RETRIEVING TOPSOIL PROPERTIES THROUGH MULTIPLATFORM AND MULTI – HYPER SPECTRAL EO DATA.

Rossi, Francesco (1,2); Casa Raffaele (3); Huang, Wenjiang (4); Laneve, Giovanni (1); Linyi, Liu (4); Pascucci, Simone (2); Pignatti, Stefano (2); Yu, Ren (4)

Organisations: 1: University of Rome Sapienza-SIA, Rome, Italy; 2: Institute of Methodologies for Environmental Analysis, Potenza, Italy; 3: University of Tuscia, Viterbo, Italy; 4: Key laboratory of Digital Earth Sciences Aerospace Information Research Institute Chinese Academy of Sciences, Beijing , China

email address: francesco.rossi@uniroma1.it

ABSTRACT

The assessment of soil variables from multispectral remote imagers is hindered by inadequate spectral resolution, therefore they are mainly used for qualitative assessments. This project represents the initial phase of retrieving topsoil properties with multiplatform and multi-hyper-spectral EO data using machine learning and multivariate regression.

The study areas are in the experimental farm in Quzhou County, Handan City, Hebei Province, China. Measurements of topsoil properties like Soil Organic Matter (%), pH, Effective Phosphorus (mg/Kg), Available Potassium (mg/Kg), and Soil Nitrogen Content (g/Kg), were carried out between 2019 and 2020.

Satellite data from European Space Agency (ESA) Sentinel-2 and the Italian Space Agency (ASI) Hyperspectral Precursor and Application Mission (PRISMA) were used.

To improve the machine learning regression of soil properties the images were co-registered and pre-processed.

The Project is still in its early stages, the dataset is small and bound to a study area. Further research will need to focus on improving the soil dataset and testing the image pre-processing methodology.

INTRODUCTION

The position and intensity of the reflectance peaks and absorptions features related to topsoil properties require high spectral resolution to resolve the spectral features of interest with high accuracy. Therefore, hyperspectral remote sensing data can be applied to acquire, in a cost-effective way, quantitative information about soil. We applied the PRISMA hyperspectral data to retrieve topsoil properties on the two experimental areas.

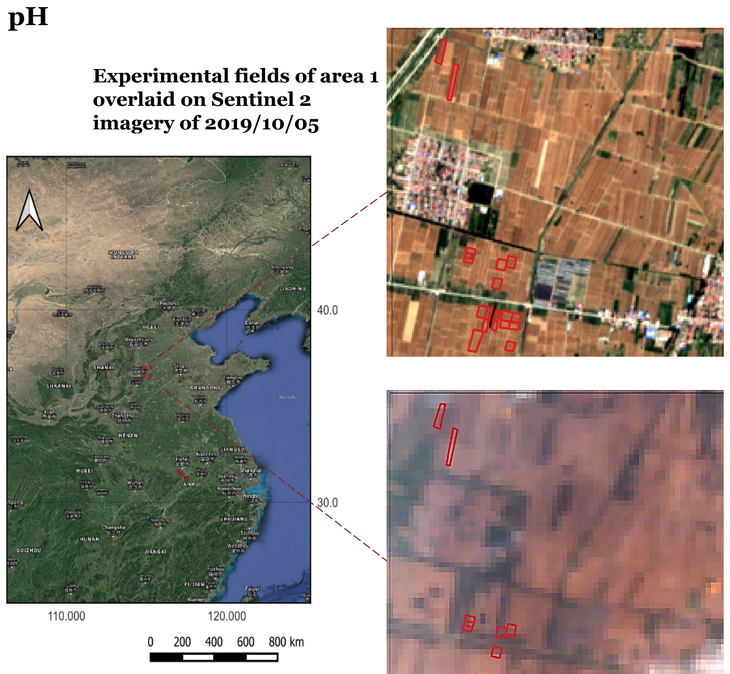
PRISMA (Hyperspectral Precursor of the Application Mission) is a mission fully funded by Italian Space Agency (ASI), launched on 22 March 2019. The instrument is a prism spectrometer, and the design is based on a push broom sensor type observation concept providing hyperspectral imagery (234 bands) at a spatial resolution of 30 m/pixel. The spectral resolution is about 12 nm in a spectral range of 400-2500 nm (VNIR and SWIR regions). Panchromatic imagery is provided along with the Hyperspectral cube, at a spatial resolution of 5 m, and is co-registered with the latter to allow testing of image fusion techniques.

OBJECTIVES

Retrieving topsoil properties with PRISMA hyperspectral data, using machine learning algorithms and multivariate regression:

- **Organic Matter (OM)**
- **Total Nitrogen (N)**
- **Effective Phosphorus (P)**
- **Available Potassium (K)**
- **pH**

Experimental fields of area 1 overlaid on Sentinel 2 imagery of 2019/10/05

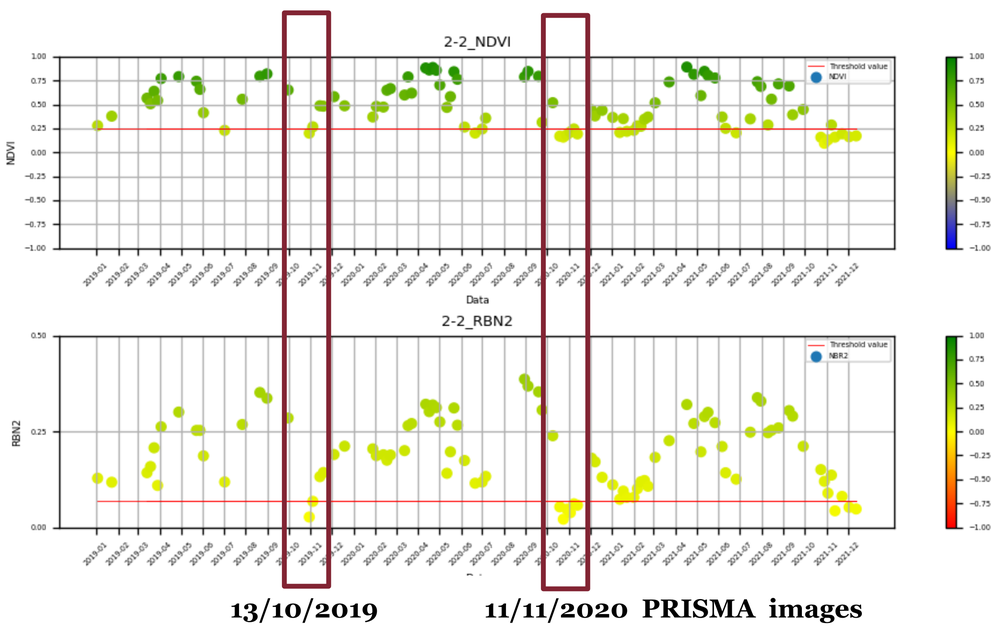


Experimental fields of area 1 overlaid on PRISMA imagery of 2019/10/13

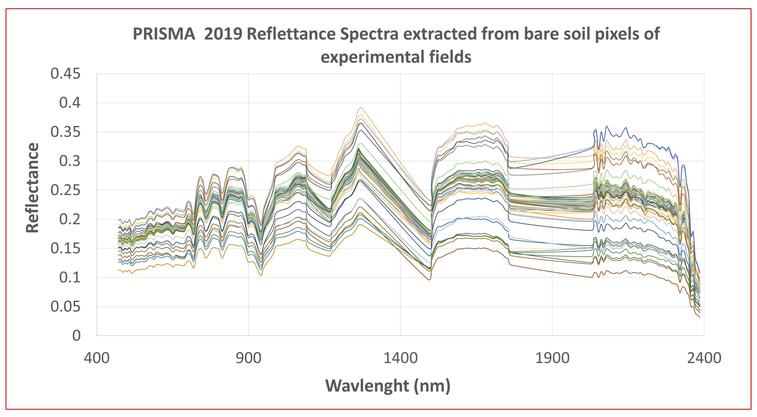
RESEARCH APPROACH

- During 2019 and 2020, measurements of topsoil properties were collected in about 50 fields.
- Topsoil properties were investigated using a five-point sampling method, and each sampling area is a 20m x 20m rectangle.
- Study of the vegetation cover to retrieve the best period of bare soils on the two experimental areas were conducted with Sentinel 2 level 2A time series.
- PRISMA imagery of 13/10/2019 and 11/11/2020 were used, they were co-registered whit Sentinel 2 images, and then preprocessed.
- Machine learning regression (PLSR and CCF) was applied to a spectra library formed by PRISMA reflectance spectra and the parameters measured in situ.

Sentinel 2 NDVI and NBR2 time series analyzed to retrieve the bare soils of the two experimental areas for PRISMA imagery



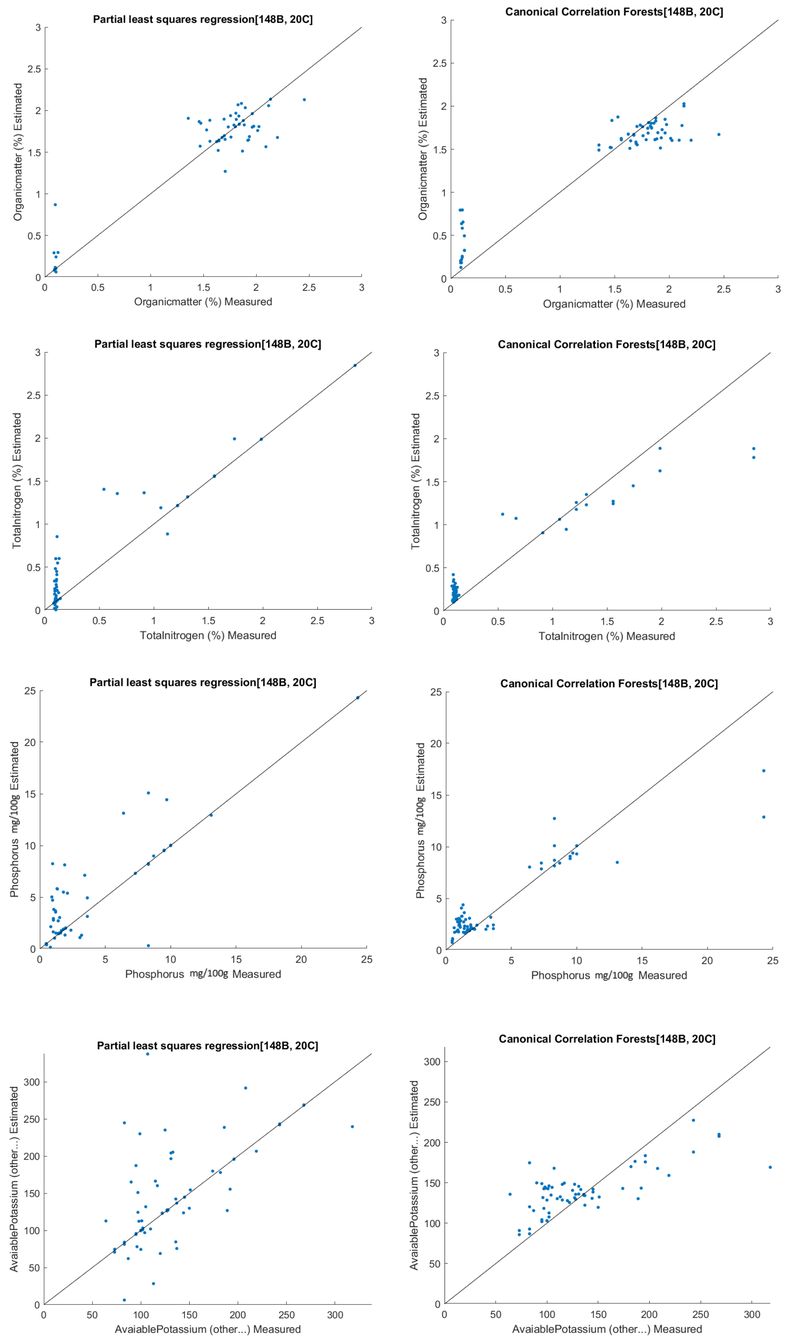
13/10/2019 **11/11/2020 PRISMA images**



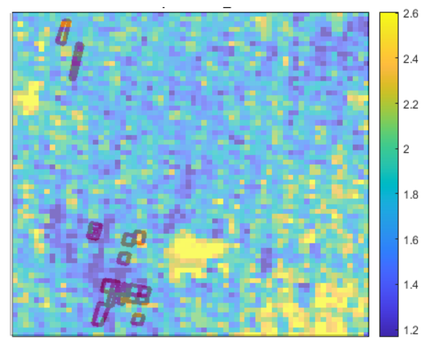
RESULTS

Variable	MLRA	MAE	RMSE	RRMSE	NRMSE	R	R2	NSE
Organic Matter	Partial least squares regression	0.11	0.17	11.97	7.05	0.98	0.95	0.95
	Canonical Correlation Forests	0.19	0.26	18.95	11.17	0.95	0.91	0.88
Total Nitrogen	Canonical Correlation Forests	0.15	0.24	54.05	8.64	0.96	0.92	0.88
	Partial least squares regression	0.17	0.28	62.58	10.01	0.93	0.86	0.83
Phosphorous	Canonical Correlation Forests	1.41	2.50	63.10	10.49	0.91	0.83	0.77
	Partial least squares regression	1.87	2.98	75.36	12.53	0.87	0.76	0.68
Avaiable Potassium	Canonical Correlation Forests	27.50	37.56	27.45	14.79	0.73	0.54	0.49
	Partial least squares regression	34.53	56.72	41.46	22.33	0.61	0.37	-9999.00
pH	Canonical Correlation Forests	0.05	0.07	0.87	16.02	0.68	0.46	0.45
	Partial least squares regression	0.07	0.12	1.56	28.72	0.36	0.13	-9999.00

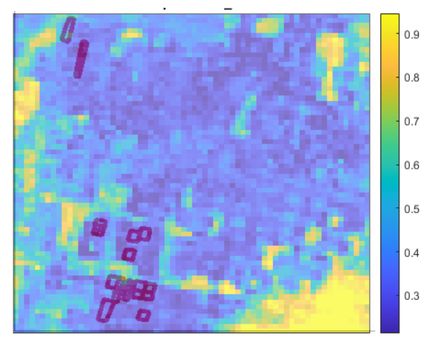
Best correlation results between measured and estimated (from PRISMA) values of soil variables



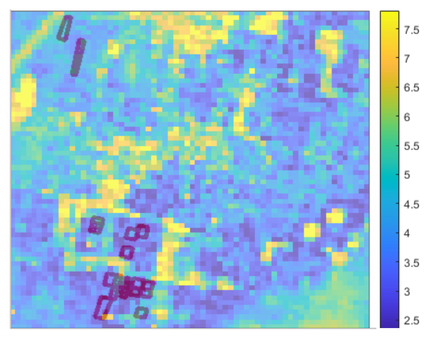
Organic Matter (%) PLSR 2019 map



Total Nitrogen (%) PLSR 2019 map



Effective Phosphorus (mg/kg) PLSR 2019 map



DISCUSSION

Concerning the PRISMA dataset used, as the bare soil on the fields of interest occurs mostly during the month of July, October, and December and many images showed the clouds and other climate phenomena (e.g., haze), only two PRISMA images of 13/10/2019 and 11/11/2020 were utilized for our purposes. On these two dates not all the fields displayed bare soil, so some spectra were not extracted because the presence of green and dry vegetation. Moreover, even if the original PRISMA data contained 234 bands only 140 of them showing low SNR were used for the retrieval of the soil variables. To these bands we applied techniques like Principal Component Analysis to reduce further the dimensionality of the spectra.

Many factors that can influence the soil spectral features like the soil texture and water content were not taken into consideration for the ML retrieval.

The major limits can be resumed in a small dataset and the values of some variables are not uniformly distributed but they can be grouped in clusters.

Not all the estimated soil biophysical variables produced comparable and accurate results. For example, the available Potassium and the pH gave the worst result. Among the tested ML algorithms, except for the Organic Matter, the CCF obtained the best results for the different variables.

CONCLUSION

This mid-term research considers the ability of the hyperspectral sensor PRISMA to predict and map topsoil variables (OM, pH, N, P, K) in two agricultural areas in China. We obtained good prediction accuracies using two hyperspectral PRISMA datasets (R2 = 0.95 for OM is the best result obtained using PLSR).

The Project is still in its early stages and the dataset is restricted to the study area and topsoil properties variability. To extend the capability of retrieving these parameters, further research will need to focus on improving the soil dataset. Moreover, the medium spatial resolution of the PRISMA images (30 m) could be a limiting factor for mapping soil properties in the two investigated area.

The image pre-processing methodology, the co-registration algorithm, and in particular the data fusion algorithm needs to be further tested.

MAJOR REFERENCE

N. Mzid, F. Castaldi, M. Tolomio, S. Pascucci, R. Casa and S. Pignatti, "Evaluation of Agricultural Bare Soil Properties Retrieval from Landsat 8, Sentinel-2 and PRISMA Satellite Data," Remote Sensing MDPI, 2022.

F. Castaldi, A. Palombo, F. Santini, S. Pascucci, S. Pignatti e R. Casa, “Evaluation of the potential of the current and forthcoming multispectral,”. REMOTE SENSING OF ENVIRONMENT, vol. 179, pp. 54-65, 2016