Digital Soil Assessments and Beyond – Minasny, Malone & McBratney (eds)
© 2012 Taylor & Francis Group, London, ISBN 978-0-415-62155-7

Using Vis-NIR hyperspectral data to map topsoil properties over bare soils in the Cap Bon region, Tunisia

C. Gomez
IRD, Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH), IRD-INRA-SupAgro, Montpellier, France

P. Lagacherie
INRA, Laboratoire d'étude des Interactions Sol Agrosystème Hydrosystème (LISAH), IRD-INRA-SupAgro, Montpellier, France

S. Bacha
Centre National de Cartographie et de Télédétection (CNCT), El Aouina, Tunis, Tunisia

ABSTRACT: The aim of this work was to examine whether Vis-NIR airborne spectroscopy could be used for mapping topsoil properties in a 300 km² Mediterranean cultivated landscape (Lebna catchment, Tunisia) that includes contrasting pedological patterns and a large proportion of bare soil surfaces. This work employed AISA-Dual Vis-NIR hyperspectral airborne data acquired with a fine spatial resolution (5 m) and fine spectral resolution (260 spectral bands from 450 to 2500 nm). Partial Least Square Regressions were applied to model the relations between soil properties and AISA-Dual spectra. The results showed that four out of the eight soil properties (clay, sand, iron, and cation-exchange capacity) were satisfactorily mapped with good precisions both for estimating local values and for capturing the spatial structures. This study highlights the complexity of the North African soil patterns and opens up the possibility of more extensive use of hyperspectral data for digital soil mapping of the successfully mapped soil properties.

1 INTRODUCTION

Among the large set of possible emerging technologies that could be used in Digital Soil Mapping (DSM), visible and near infrared (Vis-NIR, 350–2500 nm) hyperspectral imaging spectroscopy is one of the most promising methods because (i) it is derived from reflectance spectroscopy, a laboratory technique that has been proven as a good alternative to costly physical and chemical laboratory soil analysis for the estimation of a large range of soil properties (Ben-Dor & Banin, 1995); (ii) it can benefit from the increasing number of methodologies developed for Vis-NIR hyperspectral airborne imaging in soil property mapping (e.g., Selige et al., 2006, Gomez et al., 2008); (iii) it can provide a global view of the area under study at spatial resolutions appropriate for DSM (Gomez et al., under review in Geoderma); and (iv) it is particularly well adapted to Mediterranean and semi-arid areas, where bare soil surfaces are common and where dry periods allow for avoiding soil moisture perturbations of the spectrum (Lagacherie et al., 2008).

The aim of this work was to examine whether Vis-NIR hyperspectral airborne imaging spectroscopy could be used for mapping eight of the most common soil properties (clay, sand, silt, calcium carbonate, free iron, cation-exchange capacity (CEC), organic carbon (OC) and pH) over a 300 km² Mediterranean area. This work investigated the use of the partial least-square regression (PLSR) to construct the models necessary to estimate the soil properties from the Vis-NIR data. The high spatial resolution (5 m) of the imaging data used in this research is expected to provide detailed pattern recognition of the soil's heterogeneity. In addition, the large coverage (300 km²) of the imaging data used in this research is expected to provide a global view of main soil patterns.

2 MATERIAL AND METHODS

2.1 *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, Tunisia (Figure 1).

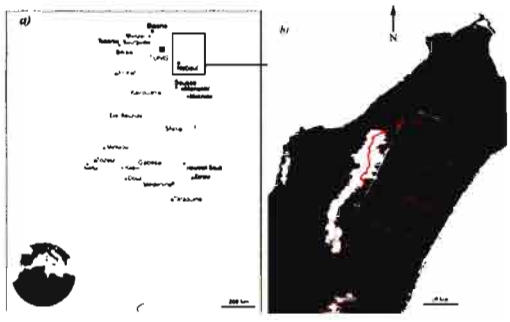

Figure 1. a) Location of the Cap Bon region in the Northern Tunisia, b) limits of the Lebna Catchment (in red) and the AISA hyperspectral image (in grey) plotted over the STRM DEM of the Cap Bon.

This 300 km² area includes the Lebna catchment (Figure 1b), which is mainly rural (>90%) and dominated by cereals in addition to legumes, olive trees, natural vegetation for breeding and vineyards. It is characterized by relief areas, with an altitude between 0 and 226 m. The main soil types are Regosols, Eutric Regosols (9.6%) predominantly associated with sandstone outcrops, Calcic Cambisol, and Vertisol predominantly formed on marl outcrops and lowlands. The southeastern region of the study area represents a more flat landscape with sandy Pliocene deposits yielding Calcosol and Rendzina.

2.2 *AISA-Dual Vis-NIR hyperspectral airborne data*

On November 2, 2010, Vis-NIR AISA-Dual hyperspectral data were acquired over the study area (12 × 25 km) with a spatial resolution of 5 m (Figure 2). The AISA-Dual airborne imaging spectrometer measures the reflected radiance in 359 non-contiguous bands covering the 400- to 2450-nm spectral domain, with 4.6 nm bandwidths between 400 and 970 nm and 6.5 nm bandwidths between 970 and 2450 nm. The instantaneous field of view (IFOV) is 24 degrees. The radiance units were converted to reflectance units using ASD spectrometer measurements of uniform surfaces (parking lots, asphalt, concrete) that were collected at the same time during the over flight. An empirical line correction method was used to calibrate each flight line to the reflectance. Topographic corrections were performed using a 30 m digital elevation model built from ASTER data and ground control points. In this study, we removed: 1) the spectral bands in the blue part of the spectral domain (between 400 and 484 nm) due to noise in these bands and 2) the spectral bands

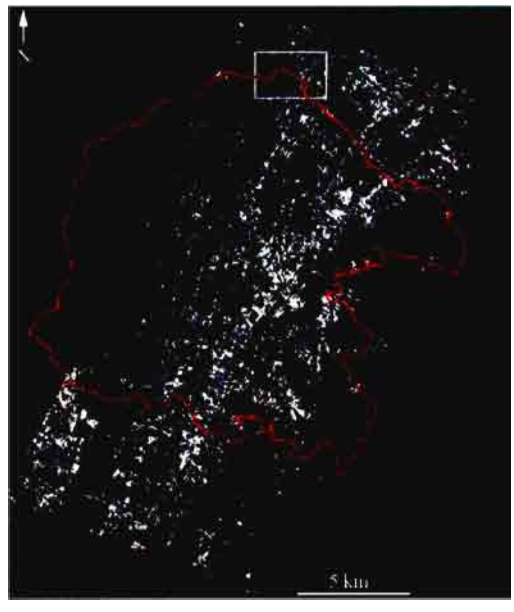

Figure 2. Location of the 129 soil samples (blue points) collected over bare soils and the Kamech catchment (full white square) plotted over the AISA-Dual image (12 × 25 km). Only the bare soils are represented (water, urban, vegetation area and mixed pixels are represented in black).

between 1339 and 1464 nm as well as between 1772 and 2004 nm due to vibrational-rotational H2O absorption bands. Consequently, 280 AISA-Dual spectral bands were retained.

When the image was acquired (November 2010), a minor part of the soil surface was covered by green vegetation, consisting mainly of olive trees, native forests, green plants and vineyards. 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 that had been visually inspected on the field. Water areas were also masked using an expert-calibrated threshold: pixels with a reflectance of less than 8% at 1665 nm were removed. Finally, urban areas were masked using a map of urban areas.

Based on the AISA-Dual data with a spatial resolution of 5 m, the bare soils represent 46.3% of our study area and 5 889 847 AISA-Dual pixels (Figure 2).

2.3 *Field sampling*

129 soil samples were collected on the Lebna catchment: 58 were collected in October 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 data acquisition in November 2010 (Figure 2). All of the samples were composed of five sub-samples collected to a depth of 5 cm at random locations within a 10 × 10 m square centered on the geographical position of the sampling plot, as recorded by a Garmin GPS instrument.

The samples were air-dried and sieved with a 2 mm sieve. After homogenizing the sample, approximately 20 g was used for soil property analysis. The determination of eight soil properties was performed using classical physico-chemical soil analysis. These properties were free iron, cation-exchange capacity (CEC), clay (granulometric fraction <2 μm), silt (granulometric fraction between 2 and 50 μm), sand (granulometric fraction between 0.05 and 2 mm), calcium carbonate ($CaCO_3$), pH and organic carbon (OC).

2.4 *Prediction models*

The Partial Least Square Regression (PLSR) was used to establish relationships between the soil variables and the AISA-Dual spectra. Before this multivariate analysis, the reflectance was converted into "absorbance" (log [1/reflectance]) and a noise reduction was achieved through standard pretreatments: a Savitzky–Golay filter with second-order polynomial smoothing and window widths of 30 nm, mean centering and variance scaling. The spectroscopic and chemometric analyses were implemented in R. The maximum number of latent variables (LV) in the PLSR was defined as 10. The optimal number of LVs was determined using prediction residual error sum of squares (PRESS) analysis, taking care to avoid under—and over-fitting. A prediction model was built for each soil property. Because a limited number of samples was available, a leave-one-out cross-validation procedure was adopted to verify the prediction capability of the PLS models. Two types of outliers were rejected from the calibration set: (i) concentration outlier, when the predicted value has a residual difference significantly greater (>2.5) than the mean of the predicted values and (ii) spectral outlier, when the sample is spectrally different from the rest of the samples. An H value of 3 based on the Mahalanobis distance, calculated on PCA-reduced data, was selected for the identification of spectral outliers.

2.5 *Models evaluation*

The prediction models were evaluated using the root mean squared error of cross-validation (RMSECV), the coefficient of determination of cross-validation (R^2cv) and the ratio of performance deviation (RPD). The RPD is the ratio between the standard deviation of the entire data set against the RMSECV. Chang and Laird (2002) defined three classes of RPD: category A (RPD > 2) describes models that can accurately predict the soil property, category B (2 > RPD > 1.4) describes models with limited predictive power and category C (RPD < 1.4) describes models that have no prediction ability.

3 RESULTS

3.1 *Soil samples study*

The soil properties exhibited contrasting variations over the study area (Table 1). The clay, sand, CEC and iron were very variable, whereas silt, $CaCO_3$, pH and carbon exhibited much smaller variations with sometimes asymmetric distributions (pH and $CaCO_3$). Most of the soil properties of these 129 soil samples were not correlated to each other, with the exception of: (i) high positive correlation between the clay content and CEC (R^2 = 0.9), (ii) high negative correlation between the clay and sand contents

Table 1. Statistical parameters of soils properties for the 129 soils samples.

	Min	Max	Mean	SD*
Iron (g/100 g)	0,3	3,3	1,6	0,5
CEC (cmol + /kg)	2,8	34,1	19,5	6,6
Clay (g/kg)	46	772	463,7	175,9
Sand (g/kg)	32	896	326	202,3
Silt (g/kg)	58	429	210,2	58,7
$CaCO_3$ (g/kg)	1	346	36,9	52,3
pH	5,8	8,8	8,2	0,5
OC (g/kg)	2,7	21,8	8,9	2,8

* *Standard deviation.*

Table 2. Correlation coefficient between the soil properties, calculated from the 129 soil samples.

	Clay	Silt	Sand	OC	pH	$CaCO_3$	CEC
Silt	0,31						
Sand	**–0,96**	**–0,56**					
OC	0,14	0,21	–0,18				
pH	0,34	0,23	–0,37	–0,11			
$CaCO_3$	–0,10	0,00	0,09	0,22	0,26		
CEC	**0,96**	0,37	**–0,94**	0,22	0,34	–0,09	
Fer	**0,77**	0,31	**–0,76**	0,07	0,08	–0,25	**0,71**

(R^2 = 0.9) and (iii) positive correlation between iron and sand contents as well as between iron and clay contents (R^2 = 0.6) (Table 2). The ranges of clay, sand, and iron contents and CEC are large (Table 1) and exhibit a centered normal distribution. Whereas the ranges of the four other properties were quite small (Table 1) and exhibited Poisson distribution.

3.2 *Prediction models results*

PLSR-based prediction models were built using the 129 AISA-Dual spectra corresponding to the location of the soil samples collected over bare soils (Figure 2). Two spectral outliers were identified and removed, and the number of concentration outliers depended on the soil property and varied between 0 and 7. The elimination of outliers from the soil database modified significantly the concentration range of the soil properties for $CaCO_3$, silt, pH and OC (Table 3).

Correct prediction models (category B), with R^2 and RPD values greater than 0.6 and 1.4 respectively, were obtained for four soil properties: iron, CEC, clay and sand contents (Figure 3). The prediction models were inaccurate for silt, $CaCO_3$, pH and OC, with R^2 values less than 0.35 (Figure 3).

3.3 *Predicted maps*

The prediction of soil properties was performed for all the bare soils of the AISA-Dual image. Only the soil properties, for which correct local predictions (Figure 3) were obtained, were mapped. Thus we created digital maps for four soil properties: free iron, CEC, clay and sand.

The predicted clay map of the entire study area showed a complex regional soil pattern (Figure 4), with predominant variations in lithology. Differences both in values and soil patterns appeared between the Pliocene area, located in the southeast corner of the image, and the Miocene area, covering the rest of the image. Pliocene area exhibits low and weakly variable topsoil clay contents, whereas the Miocene area shows a large range of clay content values. Variations within the Miocene area are also visible. They follow the geological pattern formed by the alternating sandstone and marl outcrops, yielding low (blue) and high (red) values of clay content, respectively (Figure 4). The soil patterns vary across the southeast/northwest direction, with a decreasing distance between successive sandstone outcrops and the occurrence of a large sandstone outcrop in the middle. The deposition of sandy material from the erosion of sandstone areas in the valleys that

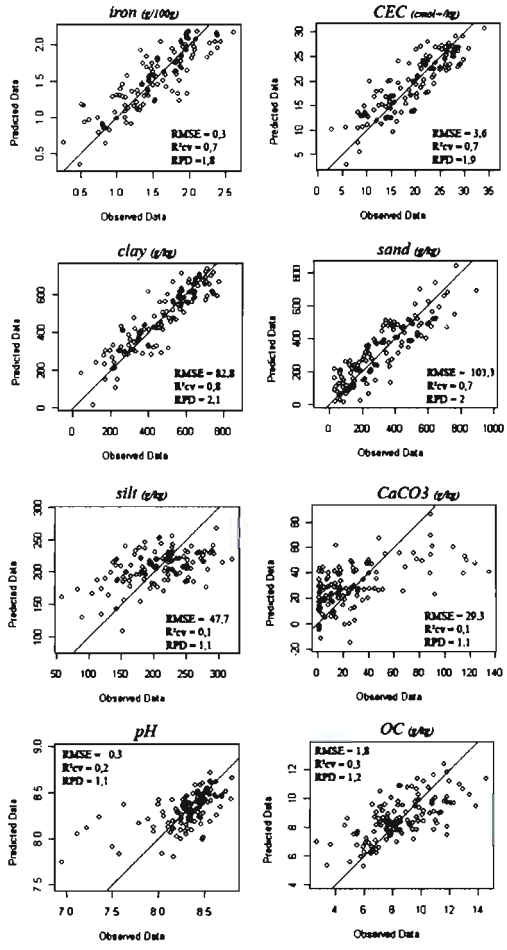

Figure 3. Plots of observed versus predicted soil properties obtained by PLSR using AISA-Dual spectra of the 129 soil samples location.

Table 3. Statistical parameters of soils properties for the soil samples used in the PLSR models after outlier removal.

	Outlier#	Min	Max	Mean	SD*
Iron (g/100 g)	7	0,26	2,62	1,5	0,48
CEC (cmol + /kg)	0	2,8	34,1	19,5	6,6
Clay (g/kg)	1	46	772	467	176
Sand (g/kg)	3	32	896	327	204
Silt (g/kg)	3	58	321	206	51,3
$CaCO_3$ (g/kg)	5	1	135	29,3	31
pH	6	6,9	8,8	8,3	0,3
OC (g/kg)	2	2,7	14,6	8,6	2,2

Number of concentration outliers.

* *Standard deviation.*

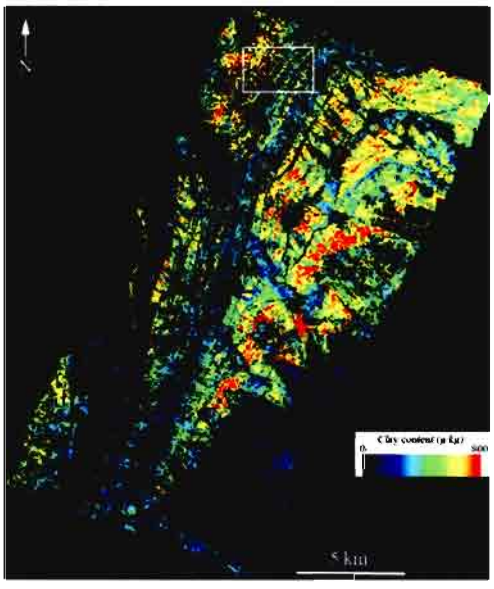

Figure 4. Clay content predicted over bare soils from AISA—Dual spectra (black areas correspond to mixed surfaces). The white square delimits the Kamech catchment.

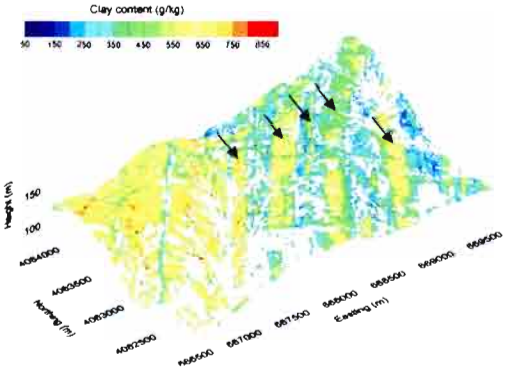

Figure 5. Zoom on predicted clay content map over bare soils of the Kamech catchment (grey areas correspond to mixed surfaces). Black arrows represent clay-rich areas. Coordinates are in UTM WGS 84.

are perpendicular to the outcrops add to the complexity of the regional soil pattern.

The predicted clay map over the Kamech catchment (Figure 5) is representative of the alternating sandstone and marl outcrops (highlighted by arrows in Figure 5). Some mixed areas also appear in transition areas between these outcrops and in shoal areas. In the Northeast corner of the study area (Figure 4) similar successions of Marl and sandstone outcrops were observed with larger distances between successive outcrops.

4 DISCUSSION

The local soil property prediction models, built from AISA-Dual Vis-NIR spectra using the 129 soil samples, allowed the estimation of four out of the eight soil properties with respect to common quality indicators (Figure 3). The accuracy of the prediction model of clay contents (R^2_{cv} = 0.8, Figure 3) is in agreement with that presented by Selige et al. (2006) and Gomez et al. (2008). The accuracy of the prediction model of sand contents (R^2_{CV} = 0.7, Figure 3) is in agreement with that reported by Selige et al. (2006). The inaccuracy of the prediction models of $CaCO_3$ contents (R^2_{CV} = 0.1, Figure 3) differs from that of Gomez et al. (2008). Our lower accuracy may be due to the low $CaCO_3$ contents in the calibration data set (between 1 to 135 g/kg after removing the outliers, Table 3) as compared to $CaCO_3$ contents of Gomez et al. (2008) (between 0.26 to 472 g/kg). The inaccurate results obtained for the prediction of OC contents differ from those reported by Stevens et al. (2010). Our lower accuracy may be due to the low variability in OC content (between 2 and 14.6 g/kg after removing the outliers, Table 3) in the calibration data set compared to the OC contents of Stevens et al. (2010) (between 5 to 50 g/kg). These poorest results can also be explained by the lower number of samples used to calibrate our model (129) compared to Stevens et al. (2010) (306 samples). Finally, no predictions based on hyperspectral data for free iron, CEC, pH and silt contents are available in the literature, so literature references cannot be used to compare and evaluate our results for these soil properties.

The large coverage (300 km²) of our AISA-Dual imaging data provides a global view of the main soil patterns (Figure 4). Successions of sandstone outcrops and marl outcrops are mapped, with a decrease of the distance between these successive outcrops from East to West. A sandy area is mapped in the Southeast part of the study area, which corresponds to Pliocene. At a local scale, the high spatial resolution (5 m) of our AISA-Dual imaging data provides detailed pattern recognition of the soil's heterogeneity, in particular for the alternance of sandstone and marl outcrops.

As correlations exist between clay, CEC, iron and sand properties (Table 2), predicted maps of these four soil properties are highly correlated as well. Nevertheless, in addition to these soil properties maps, a map of the textural class could be

obtained from the synthesis of the predicted clay and sand maps.

5 CONCLUSION

This study demonstrated that Vis-NIR hyperspectral imaging data can be used to map several key topsoil properties over large areas of bare soil. In the future, this new spatial information on topsoil properties should be used in Digital Soil Mapping both for generating complete maps of soil properties (Ciampalini et al., Submitted in DSM2012) and for improving the digital soil mapping of related subsoil properties. Moreover, diverse surface conditions including partially vegetated surfaces should be considered and treated to increase the surface of key soil properties mapping. A first way could be to use source separation methods as shown by Ouerghemmi et al. (2011). Finally the development of Vis-NIR hyperspectral sensors which are planned to be launched on board satellites within the next two years, such as PRISMA and EnMap, will extend the use of Vis-NIR hyperspectral imaging data in Digital Soil Mapping.

ACKNOWLEDGEMENTS

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 images for this study. This hyperspectral data acquisition was granted by IRD, INRA and the French National Research Agency (ANR) (ANR-O8-BLAN-C284-01) ». We are also indebted to Yves Blanca (IRD-UMR LISAH Montpellier), Zakia Jenhaoui (IRD-UMR LISAH Tunis) for the soil sampling in 2009 and 2010 over the Lebna catchment and to Hedi Hamrouni (DG/ ACTA Sol, Tunis) for his significant support to this study.

REFERENCES

Ben-Dor, E. & Banin, A. (1995). Near infrared analysis (NIRA) as a rapid method to simultaneously evaluate, several soil properties. *Soil Science Society of American Journal*, 59, 364–372.

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

Ciampalini, R., Lagacherie, P., Monestier, P., Walker E. & Gomez, C. (2012). Using a Vis-NIR hypersepctral image as a covariate for mapping properties in Northern Tunisia. Submitted at the DSM 2012 Conference Proceedings.

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., Lagacherie, P. & Coulouma, G. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis-NIR data. Under Review in *Geoderma*.

Lagacherie, P., Baret, F., Feret, J-B, Madeira Netto, 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.

Ouerghemmi, W., Gomez, C. Nacer, 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.

Selige, T., Bohner, J. & Schmidhalter, U. 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma*, 136, no1–2, pp. 235–244.

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.