relationships** between the soil properties:

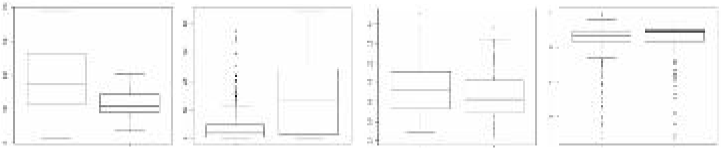
Peyne

	Clay	CaCO3	Iron	pH
Clay	1.00	-0.13	0.53	-0.04
CaCO3		1.00	-0.56	0.68
Iron			1.00	-0.27
pH				1.00

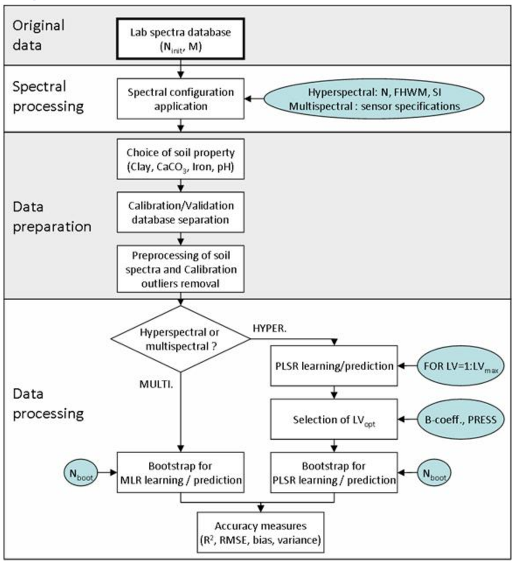
Lebna

	Clay	CaCO3	Iron	pH
Clay	1.00	0.14	**0.85**	0.31
CaCO3		1.00	-0.07	0.59
Iron			1.00	0.11
pH				1.00

- Soil property distribution among site (T: Lebna, P: Peyne):



3) METHODOLOGY:



- A spectral configuration is defined by 3 parameters, **the number of spectral bands (N), the spectral resolution (FHWM) and the spectral interval sampling (SI)**. A gaussian shape filter is used for resampling the soil spectra from 440-2400nm.
- The **hyperspectral** configurations have **FHWM = SI**, except for the reference Init_1/1 (ASD spectra)

	Init_1/1	Config_3/10	Config_5/10	Config_10/10	Config_40/40
SI (VNIR)	1	3	5	10	40
SI (SWIR)	1	10	10	10	40
FHWM (VNIR)	3	3	5	10	40
FHWM (SWIR)	10	10	10	10	40
N	1961	328	253	197	50

- The **multispectral** configurations are based on the satellite specifications spectral filters of **ASTER** and **LANDSAT-7**

ASTER spectral specifications (N=9)

Spectral bands	560	660	810	1650	2165	2205	2260	2330	2395
Spectral resolution	80	60	100	100	40	40	50	70	70

LANDSAT-7 spectral specifications (N=6)

Spectral bands	480	565	660	825	1650	2220
Spectral resolution	65	80	60	150	200	260

- 2/3 of total samples (M) are selected for calibration and 1/3 for validation
- The spectra reflectance are converted into absorbance and the data are mean-centered
- Outliers are removed after Principal Component Analysis and Mahalanobis distance computation

- **Multivariate Linear Regression** (MLR) is performed with multispectral configurations and **Partial Least Square Regressions** (PLSR) for hyperspectral ones to deal with collinearity variable issues
- The selection of the optimal number of Latent Variable (LV_{opt}) is assessed by observing PLSR prediction over a given range of LV (LV_{max} set to 10) based on 2 criteria: the minimum of Prediction Residual Error Sum of Squares (PRESS) and the divergence of the PLSR coefficients (b-coeff.)
- The prediction accuracy measures are computed with the **boostrap** procedure applied on MLR/PLSR learning and prediction with a repetition of 99 (N_{boot})

4) RESULTS:

- For **LANDSAT** configuration, prediction is inaccurate for Iron, pH and $CaCO_3$ (lack of the 2340nm band), except for Clay (P) with a high mean content and distribution
- **No significant degradation from ASD initial configuration to hyperspectral configurations until ASTER configuration** for predicting soil properties with high mean content/distribution and with a spectral signature such as Clay (P+L), Iron (P+L) and CaCO3 (P)
- **Soil properties having a short spectral absorption feature are sensitive to the spectral resolution and central band** such as CaCO3 (P+L) between Init_1/1, Config_3/10 and Config_ASTER
- **Soil properties with weak mean/distribution are unpredictable** like CaCO3 (L) and pH (L), also if they do not have a spectral signature like pH
- The **prediction performances of soil properties without spectral signature** (pH) **decrease along the configurations** since the number of spectral bands decreases
- The role of **correlation relationships between soil properties** can **increase prediction performances** like between pH (P) and CaCO3 (P), as well as **correlation between spectral bands** among spectra

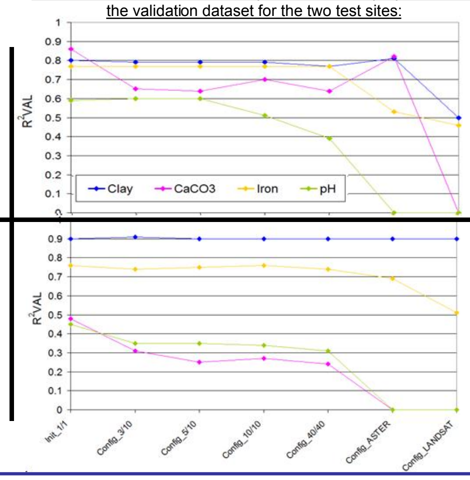
Site:

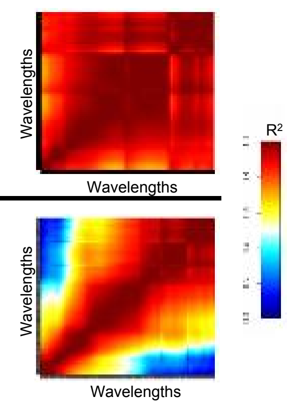
Determination coefficient for the prediction of soil properties over the validation dataset for the two test sites:

Spectral correlation coefficient among the M samples for the two test sites:

Peyne (P)

Lebna (L)





Configuration: Init 1/1

5) CONCLUSIONS:

- Prediction performances are dependent of the initial soil property mean content and distribution, soil property correlation relationships, and correlation between spectral bands that could be site-specific
- Following the good results of multispectral scenarios (ASTER), are **spectroscopic instruments over-designed for soil characterization**? A solution might be **spectral feature selection** [3]
- Perspectives: **impact of spectral configurations to hyperspectral airborne data** for soil property mapping

REFERENCES:

[1] Gomez, C., Lagacherie, P., Coulouma, G., 2008. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. Geoderma, 148(2), pp.141-148.
[2] Peng, X., Shi, T., Song, A., Chen, Y., Gao, W., 2014. Estimating Soil Organic Carbon Using VIS/NIR Spectroscopy with SVMR and SPA Methods. Remote Sensing, 6(4), pp. 2699-2717.
[3] Vohland, M., Ludwig, M., Thiele-Bruhn, S., Ludwig, B., 2014. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. Geoderma, Vol.223-225, pp.88-96.