Topsoil clay content mapping in croplands from Sentinel-2 data: Influence of atmospheric correction methods across a season time series

Cécile Gomez, Emmanuelle E. Vaudour, Jean-Baptiste Féret, Florian de Boissieu, Subramanian Dharumarajan

Topsoil clay content mapping in croplands from Sentinel-2 data: influence of atmospheric correction methods across a season time series

Cécile Gomez[1,2], Emmanuelle Vaudour[3], Jean-Baptiste Féret[4], Florian de Boissieu[4,5], Subramanian Dharumarajan[6]

[1] LISAH, Univ. Montpellier, IRD, INRAE, Institut Agro, Montpellier, France; cecile.gomez@ird.fr

[2] Indo-French Cell for Water Sciences, IRD, Indian Institute of Science, Bangalore, India

[3] Université Paris-Saclay, INRAE, AgroParisTech, UMR ECOSYS, 78850, Thiverval-Grignon, France

[4] TETIS, INRAE, AgroParisTech, CIRAD, CNRS, Université Montpellier, Montpellier, France

[5] UMR Eco&Sols, CIRAD, Montpellier France

[6] ICAR-National Bureau of Soil Survey and Land Use Planning, Hebbal, Bangalore, India

Abstract

Recent studies demonstrated the capability of Sentinel-2 (S2) data to estimate topsoil properties and highlighted the sensitivity of these estimations to soil surface conditions depending on the S2 acquisition date. These estimations are based on Bottom Of Atmosphere (BOA) reflectance images, obtained from Top Of Atmosphere (TOA) reflectance values using Atmospheric Correction (AC) methods. AC of optical satellite imagery is an important pre-processing stage before estimating biophysical variables, and several AC methods are currently operational to perform such conversion. This study aims at evaluating the sensitivity of topsoil clay content estimation to atmospheric corrections along an S2 time series. Three AC methods were tested (MAJA, Sen2Cor, and LaSRC) on a time series of eleven Sentinel-2 images acquired over a cultivated region in India (Karnataka State) from February 2017 to June 2017. Multiple Linear Regression models were built using clay content analyzed from topsoil samples collected over bare soil pixels and corresponding BOA reflectance data. The influence of AC methods was also analysed depending on bare soil pixels selections based on two spectral indices and several thresholds: the normalized difference vegetation index (NDVI below 0.25, 0.3 and 0.35) and the combination of NDVI (below 0.3) and Normalized Burned Ratio 2 index (NBR2 below 0.09, 0.12 and 0.15) for masking green vegetation, crop residues and soil moisture.

First, this work highlighted that regression models were more sensitive to acquisition date than to AC method, suggesting that soil surface conditions were more influent on clay content estimation models than variability among atmospheric corrections. Secondly, no AC method outperformed other methods for clay content estimation, and the performances of regression models varied mostly depending on the bare soil pixels selection used to calibrate the regression models. Finally, differences in BOA reflectance among AC methods for the same acquisition date led to differences in NDVI and NBR2, and hence in bare soil coverage identification and subsequent topsoil clay content mapping coverage. Thus, selecting S2 images with respect to the acquisition date appears to be a more critical step than selecting an AC method, to ensure optimal retrieval accuracy when mapping topsoil properties assumed to be relatively stable over time.

Keywords: clay content; Sentinel-2; atmospheric correction; multiple linear regression; soil property mapping; India.

1. Introduction

Soils are key to meeting global environmental sustainability challenges for food security, water security, energy sustainability, climate stability, biodiversity, and ecosystem service delivery (McBratney et al., 2014). An accurate and spatially referenced characterization of soil properties over cultivated areas, including soil organic matter, soil texture, or iron content, is essential for meeting these global environmental sustainability challenges and would also allow to help for planning agricultural engineering work such as land consolidation, drainage management, and soil erosion prevention. Particle-size distribution, also called soil texture, refers to relative amounts of sand, silt, and clay in grams per kilogram (g kg^{-1}) in a soil, the sum of these particle size fractions equaling 1000 g kg^{-1}. Soil texture is a major component of soil, as it has an important influence on water infiltration and soil stabilization (Le Bissonnais et al., 2007, 2018). To ensure soil security, adequate decisions both at global and local levels are required to favor the beneficial roles of soil (Rodrigo-Comino et al., 2020). Such decisions require accurate spatially referenced soil information systems that can be used in environmental modeling.

Spectroscopy covering the Visible, Near Infrared, and Short Wave Infrared domains (VNIR/SWIR, 400 – 2500 nm) is a technology that proved its relevance for the estimation of soil properties (Viscarra Rossel et al., 2006) as the soil reflectance spectrum results from the position and shape of absorption features of chemical constituents (“peaks”) (e.g., water molecules influence the absorption features at specific wavelengths that are the results of overtone and combination modes from the IR region) and overall spectral shape of the physical properties (e.g., texture) (Ben-Dor and Banin, 1995a, 1995b). Most of the absorptions in the VNIR/SWIR region are characteristic of clay and organic matter, and are dominated by C–H, O–H, N–H and metal–OH bonds (e.g., Clark et al., 1990). Soil physical properties, including particle size, scatter the light in a way that the spectrum shape and base line are changing (Wetzel, 1983; Chabrillat et al., 2019). According to Ben-Dor et al. (2002), an accurate estimate of a soil property can be expected from VNIR/SWIR data if the targeted soil property i) is related to a chemical species that impacts soil surface reflectance values through absorption bands (e.g., OH- ion for clay) or ii) is highly correlated with the latter (e.g., CEC correlated with clay content). In addition to these two rules, a minimum level of variability of the targeted soil property across study regions is required to be accurately estimated (Gomez et al., 2012a).

VNIR/SWIR airborne spectroscopy was successfully used to map a large range of soil properties such as iron, soil organic carbon (SOC), and clay contents over bare soil surfaces with high accuracy (e.g., Ben-Dor et al., 2002; Stevens et al., 2010; Gomez et al., 2008, 2012b; Chabrillat et al., 2019). More recently, the VNIR/SWIR multispectral satellites Sentinel-2 (S2) enabled mapping topsoil properties over bare soil surfaces, such as SOC (e.g., Gholizadeh et al., 2018; Vaudour et al., 2019a, b, 2021; Žížala et al., 2019; Castaldi et al., 2021; Dvorakova et al., 2021; Urbina-Salazar et al., 2021) and texture (e.g., Gholizadeh et al., 2018; Gomez et al., 2019; Bousbih et al., 2019). Despite lower accuracy of S2 estimates compared to airborne imaging spectroscopy, the global coverage and high revisit frequency of S2 (five days at the equator) opened perspectives for regional mapping. However, the selection of an S2 acquisition over a multi-temporal series in order to have an optimal estimation of soil properties, such as clay content, raises questions due to multiple factors influencing surface reflectance, and possibly statistical models adjusted to estimate soil properties (Vaudour et al., 2019b; Gomez et al., 2019; Castaldi et al., 2019). Indeed soil surface conditions such as soil moisture and roughness (e.g., due to tillage operations) influence soil reflectance (Chabrillat et al., 2019). The Normalized Difference Vegetation Index (NDVI, based on Red and NIR spectral bands) and the Normalized Burned Ratio 2 (NBR2) index (based on spectral bands around 1600 nm and 2200 nm) have been widely used as indicators for photosynthetic vegetation and dry crop residues, respectively (e.g., for

Landsat 8 by Demattê et al., 2018 and S2 by Castaldi et al. 2019). Nevertheless, while NBR2 follows a linear relationship with crop residue cover over dry soils, no correlation with residue cover could be found over moist soils (Dvorakova et al., 2020). So it is assumed that NBR2 reacts both to crop residues and soil moisture, where high values of NBR2 indicate soils that are moist and/or are covered by crop residues. In the way of creating a composite multi-date bare soil image based on S2 images for soil organic carbon (SOC) prediction, Dvorakova et al., (2021) and Vaudour et al. (2021) developed strategies that allow selecting S2 pixels with minimal influence of crop residues, surface roughness or soil moisture, using spectral indices, especially NDVI and NBR2 indices for detecting these disturbing factors.

Atmospheric conditions also strongly influence the Top of Atmosphere (TOA) reflectance (Level-1C products), and the choice of atmospheric correction (AC) method used to convert S2 Level-1C into Level-2A products may result in differences in terms of Bottom of Atmosphere (BOA) reflectance. These AC methods use different atmospheric models and hypotheses, which may affect the resulting BOA reflectance depending on the season, clouds, sun azimuth, and elevation, and therefore the soil property estimations. Several AC methods have been developed for multispectral images (e.g., Multi-sensor Atmospheric Correction and Cloud Screening -MACCS, Hagolle et al. (2015a), updated and renamed « MACCS-ATCOR joint algorithm » or MAJA, Lonjou et al. (2016); Sen2Cor, Gascon et al. (2017), Louis et al. (2016); Land Surface Reflectance Code -LaSRC, Vermote et al. (2016)). MAJA and Sen2Cor methods were both developed specifically for Sentinel-2 and have become widely used standard products available from Copernicus and other platforms. LaSRC, originally developed by NASA for Landsat, has been adapted recently for Sentinel-2, so it may also become a standard product in the future (e.g., considered in the Brazil Data Cube, Ferreira et al., 2020). These methods use quite different processing paths, from the atmospheric parameter estimation to the radiative transfer model and temporal information, justifying interest in their comparison. A first Atmospheric Correction Inter-comparison Exercise (ACIX) was carried out under an international collaborative initiative to compare a set of AC methods for optical sensors, including S2 (Doxani et al., 2018). However, because the exercise continued, Doxani et al. (2018) did not draw common conclusions from all the algorithms. In addition, Sola et al. (2018a, 2018b) evaluated four AC methods for S2 images, highlighting minor differences between these AC methods. Finally, to the best of our knowledge, no work has been conducted on the impact of AC methods on soil property estimation.

Following, on the one hand, the studies from Doxani et al. (2018) and Sola et al. (2018a, 2018b) evaluating AC methods for S2 images, and those from Gomez et al. (2019) and Vaudour et al. (2019b) studying the impact of acquisition dates on prediction performances of texture and SOC contents, the objective of this study is to evaluate the impact of three major AC methods (Sen2Cor, MAJA, and LaSRC) along an S2 time series on clay content estimation in the topsoil. The influence of these AC methods was also analysed depending on bare soil pixels selections based on two spectral indices and several thresholds: the NDVI (below 0.25, 0.3 and 0.35) and the combination of NDVI (below 0.3) and NBR2 (below 0.09, 0.12 and 0.15). Eleven S2 images covering a cropping cycle in 2017 in a cultivated region in India (Karnataka State) were selected for this study. Soil samples were collected over the study area and analyzed in a laboratory. Clay content was then estimated from each BOA S2 image, corrected by the three AC methods, using multiple linear regression models.

2. Materials and Methods

2.1. Study area

The Berambadi catchment is a subcatchment of the South Gundal located in the Deccan Plateau of Southern India (Figure 1a), extending over 84 km^2 (Figure 1b). Our study area is located in the eastern part of the Berambadi catchment, of which 60% of the surface is dedicated to agriculture, with a high diversity of crops (e.g., sunflower, marigold, sorghum, turmeric, maize, etc.) and high seasonal variability. The western part of the Berambadi catchment is covered by forest (not shown in Figure 1b). The Berambadi catchment belongs to the Kabini Critical Zone Observatory (AMBHAS, BVET, Tomer et al., 2015; Sekhar et al., 2016), which is part of the OZCAR network (Gaillardet et al., 2018). The climate is tropical subhumid with an average rainfall of 800 mm/year and a PET of 1100 mm (aridity index P/PET of 0.7). The monsoon dynamics drive three main seasons: dry season (winter in January and February, summer from March to May), Kharif (southwest monsoon season, from June to September), and Rabi (north-east monsoon season, from October to December). Red soils (Ferralsols and Chromic Luvisols) cover the uplands and hillslopes, while black soils (Vertisols and Vertic intergrades) are mostly found in the valley bottom (Barbiero et al., 2010). Uplands and hillslopes are mainly characterized by coarse soil texture (sandy loam) due to erosion processes, whereas valleys bottoms are mainly characterized by finer soil texture (clay) mainly caused by deposition processes (Gunnell and Bourgeon, 1997; Barbiero et al., 2010).

Three cropping seasons regulate the farm system (Robert et al., 2017). During the Kharif season (from June to September) corresponding to the rainy season, most of the cropping area are cultivated with sorghum, maize, sunflower, marigold, as well as crops grown in irrigated conditions, such as turmeric, onion, garlic, and banana. During the Rabi season (from October to January) corresponding to the winter season, most irrigated plots are cultivated with maize, horse gram, and vegetables. Finally, during the Summer season (from

February to May) corresponding to the hot and dry season, only few plots are cultivated, and almost 90% of the cropland is bare land.

[Figure 1]

2.2. Soil dataset

A total of 164 topsoil samples were collected over Berambadi in November 2019 (Figure 1b). All samples were composed of five sub-samples collected to a depth of 5 cm within a 10 m ×10 m square (one at the center and four at each corner) centered on the geographical position of the sampling plot, as recorded by a Garmin GPS instrument. After sample homogenization, approximately 20 g was devoted to soil property analysis. The samples were then air-dried and sieved with a 2 mm sieve prior to being analyzed in the laboratory. The clay fraction was determined using the pipette method as described by Piper (1966). The clay content ranged between 58 and 622 g kg^{-1} (mean = 252 g kg^{-1}, standard deviation = 124 g kg^{-1} and skewness = 5 g kg^{-1}).

2.3. Sentinel-2 images

Launched in 2015 and then 2017 the combination of both Sentinel-2 satellites (S2A and S2B) delivers a revisit period of five days at the equator. The multispectral sensors acquired information over thirteen spectral bands in the VNIR/SWIR spectral domain, with spatial resolution ranging from 10 m to 60 m.

Twenty-six images from the S2 tile 43PFP were acquired over Berambadi between February 1st and June 30th 2017. This corresponds to the summer season and beginning of Kharif season in South India, when both a maximum of bare soil pixels and a minimum of clouds can be encountered. As clay content is a perennial property in topsoil horizons, we can assume that the clay content analyzed from soil samples collected in 2019 remained unchanged compared to 2017, as assumed in previous studies (e.g., Gomez et al., 2008; Loiseau et al., 2019).

2.3.1. Atmospheric correction method

Atmospheric corrections were performed on the 26 images using three AC methods described in the next subsections (MAJA, Lonjou et al., 2016; Sen2Cor, Louis et al., 2016; LaSRC ; Vermote et al., 2016) providing three time series of atmospherically corrected images (Figure 2a). After atmospheric correction, only ten bands were retained for spectral modeling, corresponding to the four 10 m resolution bands (B2, B3, B4 and B8, see Table 1 in Dvorakova et al., 2021) and the six 20 m resolution bands (B5, B6, B7, B8A, B11 and B12, see Table 1 in Dvorakova et al., 2021). The BOA products were provided with reflectance *10000.

[Figure 2]

a. Sen2Cor

Since 2015, the Copernicus Open Access Hub has provided Level-2A products of S2 imagery data over Europe, using the Sen2Cor processor developed by European Space Agency (ESA) (ESA, 2015; Gascon et al., 2017; Louis et al., 2016). The Sen2Cor processor performs atmospheric correction, terrain and cirrus correction, and scene classification applied to TOA data.

The Sen2Cor 2.8 was applied to produce Level-2A images from Level-1C images using default settings (available in the Ground Image Processing Parameters files delivered with Sen2Cor official release, https://step.esa.int/main/snap-supported-plugins/sen2cor/sen2cor_v2-8). The Planet Digital Elevation Model was used, and cirrus and Bidirectional Reflectance Distribution Function (BRDF) corrections were deactivated. All other parameters were set to their default value. In addition to Level-2A reflectance data, Sen2Cor also produces an Aerosol Optical Thickness (AOT) map, a Water Vapour (WV) map, and a Scene Classification Map (SCM) together with Quality Indicators (QI) for cloud and snow probabilities. After conversion to Level-2A reflectance, Sen2Cor provides the four VNIR spectral bands with their native spatial resolution of 10 m and the six 20 m resolution bands, which were resampled to 10 m using nearest-neighbor interpolation.

b. MAJA

The MAJA processor (Lonjou et al., 2016; Hagolle et al., 2019) was initially developed to perform cloud detection and atmospheric correction over time series of optical images acquired at high resolution and under quasi constant viewing angles. MAJA combines Multi-Mission Atmospheric Correction and Cloud Screening (MACCS) developed by the French Centre National d'Études Spatiales (CNES) and ATCOR developed by the German Aerospace Center (DLR). This spectro-temporal AC method was developed to process images from Formosat-2, Landsat, VENµS, and S2 satellites. MAJA is based on a spectral assumption of the link between red and blue spectral bands and a temporal assumption assuming that a given neighborhood separated by a few days should yield similar surface reflectance (Hagolle et al., 2015a; 2015b). In the current study, the MAJA correction was processed with on-demand PEPS (Plateforme d'Exploitation des Produits Sentinel) processing service, which uses eight S2 acquisitions prior to each acquisition of interest to meet the temporal assumption (https://labo.obs-mip.fr/multitemp/on-demand-sentinel2-l2a-processing-with-maja-on-peps, last access on 2021-07-01). As Sen2Cor, MAJA provides spectral bands with their native spatial resolution. After Level-2A conversion, we used nearest-neighbor interpolation to convert all 20 m bands to 10 m resolution.

c. LaSRC

Landsat Surface Reflectance Code (LaSRC) is an AC method initially developed to convert TOA radiance to BOA reflectance for Landsat 8 collection, which was recently adapted to Sentinel-2 (Vermote et al., 2018). The algorithm performs atmospheric correction, assuming a Lambertian-plane-parallel atmosphere, and using the Second Simulation of the Satellite Signal in the Solar Spectrum (6S) “Urban Clean” model (Vermote et al., 2016; USGS, 2016). The atmospheric parameters required for the inversion include surface pressure (from the National Center for Environmental Prediction Global Data Assimilation System—NCEP GDAS weather model), water vapor (derived from the MODIS near-infrared channels), ozone (from NCEP GDAS), and aerosol properties (AOT and Angstrom exponent) (Claverie et al., 2018). The aerosol properties are estimated using the comparison between assumed surface reflectance ratios computed from MODIS time series and Sentinel-2 TOA reflectance ratios. The correction algorithm is applied at a spatial resolution of 10 m, which is the output resolution for all bands. Therefore, the L1C bands with a resolution of 20 and 60 m are resampled with a nearest-neighbor method in a pre-processing step.

2.3.2. Cloud mask

Pixels identified as clouds or cloud shadows were masked over the S2 time series, using the cloud mask product provided by MAJA (Figure 2a). The MAJA cloud mask is computed with a combination of mono-temporal and multi-temporal approaches (Hagolle et al., 2010) and accounts for different types of clouds (low, high, thin) and corresponding projected shadows (Baetens et al., 2019). Recent cloud mask algorithms inter-comparisons highlighted that the MAJA cloud mask algorithm provided similar to better performances over other cloud and shadow mask algorithms on a large variety of environments (Baetens et al., 2019 ; Tarrio et al., 2020).

2.3.3. Bare soil selection

S2 pixels identified as non-cultivated land were masked using a land-use map available for the study area (AMBHAS Team, 2015) (Figure 2a). These pixels correspond to urban areas, bodies of water, and natural vegetation (forest). Among the 26 images available over the study area, eleven were finally kept for this study, for which the surface of cultivated land outside clouds and cloud shadows covered more than 95% of the Berambadi catchment.

After masking pixels corresponding to non-cultivated land, bare soil were discriminated from photosynthetic vegetation, based on a thresholding applied on the normalized difference vegetation index (NDVI) calculated using the red band B04 (665 nm) and the NIR band B08 (842 nm) (Figure 2a). Three NDVI thresholds common to all images were defined based on visual interpretation from an expert with field knowledge and literature (e.g., Stevens et al., 2008; Vaudour et al., 2016; Lu et al., 2013), : (i) below 0.25, (ii) below 0.3 and (iii) below 0.35. Pixels corresponding to bare soil were also differentiated from crop residue and moist soil, applying thresholding on the Normalized Burned Ratio 2 index (NBR2) calculated using the SWIR1 band B11 (1610 nm) and the SWIR2 band B12 (2202 nm) (Figure 2a). As setting a threshold for the NBR2 index might be difficult without relevant field observation (Dvorakova et al., 2020 and 2021), four thresholds of NBR2 were tested: (i) no threshold, (ii) below 0.15, (iii) below 0.12 and (iv) below 0.09. . This study focused on the common pixels identified as bare soil for all images to compare clay predictions obtained for the different acquisitions and AC methods.

Topsoil water content is directly related to the time since last precipitations. As soil moisture affects spectra (Diek et al., 2016; Vaudour et al., 2019b) and in order to compare the NBR2 response with rainfall events, we calculated the number of days from the last rain for each S2 image to hint at the topsoil humidity conditions, based on daily rainfall measured in the Maddur village, on the West part of the Berambadi catchment. Among these eleven selected images, two images were acquired following heavy rainfall (cumulative rainfall of 25.5 mm and 18.4 mm over the last two days before the S2 data acquisition on 08-03-2017 and 04-04-2017, respectively), three images were acquired following moderate rainfall (cumulative rainfall of 6.5 mm, 5.5 mm and 9 mm over the last five days before the S2 data acquisition on 25-03-2017, 24-04-2017, and 07-05-2017), and six images were acquired after more than five days without rain (03-02-2017, 16-02-2017, 23-02-2017, 26-02-2017, 28-03-2017, and 27-04-2017).

2.4 Spectral Measures

The BOA reflectances produced with the three different AC methods were compared pairwise to estimate the spectral similarity between AC methods over the eleven dates (Figure 2b). These spectral similarity analysis were performed using R software (R Development Core Team, 2015). The coefficient of correlation $r_{i,t}$ was calculated between the BOA reflectance values as follows:

$$r_{i,t} = cor(Refl_{i,t}^{AC1}, Refl_{i,t}^{AC2}) \qquad (1)$$

where $Refl_{i,j}^{AC1}$ and $Refl_{i,j}^{AC2}$ are the BOA reflectance values obtained from AC methods 1 and 2, at band *i* for the S2 image acquired at date *t*. The coefficient of correlation *r* was calculated over the *N* bare soil pixels, for which soil sample was collected and clay content was analyzed.

As the coefficient of correlation $r_{i,t}$ does not reflect the bias in the data (due to change in albedo in our case), the bias $Rbias_{i,t}$ between two AC methods at band *i* for the S2 image acquired at date *t* was calculated as follows:

$$Rbias_{i,t} = \sum_{k=1}^{N} \frac{Refl_{i,t.k}^{AC1}}{N} - \sum_{k=1}^{N} \frac{Refl_{i,t.k}^{AC2}}{N} \qquad (2)$$

where $Refl_{i,t.k}^{AC1}$ and $Refl_{i,t.k}^{AC2}$ are the BOA reflectance values for pixel k, at band i for the S2 image acquired at date t, obtained from one AC and another, respectively, and N is the number of considered pixels. The bias $Rbias_{i,t}$ was calculated over the N bare soil pixels for which soil sample was collected and clay content was analyzed.

The spectral angle was used to analyze the spectral similarity between BOA reflectance spectra. The spectral angle $SA_{k,t}$ considers the whole spectra and not a single band as $r_{i,t}$ and $Rbias_{i,t}$ (Kruse et al., 1993). It was calculated between the BOA spectrum obtained from one AC and another, for pixel k and the S2 image acquired at date t, as follows:

$$SA_{k,t} = \frac{\sum_{i=1}^{nb} Refl_{i,t,k}^{AC1} * Refl_{i,t,k}^{AC2}}{\sqrt{\sum_{i=1}^{nb} {Refl_{i,t,k}^{AC1}}^2} \sqrt{\sum_{i=1}^{nb} {Refl_{i,t,k}^{AC2}}^2}} \quad (3)$$

where $Refl_{i,t.k}^{AC1}$ and $Refl_{i,t.k}^{AC2}$ are the BOA reflectance values at band i obtained from one AC method and another, respectively, for pixel k, and nb is the total number of spectral bands (10 in our case). $SA_{k,t}$ ranges between 0 and 1, with low values corresponding to low spectral similarity and high values corresponding to high spectral similarity. The spectral angle mapper SA_t was finally calculated for the S2 image acquired at date t, over the N bare soil pixels for which soil sample was collected and clay content was analyzed, as follows:

$$SA_t = \frac{1}{N} \sum_{k=1}^{N} SA_{k,t} \quad (4)$$

where $SA_{k,t}$ was calculated from Equation (3).

2.5. Regression model

Regression models and analysis were performed using R software (R Development Core Team, 2015), and both the ade4 (Dray and Dufour, 2007) and pls packages (Mevik and Wehrens, 2007) were used.

A Multiple Linear Regression (MLR) method was used to produce clay maps estimated from S2 images. MLR is a multivariate approach adjusting a linear relationship between a dependent (response) variable (Y-variable, i.e., clay content in the present case), and a set of predictor variables (X-variables, i.e., S2 spectra in the present case) (Tenehaus, 1998) (Figure 2d). A restrictive selection of pixels corresponding to bare soil may result in small sample size available to train and validate regression models using independant datasets. In this work, a k-fold cross-validation (CV) was used. The original dataset was randomly divided into k sub-datasets. Then, k-1 sub-datasets were used as training data, and the remaining one was used as validation data. The CV process was repeated k times, and the model performance was evaluated by averaging prediction error obtained for the k sub-datasets. The k-fold CV method can take full advantage of data, as each part of the original dataset is randomly divided and used for both training and testing. Here, 10-fold cross-validation (CV) was used to build robust methods for estimating the accuracy of MLR models and repeated 5 times.

Finally, three statistical criteria were used to assess model performances: mean absolute error (MAE), root mean square error ($RMSE_{cv}$), and Pearson correlation coefficient (R^2_{cv}) of cross validation (Figure 2e). These statistical criteria were calculated taking into account the 10-fold CV and 5 repetitions.

3. Results

3.1 Bare soil coverage analysis

The distribution of NDVI values calculated over all unmasked pixels showed positive skew for all acquisition dates and AC methods with values lower than 0.25 (Figure 3). Larger differences in NDVI distributions between the three AC methods were obtained for the S2 image acquired on 08-03-2017 (Figure 3). The median value of NDVI distributions obtained using the LaSRC method (around 0.21) was slightly higher than the median obtained using Sen2Cor and MAJA methods (around 0.19) regardless of the date (Figure 3).

[Figure 3]

The distribution of NBR2 values calculated over all unmasked pixels showed positive skew for all acquisition dates and AC methods with values lower than 0.2 (Figure 4). The median value of NBR2 distributions obtained using the LaSRC (around 0.12 along the dates) method was slightly higher than the median obtained using Sen2Cor and MAJA methods (around 0.9 and 0.10 along the dates, respectively), regardless of the date (Figure 4). The AC methods affected the NBR2 values (Figure 4) more than NDVI values (Figure 3) as NBR2 distributions differed from an AC method to another, especially for the S2 images acquired on 08-03-2017, 04-04-2017, 24-04-2017, 27-04-2017 and 07-05-2017.

[Figure 4]

Bare soil coverage selected using both NDVI and NBR2 thresholding varied depending on both acquisition dates and AC methods (Table 1). The acquisition date providing the maximum bare soil coverage varied among AC methods. As an example, using a NDVI below 0.3, the S2 image providing the maximum bare soil coverage was the one acquired on 26-02-2017 from MAJA-corrected S2 images (with 85.9%), on 08-03-2017 from Sen2Cor-corrected S2 images (with 85.9%), and on 23-02-2017 from LaSRC-corrected S2 images (with 82.7%) (Table 1). The acquisition date providing the minimum bare soil coverage also varied from an AC method to another. Still using a NDVI below 0.3, the S2 image providing the minimum bare soil coverage was the one acquired on 07-05-2017 along MAJA-corrected S2 images (with 76.3%), on 03-02-2017 along Sen2Cor-corrected S2 images (with 73.4%) and on 08-03-2017 along LaSRC-corrected S2 images (with 75.1%) (Table 1). Regardless of the NDVI and NBR2 thresholds, the image acquired on 08-03-2017 presented maximum bare soil coverage for all AC methods (range from 5.7 to 20.75%, Table 1). Regardless of the

NDVI and NBR2 thresholds, the Sen2Cor method provided the highest range of bare soil coverage along the acquition dates (range from 9.18 to 18.49%, Table 1). Finally, a majority of LaSRC-corrected S2 images provided slightly less bare soil coverage than those corrected by Sen2Cor and MAJA, regardless of the NDVI threshold or the combination of NDVI and NBR2 thresholds (Table 1).

[Table 1]

The bare soil pixels that were common for all S2 images corrected by MAJA, Sen2Cor and LaSRC covered from 67.3% of the surface based on a NDVI below 0.35 to 1.2% based on NDVI and NBR2 below 0.3 and 0.09, respectively (Table 2). This resulted in varying sample sizes of bare soil locations with clay content information, which ranged from 122 samples based on NDVI below 0.35 to 2 samples based on NDVI and NBR2 below 0.3 and 0.09, respectively (Table 2). Due to the poor bare soil coverage on 08-03-2017 using NBR2 below to 0.12 or lower, especially from the LaSRC method (Table 1, Figure 4), the bare soil coverage common to all dates did not exceed 7%, which allowed 12 or less samples (Table 2).

As the combinations of NDVI below 0.3 and NBR2 below 0.09 or 0.12, were too restrictive, this study used the four datasets reaching more than 45 collected samples identified as bare soil pixels (i.e. extracted using NDVI below 0.25, 0.3, 0.35 and using the combination of NDVI and NBR2 below 0.3 and 0.15, respectively) to train MLR models (Figures 2c and d) and apply them to the corresponding bare soil coverage. These four dataset displayed close distributions with a range between 58 and 592 $g\ kg^{-1}$, a mean around 220 $g\ kg^{-1}$, and skewness from 7 to 8.7 $g\ kg^{-1}$ (Table 2).

[Table 2]

3.2 BOA reflectance analysis based on bare soil pixels identified with NDVI below 0.3

The effect of AC methods on BOA reflectance values was investigated based on the 111 bare soil pixels for which a measured clay content value was available for training the regression models, identified with a NDVI below 0.3 (Table 2). The BOA reflectance values obtained for each image corrected by the three AC methods were compared based on these 111 bare soil pixels. Reflectance obtained with LaSRC showed lower values for almost all bands, except for B04 on 08-03-2017 (Figure 5C) and for B03 (Figure 5B). Reflectance obtained with Sen2Cor from bands B04 to B12 (665 nm to 2190 nm, red boxplots, Figures 5 from C to J) showed systematically higher values than reflectance obtained with MAJA and LaSRC (yellow and blue boxplots, respectively, Figures 5 from C to J), except on 08-03-2017 for Band 04. The largest difference in BOA reflectance obtained among the three AC methods at the same date was observed on 08-03-2017 and mainly over the visible spectral bands (Figures 5A, B, and C). The SA_t values (Equation (4)), computed for pairwise comparison among AC methods over the 111 topsoil spectra, are high, up to 0.994, showing high similarities between spectra corrected by different AC methods (data not shown).

[Figure 5]

The Pearson's correlation coefficient $r_{i,t}$ (Equation (1)) computed between the 111 topsoil spectra corrected by MAJA and Sen2Cor were very high for all bands and dates, ranging from 0.93 to 0.99 (Figure 6A1). The lowest correlations were obtained for B02 (490 nm) and B03 (560 nm), with a mean of 0.97, while the NIR and SWIR bands showed a very high correlation (> 0.99) (Figure 6A1). Along the eleven dates, the lowest correlations were obtained for the S2 image acquired on 08-03-2017 (mean of 0.99 over the spectral bands), while the highest correlations were obtained for the S2 image acquired on 16-02-2017 (mean up to 0.999 over the spectral bands) (Figure 6A1).

The correlations $r_{i,t}$ between the 111 topsoil spectra corrected by MAJA and LaSRC followed similar patterns to correlations $r_{i,t}$ calculated between the 111 spectra corrected by Sen2Cor and LaSRC (Figures 6A2 and A3). The correlations $r_{i,t}$ were very high varying from 0.90 to 0.99, except for B02 (from 0.52 and 0.90) and B03 (from 0.79 and 0.97). Along the eleven dates, the lowest correlations were obtained for the S2 image acquired on 08-03-2017 (mean of 0.88 over the spectral bands, Figures 6A2 and A3).

The $Rbias_{i,t}$ (in absolute value) between the 111 spectra corrected by MAJA and Sen2Cor ranged from 2.5 (B02 at 490 nm, on 23-02-2017) to 400.0 (B12 at 2190 nm, on 24-04-2017) for all the bands and dates (Figure 6B1). Along the 10 spectral bands, the highest $Rbias_{i,t}$ were obtained for the B12 (2190 nm) (mean of 302.2 in absolute value) while the lowest $Rbias_{i,t}$ were obtained for the B02 (490 nm) (mean of 68 in absolute value) (Figure 6B1). Along the eleven dates, the highest $Rbias_{i,t}$ were obtained for the S2 image acquired on 04-04-2017 (mean of 220.5 in absolute value), while the lowest $Rbias_{i,t}$ were obtained for the S2 image acquired on 16-02-2017 (mean of 76.3 in absolute value) (Figure 6B1).

The $Rbias_{i,t}$ (in absolute value) between the 111 spectra corrected by MAJA and LaSRC varied from 0.00 (B8A at 865 nm, on 28-03-2017) to 249.6 (B02 at 490 nm, on 08-03-2017) for all bands and dates (Figure 6B2). Along the 10 spectral bands, the highest $Rbias_{i,t}$ were obtained for the B02 (490 nm) (mean of 112.3 in absolute value) while the lowest $Rbias_{i,t}$ were obtained for the B08A (865 nm) (mean of 10.9 in absolute value) (Figure 6B2). Along the eleven dates, the highest $Rbias_{i,t}$ were obtained for the S2 image acquired on 08-03-2017 (mean of 135.9 in absolute value), while the lowest $Rbias_{i,t}$ were obtained for the S2 image acquired on 25-03-2017 (mean of 33.2 in absolute value) (Figure 6B2).

The $Rbias_{i,t}$ (in absolute value) between the 111 spectra corrected by Sen2Cor and LaSRC varied from 20 (B02 at 490 nm, on 03-02-2017) and 460.8 (B05 at 705 nm, on 08-03-

2017) for all the bands and dates (Figure 6B3). Along the 10 spectral bands, the highest $Rbias_{i,t}$ were obtained for the B05 (705 nm) (mean of 310.9 in absolute value), while the lowest biases were obtained for the B08A (865 nm) (mean of 128.8 in absolute value) (Figure 6B3). Along the eleven dates, the highest $Rbias_{i,t}$ were obtained for the S2 image acquired on 08-03-2017 (mean of 308.4 in absolute value), while the lowest $Rbias_{i,t}$ were obtained for the S2 image acquired on 16-02-2017 (mean of 100.8 in absolute value) (Figure 6B3).

[Figure 6]

3.3 Clay predictions based on bare soil pixels identified with NDVI below 0.3

The effect of AC methods on clay content estimations was firstly investigated based on bare soil pixels obtained with NDVI below 0.3. MLR models were built from each S2 image and each AC method using the 111 topsoil samples identified with NDVI below 0.3, providing 33 MLR models. Performances for the prediction of clay content strongly varied depending on the acquisition date, with R^2_{cv} ranging from 0.49 to 0.72 for MAJA, from 0.45 to 0.71 for Sen2Cor, and from 0.43 to 0.71 for LaSRC (Table 3).

The difference in R^2_{cv} obtained when comparing dates of acquisition for each AC method varied from 0.23 (obtained from MAJA-corrected S2 images) to 0.29 (obtained from LaSRC-corrected S2 images) (Table 3), suggesting a strong impact of the acquisition date on clay content estimation. Based on MAJA- and Sen2Cor-corrected S2 images, the best MLR model was obtained from the S2 image acquired on 24-04-2017 (R^2_{cv} > 0.71 and $RMSE_{CV}$ > 6.50%, Table 3), while based on LaSRC-corrected S2 images, the best MLR model was obtained from the S2 image acquired on 25-03-2017 (R^2_{cv} of 0.71 and $RMSE_{CV}$ of 6.38%, Table 3). The MLR models built from the S2 image acquired on 04-04-2017 provided poor performances for clay content prediction (R^2_{cv} < 0.55 and $RMSE_{CV}$ > 6.50%, Table 3) for all AC methods.

The difference in R^2_{cv} obtained when comparing the three corrected S2 images for a given date of acquisition varied from 0.00 (obtained on 27-04-2017) to 0.07 (obtained on 08-03-2017) (Table 3), suggesting a poor impact of the AC method selection on clay content estimation. The three AC methods provided very close regression performances for six S2 images (03-02-2017, 26-02-2017, 25-03-2017, 28-03-2017, 24-04-2017 and 27-04-2017, Table 3). Considering the five remaining S2 images, MAJA provided corrected S2 images associated to the best clay content estimations for four acquisition dates, while LaSRC provided corrected S2 images associated to the best clay content estimations for only one acquisition dates (04-04-2017, Table 3).

Finally, the largest difference in R^2_{cv} obtained when comparing AC methods for a given date of acquisition was 0.07, while the largest difference in R^2_{cv} obtained when comparing dates of acquisition for each AC method was 0.29 (Table 3). So it suggested a stronger impact of the acquisition date on clay prediction performance compared to the AC method.

[Table 3]

3.4 Influence of bare soil pixels identification on clay predictions

The analysis of AC influence on clay content estimations was extended to varying bare soil pixels identifications: with NDVI below 0.25 and 0.35, and with the combination of NDVI and NBR2 below 0.3 and 0.15, respectively. Based on these bare soil pixels identifications, MLR models were built from MAJA-, Sen2Cor- and LaSRC-corrected images.

Regardless of the AC methods and thresholdings, the poorest performances were obtained from the S2 image acquired on 08-03-2017 (Table 3). Using MAJA-corrected S2 images, the best performances were obtained from the S2 image acquired on 24-04-2017, independently from thresholding. Using Sen2Cor- and LaSRC-corrected S2 images, the best performances were obtained from S2 images acquired on 25-03-2017 and 24-04-2017, depending on thresholding (Table 3).

Regression models performances obtained based on bare soil coverage selected using NDVI below 0.3 and 0.35, were similar. Using such NDVI thresholds, the largest difference in R^2_{cv} obtained when comparing AC methods for a given date of acquisition was 0.07, while the largest difference in R^2_{cv} obtained when comparing dates of acquisition for each AC method was 0.29 (Table 3). Using the 84 topsoil samples identified with NDVI below 0.25, the largest difference in R^2_{cv} obtained when comparing AC methods for a given date of acquisition was 0.04, while the largest difference in R^2_{cv} obtained when comparing dates of acquisition for each AC method was 0.31 (Table 3). Using the 47 topsoil samples identified with NDVI below 0.3 and NBR2 below 0.15, the largest difference in R^2_{cv} obtained when comparing AC methods for a given date of acquisition was 0.08, while the largest difference in R^2_{cv} obtained when comparing dates of acquisition for each AC method was 0.26 (Table 3). These results suggest that the acquisition date has stronger influence than the AC method on clay prediction performance. Finally, applying restrictive NDVI thresholds or combining NDVI and NBR2 thresholds did not systematically improve the models performance for a given acquisition date (Table 3).

4. Discussion

Variation in bare soils coverage

The extent of bare soil coverage varied depending on acquisition date, AC method and index thresholds (Table 1), with more variability due to dates and index thresholds than AC methods. While the non-cultivated lands mask (urban areas, bodies of water, and natural vegetation forest) was unique regardless of AC methods and date of acquisition, the cloud, and land masks were consistent only among the three AC methods but not among the eleven dates, and the photosynthetic vegetation masks varied between AC methods and the date of acquisition. Indeed, the clouds were masked using one cloud mask specific to each S2 date provided by MAJA (Table 1, Section 2.3.2), and the photosynthetic vegetation was masked by applying a threshold on the NDVI calculated using the red band B04 (665 nm) and the NIR band B08 (842 nm) of each S2 image.

As clouds only affected three dates (08-03-2017, 04-04-2017, and 27-04-2017, Table 1) and bands B04 and B08 differed slightly depending on the S2 acquisition date (Figures 5C and G, respectively), the variability of bare soil coverage along the dates was mainly due to the cut-off induced by NDVI and NBR2 values, which may vary according to some changes in soil surface conditions (due to humidity, roughness), new crop growing within the dates (Vaudour et al., 2019b) or crop residues (Dvorakova et al., 2021). The variability of NDVI values for the same date depending on AC methods was also previously observed by Sola et al. (2018a, 2018b).

The choice of an AC method showed very minor influence on bare soil coverage, except for the image acquired on 08-03-2017 (e.g., from 58.22% to 78.97% using NDVI < 0.25 and LaSRC- and Sen2Cor-corrected S2 images, Table 1). For this specific date, such difference may be due to some unmasked clouds (including cirrus) over the Berambadi (Table 1) affecting the reflectance along the spectra, hence the NDVI values. Tarrio et al. (2020) showed that although MAJA detects clouds and cloud shadows fairly well, producing few errors of omission, the majority of omitted clouds for MAJA was composed of high cirrus clouds.

Variation in BOA spectra (based on bare soil pixels identified with NDVI below 0.3)

Atmospheric corrections performed with Sen2Cor resulted in higher reflectance values for bands B04 to B12 (665 nm to 2190 nm, red boxplots, Figures 5 from C to J), except for the date 08-03-2017 in Band 04 (Figure 5C), compared to others AC methods. These results are in agreement with those obtained by Sola et al. (2018a, 2018b), who reported higher BOA reflectance obtained by Sen2Cor compared to MAJA and the 6S model. Li et al. (2018) also observed an overestimation of surface reflectance by Sen2Cor, especially for bright pixels, and suggested that may be due to an AOT overestimation. Reflectance obtained with LaSRC showed systematically lower values for almost all bands, with the exception for B04 on 08-03-2017 (Figure 5C) and generally for B03 (Figure 5B), than reflectance obtained with MAJA and Sen2Cor. These results are still in agreement with Sola et al. (2018a), who reported lower BOA reflectance obtained by the 6S model compared to MAJA and Sen2Cor. Moreover, reflectance values obtained on 08-03-2017 from bands B05 to B12 (705 nm to 2190 nm, Figure 4) were the lowest of the time series, which may be explained by a higher soil moisture content on the 08-03-2017 image due to previous rainfall, likely to cause a general decrease in reflectance (Minasny et al., 2011) or some unmasked clouds (including cirrus).

Variations in performances between AC methods might be explained by several factors: i) MAJA is based on a multitemporal approach, while other methods are based on individual acquisitions and ii) the computation of AOT and water vapor as well as the radiative transfer model differs among methods, iii) each AC method uses its own cloud mask method to screen the main cloudy pixels for which AC parameter estimations (such as AOT) would probably not converge. Regarding this last factor, a gap-filling operation was performed to fill AOT values for cloudy pixels (either by using a constant value for MAJA or interpolation for Sen2Cor and LaSRC) before applying atmospheric correction with radiative transfer models. Such differences in cloudy pixels may result in local differences in correction accuracy, especially in the surroundings of shadowed or clouded pixels, and therefore contribute to a small part of the difference between the accuracy of MAJA and the other methods. However, considering the low rate of omission of MAJA cloud detection exposed in

Baetens et al. (2019), as well as the spatial homogeneity of the atmospheric correction parameters, this cannot fully explain the differences in accuracy observed between the methods.

Despite these differences observed for BOA reflectance produced by the three AC methods (Figures 5 and 6), the spectra were highly correlated regardless of the date of acquisition, except for 08-03-2017, and regardless of the bands, except the B02 (490 nm). Such high correlations are in accordance with the results obtained by Padró et al. (2017), who compared BOA reflectances obtained by several AC methods, including Sen2Cor, with in situ reflectances, and obtained high correlations regardless of the AC methods ($r^2 > 0.9$).

AC method as minor driver of prediction performances

Our MLR regression models for clay content estimations provided performances (Table 3) in accordance with the performances obtained by Shabou et al. (2015) and Gasmi et al. (2021) using multispectral Landsat images over Mediterranean contexts, and by Bellinaso et al. (2021) and Vaudour et al. (2019a) using S2 images over tropical and temperate/Mediterranean contexts, respectively.

Regardless of the bare soil coverage selection, the AC method had a limited influence on MLR performances for clay content prediction (Table 3). Indeed the MLR models built from MAJA-, Sen2Cor- and LaSRC-corrected S2 images provided close performances (Table 3). This is in accordance with Marujo et al. (2021), who expected that LaSRC and Sen2Cor would converge to produce consistent and comparable data from both sensors. However, this is not in accordance with Sola et al. (2018a), who showed MAJA provided better performances in the analysis per land cover compared to Sen2Cor.

S2 acquisition date as the main driver of prediction performances

Regardless of the bare soil coverage selection, the date of S2 image acquisition showed a stronger influence than the AC method on the performances for clay content prediction (Table 3). As previously shown by Demattê et al. (2018) and then Castaldi et al. (2019), Vaudour et al. (2019b) and Dvorakova et al. (2021), soil surface conditions impact the accuracy of SOC prediction models. This seems the case through our observed time-series, for which the dates acquired shortly after a rainfall event resulted in the lowest performance (on 08-03-2017 and 04-04-2017, Table 3), presumably due to higher soil moisture.

As the date of S2 image acquisition appears to be very important for soil property estimation, and this even more compared to the AC method, future researches could focus on identifying rules for selecting the best date of image acquisition. This was started by selecting dates with average driest condition or lower crop residues over bare soils based on NDVI and NBR2 spectral indices and enabled to obtain the best prediction performances using single-dates images (Castaldi et al., 2019; Vaudour et al., 2019b; Urbina-Salazar et al., 2021).

Similarly, rules for date selection may be used to create multi-temporal image composites composed of multidate images stacked over the same tile. From now, multi-temporal image composites are built from either pixelwise thresholding based on the minimum pixel value (Loiseau et al., 2019; Vaudour et al., 2021), average reflectance value (e.g., Demattê et al., 2018; Gasmi et al., 2019; Dvorakova et al., 2021), median reflectance value (Castaldi et al., 2021; Luo et al., 2022), or considering a trade-off between average per date-indices and maximum bare soil coverage (Vaudour et al., 2021) along a multidate satellite series. Multiple studies suggested using multi-temporal image composites to maximize bare soil coverage for soil property prediction (e.g., Demattê et al., 2018; Vaudour et al., 2021; Dvorakova et al., 2021). However, current researches do not converge towards a common conclusion about their benefits compared to single-date images, especially in terms of soil property prediction accuracies. While Gasmi et al. (2019) showed that multi-temporal image composites based on mean spectral reflectance from bare soil pixels along a Landsat-TM time series allowed to increase both the prediction accuracy of soil clay content and mapping coverage, Vaudour et al. (2021) showed that none of the multi-temporal image composites based on pixelwise or per-date bare soil reflectance along a S2 time series improved model performance for SOC prediction compared to the best single-date image. Therefore, benefit and methodologies for multi-temporal mosaicking are critical questions remaining to be explored.

5. Conclusions

The influence of three atmospheric corrections, namely MAJA, Sen2Cor and LaSRC, on eleven S2 images was evaluated based on clay content estimation over bare soil pixels. Our study highlighted the influence of the S2 acquisition date and AC method on model performances adjusted for clay content estimations, with more variability induced by the acquisition date than the AC method. Thus, the influence of the choice of an AC on the estimation of soil properties can be considered as moderate compared to soil surface conditions, such as moisture, crop residue or roughness, which may be strongly variable in space and time. As regression models performances were close from one AC method to another, this work did not allow to consider one AC method to be the best method prior estimating clay content. As Sen2Cor provided performances for clay content estimations close to MAJA and LaSRC methods, and since ESA provides corrected imagery with Sen2Cor, this AC method might be a satisfactory choice. Finally, as soil properties such as organic carbon and iron are key properties influencing soil radiometric properties in a different manner than clay, the impact of AC methods on the estimation of such topsoil properties could be further investigated to test the robustness of our conclusions.

Acknowledgments

This project was funded by the Programme National de Télédétection Spatiale (PNTS, http://www.insu.cnrs.fr/pnts), grant no. PNTS-2019-5. This research was also supported by the project ATCHA ANR-16-CE03-0006 and by the TOSCA "Cartographie Numérique des Sols (CNS)" program (grant number 3261-3264 CES Theia CartoSols) of the CNES. The authors are indebted to NBSS&LUP for soil samples collection. The Kabini Critical Zone Observatory (AMBHAS, BVET, Sekhar et al., 2016; Tomer et al., 2015, www.ambhas.com; https://mtropics.obsmip.fr/), which is part of the OZCAR network (Gaillardet et al., 2018, http://www.ozcar-ri.org/ozcar/), are also acknowledged.

References

AMBHAS Team, 2015. A Manual for Agro-Hydrological Monitoring in Pilot Experimental Watersheds. Indian Institute of Sciences and AMBHAS August 2015. 65 pages. https://watershed.karnataka.gov.in/storage/pdf-files/Sujala%20Docs/Hydrology%20Manual.pdf.

Baetens, L., Desjardins, C., Hagolle, O., 2019. Validation of Copernicus Sentinel-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure. *Remote Sensing*, 11, 433. https://doi.org/10.3390/rs11040433

Barbiero, L., Kumar, M.S.M., Violette, A., Oliva, P., Braun, J.J., Kumar, C., Furian, S., Babic, M., Riotte, J., Valles, V., 2010. Ferrolysis induced soil transformation by natural drainage in vertisols of sub-humid South India. *Geoderma*, 156, pp. 173–188. https://doi.org/10.1016/j.geoderma.2010.02.014

Bellinaso, H., Silvero, N. E., Ruiz, L. F. C., Accorsi Amorim, M. T., Rosin, N. A., de Sousa Mendes, W., de Sousa, G. P. B., Sepulveda, L. M. A., de Queiroz, L. G., Nanni, M. R.

& Demattê, J.A., 2021. Clay content prediction using spectra data collected from the ground to space platforms in a smallholder tropical area. *Geoderma*, 399, 115116. https://doi.org/10.1016/j.geoderma.2021.115116

Ben-Dor, E., Banin, A., 1995a. Near infrared analysis (NIRA) as a method to simultaneously evaluate spectral featureless constituents in soils. *Soil Science*, 159(4):259–270. https://doi.org/10.1097/00010694-199504000-00005

Ben-Dor, E, Banin, A., 1995b. Near infrared analysis (NIRA) as a rapid method to simultaneously evaluate several soil properties. *Soil Sci Soc Am Journal*, 59:364–372. 10.2136/sssaj1995.03615995005900020014x

Ben-Dor, E., Patkin, K., Banin, A., Karnieli, A., 2002. Mapping of several soil properties using DAIS-7915 hyperspectral scanner data: a case study over clayey soils in Israel. *International Journal of Remote Sensing*, 23:1043–1062. https://doi.org/10.1080/01431160010006962

Bousbih, S., Zribi, M., Pelletier, C., Gorrab, A., Lili-Chabaane, Z., Baghdadi, N., Ben Aissa, N., Mougenot, B., 2019. Soil Texture Estimation Using Radar and Optical Data from Sentinel-1 and Sentinel-2. *Remote Sensing*. 11, 1520. https://doi.org/10.3390/rs11131520

Castaldi, F., Chabrillat, S., Don, A., van Wesemael, B. 2019. Soil Organic Carbon Mapping Using LUCAS Topsoil Database and Sentinel-2 Data: An Approach to Reduce Soil Moisture and Crop Residue Effects. *Remote Sensing*. 11(18):2121. https://doi.org/10.3390/rs11182121

Castaldi, F. 2021. Sentinel-2 and Landsat-8 Multi-Temporal Series to Estimate Topsoil Properties on Croplands. *Remote Sensing*, 13, 3345. https://doi.org/10.3390/rs13173345

Chabrillat, S., Ben-Dor, E., Cierniewski, J., Gomez, C., Schmid, T., van Wesemael, B., 2019. Imaging Spectroscopy for Soil Mapping and Monitoring. Surveys in Geophysics. Springer Netherlands, pp. 1-39. https://doi.org/10.1007/s10712-019-09524-0

Clark, R.N., King, T.V., Klejwa, M., Swayze, G.A., Vergo, N., 1990. High spectral resolution reflectance spectroscopy of minerals. J. Geophys. Res. 95:12653-12680.

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, 219, pp. 145–161. https://doi.org/10.1016/j.rse.2018.09.002

Demattê, J.A.M., Fongaro, C.T., Rizzo, R., Safanelli, J.L., 2018. Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images, Remote Sensing of Environment, 212, Pages 161-175, https://doi.org/10.1016/j.rse.2018.04.047.

Dvorakova, K., Shi, P., Limbourg, Q., van Wesemael, B. 2020. Soil Organic Carbon Mapping from Remote Sensing: The Effect of Crop Residues. *Remote Sensing*, 12, 1913.

Dvorakova, K., Heiden, U., van Wesemael, B. 2021. Sentinel-2 Exposed Soil Composite for Soil Organic Carbon Prediction. *Remote Sensing*, 13, 1791. https://doi.org/10.3390/rs13091791

Diek, S., Schaepman, M.E., De Jong, R., 2016. Creating Multi-Temporal Composites of Airborne Imaging Spectroscopy Data in Support of Digital Soil Mapping. *Remote Sensing*, 8, 906. https://doi.org/10.3390/rs8110906

Doxani G., Vermote, E., Roger, J.-C., Gascon, F., Adriaensen, S., Frantz, D., Hagolle, O., Hollstein, A., Kirches, G., Li, F., Louis, J., Mangin, A., Pahlevan, N., Pflug, B., Vanhellemont, Q., 2018. Atmospheric correction inter-comparison exercise, *Remote Sensing*, 10, p. 352. https://doi.org/10.3390/rs10020352

Dray, S., Dufour, A.B., 2007. The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22, pp. 1-20. 10.18637/jss.v022.i04

ESA (E.S. Agency), 2015. Sentinel-2 User Handbook. 64 pp.

Ferreira, K. R., Queiroz, G. R., Vinhas, L., Marujo, R. F. B., Simoes, R. E. O., Picoli, M. C. A., Camara, G., Cartaxo, R., Gomes, V. C. F., Santos, L. A., Sanchez, A. H., Arcanjo, J. S., Fronza, J. G., Noronha, C. A., Costa, R. W., Zaglia, M. C., Zioti, F., Korting, T. S., Soares, A. R., Chaves, M. E. D., Fonseca, L. M. G., 2020. Earth Observation Data Cubes for Brazil: Requirements, Methodology and Products. Remote Sensing, 12(24), 4033. doi:10.3390/rs12244033

Gaillardet, J., Braud, I., Hankard, F., Anquetin, S., Bour, O., Dorfliger, N., de Dreuzy, J.R., Galle, S., Galy, C., Gogo, S., Gourcy, L., Habets, F., Laggoun, F., Longuevergne, L., Le Borgne, T., Naaim-Bouvet, F., Nord, G., Simonneaux, V., Six, D., Tallec, T., Valentin, C., Abril, G., Allemand, P., Arènes, A., Arfib, B., Arnaud, L., Arnaud, N., Arnaud, P., Audry, S., Comte, V.B., Batiot, C., Battais, A., Bellot, H., Bernard, E., Bertrand, C., Bessière, H., Binet, S., Bodin, J., Bodin, X., Boithias, L., Bouchez, J., Boudevillain, B., Moussa, I.B., Branger, F., Braun, J.J., Brunet, P., Caceres, B., Calmels, D., Cappelaere, B., Celle-Jeanton, H., Chabaux, F., Chalikakis, K., Champollion, C., Copard, Y., Cotel, C., Davy, P., Deline, P., Delrieu, G., Demarty, J., Dessert, C., Dumont, M., Emblanch, C., Ezzahar, J., Estèves, M., Favier, V., Faucheux, M., Filizola, N., Flammarion, P., Floury, P., Fovet, O., Fournier, M., Francez, A.J., Gandois, L., Gascuel, C., Gayer, E., Genthon, C., Gérard, M.F., Gilbert, D., Gouttevin, I., Grippa, M., Gruau, G., Jardani, A., Jeanneau, L., Join, J.L., Jourde, H., Karbou, F., Labat, D., Lagadeuc, Y., Lajeunesse, E., Lastennet, R., Lavado, W., Lawin, E., Lebel, T., Le Bouteiller, C., Legout, C., Lejeune, Y., Le Meur, E., Le Moigne, N., Lions, J., Lucas, A., Malet, J.P., Marais-Sicre, C., Maréchal, J.C., Marlin, C., Martin, P., Martins, J., Martinez, J.M., Massei, N., Mauclerc, A., Mazzilli, N., Molénat, J., Moreira-Turcq, P., Mougin, E., Morin, S., Ngoupayou, J.N., Panthou, G., Peugeot, C., Picard, G., Pierret, M.C., Porel, G., Probst, A., Probst, J.L., Rabatel, A., Raclot, D., Ravanel, L., Rejiba, F., René, P., Ribolzi, O., Riotte, J., Rivière, A., Robain, H., Ruiz, L., Sanchez-Perez, J.M., Santini, W., Sauvage, S., Schoeneich, P., Seidel, J.L., Sekhar, M., Sengtaheuanghoung, O., Silvera, N., Steinmann, M., Soruco, A., Tallec, G., Thibert, E., Lao, D.V., Vincent, C., Viville, D., Wagnon, P., Zitouna, R., 2018. OZCAR: the French network of critical zone observatories. Vadose Zone J. 17, 180067. *Vadose Zone Journal*, 17:180067. https://doi.org/10.2136/vzj2018.04.0067

Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., Languille, F., Alhammoud, B., Viallefont, F., Pflug, B., Bieniarz, J., Clerc, S., Pessiot, L., Trémas, T., Cadau, E., De Bonis, R., Isola, C., Martimort, P., Fernandez, V., 2017. Copernicus Sentinel-2A Calibration and Products Validation Status. *Remote Sensing*, 9, 584. https://doi.org/10.3390/rs9060584

Gasmi A., Gomez, C., Lagacherie P., Zouari, H., Laamrani, A., Chehbouni, A., 2021. Mean spectral reflectance from bare soil pixels along a Landsat-TM time series to increase both the prediction accuracy of soil clay content and mapping coverage. *Geoderma*, 388, 15, 114864. https://doi.org/10.1016/j.geoderma.2020.114864

Gholizadeh, A., Žižala, D., Saberioon, M., Borůvka, L., 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sens. Environ*. 218, 89–103. https://doi.org/10.1016/j.rse.2018.09.015

Gomez, C., Lagacherie, P., Coulouma, G. 2008. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma*, 148(2), pp. 141-148. https://doi.org/10.1016/j.geoderma.2008.09.016

Gomez, C., Lagacherie, P., Bacha, S., 2012a. Using an VNIR/SWIR hyperspectral image to map topsoil properties over bare soil surfaces in the Cap Bon region (Tunisia). In: Digital Soil Assessments and Beyond, Minasny, B., Malone, B.P., McBratney, A.B. (Eds.), Springer, pp. 387–392. DOI:10.1201/b12728-76

Gomez, C., Coulouma, G., Lagacherie, P., 2012b. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis–NIR data, *Geoderma*, 189–190, pp. 176-185. https://doi.org/10.1016/j.geoderma.2012.05.023

Gomez, C., Dharumarajan, S., Féret, J.-B., Lagacherie, P., Ruiz, L., Sekhar, M., 2019. Use of Sentinel-2 Time-Series Images for Classification and Uncertainty Analysis of Inherent Biophysical Property: Case of Soil Texture Mapping. *Remote Sensing*. 11(5), 565. https://doi.org/10.3390/rs11050565

Gunnell, Y., Bourgeon, G., 1997. Soils and climatic geomorphology on the Karnataka plateau, peninsular India. *Catena*, 29, 239–262. https://doi.org/10.1016/S0341-8162(96)00070-7

Hagolle, O., Huc, M., Pascual, D.V., Dedieu, G., 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENµS, LANDSAT and SENTINEL-2 images. *Remote Sensing of Environment*, 114, pp. 1747–1755. https://doi.org/10.1016/j.rse.2010.03.002

Hagolle, O., Huc, M., Villa Pascual, D., Dedieu, G., 2015a. A multi-temporal and multi-spectral method to estimate aerosol optical thickness over land, for the atmospheric correction of FormoSat-2, LandSat, VENµS and Sentinel-2 images. *Remote Sensing*, 7, pp. 2668–2691. https://doi.org/10.3390/rs70302668

Hagolle, O., Sylvander, S., Huc, M., Claverie, M., Clesse, D., Dechoz, C., Lonjou, V., Poulain, V., 2015b. SPOT-4 (Take 5): Simulation of Sentinel-2 Time Series on 45 Large Sites. *Remote Sensing*, 7, 12242-12264. https://doi.org/10.3390/rs70912242

Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J., Goetz, A.F.H., 1993. The spectral image processing system (SIPS) – interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44, pp. 145-163. https://doi.org/10.1016/0034-4257(93)90013-N

Le Bissonnais, Y., Blavet, D., De Noni, G., Laurent, J.-Y., Asseline, J., Chenu, C., 2007. Erodibility of Mediterranean Vineyard Soils: Relevant Aggregate Stability Methods and Significant Soil Variables. *European Journal of Soil Science*, 58, pp. 188–195. https://doi.org/10.1111/j.1365-2389.2006.00823.x

Le Bissonnais, Y., Prieto, I., Roumet, C., Nespoulous, J., Metayer, J., Huon, S., Villatoro, M., Stokes, A., 2018. Soil Aggregate Stability in Mediterranean and Tropical Agro-Ecosystems: Effect of Plant Roots and Soil Characteristics. *Plant and Soil*, 424, pp. 303–317. https://doi.org/10.1007/s11104-017-3423-6

Li, Y., Chen, J., Ma, Q., Zhang, H.K., Liu, J., 2018. Evaluation of Sentinel-2A Surface Reflectance Derived Using Sen2Cor in North America, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11 (6), pp. 1997-2021. doi: 10.1109/JSTARS.2018.2835823

Loiseau, T., Chen, S., Mulder, V.L., Román Dobarco, M., Richer-de-Forges, A.C., Lehmann, S., Bourennane, H., Saby, N.P.A., Martin, M.P., Vaudour, E., Gomez, C., Lagacherie, P., Arrouays, D., 2019. Satellite data integration for soil clay content modelling at a national scale, *International Journal of Applied Earth Observation and Geoinformation*, 82, 101905,https://doi.org/10.1016/j.jag.2019.101905.

Lonjou, V., Desjardins, C., Hagolle, O., Petrucci, B., Tremas, T., Dejus, M., Makarau, A., Auer, S., 2016. MACCS-ATCOR joint algorithm (MAJA). Proc. SPIE 10001, Remote Sensing of Clouds and the Atmosphere XXI, 1000107 (19 October 2016). https://doi.org/10.1117/12.2240935

Louis, J., Debaecker, V., Pflug, B., Main-Knorn, M., Bieniarz, J., Mueller-Wilm, U., Cadau, E., Gascon, F., 2016. Sentinel-2 Sen2Cor: L2A Processor for Users. Proceedings Living Planet Symposium 2016, Spacebooks Online, pp. 1-8.

Lu, P., Wang, L., Niu, Z., Li, L., Zhang, W., 2013. Prediction of soil properties using laboratory VIS–NIR spectroscopy and Hyperion imagery. *Journal of Geochemical Exploration*, 132, pp.26-33. https://doi.org/10.1016/j.gexplo.2013.04.003

Luo, C., Wang, Y., Zhang, X., Zhang, W., Liu, H. (2022). Spatial prediction of soil organic matter content using multiyear synthetic images and partitioning algorithms, *Catena*, 211, 106023, https://doi.org/10.1016/j.catena.2022.106023.

Marujo, R. F. B., Fronza, J. G., Soares, A. R., Queiroz, G. R., and Ferreira, K. R., 2021. Evaluating the impact of LaSRC and Sen2Cor atmospheric correction algorithms on LANDSAT-8/OLI and SENTINEL-2/MSI data over aeronet stations in Brazilian Territory, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-3-2021, 271–277, https://doi.org/10.5194/isprs-annals-V-3-2021-271-2021, 2021.

McBratney, A.B., Field, D.J., Koch, A., 2014. The dimensions of soil security. *Geoderma*, 213, pp. 203–213. https://doi.org/10.1016/j.geoderma.2013.08.013

Mevik, B.-H., Wehrens R., 2007. The pls Package: Principal Component and Partial Least Squares Regression in R. *Journal of Statistical Software*. 18, 2. 10.18637/jss.v018.i02

Minasny, B., McBratney, A.B., Bellon-Maurel, V., Roger, J.M., Gobrecht, A., Ferrand, L., Joalland, S., 2011. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. *Geoderma*, 167–168, 118–124. https://doi.org/10.1016/j.geoderma.2011.09.008

Padró, J.-C., Pons, X., Aragonés, D., Díaz-Delgado, R., García, D., Bustamante, J., Pesquer, L., Domingo-Marimon, C., González-Guerrero, Ò., Cristóbal, J., Doktor, D., Lange, M., 2017. Radiometric Correction of Simultaneously Acquired Landsat-7/Landsat-8 and Sentinel-2A Imagery Using Pseudoinvariant Areas (PIA): Contributing to the Landsat Time Series Legacy. *Remote Sensing*, 9, 1319. https://doi.org/10.3390/rs9121319 https://doi.org/10.3390/rs9121319

Piper C.S., 1966. Soil and Plant Analysis. Hans Publisher, Bombay.

R Development Core Team, 2015. The Comprehensive R Archive Network. The R Foundation for Statistical Computing, Wirtschaft Universitat, Vienna, Austria, URL: 77 http://www.r-project.org/.

Robert, M., Thomas, A., Sekhar, M., Badiger, S., Ruiz, L., Willaume, M., Leenhardt, D., Bergez, J.E., 2017. Farm typology in the Berambadi Watershed (India): Farming systems are determined by farm size and access to groundwater. *Water*, 9, 51. https://doi.org/10.3390/w9010051

Rodrigo-Comino, J., López-Vicente, M., Kumar, V., Rodríguez-Seijo, A., Valkó, O., Rojas, C., Pourghasemi, H. R., Salvati, L., Bakr, N., Vaudour, E., Brevik, E. C., Radziemska, M., Pulido, M., Di Prima, S., Dondini, M., de Vries, W., Santos, E. S., Mendonça-Santos, M. de L., Yu, Y., & Panagos, P., 2020. Soil Science Challenges in a New Era: A Transdisciplinary Overview of Relevant Topics. *Air, Soil and Water Research*. https://doi.org/10.1177/1178622120977491

Sekhar, M., Riotte, J., Ruiz, L., Jouquet, P. and Braun, J.J., 2016. Influences of Climate and Agriculture on Water and Biogeochemical Cycles: Kabini Critical Zone Observatory.

Proceeding of the Indian National Science Academy, 82 (3), pp. 833-846. DOI: 10.16943/ptinsa/2016/48488

Shabou, M., Mougenot, B., Chabaane, Z.L., Walter, C., Boulet, G., Aissa, N.B. Zribi, M., 2015. Soil Clay Content Mapping Using a Time Series of Landsat TM Data in Semi Arid Lands. *Remote Sensing*, 7, pp. 6059-6078. https://doi.org/10.3390/rs70506059

Sola, I., García-Martín, A., Sandonís-Pozo, L., Álvarez-Mozos, J., Pérez-Cabello, F., González-Audícana, M., Montorio Llovería, R., 2018a. Assessment of atmospheric correction methods for Sentinel-2 images in Mediterranean landscapes, *International Journal of Applied Earth Observation and Geoinformation*, 73, pp. 63-76. https://doi.org/10.1016/j.jag.2018.05.020

Sola, I., Álvarez-Mozos, J., González-Audícana, M., 2018b. Inter-Comparison of Atmospheric Correction Methods on Sentinel-2 Images Applied to Croplands, IGARSS 2018 - 2018 *IEEE International Geoscience and Remote Sensing Symposium*, pp. 5940-5943. DOI:10.1109/IGARSS.2018.8518890

Stevens, A., van Wesemael, B., Bartholomeus, H., Rosillon, D., Tychon, B., Ben-Dor, E., 2008. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. *Geoderma*, 144(1-2), pp.395-404. https://doi.org/10.1016/j.geoderma.2007.12.009

Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., van Wesemael, B., 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. *Geoderma*, 158:32–45. https://doi.org/10.1016/j.geoderma.2009.11.032

Tarrio, K., Tang, X., Masek, J.G., Claverie, M., Ju, J., Qiu, S., Zhu, Z., Woodcock, C.E., 2020. Comparison of cloud detection algorithms for Sentinel-2 imagery. *Science of Remote Sensing*, 2, 100010. https://doi.org/10.1016/j.srs.2020.100010

Tomer, S.K., Al Bitar, A., Sekhar, M., Zribi, M., Bandyopadhyay, S., Sreelash, K., Sharma, A.K., Corgne, S. and Kerr, Y., 2015. Retrieval and multi-scale validation of soil moisture from multi-temporal SAR data in a semi-arid tropical region. *Remote Sensing*, 7, pp. 8128–8153. https://doi.org/10.3390/rs70608128

United States Geological Survey (USGS), 2016. Product Guide. Provisional Landsat 8 Surface Reflectance Code (LaSRC) Product. Version 4.0, Department of the Interior: Reston, VA, USA, p. 36. Available online: https://landsat.usgs.gov/sites/default/files/documents/lasrc_product_guide.pdf (accessed on 28 October 2017).

Urbina-Salazar, D. Vaudour, E.,Baghdadi, N., Ceschia, E., Richer-de-Forges, A.C., Lehmann, S., Arrouays, D., 2021. Using Sentinel-2 Images for Soil Organic Carbon Content Mapping in Croplands of Southwestern France. The Usefulness of Sentinel-1/2 Derived Moisture Maps and Mismatches between Sentinel Images and Sampling Dates. *Remote Sensing,* 13, 5115, doi:10.3390/rs13245115

Vaudour, E., Gilliot, J.M., Bel, L., Lefevre, J., Chehdi, K., 2016. Regional prediction of soil organic carbon content over temperate croplands using visible near-infrared airborne hyperspectral imagery and synchronous field spectra. *International Journal of applied earth observation and geoinformation*, 49, pp.24-38. https://doi.org/10.3390/rs11182143

Vaudour, E., Gomez, C., Fouad, Y., Lagacherie, P., 2019a. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. *Remote Sensing of Environment*, 223, pp. 21–33. https://doi.org/10.1016/j.rse.2019.01.006

Vaudour, E., Gomez, C., Loiseau, T., Baghdadi, N., Loubet, B., Arrouays, D., Ali, L., Lagacherie, P., 2019b. The Impact of Acquisition Date on the Prediction Performance of Topsoil Organic Carbon from Sentinel-2 for Croplands. *Remote Sensing*, 11, 2143. https://doi.org/10.3390/rs11182143

Vaudour, E., Gomez, C., Lagacherie, P., Loiseau, T., Baghdadi, N., Urbina-Salazar, D., Loubet, B., Arrouays D., 2021. Temporal mosaicking approaches of Sentinel-2 images for extending topsoil organic carbon content mapping in croplands. International Journal of Applied Earth Observation and Geoinformation 96, 102277. https://doi.org/10.1016/j.jag.2020.102277

Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, 185, pp. 46–56. https://doi.org/10.1016/j.rse.2016.04.008

Vermote, E., Roger, J.C., Franch, B., Skakun, S., 2018. LaSRC (Land Surface Reflectance Code): Overview, application and validation using MODIS, VIIRS, LANDSAT and Sentinel 2 data’s, in: IGARSS 2018 *IEEE International Geoscience and Remote Sensing Symposium*, IEEE, Valencia, pp. 8173–8176. DOI:10.1109/IGARSS.2018.8517622

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*, 131, pp. 59–75. https://doi.org/10.1016/j.geoderma.2005.03.007

Wetzel, D.L., 1983. Near-infrared reflectance analysis, Anal. Chem., 55, pp. 1165A-1176A

Žížala, D., Minařík, R., Zádorová, T. 2019. Soil Organic Carbon Mapping Using Multispectral Remote Sensing Data: Prediction Ability of Data with Different Spatial and Spectral Resolutions. *Remote Sensing*, 11, 2947. https://doi.org/10.3390/rs11242947

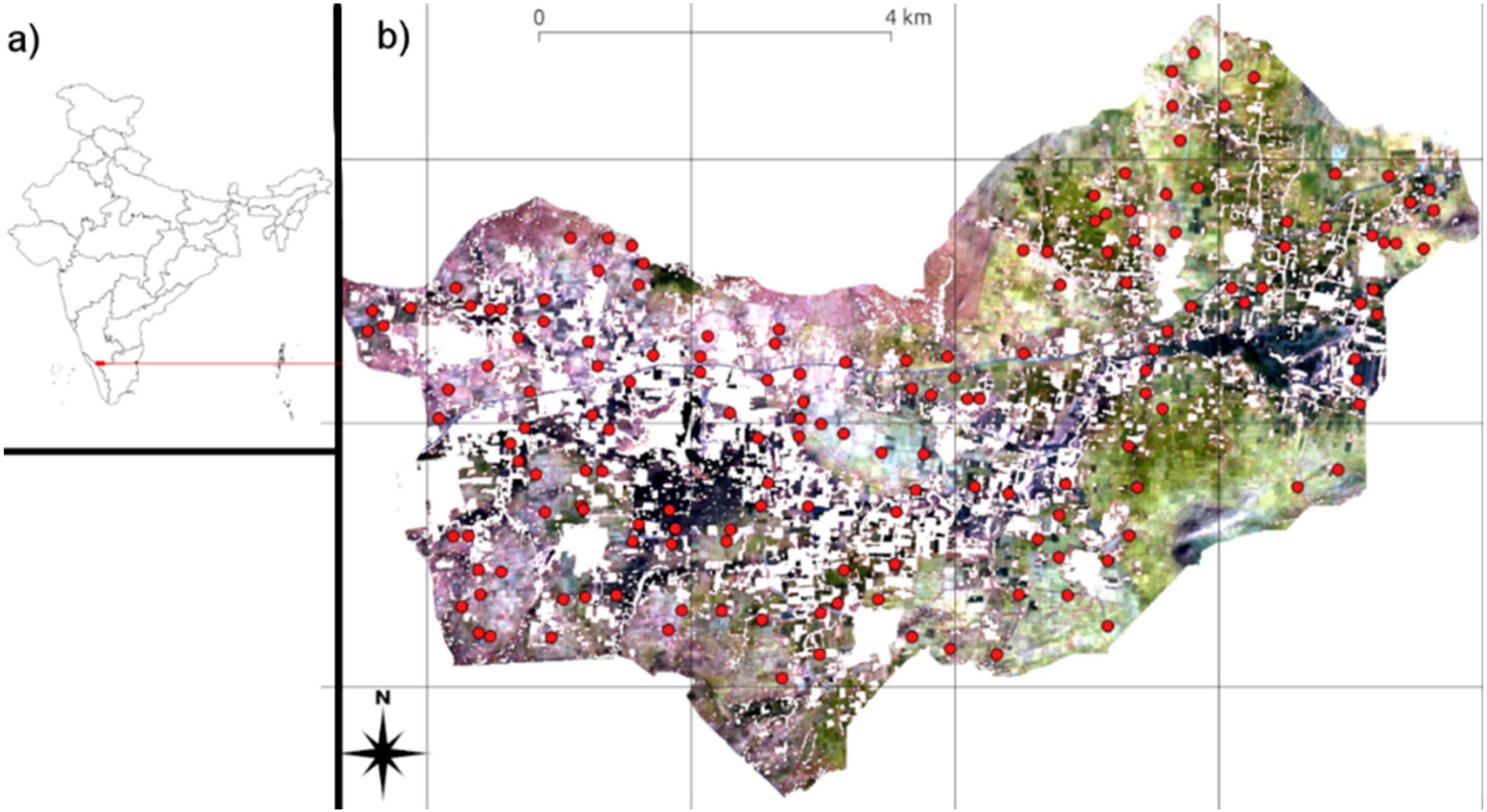

Figure 1: a) Location of the Berambadi catchment in India (red dot) and b) the 164 collected soil surface samples (red dots) plotted over the S2 image acquired on 03-02-2017 (colored composite of bands 08 (R), 04 (G), and 03 (B)). White pixels correspond to masked pixels (non-bare soils).

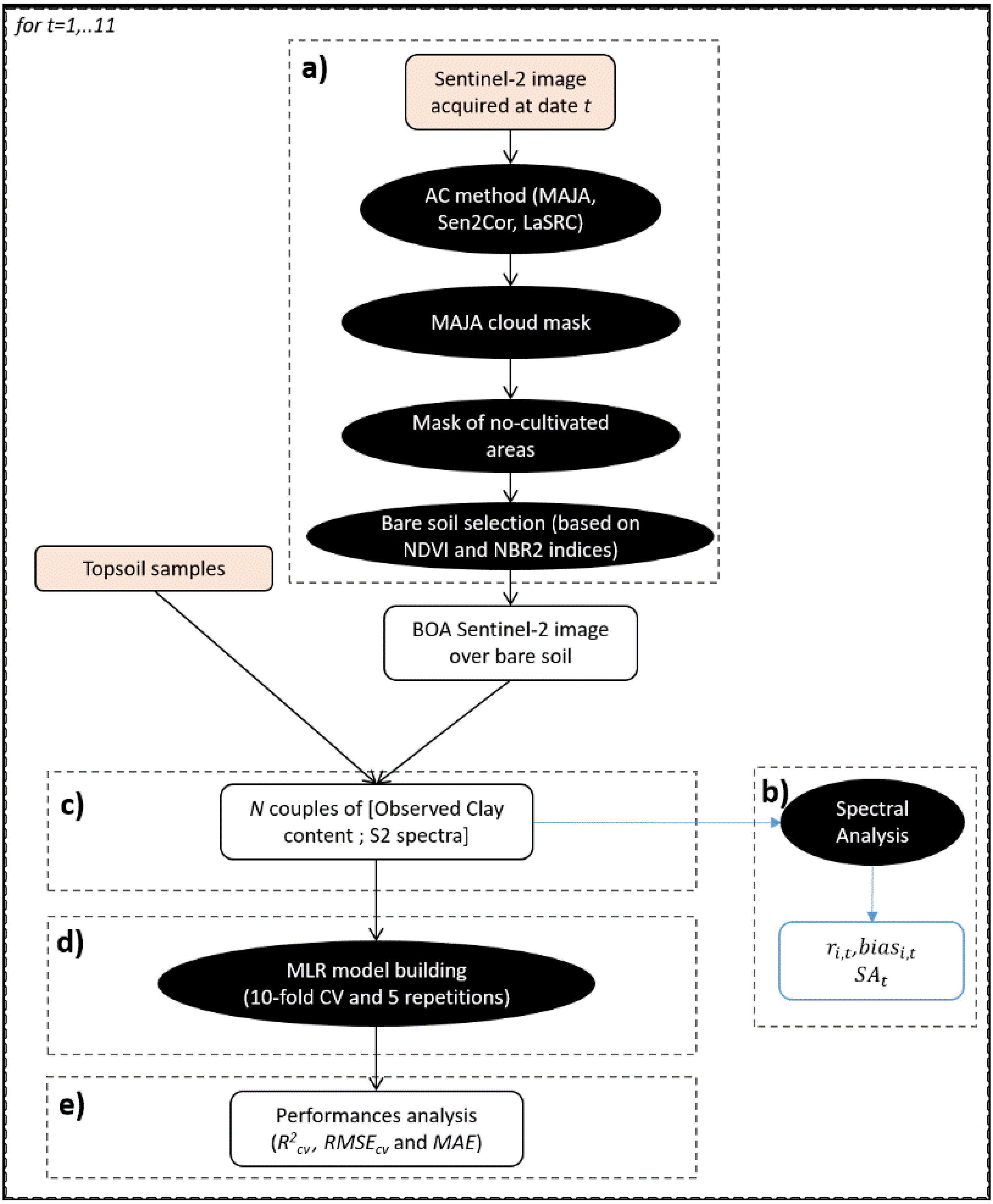

Figure 2: Workflow. a) Preparation of the BOA S2 images restricted to bare soil pixels, b) spectral analysis of BOA images across the time-series and obtained from the 3 AC methods, c) preparation of both calibration and validation datasets for d) model building and e) validation of clay predictions.

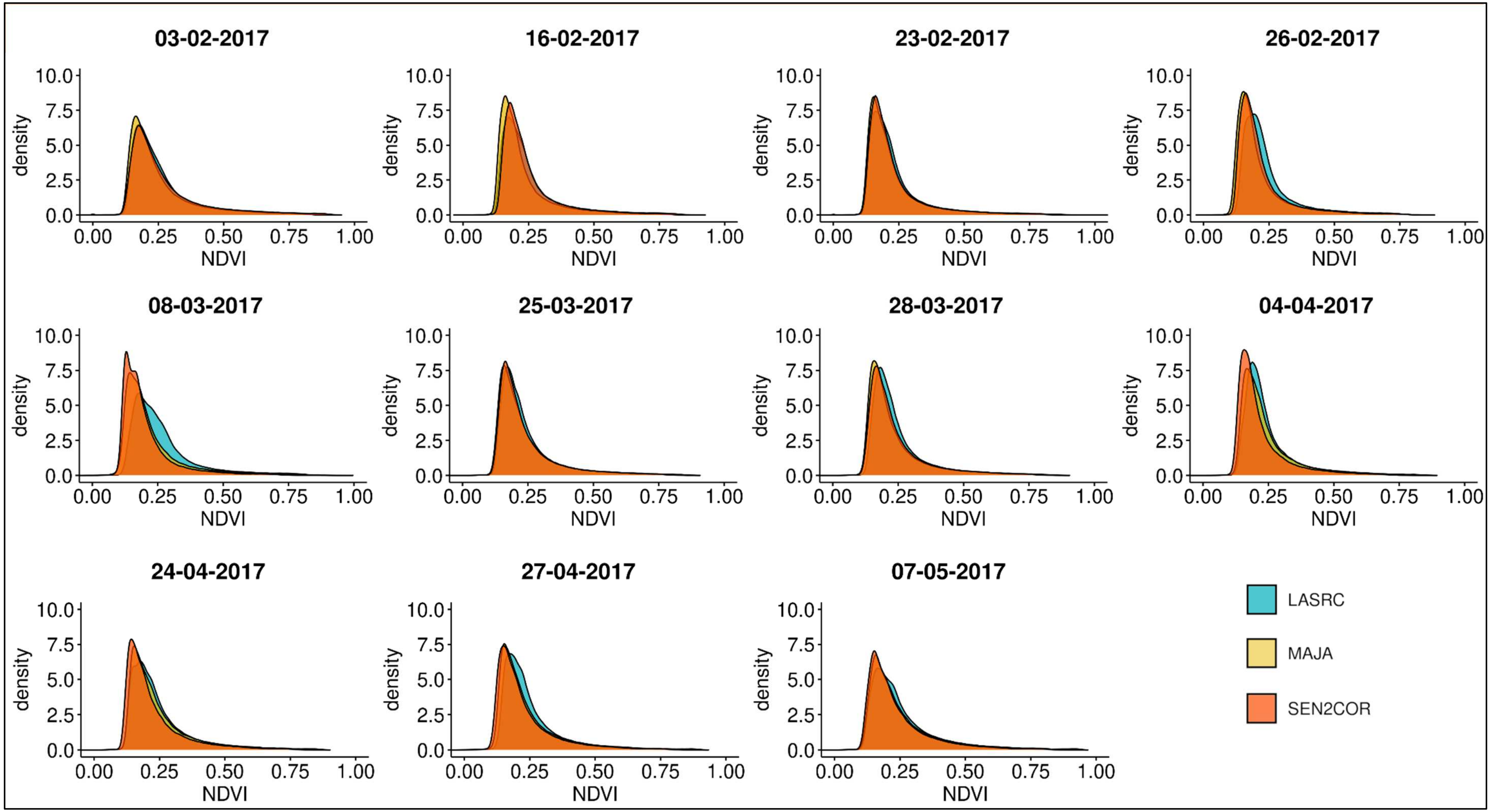

Figure 3: Histograms of NDVI values calculated for each S2 image corrected by the MAJA (yellow), Sen2Cor (orange), and LaSRC (cyan) AC methods.

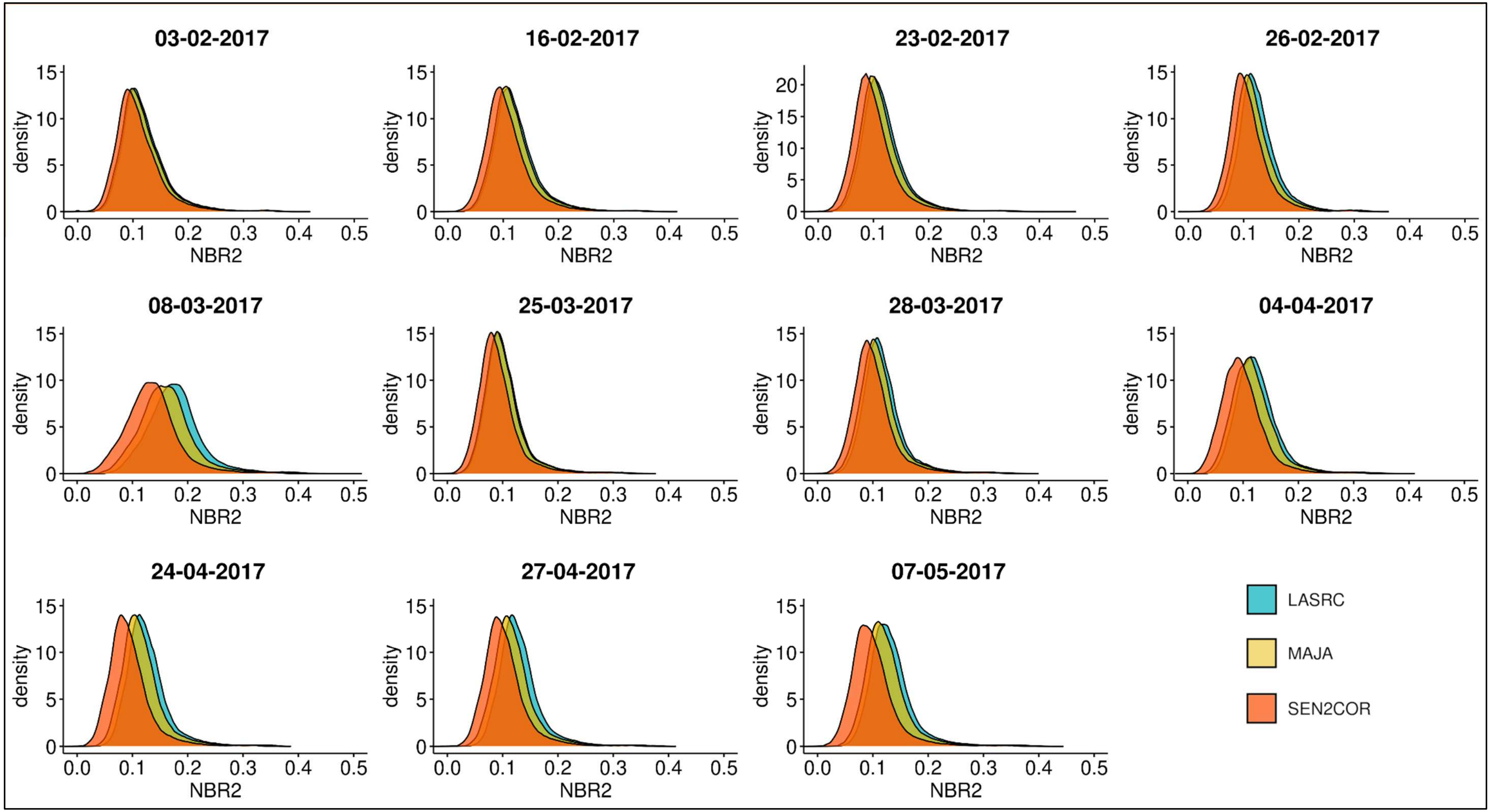

Figure 4: Histograms of NBR2 values calculated for each S2 image corrected by the MAJA (yellow), Sen2Cor (orange), and LaSRC (cyan) AC methods.

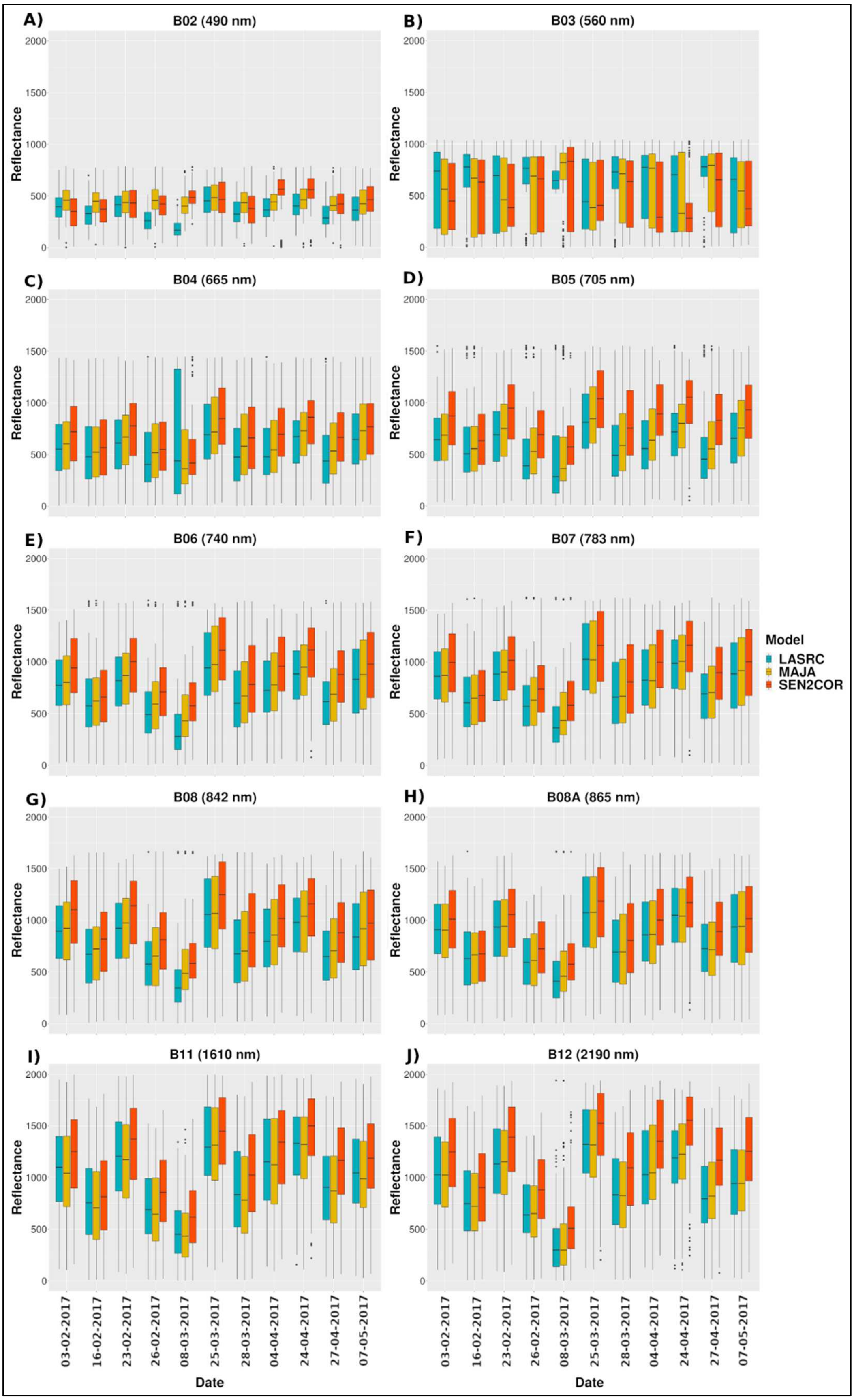

Figure 5: Distribution of BOA reflectance obtained with MAJA (yellow), Sen2Cor (orange) and LaSRC (cyan) at each spectral band and for each date.

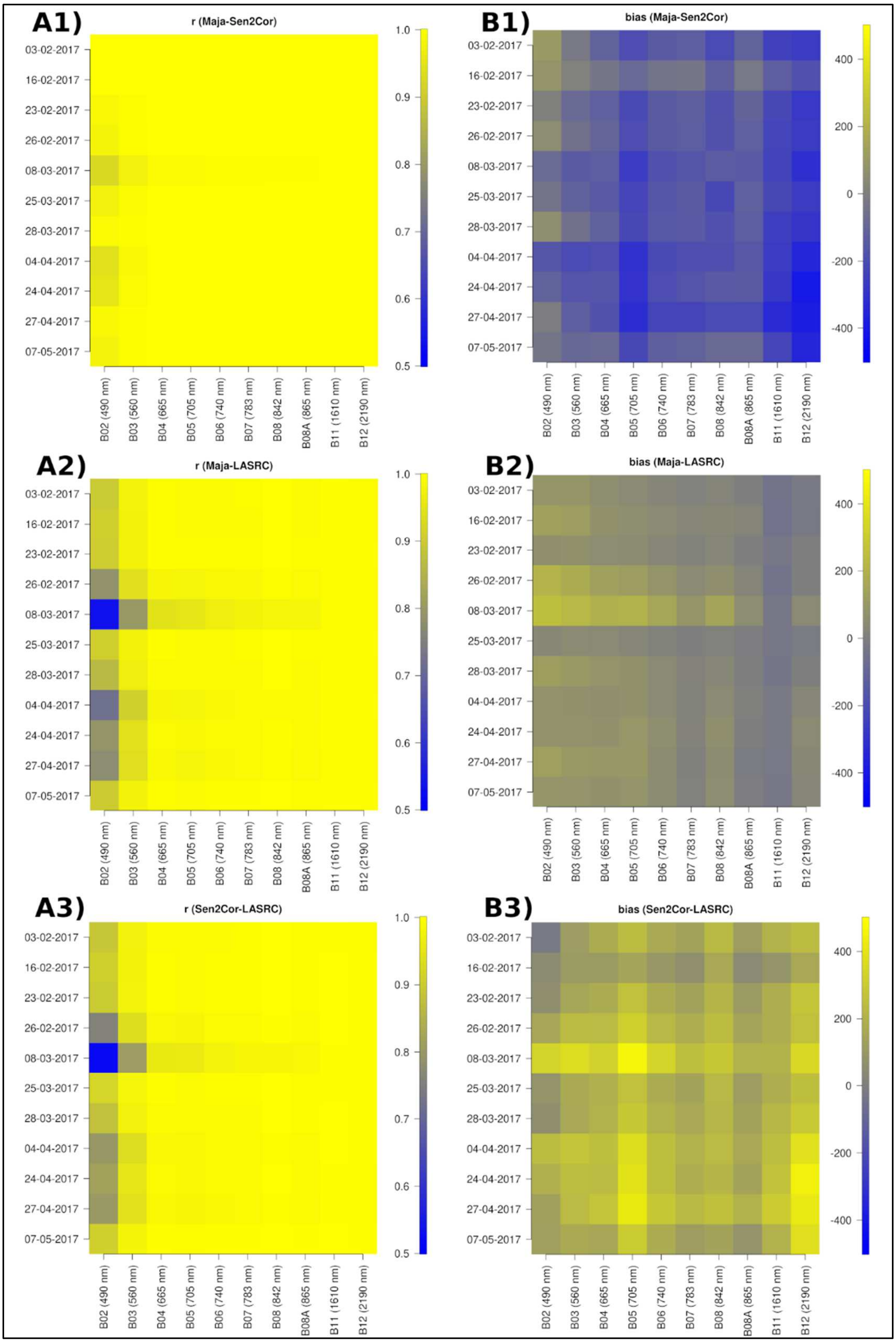

Figure 6: Coefficient of correlation $r_{i,t}$ (A1, A2 and A3) and $Rbias_{i,t}$ (BA, B2 and B3) calculated between the BOA reflectance obtained by MAJA and Sen2Cor (A1 and B1), by MAJA and LaSRC (A2 and B2), and by Sen2Cor and LaSRC (A3 and B3) for the 111 spectra.

Table 1: Proportion of masked cloud surface (%) over each S2 image. Proportion of bare soil (%) over each S2 image obtained by each AC method and the tested combinations of NDVI and NBR2 thresholds. The range (difference between the largest and smallest values of bare soil coverage) was also calculated along both AC methods and S2 dates. Highest ranges of bare soil pixels along both AC methods and S2 dates are highlighted in bold.

		S2 acquisition date (DD-MM-YYYY)											
		03-02-2017	**16-02-2017**	**23-02-2017**	**26-02-2017**	**08-03-2017**	**25-03-2017**	**28-03-2017**	**04-04-2017**	**24-04-2017**	**27-04-2017**	**07-05-2017**	**Range along the dates (%)**
	% of masked cloud surface	0	0	0	0	3	0	0	0.3	0	1.4	0	
NDVI < 0.25	**% of bare soil with MAJA**	66.45	74.5	75.18	78.7	71.78	70.43	72.16	68.42	67.69	68.38	65.03	13.67
	% of bare soil with Sen2Cor	60.48	67.85	74.6	77.42	78.97	70.59	70.28	76.4	73.96	71.97	68.89	**18.49**
	% of bare soil with LaSRC	61.01	66.78	72.74	70.09	58.22	69.26	67.64	67.79	66.3	65.92	62.7	14.52
	Range of bare soil along the 3 AC methods (%)	5.97	7.72	2.44	8.61	**20.75**	1.33	4.52	8.61	7.66	6.05	6.19	
NDVI < 0.3	**% of bare soil with MAJA**	77.9	83.4	83.8	85.9	80.5	80.8	81.8	79.8	79.1	78.6	76.3	9.6
	% of bare soil with Sen2Cor	73.4	79.9	83.5	85.6	85.9	81.1	80.9	85.1	83.3	81.3	79.5	**12.5**
	% of bare soil with LaSRC	75.2	79.1	82.7	82	75.1	80.9	80.2	80.5	78.7	79.2	75.3	7.6
	Range of bare soil along the 3 AC methods (%)	4.5	4.3	1.1	3.9	**10.8**	0.3	1.6	5.3	4.6	2.7	4.2	
NDVI < 0.35	**% of bare soil with MAJA**	84.36	88.44	88.56	90.03	85.37	86.69	87.27	86.28	85.58	84.77	83.22	6.81
	% of bare soil with Sen2Cor	80.97	86.12	88.4	90.15	89.58	87.02	86.77	90.13	88.74	86.9	85.92	**9.18**
	% of bare soil with LaSRC	82.71	85.27	87.99	87.85	83.81	87	86.67	87.29	85.82	86.46	82.93	5.28
	Range of bare soil along the 3 AC methods (%)	3.39	3.17	0.57	2.3	**5.77**	0.33	0.6	3.85	3.16	2.13	2.99	
	% of bare soil with MAJA	23.69	20.73	27.24	17.57	5.11	40.7	26.07	22.37	21.42	18.42	15.34	35.59

NDVI < 0.3 and NBR2 < 0.09	**% of bare soil with Sen2Cor**	33.89	35.86	44.66	32.25	14.81	55.19	41.83	44.34	51.25	38.96	44.38	**40.38**
	% of bare soil with LaSRC	20.22	17.26	21.82	10.42	2.08	37.83	19.17	15.16	11.83	9.65	8.53	35.75
	Range of bare soil along the 3 AC methods (%)	13.67	18.6	22.84	21.83	12.73	17.36	22.66	29.18	**39.42**	29.31	35.85	
NDVI < 0.3 and NBR2 < 0.12	**% of bare soil with MAJA**	58.51	57.64	64.29	57.21	17.21	73.28	64.02	55.51	59.32	54.8	50.48	56.07
	% of bare soil with Sen2Cor	62.82	67.21	73.72	69.59	36.37	77.75	72.56	73.38	77.52	70.72	72.09	41.38
	% of bare soil with LaSRC	55.24	52.96	59.46	47.06	10.15	71.79	57.71	48.09	47.64	42.63	39.64	**61.64**
	Range of bare soil along the 3 AC methods (%)	7.58	14.25	14.26	22.53	26.22	5.96	14.85	25.29	29.88	28.09	**32.45**	
NDVI < 0.3 and NBR2 < 0.15	**% of bare soil with MAJA**	74.78	78.25	80.58	79.76	40.01	80.35	79.75	74.97	77.06	75.21	72.43	40.57
	% of bare soil with Sen2Cor	72.51	78.5	82.55	83.55	63.42	80.87	80.17	83.84	82.75	80.05	78.9	20.42
	% of bare soil with LaSRC	72.13	74.04	78.71	74.72	27.63	80.29	77.53	73.13	73.89	72.1	68.22	**52.66**
	Range of bare soil along the 3 AC methods (%)	2.65	4.46	3.84	8.83	**35.79**	0.58	2.64	10.71	8.86	7.95	10.68	

Table 2: Proportion of common bare soil surface (%) for all S2 images corrected by MAJA. Sen2Cor and LaSRC. Number of collected topsoil samples located over these common bare soil pixels and statistics calculated for each dataset of topsoil samples. These values were calculated for each tested combinations of NDVI and NBR2 thresholds.

		Statistics on collected topsoil samples located over common bare soil pixels					
	% of common bare soil	**Number of samples**	**min (g kg^{-1})**	**max (g kg^{-1})**	**mean (g kg^{-1})**	**Standard deviation (g kg^{-1})**	**skewness (g kg^{-1})**
NDVI < 0.25	40.0%	84	58	592	212	117	8.7
NDVI < 0.3	57.2%	111	58	592	224	120	7.8
NDVI < 0.35	67.3%	122	58	592	228	190	7.0
NDVI < 0.3 and NBR2 < 0.09	1.2%	2	58	64	61	3.9	0
NDVI < 0.3 and NBR2 < 0.12	7%	12	58	449	147	109	18.0
NDVI < 0.3 and NBR2 < 0.15	21%	47	58	592	228	119	7.0

Table 3: Validation performances for the estimations of soil clay content obtained from the MLR models built from MAJA- Sen2Cor- and LaSRC-corrected S2 images. Best performances in terms of both R^2_{cv} and $RMSE_{cv}$ for each date are highlighted in bold.

			S2 acquisition date (DD-MM-YYYY)										
			03-02-2017	**16-02-2017**	**23-02-2017**	**26-02-2017**	**08-03-2017**	**25-03-2017**	**28-03-2017**	**04-04-2017**	**24-04-2017**	**27-04-2017**	**07-05-2017**
NDVI < 0.25	**MAJA**	R^2_{cv}	0.62	0.63	0.62	0.61	0.52	0.74	0.71	0.62	**0.80**	0.73	0.72
		$RMSE_{cv}$	7.77	7.56	7.82	7.66	8.74	6.44	6.94	7.86	**5.68**	6.56	6.47
		MAE	5.55	5.84	5.97	5.72	6.88	5.02	5.45	6.29	4.68	5.23	5.22
	Sen2Cor	R^2_{cv}	0.62	0.63	0.62	0.60	0.50	0.76	0.72	0.62	**0.80**	0.74	0.72
		$RMSE_{cv}$	7.86	7.68	7.77	7.81	8.93	6.34	7.00	7.74	**5.70**	6.52	6.59
		MAE	5.54	5.89	5.88	5.71	7.09	4.94	5.48	6.21	4.70	5.20	5.30
	LaSRC	R^2_{cv}	0.63	0.62	0.62	0.59	0.50	0.75	0.72	0.63	**0.77**	0.72	0.68
		$RMSE_{cv}$	7.61	7.58	7.61	7.91	8.94	6.21	6.84	7.64	**6.16**	6.50	6.99
		MAE	5.39	5.66	5.85	5.99	7.09	4.79	5.47	6.02	5.07	5.35	5.73
NDVI < 0.3	**MAJA**	R^2_{cv}	0.62	0.65	0.64	0.65	0.49	0.71	0.66	0.51	**0.72**	0.68	0.68
		$RMSE_{cv}$	7.57	7.29	7.44	7.34	8.77	6.60	7.19	8.56	**6.50**	6.98	6.89
		MAE	5.67	5.57	5.62	5.41	6.98	5.18	5.68	6.79	5.17	5.53	5.73
	Sen2Cor	R^2_{cv}	0.61	0.64	0.61	0.64	0.45	0.71	0.67	0.48	**0.71**	0.68	0.67
		$RMSE_{cv}$	7.68	7.35	7.54	7.38	9.15	6.63	7.15	8.73	**6.56**	6.97	6.98
		MAE	5.70	5.63	5.66	5.40	7.30	5.20	5.62	6.89	5.24	5.48	5.84
	LaSRC	R^2_{cv}	0.62	0.62	0.62	0.64	0.43	**0.71**	0.66	0.55	0.71	0.68	0.64
		$RMSE_{cv}$	7.43	7.31	7.38	7.44	9.26	**6.38**	7.17	8.20	6.62	6.70	7.10
		MAE	5.50	5.54	5.61	5.51	7.40	4.94	5.68	6.44	5.34	5.39	5.88
NDVI < 0.35	**MAJA**	R^2_{cv}	0.63	0.64	0.61	0.65	0.51	0.74	0.71	0.55	**0.75**	0.70	0.70
		$RMSE_{cv}$	7.47	7.42	7.59	7.42	8.51	6.32	6.72	8.27	**6.14**	6.68	6.79
		MAE	5.59	5.60	5.69	5.59	6.76	4.91	5.28	6.55	4.79	5.22	5.61
	Sen2Cor	R^2_{cv}	0.62	0.63	0.62	0.62	0.47	0.73	0.70	0.51	**0.75**	0.70	0.69

		$RMSE_{cv}$	7.61	7.51	7.55	7.57	8.88	6.34	6.77	8.42	**6.22**	6.68	6.81
		MAE	5.68	5.69	5.72	5.65	7.04	4.91	5.26	6.62	4.87	5.20	5.68
	LaSRC	***R^2_{cv}***	0.60	0.61	0.61	0.63	0.45	**0.74**	0.70	0.58	0.74	0.69	0.66
		$RMSE_{cv}$	7.55	7.42	7.45	7.59	8.97	**6.21**	6.80	7.92	6.30	6.81	7.03
		MAE	5.58	5.54	5.53	5.76	7.11	4.80	5.35	6.24	4.97	5.32	5.75
NDVI < 0.3 and NBR2 < 0.15	**MAJA**	***R^2_{cv}***	0.70	0.73	0.71	0.70	0.58	0.75	0.66	0.62	**0.77**	0.73	0.73
		$RMSE_{cv}$	6.91	7.22	7.06	7.69	9.03	6.98	7.93	8.84	**6.82**	7.48	7.36
		MAE	5.73	6.14	5.94	6.28	7.65	5.80	6.47	7.32	5.60	6.31	6.34
	Sen2Cor	***R^2_{cv}***	0.70	0.73	0.71	0.72	0.54	**0.78**	0.67	0.62	0.77	0.73	0.72
		$RMSE_{cv}$	6.97	7.09	7.26	7.54	9.50	**6.88**	7.77	8.29	7.03	7.63	7.23
		MAE	5.80	5.99	6.24	6.14	7.97	5.71	6.42	6.82	5.75	6.43	6.14
	LaSRC	***R^2_{cv}***	0.73	0.74	0.72	0.71	0.52	**0.78**	0.68	0.67	0.69	0.71	0.74
		$RMSE_{cv}$	5.98	6.77	6.46	7.11	9.53	**6.81**	7.78	8.43	8.41	8.04	7.49
		MAE	4.99	5.71	5.43	5.70	7.86	5.38	6.50	7.17	7.11	6.80	6.37