

Predicting soil nutrients with PRISMA hyperspectral data at the field scale: the Handan (south of Hebei Province) test cases

Francesco Rossi, Raffaele Casa, Wenjiang Huang, Giovanni Laneve, Liu Linyi, Saham Mirzaei, Simone Pascucci, Stefano Pignatti & Ren Yu

To cite this article: Francesco Rossi, Raffaele Casa, Wenjiang Huang, Giovanni Laneve, Liu Linyi, Saham Mirzaei, Simone Pascucci, Stefano Pignatti & Ren Yu (28 May 2024): Predicting soil nutrients with PRISMA hyperspectral data at the field scale: the Handan (south of Hebei Province) test cases, *Geo-spatial Information Science*, DOI: [10.1080/10095020.2024.2343021](https://doi.org/10.1080/10095020.2024.2343021)

To link to this article: <https://doi.org/10.1080/10095020.2024.2343021>



© 2024 Wuhan University. Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 28 May 2024.



Submit your article to this journal [↗](#)



Article views: 287











View related articles [↗](#)



View Crossmark data [↗](#)

Predicting soil nutrients with PRISMA hyperspectral data at the field scale: the Handan (south of Hebei Province) test cases

Francesco Rossi ^{a,b}, Raffaele Casa ^c, Wenjiang Huang ^d, Giovanni Laneve ^a, Liu Linyi ^d,
Saham Mirzaei ^b, Simone Pascucci ^b, Stefano Pignatti ^b and Ren Yu^d

^aSchool of Aerospace Engineering, Sapienza University of Rome, Rome, Italy; ^bNational Research Council, Institute of Methodologies for Environmental Analysis, Tito Scalo, Italy; ^cUniversity of Tuscia-DAFNE, Viterbo, Italy; ^dState Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China

ABSTRACT

This research investigates the suitability of PRISMA and Sentinel-2 satellite imagery for retrieving topsoil properties such as Organic Matter (OM), Nitrogen (N), Phosphorus (P), Potassium (K), and pH in croplands using different Machine Learning (ML) algorithms and signal pre-treatments. Ninety-five soil samples were collected in Quzhou County, Northeast China. Satellite images captured soil reflectance data when bare soil was visible. For PRISMA data, a Linear Mixture Model (LMM) was used to separate soil and Photosynthetic Vegetation (PV) end-members, excluding Non-Photosynthetic Vegetation (NPV) using Band Depth values at the 2100 nm absorption peak of cellulose. Sentinel-2 bare soil reflectance spectra were obtained using thresholds based on NDVI and NBR2 indices. Results showed PRISMA data provided slightly better accuracy in retrieving topsoil nutrients than Sentinel-2. While no optimal predictive algorithm was best, absorbance data proved more effective than reflectance. PRISMA results demonstrated potential for predicting soil nutrients in real scenarios.

ARTICLE HISTORY

Received 21 March 2023
Accepted 9 April 2024

KEYWORDS

PRISMA; soil properties; bare soil; available phosphorus; available potassium; total nitrogen

1. Introduction

The knowledge of the agricultural soil characteristics is one of the crucial agronomical types of information that can seriously impact food production and the agricultural ecosystem efficiency (Kopittke et al. 2019; Rojas et al. 2016). Indeed, a good knowledge of the soil attributes could lead to a better knowledge of the soil fertility and to an improvement in agricultural management. Additionally, according to recent reports (FAO and ITPS 2015), soil losses and soil deterioration (i.e. degradation, contamination, and nutrient depletion) determine a loss in soil quality causing yield reduction. According to the Xu et al. (2018) and Lefèvre et al. (2017) finding, which analyzed the carbon storage in China's terrestrial ecosystems, the topsoil (0–20 cm) nutrients and texture explained 70% of the variation in the spatial pattern of the Soil Organic Carbon (SOC) density. There is in fact a positive relationship between the amount of organic carbon in soil and the availability of nutrients (Lefèvre et al. 2017), since soil OM stocks large amounts of plant nutrients, such as nitrogen and phosphorus. Other nutrients, like potassium, are more linked to the mineralogy of the soil, in particular, to the clay amount and clay

mineralogy. Nutrients are added regularly to soils as fertilizers for the purpose of supporting crop growth, to avoid their depletion. The relative ability of soils to store one group of nutrients, the cations, is referred to as cation exchange capacity; when soil has higher levels of organic carbon, it is better able to retain nutrients and has an increased capacity to exchange cations. According to Zhang et al. (2022) climate, soil texture, and land use are factors influencing SOC. Generally, finer textures like clay and silt have a greater microbial activity they hold nutrients and water better, thus providing good conditions for plant growth. Coarse soils are better aerated, and the presence of oxygen results in a more rapid decay of OM.

The assessment and monitoring of the agricultural topsoil characteristics are, therefore, critical issues that should be taken into consideration to assess soil quality indicators and to slow or mitigate soil quality degradation. Soil quality indicators require a comprehensive evaluation, which includes the nutrients present in the soil such as nitrogen, phosphorus, and potassium. These nutrients play a crucial role in agricultural productivity and sustainable management, allowing for the optimal use of fertilizers.

CONTACT Stefano Pignatti  stefano.pignatti@imaa.cnr.it

© 2024 Wuhan University. Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Remote sensing data have been widely used to retrieve SOC or soil OM, mineralogical and textural properties (clay, sand, and silt content), and in more limited cases, also soil nutrients. Among the listed soil parameters, OM/SOC is one of the most relevant, for its clear agronomical importance as impacting directly soil fertility. A complete review of the remote sensing data application for OM/SOC retrieval can be found in Angelopoulou et al. (2019). In recent years, soil spectroscopy has been used as an alternative to soil sampling and wet soil analysis for the assessment of most of the abovementioned soil properties (Viscarra et al. 2022; Liu et al. 2019; Mirzaei et al. 2022; Rasooli et al. 2023; Rossel et al. 2022; Sujatha and Jaidhar 2024). The relevant wavelengths for OM/SOC estimations are in the VIS region around 550–700 nm, in the NIR region around 850 nm, and in the SWIR (Short Wave InfraRed) region from 1700 to 1900 nm and from 2100 to 2400 nm. The SOC in the topsoil, significantly affects the shape and nature of soil reflectance spectra (Gomez, Viscarra Rossel, and McBratney 2008; Liu et al. 2019). In the SWIR spectral region, the two main organic compounds affecting the reflectance are lignin (between 1600 and 1800 nm and at 2100 nm) and cellulose (at 2100 nm) (Ben-Dor, Inbar, and Chen 1997; Gholizadeh et al. 2020, 2021). Furthermore, biochemical components of OM like chlorophyll, cellulose, starch, oil, pectin, lignin, and humic acids (Ward et al. 2020) impact on the Visible and Near-Infrared (VNIR) and SWIR spectral regions.

Recent literature reports a variety of multivariate methods to retrieve soil parameters such as texture, soil moisture content (SMC), and OM from remote sensing optical multispectral and hyperspectral data. Multispectral data set (e.g. S2 and Gaofeng (GF) series) succeeded in the estimation, but only for specific limited test areas showing bare soil (i.e. seed beds conditions), of soil properties of agronomical interest like SOC (Castaldi et al. 2018; Gholizadeh et al. 2018; Guo et al. 2020; Haghi, Pérez-Fernández, and Robertson 2021; Saygin et al. 2023; Wang et al. 2021). Regarding remote sensing technologies to infer soil characteristics, different examples report the use of multispectral optical remote data, while hyperspectral remote sensing is still to be fully explored for operational use within downstream services for agricultural soil mapping (Chabrilat et al. 2019; Stevens et al. 2013). However, the minimization of disturbing factors, even at the subpixel scale, remains the main challenge for the topsoil properties estimation. The search for ideal soil conditions reduces the time window to get a useful image. Nevertheless, the quantitative estimation of soil variables of bare soil retrieved by multispectral imageries shows inherent limitations, since these data are covering, with very broad spectral band-passes, the spectral ranges mostly related to the soil chromophores,

particularly in the SWIR spectral range. The current availability of different ongoing hyperspectral missions (e.g. GF-5, PRISMA, EnMAP, HYSUI, etc.), providing in the SWIR a high SNR and spectral sampling, will doubtless allow a more accurate estimation of most soil variables as compared to current multispectral sensors. In this regard, according to finding of (Castaldi et al. 2016), hyperspectral data provide better results for clay, sand, and, especially OM estimation than multispectral data. Of course, improvements in the number of soil attribute retrieval and their retrieval accuracy will depend on the quality (SNR) and data availability of the new generation hyperspectral sensors (e.g. the NASA SBG, ESA CHIME, and ASI PRISMA-2 G missions (NASA 2023; ESA 2023; ASI n.d.)).

Studies dealing with the monitoring of soil nutrients based on the EO imagery, to the best of our knowledge, are instead only a few. This is probably because N, K, and P do not show distinctive spectral features in the VNIR-SWIR (i.e. from 400 to 2500 nm) spectral region. Soil N estimation from spectroscopy has been more studied than other nutrients. In the study of Song et al. (2018), the authors used different regression models from the Chinese Environmental 1A (HJ1A) satellite data along with 1,297 soil samples in Zengcheng, north of the Pearl River Delta (China), to assess the soil concentration of the total nitrogen (TN), the soil available phosphorus (AP), and the soil available potassium (AK). The study of Song et al. (2018) concluded that the application of a back-propagation neural network model was the most efficient method for mapping and monitoring soil nutrients at a regional scale. In 2019, Yu et al. (2018) retrieved soil nutrients in Shenzha County of the Qiangtang Plateau, north-western Qinghai – Tibet Plateau. They used field and satellite hyperspectral data to demonstrate that the 1720–1738 nm spectral range has significant correlations with the analyzed soil constituencies and that the applied stepwise regression model produced root mean square error (RMSE) values of 68.9%, 46.3%, 31.4%, and 45.5% for SOC, TN, total phosphorus, and total potassium, respectively. More recent studies conducted in China in 2021 and 2022 describe soil spectroscopy applications to derive soil nutrients. In particular, the study of Li et al. (2022), conducted in 2022, used about 490 soil samples (1 km × 1 km sampling strategies) collected over the entire Quzhou County and Hebei Province in China. Soil samples were analyzed using FTIR lab spectroscopy (4000 to 600 cm⁻¹; wavelengths: 2500–17000 nm) and used to describe the topsoil OM and TN content. By combining multivariate methods and variable selection techniques, H. Li et al. (2022) assessed SOC and TN obtaining an R² of 0.72 and 0.74, respectively. While the study of Guo et al. (2021) explored the VNIR-SWIR range on more than

100 samples which were measured in the lab to assess TN, AK, and AP content of the soil samples. Results showed for OM, AK, and AP, gained an R^2 of 0.90, 0.75, and 0.74, respectively. They conclude that the analysis of soil spectroscopy in the laboratory demonstrates promising potential, whereas the application of hyperspectral imagery should exhibit substantial discriminatory capabilities for mapping and monitoring soil nutrients at a regional scale. More recently, Gasmi et al. (2022) used a PRISMA L2D image on a test site in Morocco to assess soil nutrients with 107 sampling points (Gasmi et al. 2022). They used the RF models based on the RF-embedded feature selection technique obtaining good predictive results for OM ($R^2_{val} = 0.69$, and RPIQ = 2.56), soil P_2O_5 ($R^2_{val} = 0.44$, RMSE = 44.10 ppm, and RPIQ = 0.75), and soil K_2O ($R^2_{val} = 0.51$, RMSE = 159.29 ppm, and RPIQ = 1.34), concluding that hyperspectral satellite imagery can be an efficient tool for monitoring soil nutrients, and to fertilization recommendations in the context of precision agriculture.

Regarding the retrieval methods used to infer soil properties from EO data, they range from spectral properties to multivariate statistical methods. As regards soil spectral properties, the soil scattering and absorption effects are used to derive spectral indices, and continuum removal techniques to retrieve peculiar physical or chemical properties. Similarly, the more recent ML and transfer knowledge methods that take advantage of the available global or regional soil spectral libraries are to be still fully explored for their generalization and applicability in wide regions (Bao et al. 2020, Li, et al. 2024; Karray et al. 2023; Kokkas et al. 2023). Among the above-mentioned approaches, statistical multivariate approaches (e.g. PCA and PLSR) as well as the Machine Learning Regression Algorithms (MLRA) techniques are often explored to infer soil optical properties (Castaldi et al. 2016; Chabrilat et al. 2019; Viscarra et al. 2022). However, their applicability has been still limited to relatively local areas for which results are depending on illumination and viewing geometries, atmospheric conditions, and sensor spectral and spatial resolutions. Notably, only some recent studies, such as that of Wu et al. (2023) and Ou et al. (2023), have adopted a semi-empirical approach by combining a radiative model with empirical data.

In this study, field-measured soil samples collected in 96 fields between 2019 and 2020 in Quzhou County and Hebei Province in China, along with PRISMA hyperspectral images acquired between 2019 and 2022 on these areas were used to retrieve OM, N, K and P topsoil properties. Multispectral S2 time series, given their relatively high revisit frequency (five days), were applied to identify the optimal time window for hyperspectral PRISMA acquisitions depicting bare soil conditions. Pre-treatment techniques were applied to:

(a) retrieve soil pixels unaffected by vegetation at sub-pixel scale and (b) normalize spectra that have been used before applying statistical ML regression techniques in order to retrieve soil nutrients at the field scale.

2. Material and methods

The flowchart of the methodology is illustrated in Figure 1 and its steps are briefly described afterward. These include: (a) soil sample collection and analysis (Section 2.2); (b) pre-processing of the hyperspectral remote sensing satellite data (Section 2.6) and (c) the prediction of soil properties based on bare soil spectra from collected satellite data (Section 2.7).

2.1. Study site

The study site is located in Quzhou County, Handan City, in the southern part of Hebei Province, China (Figure 2). Two agricultural test sites were selected for the estimation of topsoil properties one in the North-East corner (Lat. 36°50'59.01"N; Lon. 115°1'57.76" E) and one in the South-West corner of the area (Lat. 36°42'40.90"N; Lon. 114°55'38.73"E). The region is characterized by a warm temperate semi-humid continental monsoon climate. The average annual air temperature is around 13.1°C. July is the hottest month with an average temperature of 26.8°C and an extreme maximum temperature of 41.1°C, while January is the coldest month with an average temperature of -2.9°C and an extreme minimum temperature of -19.9°C. The average annual precipitation is 556.2 mm, with precipitation mainly concentrated from July to September, accounting for one-third of the annual rainfall. The rain and heat in the same period are very favorable to agricultural production, and the land production potential is high. The surface soil is primarily sandy loam, clay, and light loam, whereas the deeper layers are primarily sand and clay.

2.2. Field data collection/field soil samples

Between 2019 and 2020, a total of 95 topsoil samples analyzes of OM, TN, AP, AK, and pH were done on 50 fields of the study area (Figure 3). The topsoil properties were investigated using a five-point sampling method, and the geographic coordinates of each sample were determined using a GPS device. For each sampling area, first, all plants were removed, stones, etc. from the surface of the soil and dug out small pits using tools. Then an appropriate amount of topsoil was taken along the cut surface from the bottom upwards. Soil was sampled to a depth of approximately 20 cm.

The soil obtained from the five points was mixed, placed in plastic bags, and sent to the

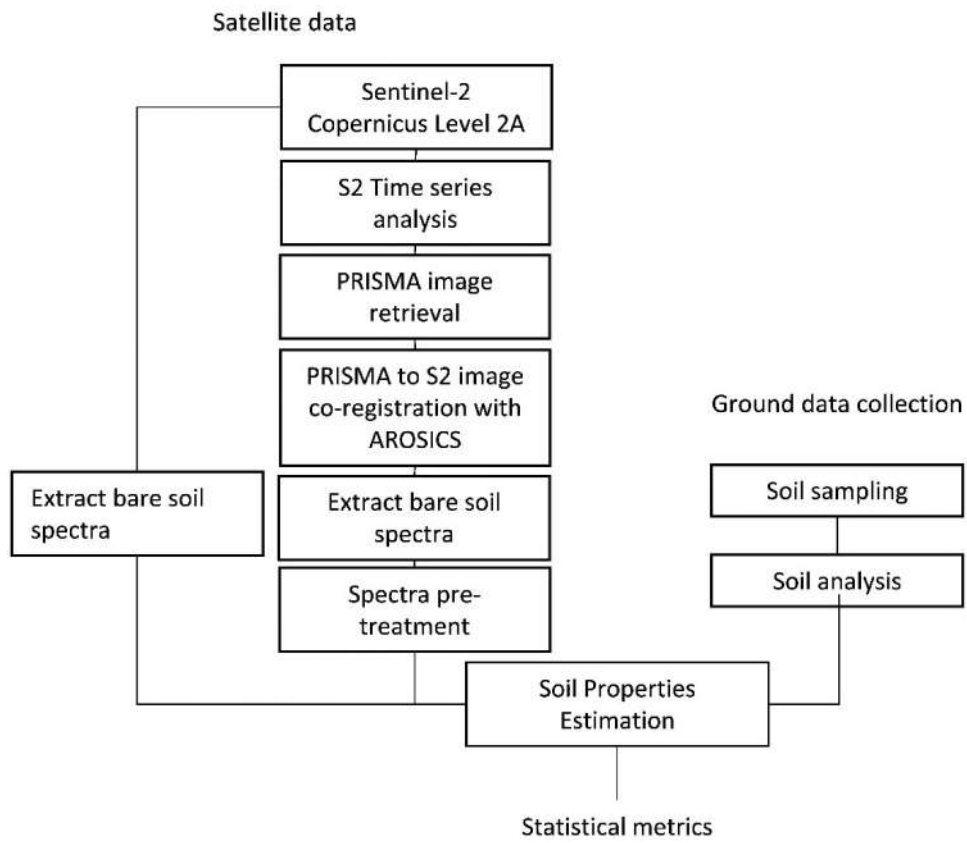


Figure 1. Flowchart of the proposed methodology.

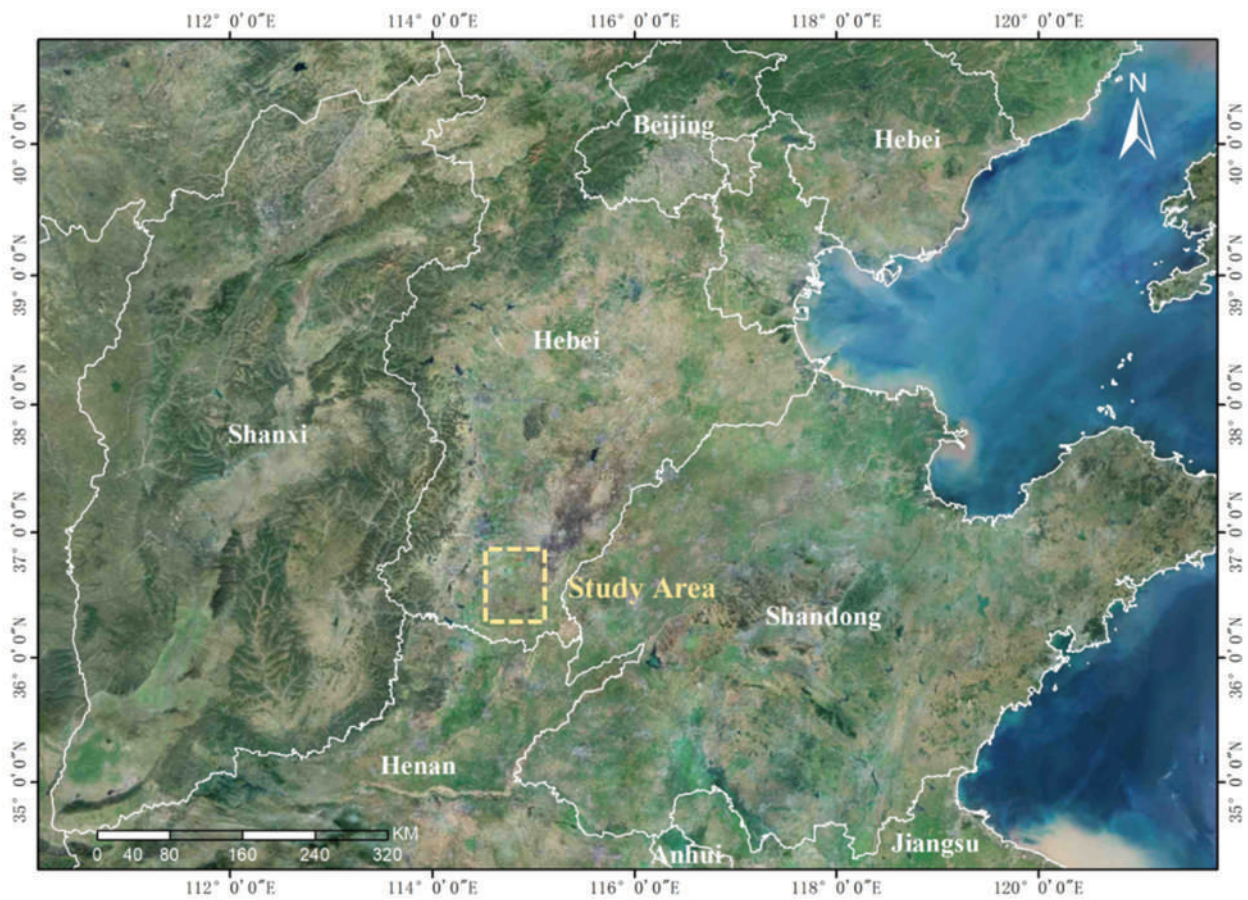


Figure 2. Location of the study area.

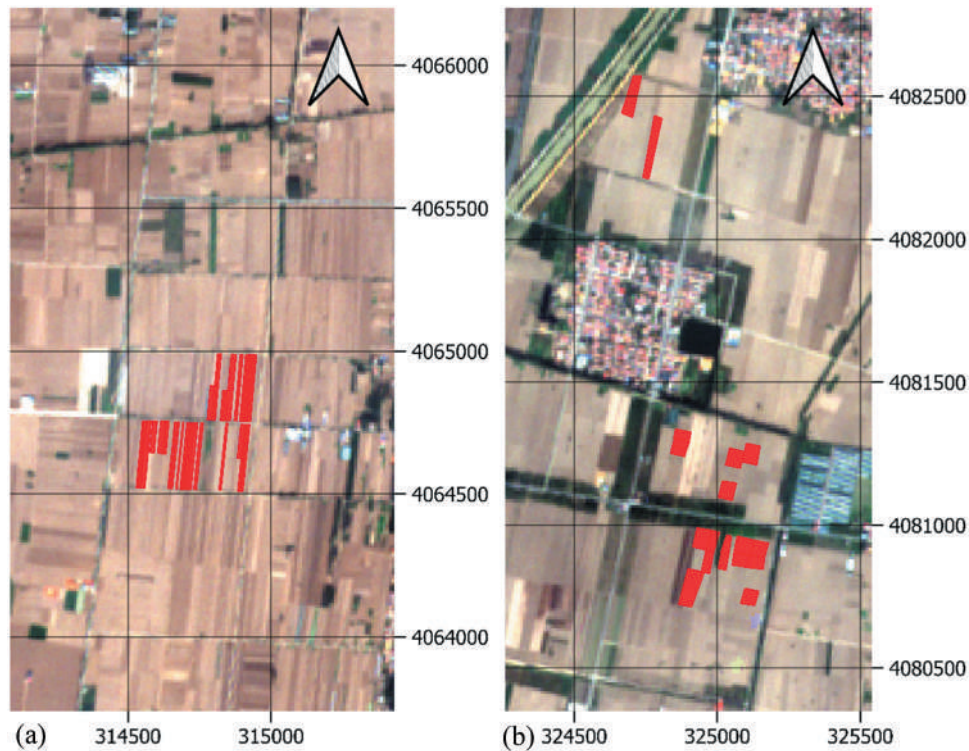


Figure 3. Red polygons of the 50 study fields. (a) Fields in the south-west corner of the study area and (b) fields in the north-east corner. Data source Sentinel-2 19-OCT-2020 RGB image. Reference system WGS84/UTM zone 50N.

laboratory for testing. The soil samples collected underwent a natural air-drying process followed by laboratory sieving to effectively remove any impurities. The determination of OM content in the soil samples involved the utilization of the potassium dichromate oxidation method with external heating. The quantification of TN in the soil was conducted using the semi-micro Kjeldahl method (Bremner 1960). The measurement of AP and AK relied on the application of the molybdenum antimony anti-colorimetric method and flame photometry method, respectively. Furthermore, the assessment of soil pH was accomplished through the utilization of the potentiometric method. The laboratory analysis statistics are reported in the following table (Table 1) (Ren et al. 2023)

2.4. Remote imagery

“PRecursore IperSpettrale della Missione Applicativa” (PRISMA) is a satellite mission put in orbit in 2019 by

the Italian Space Agency (ASI), which acquires hyper-spectral and panchromatic images of the Earth as a novel contribution to remote sensing applications such as resource management and environmental monitoring (Loizzo et al. 2018). The satellite was built for the ASI by OHB Italia Spa as the main contractor. Leonardo Space & Airborne Systems was responsible for the payload instruments which include a VNIR-SWIR imaging spectrometer and panchromatic camera PAN (Cogliati et al. 2021). The main Payload key technical features are described in the (Loizzo et al. 2018; Pignatti et al. 2022) and here briefly listed in Table 2.

PRISMA is a push broom imaging spectrometer based on prism technology to obtain several spectral bands of the same strip on the ground. The imaging spectrometer covers the nominal 400–2500 nm spectral range with two separated instruments (Table 2): the VNIR spectrometer features 66 spectral bands from about 400 nm to 1000 nm and the SWIR detector provides 173 spectral bands between 943 and 2500

Table 1. Results of the lab analysis on Quzhou County soils, for the year 2019 on the left side, and year 2020 on the right. The topsoil properties investigated are OM, TN, AP, AK and pH.

Year	2019					2020				
Properties	OM (%)	TN (g/kg)	AP (mg/kg)	AK (mg/kg)	pH(-)	OM (%)	TN (g/kg)	AP (mg/kg)	AK (mg/kg)	pH(-)
Min	1.18	0.71	4.87	85.00	7.71	1.12	0.86	5.43	62.00	7.54
Max	2.46	1.44	36.33	318.00	8.17	2.75	1.56	74.20	243.00	8.17
Mean	1.81	1.03	15.68	138.40	7.91	1.86	1.16	18.66	121.53	7.96
Std	0.23	0.14	7.74	46.64	0.10	0.43	0.20	12.12	46.91	0.12

Table 2. PRISMA Payload key technical features.

	VNIR	SWIR	PAN
Spectral range	400-1010nm	920-2500nm	400-750 nm
Spectral resolution	12nm	12nm	
Spectral bands	66	173	1
SNR	>200:1 on 400–1000 nm > 600:1 @ 650 nm	>200:1 on 1000–1750 nm > 400:1 @ 1550 nm > 100:1 on 1950–2350 nm > 200:1 @ 2100 nm	>240:1
MTF	@ Nyquist Frequency > 0.3	@ Nyquist Frequency > 0.3	@ Nyquist Frequency > 0.2
Swath width		30 Km; 2.77°	
Spatial resolution	30 m	30 m	5 m
Telescope aperture		210 mm (diameter)	
Orbital altitude		615 Km	

nm with a nominal spectral sampling interval under 12 nm and bandwidth under 12 nm. PRISMA is flying on a Sun-Synchronous Low Earth Orbit at an altitude of 615 km with an inclination of 97.85° and an orbit period of 98 minutes. The expected operational mission lifetime is five years. The nominal orbit re-visit time is 29 days (Nadir-looking) with a re-look capability for a specific target of seven days with off-nadir viewing. The geographic coverage is between 70°S-70° N latitude (at equinoxes) and 180°W-180°E longitude. The size of a single image is 30 × 30 km with a Ground Sampling Distance (GSD) of 30 m (VNIR-SWIR) and 5 m (PAN).

The PRISMA products are stored in HDF-EOS5 (Hierarchical Data Format – Earth Observing System). The files include all information related to the platform, instrument engineering parameters, and information necessary for the geometric processing. Images are provided by ASI (<https://prisma.asi.it>) at different processing levels: (a) the L1 (TOA radiance units); (b) the L2C (ground reflectance without geometric correction); and (c) L2D (geocoded ground reflectance). Details of the ASI standard processing and HDF-EOS5 information are available in the PRISMA Products Specification Document (ASI 2020) created as a “guide” to understanding the content of the PRISMA products.

In this research, S2 data are being mostly utilized to study the crop vegetation cover to retrieve the best phase of bare soils in the two experimental areas (Figure 4).

Using the Python package Sentinel-Sat, all the available S2 images (with a cloud coverage of less than 30%) from 2019 to 2021 were downloaded from the Copernicus Open Access Hub (available at: <https://scihub.copernicus.eu/dhus/>). Pixels affected by clouds and clouds’ shadows were masked using the Scene Classification Layer (SCL) included in the products. Then for each polygon, the time series mean value of the NDVI (Eq. (1)) and Normalized Burn Ratio (NBR2; Eq. (2)) were computed to detect respectively the green vegetation pixels and the NPV.

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} = \frac{B8 - B4}{B8 + B4} \quad (1)$$

$$NBR2 = \frac{\rho_{SWIR1} - \rho_{SWIR2}}{\rho_{SWIR1} + \rho_{SWIR2}} = \frac{B11 - B12}{B11 + B12} \quad (2)$$

During October and November 2019–2022, both spectral indexes showed the lowest values for the study site. So, from the PRISMA mission website (<http://prisma.asi.it>) L2D (current version, i.e. 2.0.5) images of test area acquired at the date of 13 October 2019, 31 October 2022, and 5 November 2022 were downloaded and applied for this study. Also, the S2 product data in the same period were downloaded. All the applied satellite imagery are listed in Table 3:

Even though L2D PRISMA images supplied are geocoded (in our case with Datum: WGS-84 and Projection: UTM 50 N), there is a residual, non-constant small shift in all the images when compared to S2 images and vectorized shape file of the fields of interest.

Consequently, using the package “An Automated and Robust Open-Source Image Co-Registration Software for Multi-Sensor Satellite Data” AROSICS (Scheffler 2017) implemented in a Python script, PRISMA images were co-registered with the S2 ones. A local co-registration approach was used, and the geometric shift was calculated for each point of a regular grid. The average geometric shift corrected was between 50 and 160 m. The two co-registered VNIR and SWIR hyperspectral datasets were merged into a single image file and resampled in a grid of pixels of 30 m × 30 m aligned with the S2 raster grid with a Nearest Neighbour resampling algorithm. Finally, they were converted to the ENVI image format. The images in the two spectral regions (VNIR and SWIR) present an overlapping of the spectral bands in the 930–998 nm spectral region. We chose to maintain all the bands and remove them at a later stage.

By using the tool “image to image registration” of the ENVI v.5.0. software (Exelis Visual Information Solutions, Boulder, Colorado, USA) and manually selecting tie points between the S2 images and the PRISMA co-registered images, we assessed a RMSE of 0.509 pixel.

2.5. Bare soil satellite data

We created two different bare soil spectral libraries using respectively PRISMA and S2 data. These spectral

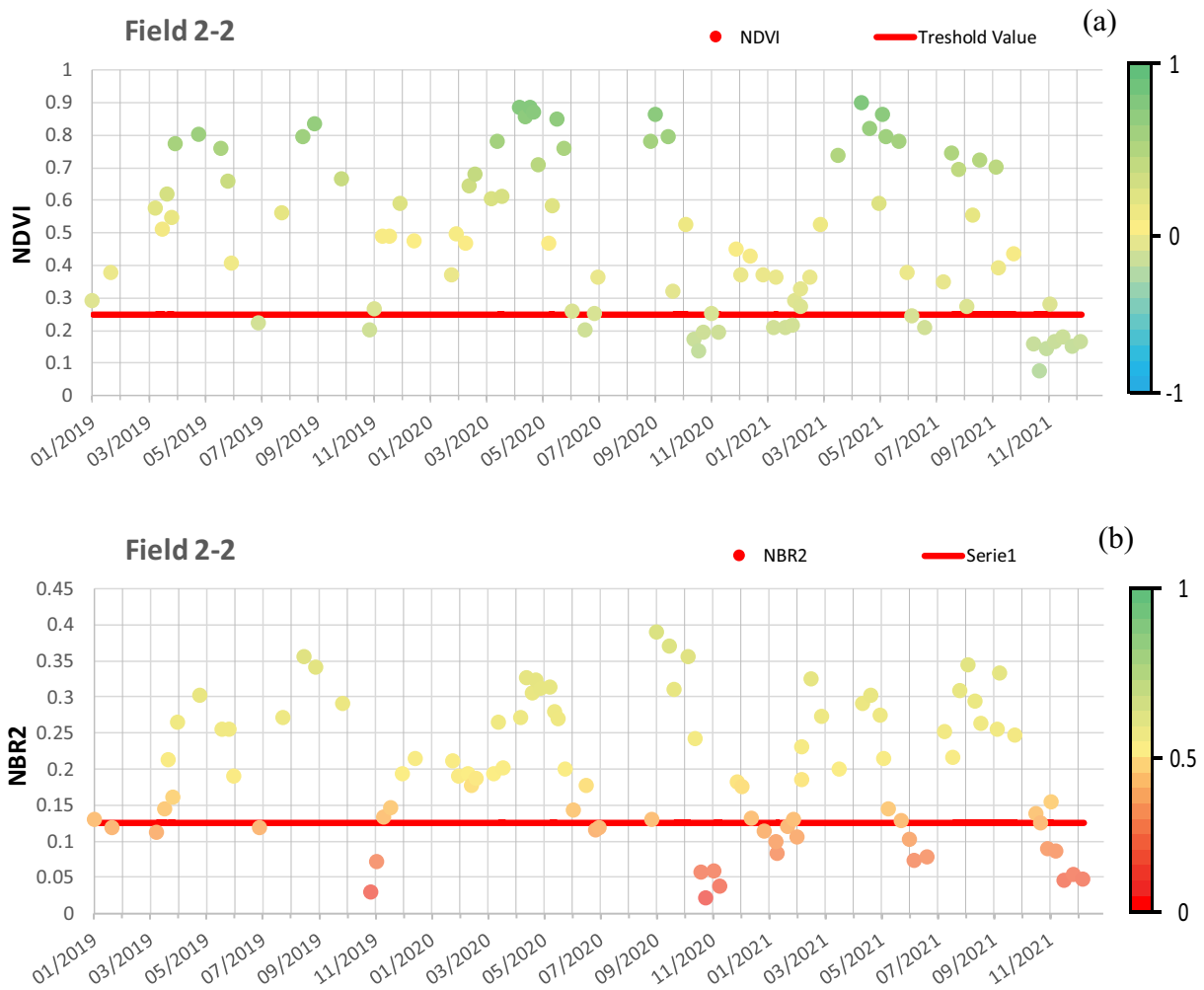


Figure 4. Example of time series analysis, field 2–2 mean value of (a) NDVI and (b) NBR2 time series.

Table 3. Dates of PRISMA and Sentinel-2 used in this research.

PRISMA	Sentinel-2
13 October 2019	5 October 2019
31 October 2022	15 October 2019
5 November 2022	19 October 2020
	30 October 2020

libraries were built in a way that each polygon field “ID” is associated with the values of the soil properties (OM, TN, AP, AK, pH) and the spectral profile. The flowchart of the procedure used to build a spectral library from PRISMA is illustrated in Figure 5. The procedure is described below.

From PRISMA images, the vegetation spectra were derived differently with respect to the S2 ones. We decided to explore the synergic use of linear mixing model (LMM), technique to separate soil and vegetation (PV) and then, taking advantage of the PRISMA spectral resolution, to assess NPV presence by the cellulose band depth at 2100 nm. Considering the high correlation between the spectra of pure bare soil and cellulose, the LMM trained with three endmembers (i.e. PV, bare soil and NPV) could not properly assess the endmembers abundances. The LMM model

assumes that the representation of endmember spectra is a multivariate distribution, e.g. the normal compositional model (Eches et al. 2010; Ren et al. 2023). In particular, Li et al. (2016) confirmed that poor results are obtained by discriminating PV, NPV using an unmixing procedure. Through the use of the “Linear Spectral Unmixing” tool in ENVI, PV was separated from soil spectra (Eq. (3)). Assuming a linear combination of the Endmembers, a spectrum of bare soil and one of green vegetation (PV) were selected from the images. Through the tool, the fractions f_{veg} and f_{soil} constrained to a unit sum were set.

$$R_{TOT} = f_{veg}R_{veg} + f_{soil}R_{soil} \quad (3)$$

where $f_{soil} = 1 - f_{veg}$

Subsequently by applying Eq. 3, an image of spectra purified from the vegetation component was obtained. The simplified unmixing approach based on two endmembers is justified by the fact that we are mainly interested in detecting the presence of vegetation. The resulting image was masked using the raster with the fraction of each Endmember excluding all the pixels that have a fraction of soil (or not green vegetation) less than 85%.

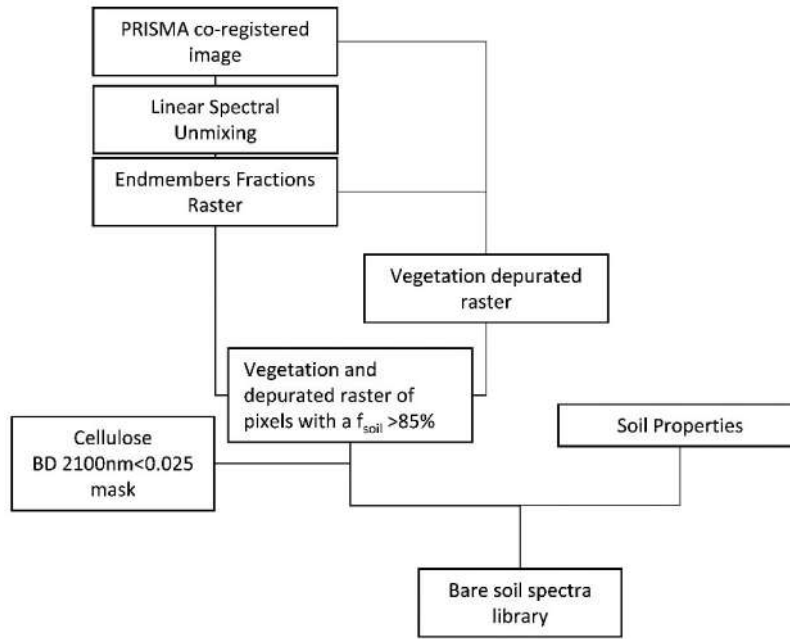


Figure 5. Flowchart of the bare soil spectra extraction procedure.

$$R_{soil} = \frac{R_{TOT} - (1 - f_{soil}) \cdot R_{veg}}{f_{soil}} \quad (4)$$

This threshold value of 85% was selected on the bases of two criteria: (a) to not exclude an excessive number of pixels and (b) to not take pixels with a high unmixing RMSE.

To exclude the so derived bare soil pixels with a significant presence of NPV, we have developed in IDL a code that derives the Band Depth (BD) centered at the absorption peak of cellulose at 2100 nm by using the 2016 nm and 2214 nm as shoulders of the cellulose absorption peak. A preliminary validation procedure of the retrieved BD values (data not shown; PRISMA product from ASI project TEHRA no. F83C22000160005) indicated that a R^2 of about 0.66 could be obtained between cellulose BD at 2100 nm and the number of residues spread on 60×60 m areas obtained by classifying ground nadir true color photos. For this study the threshold of 0.025, corresponding to pixels having more than 35% of cellulose, has been used to exclude the pixels dominated by NPV. To account for pixels that are not completely within a polygon, a script in R language was used. So, it was possible, through the “*exactextractr*” package in R code (Baston 2022), to extract a mean of the pixels that intersects the polygon, weighted by the polygon coverage fraction of every single pixel.

In the case instead of S2, bare soil pixels were selected as those for which the NDVI is less than 0.15 and the NBR2 is less than 0.08. Then, with the same script in R it was possible to extract a weighted mean of the pixels for each field.

2.6. Spectral pre-treatment

Spectral pre-treatment can have an impact on the result of ML regression, so several spectral pre-processing methods have been introduced. For example, Stevens et al. (2010) used first and second derivatives, Savitzky-Golay (SG) smoothing and derivatives to measure SOC in cropland. Rinnan et al. (2009), reviewed the widespread pre-processing techniques for near-infrared spectra in chemometrics.

As pre-treatment the following methods were tested: (a) the SG smoothing, (b) 1st derivative, (c) Absorbance ($\log(1/R)$), (d) Median Filter (MC), and (e) Standard Normal Variate (SNV). They were tested on the PRISMA dataset following the order illustrated in the flowchart of Figure 6. In the case of the S2 dataset, considering that it has only 12 bands, no pre-treatment was applied.

Details of the input parameters of the spectral pre-treatment methods are reported in Table 4.

2.7. Machine learning retrieval algorithms

The MLR was done using a plugin of the MATLAB software package called Automated Radiative Transfer Models Operator (ARTMO) (Caicedo et al. 2014). ARTMO simplifies the retrieval of biophysical parameters from remote observations with a MATLAB graphical user interface (GUI) environment. In particular, the MLRAs toolbox enables analyzing the predictive power of various algorithms in a systematic approach. It contains both linear and nonlinear regression algorithms. The ones we used in this research are listed in Table 5.

The Toolbox also has different dimensionality reduction techniques, we chose to use the Principal

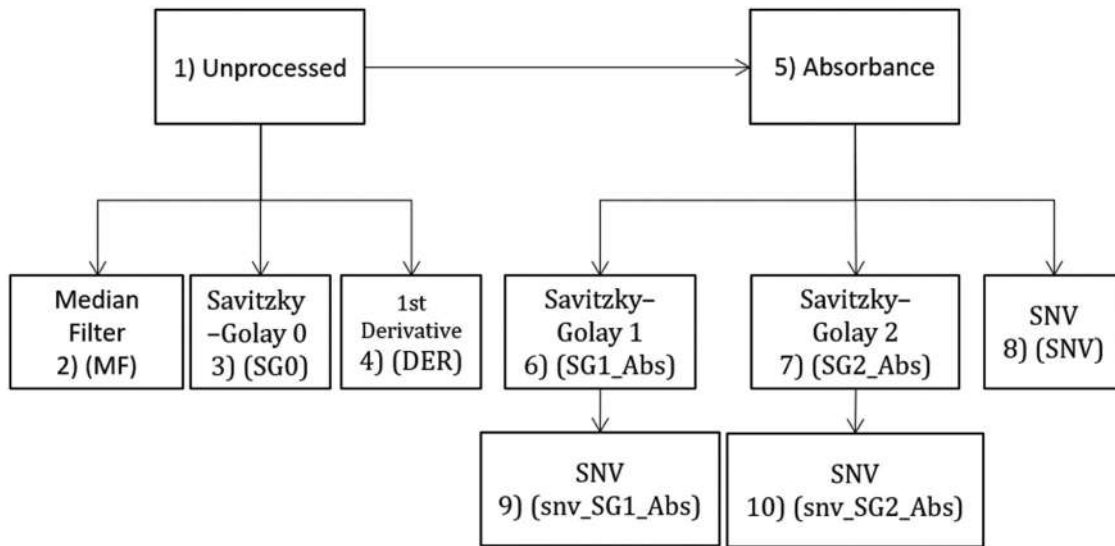


Figure 6. Flowchart of the spectra pre-treatment methods: 1) Unprocessed no pre-processing of spectral reflectance values R. 2) Median Filter (MF). 3) Savitzky–Golay filter 0st derivative applied to the reflectance values (SG0). 4) 1st derivative of the reflectance values. 5) Absorbance obtained from del log of the reflectance. 6) Savitzky–Golay filter applied to the absorbance values (SG1_Abs). 7) Savitzky–Golay filter 2nd derivative applied to the absorbance values (SG2_Abs). 8) Standard Normal Variate (SNV). 9) SNV applied to Savitzky–Golay filter 1st derivative of the absorbance (snv_SG1_Abs). 10) SNV applied to Savitzky–Golay filter 2nd derivative of the absorbance (snv_SG2_Abs).

Table 4. Spectral pre-treatments used in this research.

Name	Description
Unprocessed	No pre-processing of spectral reflectance values R
Absorbance	$\log\left(\frac{1}{R}\right)$
Median Filter (MF)	Window length=3
Savitzky–Golay 0 (SG0)	Window length=6; polynomial order=3; derivative order=0
Savitzky–Golay 1 (SG1)	Window length=6; polynomial order=3; derivative order=1
Savitzky–Golay 2 (SG2)	Window length=6; polynomial order=3; derivative order=2
1st Derivative	The first derivative of reflectance
Standard Normal Variate (SNV)	

At the end, of the original 239 bands we removed: (a) the bands in the atmospheric water absorption regions (corresponding to the 1328–1523 nm and 1812–2044 nm spectral ranges); and (b) the bands in the overlapping zone (i.e. 930–998 nm). The resulting dataset contained 204 of the original 239 reflectance bands. An example of the different pre-processing treatments applied to Soil Reflectance Spectra is shown in Figure 7.

Table 5. Different MLRAs applied in this research are divided into groups.

Group	Machine Learning Retrieval Algorithm
Linear Models	Least Squares Linear Regression Partial Least Squares Regression Regularized Least-Squares Regression Principal Components Regression Elastic Net Regression
Splines and Polynomials	Adaptive Regression Splines
Tree Models	Bagging trees Random Forest
Kernel Methods	Canonical Correlation Forests Support Vector Regression Kernel Ridge Regression
Gaussian Processes	Gaussian Processes Regression

65% training and 35% test set so to have test sets with the same number of elements (16), between the two.

To assess the performance of the predictive models we used the Coefficient of Determination (R^2) (Eq. (5)), the RMSE (6), the Residual Deviation of Prediction (RPD) (7) and the Ratio of Performance to InterQuartile (RPIQ) (Eq. (8)) defined as:

$$R^2 = 1 - \frac{\sum \text{Squared Residual (SSR)}}{\text{Total } \sum \text{ of Squares (SST)}} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} (\hat{y} - y)^2} \quad (6)$$

$$\text{RPD} = \frac{\text{STD}}{\text{RMSE}} \quad (7)$$

Component Analysis (PCA) reducing the dimensionality of the data from 204 bands to 15 principal components in the case of PRISMA. For the validation, the dataset obtained from PRISMA was divided with an 80/20 split: 80% training and 20% test set. Instead, the dataset obtained from S2 was divided into a 65/35 split

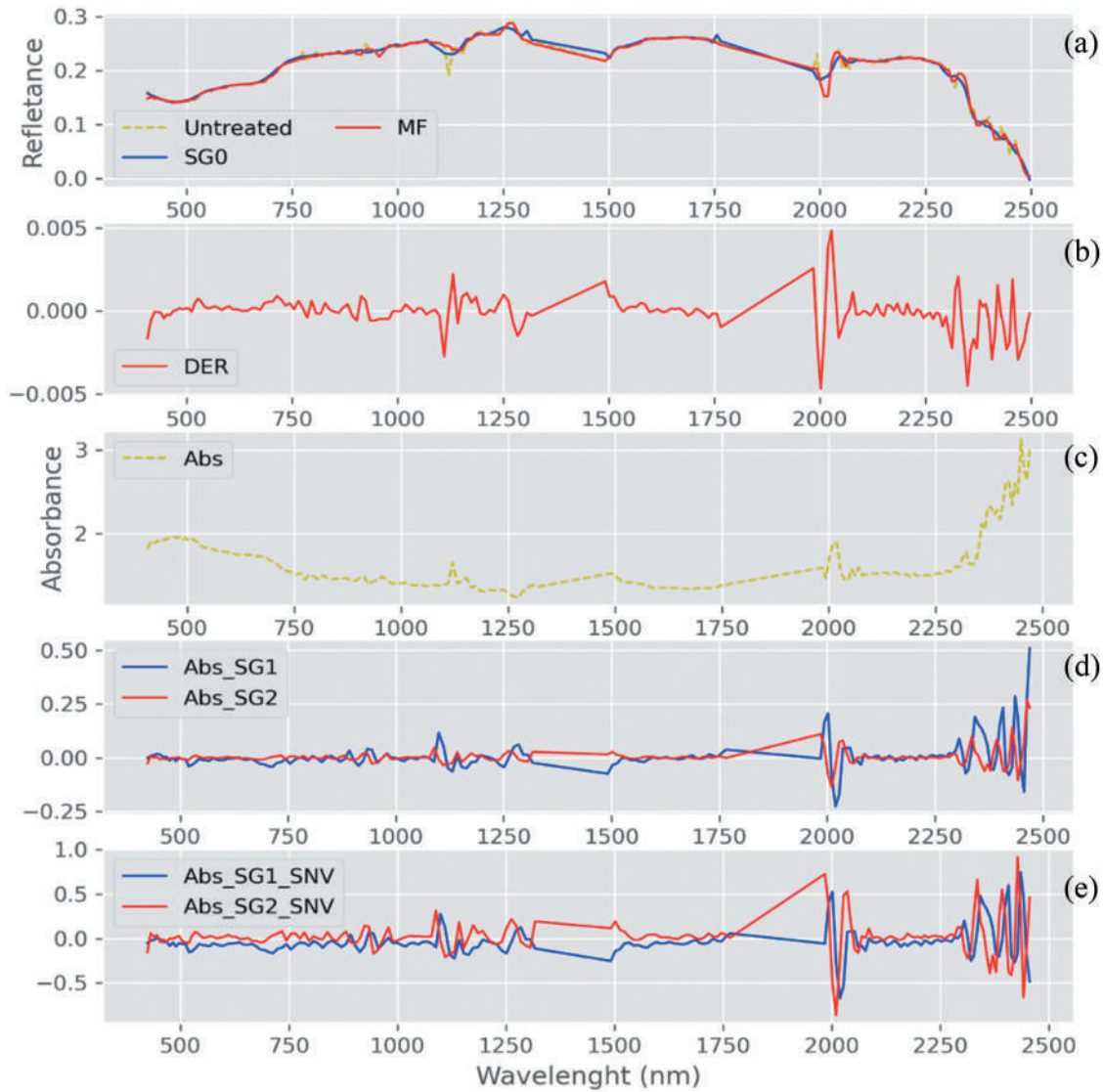


Figure 7. Pre-processing techniques used in this research applied to the same spectra. In the first graph (a), raw/unprocessed spectra (green dashed line) are plotted along the same spectra pre-processed with a Savitzky-Golay filter (blue line) and a Median filter (red line). In the second (b) and third (c) respectively the first-order derivate and Absorbance of the same unprocessed spectra. In the fourth graph (d) Savitzky-Golay's first derivative (blue line) and second derivative (red line) of the absorbance spectra previously mentioned. In the last plot e), Standard Normal Variate transformation is applied to the spectra of plot (c).

$$RPIQ = \frac{IQ}{RMSE} \quad (8)$$

- where y_i and \hat{y}_i refer to the measured value and the corresponding estimated value, respectively; \bar{y} is the average of the measured values.
- n denotes the number of samples.
- STD is the standard deviation of the test values.
- IQ is the interquartile distance of the test values.

R^2 represents the fraction of variance for a dependent variable that is explained by an independent variable or variables in a regression model. The RPD is the ratio of the standard deviation of the measured values to the RMSE. In soil science, the RPD was used as an index of quality (Chang et al. 2001). When $RPD > 2.0$, the estimation model is reliable, values of RPD between 1.4 and 2.0 suggest that the model could be

improved, and $RPD < 1.4$ indicates that the model poorly predict the soil properties (Bellon Maurel and Mcbratney 2011; Chang et al. 2001; Minasny 2013).

3. Results

In the following two sections are presented the main results of the bare soil extraction (i.e. linear spectral unmixing) and the machine learning retrieval algorithms applied to retrieve the soil variables for the Chinese study area.

3.1. Bare soil satellite results

The product of the spectral unmixing process, applied to PRISMA images, is a raster that shows the percentage of each Endmember. Figure 8 indicates that using a linear spectral unmixing approach with two

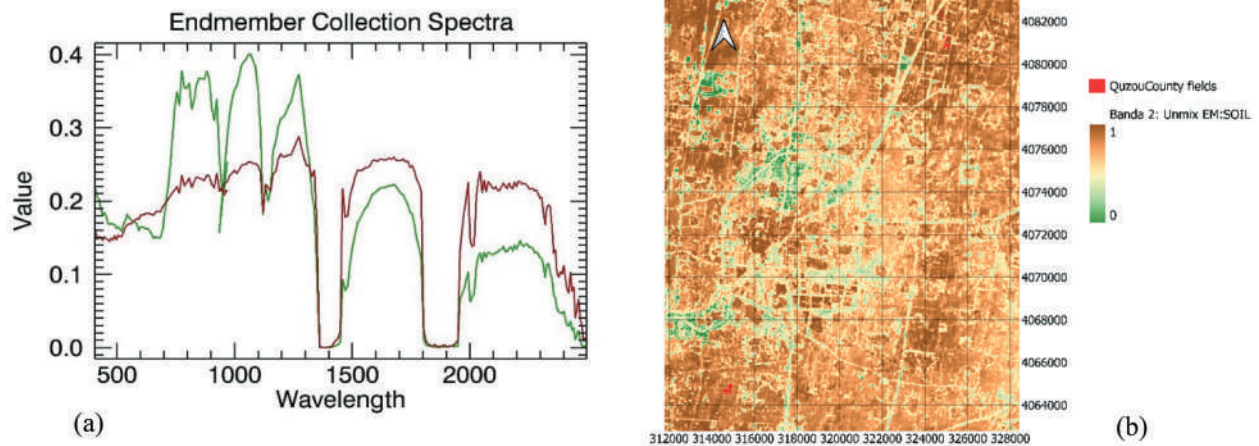


Figure 8. (a) Example of PRISMA soil (red) and vegetation (green) Endmembers Spectra. (b) Green to Red fraction of soil in the study area was obtained from PRISMA 13-October-2019. Reference system WGS84/UTM zone 50N.

endmembers produced satisfactory outcomes while maintaining the sum 1 constraint. The agricultural fields located in the North-East and South-West corners of Figure 8(b) (brown color near one) show the highest values of soil abundance fraction.

To determine the appropriate threshold for soil percentage, we analyzed the histogram in Figure 9, which displays the number of pixels with a given soil percentage. It is evident that increasing the threshold value would result in discarding a large portion of the pixels, while decreasing it would retain pixels with a higher RMSE (as shown in Figure 9).

To discharge the PRISMA pixels with a dominant NPV response, we decided to take advantage of the PRISMA spectral resolution assessing the NPV presence by the cellulose band depth at 2100 nm selecting as threshold a DB of 0.025 corresponding to about a 35% of NPV. Of the 95 ground measurements acquired between 2019 and 2020 only 84 and 47, respectively in the case of PRISMA and S2, remained after this procedure. The descriptive statistics of the two datasets created are shown in Table 6. The different number of samples in the two databases is related to: (a) the different spatial resolution, and (b) the different methods used to select the laboratory dataset by images. Specifically, subpixel with Spectral Unmixing of PRISMA and establishment of thresholds on the multispectral indices of S2.

3.2. Soil properties retrieval using MLRA

For each one of the analyzed biophysical variables, 10 pre-processing methods were utilized and for each of them, 12 different MLRA were tested for a total of 120 possible combinations. The results show a great variability between models and pre-processing.

Regarding OM analysis, the regressions did not provide good results on both datasets. The best predictive model from the PRISMA data estimations was with the

use of the Support Vector Regression (SVR) and the MF spectra that resulted in $R^2 = 0.39$ and $RMSE = 0.24\%$. The second best more accurate predictions show $R^2 = 0.57$ but the RMSE increased to 0.26% thus a lower RPD. These results are slightly better with respect to the ones obtained from the S2 data for which the best performing algorithm was the PLSR that resulted in an $R^2 = 0.26$ and $RMSE = 0.37\%$. From the results, every RPD value is poor, close to 1, meaning that the RMSE values are quite similar to STD of datasets, as displayed in Table 7 and Figure 10. Then, we can observe that the different combinations of preprocessing and algorithms show a reduced capability to predict OM so that further models need to be investigated to retrieve properly the OM content in agricultural topsoil with an accuracy comparable to the Mzid et al. (2022) and Angelopoulou et al. (2023) research that predicted OM in the Mediterranean areas with a higher RPD.

In the case of TN, the best regression results for PRISMA were obtained using the SVR combined with SG2_Abs spectra, which gave an R^2 of 0.58, an RMSE of 0.13 (g/kg), and an RPD of 1.56. The obtained RPD value is above the threshold of 1.4 for which a model is considered acceptable (Bellon Maurel and Mcbratney 2011; Chang et al. 2001; Minasny 2013). With the S2 dataset, we have achieved similar results with the best-performing algorithm, PLSR: $R^2 = 0.47$, $RMSE = 0.16$ (g/kg), and $RPD = 1.31$. In the comparison, however, it is necessary to consider different factors above all the spatial resolution of the sensors, worse in the case of PRISMA (30 m) against the 10/20/60 m of the S2 as well as the fact that the satellite images are acquired in different days of the year with different atmospheric conditions, soil moisture and residual vegetation covers. Table 8 summarizes all the best algorithms for each pre-processing along with graphs in Figure 11 that shows the estimated values against the measured for the two best pairs of pre-processing/algorithm.

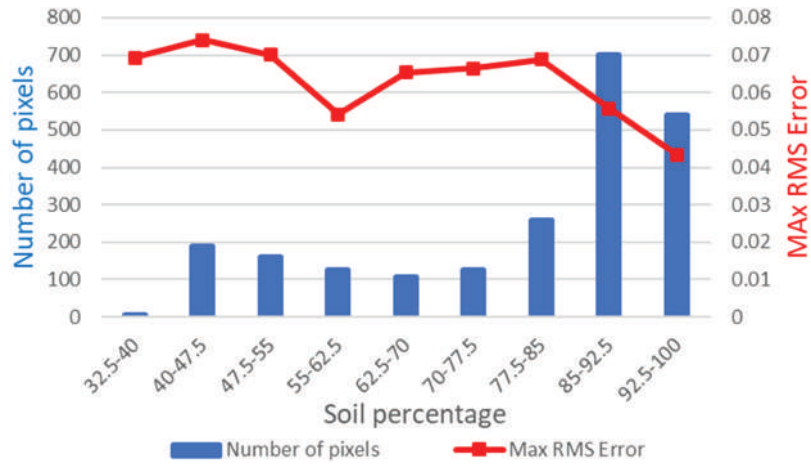


Figure 9. Histogram of the number of pixels inside the study fields that have a certain soil percentage. The line represents the RMSE of the spectral unmixing.

Table 6. Descriptive statistics of the soil properties datasets.

	PRISMA Bare Soil spectral library (n=84)					S2 Bare Soil spectral library (n=47)				
	OM (%)	N (g/kg)	AP (mg/kg)	AK (mg/kg)	pH(-)	OM (%)	N (g/kg)	AP (mg/kg)	AK (mg/kg)	pH(-)
Min	1.121	0.705	4.866	62.000	7.540	1.121	0.782	4.866	62.000	7.540
Max	2.750	1.565	74.201	268.000	8.170	2.721	1.565	74.201	318.000	8.170
Mean	1.850	1.102	17.741	127.012	7.932	1.793	1.104	19.683	140.660	7.932
Std	0.345	0.183	10.426	42.814	0.117	0.376	0.198	11.338	62.569	0.130

*Min: Minimum; Max: Maximum; Mean; Std: standard deviation, S2: Sentinel-2.

Table 7. Best-performing regression results for the different pre-processing techniques for Organic Matter (OM).

	Preprocessing	MLRA	R^2_{cal}	RMSE _{val} (%)	R^2_{val}	RPD	RPIQ
	MF	Support Vector Regression	0.24	0.24	0.39	1.26	1.74
	SG0	Partial least squares regression	0.75	0.40	0.11	0.77	1.06
	DER	Support Vector Regression	0.29	0.27	0.28	1.14	1.58
	Absorbance	Gaussian Processes Regression	0.77	0.41	0.11	0.74	1.03
	SNV	Kernel ridge Regression	0.28	0.26	0.57	1.16	1.61
	SG1_Abs	Support Vector Regression	0.36	0.33	0.07	0.93	1.29
	SG2_Abs	Support Vector Regression	0.29	0.28	0.18	1.10	1.52
	snv_SG1_abs	Random Forest	0.80	0.29	0.13	1.04	1.44
	snv_SG2_abs	Bagging trees	0.81	0.33	0.00	0.93	1.29
S2	S2_Refl	Partial least squares regression	0.52	0.37	0.26	0.94	0.60
	S2_Absor	Canonical Correlation Forests	0.92	0.29	0.25	1.18	0.75

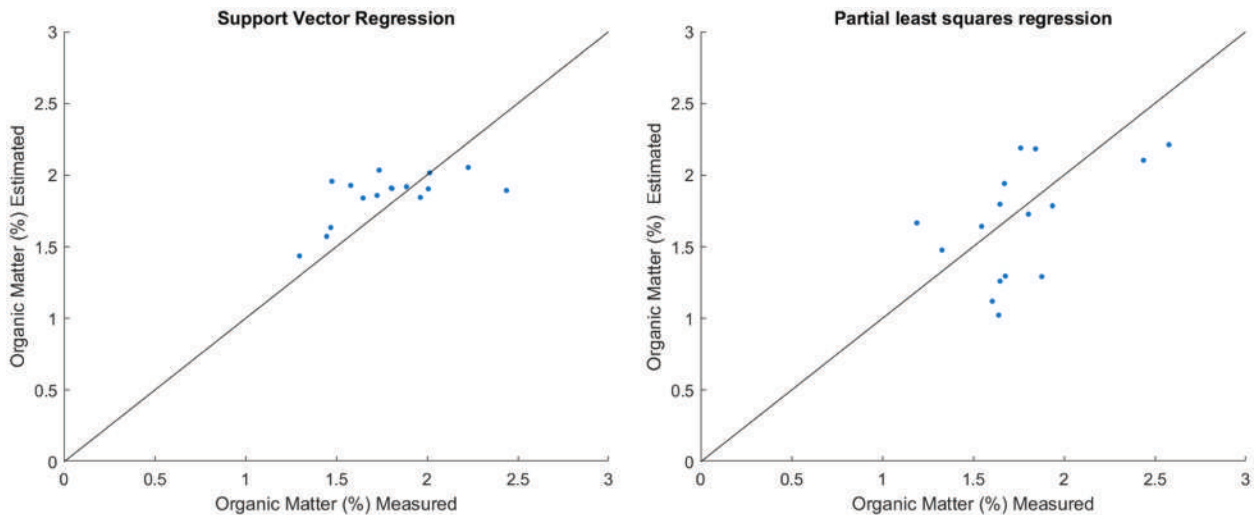
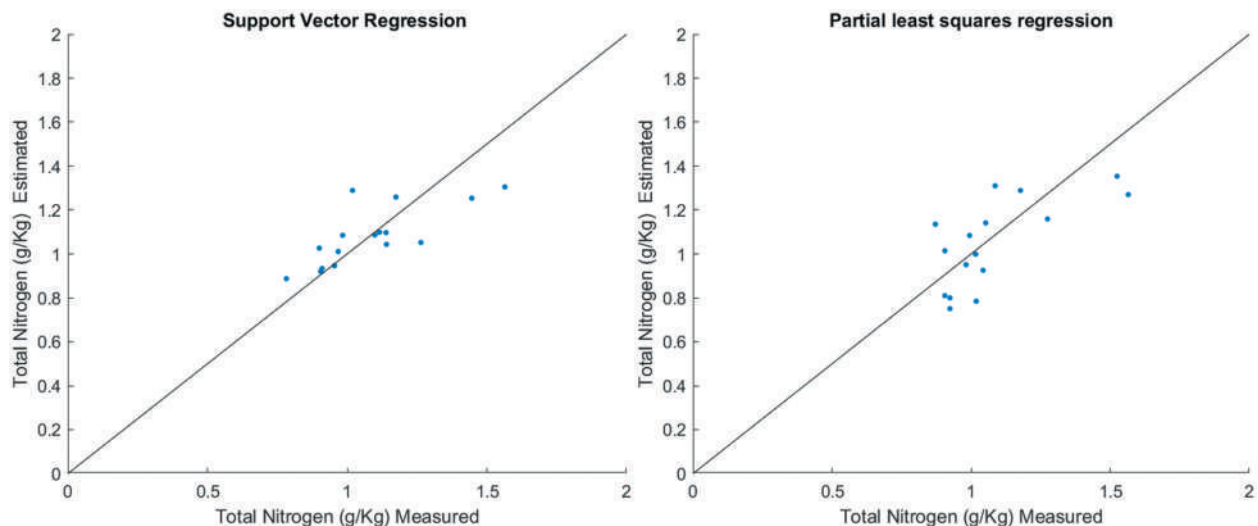


Figure 10. Best performing MLRA for the organic matter (OM). On the left SVR was applied on the PRISMA dataset ($R^2 = 0.39$), and on the right PLSR was applied on the S2 dataset ($R^2 = 0.26$).

Table 8. Best-performing regression results for the different pre-processing techniques for Total Nitrogen (TN).

	Preprocessing	MLRA	R^2_{cal}	RMSE _{val} (%)	R^2_{val}	RPD	RPIQ	
PRISMA	Untreated	Adaptive Regression Splines	0.28	0.17	0.43	1.25	1.26	
	MF	Support Vector Regression	0.30	0.14	0.54	1.50	1.51	
	SG0	Support Vector Regression	0.33	0.18	0.25	1.18	1.19	
	DER	Support Vector Regression	0.34	0.18	0.22	1.14	1.14	
	Absorbance	Adaptive Regression Splines	0.23	0.17	0.34	1.21	1.21	
	SNV_Abs	Support Vector Regression	0.32	0.18	0.20	1.15	1.16	
	SG1_Abs	Support Vector Regression	0.34	0.18	0.26	1.17	1.17	
	SG2_Abs	Support Vector Regression	0.30	0.13	0.58	1.56	1.57	
	snv_SG1_abs	Bagging trees	0.77	0.18	0.23	1.16	1.17	
	snv_SG2_abs	Support Vector Regression	0.38	0.18	0.27	1.17	1.18	
	S2	S2_Refl	Partial least squares regression	0.60	0.16	0.47	1.31	1.15
		S2_Absor	Random Forest (TreeBagger)	0.71	0.16	0.44	1.34	1.18

**Figure 11.** Best performing MLRA for the Total Nitrogen. On the left SVR was applied on the PRISMA dataset ($R^2 = 0.58$), and on the right PLSR was applied on the S2 dataset ($R^2 = 0.47$).

For the AP (Table 9 and Figure 12) the regressions on the PRISMA dataset provided the best results with a combination of RF and snv_SG2_abs spectra thus obtaining an $R^2 = 0.60$, RMSE = 4.70 (mg/kg) and RPD = 1.58. The next more accurate predictions are lower, with an RPD of 1.36 that corresponds to the case of PLSR applied on Absorbance. The S2 dataset performed significantly worse as they obtained using even the best algorithm, the SVR, values of R^2 of 0.3, RMSE = 11.34 (mg/kg), and RPD = 1.03. These results point out that with the S2 dataset, no algorithm was apt to predict the AP with a sufficient accuracy (Figure 12).

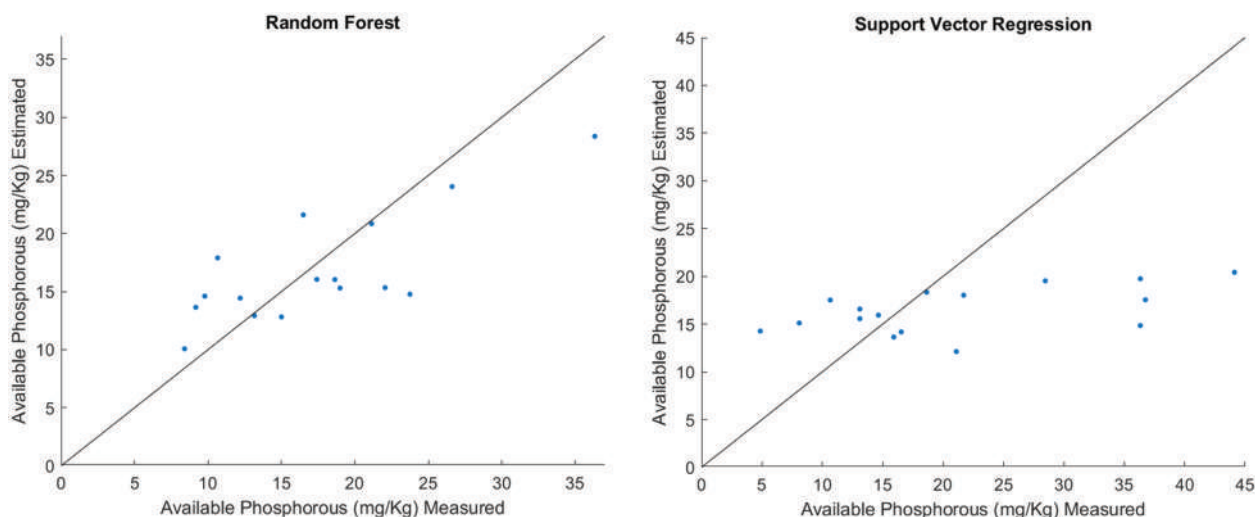
In the case of the AK (Table 10), the regressions on the PRISMA dataset, using RF and SNV_SG1_Abs spectra, provided an R^2 of 0.59, RMSE of 24.59 (mg/Kg) and RPD of 1.37. Similar results were obtained with the S2 dataset as by using the Kernel ridge Regression we obtained values of R^2 of 0.46, RMSE of 46.56 (mg/kg), and RPD of 1.21. The fact that the S2 dataset obtained similar RPD and RPIQ than PRISMA (Table 10) despite having a worse R^2 is related mainly to the test dataset which has a higher STD and IQ, respectively of 56.65 (mg/kg) and 72.25 (mg/kg) instead of 33.70 (mg/kg) and 40.75 (mg/kg).

This is also confirmed by the obtained RMSE values which are almost double that of PRISMA. Plots of Figure 13 show the estimated values against the measured ones for the two best pairs of pre-processing algorithms.

Concerning the pH results, the regressions did not obtain satisfying results on S2 dataset as the KRR showed $R^2 = 0.23$, RMSE = 0.1, and RPD = 1.13. While PRISMA relatively better results were obtained using the Regularized least-squares regression (RLSR) applied to the reflectance spectra with an associated R^2 of 0.25, RMSE of 0.07, and RPD of 1.34 as shown in Table 11. For the analyzed data, it seems that for the pH and OM cases the optimal prediction model shows a relationship between the reflectance spectra and the soil properties, but the relative RMSE is slightly lower than the STD of the test sample, thus determining poor RPD values. Moreover, data set for pH and OM show a reduced range of values (Table 1), that could severely impact on the prediction models efficiency. In Figure 14 are presented the maps for the analyzed soil variables (OM, TN, AP, AK, pH) as obtained by applying the different optimal approaches using PRISMA imagery of 5 November 2022.

Table 9. Available phosphorus best-performing regression results for the different pre-processing techniques.

	Preprocessing	MLRA	R^2_{cal}	RMSE _{val} (%)	R^2_{val}	RPD	RPIQ	
PRISMA	Untreated	Partial least squares regression	0.87	8.24	0.34	0.90	1.29	
	MF	Least squares linear regression	0.18	8.50	0.17	0.87	1.25	
	SG0	Partial least squares regression	0.67	7.43	0.15	1.00	1.43	
	DER	Partial least squares regression	0.84	14.02	0.18	0.53	0.76	
	Absorbance	Partial least squares regression	0.74	5.46	0.53	1.36	1.94	
	SNV	Least squares linear regression	0.19	6.26	0.26	1.19	1.69	
	SG1_Abs	Random Forest	0.71	6.27	0.29	1.19	1.69	
	SG2_Abs	Partial least squares regression	0.82	13.51	0.27	0.55	0.78	
	snv_SG1_abs	Bagging trees	0.72	6.77	0.12	1.10	1.57	
	snv_SG2_abs	Random Forest	0.59	4.70	0.60	1.58	2.26	
	S2	S2_Refl	Support Vector Regression	0.02	11.34	0.30	1.03	1.53
		S2_Absor	Support Vector Regression	0.02	11.32	0.30	1.03	1.53

**Figure 12.** Best performing MLRA for the available phosphorus. On the left Least squares linear regression applied on the PRISMA dataset $R^2 = 0.63$, and on the right Support Vector Regression applied on the S2 dataset $R^2 = 0.3$.**Table 10.** Available K best-performing regression results for the different pre-processing techniques.

	Preprocessing	MLRA	R^2_{cal}	RMSE _{val} (%)	R^2_{val}	RPD	RPIQ
PRISMA	Untreated	Partial least squares regression	0.83	27.97	0.28	1.21	1.46
	MF	Canonical Correlation Forests	0.95	29.79	0.21	1.13	1.37
	SG0	Canonical Correlation Forests	0.96	29.86	0.20	1.13	1.36
	DER	Partial least squares regression	0.87	34.23	0.43	0.98	1.19
	Absorbance	Canonical Correlation Forests	0.96	28.75	0.25	1.17	1.42
	SNV	Adaptive Regression Splines	0.51	55.68	0.20	0.61	0.73
	SG1_Abs	Bagging trees	0.81	29.67	0.32	1.14	1.37
	SG2_Abs	Bagging trees	0.79	28.94	0.38	1.16	1.41
	snv_SG1_abs	Random Forest	0.85	24.59	0.59	1.37	1.66
	snv_SG2_abs	Bagging trees	0.78	27.65	0.48	1.22	1.47
	S2	S2_Refl	Kernel ridge Regression	1.00	46.56	0.46	1.21
S2_Absor		Partial least squares regression	0.59	42.60	0.46	1.33	1.81

3. Discussion

This study shows one of the first attempts to evaluate the potential of satellite hyperspectral remote sensing data for the retrieval of topsoil properties and different nutrients. For this purpose, PRISMA hyperspectral images, acquired in season on China's agricultural bare soils, were used to build topsoil OM, N, P, and K estimation models. These models were compared with laboratory analysis conducted on 95 soil samples, collected between 2019 and 2020, in terms of prediction accuracy.

Topsoil SOC content in agricultural fields can be estimated and, in general a high spectral resolution should support the detection of its specific wavelengths in the VNIR, and the SWIR spectral regions as shown by the recent literature (Gomez, Viscarra Rossel, and McBratney 2008; Liu et al. 2019). The characteristic spectral absorption features for the SOC are located mainly in the VIS spectral region around 550–700 nm in the NIR region around 850 nm, and secondly in the SWIR region from 1700–1900 nm and 2100–2400 nm. Nevertheless, based on the literature (Ben-Dor, Inbar, and Chen 1997), the

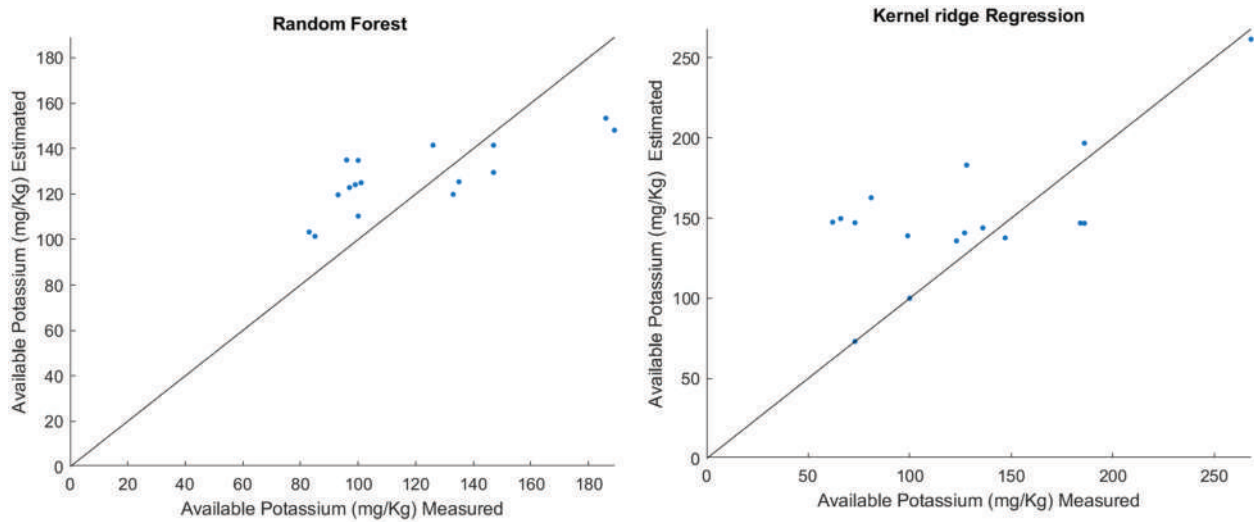


Figure 13. Best performing MLRA for the Available Potassium. On the left Gaussian Processes Regression applied on the PRISMA dataset $R^2 = 0.59$, on the right Kernel ridge Regression applied on the S2 dataset $R^2 = 0.46$.

prediction of soil properties like OM using hyperspectral remote sensing has constraints that are mainly connected to the low SNR and to the coarse spectral and spatial resolutions along with the atmospheric effects, and soil surface effects (i.e. vegetation cover, soil roughness and moisture). Nonetheless, for the selected test area, the spectral resolution of PRISMA imagery, for the study area, did not offer improvements in the SOC estimation with respect to the S2 multispectral sensor. As a matter of fact, both sensors are not able to provide a regression model apt to describe the site variability for SOC. This may be related to three different factors. One of them is the scarce variability of SOC amount in the topsoil of the collected samples (Table 3) that could not be fully described by the regression algorithms. The very limited variability of soil properties, in the agricultural fields employed in this research, is also a severe issue for soil pH estimation, as in that in the area it ranges between 7.5 and 8.1 thus negatively impacting on all of the algorithm's performance used for its prediction. The estimation of OM and TN contents on a laboratory scale using mid-infrared (MIR) reflectance spectroscopy in the same region of Quzhou County was attempted by H. Li et al. (2022). They obtained a higher R^2 with respect to our results by using a sensor with a wider spectral range (2500–25,000 nm) than PRISMA, even though the data set they used showed a comparable range of STD and mean value of OM. In this research, we have a maximum value of OM of about 2.75%, which converted to SOC by the “Van Bemmelen factor” (Minasny et al. 2020) is equivalent to a SOC of 1.58% which corresponds to a SOC value of $15.8 \text{ g}\cdot\text{kg}^{-1}$. This range of value is very limited, compared to other studies, like the one of S. Wang et al. (2022), which used SOC concentrations ranging between 0.5 and $700 \text{ g}\cdot\text{kg}^{-1}$. The second factor to consider is the SMC.

In fact, the images may contain reflectance spectra of soil with different SMC. C. Wang and Pan (2016) found that SMC affects the relationship between reflectance and OM. When using the calibration model derived from air-dried samples to predict the OM of moist samples, a decrease occurred in the accuracy for all SMC groups compared with that of dried samples. The third factor is the time frame constraints to have images with bare soil on the selected fields. Then, only winter images were selected. This constrains determines: (a) the availability of images characterized by low SNR typical for the autumn-winter acquisitions, and (b) the complexity of wet atmospheric profiles lowering the atmospheric transmittance in the VIS spectral region.

The results have shown a good potential of the PRISMA data to map Nitrogen content (Table 5 and Figure 7). Despite laboratory analysis of SOC consistently shows better metrics ($R^2 = 0.80$) (Li et al. 2022) compared to the ones of N, in this research the optimal regression model obtained by PRISMA for N shows slightly higher values (i.e. $R^2 = 0.58$, RMSE = 0.13 g/kg, and an RPD = 1.56) than the SOC retrieval. The most performing algorithm is the SVR applied to the smoothed absorbance (SG second order derivative filtering) that finds a hyperplane that best fits the data points. These values, even though they depict discrete prediction accuracy, are similar to the ones predicted for the S2 bare soil pixels selected using threshold values of $\text{NDVI} < 0.15$ and $\text{NBR2} < 0.08$. Soil N content is often highly correlated with that of SOC (Martin et al. 2002) and this relation is also valid for our dataset (Figure 15).

Absorbance of the N-specific features present in the soil spectra is not as strong as the absorbance of C bond as the mass of SOC in soil is generally one order of magnitude greater than that of N (Martin et al. 2002). It was explained that N is best predicted by its correlation

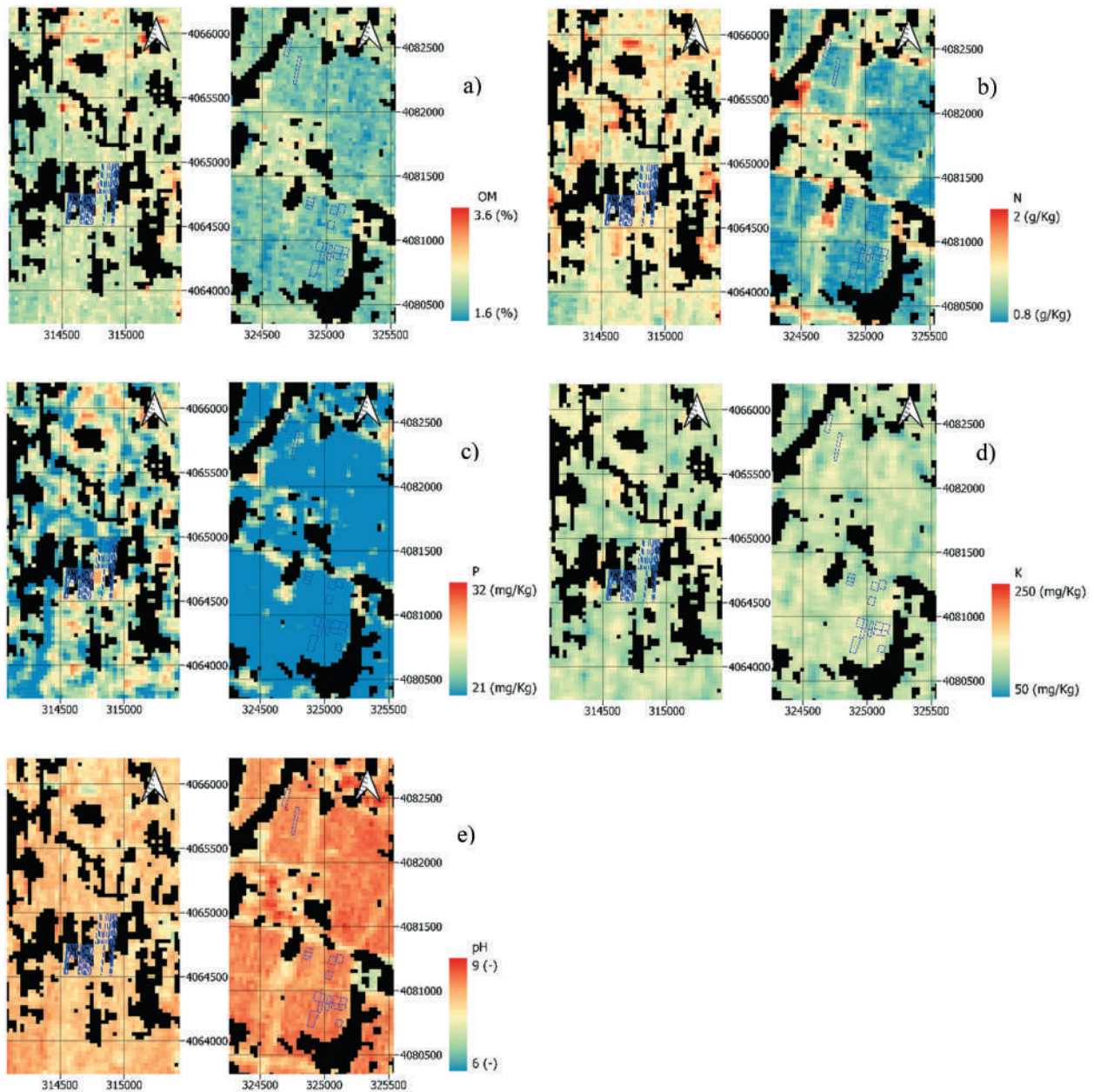


Figure 14. Prediction maps of a) OM, b) TN, c) EP, d) AP and e) pH obtained applying the best performing MLRA and preprocessing on the PRISMA image of 05-November-2022. For each subplot on the left are the fields in the South-West corner of the study area and on the right the fields in the North-East corner. To all the images is applied in black the mask used to exclude the pixels containing cellulose and vegetation, while the blue rectangular show the shape of the sampling fields. Reference system WGS84/UTM zone 50N.

Table 11. pH best-performing regression results for the different pre-processing techniques.

	Preprocessing	MLRA	R^2_{cal}	RMSE _{val} (%)	R^2_{val}	RPD	RPIQ
PRISMA	Untreated	Regularized least-squares regression	0.65	0.07	0.25	1.34	1.82
	MF	Kernel ridge Regression	0.53	0.08	0.32	1.19	1.61
	SG0	Adaptive Regression Splines	0.26	0.07	0.47	1.29	1.75
	DER	Partial least squares regression	0.70	0.10	0.47	1.00	1.37
	Absorbance	Partial least squares regression	0.65	0.08	0.53	1.20	1.63
	SNV	Partial least squares regression	0.65	0.11	0.39	0.90	1.22
	SG1_Abs	Partial least squares regression	0.69	0.09	0.46	1.11	1.51
	SG2_Abs	Gaussian Processes Regression	0.92	0.07	0.44	1.31	1.78
	snv_SG1_abs	Bagging trees	0.81	0.08	0.31	1.24	1.69
	snv_SG2_abs	Gaussian Processes Regression	0.78	0.07	0.38	1.30	1.77
S2	S2_Refl	Kernel ridge Regression	0.55	0.11	0.23	1.13	1.45
	S2_Absor	Kernel ridge Regression	0.54	0.11	0.23	1.12	1.44

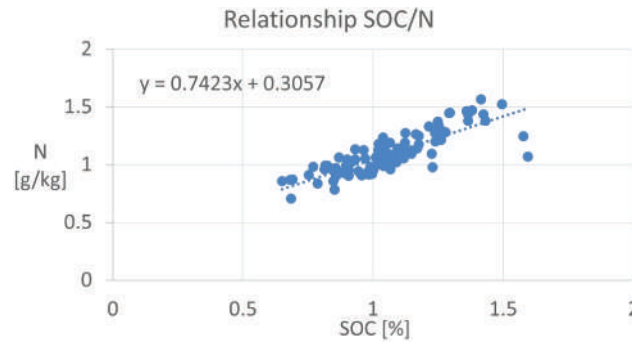


Figure 15. Relationship between SOC and N for the soil lab analysis samples.

with SOC if a large SOC-to-N correlation exists (Martin et al. 2002). These slightly better retrieval values concerning the SOC ones could be due to the fact that N, according to finding of Patel et al. (Patel et al. 2020), influences the SWIR spectral region, in particular for the bands located at about 1899 nm and 2195 nm, which is less sensitive to the atmospheric issue than the VIS spectral range. To explain the different results obtained for SOC and N, although they are strongly correlated, we could probably refer to the vegetation suppression procedures described in section 0, where PRISMA spectra for bare soil were derived by subtracting the subpixel contribution of vegetation. Considering that in the VNIR spectral range vegetation spectra show great variability in terms of reflectance values, its subtraction could have affected the soil spectral features related to SOC that mainly occur in VNIR range (Shi et al. 2020), but not the N that mainly occurs in the SWIR range. Therefore, SOC retrieval could be slightly better than N.

Similarly, to SOC and N, other nutrients play an important role in soil fertility. Among them, this study considers K and P for which the soil spectroscopy literature still does not report extensive studies and reviews. This research is clearly one of the first attempts of applying hyperspectral data, and particularly by using PRISMA, to define a prediction method for soil nutrients. According to finding of Ji et al. (Ji et al. 2014), nutrients like soil K and P don't show distinctive spectral features so their prediction needs to rely on a regression procedure associated with the whole spectrum. Moreover, such soil nutrients usually occur at low concentrations. Even though a not clear understanding of the useful spectral features is present in literature, e.g. Guo et al. (2021), based on laboratory spectral data, reports that for AK both VNIR and SWIR spectral regions are essential. In particular, Guo et al. (2021) recognized ferrihydrite, goethite, amine (N-H), OM, free water (O-H), cellulose, lignin, starch, the first overtone of O-H stretch, Al-OH or Mg-OH as the main spectral features. Attributes that are very similar to those related to K, even if within more complex element types. We have compared the ability of different linear, and nonlinear ML approaches in predicting

soil nutrient contents (P and K). For both nutrients, we have achieved sufficient prediction performances (Tables 7, 9) depicted by $R^2 = 0.60$, RPD = 1.58, and $R^2 = 0.59$, RPD = 1.37, for P and K respectively. For both retrievals, the filtered absorbance data set appears as the most performing. For the P prediction, the combination of SNV applied to the SG second derivative smoothing of the absorbance pre-treatment with the use of RF is the most performing combination, while for the K prediction the SNV applied to the SG first derivative smoothing of the absorbance and random forest, produced the better results. In both retrieval cases (i.e. P and K) the pre-processing is similar, the only difference is represented by the derivative order, while the better MLRA is the RF with an ensemble of decision trees in which each tree is learned from a random subset of samples from the training set. The prediction done by using PRISMA for both nutrients is better than those obtained by S2, both in terms of RMSE and RPD. The predicted accuracy for nutrients retrieval (which statistical values are shown in Table 3) in terms of RMSE (24.59 mg kg^{-1} for P, 4.70 mg kg^{-1} for K) is comparable to the one given by Song et al. (2018), and Yu et al. (2018) on different Chinese test sites. The study of Song et al. (2018), conducted on a wider regional scale (Zengcheng city with a larger sample data set, i.e. a total of 1.297 soil database showing an AP mean of 74.81 mg kg^{-1} with a standard deviation of 53; AK mean of $100.55 \text{ mg kg}^{-1}$ with a standard deviation of 89.17), has predicted, using an RF model, the P soil concentration with an RMSE of 43.21 mg kg^{-1} and K with an RMSE of 72.97 mg kg^{-1} .

From our PRISMA results, we could not identify a single predictive algorithm or a group as the most performing (Table 4) as their performances are not constant among the analyzed soil parameters pertaining to the Quzhou County area. The use of the absorbance data set, instead, resulted in the most performing pre-treatment when applying different filtering procedures for each variable. The average R^2 values for P retrieved on the absorbance spectra pre-treating methods is 1.65 times the reflectance ones and for K is 1.32. However, in

the case of N, there is no such difference among pre-treatments. In general, the linear models as well as the kernel or Gaussian process approaches seem to provide more accurate retrievals of soil nutrients content when it is not possible to identify peculiar spectral features associated to these nutrients.

The soil nutrient maps showed higher concentrations of OM, N, P, K in the South-West sectors of the investigated area, whereas the lower concentrations were exposed in the North-East. The pH case is the opposite as the higher values are present in the North-East zone of the research area. This trend is in accordance with the laboratory data analysis used to build the spectral library for the machine learning regression. The soil nutrient maps also showed less scattering of high values in the North-East corner of the study area, where a reduced number of discarded pixels are present, thus indicating that the South-West fields may contain a more complex spectral signature.

The importance of the SWIR bands in the development of performing prediction models, especially for P and K, confirms the potential of satellite hyperspectral sensors for soil properties/constituencies estimation, as shown in this early study using PRISMA imagery. In particular, the SWIR spectral region appears as an important advance with respect to the multispectral sensors. Moreover, PRISMA quality allows to process at the subpixel level by applying vegetation suppression when the vegetation abundance within the pixel is below 15%.

4. Conclusions

We investigated the suitability of PRISMA and S2 bare soil images for predicting the TN, AP, AK contents of the topsoil in croplands based on different ML regression algorithms and reflectance spectra pre-treatments. PRISMA bare soil spectra were obtained by eliminating the vegetation influence on the pixel spectra response based on an unmixing procedure. Vegetation contribution was subtracted when abundance was below 15% within the pixel, while pixels dominated by NPV spectra response were discharged on the basis of the BD of the cellulose absorption peak at 2100 nm. Moreover, we compared PRISMA hyperspectral data results with those obtained using S2 multispectral images for a test area in China. The results show a better prediction accuracy for PRISMA than S2 data.

Moreover, in this study, good accuracies using PRISMA datasets (RPD values of 1.58, 1.37, and 1.56 for P, K, and TN, respectively) were obtained. As expected, with respect to the laboratory cases, a slight worsening in terms of accuracy was observed using the spectra extracted from the real hyperspectral and multispectral spaceborne sensors. PRISMA

imagery resolution (30 m) is, therefore, not a limiting factor for soil properties mapping especially when a vegetation correction is applied instead of the classical pixel filtering procedure based on the threshold applied to NDVI and NBR2 spectral indices.

Our results confirm PRISMA satellite hyperspectral data potential to predict topsoil nutrients in a real (i.e. not using laboratory soil samples) test case obtaining even a higher accuracy with respect to the S2 multispectral sensor performance, which has a limited spectral sampling in the SWIR region. A more accurate unmixing approach should be considered, besides vegetation and bare soil. Maybe it will be good to include also dry vegetation and shading components.

The next full availability of a hyperspectral constellation (e.g. available PRISMA and EnMAP) and the forthcoming satellite missions (e.g. NASA's SBG, ASI-PRISMA-2, ESA-CHIME), hopefully assuring a higher SNR in the SWIR spectral region and a higher acquisition frequency better covering the short time frame in which bare soil is exposed, will certainly enhance the production of soil fertility maps and the monitoring of the soil nutrients and SOC as compared to existing available possibilities.

Acknowledgments

Italian Space Agency (ASI), project TEHRA "Topsoil properties Estimation from Hyperspectral Remote sensing for Agriculture" [Contract/Agreement No. 2022-6-U.0, CUP 451 no. F83C22000160005]. Italian Space Agency, project PRISCAV "PRISMA CALibration/Validation" [grant number 2019-5-HH.0].

Disclosure statement

This is to acknowledge any financial interest or benefit that has arisen from the direct applications of your research. Further guidance on what is a conflict of interest and how to disclose it.

Funding

This study was co-funded by the European Space Agency Dragon5 project [ESA Contract number 4000135216/21/I-NB]. This study was co-funded by the Italian Space Agency THERA project [grant number DC-UOT-2019-061].

Notes on contributors

Francesco Rossi is a Ph.D. student at the Sapienza University of Sapienza of Rome, Italy. His research interest is hyperspectral remote sensing, to monitor soils and crops.

Raffaele Casa received the Ph.D. degree from the University of Dundee, Scotland, U.K., in 2003. He is an Associate

Professor in agronomy and experimental methods in agriculture with the University of Tuscia, Viterbo, Italy. Since more than 15 years his research has focused on the application of remote sensing and precision agriculture for the improvement of the agronomic management of crops. He has been a member of the Mission Advisory Board of the Sentinel-2 Mission of the European Space Agency and is now a member of the Earth Observation Scientific Advisory Committee of the Italian Space Agency. As a member of the Group of Experts on precision agriculture with the Italian Ministry of Agriculture (2015–2016), he has coauthored the policy document on the National Guidelines for Precision Agriculture in Italy. He is a member of the Directive Committee of the Italian Society of Agronomy 2016–2017.

Wenjiang Huang received the Ph.D. degree in cartography and GIS from Beijing Normal University, Beijing, China, in 2005. He is currently a Professor with the Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing. His research interests include quantitative remote sensing research and application in vegetation.

Giovanni Laneve received the Laurea degree in aeronautic engineering from the Università di Napoli, Naples, Italy, in 1985, and the Laurea degree in aerospace engineering from the Università di Roma “La Sapienza,” Rome, Italy, in 1988. He has authored more than 220 scientific papers. His research interests include aeronomy, satellite thermal control, mission design, new algorithms for the exploitation of satellite images, satellite remote sensing applications for fire management, applications of satellite data for the African regions, and studies on environmental and disaster monitoring.

Liu Linyi received the B.S. degree in geographic information system from Capital Normal University, Beijing, China, in 2015. He is currently pursuing the Ph.D. degree with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing. His current research interests include crop disease monitoring, remote sensing, and geographic information systems.




Saham Mirzaei is a postdoctoral researcher at the CNR-IMAA of Italy. His research interest is hyperspectral remote sensing and spectroscopy to soil study.

Simone Pascucci earned a bachelor’s degree in environmental sciences from the University of La Tuscia in 2008. Since 2002, he has served as a technical expert in remote sensing at CNR-IMAA, focusing on various applications such as environmental, agronomic, marine, and urban studies. His research encompasses preprocessing hyperspectral images, characterizing contaminated sites, analyzing spectral properties of urban materials, and precision agriculture.

Stefano Pignatti has been a Researcher with the National Research Council, Italy, since 1995. At present Director of Research of the Institute of Methodologies for Environmental Analysis (CNR IMAA), Rome, Italy. His research interests include hyperspectral remote sensing both from airborne (VSWIR and TIR) and satellite platforms (PRISMA, CHIME, SBG-TIR). His current activity deals with data calibration as well as the exploitation of hyperspectral remote sensing in environmental applications, including development of novel approaches for the provision of new soil products from remote sensing.

Yu Ren Yu Ren received the B.S. degree in application and technology of remote sensing and the M.S. degree in geology from the Chengdu University of Technology, China, in 2015 and 2018, respectively. She is currently pursuing the Ph.D. degree in cartography and geographic information systems with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. Her current research interests include crop nitrogen stress and disease monitoring, and quantitative remote sensing of vegetation.

ORCID

Francesco Rossi  <http://orcid.org/0009-0000-2725-1502>
 Raffaele Casa  <http://orcid.org/0000-0003-3091-7680>
 Wenjiang Huang  <http://orcid.org/0000-0003-1710-8301>
 Giovanni Laneve  <http://orcid.org/0000-0001-6108-9764>
 Liu Linyi  <http://orcid.org/0000-0003-4587-2489>
 Saham Mirzaei  <http://orcid.org/0000-0002-8724-1725>
 Simone Pascucci  <http://orcid.org/0000-0002-8311-8615>
 Stefano Pignatti  <http://orcid.org/0000-0002-0587-8926>

Data availability statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

- Angelopoulou, T., S. Chabrillat, S. Pignatti, R. Milewski, K. Karyotis, M. Brell, T. Ruhtz, D. Bochtis, and G. Zalidis. 2023. “Evaluation of Airborne HySpex and Spaceborne PRISMA Hyperspectral Remote Sensing Data for Soil Organic Matter and Carbonates Estimation.” *Remote Sensing* 15 (4). <https://doi.org/10.3390/rs15041106>.
- Angelopoulou, T., N. Tziolas, A. Balafoutis, G. Zalidis, and D. Bochtis. 2019. “Remote Sensing Techniques for Soil Organic Carbon Estimation: A Review.” *Remote Sensing* 11 (6): 676. <https://doi.org/10.3390/rs11060676>.
- ASI. 2020. “PRISMA Products Specification Document Issue 2.3.”
- ASI. n.d. “PRISMA 2G.” Accessed August 23, 2023. <https://www.asi.it/2022/02/prisma2g-siglatto-il-contratto-tra-agenzia-spaziale-italiana-e-thales-alenia-space-per-lavvio-della-fase-di-studio/destinata-allo-sviluppo-del-satellite-di-seconda-generazione/>.
- Bao, Y., X. Meng, S. Ustin, X. Wang, X. Zhang, H. Liu, and H. Tang. 2020. “Vis-SWIR Spectral Prediction Model for Soil Organic Matter with Different Grouping Strategies.” *Catena (Elsevier B V)* 195. <https://doi.org/10.1016/j.catena.2020.104703>.
- Baston, D. 2022. “Exactextract: Fast Extraction from Raster Datasets Using Polygons.” <https://www.asi.it/2022/02/prisma2g/siglatto-il-contratto-tra-agenzia-spaziale-italiana-e-thales-alenia-space-per-lavvio-della-fase-di-studio-destinata-allo-sviluppo-del-satellite-di-seconda-generazione/>.
- Bayer, A. D., M. Bachmann, D. Rogge, A. Müller, and H. Kaufmann. 2016. “Combining Field and Imaging Spectroscopy to Map Soil Organic Carbon in a Semiarid Environment.” *IEEE Journal of Selected Topics in Applied*

- Earth Observations & Remote Sensing* 9 (9): 3997–4010. <https://doi.org/10.1109/JSTARS.2016.2585674>.
- Bellon Maurel, V., and A. Mcbratney. 2011. “Near-Infrared (NIR) and Mid-Infrared (MIR) Spectroscopic Techniques for Assessing the Amount of Carbon Stock in Soils—Critical Review and Research Perspectives.” *Soil Biology & Biochemistry* 43:1398–1410. <https://doi.org/10.1016/j.soilbio.2011.02.019>.
- Ben-Dor, E., Y. Inbar, and Y. Chen. 1997. “The Reflectance Spectra of Organic Matter in the Visible Near-Infrared and Short Wave Infrared Region (400–2500 Nm) During a Controlled Decomposition Process.” *Remote Sensing of Environment* 61 (1): 1–15. [https://doi.org/10.1016/S0034-4257\(96\)00120-4](https://doi.org/10.1016/S0034-4257(96)00120-4).
- Bremner, J. M. 1960. “Determination of Nitrogen in Soil by the Kjeldahl Method.” *The Journal of Agricultural Science* 55:11–33. <https://doi.org/10.1017/S0021859600021572>.
- Caicedo, J. P. R., J. Verrelst, J. Muñoz-Mari, J. Moreno, and G. Camps-Valls. 2014. “Toward a Semiautomatic Machine Learning Retrieval of Biophysical Parameters.” *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 1249–1259. <https://doi.org/10.1109/JSTARS.2014.2298752>.
- Castaldi, F., A. Hueni, S. Chabrillat, K. Ward, G. Buttafuoco, B. Bomans, K. Vreys, M. Brell, and B. Wesemael. 2018. “Evaluating the Capability of the Sentinel 2 Data for Soil Organic Carbon Prediction in Croplands.” *ISPRS Journal of Photogrammetry & Remote Sensing* 147:267–282. <https://doi.org/10.1016/j.isprsjprs.2018.11.026>.
- Castaldi, F., A. Palombo, F. Santini, S. Pascucci, S. Pignatti, and R. Casa. 2016. “Evaluation of the Potential of the Current and Forthcoming Multispectral and Hyperspectral Imagers to Estimate Soil Texture and Organic Carbon.” *Remote Sensing of Environment* 179:54–65. <https://doi.org/10.1016/j.rse.2016.03.02>.
- Chabrillat, S., E. Ben-Dor, J. Cierniewski, C. Gomez, T. Schmid, and B. van Wesemael. 2019. “Imaging Spectroscopy for Soil Mapping and Monitoring.” *Surveys in Geophysics* 40 (3): 361–399.
- Chang, C.-W., D. Laird, M. Mausbach, and C. Hurburgh. 2001. “Near-Infrared Reflectance Spectroscopy—Principal Components Regression Analyses of Soil Properties.” *Soil Science Society of America Journal* 65:480–490. <https://doi.org/10.2136/sssaj2001.652480x>.
- Cogliati, S., F. Sarti, L. Chiarantini, M. Cosi, R. Lorusso, E. Lopinto, F. Miglietta, et al. 2021. “The PRISMA Imaging Spectroscopy Mission: Overview and First Performance Analysis.” *Remote Sensing of Environment* 262:112499. <https://doi.org/10.1016/j.rse.2021.112499>.
- Eches, O., N. Dobigeon, C. Mailhes, and J.-Y. Tourneret. 2010. “Bayesian Estimation of Linear Mixtures Using the Normal Compositional Model. Application to Hyperspectral Imagery.” *IEEE Transactions on Image Processing* 19:1403–1413. <https://doi.org/10.1109/TIP.2010.2042993>.
- ESA. n.d. “CHIME.” Accessed August 23, 2023. <https://www.asi.it/2022/02/prisma2g-siglato-il-contratto-tra-agenzia-spaziale-italiana-e-thales-alenia-space-per-lavvio-della-fase-di-studio-destinata-allo-sviluppo-del-satellite-di-seconda-generazione/>.
- FAO and ITPS. 2015. “Status of the World’s Soil Resources (SWSR) – Main Report.” In *Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils*, Rome, Italy.
- Gasmi, A., C. Gomez, A. Chehbouni, D. Dhiba, and M. El Gharous. 2022. “Using PRISMA Hyperspectral Satellite Imagery and GIS Approaches for Soil Fertility Mapping (FertiMap) in Northern Morocco.” *Remote Sensing* 14 (16). <https://doi.org/10.3390/rs14164080>.
- Gholizadeh, A., C. Neumann, S. Chabrillat, B. van Wesemael, F. Castaldi, L. Borůvka, J. Sanderman, A. Klement, and C. Hohmann. 2021. “Soil Organic Carbon Estimation Using VNIR–SWIR Spectroscopy: The Effect of Multiple Sensors and Scanning Conditions.” *Soil and Tillage Research* 211:105017. <https://doi.org/10.1016/j.still.2021.105017>.
- Gholizadeh, A., M. Saberioon, R. A. Viscarra Rossel, L. Borůvka, and A. Klement. 2020. “Spectroscopic Measurements and Imaging of Soil Colour for Field Scale Estimation of Soil Organic Carbon.” *Geoderma* 357 (16–7061): 113927. <https://doi.org/10.1016/j.geoderma.2019.113972>.
- Gholizadeh, A., D. Žižala, M. Saberioon, and L. Borůvka. 2018. “Soil Organic Carbon and Texture Retrieving and Mapping Using Proximal, Airborne and Sentinel-2 Spectral Imaging.” *Remote Sensing of Environment* 218–103:89. <https://doi.org/10.1016/j.rse.2018.09.015>.
- Gomez, C., R. Oltra-Carrió, S. Bacha, P. Lagacherie, and X. Briottet. 2015. “Evaluating the Sensitivity of Clay Content Prediction to Atmospheric Effects and Degradation of Image Spatial Resolution Using Hyperspectral VNIR/SWIR Imagery.” *Remote Sensing of Environment* 164:1–15. <https://doi.org/10.1016/j.rse.2015.02.019>.
- Gomez, C., R. A. Viscarra Rossel, and A. B. McBratney. 2008. “Soil Organic Carbon Prediction by Hyperspectral Remote Sensing and Field Vis-NIR Spectroscopy: An Australian Case Study.” *Geoderma* 146 (3): 403–411. <https://doi.org/10.1016/j.geoderma.2008.06.011>.
- Guo, Y., J. He, S. Li, G. Zheng, and L. Wang. 2020. “Evaluating the Feasibility of GF-1 Remote Sensing Comparison with Hyperspectral Data for Soil Organic Carbon Prediction and Mapping.” *IOP Conference Series: Earth and Environmental Science* 545 (1). <https://doi.org/10.1088/1755-1315/545/1/012016>.
- Guo, P., T. Li, H. Gao, X. Chen, Y. Cui, and Y. Huang. 2021. “Evaluating Calibration and Spectral Variable Selection Methods for Predicting Three Soil Nutrients Using Vis-NIR Spectroscopy.” *Remote Sensing* 13 (19). <https://doi.org/10.3390/rs13194000>.
- Haghi, R. K., E. Pérez-Fernández, and A. H. J. Robertson. 2021. “Prediction of Various Soil Properties for a National Spatial Dataset of Scottish Soils Based on Four Different Chemometric Approaches: A Comparison of Near Infrared and Mid-Infrared Spectroscopy.” *Geoderma (Elsevier B V)* 396. <https://doi.org/10.1016/j.geoderma.2021.115071>.
- Jing, Z., J. Dongli, D. Dong, M. Jinjie, L. Hongwei, and B. Yaonan. 2020. “Temporal Paradox in Soil Potassium Estimations Using Spaceborne Multispectral Imagery.” *Catena (Elsevier B V)* 194. <https://doi.org/10.1016/j.catena.2020.104771>.
- Ji, W., Z. Shi, J. Huang, and L. Shuo. 2014. “In situ Measurement of Some Soil Properties in Paddy Soil Using Visible and Near-Infrared Spectroscopy.” *PLOS ONE* 9:e105708. <https://doi.org/10.1371/journal.pone.0105708>.
- Karray, E., H. Elmannai, E. Toumi, M. Hedi Gharbia, S. Meshoul, H. Aichi, and Z. Ben Rabah. 2023. “Evaluating the Potentials of PLSR and SVR Models for Soil Properties Prediction Using Field Imaging, Laboratory VNIR Spectroscopy and Their Combination.” *CMES - Computer Modeling in*

- Engineering and Sciences (Tech Science Press)* 136 (2): 1399–1425. <https://doi.org/10.32604/cmes.2023.023164>.
- Keshavarzi, A., F. Kaya, L. Başayığit, Y. Gyasi-Agyei, J. Rodrigo-Comino, and A. Caballero-Calvo. 2023. “Spatial Prediction of Soil Micronutrients Using Machine Learning Algorithms Integrated with Multiple Digital Covariates.” *Nutrient Cycling in Agroecosystems (Institute for Ionics)* 127 (1): 137–153. <https://doi.org/10.1007/s10705-023-10303-y>.
- Kokkas, S., N. L. Tsakiridis, N. Tziolas, K. Karyotis, N. Samarinas, and G. C. Zalidis. 2023. *Topsoil Organic Carbon Estimations in Greece via Deep Learning and Open Earth Observation Data*. 5297–5300. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/IGARSS52108.2023.10282347>.
- Kopittke, P. M., N. W. Menzies, P. Wang, B. A. McKenna, and E. Lombi. 2019. “Soil and the Intensification of Agriculture for Global Food Security.” *Environment International* 132:105078. <https://doi.org/10.1016/j.envint.2019.105078>.
- Lefèvre, C., R. Fatma, A. Viridiana, and WieseLies L. 2017. *Soil Organic Carbon: The Hidden Potential*. Roma: Wiese Lies.
- Liu, S., H. Shen, S. Chen, X. Zhao, A. Biswas, X. Jia, Z. Shi, and J. Fang. 2019. “Estimating Forest Soil Organic Carbon Content using vis-NIR Spectroscopy: Implications for Large-Scale Soil Carbon Spectroscopic Assessment.” *Geoderma* 348 (16–7061): 37–44. <https://doi.org/10.1016/j.geoderma.2019.04.003>.
- Li, H., J. Wang, J. Zhang, T. Liu, G. E. Acquah, and H. Yuan. 2022. “Combining Variable Selection and Multiple Linear Regression for Soil Organic Matter and Total Nitrogen Estimation by DRIFT-MIR Spectroscopy.” *Agronomy* 12 (3): 638. <https://doi.org/10.3390/agronomy12030638>.
- Li, J., Z. Wu, K. Tan, W. Ma, X. Wang, and Y. Zhu. 2024. “GF-5 Hyperspectral Inversion of Soil Parameters Using a VAE Style-2 Based Spectral Fusion Model 3 4 Depin Ou.” <https://www.asi.it/2022/02/prisma2g-siglato-il-contratto-tra-agenzia-spaziale-italiana-e-thales-alenia-space-per-lavvio-della-fase-di-studio-destinata-allo-sviluppo-del-satellite-di-seconda-generazione/>.
- Li, X., G. Zheng, J. Wang, C. Ji, B. Sun, and Z. Gao. 2016. “Comparison of Methods for Estimating Fractional Cover of Photosynthetic and Non-Photosynthetic Vegetation in the Otindag Sandy Land Using GF-1 Wide-Field View Data.” *Remote Sensing* 8 (8). <https://doi.org/10.3390/rs8100800>.
- Loizzo, R., R. Guarini, F. Longo, T. Scopa, R. Formaro, C. Facchinetti, and G. Varacalli. 2018. “Prisma: The Italian Hyperspectral Mission.” *IGARSS 2018 – 2018 IEEE International Geoscience and Remote Sensing Symposium*, (2153–7003): 175–178. <https://doi.org/10.1109/IGARSS.2018.8518512>.
- Ma, Y., B. Minasny, J. A. M. Dematté, and A. B. McBratney. 2023. “Incorporating Soil Knowledge into Machine-Learning Prediction of Soil Properties from Soil Spectra.” *European Journal of Soil Science* 74 (6): John Wiley and Sons Inc. <https://doi.org/10.1111/ejss.13438>.
- Martin, P. D., D. F. Malley, G. Manning, and L. Fuller. 2002. “Determination of Soil Organic Carbon and Nitrogen at the Field Level Using Near-Infrared Spectroscopy.” *Canadian Journal of Soil Science* 82 (4): 413–422. <https://doi.org/10.4141/S01-054>.
- Minasny, B. September 10, 2013. “Why Calculating RPD Is Redundant”.
- Minasny, B., A. B. McBratney, A. M. J.-C. Wadoux, E. N. Acoeb, and T. Sabrina. 2020. “Precocious 19th Century Soil Carbon Science.” *Geoderma Regional* e00306. <https://doi.org/10.1016/j.geodrs.2020.e00306>.
- Mirzaei, S., A. Darvishi Bolorani, H. Ali Bahrami, S. Kazem Alavipanah, A. Mousivand, and A. Mounem Mouazen. 2022. “Minimising the Effect of Moisture on Soil Property Prediction Accuracy Using External Parameter Orthogonalization.” *Soil and Tillage Research* 215. <https://doi.org/10.1016/j.still.2021.105225>.
- Mzid, N., F. Castaldi, M. Tolomio, S. Pascucci, R. Casa, and S. Pignatti. 2022. “Evaluation of Agricultural Bare Soil Properties Retrieval from Landsat 8, Sentinel-2 and PRISMA Satellite Data.” *Remote Sensing* 14 (3). <https://doi.org/10.3390/rs14030714>.
- NASA. n.d. “SBG.” Accessed August 23, 2023. <https://www.asi.it/2022/02/prisma2g-siglato-il-contratto-tra-agenzia-spaziale-italiana-e-thales-alenia-space-per-lavvio-della-fase-di-studio-destinata-allo-sviluppo-del-satellite-di-seconda-generazione/>.
- Ou, D., K. Tan, J. Li, Z. Wu, L. Zhao, J. Ding, X. Wang, and B. Zou. 2023. “Prediction of Soil Organic Matter by Kubelka-Munk Based Airborne Hyperspectral Moisture Removal Model.” *International Journal of Applied Earth Observation and Geoinformation*, 124. Elsevier B.V. <https://doi.org/10.1016/j.jag.2023.103493>.
- Patel, A., J. Ghosh, S. Pande, and S. Sayyad. 2020. “Deep-Learning-Based Approach for Estimation of Fractional Abundance of Nitrogen in Soil from Hyperspectral Data.” *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 13. <https://doi.org/10.1109/JSTARS.2020.3039844>.
- Pignatti, S., A. Amodeo, M. F. Carfora, R. Casa, L. Mona, A. Palombo, S. Pascucci, M. Rosoldi, F. Santini, and G. Laneve. 2022. “PRISMA L1 and L2 Performances within the PRISCAV Project: The Pignola Test Site in Southern Italy.” *Remote Sensing* 14 (9). <https://doi.org/10.3390/rs14091985>.
- Rasooli, N., M. H. Farpour, M. Mahmoodabadi, and I. Esfandiarpour-Boroujeni. 2023. “Vis-NIR Spectroscopy As an Eco-Friendly Method for Monitoring Pedoenvironmental Variations and Pedological Assessments in Lut Watershed, Central Iran.” *Soil and Tillage Research* 233:105808. <https://doi.org/10.1016/j.still.2023.105808>.
- Ren, L., D. Hong, L. Gao, X. Sun, M. Huang, and J. Chanussot. 2023. “Orthogonal Subspace Unmixing to Address Spectral Variability for Hyperspectral Image.” *IEEE Transactions on Geoscience & Remote Sensing*. <https://doi.org/10.1109/TGRS.2023.3236471>.
- Rinnan, Å., F. van den Berg, and S. Balling Engelsen. 2009. “Review of the Most Common Pre-Processing Techniques for Near-Infrared Spectra.” *TrAC Trends in Analytical Chemistry* 28 (10): 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>.
- Rojas, R. V., M. Achouri, J. Maroulis, and L. Caon. 2016. “Healthy Soils: A Prerequisite for Sustainable Food Security.” *Environmental Earth Sciences (Springer Verlag)* 75 (3). <https://doi.org/10.1007/s12665-015-5099-7>.
- Rossel, R. A. V., T. Behrens, E. Ben-Dor, S. Chabrillat, J. Alexandre Melo Dematté, Y. Ge, C. Gomez, et al. 2022. “Diffuse Reflectance Spectroscopy for Estimating Soil Properties: A Technology for the 21st Century.” *European Journal of Soil Science* 73 (4). (John Wiley and Sons Inc). <https://doi.org/10.1111/ejss.13271>.

- Saygin, F., H. Aksoy, P. Alaboz, and O. Dengiz. 2023. "Different Approaches to Estimating Soil Properties for Digital Soil Map Integrated with Machine Learning and Remote Sensing Techniques in a Sub-Humid Ecosystem." *Environmental Monitoring and Assessment* 195(9). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s10661-023-11681-0>.
- Scheffler, D., A. Hollstein, H. Diedrich, K. Segl, and P. Hostert. 2017. "AROSICS: An Automated and Robust Open-Source Image Co-Registration Software for Multi-Sensor Satellite Data." *Remote Sensing* 9 (7): 676. <https://doi.org/10.5281/zenodo.3743085>.
- Shi, P., F. Castaldi, B. Van Wesemael, and K. Van Oost. 2020. "Large-Scale, High-Resolution Mapping of Soil Aggregate Stability in Croplands Using APEX Hyperspectral Imagery." *Remote Sensing* 12 (4). <https://doi.org/10.3390/rs12040666>.
- Song, Y.-Q., X. Zhao, H.-Y. Su, B. Li, Y.-M. Hu, and X.-S. Cui. 2018. "Predicting Spatial Variations in Soil Nutrients with Hyperspectral Remote Sensing at Regional Scale." *Sensors* 18 (9). <https://doi.org/10.3390/s18093086>.
- Stevens, A., M. Nocita, G. Toth, L. Montanarella, and B. Van Wesemael. 2013. "Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy." *PLOS ONE* 8 (6). <https://doi.org/10.1371/journal.pone.0066409>.
- Stevens, A., T. Udelhoven, A. Denis, B. Tychon, R. Lioy, L. Hoffmann, and B. van Wesemael. 2010. "Measuring Soil Organic Carbon in Croplands at Regional Scale Using Airborne Imaging Spectroscopy." *Geoderma* 158 (1): 32–45. <https://doi.org/10.1016/j.geoderma.2009.11.032>.
- Sujatha, M., and C. D. Jaidhar. 2024. "Machine Learning-Based Approaches to Enhance the Soil Fertility—A Review." *Expert Systems with Applications*. Elsevier Ltd. <https://doi.org/10.1016/j.eswa.2023.122557>.
- Viscarra, R., A. Raphael, T. Behrens, E. Ben-Dor, S. Chabrilat, J. Alexandre Melo Demattê, Y. Ge, C. Gomez, et al. 2022. "Diffuse Reflectance Spectroscopy for Estimating Soil Properties: A Technology for the 21st Century." *European Journal of Soil Science* 73 (4): e13271. <https://doi.org/10.1111/ejss.13271>.
- Wang, S., K. Guan, C. Zhang, D. Lee, A. J. Margenot, Y. Ge, J. Peng, W. Zhou, Q. Zhou, and Y. Huang. 2022. "Using Soil Library Hyperspectral Reflectance and Machine Learning to Predict Soil Organic Carbon: Assessing Potential of Airborne and Spaceborne Optical Soil Sensing." *Remote Sensing of Environment* 271:112914. <https://doi.org/10.1016/j.rse.2022.112914>.
- Wang, C., and X. Pan. 2016. "Calibration; Germanium; Near Infrared; Near Infrared Spectroscopy; Projection Systems; Spectrometers." *Journal of Near Infrared Spectroscopy* 3:231–241. <https://doi.org/10.1255/jnirs.1184>.
- Wang, K., Y. Qi, W. Guo, J. Zhang, and Q. Chang. 2021. "Retrieval and Mapping of Soil Organic Carbon Using Sentinel-2A Spectral Images from Bare Cropland in Autumn." *Remote Sensing* 13 (6). <https://doi.org/10.3390/rs13061072>.
- Ward, K. J., S. Chabrilat, M. Brell, F. Castaldi, D. Spengler, and S. Foerster. 2020. "Mapping Soil Organic Carbon for Airborne and Simulated EnMAP Imagery Using the LUCAS Soil Database and a Local PLSR." *Remote Sensing* 12 (20). <https://doi.org/10.3390/rs12203451>.
- Wu, F., K. Tan, X. Wang, J. Ding, and Z. Liu. 2023. "A Novel Semi-Empirical Soil Multi-Factor Radiative Transfer Model for Soil Organic Matter Estimation Based on Hyperspectral Imagery." *Geoderma* (Elsevier B.V.) 437. <https://doi.org/10.1016/j.geoderma.2023.116605>.
- Xu, L., G. Yu, N. He, Q. Wang, Y. Gao, D. Wen, S. Li, S. Niu, and J. Ge. 2018. "Carbon Storage in China's Terrestrial Ecosystems: A Synthesis." *Scientific Reports* 8 (1): 2806.
- Yu, H., B. Kong, G. Wang, R. Du, and G. Qie. 2018. "Prediction of Soil Properties Using a Hyperspectral Remote Sensing Method." *Archives of Agronomy and Soil Science* 64 (4): 546–549. Edited by Taylor & Francis. <https://doi.org/10.1080/03650340.2017.1359416>.
- Zhang, Y., P. Li, X. Liu, L. Xiao, T. Li, and D. Wang. 2022. "The Response of Soil Organic Carbon to Climate and Soil Texture in China." *Frontiers of Earth Science* 16. <https://doi.org/10.1007/s11707-021-0940-7>.