



**HAL**  
open science

## Evaluating the sensitivity of clay content prediction to atmospheric effects and degradation of image spatial resolution using Hyperspectral VNIR/SWIR imagery

Cecile Gomez, R. Oltra-Carrió, S. Bacha, Philippe Lagacherie, X. Briottet

### ► To cite this version:

Cecile Gomez, R. Oltra-Carrió, S. Bacha, Philippe Lagacherie, X. Briottet. Evaluating the sensitivity of clay content prediction to atmospheric effects and degradation of image spatial resolution using Hyperspectral VNIR/SWIR imagery. *Remote Sensing of Environment*, 2015, 164, pp.1-15. 10.1016/j.rse.2015.02.019 . hal-01225550

**HAL Id: hal-01225550**

**<https://hal.science/hal-01225550>**

Submitted on 22 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Evaluating the sensitivity of clay content prediction to atmospheric effects and degradation of image spatial resolution using Hyperspectral VNIR/SWIR imagery

C. Gomez <sup>a,\*</sup>, R. Oltra-Carrió <sup>b</sup>, S. Bacha <sup>c</sup>, P. Lagacherie <sup>d</sup>, X. Briottet <sup>e</sup>

<sup>a</sup> IRD, UMR LISAH (INRA-IRD-SupAgro), F-34060 Montpellier, France

<sup>b</sup> CESBIO, Toulouse, France

<sup>c</sup> CNCT, Tunis, Tunisia

<sup>d</sup> INRA, UMR LISAH (INRA-IRD-SupAgro), F-34060 Montpellier, France

<sup>e</sup> ONERA, The French Aerospace Lab, 31055 Toulouse, France

Visible, near-infrared and short wave infrared (VNIR/SWIR, 0.4–2.5  $\mu\text{m}$ ) hyperspectral satellite imaging is one of the most promising tools for topsoil property mapping for the following reasons: i) it is derived from a laboratory technique that has been demonstrated to be a good alternative to costly physical and chemical laboratory soil analysis for estimating a large range of soil properties; ii) it can benefit from the increasing number of methodologies developed for VNIR/SWIR hyperspectral airborne imaging; and iii) it provides a synoptic view of the study area. Despite the significant potential of VNIR/SWIR hyperspectral airborne data for topsoil property mapping, the transposition to satellite data must be evaluated. The objective of this study was to test the sensitivity of clay content prediction to atmospheric effects and to degradation of spatial resolution. This study may offer an initial analysis of the potential of future hyperspectral satellite sensors, such as the HYPerspectral X Imagery (HYPXIM), the Spaceborne Hyperspectral Applicative Land and Ocean Mission (SHALOM), the PRecursores IperSpettrale della Missione Applicativa (PRISMA), the Environmental Mapping and Analysis Program (EnMAP) and the Hyperspectral Infrared Imager (HyspIRI), for soil applications. This study employed VNIR/SWIR AISA-DUAL airborne data acquired in a Mediterranean region over a large area (300  $\text{km}^2$ ) with an initial spatial resolution of 5 m. These hyperspectral airborne data were simulated at the top of the atmosphere and aggregated at six spatial resolutions (10, 15, 20, 30, 60 and 90 m) to correlate with the future hyperspectral satellite sensors. The predicted clay content maps were obtained using the partial least squares regression (PLSR) method. The large area of the studied region allows analysis of different pedological patterns of soil composition and spatial structures. Our results showed the following: (i) when a correct compensation of atmosphere effects was performed, only slight differences were detected between clay maps retrieved from the airborne imagery and those from spaceborne imagery (both at 5 m of spatial resolution); (ii) the PLSR models, built from data with 5 to 30 m spatial resolutions, performed well, and allowed clay mapping, although variations in clay content related to short scale succession of parent material was imperfectly captured beyond 15 m of spatial resolution; (iii) the PLSR models built from data with 60 and 90 m spatial resolutions were inaccurate, and did not enable clay mapping; and (iv) the two latter results could be explained by the combination of a small short-scale clay content variability and small field sizes observed in the study area. Therefore, in the Mediterranean and under the spectral specifications of the AISA-DUAL airborne sensor, most of the future hyperspectral satellite sensors (four of the five sensors examined in this study) will be potentially useful for clay content mapping.

## 1. Introduction

In soil science, laboratory VNIR/SWIR spectroscopy has been demonstrated to be an alternative to costly physical and chemical soil analysis for the estimation of a large range of soil properties (e.g., Ben-Dor &

Banin, 1995; Cécillon et al., 2009; Viscarra Rossel, Walvoort, McBratney, Janik, & Skjemstad, 2006). One way to retrieve soil properties is to use known specific absorption features (e.g., Clark & Roush, 1984; Lagacherie, Baret, Feret, MadeiraNetto, & Robbez-Masson, 2008). For example, illite, kaolinite and montmorillonite clay materials have an absorption peak around 2.206  $\mu\text{m}$ , corresponding to the combination of OH stretch and OH-Al bending modes (e.g., Chabrilat, Goetz, Krosley, & Olsen, 2002). Illite clay material also has absorption features near

\* Corresponding author.  
E-mail address: cecile.gomez@ird.fr (C. Gomez).

2.3 and 2.4  $\mu\text{m}$  (Clark, 1999), and  $\text{CaCO}_3$  has an absorption peak around 2.348  $\mu\text{m}$  corresponding to  $\text{CO}_3$  overtone vibrations (Gaffey, 1987). Moreover, various chemometric methods have been used to relate soil VNIR/SWIR spectra to soil properties, including multiple regression analysis (Ben-Dor & Banin, 1995), stepwise multiple linear regression (Shibusawa, Imade Anom, Sato, Sasao, & Hirako, 2001), multivariate adaptive regression splines (Shepherd & Walsh, 2002), principal components regression (e.g., Chang, Laird, Mausbach, & Hurburgh, 2001) and support vector machine regression (Stevens et al., 2010). The partial least-squares regression (PLSR) method is the most common multivariate statistical technique used for spectral calibration and prediction of soil properties (e.g., Chang & Laird, 2002; McCarty, Reeves, Reeves, Follett, & Kimble, 2002). These chemometric methods allow for the prediction of soil properties, independent of whether they have a spectral response. For example, pH which is not expected to have a direct spectral response, has been fairly predicted in several cases using chemometric methods, due to covariation to spectrally active soil constituents, such as organic matter and clay (e.g., Chang et al., 2001; Islam, Stingh, & McBratney, 2003).

Because knowledge of laboratory spectroscopy has improved, the number of studies using VNIR/SWIR hyperspectral airborne imaging in topsoil property mapping has also increased (e.g., Ben-Dor, Patkin, Banin, & Karnieli, 2002; Gomez, Lagacherie, & Coulouma, 2008; Selige, Bohner, & Schmidhalter, 2006; Stevens et al., 2010). VNIR/SWIR hyperspectral airborne imaging has been considered a promising technology for increasing the accuracy of digital mapping of topsoil properties targeted by the Global Soil Map (GSM) project ([www.globalsoilmap.net](http://www.globalsoilmap.net)) (Lagacherie & Gomez, 2014). The GSM project has proposed the construction of a new digital soil map of the world at a spatial resolution of 90 m to assist in the decision-making process for a range of global issues, such as food production, climate change and environmental degradation (Sanchez et al., 2009). In this context, VNIR/SWIR hyperspectral airborne imaging can provide a synoptic view of the area under study at spatial resolutions appropriate for topsoil property mapping (Gomez, Coulouma, & Lagacherie, 2012) and is particularly well-adapted to semi-arid areas in which bare soil surfaces are common, and dry periods enable avoidance of soil moisture perturbations of the spectrum (Lagacherie et al., 2008).

Two spaceborne sensors, PRISMA (PRecursore IperSpettrale della Missione Applicativa) and EnMAP (Environmental Mapping and Analysis Program, <http://www.enmap.org/>) are expected to be launched in the near future. PRISMA is an Italian hyperspectral sensor to be launched in 2017 (Lopinto & Ananasso, 2013). The spatial resolution initially considered for this sensor was between 20 and 30 m over the spectral range of 0.4 to 2.5  $\mu\text{m}$ , and between 2.5 and 5 m in the panchromatic band (Giampaolo et al., 2008). Recently, the spatial resolution of PRISMA has been highlighted as 30 m in the 0.4 to 2.5  $\mu\text{m}$  interval, and 5 m in the panchromatic band (The Prisma Mission, Nota Informativa, Agenzia Spaziale Italiana, DC-OST-2009-124-04/09/09/Retrieved 23 August 2013). EnMAP is a German hyperspectral sensor, set for launch in 2017, with a spatial resolution of 30 m over the spectral range of 0.4 to 2.5  $\mu\text{m}$  (Stuffer et al., 2007). Moreover, the following three new advanced hyperspectral sensors are under study: HyspIRI (Hyperspectral Infrared Imager), SHALOM (Spaceborne Hyperspectral Applicative Land and Ocean Mission) and HYPXIM (HYPerspectral X Imagery). HyspIRI is an American hyperspectral sensor with a spatial resolution of 60 m over the spectral range of 0.38 to 2.5  $\mu\text{m}$  ([hyspiri.jpl.nasa.gov](http://hyspiri.jpl.nasa.gov)). SHALOM is an Italy–Israel initiative with a spectral range of 0.4 to 2.5  $\mu\text{m}$ , and a spatial resolution initially defined at 15 m (Bussoletti, 2012), which recently evolved from 15 m to 10 m (Ben-Dor, Kafri, & Varacalli, 2014). HYPXIM is a French hyperspectral sensor with a spatial resolution better than 8 m over the spectral range of 0.5 to 2.5  $\mu\text{m}$ , and of 1.8 m in the panchromatic band (Briottet et al., 2013; Carrere et al., 2013).

Despite the significant potential of the VNIR/SWIR hyperspectral airborne data for mapping several soil properties, the application of

methods developed for airborne data (such as the PLSR method) to satellite data is not a straightforward process. Several factors can affect topsoil property mapping via satellite sensors, including a lower signal to noise ratio, atmospheric effects, different spectral characteristics (varying band centers and spectral resolutions) and coarser spatial resolutions. At the moment, the future spaceborne sensors plan to acquire reflectance data with spectral resolution of approximately 10 nm (Staenz, Mueller, & Heiden, 2013), which is close to the resolution used by airborne sensors HYMAP (approximately 15 nm, Cocks, Jenssen, Stewart, Wilson, & Shields, 1998; Kruse et al., 1999) or AISA-DUAL (approximately 5 nm and 7 nm, Specim Society, <http://www.channelsystems.ca/>). So the application of methods to satellite data for mapping several soil properties should not be affected by the spectral resolution. Conversely, spatial resolutions planned for future spaceborne sensors (between 8 to 60 m) are coarser than the ones used by airborne sensors, such as HYMAP or AISA-DUAL (approximately 5 m). The various effects of the spatial resolution degradation on multi-spectral remote sensing data have been studied, including flux prediction (e.g., Kustas & Norman, 2000), crop area estimation and crop growth monitoring (e.g., Duveiller & Defourny, 2010) and agricultural land monitoring (e.g., Ismail, Mutanga, Kumar, & Urmilla, 2008; Pax-Lenney & Woodcock, 1997). Some spatial resolution degradation studies have also focused on VNIR/SWIR hyperspectral data for vegetation studies (e.g., Nijland, Addink, De Jong, & Van der Meer, 2009; Rahman, Gamon, Sims, & Schmidts, 2003; Schaaf, Dennison, Fryer, Roth, & Roberts, 2011; Thorp, French, & Rango, 2013; Zhang, Middleton, Gao, & Cheng, 2012). However, to our knowledge, only a few studies on the impact of degraded spatial resolution have used VNIR/SWIR hyperspectral data for fire detection, temperature retrieval (Lugassi, Ben-Dor, & Eshel, 2010) and urban studies (Jensen & Cowen, 1999; Roberts, Quattrochi, Hulley, Hook, & Green, 2012), with none focused on soil science applications.

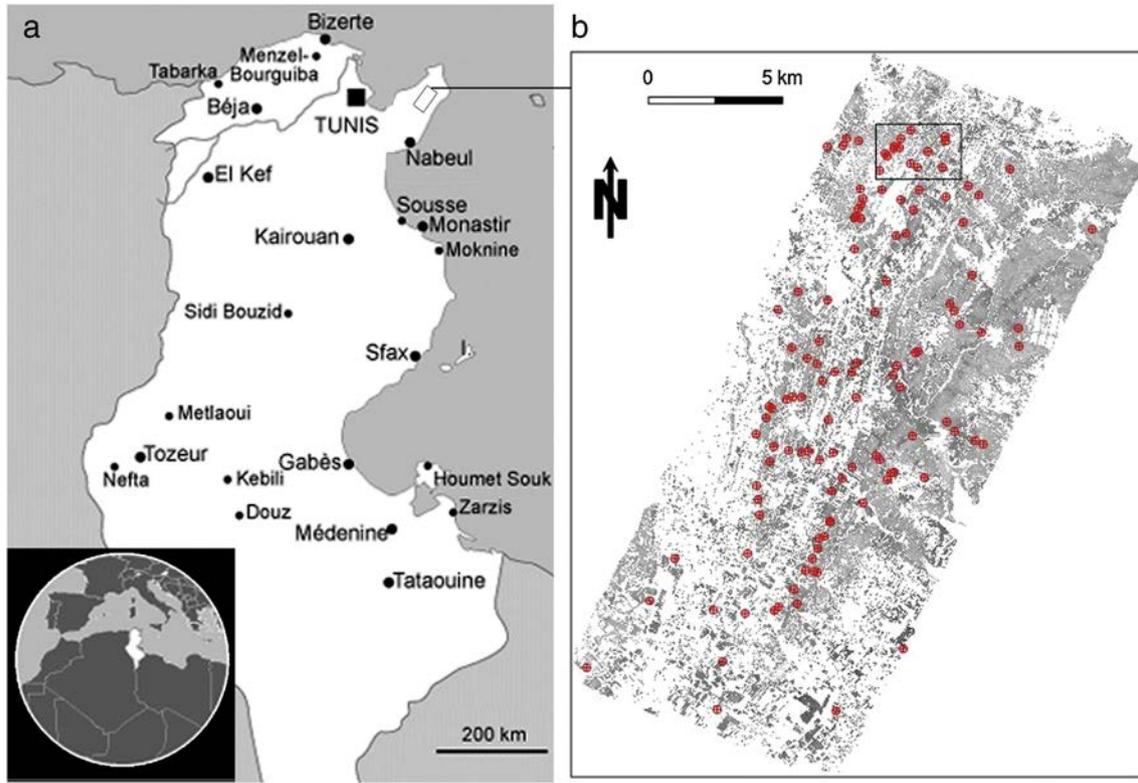
Moreover, few studies have reported the use of VNIR/SWIR hyperspectral satellite data for topsoil property mapping (e.g., Gomez, Viscarra Rossel, & McBratney, 2008; Weng, Gong, & Zhu, 2008; Zhang, Li, & Zheng, 2009). This scarcity of studies using satellite data in soil science, compared with studies using airborne data, is due to the existence of only one VNIR/SWIR hyperspectral satellite sensor (the Hyperion sensor on board the EO-1 satellite launched in 2002), which has a low signal to noise ratio (e.g., ~50:1 from 2.1 to 2.4  $\mu\text{m}$ ) (Folkman, Pearlman, Liao, & Jarecke, 2001).

The objective of this study was to evaluate the sensitivity of VNIR/SWIR hyperspectral-based soil property prediction to atmospheric effects and degradation in spatial resolution. The selected property was the topsoil clay content, which is one of the basic soil properties used by soil surveyors to describe soil types. It is also a key driver of soil erosion processes (Le Bissonnais, 1996), and one of the primary soil properties included in the GlobalSoilMap specifications (Arrouays et al., 2014). This study employed the VNIR/SWIR AISA-DUAL hyperspectral airborne data acquired over a large area (300  $\text{km}^2$ ) in a Mediterranean region, with an initial spatial resolution of 5 m. The predicted clay content maps were obtained using the PLSR method.

The study area is described in Section 2, and the airborne, simulated satellite and field data are described in Section 3. The methodology of the clay content estimation and statistical issues are presented in Section 4. Finally, the results are presented in Section 5 and are discussed in Section 6.

## 2. Study area

The study area is located in the Cap Bon region in northern Tunisia (36°24'N to 36°53'N; 10°20'E to 10°58'E), 60 km east of Tunis (Fig. 1a). This 300  $\text{km}^2$  area includes the Lebna catchment, which is primarily rural (>90%), and devoted to cereals, legumes, olive trees, natural vegetation for breeding and vineyards. This area is characterized by rolling land, with an elevation between 0 and 226 m. The climate varies



**Fig. 1.** a) Map of Tunisia and location of the study area over the Cap Bon region (black polygon) and b) the AISA-DUAL hyperspectral image at  $2.206 \mu\text{m}$ , with location of the Kamech catchment (in the rectangle). In white are the masked areas (vegetated, urban, and water areas). Red targets represent locations of collected soil samples of *CalBase\_5m*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

from humid to semi-arid, with an inter-annual precipitation of 600 mm and an annual potential evapotranspiration of 1500 mm. The soil pattern of the Lebna catchment primarily arises from variations in the lithology. The changes in the landscape between the Miocene sandstone and the marl outcrops induce significant variations in the physical and chemical soil properties (Zante, Collinet, & Pepin, 2005). Furthermore, the distance between the successive sandstone outcrops decreases in the direction of sea-mountains, which results in variations in the soil property patterns. The soil materials were redistributed laterally along the slopes during the Holocene, which adds to the complexity of the soil patterns. The main soil types are Regosols, Eutric Regosols (9.6%) preferentially associated with sandstone outcrops, Calcic Cambisol and Vertisol, preferentially formed on marl outcrops and lowlands. The southeastern region of the study area has a flatter landscape, with sandy Pliocene deposits yielding Calcosol and Rendzina.

A sub-catchment of the study area was used to examine the short scale variations in the clay contents. This sub-catchment is a  $6.67 \text{ km}^2$  area centered on the Kamech catchment (Fig. 1b). The Kamech experimental catchment belongs to a long-term environmental research observatory called OMERE (Mediterranean observatory of water and rural environment), which aims to study the anthropogenic effects on water and sediment budgets at the catchment scale (e.g., Mekki, Albergel, Ben Mechlia, & Voltz, 2006; Raclot & Albergel, 2006). The Kamech catchment represents an area (15 km from the coast) in which the soil pattern shows strong variations on a small scale, with a rapid succession of sandstone and shale. This area had a high percentage of bare soils during the image acquisition (43.4%), and exhibited contrasting soil patterns.

### 3. Airborne, simulated satellite and field data

#### 3.1. Hyperspectral airborne radiance data

On November 2, 2010, an AISA-DUAL hyperspectral image was acquired over the study area ( $12 \times 25 \text{ km}$ ) via plane, at an altitude of

3600 m, with a field of view (FOV) of  $24^\circ$  and a spatial resolution of 5 m (Fig. 1b). The AISA-DUAL airborne imaging spectrometer measures the at-flight radiances ( $L_{flight\_AISA\_5m}$ ) in 359 non-contiguous bands, covering the  $0.40 \mu\text{m}$  to  $2.45 \mu\text{m}$  spectral domain, with  $0.0046 \mu\text{m}$  bandwidths between  $0.40 \mu\text{m}$  and  $0.97 \mu\text{m}$ , and  $0.0065 \mu\text{m}$  bandwidths between  $0.97 \mu\text{m}$  and  $2.45 \mu\text{m}$ . The field of view of the instrument is  $24^\circ$ . On November 2, 2010, the visibility was estimated at 40 km, using weather reports and visual estimations from the airplane. The cloud cover was 0% during the flight-time, and neither the in-situ atmospheric profiles nor the aerosol data were available during the campaign.

#### 3.2. Simulation of the satellite radiances

##### 3.2.1. Simulation of the "Top of Atmosphere" radiances

The Top of Atmosphere (TOA) synthetic data were obtained by considering the entire atmosphere layer, requiring the definition of an atmosphere profile with a given water vapor content, and the type and abundance of the aerosols. From these inputs, the TOA radiances ( $L_{TOA\_5m}$ ) were simulated from the at-flight radiances ( $L_{flight\_AISA\_5m}$ ) (Fig. 2), according to the following equation:

$$L_{TOA\_5m} = (\tau_{dir} + \tau_{dif})L_{flight\_AISA\_5m} + L_{atm} \quad (1)$$

where  $\tau_{dir}$  and  $\tau_{dif}$  are the atmospheric direct and diffuse transmittance, respectively, and  $L_{atm}$  is the upwelling atmospheric radiance (without the contribution of the land). COMANCHE is a radiative transfer model used to solve Eq. (1) (Miesch et al., 2005). COMANCHE uses an analytical formulation of the upwelling radiance at the sensor level, in which the atmospheric parameters are independent of the ground parameters. The MODTRAN 4 radiative transfer code (Berk et al., 1999) was used to compute most of the atmosphere parameters (e.g., the columnar water vapor amount), except for the Earth atmosphere coupling irradiance

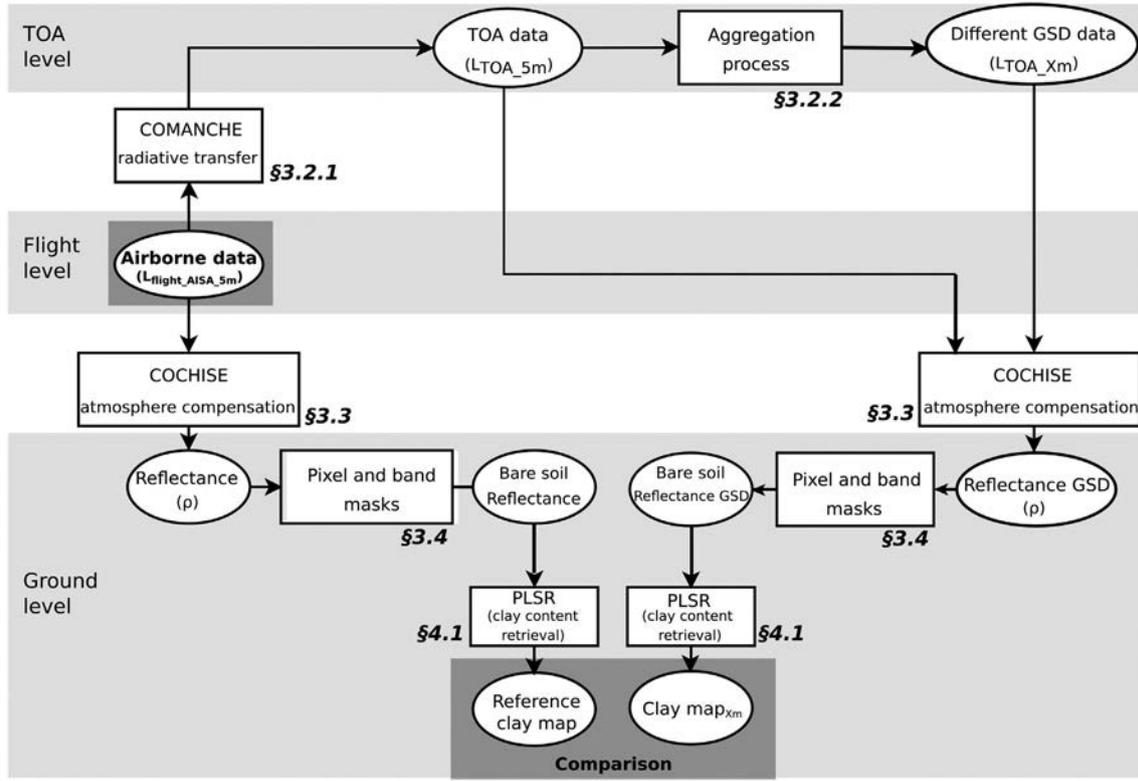


Fig. 2. Procedure overview.

and the environment upwelling radiance, which were estimated with a Monte Carlo kernel.

In the COMANCHE tool, a rural aerosol model was used with a 40 km visibility, and the United States Standard 1976 atmosphere model from MODTRAN (Berk et al., 1999). This model was chosen because it had the best fit with the value of the total water vapor content, estimated at  $1.5 \text{ g/cm}^2$  by MOD07 (<https://lpdaac.usgs.gov/>).

### 3.2.2. Spatial resolution degradation

An aggregation process was applied to the geometrically corrected TOA radiances ( $L_{TOA\_5m}$ ) (Fig. 2). As proposed in the following Eq. (2), the flux of the aggregated surface is equal to the addition of the flux of each  $i$  element forming the entire aggregated surface:

$$\Omega_{agg} L_{TOA\_xm} = \sum_{i=1}^N \Omega_i L_{TOA\_5m,i} \quad (2)$$

where  $\Omega$  is the solid angle corresponding to the instantaneous FOV related to the different ground sampling distance (GSD) values, and the index  $agg$  refers to the aggregated surface, which is composed of  $N$   $i$  elements. If the  $N$  solid angles,  $\Omega_i$ , are equivalent, then the total aggregated solid angle is given by Eq. (3). Finally, the aggregated radiance is defined by the average of the  $i$  radiances, as shown in Eq. (4). Therefore, the new aggregated pixels were obtained by averaging all of the pixel values that contributed to the output pixel. This exercise was performed using the Resampling Aggregate tool of the ENVI application using the following equations:

$$\Omega_{agg} = N\Omega_i \quad (3)$$

$$L_{TOA\_xm} = \frac{1}{N} \sum_{i=1}^N L_{TOA\_5m,i} \quad (4)$$

The selected ground sampling distances were driven by the GSD of the airborne acquired data (5 m). Then, six new TOA radiance images ( $L_{TOA\_xm}$ ) were produced at the following GSD values: 10, 15, 20, 30, 60 and 90 m. These values were close to those from the spaceborne missions, as follows: HYPXIM (10 m), SHALOM (15 m), PRISMA (20 m, 30 m), EnMAP (30 m) and HypSIPI (60 m), and from the value of the Global Soil Map (GSM) project ([www.globalsoilmap.net](http://www.globalsoilmap.net)), set at 90 m.

### 3.3. Hyperspectral reflectance data

From the airborne radiances ( $L_{flight\_AISA\_5m}$ ), the simulated TOA radiances ( $L_{TOA\_xm}$ ) described in Sections 3.1 and 3.2 and the knowledge of the atmospheric conditions, the surface reflectance,  $\rho$ , was retrieved using the atmosphere compensation tool COCHISE, which is an inverse radiative transfer model (Miesch et al., 2005) (Fig. 2), based on the following equation:

$$L_{TOA\_xm}(i, j) = L_{atm} + \tau_{dir} \cdot \left[ \frac{E_0 \cdot \rho_{xm\_SAT}(i, j)}{\pi \cdot (1 - \rho_C(i, j) \cdot S)} \right] + \tau_{dif} \cdot \left[ \frac{E_0 \cdot \rho_F(i, j)}{\pi \cdot (1 - \rho_C(i, j) \cdot S)} \right] \quad (5)$$

where  $E_0$  is the distance corrected extra-terrestrial solar irradiance,  $\rho_{xm\_SAT}(i, j)$  is the reflectance to be retrieved at pixel  $(i, j)$ ,  $\rho_F(i, j)$  and  $\rho_C(i, j)$  are the average albedo values given by the convolution with the environmental functions to consider the contribution of the environmental upwelling radiance (F) and the earth atmosphere coupling irradiance (G), respectively.  $S$  is the atmospheric spherical albedo. These terms are estimated using a Monte Carlo approach. The remaining atmospheric parameters are computed from the MODTRAN 4 radiative transfer code (Berk et al., 1999). The inversion process is iterative. In the initial step, the adjacency effects are neglected, and a first estimate of  $\rho_{xm\_SAT}(i, j)$  is obtained. Then, Eq. (5) is used to retrieve the surface reflectance from the simulated TOA radiances ( $L_{TOA\_xm}$ ). Eq. (5) is used

again to retrieve the surface reflectance from the airborne radiances ( $L_{flight\_AISA\_5m}$ ), and in this case,  $\rho_{xm\_SAT}(i, j)$  is replaced by  $\rho_{flight\_5m}(i, j)$ .

Finally, to remove the noise effects introduced by this process, the reflectance spectrum of each pixel was smoothed using the following filter:

$$\rho_i^{smooth} = \frac{1}{N} \sum_{j=i-m}^{i+m} \rho_j, \quad m = \frac{N-1}{2} \quad (6)$$

where  $\rho_i^{smooth}$  is the surface reflectance in the channel number  $i$  after the application of the filter, and  $N$  is an odd integer indicating the number of wavelength neighbors to be averaged. In our case,  $N$  is equal to 3.

Thus, eight reflectance images were obtained. One image was obtained at a 5 m resolution from the acquired airborne radiances and was called *5m\_AISA*. Seven images were obtained from the simulated TOA radiances at the different GSD values (see Section 3.2.2) and were called *5m\_HYPXIM*, *10m\_HYPXIM-SHALOM*, *15m\_SHALOM*, *20m\_PRISMA*, *30m\_PRISMA-EnMAP*, *60m\_HyspIRI* and *90m\_GSM*.

#### 3.4. Band selections and urban, vegetation and water masks

Before the implementation of the clay content retrieval procedure, noisy or troublesome spectral bands and pixels must be removed (Fig. 2). We removed the following: a) 20 spectral bands in the blue part of the spectral domain from 0.4 to 0.484  $\mu\text{m}$  due to instrumental noise, b) 5 bands from 0.747  $\mu\text{m}$  to 0.766  $\mu\text{m}$ , which correspond to  $\text{O}_2$  band absorption, c) 12 bands from 0.952  $\mu\text{m}$  to 1.019  $\mu\text{m}$  due to  $\text{H}_2\text{O}$  band absorption and instrument defect, d) 14 bands from 1.094  $\mu\text{m}$  to 1.176  $\mu\text{m}$  due to  $\text{H}_2\text{O}$  band absorption, e) 21 bands from 1.339  $\mu\text{m}$  to 1.465  $\mu\text{m}$  due to  $\text{H}_2\text{O}$  band absorption and f) 38 bands from 1.773  $\mu\text{m}$  to 2.005  $\mu\text{m}$  due to  $\text{H}_2\text{O}$  band absorption. Therefore, 249 spectral bands between 0.489  $\mu\text{m}$  and 2.451  $\mu\text{m}$  were retained.

Considering that vegetation coverage is a limiting factor for soil property mapping, the current hyperspectral data treatments do not allow comprehensive mapping of an extensive region (e.g., Gomez, Lagacherie, et al., 2008; Schwanghart & Jarmer, 2011; Selige et al., 2006; Stevens et al., 2010). Only two studies in remote sensing soil mapping have focused on diversified surface conditions, including partially vegetated surfaces (Bartholomeus et al., 2010; Ouerghemmi, Gomez, Naceur, & Lagacherie, 2011), and performed marginally well. Therefore, in the absence of effective techniques for predicting soil properties over semi-vegetated pixels, having a suitable mapping surface remains an important criterion of hyperspectral image quality. 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  $\mu\text{m}$  and 0.799  $\mu\text{m}$ . Areas of water were also masked using an expert-calibrated threshold. Pixels with a reflectance of less than 8% at 1.665  $\mu\text{m}$  were removed. Finally, 13 urban areas were identified by visual inspection and were also masked.

When the hyperspectral airborne image was acquired (November 2010), a major portion of the soil surface was covered by green vegetation, primarily consisting of olive trees, native forests, green plants and vineyards. Based on the hyperspectral image, with a spatial resolution of 5 m, bare soils represented 43.4% of our study area. A degradation of spatial resolution leads to a slight decrease in the suitable surface area that can be mapped by VNIR/SWIR spectroscopy. From 5 to 30 m, this decrease in suitable surface is approximately 4.4% of suitable surface area, and from 30 to 90 m, this decrease is approximately 6.4%. Based on the images at 10 m, 15 m, 20 m, 30 m, 60 m and 90 m of spatial resolution, bare soils represent 42.6%, 41.5%, 40.6%, 39%, 35.2% and 32.6%, respectively.

#### 3.5. Field data

One hundred twenty soil samples were collected over the study area (Fig. 1b). Among this sample set, 50 were collected in June 2008, 30 in October 2009 and 41 in November 2010. All of these soil samples were collected in fields that were bare during the hyperspectral airborne data acquisition in November 2010. All of the samples were composed of five sub-samples, which were collected to a depth of 5 cm at random locations within a  $10 \times 10 \text{ m}^2$  square centered on the geographical position of the sampling plot, as recorded by a Garmin GPS instrument. Because soil samples were collected over plowed fields, the collected soils from 0 to 5 cm can be considered as having mixed and homogeneous soil properties. After homogenizing the sample, approximately 20 g was allocated for clay content analysis. The initial samples were air-dried and sieved with a 2 mm sieve prior to being transported to the laboratory for analysis. The determination of the clay content (granulometric fraction  $< 2 \mu\text{m}$ ) was determined using a pipette method (method NF X 31-107, particle size distribution by sedimentation, Baize & Jabiol, 1995). The clay content of the 120 soil samples varied between 108 and 772 g/kg, and followed a normal distribution (Table 1).

A degradation in the spatial resolutions leads to a slight decrease in the suitable surface area that can be mapped by VNIR/SWIR spectroscopy (Section 3.4), and leads to a decrease in the number of soil samples associated with bare soil pixels. At 5 m, 120 soil samples were available over the bare soil pixels. This 120 sample database is called *CalBase-5m*. At 60 and 90 m, only 67 and 65 soil samples, respectively, were available over the bare soil pixels. These databases are called *CalBase-60m* and *CalBase-90m*, respectively. The *CalBase-5To30m* database consists of 82 soil samples, which correlate with the images from 5 to 30 m of resolution. Although the number of reference soil samples varied depending on the spatial resolution, the distribution and statistical characteristics of the soil datasets did not vary. The clay content of the *CalBase-5To30m* database varied from 122 to 772 g/kg, and followed a normal distribution (Table 1). The clay content of the *CalBase-60m* and *CalBase-90m* database varied from 122 to 772 g/kg and to 764 g/kg, respectively, and both followed a normal distribution (Table 1).

## 4. Methods

### 4.1. Clay content prediction

Partial least squares regression (PLSR) is a multivariate model commonly used for soil property estimation. The PLSR method specifies a linear relationship between a set of dependent (response) variables ( $Y$ -variables, the clay content in our case), and a set of predictor variables ( $X$ -variables, the spectra) (Tenenhaus, 1998). A detailed description of the PLSR model can be found in Wold, Sjöström, and Eriksson (2001). The general concept of PLSR is to extract the orthogonal or latent predictor variables, accounting for the maximum amount of the variation of the  $Y$ -variables. A PLSR model is developed from a training set of  $N$  observations (the number of spectra in the calibration dataset), with  $K$   $X$ -variables (the number of wavelengths in the spectra), and  $M$   $Y$ -variables (the number of soil properties). The training data form matrices  $X$  and  $Y$ , of dimensions  $(N \times K)$  and  $(N \times M)$ , respectively. As with all of the factorial methods, the main goals of PLSR are the following: i) to locate a subspace of the spectral space  $\mathbb{R}^K$  on which the spectra are projected, yielding a matrix of  $N$  scores  $T$  ( $N \times k$ ); and ii) to perform

**Table 1**  
Main descriptive statistics of the soil dataset used.

| Dataset        | Number of data | Mean $\pm$ st.dev | Min | Max | Skewness |
|----------------|----------------|-------------------|-----|-----|----------|
| CalBase-5m     | 120            | 466 $\pm$ 174     | 108 | 772 | -0.11    |
| CalBase-5To30m | 82             | 481 $\pm$ 173     | 122 | 772 | -0.19    |
| CalBase-60m    | 67             | 489 $\pm$ 173     | 122 | 772 | -0.15    |
| CalBase-90m    | 65             | 484 $\pm$ 170     | 122 | 764 | -0.23    |

a linear regression between T and Y. The maximum number of latent predictor variables was set to 10 and was determined using the Predicted Residual Sum of Squares (PRESS) statistic, taking care to avoid under- and over-fitting.

An analysis was performed to detect the outliers in the calibration dataset. Outliers are commonly defined as observations that are inconsistent with the majority of the data (Chiang, Pell, & Seasholtz, 2003; Pearson, 2002), such as observations that deviate significantly from normal values. An outlier can be defined as the following: (i) a spectral outlier when the sample is spectrally different from the remainder of the samples or (ii) a concentration outlier when the predicted value has a residual difference significantly greater than the mean of the predicted values. To identify the spectral outliers, we used the principle of the Mahalanobis distance (Mark & Tunnell, 1985) and applied it to principal component analysis reduced data. The samples with a Mahalanobis distance greater than 3 were identified as outliers and removed from the calibration dataset. For the case of normally distributed data, 99% of the data will be included in an interval of  $\pm 3\sigma$ , which will be less than a threshold of 3.

Prior to the quantitative statistical analysis, the reflectance was converted into a "pseudo absorbance" ( $\log [1/\text{reflectance}]$ ), and a standard normal variate correction was achieved for removal of additive and multiplicative effects (Barnes, Dhanoa, & Lister, 1993).

The prediction capability of the PLSR was analyzed using two different procedures, which were dependent on the size of the dataset (the number of spectra associated with the clay content). *Type 1* was applied to the *CalBase-5m* and *CalBase-5To30m* datasets, which contain more than 80 variables. In this case, the dataset was divided into two groups, one group for calibration of the PLSR model (3/4 of the dataset), and a second group used for validation (1/4 of the database). The Y values were sorted in an ascending order. The method starts by selecting the sample with the lower clay content, and placing it in a validation set. Then, the next three samples are placed in the calibration set, and the procedure is continued by alternately placing the following sample in the validation set and the next three samples in the calibration set. Such a process would ensure a relatively equal distribution of the samples in both sub-datasets. *Type 2* is applied to datasets with less than 80 reference samples (*CalBase-60m* and *CalBase-90m*). In this case, a subdivision in calibration and validation datasets would provide too little data in the calibration set. So all of the data were used in the calibration dataset, and no independent validation set was used.

In both types of procedures, a leave-one-out cross-validation procedure was adopted to verify the prediction capability of the PLSR model for the calibration set.  $N - 1$  samples were used to build the regression model from all of the  $N$  samples within the dataset. Based on this model, the value for the clay content of the sample not used in the development of the model was predicted. This procedure was repeated for all of the  $N$  samples, resulting in predictions for all of the samples.

#### 4.2. Performance of the clay content prediction

The performance of the prediction was evaluated using the coefficient of correlation ( $R^2_{\text{cal}}$ ) of the predicted values against the measured values, and the root mean square errors of calibration (RMSEC). Moreover, for *Type 1*, the correlation coefficient of validation ( $R^2_{\text{val}}$ ) and the root mean square errors of the prediction in the validation set (RMSEP) were also measured. The ratio of the performance to the deviation (RPD), which is the ratio between the standard deviation in the validation set and the RMSEP, was used. Thresholds of the RPD values used to assess the accuracy of prediction were proposed by Chang et al. (2001). RPD values greater than 2 indicate models with excellent prediction capability, values between 2 and 1.4 indicate intermediate models and values below 1.4 indicate unreliable models. However, although RPD is largely used in soil spectroscopy, these thresholds were not determined statistically, and RPD does not describe the range of variation correctly, particularly for data with a non-normal

distribution. Therefore, the ratio of performance to interquartile (RPIQ), which is the ratio of the interquartile ( $IQ = Q3 - Q1$ ) to the RMSEP recently proposed to represent the spread of the population (Bellon-Maurel, Fernandez-Ahumada, Palagos, Roger, & McBratney, 2010), was also used.

Finally, the Variable Importance in the Projection (VIP) (Chong & Jun, 2005; Wold et al., 2001) and the PLSR b-coefficients (Haaland & Thomas, 1988) were used to study the significant wavelengths used in PLSR. A wavelength is considered to be significant when both its b-coefficient and VIP value are sufficiently large (Wold et al., 2001). The thresholds for the VIP were set to 1 (following the recommendations by Chong & Jun, 2005), and the thresholds for the b-coefficients were based on their standard deviations (Viscarra Rossel, Jeon, Odeh, & McBratney, 2008). The chemometric developments were implemented in R (version 1.17) using the pls package (Mevik & Wehrens, 2007).

#### 4.3. Spatial structure issues

A comparison between two digital images can be performed using a simple method, which compares the two images on a pixel by pixel basis. However, this method does not allow for analysis of the structure of the predicted values in a spatial dimension. Thus, variograms were used to compare the ability of the prediction models to accurately reproduce the spatial structures of the soil property in the study area. Variograms measure the spatial dependence of the soil properties using semi-variance. The average variance between any pair of sampling points (i.e., the semivariance) for a soil property, Y, at any vector of distance  $h$  apart can be determined using the following formula (Webster & Oliver, 1990):

$$\gamma(h) = \frac{1}{2m} \sum_{i=1}^m \{Y(x_i) - Y(x_i + h)\}^2 \quad (7)$$

where  $\gamma(h)$  is the average semi-variance of the soil property,  $m$  is the number of pairs of sampling points,  $Y$  is the soil property content,  $x$  is the coordinate of the point and  $h$  is the lag (the distance that the pairs are apart). Variograms provide evidence of spatial autocorrelation when the semi-variances are lower at smaller lags than at larger lags, i.e., the sampling locations located close to each other have similar values. The autocorrelation of clay distribution can be studied by means of nugget (unexplained variability referring to noise in the data), partial sill (structural component of the total variance) and range (range of the variogram) analysis. The geostatistical operations were performed with R (version 1.17), using the gstat package (Pebesma, 2004).

To evaluate the rate of soil property variability that was theoretically lost as the spatial resolution increased, the variability in the predicted clay contents for each aggregated image was estimated from the map of clay content predicted using the *5m\_AISA* airborne data. For this purpose, grids corresponding to square surfaces, centered on the observed sites with sizes corresponding to the tested spatial resolutions, were created, and the inner-grid variability,  $V$ , was calculated according to the following equation:

$$V = \frac{1}{N} \sum_{i=1}^N \frac{1}{p} \sum_{j=1}^p \sqrt{(z_i(x_j) - \bar{z}_i)^2} \quad (8)$$

where,  $N$  is the number of calibration soil samples,  $p$  is the number of pixels in the grids,  $z_i(x_j)$  is the predicted soil property, i.e., the clay content of the pixel  $x_j$  in the grid centered on the  $i$ th calibration sample, and  $\bar{z}_i$  is the mean of the predicted clay contents inside the grid centered on the  $i$ th calibration sample. The number of pixels,  $p$ , depends on the studied spatial resolution. Because the original spatial resolution was 5 m and the grids must be centered in the calibration pixels, the variability could not be measured exactly for the GSD values proposed in this study. The following five sizes were constructed:  $3 \times 3$  pixels ( $p = 9$ ,

spatial resolution of 15 m),  $5 \times 5$  pixels ( $p = 25$ , spatial resolution of 25 m),  $7 \times 7$  pixels ( $p = 49$ , spatial resolution of 35 m),  $13 \times 13$  pixels ( $p = 169$ , spatial resolution of 65 m) and  $19 \times 19$  pixels ( $p = 361$ , spatial resolution of 95 m) (Fig. 3).

## 5. Results

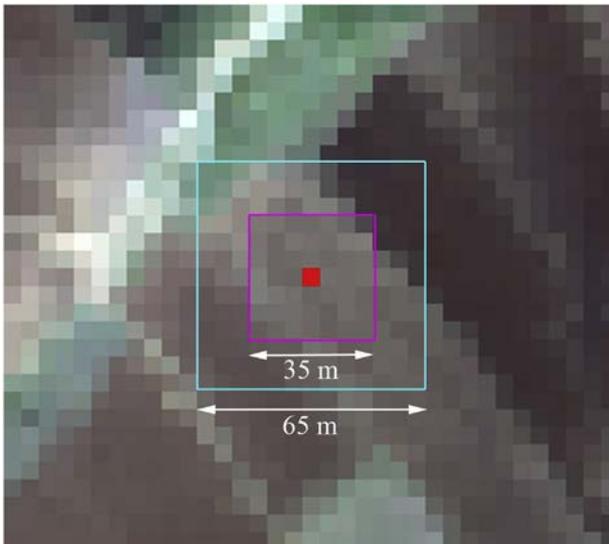
### 5.1. Preliminary results

A reference PLSR model, following *Type 1* (Section 3.4), was built from the *5m\_AISA* airborne image and the 120 clay content values associated with the 120 available soil samples. Ninety samples were assigned to the calibration database, and thirty to the validation database. The performance of this model was relatively accurate, with an  $R^2_{cal}$  value of approximately 0.74, an  $R^2_{val}$  value of approximately 0.75, a RPIQ of 3.2 and an RMSEP value of approximately 86 g/kg (Table 2), as previously shown in Gomez, Lagacherie, and Bacha (2012). The inner-pixel variability,  $V$  (Eq. 8), ranges from 27 g/kg (with  $N = 82$  grids of  $15 \times 15$  m<sup>2</sup> and  $p = 9$ ) to 66 g/kg (with  $N = 65$  grids of  $95 \times 95$  m<sup>2</sup> and  $p = 361$ ) (Fig. 4).

This reference PLSR model was applied to all the spectra over bare soil from the *5m\_AISA* airborne image to produce a predicted clay content map. The model is referred to as the *5m\_AISA\_Clay* map (Fig. 5a).

Contrasting clay content and soil patterns appear between the Pliocene area, located in the southeast corner of the image, and the Miocene area, covering the remainder of the image. The Pliocene area exhibits low and weakly variable topsoil clay content, whereas the Miocene area shows a large range of clay content values. Variations within the Miocene area are also visible, and follow the geological pattern formed by the alternating sandstone and marl outcrops, yielding low (blue) and high (red) values of clay content, respectively (Fig. 5a). The predicted clay content values of the *5m\_AISA\_Clay* map followed a normal distribution, were centered on 447 g/kg and had a standard deviation of 160 g/kg, and a skewness of 0.17 (Fig. 5c in blue).

The empirical variogram of the predicted clay content values exhibited spatial structures with clear increases in the semi-variances as distance between pairs of points increased (Fig. 5d). An exponential model was sufficient to fit the variogram of the predicted clay content values (blue line on Fig. 5d). The theoretical variogram is characterized by a nugget of approximately 3900, a sill of approximately 16,000 and a range of approximately 380.



**Fig. 3.** Grid of  $7 \times 7$  pixels (pink square) and  $13 \times 13$  pixels (cyan square) which correspond respectively to 35 m and 65 m of spatial resolution. The pixel in red indicates the center location a calibration site. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 5.2. Sensitivity to atmospheric effects

A PLSR model was built following *Type 1*, using the 120 spectra associated with the bare soil pixels of the *5m\_HYPXIM* image, for which soil samples were collected and clay contents were measured. The models built from the *5m\_HYPXIM* and *5m\_AISA* images were compared (Table 2). The best model was obtained from the *5m\_AISA* image (Table 2,  $R^2_{val} \approx 0.75$  and RPIQ  $\approx 3.2$ ). However, the performance of the model obtained from the satellite-simulated data at 5 m is only slightly less accurate (Table 2,  $R^2_{val} \approx 0.71$  and RPIQ  $\approx 3$ ). Therefore, the atmosphere appears to slightly affect the performance of the PLSR-based prediction models.

No significant difference in the b-coefficients appears between the models (Fig. 6a). Only the b-coefficients between 0.932 and 1.056  $\mu\text{m}$  and 1.182 and 1.25  $\mu\text{m}$  were more important for the PLSR-model built from the satellite-simulated data than from the airborne data. No significant difference in the VIP values appears between both models (Fig. 6b). Only the VIP values between 1.5 and 1.77  $\mu\text{m}$  were more important for the PLSR model built from the airborne data than from the satellite-simulated data. However, the same important wavelengths were identified regardless of the PLSR-model used. Finally, the analysis of the combination of the VIP and the b-coefficients shows that the most important bands are located from 2.2 to 2.4  $\mu\text{m}$  in both images, including the spectral absorption band of clay at 2.206  $\mu\text{m}$  (Fig. 6a and b).

The PLSR model built following *Type 1*, and using the 120 spectra associated with the bare soil pixels of the *5m\_HYPXIM* image, was then applied to all the spectra over bare soil in the *5m\_HYPXIM* image. Therefore, a predicted clay content map was obtained for the study area from this PLSR model and was called the *5m\_HYPXIM\_Clay\_1* map (Fig. 5b). No significant difference can be observed between maps *5m\_HYPXIM\_Clay\_1* and *5m\_AISA\_Clay* via visual inspection (Fig. 5a and b). The same large pedological units can be identified, indicating that both of the PLSR models made the same relative predictions. Moreover, no difference occurs in the absolute prediction, considering the distribution of the predicted clay content over the entire area (Fig. 5c). The predicted clay content values of the *5m\_HYPXIM\_Clay* map, follow a normal law centered on 449 g/kg, a standard deviation of 155 g/kg and a skewness of 0.19 (Fig. 5c). Finally, no difference of prediction appears, and the same pedological variations occur in the semi-variogram of the study area (Fig. 5d). An exponential model was sufficient to fit the variogram of the predicted clay content values (pink line on Fig. 5c). Moreover, the theoretical variogram is characterized by a nugget of approximately 3800, a sill of approximately 15,000 and a range of approximately 380 m.

### 5.3. Sensitivity to degradation of spatial resolutions

The analysis of the PLSR models, built from the seven simulated-satellite images, allowed the identification of two groups of sensors with specific behaviors.

A first group of spatial resolutions, from 5 to 30 m, corresponding to HYPXIM, SHALOM, PRISMA and EnMAP sensors, offers accurate PLSR models with  $R^2_{cal} > 0.65$ , RMSEC  $< 100$  g/kg and  $R^2_{val} > 0.6$  (Table 2). The number of outliers remains stable (between 1 and 3) regardless of the spatial resolution of this first group of sensors. Slight differences in the prediction performance were observed in the validation dataset. The spatial resolutions of 5 and 30 m yielded similar prediction performances, whereas the spatial resolution of 20 m provided a better performance, with an  $R^2_{val}$  value of 0.81 and RPIQ value of 3.7. The spatial resolution of 15 m provided the worst performance, with an  $R^2_{val}$  value of 0.7 and RPIQ value of 3 (Table 2). Therefore, these results indicate that between 5 and 30 m of spatial resolution, the various model performances do not decrease when the spatial resolution becomes coarser, and the degradation in spatial resolutions of the images does not linearly affect model performances. Finally, no significant difference in the VIP and b-coefficient values appears between the first group

**Table 2**  
Results of the PLSR models.

|                   | Type of PLSR model | Number of latent variables | Number of calibration data | $R^2_{cal}$ | RMSEC (g/kg) | Number of validation data | $R^2_{val}$ | RMSEP (g/kg) | RPD | RPIQ | Name of predicted clay map |
|-------------------|--------------------|----------------------------|----------------------------|-------------|--------------|---------------------------|-------------|--------------|-----|------|----------------------------|
| 5m_AISA           | Type_1             | 5                          | 90                         | 0.74        | 86           | 30                        | 0.75        | 86           | 2   | 3.2  | 5m_AISA_Clay               |
| 5m_HYPXIM         | Type_1             | 6                          | 90                         | 0.74        | 86           | 30                        | 0.71        | 94           | 1.9 | 3    | 5m_HYPXIM_Clay_1           |
| 5m_HYPXIM         | Type_1             | 6                          | 62                         | 0.68        | 99           | 20                        | 0.71        | 90           | 1.9 | 3    | 5m_HYPXIM_Clay_2           |
| 10m_HYPXIM-SHALOM | Type_1             | 6                          | 62                         | 0.71        | 94           | 20                        | 0.6         | 105          | 1.6 | 2.6  | 10m_HYPXIM_Clay            |
| 15m_SHALOM        | Type_1             | 6                          | 62                         | 0.67        | 99           | 20                        | 0.7         | 91           | 1.9 | 3    | 15m_SHALOM_Clay            |
| 20m_PRISMA        | Type_1             | 6                          | 62                         | 0.72        | 91.22        | 20                        | 0.81        | 73           | 2.3 | 3.7  | 20m_PRISMA_Clay            |
| 30m_PRISMA-ENMAP  | Type_1             | 6                          | 62                         | 0.66        | 100          | 20                        | 0.7         | 90           | 1.9 | 3    | 30m_PRISMA-ENMAP_Clay      |
| 60m_HyspIRI       | Type_2             | 4                          | 67                         | 0.21        | 156          |                           |             |              |     |      |                            |
| 90m_GSM           | Type_2             | 4                          | 65                         | 0.36        | 135          |                           |             |              |     |      |                            |

models (Fig. 7a and b). The most important bands are located from 1.6 to 1.8  $\mu\text{m}$  and from 2.2 to 2.4  $\mu\text{m}$  for all of the models, including the spectral absorption band of clay at 2.206  $\mu\text{m}$  (Fig. 7a and b). The spectral range from 1.6 to 1.8  $\mu\text{m}$  can include information on soil moisture at the acquisition time, which may be related to clay. The visualization of three spectra, extracted from the 5m\_HYPXIM and 30m\_PRISMA-EnMAP images, and corresponding to 3 different clay contents (122, 489 and 772 g/kg), show only a slight difference of reflectance (Fig. 8). The most important difference between spectra at both spatial resolutions comes from the albedo. Nevertheless, this difference of albedo is not the same for each target. A higher albedo exists for the 30m\_PRISMA-EnMAP image than for the 5m\_HYPXIM image, for targets with clay contents of 489 and 772 g/kg. An inversely lower albedo exists for the 30m\_PRISMA-EnMAP image than for the 5m\_HYPXIM image, for targets with clay content of 122 g/kg (Fig. 8).

A second group of spatial resolutions corresponding to the HyspIRI sensor and GSM requirement, and greater than 30 m, produced inaccurate PLSR models, with an  $R^2_{cal} < 0.4$  and  $\text{RMSEC} > 130$  g/kg (Table 2). The analysis of the combination of the VIP and b-coefficients shows that only approximately ten spectral bands, centered at 2.206  $\mu\text{m}$ , were significant for the model at 60 m, and no significant spectral bands were identified at 90 m (Fig. 9a and b).

Each PLSR model built from the 82 soil samples of the common dataset was applied to all the spectra over bare soil for its corresponding simulated-satellite image, generating predicted clay maps called 5m\_HYPXIM\_Clay\_2, 10m\_HYPXIM-SHALOM\_Clay, 15m\_SHALOM\_Clay, 20m\_PRISMA\_Clay and 30m\_PRISMA-EnMAP\_Clay. Because the PLSR

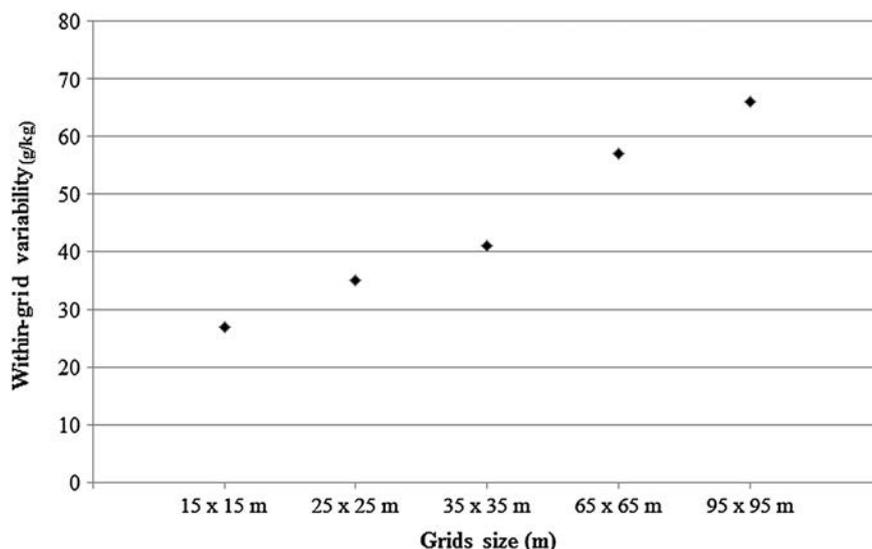
models built from the simulated-satellite images at 60 and 90 m were inaccurate, no predicted clay maps were built at these resolutions. Regardless of the spatial resolution from 5 to 30 m, no large difference occurs in the absolute prediction in regard to the distribution of the predicted clay content over the entire area (Fig. 10).

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. The 5m\_HYPXIM\_Clay\_2 map, which was restricted to the Kamech catchment, shows these variations in the soil pattern (Fig. 11a). Certain mixed areas also appear in transition areas between these outcrops, and in shoal areas. The reduction in the spatial resolution resulted in a decrease in the number of bare soil pixels that can be used in VNIR/SWIR spectrometry, providing only a partial view of the succession of the clay-rich areas and clay-poor areas (Fig. 11). Particularly, clay-rich areas with a width less than 60 m disappear when a spatial resolution of 30 m is used (Fig. 11e). For the intermediate spatial resolutions (10 to 20 m), marl and sandstone outcrops can still be observed, but the boundaries between these outcrops are unclear (Fig. 11b, c and d).

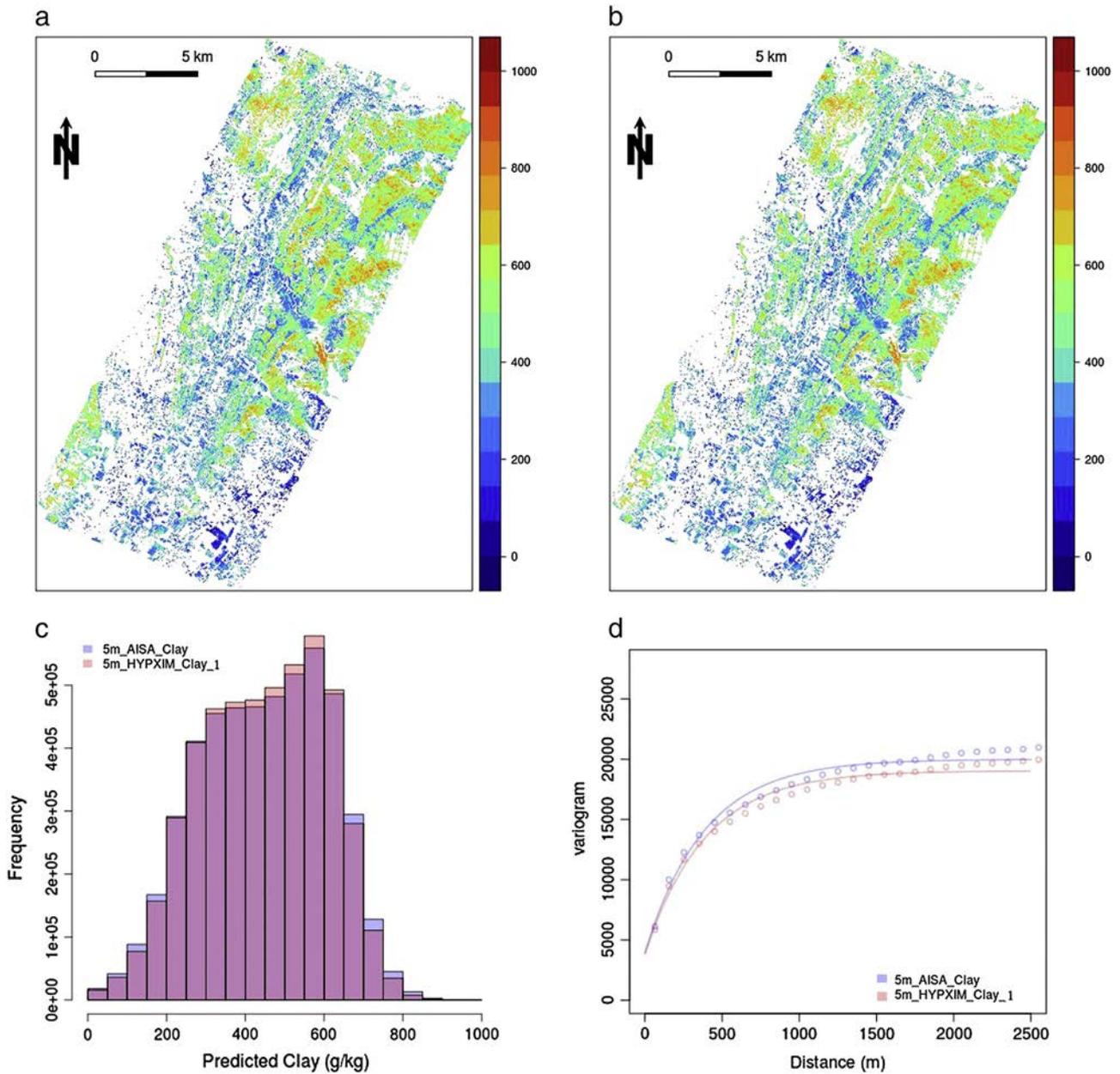
## 6. Discussion

### 6.1. Impact of atmosphere compensation

From our simulations of the satellite data, the atmosphere compensation appears to only slightly affect the performances of the



**Fig. 4.** Within-grid variability  $V$  of the clay contents predicted from the AISA\_5m image over surfaces of 15 × 15 m, 25 × 25 m, 35 × 35 m, 65 × 65 m and 95 × 95 m.



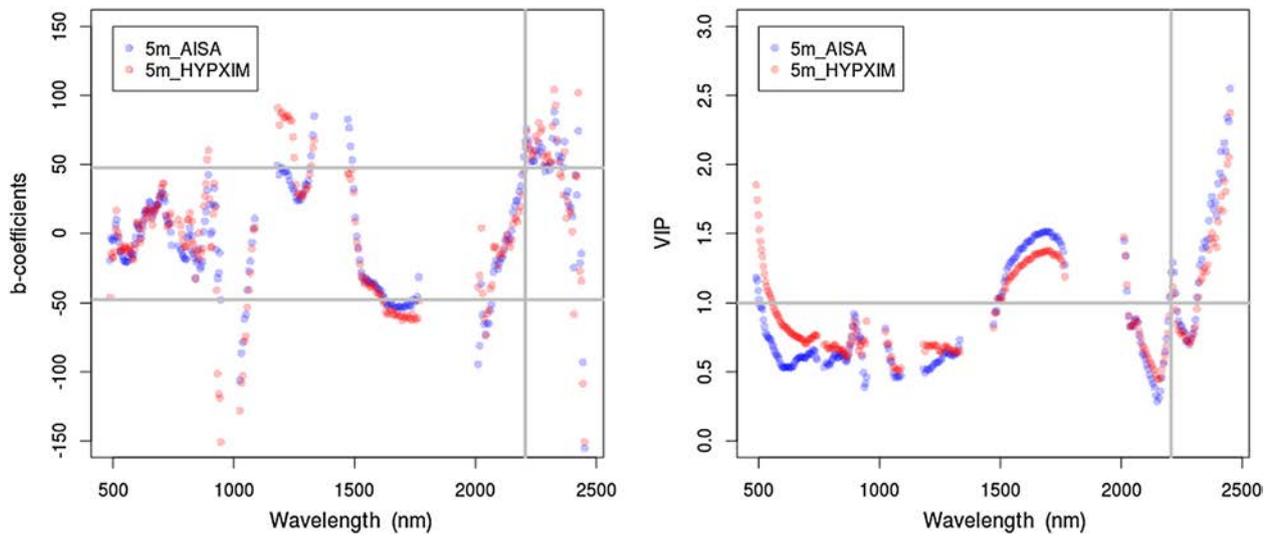
**Fig. 5.** a) 5m\_AISA\_Clay map, b) 5m\_HYPXIM\_Clay\_1 map, c) histograms and d) empirical variograms (points) and theoretical variograms fitted (lines) of clay predictions of the 5m\_AISA\_Clay (blue) and 5m\_HYPXIM\_Clay\_1 (red) maps (with lag of 100 m). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

PLSR-based prediction models. The best model was obtained from the AISA-DUAL airborne image. The performance of the PLSR model obtained from the satellite-simulated data at 5 m is only slightly less accurate. The most important difference in the performance may occur from a misunderstanding of the atmosphere. However, in our case study, the water vapor content is very low because such an error in this parameter will not induce a significant difference in the PLSR performance. Furthermore, the visibility during the flight was very high (40 km, close to Rayleigh conditions). Therefore, the effect of the aerosols on the simulated satellite and the retrieved reflectance data can both be considered negligible. For such a low aerosol load, the type of aerosol has no effect. Considering the previous and numerous validation studies conducted using COMANCHE (Miesch et al., 2005) and COCHISE (using Modtran kernel), the errors introduced by the models themselves can be neglected. The results obtained from the satellite-simulated data at 5 m correctly reflect the performances that will be

obtained from future satellite images at 5 m of spatial resolution, and under these atmospheric conditions.

## 6.2. Short-scale clay content variations as a driver of spatial resolutions impact

The spatial resolutions from 5 to 30 m appear to only slightly affect the performances of the PLSR-based prediction models (Table 2). The results of the calibration and validation datasets indicate that the PLSR model performances do not decrease when the spatial resolution becomes coarser, and the degradation in spatial resolution of the image does not linearly affect the model performance. The spatial resolutions of 5 and 30 m yielded similar prediction performances, whereas the spatial resolution of 20 m provided better performance. The slight differences in the performances may arise from the calibration dataset, which contains an acceptable, but not a significant number of data to



**Fig. 6.** b-Coefficients (left) and VIP (right) values for both PLSR models built from the *5m\_HYPXIM* and *5m\_AISA* images, using 120 soil samples. The spectral absorption band of clay at 2.206  $\mu\text{m}$  is located at the vertical gray line over both plots. The horizontal lines indicate thresholds of b-coefficients (their standard deviation) and VIP (threshold at 1).

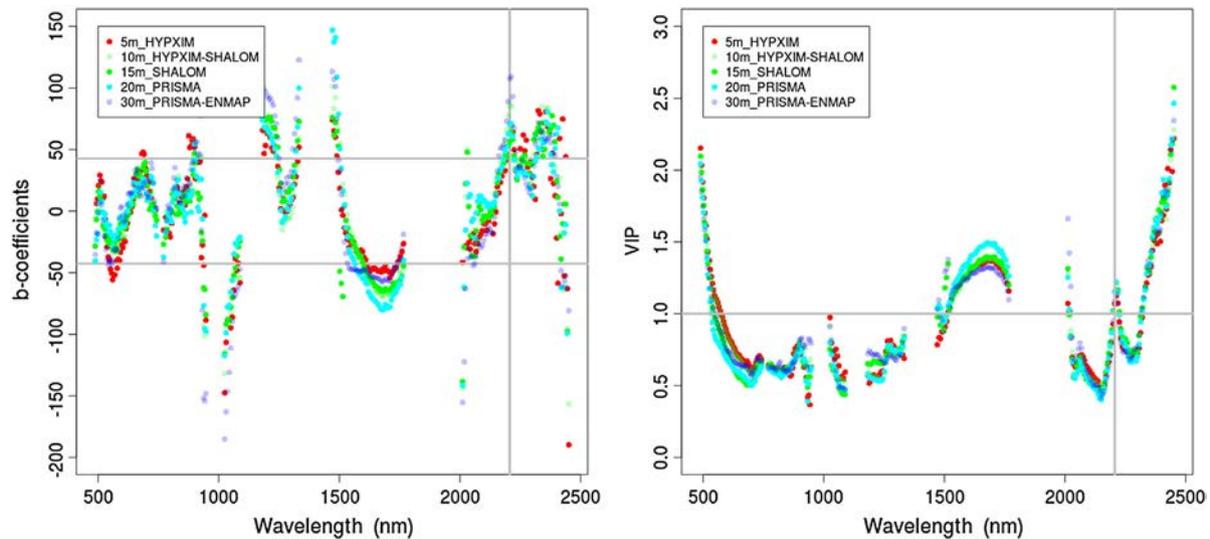
calibrate the models. The slight effect of the spatial resolutions on the PLSR model performances may be explained by the low rate of short-scale clay content variations, which represents only 22% of the whole variance of clay content in our study area. Indeed, the clay samples collected for characterizing surfaces of  $10 \times 10 \text{ m}^2$  were also representative of surfaces of  $30 \times 30 \text{ m}^2$  because differences in clay contents remain low within such small areas.

### 6.3. Field size as a driver of spatial resolutions impact

The results revealed a failure in the performances of the PLSR-based prediction models beyond 30 m of spatial resolution, which can be attributed to two cumulative errors. The first error may arise from the Y-variables of the PLSR models (clay contents), and the second error may be due to the X-variables (soil spectra). The error due to the Y-variables (clay contents) is linked to the natural pedological variations in the study area. The soil samples collected over a  $10 \times 10 \text{ m}^2$

surface would not be representative of surfaces larger than 30 m. The clay content measured over the  $10 \times 10 \text{ m}^2$  surface would be excessively different from the clay content of the  $90 \times 90 \text{ m}^2$  surface. However, the error in the Y-variables cannot completely explain the poor performances observed beyond 30 m because the inner-grid variability of clay content, discussed above, increases no more than linearly beyond 30 m (Fig. 4). The error due to the X-variables (soil spectra) is mainly driven by the changes in the topsoil surface conditions, which add a “noise” to the calibration spectra. In the study area, the mean area of the cultivated fields is 0.56 ha (Jenhaoui, Raclot, & Lamachère, 2008). Therefore, a pixel of  $60 \times 60 \text{ m}^2$  or  $90 \times 90 \text{ m}^2$  can straddle two fields, and can include two types of surface roughness and soil humidity due to the different types of plowing (Fig. 3).

It is also worth noting that degradation in spatial resolution leads to a slight decrease in the suitable surface area that can be mapped using VNIR/SWIR spectroscopy (Section 3.4), and leads to a decrease in the number of soil samples associated with bare soil pixels. When the pixels



**Fig. 7.** b-Coefficients (left) and VIP (right) values for PLSR models built from simulated-satellite images, using the 82 soil samples of the *CalBase-5To30m* dataset. The spectral band of clay at 2.206  $\mu\text{m}$  is located at the vertical gray line over both plots. The horizontal lines indicate thresholds of VIP (threshold at 1) and b-coefficients (calculated from the b-coefficients of the model based on the *5m\_HYPXIM* data).

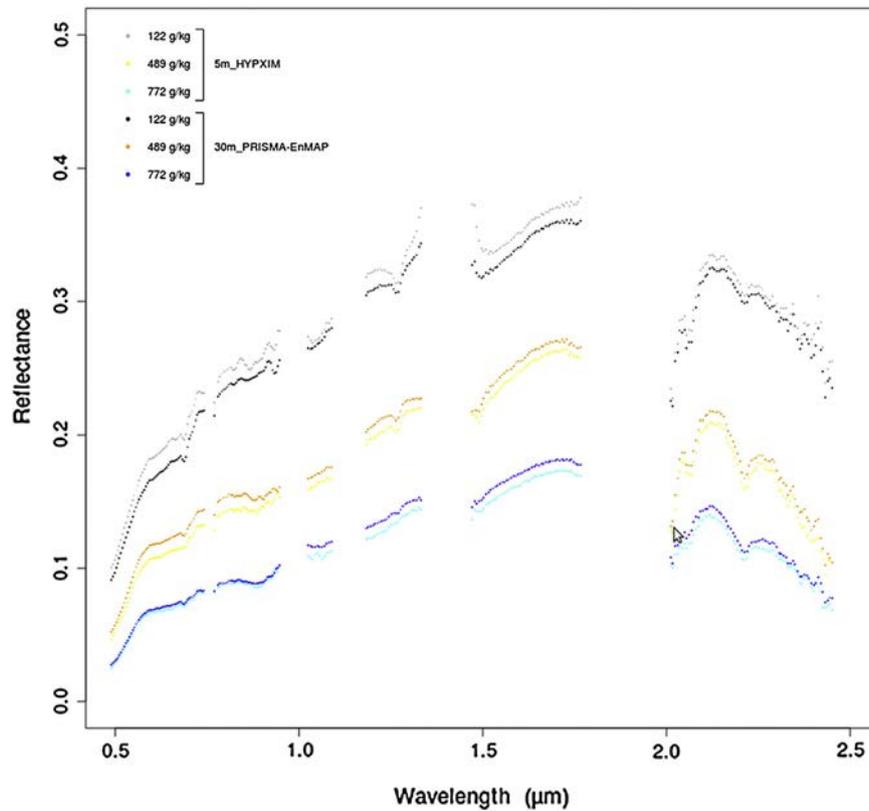


Fig. 8. Reflectance spectra of 3 pixels corresponding to 3 different clay contents (122, 489 and 772 g/kg) extracted from the 5m\_HYPXIM and 30m\_PRISMA-EnMAP images.

of the study area become coarser, an increasing number of pixels become inadequate for calibration and prediction because they more often include heterogeneous land use. The smaller the field size, the more prominent the phenomenon is. VNIR-SWIR airborne spectroscopy has proven to be a powerful tool for estimating soil properties over pixels covered by bare soils, whereas the applicability of spectroscopic techniques decreases when the pixels are mixed, such as when a soil surface is partially covered with vegetation (Bartholomeus et al., 2010; Ouerghemmi et al., 2011). However, this limitation would not

be insurmountable if the development of the unmixing methods becomes an extensive research area.

#### 6.4. Exploration of results for future hyperspectral-satellite-based soil predictions

The relation between spatial resolutions and performance of clay content predictions found in this study is specific to this area, and might differ for other areas, soil properties or pedological contexts.

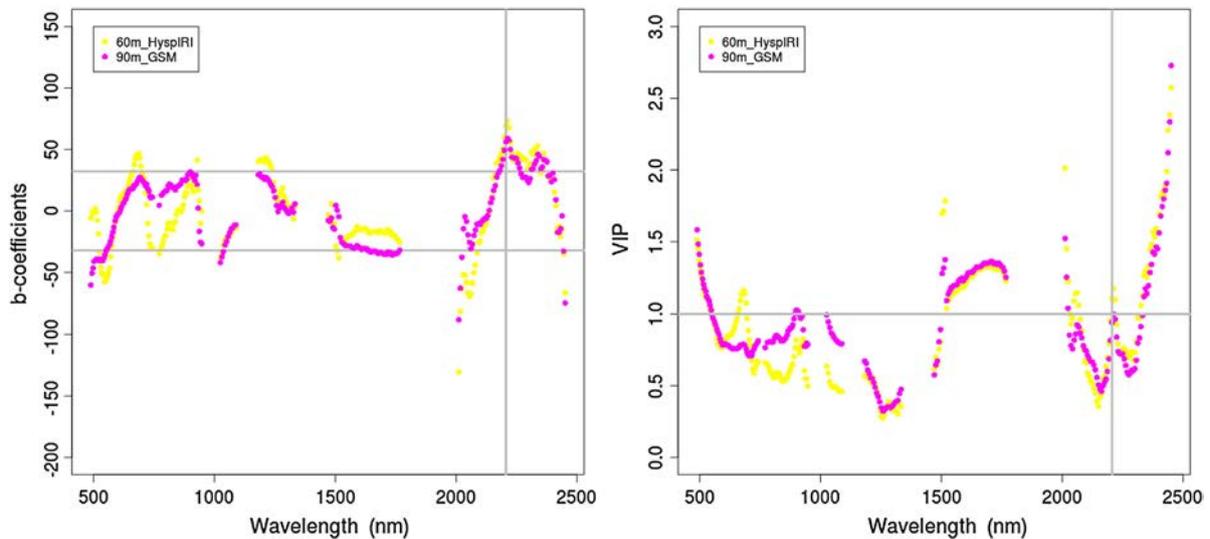
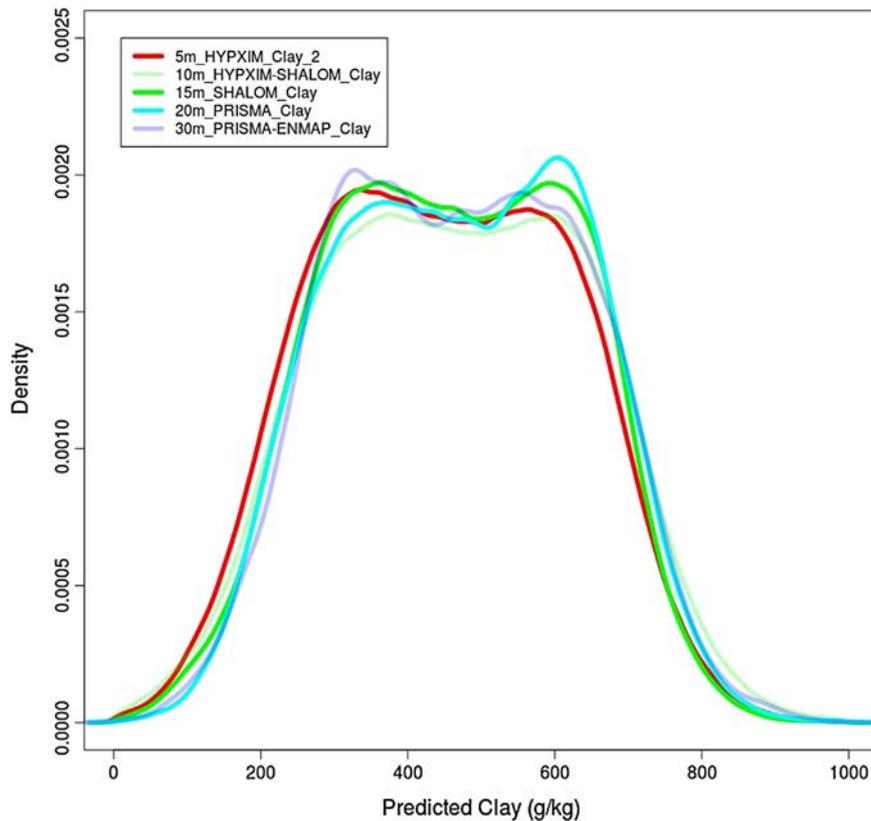


Fig. 9. b-Coefficients (left) and VIP (right) values for PLSR models built from the simulated-satellite images at 60 and 90 m of spatial resolution, using the soil samples of the CalBase-60m and CalBase-90m datasets respectively. The spectral band of clay at 2.206  $\mu\text{m}$  is located at the vertical gray line over both plots. The horizontal lines indicate thresholds of VIP (threshold at 1) and b-coefficients (calculated from the b-coefficients of the model based on the 60m\_HyspIRI data).



**Fig. 10.** Histograms of clay predictions obtained over the entire area, from the 5m\_HYPXIM, 10m\_HYPXIM-SHALOM, 15m\_SHALOM, 20m\_PRISMA and 30m\_PRISMA-ENMAP images using the 82 soil samples of the CalBase-5To30m dataset.

Nevertheless, the slight effect of the spatial resolutions from 5 to 30 m on the PLSR model performances can reasonably be extrapolated to areas where the soil properties exhibit moderate spatial variabilities at short distances, as was observed in this case study. An analysis of the variograms of the clay contents gathered by McBratney and Pringle (1999) (Fig. 2b, p 132) showed that 77% of these variograms had a smaller semi-variance at a lag value of 30 m than the variogram in our case study. This finding suggests that the results obtained in this study could be extrapolated over a significant portion of the planet. Similarly, the decrease in performances observed beyond 30 m is closely related with the small sizes of fields in our study area (mean size = 0.56 ha). However, these may represent important surfaces, such as those covered by traditional and familial agrosystems.

Finally, by identifying two main drivers of spatial resolution impact, namely short scale soil variability and field size, this study paves the way to an a priori assessment of the performances of hyperspectral-satellite-based soil predictions from measurable properties of continental surfaces. More studies similar to the one presented in this paper will be necessary to strengthen this assessment.

#### 6.5. Comparisons with other targeted properties of continental surfaces

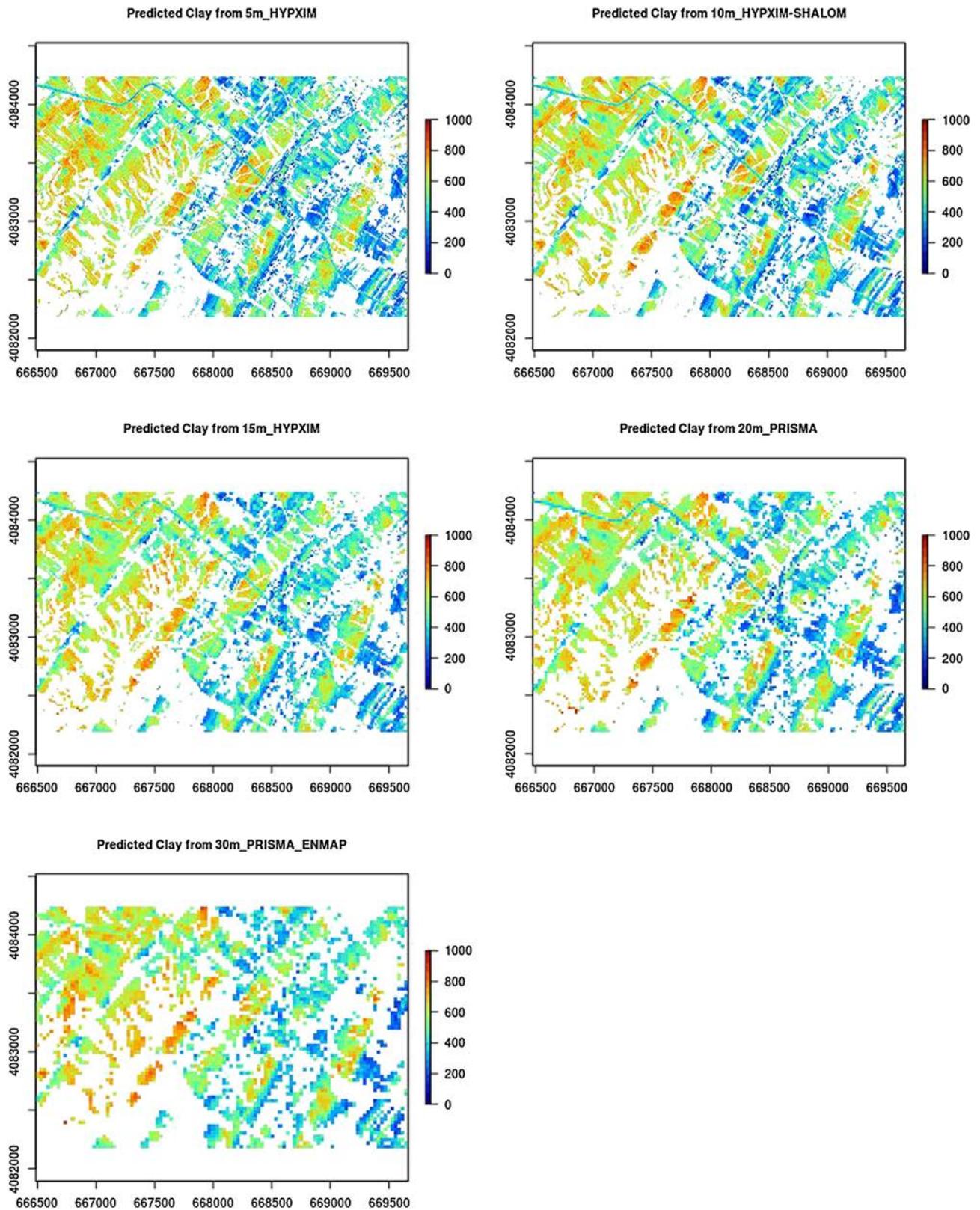
Based on a literature review, although different spatial resolutions are estimated to be optimal depending on the application, these optimal resolutions are all included from 6 to 20 m. Rahman et al. (2003) found that a pixel size of 6 m or less would be optimal for studying a canopy-to watershed-level ecosystem, and a pixel size of 20 m or more would imply a failure to identify the variability of the plants. Thorp et al. (2013) showed that the optimal spatial resolution for studying vegetation features in their study area was less than 15 m. Therefore, the soil does not present high specificities regarding spatial resolutions of future hyperspectral satellite sensors.

#### 6.6. Future research

Residue cover, soil roughness and moisture content are environmental factors that are known to affect soil spectra collected in the field and in laboratory (e.g., Liu et al., 2002; Ouerghemmi et al., 2011). Therefore, these environmental factors are expected to also affect the soil spectra recorded by imaging sensors, including proxy, Unmanned Aerial Vehicles, airborne or satellite sensors. Currently, few publications address the consideration of these environmental factors in spectra collection and prediction models (e.g., Haubrock et al., 2008; Lagacherie et al., 2008). Futures research should focus on how spectral information from surface soils and soil prediction models is affected by disturbances in soil surface conditions, such as roughness, residues and variations in soil moisture, and at different measurement scales (i.e., different spatial resolutions). Taking into account environmental factors may help the research community to build more performant prediction models, in waiting the hyperspectral satellite data.

### 7. Conclusions

Aerial Vis-NIR hyperspectral imaging sensors have previously shown their effectiveness for use in soil surface mapping. Therefore, the availability of future hyperspectral satellite data might offer considerable opportunities for mapping topsoil properties over large areas. In this study, we found that spatial resolutions from 5 to 30 m provided suitable surface clay contents maps in our Mediterranean study area. We also demonstrated that the impact of spatial resolution is dependent on the short-scale spatial variability of the soil property and field size. Thorp et al. (2013) found that the spatial resolution is of secondary importance to the spectral resolution for mapping green vegetation, nonphotosynthetic vegetation and bare soil using a Multiple Endmember Spectral Mixture (MESMA). All the future hyperspectral spaceborne sensors are expected



**Fig. 11.** a) 5m\_HYPXIM\_Clay\_2 map, b) 10m\_HYPXIM-SHALOM\_Clay map, c) 15m\_SHALOM\_Clay map, d) 20m\_PRISMA\_Clay map and e) 30m\_PRISMA-ENMAP\_Clay map over the Kamech catchment (upper rectangle in Fig. 1b).

to acquire reflectance data with similar spectral resolution of approximately 10 nm (Staenz et al., 2013), which is close to the resolution used by Vis-NIR hyperspectral imaging airborne sensors, such as HYMAP or AISA-DUAL. As these airborne sensors have already shown their potentials for soil surface property mapping, we can be confident

in the performances of future hyperspectral spaceborne sensors. Nevertheless, it would be interesting to study the sensitivity of soil property prediction results to degradation in spectral resolutions, as has been initiated by Casa, Castaldi, Pascucci, Palombo, and Pignatti (2013), as well as the signal to noise ratio. This could potentially improve the design of

future spaceborne sensors, in terms of spectral resolutions and signal to noise ratio. Finally, the optimal spatial resolutions found in this study are specific to this area, and likely differ in other areas with different soil properties. To reinforce these conclusions and the guideline, this study could be enlarged to include additional soil properties and additional pedological settings.

## Acknowledgments

This research was granted by the TOSCA-CNES project «HUMPER-Mission HYPXIM : Apport de la résolution spatiale de la mission HYPXIM pour l'étude des propriétés pérennes des sols et de leur humidité de surface» (2013–2014). This research was also supported by the French National Research Agency (ANR) through the ALMIRA project (ANR-12-TMED-0003). The authors are indebted to UMR LISAH (IRD, France) and to CNCT (Centre National de Cartographie et de Télédétection, Tunisia), for providing the AISA-DUAL hyperspectral images for this study. This hyperspectral data acquisition was granted by IRD (Institut de Recherche pour le Développement), INRA (Institut National de la Recherche Agronomique) and the French National Research Agency (ANR) (ANR-O8-BLAN-C284-01).

## References

Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B. M., Hong, S. Y., et al. (2014). Chapter three – GlobalSoilMap: Toward a fine-resolution global grid of soil properties. In Donald L. Sparks (Ed.), *Advances in Agronomy, Volume 125*. (pp. 93–134). Academic Press.

Baize, D., & Jabiol, B. (1995). *Guide pour la description des sols*. Paris: INRA édition.

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. *Journal of Near Infrared Spectroscopy*, 1, 185–186.

Bartholomeus, H., Kooistra, L., Stevens, A., Van Leeuwen, M., VanWesemael, B., Ben-Dor, E., et al. (2010). Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, 13(1), 81–88.

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 in Analytical Chemistry*, 29(9), 1073–1081.

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

Ben-Dor, E., Kafri, A., & Varacalli, G. (2014). SHALOM: An Italian-Israeli hyperspectral orbital mission – Update. *International Geoscience and Remote Sensing Symposium, Quebec, Canada, July 13–18, 2014*.

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.

Berk, A., Anderson, G. P., Bernstein, L. S., Acharya, P. K., Dothe, H., Matthew, M. W., et al. (1999). MODTRAN 4 radiative transfer modeling for atmospheric correction. *Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research III*, 3756, 348–353.

Briottet, X., Marion, R., Carrere, V., Jacquemoud, S., Bourguignon, A., Chami, M., et al. (2013). HYPXIM: HYPXIM: A second generation high spatial resolution hyperspectral satellite for dual applications. *5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), June 2013, Gainesville, Florida, USA*.

Bussoletti, E. (2012). Space observations for agriculture and food support. *Inter-Agency Meeting on Outer Space Activities: 2012. Thirty-second session, 7–9 March 2012. Rome, Italy*.

Carrere, V., Bourguignon, A., Briottet, X., Chami, M., Chevrel, S., Jacquemoud, S., et al. (2013). *The French Hyperspectral Earth Observation Science/Defense mission HYPXIM – A second generation high spectral and spatial resolution imaging spectrometer*. IGARS, 21–26 July 2013, Melbourne, Australia.

Casa, R., Castaldi, F., Pascucci, S., Palombo, A., & Pignatti, S. (2013). A comparison of sensor resolution and calibration strategies for soil texture estimation from hyperspectral remote sensing. *Geoderma*, 197–198(2013), 17–26.

Cécillon, L., Barthès, B. G., Gomez, C., Ertlen, D., Genot, V., Hedde, M., et al. (2009). Assessment and monitoring of soil conditions using indexes based on near infrared reflectance (NIR) spectroscopy. *European Journal of Soil Science*, 60, 770–784.

Chabrilat, S., Goetz, A. F. H., Krosley, L., & Olsen, H. W. (2002). Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. *Remote Sensing of Environment*, 82(2), 431–445.

Chang, C. -W., & Laird, D. A. (2002). Near-infrared reflectance spectroscopic analysis of soil C and N. *Soil Science*, 167(2), 110–116.

Chang, C. -W., Laird, D. A., Mausbach, M. J., & Hurlburgh, C. R., Jr. (2001). Near-infrared reflectance spectroscopy—Principal components regression analysis of soil properties. *Soil Science Society of America Journal*, 65, 480–490.

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

Chong, I. G., & Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 78(2005), 103–112.

Clark, R. N. (1999). Chapter 1: Spectroscopy of rocks and minerals, and principles of spectroscopy. In A. N. Rencz (Ed.), *Manual of Remote Sensing. Remote Sensing for the Earth Sciences, Volume 3*. (pp. 3–58). New York: John Wiley and Sons.

Clark, R. N., & Roush, T. L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research*, 89(B7), 6329–6340.

Cocks, T., Janssen, R., Stewart, A., Wilson, I., & Shields, T. (1998). The HyMap Airborne Hyperspectral Sensor: The system, calibration and performance. In M. Schaeppman, D. Schlöpfer, & K. I. Itten (Eds.), *Proc. 1st EARSeL Workshop on Imaging Spectroscopy. 6–8 October 1998, Zurich, EARSeL, Paris*. (pp. 37–42).

Duveiller, G., & Defourny, P. (2010). A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing. *Remote Sensing of Environment*, 114, 2637–2650.

Folkman, M., Pearlman, J., Liao, L., & Jarecke, P. (2001). EO1/Hyperion hyperspectral imager design, development, characterization and prediction. In W. L. Smith, & Y. Yasuoka (Eds.), *Hyperspectral Remote Sensing of the Land and Atmosphere. SPIE Proc, Vol. 4151*. (pp. 40–51).

Gaffey, S. J. (1987). Spectral reflectance of carbonate minerals in the visible and near infrared (0.35–2.55  $\mu\text{m}$ ): Anhydrous carbonate minerals. *Journal of Geophysical Research*, 92.

Giampaolo, P., Cisbani, A., De Cosmo, V., Galeazzi, C., Labate, D., & Melozzi, M. (2008). Hyperspectral instruments for earth observation. *International conference on Space Optics, October 14–17 Toulouse, France*.

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

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), 141–148.

Gomez, C., Viscarra Rossel, R. A., & McBratney, A. B. (2008). Soil organic carbon prediction by hyperspectral remote sensing and field VNIR/SWIR spectroscopy: An Australian case study. *Geoderma*, 146(3–4), 403–411.

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

Haubrock, S. -N., Chabrilat, S., Kuhnert, M., Hostert, P., & Kaufmann, H. (2008). Surface soil moisture quantification and validation based on hyperspectral data and field measurements. *Journal of Applied Remote Sensing*, 2(1).

Islam, K., Stingh, B., & McBratney, A. B. (2003). Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Australian Journal of Soil Research*, 41, 1101–1114.

Ismail, R., Mutanga, O., Kumar, L., & Urmilla, B. (2008). Determining the optimal spatial resolution of remotely sensed data for the detection of *Sirex noctilio* infestations in pine plantations IN KwaZulu-Natal, South Africa. *South African Geographical Journal*, 90(1), 22–30.

Jenhouji, J., Raclot, D., & Lamachère, J. M. (2008). *Le parcellaire et l'occupation du sol entre 2004 et 2008 sur le bassin versant de Kamech (Cap Bon, Tunisie)*. Tunis: UMR LISAH, IRD (15 pp. + annexes).

Jensen, J. R., & Cowen, D. C. (1999). Urban/suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering and Remote Sensing*, 65(5), 611–622.

Kruse, F. A., Boardman, J. W., Lefkoff, A. B., Young, J. M., Kiernin-Young, K. S., & Cocks, T. (1999). The 1999 USA HyMap Group Shoot. *Proceedings, ISSSR'99, 31 October–4 November 1999, Las Vegas, Nevada*.

Kustas, W. P., & Norman, J. M. (2000). Evaluating the effects of subpixel heterogeneity on pixel average fluxes. *Remote Sensing of Environment*, 74, 327–342.

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 Sensing of Environment*, 112(3), 825–835.

Lagacherie, P., Ruth, A., Gomez, C., Bacha, S., Coulouma, G., Hamrouni, H., Mekki, I., & Bacha, S. (2013). Combining hyperspectral imagery and legacy measured soil profiles to map subsurface soil properties in a Mediterranean area (Cap-Bon, Tunisia). *Geoderma*, 209–210, 168–176.

Lagacherie, P., & Gomez, C. (2014). What can GlobalSoilMap expect from Vis–NIR hyperspectral imagery in the near future? In Dominique Arrouays, Neil McKenzie, Jon Hempel, Anne Richer de Forges, & Alex B. McBratney (Eds.), *GlobalSoilMap: Basis of the global spatial soil information system* (pp. 387–392). CRC Press.

Le Bissonnais, Y. (1996). Soil aggregate stability and assessment of soil crustability and erodibility: I. Theory and methodology. *European Journal of Soil Science*, 47, 425–431.

Liu, L., Baret, F., Xingfa, G., Quingxi, T., Lanfen, Z., & Bing, Z. (2002). Relating soil surface moisture to reflectance. *Remote Sensing of Environment*, 81, 238–246.

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*.

Lugassi, R., Ben-Dor, E., & Eshel, G. (2010). A spectral-based method for reconstructing spatial distributions of soil surface temperature during simulated fire events. *Remote Sensing of Environment*, 114(2), 322–331.

- Mark, H. L., & Tunnell, D. (1985). Qualitative near-infrared reflectance analysis using Mahalanobis distances. *Analytical Chemistry*, 57, 1449–1456.
- McBratney, A. B., & Pringle, M. J. (1999). Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. *Precision Agriculture*, 1, 219–236.
- McCarty, G. W., Reeves, J. B., III, Reeves, V. B., Follett, R. F., & Kimble, J. M. (2002). Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurements. *Soil Science Society of America Journal*, 66, 640–646.
- 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. *Physics and Chemistry of the Earth*, 31, 1048–1061.
- Mevik, B. -H., & Wehrens, R. (2007). The pls Package: Principal component and partial least squares regression in R. *Journal of Statistical Software*, 18(2), 1–24.
- Miesch, C., Poutier, L., Achard, V., Briottet, X., Lenot, X., & Boucher, Y. (2005). Direct and inverse radiative transfer solutions for visible and near-infrared hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 43(7), 1552–1562.
- Nijland, W., Addink, E. A., De Jong, S. M., & Van der Meer, F. D. (2009). Optimizing spatial image support for quantitative mapping of natural vegetation. *Remote Sensing of Environment*, 113, 771–780.
- Ouergemmi, 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.
- Pax-Lenney, M., & Woodcock, C. E. (1997). The effect of spatial resolution on the ability to monitor the status of agricultural lands. *Remote Sensing of Environment*, 61, 210–220.
- Pearson, R. K. (2002). Outliers in process modeling and identification. *IEEE Transactions on Control Systems Technology*, 10(1), 55–63.
- Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences*, 30, 683–691.
- Raclot, D., & Albergel, J. (2006). Runoff and water erosion modelling using WEPP on a Mediterranean cultivated catchment. *Physics and Chemistry of the Earth*, 31, 1038–1047.
- Rahman, A. F., Gamon, J. A., Sims, D. A., & Schmidts, M. (2003). Optimum pixel size for hyperspectral studies of ecosystem function in southern California chaparral and grassland. *Remote Sensing of Environment*, 84, 192–207.
- Roberts, D. A., Quattrochi, D. A., Hulley, G. C., Hook, S. J., & Green, R. O. (2012). Synergies between VSWIR and TIR data for the urban environment: An evaluation of the potential for the Hyperspectral Infrared Imager (HyspIRI) Decadal Survey mission. *Remote Sensing of Environment*, 117(2012), 83–101.
- Sanchez, P., Ahamed, S., Carré, F., Hartemink, A., Hempel, J., Huising, J., et al. (2009). Digital soil map of the world. *Science*, 325, 680–681.
- Schaaf, A. N., Dennison, P. E., Fryer, G. K., Roth, K. L., & Roberts, D. A. (2011). Mapping plant functional types at multiple spatial resolutions using imaging spectrometer data. *GIScience & Remote Sensing*, 48(3), 324–344.
- Schwanghart, W., & Jarmer, T. (2011). Linking spatial patterns of soil organic carbon to topography – A case study from south-eastern Spain. *Geomorphology*, 126, 252–263.
- 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.
- Shepherd, K. D., & Walsh, M. G. (2002). Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal*, 66, 988–998.
- Shibusawa, S., Imade Anom, S. W., Sato, S., Sasao, A., & Hirako, S. (2001). Soil mapping using the real-time soil spectrophotometer. In G. Grenier, & S. Blackmore (Eds.), *ECPA 2001. Third European Conference on Precision Agriculture, Vol. 1.* (pp. 497–508). Agro Montpellier.
- Staez, K., Mueller, A., & Heiden, U. (2013). Overview of terrestrial imaging spectroscopy missions. *Geoscience and Remote Sensing Symposium (IGARSS), 2013* (pp. 3502–3505). IEEE International.
- Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Liroy, R., Hoffmann, L., et al. (2010). Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. *Geoderma*, 158, 1–2.
- Stuffer, T., Kaufmann, H., Hofer, S., Förster, K. -P., Schreier, G., Müller, A., et al. (2007). The EnMAP hyperspectral imager – An advanced optical payload for future applications in Earth observation programmes. *Acta Astronautica*, 61(1–6), 115–120.
- Tenenhaus, M. (1998). *La régression PLS*. Paris: Editions Technip (254 pp.).
- Thorp, K. R., French, A. N., & Rango, A. (2013). Effect of image spatial and spectral characteristics on mapping semi-arid rangeland vegetation using multiple endmember spectral mixture analysis (MESMA). *Remote Sensing of Environment*, 132, 120–130.
- 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 Research*, 46, 1–16.
- 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, 59–75.
- Webster, R., & Oliver, M. A. (1990). *Statistical methods in soil and land resource survey*. Oxford: Oxford University Press (316 pp.).
- Weng, Y., Gong, P., & Zhu, Z. (2008). Soil salt content estimation in the Yellow River delta with satellite hyperspectral data. *Canadian Journal of Remote Sensing*, 34(3), 259–270.
- Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of Chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 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 pp. + 6 annexes).
- Zhang, T., Li, L., & Zheng, B. (2009). Partial least squares modeling of Hyperion image spectra for mapping agricultural soil properties. *Proceedings of SPIE – The International Society for Optical Engineering*, 7454.
- Zhang, Q. Y., Middleton, E. M., Gao, B. C., & Cheng, Y. B. (2012). Using EO-1 Hyperion to simulate HyspIRI products for a coniferous forest: The fraction of PAR absorbed by chlorophyll (fAPAR(chl)) and leaf water content (LWC). *IEEE Transactions on Geoscience and Remote Sensing*, 50(5), 1844–1852.