Sensitivity of clay content prediction to spectral configuration of VNIR/SWIR imaging data, from multispectral to hyperspectral scenarios

Cécile Gomez, Karine R.M. Adeline, Sinan Bacha, Bart Driessen, Nathalie Gorretta, Philippe Lagacherie, Jean-Michel Roger, Xavier Briottet

HAL Id: hal-02118865
https://hal.science/hal-02118865

Submitted on 17 Nov 2021

Sensitivity of clay content prediction to spectral configuration of VNIR/SWIR imaging data, from multispectral to hyperspectral scenarios

C. Gomez[a,*], K. Adeline[a,e], S. Bacha[b], B. Driessen[a], N. Gorretta[c], P. Lagacherie[d], J.M. Roger[c], X. Briottet[e]

[a]IRD, UMR LISAH (INRA-IRD-SupAgro), F-34060 Montpellier, France
[b]CNCT, Tunis, Tunisia
[c]IRSTEA, UMR ITAP, BP 5095, 34196 Montpellier Cedex 5, France
[d]INRA, UMR LISAH (INRA-IRD-SupAgro), F-34060 Montpellier, France
[e]ONERA, The French Aerospace Lab, 31055 Toulouse, France

The use of digital soil mapping, with the help of spectroscopic data, provides a non-destructive and cost-efficient alternative to soil property laboratory measurements. Visible, near-infrared and short wave infrared (VNIR/SWIR, 400–2500 nm) hyperspectral imaging is one of the most promising tools for topsoil property mapping. The aim of this study was to test the sensitivity of soil property prediction results to coarsening image spectral resolution. This may offer an analysis of the potential of forthcoming hyperspectral satellite sensors, e.g., HYPerspectral X IMagery (HYPXIM) or Environmental Mapping and Analysis Program (EnMAP), and existing multispectral sensors, e.g., SENTINEL-2 Multispectral Sensor Instrument (MSI) or LANDSAT-8 Operational Land Imager (OLI), for soil properties mapping. This study used VNIR/SWIR hyperspectral airborne data acquired by the AISA-DUAL sensor (initial spectral and spatial resolutions of approximately 5 nm and 5 m, respectively) over a 300 km^2 Mediterranean rural region. Ten spectral configurations were built and divided in the following two groups: i) six spectral configurations corresponding to simulated sensors with regular spectral resolution from 5 nm to 200 nm (i.e., the Full Width at Half Maximum (FWHM) remains constant throughout the considered spectral domain; this includes the simulation of the forthcoming HYPXIM and EnMAP hyperspectral satellites) and ii) four spectral configurations corresponding to existing multispectral sensors with irregular spectral re-solution (i.e., the FWHM differs from spectral sampling interval; Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), SENTINEL-2 MSI, LANDSAT-7 Enhanced Thematic Mapper (ETM+) and LANDSAT-8 OLI). The soil property studied in this paper is the clay content, defined as the percentage of granulometric fraction finer than 2 μm by weight of the soil, which will be estimated using the partial least squares regression method. Our results showed that i) spectral configurations with regular spectral resolutions from 5 to 100 nm provided similar and good clay content prediction performances (R^2_{val} > 0.7 and RPIQ > 3) and allowed clay mapping with correct short-scale variations, ii) the spectral configuration with a regular spectral resolution of 200 nm provided unsatisfactory clay content prediction performance (R^2_{val} ≃ 0.01 and RPIQ ≃ 1.65) and iii) the ASTER sensor was the only existing multispectral sensor that provided both correct performance of clay content estimation (R^2_{val} ≃ 0.8 and RPIQ ≃ 3.72) and correct clay mapping. Therefore, clay mapping by the ASTER multispectral data should be pursued while awaiting the launch of forthcoming hyperspectral satellite sensors (e.g., HYPXIM and EnMAP), which will be good candidates for future large clay mapping campaigns over bare soils.

1. Introduction

Soil provides major services such as provisions of food, fiber, carbon sequestration, water purification and storage, soil contaminant reduction, climate regulation, nutrient cycling, biological habitats and gene pools. However, demographic pressure and climate change impact these key environmental functions which must be monitored, explored and studied. Many models and indicators that represent these functions are now available (Sanchez et al., 2009). To be fully operational for assisting decisions at local, national and global levels, precise spatially

referenced soil information is required as input in these models and indicators. To address this situation, hyperspectral visible, near-infrared and short-wave infrared (VNIR/SWIR, 400–2500 nm, with > 100 spectral bands) imagery can be considered as an adequate technology for accurate mapping and monitoring of some key soil surface properties (e.g., Ben-Dor et al., 2002; Selige et al., 2006; Stevens et al., 2010; Gomez et al., 2012a). Accurate local estimates were obtained by hyperspectral VNIR/SWIR imagery over bare soil surfaces for soil properties: i) related to a chemical component that impacts soil surface reflectance through absorption bands (e.g., OH^- ions for clay) or ii) highly correlated with the latter (e.g., Cation Exchange Capacity when it is correlated with, for example, clay content) (Ben-Dor et al., 2002). Moreover, recent studies showed that to be predictable, the soil properties have also to have a quite high amount of variability over the study area (Gomez et al., 2012a, 2012b). Nevertheless hyperspectral VNIR/SWIR imagery cannot be extended to large surface mapping or to temporal monitoring because of the expensive cost and the low availability of hyperspectral VNIR/SWIR imaging data.

Only one hyperspectral VNIR/SWIR satellite sensor exists, which is the HYPERION sensor with a spatial resolution of 30 m, a spectral resolution of 10 nm and a swath of 7.5 km (Folkman et al., 2001). Other existing hyperspectral VNIR/SWIR imaging sensors are airborne sensors, such as the HyMap, AISA-DUAL, Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and HySpex sensors, with spectral resolutions between 5 and 10 nm, spatial resolutions of approximately 2 to 5 m (depending on the flight altitude) and flight prints generally inferior of 400 m^2 (depending on the study case). And at least five hyperspectral VNIR/SWIR satellite sensors are planned to be launched next few years (Table 1).

In addition to the hyperspectral imaging sensors, two others categories of VNIR/SWIR imaging sensors exist: multispectral (< 10 bands) and superspectral (10 < bands < 100). Several VNIR/SWIR multispectral satellite sensors are available, such as the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) (Abrams and Hook, 2003), LANDSAT-7 Enhanced Thematic Mapper (ETM +) and LANDSAT-8 Operational Land Imager (OLI) sensors, which were launched in 1999 and 2013, respectively (Masek et al., 2001; Roy et al., 2014). The World-View 3 (Kruse and Perry, 2013) and SENTINEL-2 Multispectral sensor Instrument (MSI) (Baillarin et al., 2012) sensors are both VNIR/SWIR superspectral satellite sensors, which were launched in 2014 and 2015, respectively.

In advance of a diversity of emerging and existing VNIR/SWIR satellite sensors and to confront the lack of soil maps around the world, the potential of VNIR/SWIR satellite sensors for soil properties mapping must be studied. The effect of spectral resolutions effects on minerals and plants identification were studied by Swayze et al. (2003), by simulations of four imaging spectrometers (including AVIRIS sensor). Van Der Meer et al. (2014) demonstrated the relevance of using bands ratios based on simulated SENTINEL-2 MSI data for ferric iron, ferrous iron, laterite, gossan, ferrous silicate and ferric oxides mapping. The effect of coarsening spatial resolution (from 5 m to 60 m) on the accuracy of clay content (defined as the percentage of granulometric fraction inferior to 2 μm by weight of the soil, Baize and Jabiol, 1995) prediction models was studied by Gomez et al. (2015). They found that, up to a spatial resolution of 30 m, clay mapping was still possible, but beyond a spatial resolution of 15 m, clay content variations due to short-scale successions of parent materials were not precisely captured. In addition, spatial resolutions of 60 m or coarser were not suitable for clay content mapping over areas characterized by small short-scale clay content variability and small field sizes. The effect of coarsening spectral resolution (from 1 nm to 200 nm) on the accuracy of soil properties prediction models was studied only from laboratory spectral databases. Castaldi et al. (2016) conducted a study using several laboratory spectral databases to compare the performances of soil texture and soil organic content estimation from present (EO-1 ALI and Hyperion, LANDSAT-8 OLI, SENTINEL-2 MSI) and forthcoming (EnMAP, PRISMA and HyspIRI) multi and hyperspectral sensors. Adeline et al. (2017) used a laboratory spectral database to compare estimation performances of four soil properties (with different spectral absorption features due to their various physico-chemical interactions with soil substrates), clay content, free iron oxides, calcium carbonate and pH, from seven spectral configurations (number of spectral bands decreasing from 328 to 10 and coarsening spectral resolution from 3 nm to 200 nm). Concerning the clay content, Castaldi et al. (2016) and Adeline et al. (2017) demonstrated that coarsening spectral resolution on lab spectra induces a small decrease in prediction model performance, as this soil property has large and pronounced spectral features. Clay content is often used as a tested soil property since its estimation by VNIR/SWIR spectroscopy is driven by both:

- An absorption band centered around 2200 nm, as clay granulometric fractions is correlated to clay minerals which induce an absorption band centered around 2200 nm due to the combination of vibrations associated with the OH bond and the OH−Al−OH bonds (e.g., Hunt et al., 1971; Chabrillat et al., 2002; Kariuki et al., 2004),
- And the general shape of the spectrum as the particle size influences both spectral intensity and absorption bands depth (Baumgardner et al., 1985; Ben-Dor and Banin, 1995). At fine particle sizes, surface scattering dominates so albedo is high and the expression of absorption features worsens as path length (transmission through particles) in minerals is short. And the more the grain size increases, the more the surface to volume ratio decreases, so albedo decreases and absorption begins to dominate as path length increases in minerals. Thus, spectrum with high content of clay fraction will tend to have higher albedo than spectrum of sandy or loamy soil. Finally, the particle size usually doesn't affect the absorption bands position (Ben-Dor and Banin, 1995).

The present study complements both previous works (Adeline et al., 2017 and Castaldi et al., 2016) by assessing the effect of spectral resolution on clay topsoil property estimation. We simulated both artificial sensors (characterized by regular spectral resolutions) and existing multispectral sensors (characterized by irregular spectral resolutions). All these sensor simulations were based on real airborne hyperspectral VNIR/SWIR data acquired over landscapes at a 5-m spatial resolution (AISA-DUAL hyperspectral sensor). The simulation of sensors allowed us to assess the influence of the spectral resolution on estimated soil property, independently to others specifications (e.g. spatial resolutions, acquisition dates, signal to noise ratio).

Table 1
Characteristics of planned hyperspectral VNIR/SWIR satellite sensors.

| Sensor name	Sensor acronym	Nationality	Spatial resolution	Spectral resolution	References
PRecursore IperSpettrale della Missione Applicativa	PRISMA	Italian	30 m	10 nm	Lopinto and Ananasso, 2013
Environmental Mapping and Analysis Program	EnMAP	German	30 m	from 6.5 nm to 10 nm	Guanter et al., 2015 http://www.enmap.org/
HYPerspectral X Imagery	HYPXIM	French	8 m	from 10 nm to 14 nm	Lefèvre-Fonollosa et al., 2016
Spaceborne Hyperspectral Applicative Land and Ocean Mission	SHALOM	Italy-Israel	10 m	10 nm	Ben-Dor et al., 2014
Hyperspectral Infrared Imager	HyspIRI	American	60 m	10 nm	Lee et al., 2015
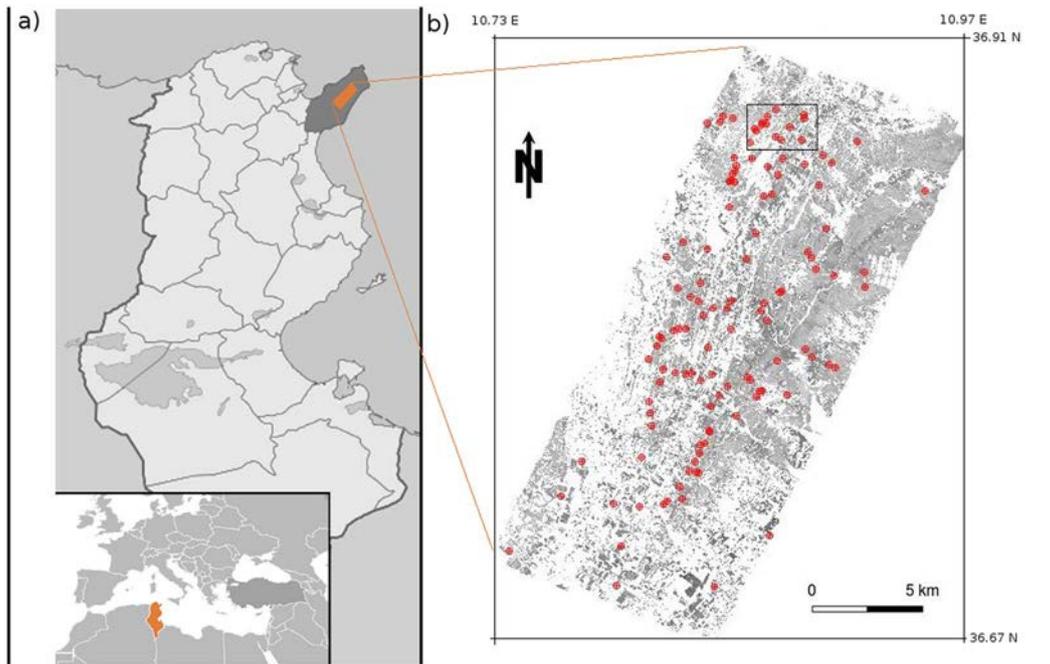
Fig. 1. a) Map of Tunisia and the location of the study area over the Cap Bon region (orange polygon), and b) the AISA-DUAL hyperspectral image at 2206 nm acquired over the Lebna catchment. The Kamech sub-catchment is delimited by the black rectangle. In white are the masked areas (vegetated, urban and water areas). Red targets represent the locations of the 129 collected soil samples. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Material

2.1. Study site

A subarea of the Cap Bon region in northern Tunisia including the catchment named Lebna, was selected as the test area for this study. It covers approximately 300 km^2 as pictured in Fig. 1a. For more details, the reader is referred to Gomez et al., 2012b and Gomez et al., 2015. Hereafter, only its main characteristics are included for a better understanding.

This study site was initially chosen for its variety in soil type and soil distribution patterns due to variability in lithology. The soil pattern is characterized as significant variations in chemical and physical soil properties due to 1) both marl outcrops and Miocene sandstones (Zante et al., 2005) and 2) successive sandstone outcrops for which the distance between them decreases steeply along an East-West direction. The major part of the study site is rural with a cultivation of mainly cereals and legumes, which implies the presence of a large coverage of bare soil during the plowing period (autumn) appropriate for hyperspectral campaign dedicated to soil mapping.

2.2. Hyperspectral airborne data

Airborne VNIR/SWIR hyperspectral data were acquired over the study site by the AISA-DUAL imaging sensor on November 2nd, 2010 over a 24 × 12 km area (Fig. 1b; previously studied in e.g., Gomez et al., 2012b and Gomez et al., 2015). The AISA-DUAL sensor covers the spectral range from 400 nm to 2450 nm with 4.6 nm bandwidth between 400 and 970 nm and with 6.5 nm bandwidth between 970 and 2450 nm, such that the reflected radiance is acquired with 359 spectral bands. The spatial resolution during the acquisition was 5 m and the topographic effects were corrected using a digital elevation model built from ASTER data and some ground control points.

At the same time of the AISA-DUAL data acquisition, reflectance was measured on five uniform surfaces over the flight print (parking lots, asphalt and concrete) using an ASD field spectrometer to apply an atmospheric correction (see Section 3.3).

To perform a fine analysis of the spectral resolution impact on clay content mapping, a subarea of the Lebna catchment, covering 6.67 km^2, was further selected (black rectangle, Fig. 1b) which includes the Kamech sub-catchment. This subarea was selected due to its high contrasting soil patterns and its status of long-term environmental research observatory (called OMERE, Mediterranean observatory of water and rural environment) for anthropogenic impacts study (mainly the agricultural impact) on water and sediment budgets at the catchment scale (e.g., Mekki et al., 2006; Raclot and Albergel, 2006).

2.3. Soil sample database

129 ground soil samples were collected over the Lebna study site (Fig. 1b), in fields that were bare during the hyperspectral data acquisition in November 2010. Among the 129 ground soil samples, 58 were collected in October 2008, 30 were collected in October 2009, and 41 were collected in November 2010. Each soil sample was composed of five soil sub-samples. And these five sub-samples were collected to a depth of 5 cm and within a 10 × 10 m square centered on the geographical position of the sampling plot as recorded by a Garmin GPS instrument.

Each soil sample was homogenized in the laboratory, and approximately 20 g was devoted to the soil property analysis. Each soil sample was air-dried and sieved with a 2 mm sieve prior to being transported to the laboratory for analysis. The clay content (granulometric fraction < 2 μm) was determined using a pipette method (method NF X 31-107, particle size distribution by sedimentation, Baize and Jabiol, 1995). The clay content of the 129 soil samples varied between 46 and 777 g/kg and followed a normal distribution.

3. Method

A schematic overview of the process required to map the soil clay content using coarsening spectral resolution is displayed in Fig. 2. This process was composed of five steps, described in the following sections: application of a spatial mask (Section 3.1), resampling of the spectra to different configurations (Section 3.2), atmospheric correction (Section
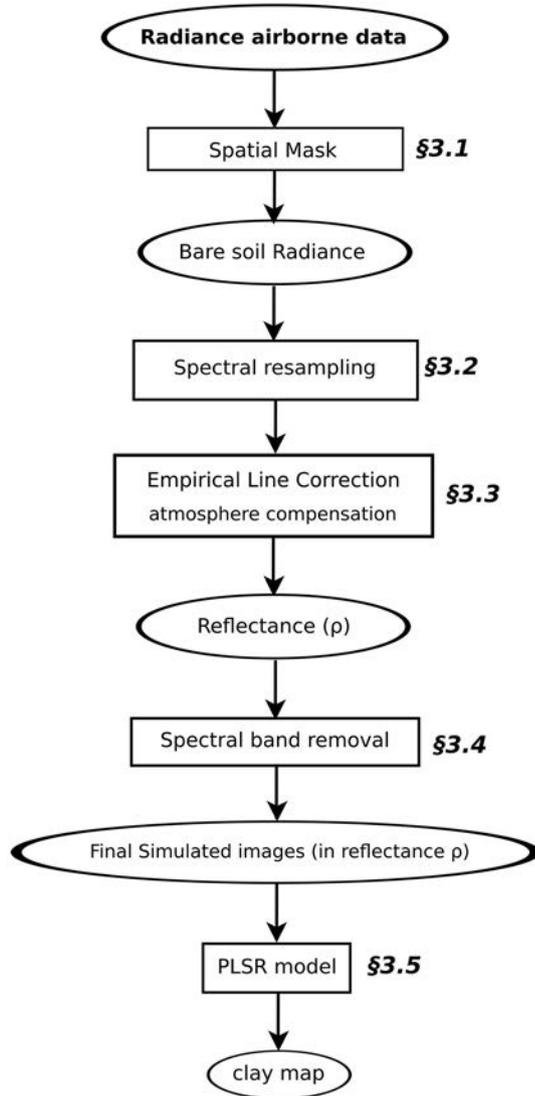
Fig. 2. The conceptual model, showing the process from airborne radiance data to clay prediction maps.

3.3), the band removal (Section 3.4) and the construction of a Partial Least Squares Regression (PLSR) model (Section 3.5). All of the treatments and analyses were implemented in R (Version 1.17).

3.1. Spatial mask

The study site was covered by bare soil, urban activity, water and vegetation consisting mainly of olive trees, native forests, green plants and vineyards. To mask the no-soil pixels over the AISA-DUAL data, the following process was applied, as previously described by Gomez et al. (2015): "To isolate the bare soil areas, pixels with normalized difference vegetation index (NDVI) values over an expert-calibrated threshold were masked. A value of 0.20 was determined after considering twenty parcels, which were visually inspected in the field. The NDVI was retrieved using bands at 0.672 μm and 0.799 μm. Areas of water were also masked using an expert-calibrated threshold. Pixels with a reflectance of less than 8% at 1.665 μm were removed. Finally, 13 urban areas were identified by visual inspection and were also masked." Finally, when the AISA-DUAL data were acquired (on November 2nd, 2010), bare soils covered 46.3% of the entire study site (Fig. 1b) and 49.2% of the Kamech area (black rectangle in Fig. 1b).

3.2. Spectral resampling

The spectral configuration was defined by three parameters: the number of spectral bands (N), the spectral resolution which is also called the Full Width at Half Maximum ($FWHM$) and the spectral sampling interval (SSI) (Swayze et al., 2003). Ten simulations of spectral configurations were built and divided in the following two groups:

- The simulation of six sensors with regular spectral resolution, which corresponds to a constant FWHM throughout the VNIR and SWIR spectral domains. Each simulation with regular spectral resolution will be noted as ~X/Y, where X is the spectral resolution in VNIR and Y in SWIR (Table 2). The spectral configuration with an FWHM of 4.68 nm and 12.58 nm in the VNIR and SWIR, respectively, approximated the EnMAP sensor characteristics (noted as ~5/10). In addition, the spectral configuration with an FWHM of 9.36 nm and 12.56 nm in the VNIR and SWIR, respectively, closely matched the spectral characteristics of HYPXIM, HyspIRI and HYPERION (noted as ~10/10). These six simulations and the AISA-DUAL initial resolution are called the RSR-Sensors group ("Regular Spectral Resolution Sensors") and range from hyperspectral (> 100 spectral bands) to multispectral (< 10 spectral bands) remote sensing configurations.
- The simulation of four existing multispectral sensors with irregular spectral resolution and multispectral configurations: ASTER, SENTINEL-2 MSI, LANDSAT-7 ETM + and LANDSAT-8 OLI (Table 3). These four simulations are called the ISR-Sensors group ("Irregular Spectral Range Sensors") and range from superspectral (10 < spectral bands < 100) to multispectral (< 10 spectral bands) remote sensing configurations.

To build each spectral configuration of the RSR-Sensors group (Table 3), the AISA-DUAL radiance spectra were resampled with Gaussian filters whose tails were cut to twice their width, following the filter response function $G(x)$:

$$G(x) = \exp\left(\frac{-(x - x_0)^2}{2 \cdot \sigma^2}\right) \quad \text{with} \quad \sigma = \frac{FHWM}{2 \cdot \sqrt{2 \cdot \log(2)}} \tag{1}$$

where x is the spectral step determined by the spectral sampling

Table 2
Spectral characteristics of the RSR-Sensors group. N represents the number of initial bands; $FWHM$ (full width at half maximum) is a measure of the spectral resolution; N_{final} represents the number of bands used in the prediction models after the bands were removed (see Section 3.4); N_{VNIR} and N_{SWIR} represent the number of bands used in the prediction models in the VNIR and SWIR spectral domains, respectively.

Configuration	N	$FWHM$ VNIR [nm]	$FWHM$ SWIR [nm]	N_{final} [N_{VNIR}, N_{SWIR}]	(H) Hyperspectral (S) Superspectral (M) Multispectral
AISA-DUAL	359	4.68	6.28	247 [92,152]	H
~5/10 (EnMAP)	240	4.68	12.56	168 [92,76]	H
~10/10 (HYPXIM/HyspIRI/HYPERION)	177	9.36	12.56	123 [47,76]	H
~37/37	53	37.44	37.68	37 [12,25]	S
~60/60	31	60.84	62.8	22 [8,14]	S
~100/100	19	98.28	100.48	13 [5,8]	S
~200/200	9	201.24	200.96	4 [2,2]	M

Table 3

Spectral characteristics of the ISR-Sensors group. N represents the number of initial bands; *FWHM* (full width at half maximum) is a measure of the spectral resolution; N_{final} represents the number of bands used in the prediction models after the bands were removed (see Section 3.4); N_{VNIR} and N_{SWIR} represent the number of bands used in the prediction models in the VNIR and SWIR spectral domains, respectively. Underlines spectral bands correspond to removed spectral bands.

Configuration	N	Central Wavelength (nm)	FWHM (nm)	N_{final} [N_{VNIR}, N_{SWIR}]
SENTINEL-2 MSI	13	443 (Coastal Aerosols)	20	11 [9,2]
		490	65	
		560	35	
		665	30	
		705	15	
		740	15	
		783	20	
		842	115	
		865	20	
		945	20	
		1375 (cirrus)	30	
		1610	90	
		2190	180	
ASTER	9	560	80	9 [3,6]
		660	60	
		810	100	
		1650	100	
		2165	40	
		2205	40	
		2260	50	
		2330	70	
		2395	70	
LANDSAT-8 OLI	8	443 (Coastal Aerosols)	16	6 [4,2]
		482	60	
		561	57	
		655	37	
		865	28	
		1609	84	
		2200	187	
		1363 (cirrus)	20	
LANDSAT-7 ETM +	6	480	65	6 [4,2]
		565	80	
		660	60	
		825	150	
		1650	200	
		2220	260	

interval (*SSI*), x_0 is the mean of the filter and equals the wavelength at which the resampling was performed, and σ is the width of the filter.

To build each spectral configuration of the ISR-Sensors group (Table 3), the AISA-DUAL radiance spectra were resampled using the Relative Spectral Response Filters (RSRF) of each sensor. The RSRF of the SENTINEL-2 MSI, LANDSAT-7 ETM + and -8 OLI, and ASTER sensors were extracted from https://sentinel.esa.int/, http://landsat.usgs.gov/instructions.php, https://asterweb.jpl.nasa.gov/characteristics.asp, respectively.

The 5 field ASD reflectance measurements, collected over uniform surfaces (parking lots, asphalt, and concrete), were also resampled to these ten spectral configurations.

3.3. Atmospheric correction

The AISA-DUAL radiance spectra were converted to reflectance spectra using an empirical line correction method (Farrand et al., 1994) based on the five ASD spectrometer measurements of uniform surfaces (parking lots, asphalt, and concrete) and the darkest pixel of the image associated to a synthetic "zero-value" reflectance. The empirical line correction method hinges on the establishment of a linear relationship between the 6 reflectance measurements (5 field spectrometer measurements and the synthetic "zero-value" reflectance over the darkest pixel) and the 6 corresponding AISA-DUAL radiance measurements (acquired over the uniform surfaces and the darkest pixel) for each

wavelength. The resulting regression equations per wavelength were used to convert the AISA-DUAL and the ten simulated radiances to reflectance.

3.4. Removal of the spectral bands

As previously identified by Gomez et al. (2015), several spectral bands have to be removed due to instrumental noise (in the blue region, from 400 to 484 nm), instrument defect (from 952 nm to 1019 nm), O_2 absorption bands (from747 nm to 766 nm) and H_2O absorption bands (from 1094 nm to 1.176 nm; from 1339 nm to 1465 nm; from 1773 nm to 2005 nm). The final number of bands, N_{final}, for each spectral configuration is presented in Tables 2 and 3.

3.5. PLSR model

The partial least squares regression (PLSR) method was used to establish relationships between the soil clay content and the VNIR/SWIR imaging data (AISA-DUAL and simulated data). The usefulness of the PLSR method comes from its ability to analyze data with many, noisy, and collinear variables in both predictor variables (X-variables, the spectra) and response variables (Y-variables, the clay content in our case) (Wold et al., 1984). The general concept of PLSR is to extract the orthogonal predictor variables (also called latent variables, LV) accounting for the maximum amount of variation of the Y-variables. A detailed description of the PLSR model can be found in Geladi and Kowalski (1986).

Before building PLSR models, the reflectance spectra (X-variables) were first converted into "pseudo absorbance" (log [1/reflectance]) and noise reduction was achieved using a standard normal variate correction to remove additive and multiplicative effects (Barnes et al., 1993). Second, the dataset was divided into two groups. One group corresponding to 3/4 of the dataset is used for PLSR calibration and one group corresponding to 1/4 of the database is used for PLSR validation. The response variables (Y, clay content in our case) are sorted in ascending order. First, the lower clay content is placing in a validation set and the next three samples are placed in the calibration set. Then the procedure is pursued by alternately putting the following sample in the validation set and the next three samples in the calibration set. This subdivision ensures a similar distribution of clay contents in both calibration and validation set.

A leave-one-out cross-validation procedure was adopted to verify the prediction capability of each PLSR model for the calibration dataset (Wold, 1978). Each time, $N - 1$ samples from all N samples within the calibration dataset were used to build the regression model. Based on this model, the value for the soil property of the sample that was not used to develop the model was predicted. This procedure was repeated for all N samples. The optimal number of Latent Variables corresponded to the first local minimum of the root mean square errors of cross-validation (RMSECV) to avoid under-fitting, and do not exceed 10 to avoid over-fitting. The optimal PLSR model was then applied to the validation dataset.

An outlier is defined as a variable (predictor or response variable) that is not consistent with the majority of the data (Chiang et al., 2003; Pearson, 2002). The presence of spectral outliers in the calibration dataset, which are samples spectrally different from other samples, has been studied using the Mahalanobis distance (Mark and Tunnell, 1985). Samples associated to a Mahalanobis distance higher than 3 were considered as spectral outliers and were removed from the calibration dataset.

3.6. Performance of clay prediction and mapping

The prediction performances of the PLSR models were evaluated using: the coefficient of determination R_{cal}^2 (predicted values against measured values in the calibration set), coefficient of determination R_{val}^2

(predicted values against measured values in the validation set), the root mean square errors of calibration (RMSEC), the root mean square errors in the validation set (RMSEP), the bias of prediction, the ratio of the performance to the deviation (RPD, ratio between the standard deviation in the validation set and the RMSEP, Chang et al., 2001) and the ratio of performance to interquartile (RPIQ, ratio of the inter-quartile to the RMSEP, Bellon-Maurel et al., 2010).

The Variable Importance in the Projection (VIP) (Wold et al., 2001; Chong and Jun, 2005) and the PLSR b-coefficients (Haaland and Thomas, 1988) were used to study the significant wavelengths used in PLSR. A wavelength n was considered to have a significant importance if it fulfilled the two following conditions (Wold et al., 2001):

- Its b-coefficient was larger than the standard deviation of the b-coefficients for all the spectral bands (Viscarra Rossel et al., 2008)
- Its variable importance in the projection (VIP) was higher than 1 (Chong and Jun, 2005; Wold et al., 2001), following the equation:

$$VIP_n = N \cdot \sum_{p=1}^{p_{opt}} w_{p,n}^2 \cdot R_p^2 \qquad (2)$$

where $w_{p,n}^2$ is the loading weight for the pth latent variable at wavelength n, p_{opt} is the optimal number of latent variables, R_p^2 is the coefficient of determination obtained for the pth latent variable and N is the number of wavelengths.

Once each PLSR model was built using the calibration dataset corresponding to each spectral configuration, each PLSR model was applied to all of the spectra over bare soil from the corresponding spectral configuration to produce predicted clay content map. To study and compare the ability of the PLSR prediction models to accurately reproduce the spatial structures of the clay contents over the study site, variograms were calculated using semi-variance γ following the formula (Webster and Oliver, 1990):

$$\gamma(h) = \frac{1}{2m} \sum_{i=1}^{m} \{ Y(x_i) - Y(x_i + h) \} \qquad (3)$$

where $Y(x_i)$ and $Y(x_i + h)$ are the observed values of Y at x_i and $x_i + h$, h is the lag (the distance from which the pairs are apart) and m is the number of paired comparisons separated by the lag h (Webster and Oliver, 1990). A variogram measures the spatial dependence of the studied soil property (in our case: clay content). A spatial autocorrelation in clay contents Y (i.e., the sampling locations located close to each other have similar values Y) is highlighted by the variogram when the semi-variance γ is lower at smaller lags than at larger lags.

4. Results

4.1. Preliminary results

A reference PLSR model was built from the AISA-DUAL spectra and observed clay content values associated to the 129 available soil samples. Four spectral outliers were removed from the calibration database of this reference model, and 4 latent variables were selected following the rule of the first local minimum of the RMSECV (Table 4, Fig. 3a). The performance of the model was accurate, with R_{cal}^2 and R_{val}^2 values of 0.77 and an RMSEP value of 82 g/kg (Table 4).

This reference PLSR model was applied to all AISA-DUAL spectra over bare soil to produce a predicted clay content map (Fig. 4a). As previously observed in Gomez et al. (2012b), the Kamech catchment is characterized by strong variations in soil patterns on a small scale, with a close succession of clay-rich areas and clay-poor areas oriented northwest/southeast, corresponding to marl and sandstone outcrops, respectively (Fig. 4a). The predicted clay content values over the Kamech catchment followed a normal distribution (Fig. 3c) and were centered on 503 g/kg with a standard deviation of 157 g/kg and a skewness of − 0.004 g/kg. Finally, the empirical variogram of the

predicted clay content values exhibited spatial structures with clear increases in the semi-variances as the distance between pairs of points increased (Fig. 3d).

4.2. Sensitivity to spectral resolution with regular spectral range

A PLSR model was built for each spectrally resampled VNIR/SWIR image belonging to the RSR-Sensors group, which led to six PLSR models. The number of outliers remained quite stable (between 3 and 7) regardless of the RSR-Sensor. Among the 6 simulated spectral configurations of the RSR-Sensors group, the five RSR-Sensors with spectral resolution from ∼5/10 to ∼100/100 nm offered accurate PLSR models with $R_{cal}^2 > 0.73$, RMSEP < 94 g/kg and $R_{val}^2 > 0.71$ (Table 4). The performance indicators, R_{cal}^2, R_{val}^2 and RMSEP, obtained with these five RSR-Sensors varied only slightly and indicate that a spectral resolution of ∼5/10 to ∼100/100 nm does not strongly affect prediction performances. The slight difference in performance indicator values within this group may be caused by the selected number of latent variables following the rule of the first local minimum of the RMSECV, which is slightly different for each spectral configuration (Table 4 and Fig. 3a). The PLSR model built from the regular spectral range of ∼200/200 nm produced an inaccurate PLSR model, with an $R_{cal}^2 < 0.4$, RMSEP > 130 g/kg and $R_{val}^2 < 0.1$ (Table 4).

The performance indicators R_{val}^2, RMSEP, RPD and RPIQ indicate that the simulated EnMAP spectral resolution (∼5/10 m) provides quite better accuracy than the initial AISA-DUAL spectral resolution. This improvement may be explained by a smoothing of the spectra due to the Gaussian filter used to simulate coarser spectral resolution, which involves a noise reduction (Fig. 5).

None of the spectral bands of the reflectance data at ∼200/200 nm may be considered as significant based on the analysis of the VIP and b-coefficients (Fig. 3b). Concerning the five other spectral configurations, the spectral range between 2000 and 2500 nm included the wavelengths with the highest amount of significance in the PLSR models (highest b-coefficients and VIP) (Fig. 3b). In particular, the spectral range at approximately 2200 nm, corresponding to an absorption feature due to the combination of OH-Al bending and OH stretching modes, was used by all PLSR models (Fig. 3b). The spectral domains at approximately 720 nm and 1490 nm were also used by these five PLSR models (Fig. 3b). Both spectral domains may correspond to the spectral baseline which must appear in the regression models (Roger et al., 2010).

The six PLSR models, built from each spectral configuration, were applied to all spectra over bare soil to produce six predicted clay content maps over the Kamech catchment. All spectral configurations provided predicted clay contents over the Kamech catchment following normal distributions (Fig. 3c). The clay contents predicted using the spectral configuration of ∼200/200 nm exhibited a lower standard deviation than the ones predicted by the five other spectral configurations (Fig. 3c). This can also be shown by the experimental variogram predicted from the spectral configuration of ∼200/200 nm, which exhibited a lower sill (Fig. 4d). This low variability of estimated clay was due to the PLSR model tending to converge towards the mean of the calibration data y, as not enough spectral information is contained in the ∼200/200 nm configuration to predict clay content with accuracy. The five other predicted clay maps exhibited correct short range variations in soil patterns, with a close succession of clay-rich areas and clay-poor areas (Fig. 4b and c), as observed in the reference map (Fig. 4a). Similarly, the experimental variograms look very similar to the one produced with the reference AISA-DUAL data (Fig. 3d).

4.3. Sensitivity to the spectral resolution with irregular spectral range

Among the four spectral configurations of the ISR-Sensors, the spectral configurations of ASTER and SENTINEL-2 MSI seemed to offer accurate PLSR models for clay prediction, corresponding to

Configuration name	R^2_{cal}	RMSEC (g/kg)	R^2_{val}	RMSEP (g/kg)	Bias (g/kg)	RPD	RPIQ	# of latent variables	# of outliers removed
AISA-DUAL	0.77	82	0.77	82	− 20	2.14	3.45	4	4
~5/10 (EnMAP)	0.73	90	0.78	81	− 16	2.18	3.52	4	4
~10/10 (HYPXIM/HyspIRI/HYPERION)	0.75	88	0.73	90	− 15	1.97	3.19	5	7
~37/37	0.74	88	0.74	90	− 8	1.99	3.21	5	3
~60/60	0.75	85	0.71	93	− 12	1.9	3.07	6	4
~100/100	0.75	87	0.71	94	− 11	1.9	3.04	5	3
~200/200	0.35	139	0.01	173	15	1.02	1.65	3	4
SENTINEL-2 MSI	0.7	95	0.71	94	− 19	1.9	3.06	8	3
ASTER	0.69	95	0.8	77	− 4	2.3	3.72	5	2
LANDSAT-8 OLI	0.29	145	0.23	152	8	1.16	1.87	3	3
LANDSAT-7 ETM +	0.27	147	0.09	166	7	1.07	1.72	3	3

$R^2_{cal} > 0.69$, RMSEP of 95 g/kg and $R^2_{val} > 0.71$ (Table 4). Between these two sensors, the PLSR model built from the spectral configuration of ASTER provided a more accurate clay prediction than the PLSR model built from the spectral configurations of SENTINEL-2 MSI, indicated by a lower bias and a higher RPIQ and RPD (Table 4). The PLSR models built from spectral configurations of LANDSAT-7 ETM + and
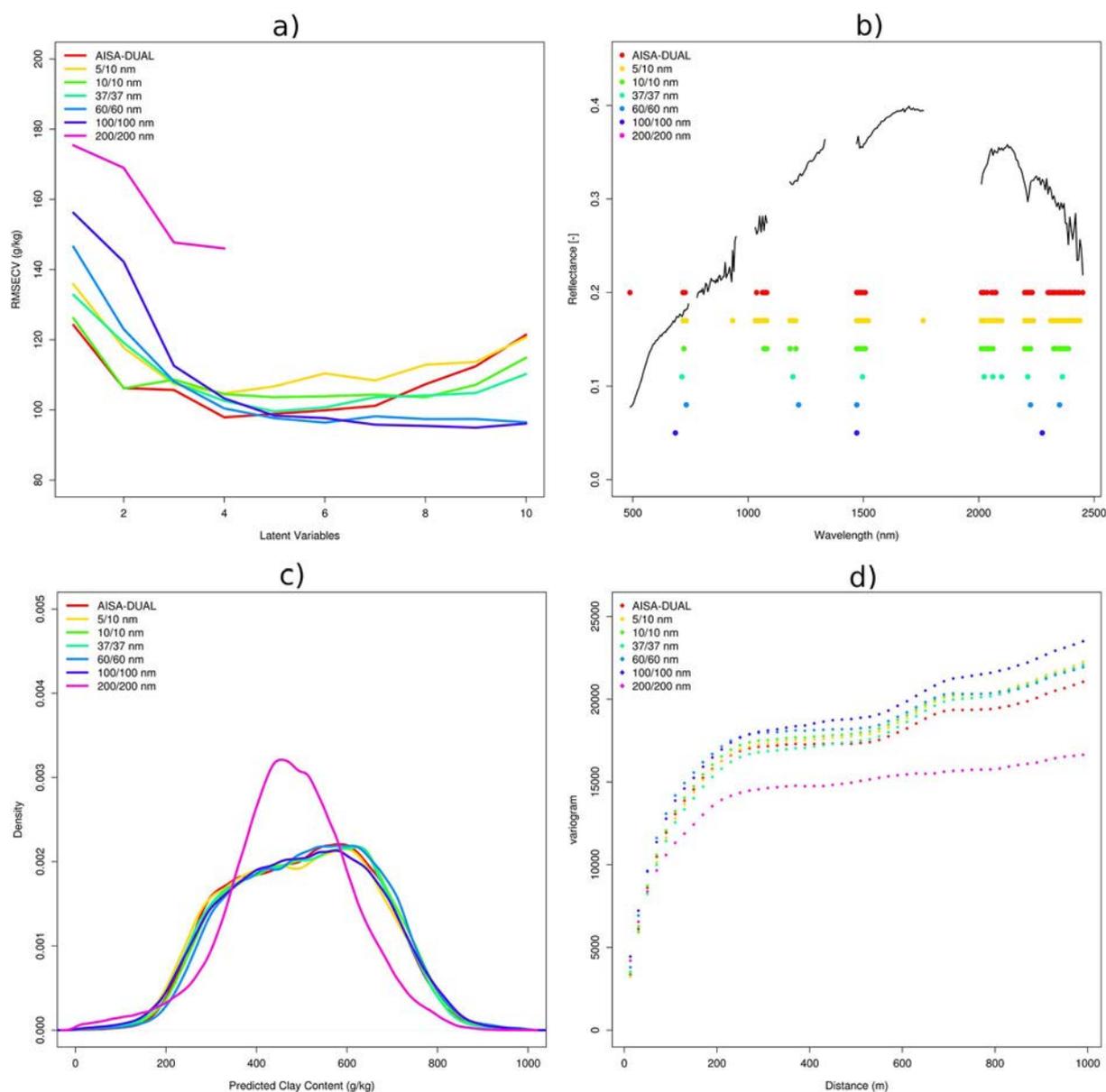

Fig. 3. a) RMSECV as a function of the latent variables retained in the PLSR models, b) significant spectral bands (colored points) used in the PLSR models, c) histograms of predicted clay content and d) empirical variograms with a lag of 20 m (points) of predicted clay content obtained from the six RSR-Sensors and the AISA-DUAL data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
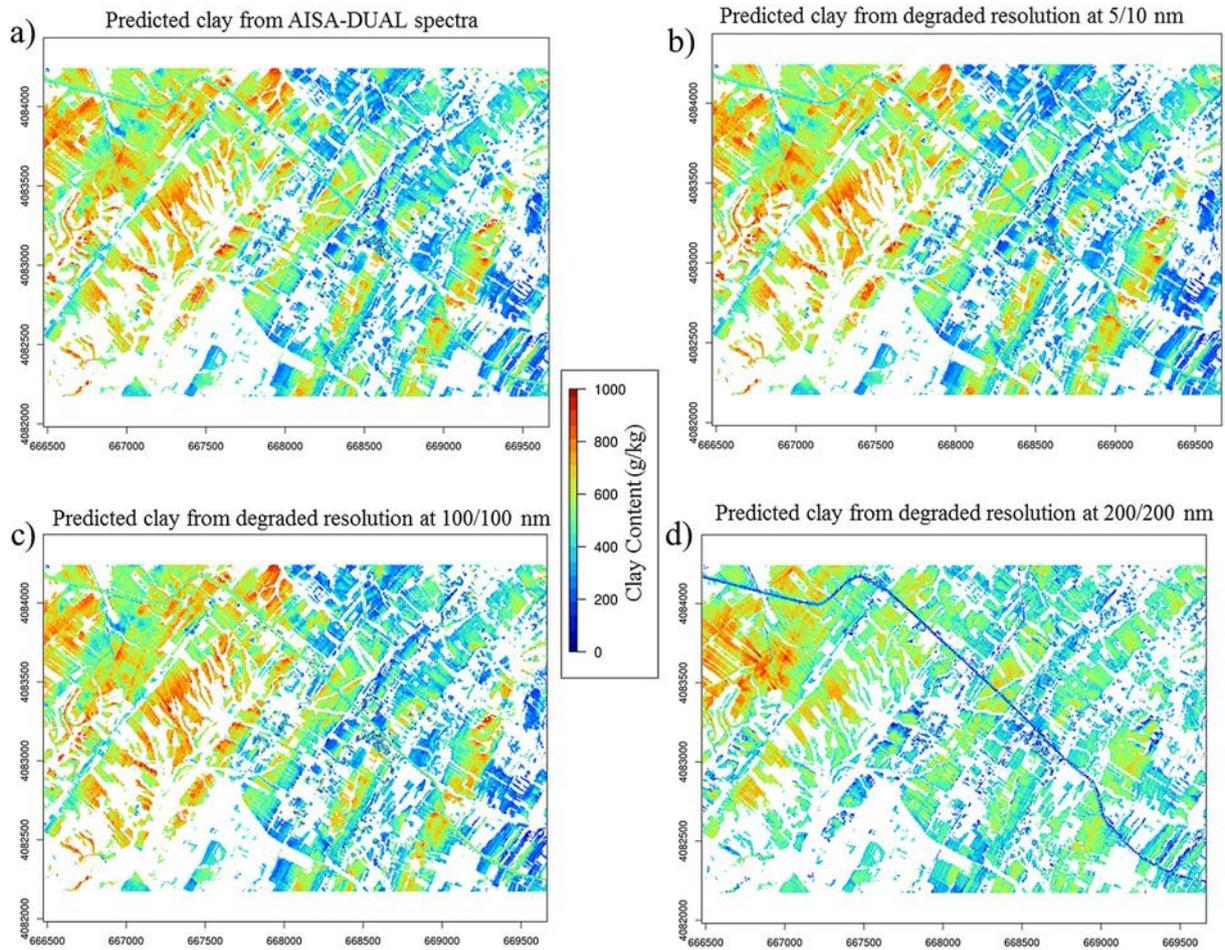
Fig. 4. Maps of the clay contents over the Kamech catchment predicted from the PLSR models using a) AISA-DUAL data and spectral configurations at b) 5/10 nm, c) 100/100 nm and d) 200/200 nm.
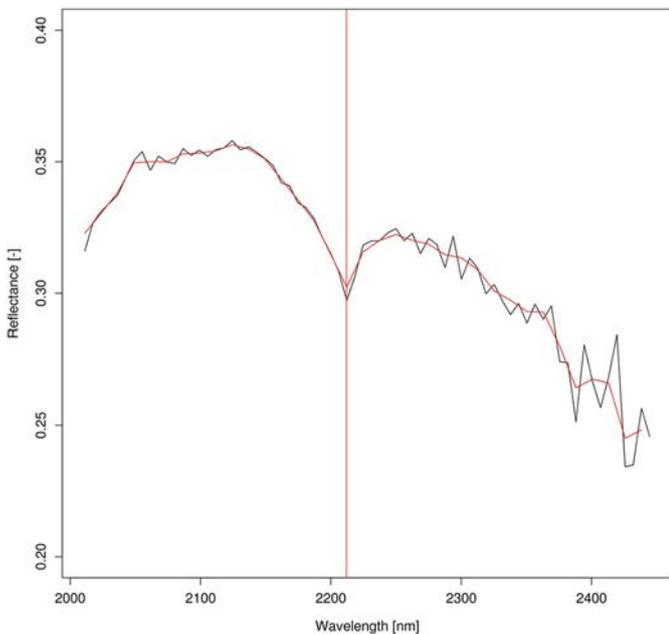

Fig. 5. The black and red spectra represent the AISA-DUAL reflectance of a random pixel and its spectrum resampled at ~5/10 nm (EnMAP configuration), respectively. The vertical red line represents the center of the clay absorption band (at 2210 nm). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

LANDSAT-8 OLI performed the worst, with an $R_{cal}^2 \leq 0.4$, RMSEP > 120 g/kg and $R_{val}^2 < 0.45$ (Table 4). In addition, 2 to 3 spectral outliers were removed from the calibration database for all tested ISR-Sensors (Table 4).

Following the rule of the first local minimum of the RMSECV, 3 to 8 latent variables were selected for these four PLSR models (Fig. 6a). For the SENTINEL-2 MSI spectral resolution, the first local minimum of RMSECV was difficult to detect (Fig. 6a). Moreover, no significant band was identified for the PLSR model built from the spectral configuration of SENTINEL-2 MSI, even though the PLSR model with 8 latent variables is quite accurate in terms of the performance indicators (Table 4, Fig. 6b). The combination of 1) the lack of a clear first local minimum of RMSECV and 2) the lack of any significant band means that the PLSR model built from the SENTINEL-2 MSI spectral configuration is inaccurate for an LV < 8 and overfit the predictions for an LV ≥ 8. Therefore, from these analyses, no PLSR models can be considered as accurate for clay content estimation and mapping using the SENTINEL-2 MSI spectral resolution.

For three of the four spectral configurations of the ISR-Sensors (ASTER, LANDSAT-7 ETM + and LANDSAT-8 OLI), the wavelength located at approximately 2200 nm, corresponding to an absorption feature caused by the combination of OH-Al bending and OH stretching modes, was one of the most significant bands (highest b-coefficients and VIP) (Fig. 6b). The spectral configurations of LANDSAT-7 ETM + and LANDSAT-8 OLI also used a band centered at 1650 and 1609 nm, respectively, which may correspond to the spectral baseline that appears in the regression models (Roger et al., 2010). The visible spectral
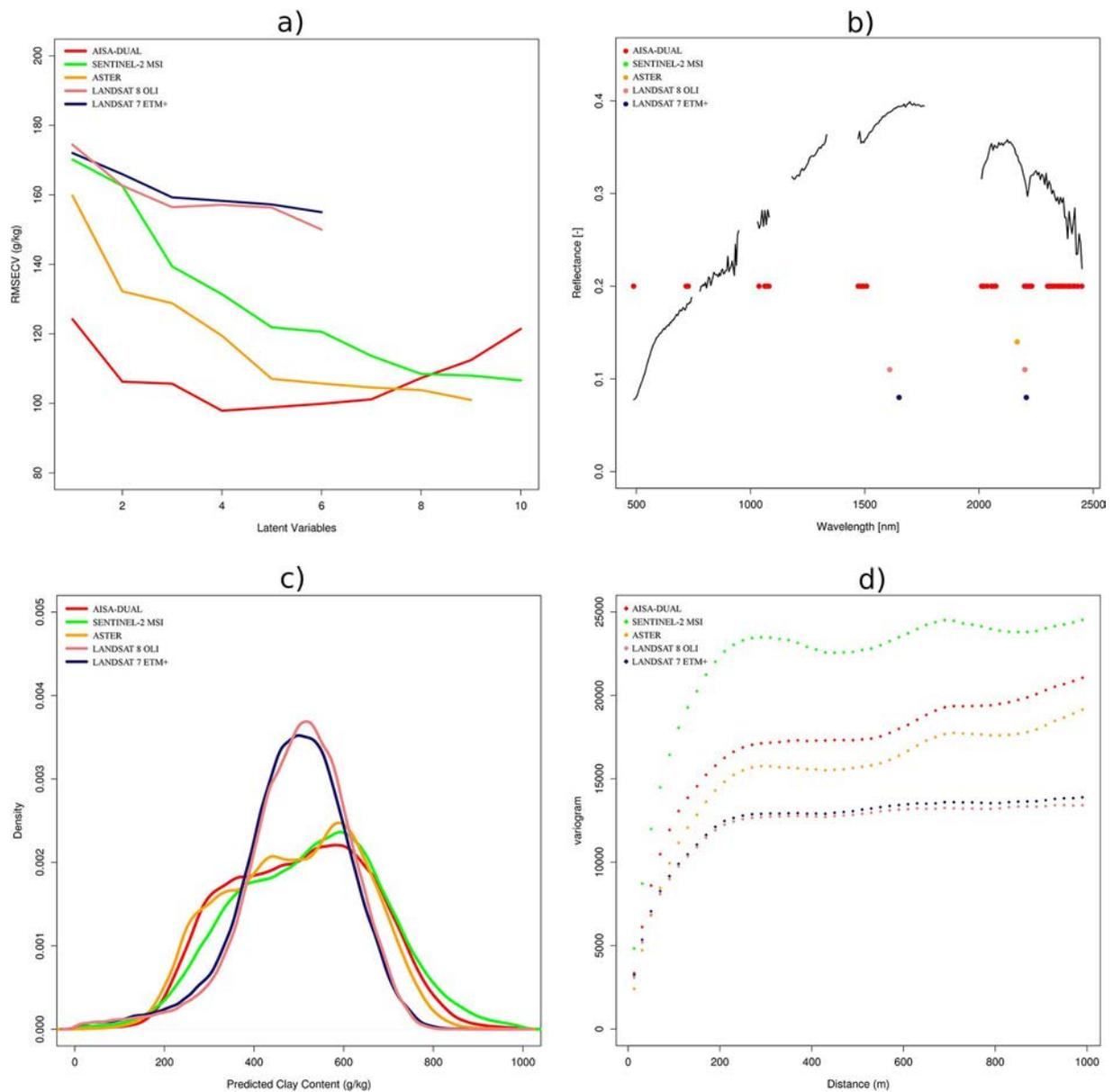
Fig. 6. a) RMSECV as a function of the latent variables retained in the PLSR models, b) significant spectral bands (colored points) used in the PLSR models, c) histograms of the predicted clay content and d) empirical variograms with a lag of 20 m (points) of predicted clay content obtained from the four ISR-Sensors and the AISA-DUAL sensor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

domain (400–1000 nm) was not considered as an important spectral domain by any of these ISR-sensors (Fig. 6b).

The four PLSR models built from spectral configurations of the ISR-Sensors were applied to all spectra over bare soil to produce four predicted clay content maps over the Kamech catchment (Fig. 7). These spectral resolutions provided predicted clay contents over the Kamech catchment that followed a normal distribution (Fig. 6c). The clay contents predicted using the spectral configurations of LANDSAT-7 ETM + and LANDSAT-8 OLI exhibited a lower standard deviation than the two other spectral resolutions (Fig. 6c). This means that the pedological variability of the Kamech catchment is not represented in the predicted clay content maps obtained using both spectral configurations (Fig. 7c and d), such as for the spectral resolution of 200/200 nm. The distribution of clay contents predicted using the spectral resolution of SENTINEL-2 MSI shifted to slightly higher values compared to those obtained from the spectral resolution of ASTER (Fig. 6c). This confirms that the PLSR model using the SENTINEL-2 MSI spectral resolution

cannot be considered as accurate for clay content estimation and mapping.

Finally, the spatial structures of estimated clay contents obtained by LANDSAT-7 ETM + and LANDSAT-8 OLI spectral configurations exhibited very low sills (Fig. 6d). By contrast, the semi-variance obtained by SENTINEL-2 MSI exhibited very high sill and strong variations. Finally, the variogram obtained by the ASTER simulated data was close to the one obtained by the AISA-DUAL data.

5. Discussion

5.1. Clay absorption feature as a driver of spectral configuration impact

The presence of the clay absorption feature centered around 2200 nm in the tested spectral configuration seems to be the main driver of clay content prediction performance. From the literature, the width of the clay absorption feature centered around 2200 nm is
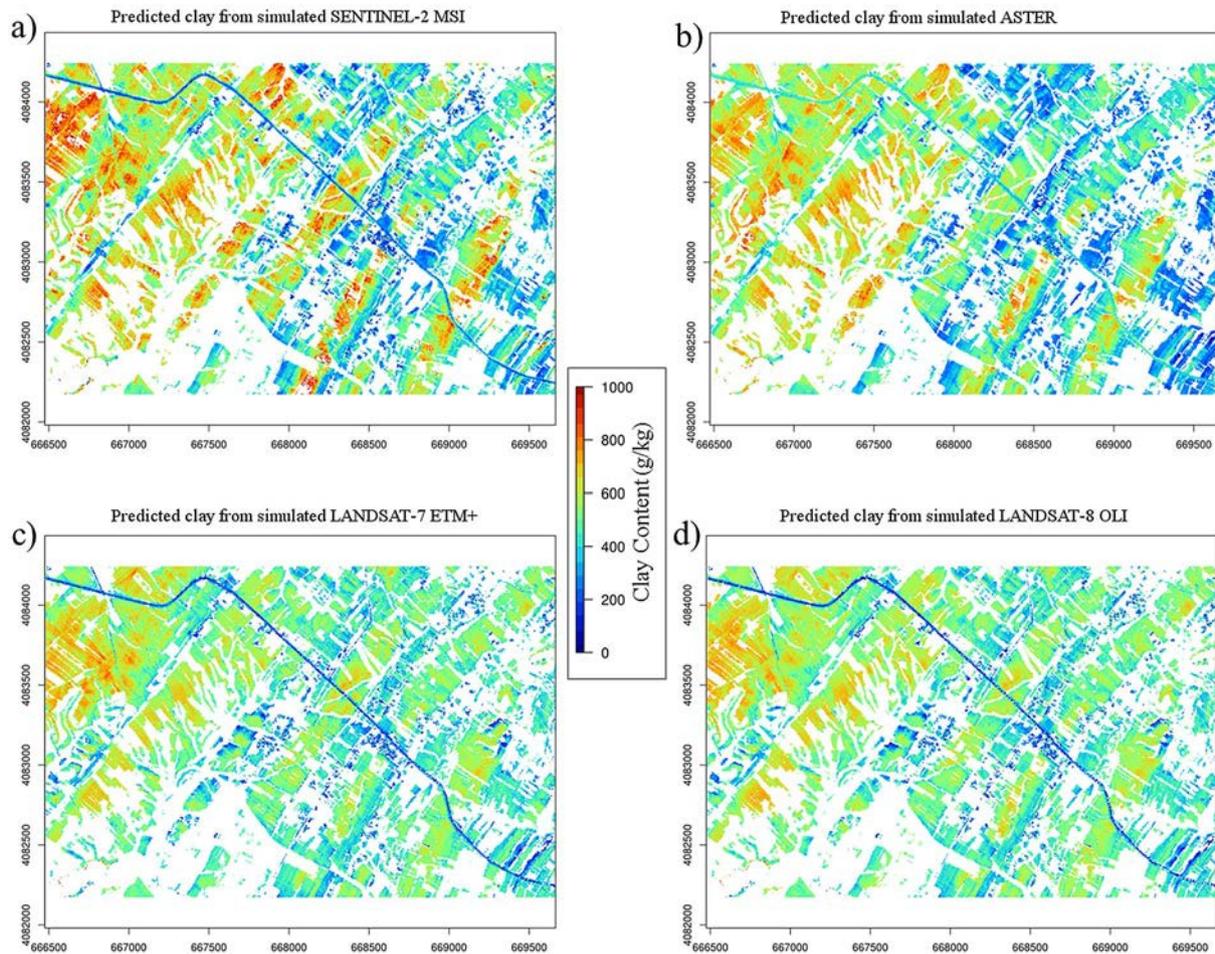
9

Fig. 7. Maps of clay contents over the Kamech catchment predicted from the PLSR models using spectral configurations of a) SENTINEL-2 MSI, b) ASTER, c) LANDSAT-7 ETM +, and d) LANDSAT-8 OLI.

between 92 nm (Levin et al., 2007) and 126 nm (Lagacherie et al., 2008).

5.1.1. Case of regular spectral resolution sensors (including simulated forthcoming hyperspectral satellites)

Correct performances of clay prediction were obtained from the PLSR models built using regular spectral resolution up to ~100/100 nm (Table 4). Weak performances of clay prediction obtained from the PLSR model using regular spectral resolution of ~200/200 nm may be explained by the fact that the clay absorption feature was included in only one spectral band (Fig. 8a), which cannot reflect depth and width of the absorption feature. However, for higher FHWM (regular spectral resolution from ~5/10 nm to 100/100 nm), the clay absorption feature was discriminated using several bands (e.g., using three bands at the spectral resolution of ~100/100 nm, Fig. 8a).

Castaldi et al. (2016) showed that good clay prediction performances are obtained using laboratory spectra with bandwidths from 10 to 160 nm. In addition, Adeline et al. (2017) showed that good clay prediction performances are obtained using laboratory spectra with bandwidths up to 200 nm. The measurement differences between the laboratory spectroscopy and imaging data (light sources, spatial resolution, instrumental noise, atmospheric conditions, purity of the pixels, etc.) may explain the performance differences between our results and those of Castaldi et al. (2016) and Adeline et al. (2017). Imaging spectra may contain perturbations which alter the spectra quality and reduce the performance of prediction when the spectral resolution decreases.

5.1.2. Case of irregular spectral resolution sensors (simulated existing multispectral satellites)

The ASTER simulated sensor is the only multispectral simulated sensor which provides accurate prediction of clay content ($R^2_{val} = 0.8$, Table 4). This accurate performance of clay prediction may be explained by the fact that the clay absorption feature is represented by three ASTER spectral bands at 2165, 2205 and 2260 nm with FWHM < 70 nm (Table 3, Fig. 8b). In addition, the weak performances of clay prediction models built using spectra sampled from SENTINEL-2 MSI and from both LANDSAT spectral resolutions (Table 4) may be explained by the fact that the clay absorption feature is represented by only one spectral band of the SENTINEL-2 MSI sensor and of both LANDSAT sensors (Fig. 8b).

Multispectral remote sensing data have seen very little use in soil science. Many of the multispectral data studies focused on the classification of pedological groups (e.g., Escadafal and Pouget, 1987). Other multispectral data studies tested mapping of quantitative physico-chemical soil properties with modest success. Shabou et al. (2015) achieved a R^2 of 0.73 for clay content estimation using LANDSAT-TM5 data and a spectral index called MID-Infrared index. These results do not match with ours. However, their model was calibrated using only 30 samples, which is low and may produce unrobust model or overfitted model (as is the case of our model obtained from the SENTINEL-2 MSI configuration).

5.2. Extension to other soil properties of continental surfaces

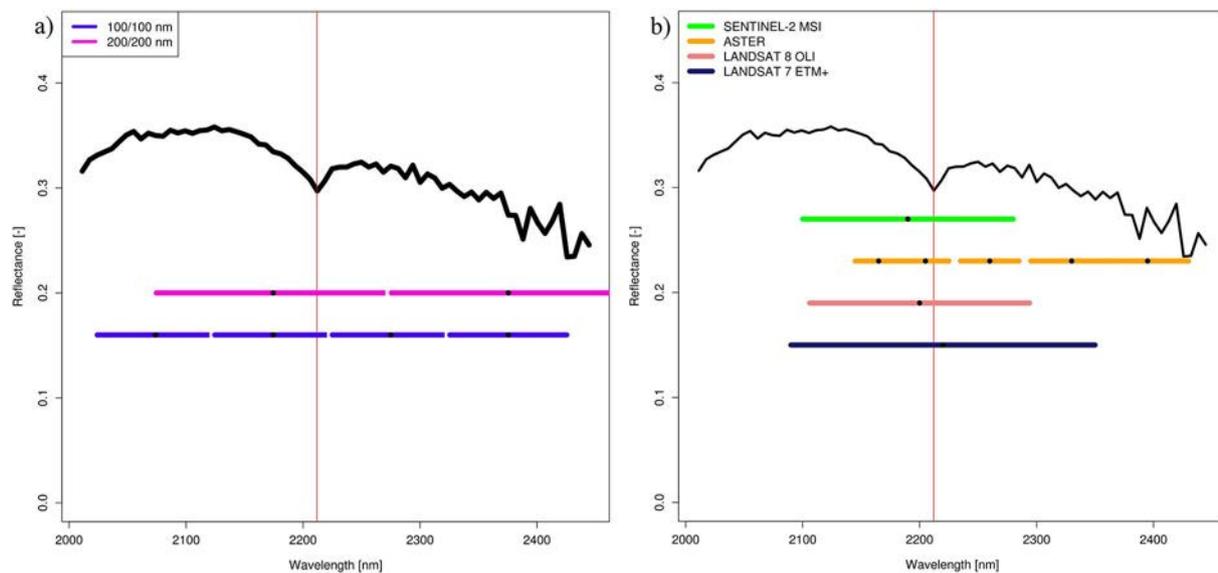Adeline et al. (2017) also demonstrated that, in the laboratory

10

Fig. 8. Spectral bands of the a) two RSR-sensors at ~100/100 nm and ~200/200 nm and b) of each IRS-Sensor. Black circles represent the central wavelength of the spectral bands. The black spectrum represents the AISA-DUAL reflectance of a random pixel. The vertical red line represents the center of the clay absorption band (at 2210 nm). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

VNIR/SWIR spectroscopy, the sensitivity of soil properties to instrumental spectral configurations depends on the chemical absorption features and the correlations with other soil properties. This demonstration was done on three soil properties with spectral absorption responses (clay, $CaCO_3$ and iron) and one soil property with no spectral absorption response (pH) (Adeline et al., 2017). A coarsening of spectral resolution may induce a small decrease in model performance of a soil property when it has large and pronounced spectral features and when it is correlated with other soil properties also characterized by spectral features. In addition, a coarsening of spectral configuration may induce a large decrease in the model performance of a soil property when it has no spectral feature, as these soil properties may only rely on the beneficial effect of correlations with other soil properties having spectral features. Following these observations, iron, which has a large and pronounced spectral response (Demattê et al., 2004), would be a soil property correctly predicted at coarse spectral resolution. $CaCO_3$, which has a pronounced but narrow spectral response (Gaffey, 1986), would be a soil property more sensitive to spectral resolution coarsening. In addition, the performance of pH prediction, which is a soil property without spectral features, would be low if no correlation between it and other soil properties having spectral features exists.

5.3. Indirect impact of mixed vegetated pixels

Beyond the spectral resolution, the prediction model performance depends on the composition of the pixel for which the spectra (X_{cal}) are used to calibrate the prediction model. Indeed, these pixels must be covered by bare soils, as materials other than soil, such as the vegetation (dry and green), are perturbing factors of spectral measurements and prevent prediction model building and application (Bartholomeus et al., 2011; Ouerghemmi et al., 2011). Therefore, a major step of image preprocessing is the masking of pixels other than bare soil (vegetation, urban, and water areas), which allows for selecting pixels for i) the calibration of the prediction models and ii) the application of the model to the entire bare soil pixels. In hyperspectral studies, the NDVI and Cellulose Absorption Index (CAI) spectral indexes are frequently used to identify green and dry vegetation pixels, respectively, and mask them (Madeira Netto et al., 2007). The NDVI index can be calculated by all sensors simulated in this paper. However, the CAI index needs spectral information measured at 2000, 2100 and 2200 nm, which are not measured by any of the ISR sensors studied in this work or any of the

RSR artificial sensors from the 100/100 nm spectral resolution. Dry vegetation is usually masked on LANDSAT data using the Modified Soil Adjusted Crop Residue Index (MSACRI) index (Bannari et al., 2000) and on ASTER data using the Lignin-Cellulose Absorption (LCA) index (Daughtry et al., 2005). Therefore, the masking step of pixels covered by some dry vegetation may be an additional drawback for the potential of sensors with low spectral resolution. It may be interesting, in futures works, to estimate the impact of this step (and the impact of these indexes on dry vegetation identification) on the prediction performances.

In our study, the AISA-DUAL data were acquired after the plowing work which allows to remove the dry vegetation (culture residue of the last year) or at least to mix the soil with these residues. Therefore, the impact of the masking step of pixels covered by some dry vegetation was not studied in this work and would have to be done in further studies.

6. Conclusion

The launch of the forthcoming sensors, in addition to existing ones, will produce an increasing amount of VNIR/SWIR data over the world, and soil surface quality could be mapped over larger areas than it is currently. This work investigated the sensitivity of spectral configurations on clay content estimation and mapping to identify adequate sensor(s) for soil mapping. Performances of the PLSR models built from six simulated sensors (with regular spectral ranges from 5 nm to 200 nm) and four existing multispectral sensors (with irregular spectral ranges relating the characteristics of ASTER, SENTINEL-2 MSI, LANDSAT-7 ETM + and LANDSAT-8 OLI) were analyzed. Moreover, spatial structures of maps obtained using these PLSR models were also analyzed. Analysis of the PLSR model performances highlighted that the RSR sensors with spectral configurations up to 100/100 nm and the ASTER spectral configuration may allow for the estimation of clay content. In addition, pedological patterns of estimated clay content are preserved with spectral configurations up to 100/100 nm and with the ASTER spectral configuration. Finally, performances of satellites (forthcoming and existing) for soil property predictions may depend on the spectral absorption features of the studied soil properties. In addition, studies limited to spectral configuration evaluations (i.e., Castaldi et al., 2016; Adeline et al., 2017), including this paper, or to spatial resolution evaluations (Gomez et al., 2015) may not give an exact view of the potential of VNIR/SWIR sensors. Coupling spectral and spatial

resolutions would provide further and necessary analysis of the potential of forthcoming and existing sensors.

Acknowledgments

References

Abrams, M., Hook, S., 2003. ASTER Users Handbook. Version 2. Jet Propulsion Laboratory. 4800 Oak Grove Dr. Pasadena, CA. pp. 135. https://asterweb.jpl.nasa.gov/content/03_data/04_Documents/aster_user_guide_v2.pdf.

Adeline, K., Gomez, C., Gorretta, N., Roger, J.M., 2017. Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data. Geoderma 288, 143–153.

Baillarin, S.J., Meygret, A., Dechoz, C., Petrucci, B., Lacherade, S., Tremas, T., Isola, C., Martimort, P., Spoto, F., 2012. Sentinel-2 level 1 products and image processing performances. In: 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, pp. 7003–7006.

Baize, D., Jabiol, B., 1995. Guide pour la description des sols. INRA édition, Paris.

Bannari, K., Haboudane, D., McNairn, H., Bonn, F., 2000. Modified Soil Adjusted Crop Residue Index (MSACRI): a new index for mapping crop residue. In: Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International, Honolulu, pp. 2936–2938.

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1993. Correction to the description of standard normal variate (snv) and de-trend transformations in practical spectroscopy with applications in food and beverage analysis – 2nd edition. J. Near Infrared Spectrosc. 1, 185–186.

Bartholomeus, H., Kooistra, L., Stevens, A., Van Leeuwen, M., VanWesemael, B., Ben-Dor, E., Tychon, B., 2011. Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. Int. J. Appl. Earth Obs. Geoinf. 13 (1), 81–88.

Baumgardner, M.F., Silva, L.F., Biehl, L.L., Stoner, E.R., 1985. Reflectance properties of soils. In: Advances in Agronomy. 38. pp. l–44.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., McBratney, A., 2010. Prediction of soil attributes by NIR spectroscopy. A critical review of chemometric indicators commonly used for assessing the quality of the prediction. Trac-Trends Anal. Chem. 29 (9), 1073–1081.

Ben-Dor, E., Banin, A., 1995. Near infrared analysis (NIRA) as a simultaneous method to evaluate spectral featureless constituents in soils. Soil Sci. 159, 259–269.

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. Int. J. Remote Sens. 23, 1043–1062.

Ben-Dor, E., Kafri, A., Varacalli, G., 2014. SHALOM: an Italian-Israeli hyperspectral orbital mission – update. In: International Geoscience and Remote Sensing Symposium, Quebec, Canada, July 13–18, 2014.

Castaldi, F., Palombo, A., Santini, F., Pascucci, S., Pignatti, S., Casa, R., 2016. Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. Remote Sens. Environ. 179, 54–65.

Chabrillat, S., Goetzd, A.F.H., Krosley, S., Olsen, H.W., 2002. Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. Remote Sens. Environ. 82, 431–445.

Chang, C.-W., Laird, D.A., Mausbach, M.J., Hurburgh Jr., C.R., 2001. Near-infrared reflectance spectroscopy—principal components regression analysis of soil properties. Soil Sci. Soc. Am. J. 65, 480–490.

Chiang, L.H., Pell, R.J., Seasholtz, M.B., 2003. Exploring process data with the use of robust outlier detection algorithms. J. Process Control 13 (5), 437–449.

Chong, I.G., Jun, C.H., 2005. Performance of some variable selection methods when multicollinearity is present. Chemom. Intell. Lab. Syst. 78 (2005), 103–112.

Daughtry, C.S.T., Hunt Jr, E.R., Doraiswamy, P.C., McMurtrey, I.I.I.J.E., 2005. Remote sensing the spatial distribution of crop residues. Agron. J. 97, 864–871.

Demattê, J.A.M., Campos, R.C., Alves, M.C., Fiorio, P.R., Nanni, M.R., 2004. Visible-NIR reflectance: a new approach on soil evaluation. Geoderma 121, 95–112.

Escadafal, R., Pouget, M., 1987. Cartographie des formations superficielles en zone aride (Tunisie méridionale) avec Landsat TM, Source Photo-Interprétation. 4. pp. 9–15.

Farrand, W.H., Singer, R.B., Merényi, E., 1994. Retrieval of apparent surface reflectance from AVIRIS data: a comparison of empirical line, radiative transfer and spectral mixture methods. Remote Sens. Environ. 47, 311–321.

Folkman, M., Pearlman, J., Liao, L., Jarecke, P., 2001. EO1/Hyperion hyperspectral imager design, development, characterization and prediction. In: Smith, W.L., Yasuoka, Y. (Eds.), Hyperspectral Remote Sensing of the Land and Atmosphere. SPIE

Proceeding. 4151. pp. 40–51.

Gaffey, S.J., 1986. Spectral reflectance of carbonate minerals in the visible and near infrared (0.35–2.55 pm): calcite, aragonite and dolomite. J. Geophys. Res. Solid Earth 71, 151–162.

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185, 1–17.

Gomez, C., Coulouma, G., Lagacherie, P., 2012a. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis–NIR data. Geoderma 189–190, 176–185.

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

Gomez, C., Oltra Carrio, R., Lagacherie, P., Bacha, S., Briottet, X., 2015. Sensitivity of soil property prediction obtained from hyperspectral Vis-NIR imagery to atmospheric effects and degradation in image spatial resolutions. Remote Sens. Environ. 164, 1–15.

Guanter, L., Kaufmann, H., Segl, K., Förster, S., Rogass, C., Chabrillat, S., Küster, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Müller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., van der Linden, S., Leitao, P.J., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., Sang, B., 2015. The EnMAP spaceborne imaging spectroscopy mission for earth observation. Remote Sens. 7 (7), 8830–8857.

Haaland, D.M., Thomas, E.V., 1988. Partial least-squares methods for spectral analyses. 2. Application to simulated and glass spectral data. Anal. Chem. 60 (11), 1202–1208.

Hunt, G.R., Salisbury, J.W., Lenhoff, C.J., 1971. Visible and near-infrared spectra of minerals and rocks: III. Oxides and hydroxides. Mod. Geol. 2, 195–205.

Kariuki, P.C., Woldai, T., Van Der Meer, F., 2004. Effectiveness of spectroscopy in identification of swelling indicator clay minerals. Int. J. Remote Sens. 25, 455–469.

Kruse, F., Perry, S., 2013. Mineral mapping using simulated Worldview-3 short-wave-infrared imagery. Remote Sens. 5 (2013), 2688–2703.

Lagacherie, P., Baret, F., Feret, J.-B., MadeiraNetto, J., Robbez-Masson, J.-M., 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. Remote Sens. Environ. 112 (3), 825–835.

Lee, C.M., Cable, M.L., Hook, S.J., Green, R.O., Ustin, S.L., Mandl, D.J., Middleton, E.M., 2015. An introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory activities. Remote Sens. Environ. 167, 6–19.

Lefèvre-Fonollosa, M.J., Bajouk, T., Briottet, X., Carrère, V., Delacourt, C., Feret, J.B., Gastellu-Etchegorry, J.P., Gomez, C., Jacquemoud, S., Le Dantec, N., Marion, R., Petit, T., Weber, C., 2016. Preparing the Future: The HYPXIM Mission, 36th EARSeL Symposium 2016, Bonn, Germany, 20–24 June 2016.

Levin, N., Kidron, G.J., Ben-Dor, E., 2007. Surface properties of stabilizing coastal dunes: combining spectral and field analyses. Sedimentology 54, 771–788.

Lopinto, E., Ananasso, C., 2013. The Prisma Hyperspectral Mission. 33rd EARSeL Symposium, Towards Horizon 2020: Earth Observation and Social Perspectives, Matera (Italy), 3–6 June 2013.

Madeira Netto, J.S.R., Robbez-Masson, J.M., Martins, E., 2007. Visible–NIR hyperspectral imagery for discriminating soil types in the La Peyne watershed (France). In: Lagacherie, P., Mc Bratney, A., Voltz, M. (Eds.), Digital Soil Mapping: An Introductory Perspective. Elsevier.

Mark, H.L., Tunnell, D., 1985. Qualitative near infrared reflectance analysis using Mahalanobis distances. Anal. Chem. 57 (7), 1449–1456.

Masek, J.G., Honzak, M., Goward, S.N., Liu, P., Pak, E., 2001. Landsat-7 ETM + as an observatory for land cover. Remote Sens. Environ. 78 (1), 118–130.

Mekki, I., Albergel, J., Ben Mechlia, N., Voltz, M., 2006. Assessment of overland flow variation and blue water production in a farmed semi-arid water harvesting catchment. Phys. Chem. Earth 31, 1048–1061.

Ouerghemmi, W., Gomez, C., Naceur, S., Lagacherie, P., 2011. Applying blind source separation on hyperspectral data for clay content estimation over partially vegetated surfaces. Geoderma 163 (3–4), 227–237.

Pearson, R.K., 2002. Outliers in process modeling and identification. IEEE Trans. Control Syst. Technol. 10 (1), 55–63.

Raclot, D., Albergel, J., 2006. Runoff and water erosion modelling using WEPP on a Mediterranean cultivated catchment. Phys. Chem. Earth 31, 1038–1047.

Roger, J.M., Palagos, B., Bertrand, D., Fernandez-Ahumada, E., 2010. CovSel: variable selection for highly multivariate and multi-response calibration. Application to IR spectroscopy. Chemom. Intell. Lab. Syst. 106 (2), 27.

Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z., 2014. Landsat-8: science and product vision for terrestrial global change research. Remote Sens. Environ. 145, 154–172.

Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., Mendonça-Santos, M.D.L., Minasny, B., Montanarella, L., Okoth, P., Palm, C.A., Sachs, J.D., Shepherd, K.D., Vågen, T.-G., Vanlauwe, B., Walsh, M.G., Winowiecki, L.A., Zhang, G.-L., 2009. Digital soil map of the world. Science 325, 680–681.

Selige, T., Bohner, J., Schmidhalter, U., 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. Geoderma 136 (1–2), 235–244.

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 Sens. 7, 6059–6078.

Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., Wesemael, B., 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. Geoderma 158, 1–2.

Swayze, G.A., Clark, R.N., Goetz, A.F.H., Chrien, T.G., Gorelick, N.S., 2003. Effects of spectrometer band pass, sampling, and signal-to-noise ratio on spectral identification using the Tetracorder algorithm. J. Geophys. Res. 108 (E9), 5105.

Van Der Meer, F.D., Van Der Merff, H.M.A., Van Ruitenbeek, F.J.A., 2014. Potential of ESA's Sentinel-2 for geological applications. Remote Sens. Environ. 148, 124–133.

Viscarra Rossel, R.A., Jeon, Y.S., Odeh, I.O.A., McBratney, A.B., 2008. Using a legacy soil sample to develop a mid-IR spectral library. Soil Res. 46, 1–16.

Webster, R., Oliver, M.A., 1990. Statistical Methods in Soil and Land Resource Survey. Oxford University Press, Oxford, pp. 316.

Wold, S., 1978. Cross-validatory estimation of the number of components in factor and principal components models. Technometrics 20, 397–405.

Wold, S., 1978. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. SIAM J. Sci. Stat. Comput. 5 (3), 735–743.

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of Chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130.

Zante, P., Collinet, J., Pepin, Y., 2005. Caractéristiques pédologiques et hydrométéorologiques du bassin versant de Kamech, Cap Bon, Tunisie. UMR LISAH IRD Tunis, DG ACTA Direction des Sols Tunis, INRGREF Tunis. (21 p. + 6 annexes).