Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy

Antoine Stevens [a,*], Thomas Udelhoven [b], Antoine Denis [c], Bernard Tychon [c], Rocco Lioy [d], Lucien Hoffmann [b], Bas van Wesemael [a]

[a] Department of Geography, Université catholique de Louvain, 3 place Pasteur, B-1348 Louvain-La-Neuve, Belgium
[b] Departement Environment and Agro-biotechnologies, Centre de Recherche Public-Gabriel Lippmann, 41 rue du Brill, L-4422 Belvaux, Grand-Duchy of Luxembourg
[c] Department of Environmental Sciences and Management, University of Liège, Campus of Arlon, 185 avenue de Longwy, B-6700 Arlon, Belgium
[d] CONVIS Herdbuch Service Élevage et Génétique-Société coopérative, B.P. 313 L-9004 Ettelbruck, Grand-Duchy of Luxembourg

ARTICLE INFO

ABSTRACT

Conventional sampling techniques are often too expensive and time consuming to meet the amount of data required in soil monitoring or modelling studies. The emergence of portable and flexible spectrometers could provide the large amount of spatial data needed. In particular, the ability of airborne imaging spectroscopy to cover large surfaces in a single campaign and to study the spatial distribution of soil properties with a high spatial resolution represents an opportunity for improving the monitoring of soil characteristics and soil threats such as the decline of soil organic matter in the topsoil. However, airborne imaging spectroscopy has been generally applied over small areas with homogeneous soil types and surface conditions. Here, five hyperspectral images acquired with the AHS-160 sensor (430 nm–2540 nm) were analysed with the objective to map soil organic carbon (SOC) at a regional scale. The study area, covering a surface of ~420 km^2 and located in Luxembourg, is characterized by different soil types and a high variation in SOC contents. Reflectance data were related to surface SOC contents of bare croplands by means of 3 different multivariate calibration techniques: partial least square regression (PLSR), penalized-spline signal regression (PSR) and support vector machine regression (SVMR). The performance of these statistical tools was tested under different combinations of calibration/validation sets (global and local calibrations stratified according to agro-geological zones, soil type and image number). Under global calibration, the Root Mean Square Error in the Predictions reached 5.3–6.2 g C kg^{-1}. Under local calibrations, this error was reduced by a factor up to 1.9. SOC maps of bare agricultural fields were produced using the best calibration model. Two map excerpts were shown, which display intra- and inter-field variability of SOC contents possibly related to topography and land management.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

The concern about the role of soil in the global carbon budget and the effects of soil organic carbon (SOC) decline on soil quality stimulated research on the development of efficient and accurate methods to assess SOC stocks and fluxes (e.g. Smith, 2004). SOC stocks are characterized by a slow temporal dynamic and a high spatial variability, which complicates the detection of changes due to environmental or human factors. This can be overcome by using a high sampling density. Since such strategy is labor-intensive, this has been done only in a limited set of well-documented areas or long-term SOC experiments, which are widely used to calibrate models (see e.g. Powlson et al., 1998). Obviously, the monitoring of key soil properties (SOC, organic N, electrical conductivity, etc.) for modeling soil processes or survey soil conditions suffers from the lack of techniques able to provide data with a high spatial and temporal resolution.

Comparatively faster than traditional measuring techniques, spectroscopy can exploit the information carried by reflectance in the visible and near-infrared (VNIR: 400–1100 nm) and shortwave infrared (SWIR: 1100–2500 nm) part of the electro-magnetic spectrum to measure soil properties. The emergence of portable and flexible spectrometers that can either be used directly in the field or mounted on aircrafts could be the solution to provide the large amount of spatial data required by soil monitoring. Spectra can now be recorded either in the laboratory or in the field (e.g. Stevens et al., 2008). They can be collected one point at a time (point spectrometry), as an array of measurements (on-line spectrometry; e.g. a sensor is mounted on a tractor and its position recorded by a GPS receiver) or as a two dimensional array of measurements (imaging spectrometry; e.g. the sensor records the surface aboard a plane or satellite).

* Corresponding author.
 E-mail address: antoine.stevens@uclouvain.be (A. Stevens).

While the first studies on the relationship between soil properties and soil spectra were qualitative in nature (eg. Stoner and Baumgardner, 1981), further investigations adopted with success a quantitative approach using multivariate statistic tools (e.g. Dalal and Henry, 1986; Ben-Dor and Banin, 1995). Nowadays, the potential of soil quantitative spectral analysis has been repeatedly demonstrated in the laboratory (e.g. Malley et al., 2004; Rossel et al., 2006). Laboratory spectroscopy is often called upon as a viable alternative for several routine agronomic soil analyses (e.g. Cohen et al., 2007). Due to its capacity to determine several soil properties simultaneously, VNIR–SWIR spectroscopy has been applied successfully to build soil quality indicators (Idowu et al., 2008) and assess other soil functions or threats (Cécillon et al., 2008). SOC is one of the main soil chromophores (Ben-Dor et al., 1999) and is correlated with other chromophores such as clays and iron oxide. Hence, SOC concentration can be generally determined with a good accuracy compared to other soil properties (see Table 1 in Rossel et al., 2006).

Early experiments such as the works of Henderson et al. (1989) and Clark et al. (1990) opened the way to the quantitative mapping of rocks, mineral and soil properties by means of airborne or satellite-borne imaging spectrometers. Ben-Dor et al. (2002) were the first to use multivariate calibration statistics to map soil characteristics from remotely-sensed data. Since then, several authors adopted the same approach to map soil properties such as C, N and clay content, pH or Cation Exchange Capacity (e.g. Selige et al., 2006; De Tar et al., 2008; Lagacherie et al., 2008). SOC content, in particular, has been often related to reflectance data acquired by airborne spectrometers. For instance, Ben-Dor et al. (2002) obtained a R^2 of 0.83 between predicted and observed values of soil organic matter (SOM) in clay soils from Israel using DAIS-7915 airborne data (400–2500 nm). Testing the performance of the CASI airborne hyperspectral sensor (400–950 nm) to map SOC contents of agricultural fields, Uno et al. (2005) and Stevens et al. (2006) obtained a R^2 of respectively 0.74 and 0.85. Selige et al. (2006) achieved slightly better results ($R^2 = 0.9$) with the HyMap sensor (420–2480 nm) while Patzold et al. (2008) obtained a R^2 of 0.74 with the same sensor and a very limited set of composite soil samples. Less satisfactory prediction models of SOM were obtained by Bajwa and Tian (2005) with the RDACS/H-3 sensor (471–828 nm; $R^2 = 0.66$) and De Tar et al. (2008) with the AVNIR sensor (429–1010 nm; $R^2 = 0.48$). SOC predictions based on space-borne sensors are probably also less accurate due to lower signal-to-noise ratio and spatial resolution, as showed by Gomez et al. (2008) with Hyperion images ($R^2 = 0.51$).

A review of the use of imaging spectroscopy for soil science application is given by Ben-Dor et al. (2008). Three major constraints to the widespread use of remote sensing for soil applications can be indentified (Ben-Dor et al., 2008): (i) atmospheric absorptions interfering with the spectral measure, (ii) low signal-to-noise ratio due to a short integration time over the target area and (iii) spatial variation in surface soil properties (e.g. soil roughness, crusting, vegetation residue, moisture) which induces a spectral variability not directly related to the property studied. These disturbing factors challenge the achievement of robust calibrations and repeatable results (Lagacherie et al., 2008; Stevens et al., 2008).

Due to the constraints mentioned above, airborne imaging spectroscopy was thus far applied over small areas (Uno et al., 2005; Stevens et al., 2006), homogeneous soil types (Ben-Dor et al., 2002) and a relatively small set of agricultural fields (Bajwa and Tian, 2005; De Tar et al., 2008) or soil samples (Selige et al., 2006; Patzold et al., 2008). We propose in this paper to test the ability of imaging spectroscopy to predict SOC over a large area (~420 km²) with different soil types and a gradient in SOC contents. Hyperspectral airborne data were related to surface SOC contents of bare cropland soils by means of 3 different multivariate calibration techniques: Partial Least Square Regression (PLSR), Penalized-spline Signal Regression (PSR) and least square Support Vector Machine Regression

(SVMR). To the best of our knowledge, the two latter multivariate calibration techniques were not yet used to analyse a soil property in the remote sensing domain. As shown by Marx and Eilers (2003), PSR may give more stable predictions under changing environmental conditions than PLSR. At large scales, this technique may therefore hold an advantage over classical PLSR calibration of which predictions may be very sensitive to uncontrollable factors affecting soil reflectance such as moisture content or roughness (Stevens et al., 2008). SVMR is a machine learning method has recently gained widespread acceptance in data-driven non-linear modelling applications. The method has several desirable characteristics, including good generalization abilities and robustness of the regression function, and the ability to deal with sparse data (Venkoba Rao and Gopalakrishna, 2009). Calibrations were run either on all samples collected in the study area (referred later to as "global calibration") or separately on different sets of samples grouped according to soil class, agro-geological regions or image number (referred later to as "local calibrations"). Then, the best multivariate model was chosen to produce SOC maps of bare agricultural fields.

2. Materials and methods

2.1. Study area

The study area consisted of a north-south transect of ~7 km width and ~60 km length (NW corner: 50°03′N 6°03′E; SE corner: 49°33′N 6°12′E), crossing 4 of the 5 agro-geological regions of the Grand-Duchy of Luxembourg (Fig. 1). The country has a temperate semi-oceanic climate with mean temperatures (Luxembourg city) ranging from 0.7 °C in January to 17 °C in July. In the Oesling region in the north both extremes are slightly lower. The mean annual rainfall is approximately 750 mm. Dominant soil types belong to the World Reference Base groups Cambisols, Luvisols, Arenosols, and Calcisols (FAO, 1998). Although of a small area (2586 km²), Luxembourg is characterized by a remarkably diverse physiogeography. The Oesling region, covering one third of the country and with a mean altitude of 450 m, is part of the Ardennes massif. This is a rather homogeneous area on a substrate of Devonian slate with predominantly shallow sandy and sandy-loam soils (Dystric Regosols and Dystric Cambisols). The most common crop rotation in the Oesling area is a 6 to 8 year-rotation with cultivation of cereals for 3 or 4 years, followed by field forage for another 3 or 4 years. The southern part of Luxembourg, the Gutland, includes (i) the Minette basin in the south characterized by red loamy/clay soils (Haplic and Gleyic Luvisols), formed by marls and sandstones covered with ferruginous sediments; (ii) central part of the Gutland (Central zone) with a substratum of Secondary sandstone and sandy soils (mostly Haplic Luvisols); (iii) the Redange-Diekirch area with soils of the loam-loess type on red sandstone and sandstone (Dystric Cambisols). This region has a more varied topography and an average elevation of 244 m. The most common crop rotation in the Gutland is a three years rotation with winter wheat, winter barley and silage maize.

2.2. Data collection and soil analysis

Ground sampling was realized on 4–9th October 2007. This time period was selected to ensure a high proportion of bare soils within the study area after harvest and ploughing of most of maize fields. Fifteen fields were retained (3–4 by agro-geological region) for intensive soil sampling (6 to 23 samples per field). Moreover, 83 additional samples were collected in 36 scattered fields (2–3 samples per field) to include, as much as possible, soil and spectral variability in the calibration models. The total amount of rainfall during the previous seven-day period before the overflight was 23.2 mm with an average air temperature of 10.3 °C (Luxembourg city). There had been no rain during at least three days before the overflight.
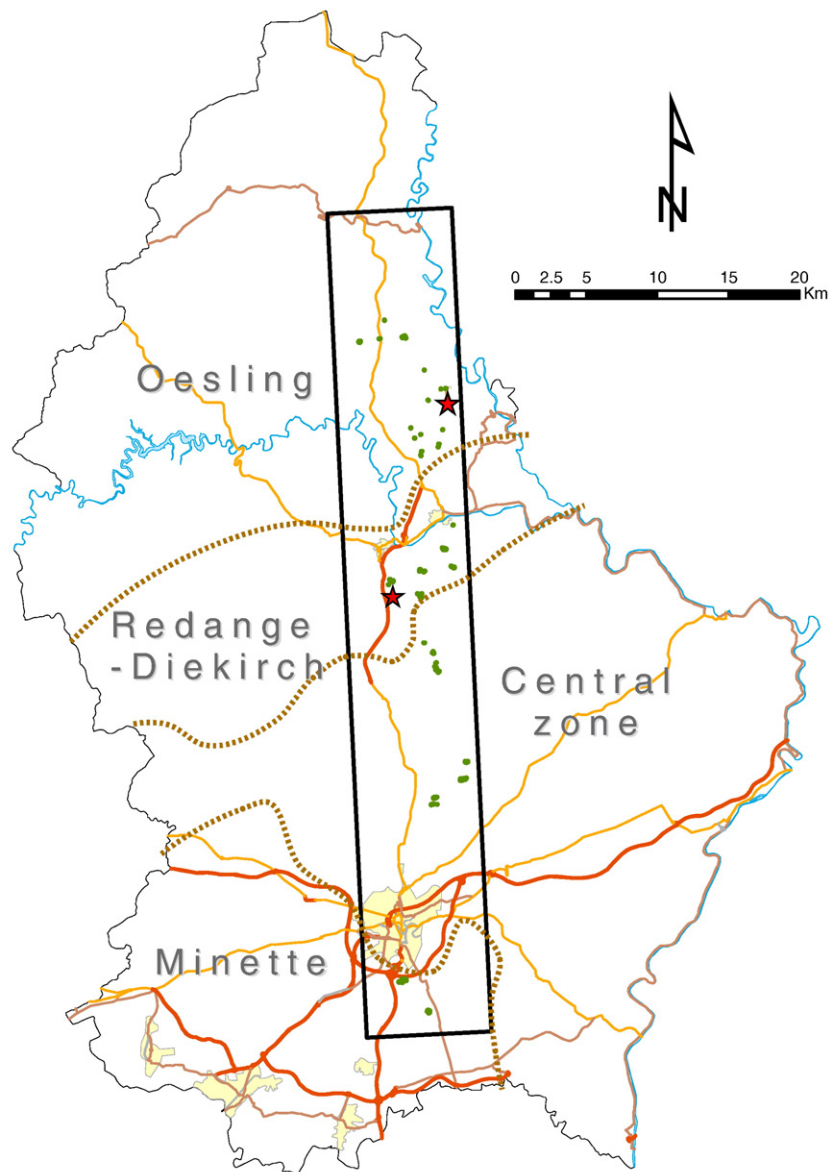
Fig. 1. Map of the Grand-Duchy of Luxembourg, with sampling points (dots), flight line (bold rectangle), and limits of agro-geological zones (dotted lines). The two stars are locations for which SOC content of bare fields was mapped (see Fig. 8).

A total of 325 soil samples were taken (49–97 samples in each agro-geological zone) for SOC analyses. These samples were composed of 10 sub-samples collected to a depth of 5 cm at random locations within a 7.5 m wide square centered on the geographical position of a sampling plot (recorded by a GPS receiver in differential mode). Soil organic carbon of air-dried and sieved (2 mm) samples was analyzed by dry-combustion with a LECO CN analyzer. Three samples containing carbonates (detected using effervescence to 1 M HCl) were removed from the dataset. For moisture analyses, 159 arbitrarily-selected soil samples were taken the day of the over flight within the sampling plot in the very first millimeters (up to 1 cm) of the soil surface and put in hermetic plastic bag. Soil moisture was determined gravimetrically. Ferrous oxide concentrations were measured on a subset of 21 samples (3–4 in each soil type) using dithionite extraction.

For each of the 325 sample plots, the corresponding soil types were extracted from a 1/100,000 soil map of the Grand Duchy of Luxembourg containing 27 soil classes. Sampling plots belong to 11 soil classes, which can be reduced to 6 main textural classes (clay, loamy-clay, loam, sandy-loam, sand, colluvium and alluvium) and one

marginal class containing only 2 samples (this unit contains soils located in the valley bottom where springs occur).

2.3. Imaging spectroscopy

The Airborne Hyperspectral Sensor 160 (AHS) onboard of a CASA 212-200 aircraft recorded, under a cloudless sky, five hyperspectral images with a swath of 1.96 km (~30% of across track overlapping), length of ~60 km and spatial resolution of 2.6×2.6 m. The AHS is a whiskbroom point scanner with an Instantaneous Field Of View of 2.5 mrad, Field Of View of 90° (750 pixels with 16% overlap) and 63 spectral bands covering the range between 430 and 2540 nm. The AHS is configured to provide 20 bands with a Full Width at Half Maximum (FWHM) of 30 nm between 430 and 1030 nm (Visible and Near Infrared; VNIR), 42 bands with a FWHM of 18 nm between 1994 and 2540 nm (Short Wave Infrared; hereafter SWIR2), complemented by an isolated band centered at 1600 nm with a FHWM of 90 nm (hereafter, SWIR1).

At-sensor radiance data with corresponding geographical positions were processed by the Central Data Processing Center of the

Vlaamse Instelling voor Technologische Onderzoek (CDPC-VITO) at Mol (Belgium), which provided geometrically and atmospherically corrected at-surface reflectance values. The geometric correction was performed by means of direct georeferencing (i.e. direct measurement of the position and orientation of a sensor using GPS in differential mode and Inertial Navigation System measurement data) over a digital elevation model. The output grid contained positions and viewing geometry parameters needed for the atmospheric correction. Atmospheric influences on radiances were then removed with the MODTRAN4 radiative transfer code (Berk et al., 1999). Visibility and water vapor parameters of MODTRAN4 were estimated from the image itself using methods of Richter et al. (2006) and Rodger and Lynch (2001). The final step consisted in the resampling of atmospherically-corrected data.

Images were manipulated using IDL language and ENVI software (ITT VIS, Boulder, CO). First, a mask was applied to the data cube to keep pixels of bare soil surface only. The soil mask was created as follows: in a first step soils were separated from green vegetation (crops and forest) and urban/concrete structures using maximum likelihood classification, following a post classification step to remove isolated pixels from the mask using a median filter (3 by 3 pixel window). A manual control of the classification results was accomplished to remove residual small roads from the soil mask. In a further step bare soil pixel were separated from soils with yellow crop residues or minor green vegetation cover using k-means clustering. Then, a filter of 3 by 3 pixel window (corresponding to the size of sampling plots) was passed through each masked image. Spectral data of the sampling sites were extracted from their corresponding positions in each masked image. Some sampling sites were covered with residual vegetation or affected by shadow from surrounding trees and were excluded from the database (19 out of 325). A total of 413 spectra were acquired. However, some plots appeared twice in the dataset due to images overlapping. Therefore, a second dataset was created by comparing the position of redundant spectra and keeping those with the lowest view zenith angle. This final dataset contained 306 spectra. A Signal-to-Noise Ratio (SNR) for each spectral band was estimated from the imagery by dividing the average reflectance values of homogeneous (black-, gray- and white-colored) targets by their standard deviation, according to the method of Ben-Dor and Levin (2000).

2.4. Statistical analyses

2.4.1. Signal preprocessing

Several mathematical preprocessing steps were applied to the raw reflectance values. These pre-treatments included: (i) absorbance ($-\log R$), (ii) 1st and 2nd derivatives, (ii) 1st and 2nd gap derivatives (Norris and Williams, 1984), (iii) Savitzky–Golay smoothing and derivatives (Savitzky and Golay. 1964), (iv) Whittaker smoothing (Eilers, 2003), (v) standard normal variate transformation and detrending (Barnes et al., 1989) or (vii) a combination of the previous. These transformations aim to decrease the noise and enhance possible spectral features linked to the property studied. For instance, Vasques et al. (2008) found that Savitzky–Golay algorithm consistently improved the ability of their models to predict soil carbon. Each pre-treatment was then calibrated to SOC with a multivariate model.

2.4.2. Calibration and validation sets

As a general rule, two thirds of the dataset have been chosen by random stratified sampling and used in the calibration set. The rest was used for validation purposes. This validation set cannot be considered to be completely independent from the calibration set due to the spatial autocorrelation in spectral data or SOC content arising between samples situated in close proximity or belonging to the same field. As demonstrated by Brown et al. (2005), a random selection of non-independent validation samples may overestimate the true accuracy in

comparison with independent samples. Such problem is likely to put in doubt the reliability of validation results if the calibration set is not sufficiently diverse in terms of SOC contents or spectral characteristics to represent their actual variation within the study area. However, the sampling of 83 locations (out of 325) in 36 scattered fields may have decreased this problem of non-independence.

Several combinations of calibration/validation sets were compared. First, a "global" calibration set containing spectral data of the entire study area was created. Sampling strata were the principal soil type (clay, silty-clay, silt, sandy-loam, sand) and agro-geological region ("Oesling", "Minette", "Central", "Redange-Diekirch"). Secondly, a series of "local" calibration/validation sets regrouped by soil type, agro-geological region and image number were constructed in order to produce a statistical model for each group separately. Sampling strata were soil types, agro-geological regions and image number. This strategy has been adopted because using spectral data from heterogeneous areas in terms of geology or soil type is known to diminish the
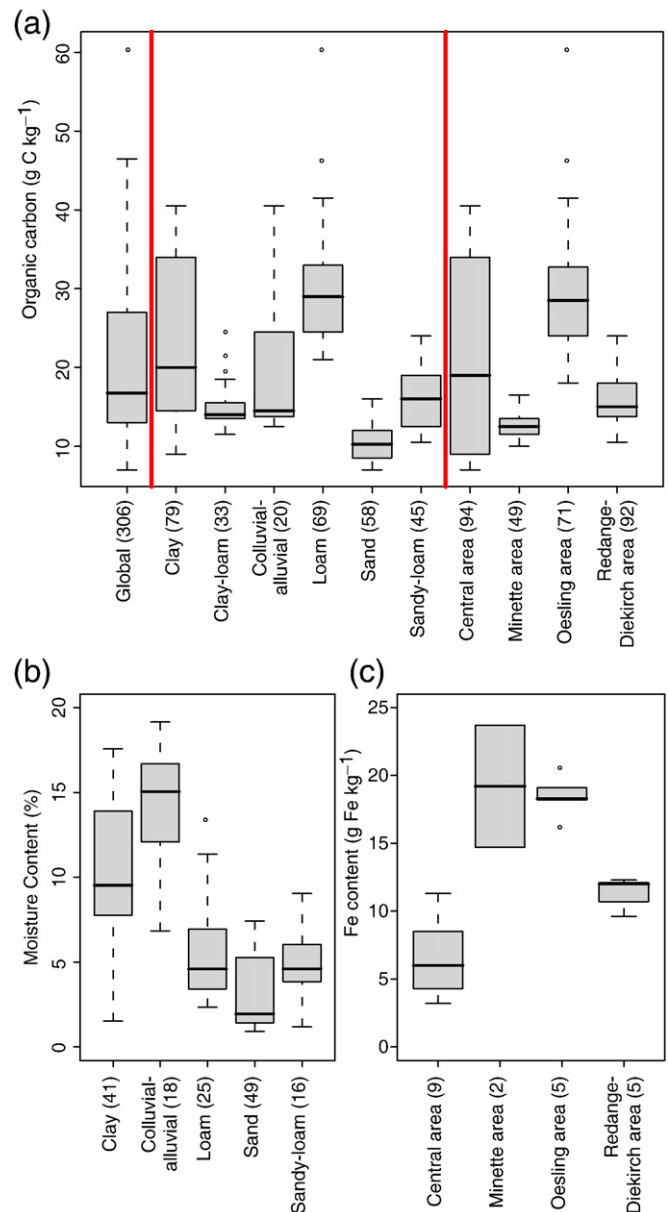

Fig. 2. Box -and-whisker plots of (a) OC content in g C kg^{-1} as a function of soil type and agro-geological region, (b) gravimetric moisture content in % as a function of soil type and (c) ferrous oxides content in g Fe kg^{-1} as a function of agro-geological regions. Number of samples is mentioned within brackets.
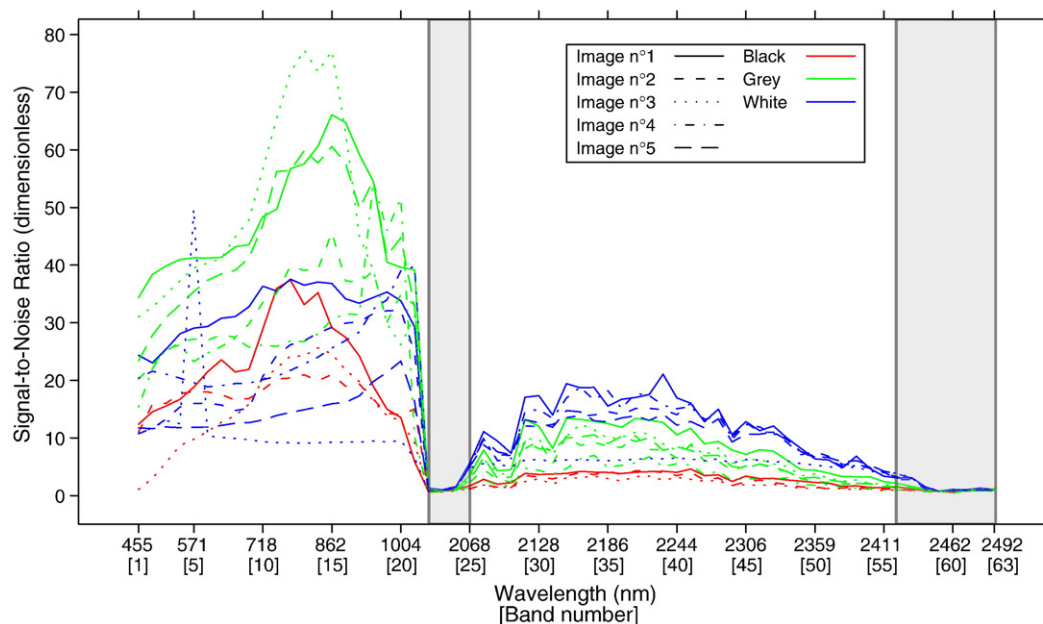
Fig. 3. Signal-to-Noise Ratio of the AHS-160 sensor (dimensionless) as a function of wavelength (band number within brackets), image number and target color (black, gray, white).

predictive ability of VNIR–SWIR spectroscopy (e.g. Udelhoven et al., 2003). The rationale behind the use of calibration based on each image is to estimate whether small differences in light or atmospheric conditions between images may influence the quality of the prediction. These variations cannot be perfectly modeled by atmospheric corrections (Ben-Dor et al., 2004), resulting possibly in non-comparable spectra between images. Each image crossed the 4 agro-geological regions (Fig. 1) and covered a large range of SOC content and soil types. Samples featuring within the overlap of neighboring images (i.e. redundant spectra) were included in these image-based calibrations.

2.4.3. Multivariate techniques

Spectral data were analyzed with 3 different multivariate techniques: (i) Partial Least Square Regression (PLSR), (ii) Penalized-spine Signal Regression (PSR) and (iii) Support Vector Machine Regression (SVMR). Statistical manipulations were carried out with the R software (R Development Core Team, 2007) and, in particular, with the PLS package of Wehrens and Mevik (2007) and PSR S-plus functions proposed in Marx and Eilers (2002). For the SVMR analysis, the SVM-KM Matlab Toolbox of Canu et al. (2005) was used.

2.4.3.1. Partial Least Square Regression (PLSR). PLSR projects predictors (X variables) and response (Y variable) into a low-dimensional space (i.e. a set of orthogonal variables called Latent Variables, maximizing the covariance between X- and Y-scores). A detailed description of the PLS algorithm is given by Wold et al. (2001). PLS regressions were fitted with the classical orthogonal scores algorithm (Wehrens and Mevik, 2007). A Principal Component Analysis was run prior to the PLSR to calculate for each spectrum the standardized Mahalanobis distance (H) to the average spectrum (Shenk and Westerhaus, 1991). Spectra with $H > 3$ (X-outliers) were removed. In order to avoid the problem of overfitting, a critical step in the algorithm is the determination of the appropriate number of Latent Variables. This is usually determined by minimizing the value of the Predictive Residual Sum of Square (PRESS) by leave-one-out cross-validation. Among other drawbacks, PRESS values do not present a clear global minimum, which compels to resort to subjective decision rules like 'the first local minimum' or 'the start of the plateau' (Faber and Rajko, 2007). Therefore, we implemented in R the randomization-test described by Faber and Rajko (2007), which allows to check the statistical significance of each component entering the model. The critical value α of the test was set to 0.1. Y-outliers were detected after

the PLSR based on the ratio between residuals and the Root Mean Square Error in the Calibration set (RMSEC), also called t-statistic. Samples with a t-statistic (in absolute value) greater than 2.5 were considered as outliers. Then, similarly to the algorithm proposed by Koshoubu et al. (2001), calibrations were run in a loop until the number of outliers reaches zero. The best pre-treatment was considered to be the one with the highest Ratio of Performance to Deviation (RPD). RPD is the ratio between Standard Deviation (SD) of the validation set divided by the RMSEP (RMSE in the validation set). Chang et al. (2001) defined 3 categories of models according to their RPD values: (i) models in category A (RPD > 2) are considered to accurately predict a given
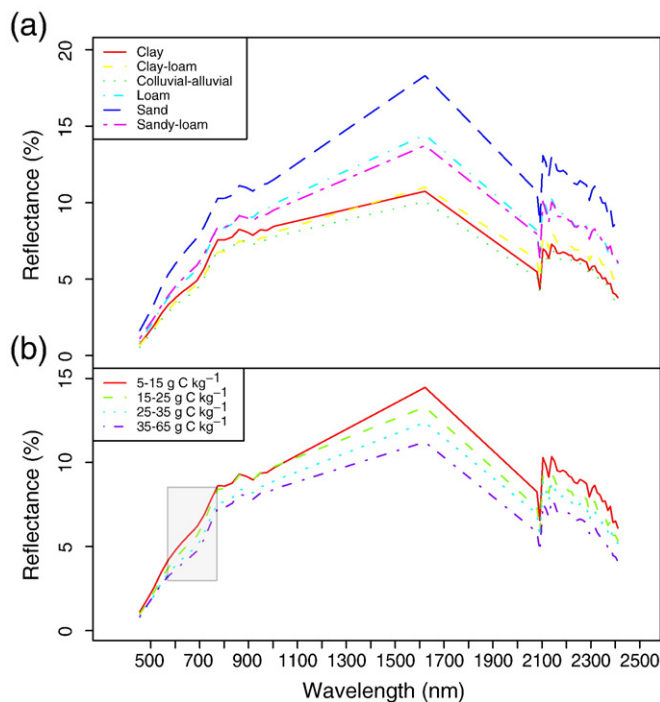

Fig. 4. Reflectance (in %) of mean spectrum as a function of (a) soil type and (b) OC class. The grey box in figure (b) highlights a small absorption feature (see text for explanations).

Table 1
Calibration and validation performance statistics for PLSR, PSR and SVMR under different spectral ranges and calibration sets.

| MC type[a]	Spectral range[b]	Calibration	N cal[c]	N val[c]	N to t[c]	Outliers[d]	SD[e]	RMSEC[f]	RMSEP[f]	RPD[g]	R^2 cal	R^2 val	Bias[h]
PLSR	VNIR–SWIR 1	Global	199	100	306	7	9.13	6.94	6.08	1.50	0.44	0.54	−0.19
		Region	198	102	306	6	10.43	3.37	4.10	2.54	0.86	0.84	0.65
		Soil	197	102	304	5	9.82	3.20	3.22	3.05	0.86	0.88	0.00
		Image	267	134	413	12	8.33	4.34	4.25	1.96	0.74	0.72	0.21
PLSR	VNIR–SWIR1–SWIR 2	Global	200	100	306	6	9.13	6.83	6.22	1.47	0.47	0.53	−0.70
		Region	198	103	306	5	10.38	2.93	4.02	2.58	0.89	0.84	0.12
		Soil	197	102	304	5	9.82	3.01	3.56	2.76	0.89	0.86	0.31
		Image	270	137	413	6	8.28	4.37	4.36	1.90	0.75	0.71	0. 17
PSR	VNIR–SWIR1–SWIR2	Global	195	100	306	11	9.13	5.92	5.75	1.59	0.57	0.59	−0.19
		Region	197	101	306	8	10.45	2.28	3.82	2.74	0.93	0.86	0.50
		Soil	196	101	304	7	9.87	2.52	3.13	3.15	0.91	0.89	−0.12
		Image	250	134	413	29	8.32	2.85	4.58	1.82	0.88	0.70	0.56
PSR	VNIR–SWIR1–SWIR2	Global	198	101	306	7	9.09	6.11	5.57	1.63	0.53	0.62	0.01
		Region	195	102	306	9	10.43	2.66	3.68	2.83	0.91	0.87	0.32
		Soil	188	101	304	15	9.87	2.54	3.20	3.08	0.91	0.88	0.01
		Image	257	135	413	21	8.20	3.56	3.99	2.06	0.81	0.75	0.20
SVMR	VNIR–SWIR1	Global	198	102	306	6	10.53	2.53	5.37	1.96	0.94	0.73	0.40
		Region	199	102	306	5	10.53	1.81	4.30	2.45	0.97	0.83	0.38
		Soil	200	101	304	3	10.72	1.96	4.75	2.26	0.96	0.79	−0.11
		Image	288	144	441	9	9.60	3.39	4.87	1.97	0.88	0.74	0.21
SVMR	VN IR–SWIR 1–SWIR 2	Global	201	102	306	3	10.53	2.13	5.35	1.97	0.97	0.74	0.55
		Region	203	102	306	1	10.53	2.01	4.71	2.24	0.96	0.79	0.34
		Soil	201	101	304	2	10.72	0.43	4.20	2.55	0.99	0.84	0.30
		Image	289	144	441	8	9.60	1.90	5.24	1.83	0.97	0.69	−0.08

[a] Multivariate Calibration type (PLSR: Partial Least Square Regression, PSR: Penalized-spline Signal Regression, SVMR: Support Vector Machine Regression).
[b] Spectral range used by the multivariate calibration model: VNIR(430–1030 nm), SWIR1 (1600 nm) and SWIR2 (2080–2411 nm).
[c] Number of samples in the calibration/validation sets and total number of samples.
[d] Total number of spectral (X) and calibration (Y) outliers.
[e] Standard Deviation in the validation set in g C kg^{-1}.
[f] Root Mean Square Error in the calibration (RMSEC) and validation (RMSEP) sets in g C kg^{-1}.
[g] Ratio of performance to deviation.
[h] Bias in the validation set in g C kg^{-1}.

property, (ii) models in category B (1.4 < RPD < 2) belong to an intermediate class and (iii) models in category C (RPD < 1.4) have no prediction ability.

2.4.3.2. Penalized-spline Signal Regression (PSR). Developed by Marx and Eilers (1999), PSR is able – like PLSR – to solve a multivariate calibration problem in which the predictors are highly correlated and their number is exceeding the number of observations. However, while in PLSR the order of predictors (i.e. wavelengths in spectroscopy) has no influence on the model, PSR forces the coefficient of the regression to vary smoothly across the wavelengths. This is done by projecting the coefficients onto a set of smooth functions (B-splines). Marx and Eilers (1999) used several published or publicly available spectral data to demonstrate that this rather simple solution to a multivariate calibration problem offers comparable performance relative to Principal Components Regression (PCR) and PLSR. Furthermore, regression coefficients are easier to interpret due to the absence of noisy features. Marx and Eilers (2002) compared the stability of PLSR, PCR and PSR in determining the composition of a mixture of water, ethanol and isopropanol measured under several temperature conditions, which caused a shifting of spectra across temperature. Their results suggests that PSR tends to be less sensitive to changing temperature (i.e. predictions are less biased) than PCR and PLSR, especially in the case of derivative spectra. In the remote sensing domain, this clearly represents an advantage since the spectral information in an hyperspectral image may not be fully consistent due either to variations in the measuring environment (atmospheric attenuation) during the recording process or spatial variability in surface properties (e.g. roughness, moisture content) within a large study area. The smoothness is controlled by imposing a penalty on the difference between adjacent B-spline coefficients (i.e. using Penalized-splines, see Eilers and Marx, 1996). Several PSR parameters must be fixed, of which (i) d, the order of penalty difference (usually between 0 and 3), (ii) the degree of B-splines and (iii) the number of intervals between knots (the point where B-splines join). After several tests, it was found that better results were obtained with $d = 3$, degree = 4, and intervals = 5. Performance indicators as well as outlier detection algorithms were equivalent to the ones used in PLSR.

2.4.3.3. Support Vector Machine Regression (SVMR). SVMR represent a different model class compared with the previous techniques since it is based on statistical learning theory (Vapnik, 1995). The most valuable properties of SVMs are their ability to handle large input spaces efficiently, to deal with noisy patterns and multi-modal class distributions, and their restriction on only a subset of training data in order to fit a (non-linear) function. The SVMR methodology is described in detail in Schölkopf and Smola (2002). In principle, an input vector X is mapped from the input domain into a higher dimensional feature space via a kernel function, where data are spread out in a way that facilitates the finding of an interpolation function (Vapnik 1995). The kernel function allows mapping the data implicitly into the feature space, where the SVMR is trained without needing to represent the feature vectors explicitly (Cristianini and Shawe-Taylor, 2003). The goal is to identify an interpolation function that is has at the same time at most deviation ε from the observations in the training data set and that is at the same time as flat as possible. This means that errors less than ε are neglected while larger deviations are not accepted. This corresponds to fitting a tube with radius ε to the training data using boundary samples, the so-called *support vectors* (SV). A trade-off has to be found between model complexity and points lying outside of the tube, which is solved using Quadratic Programming (QP) techniques (Schölkopf and Smola, 2002). The optimization requires fixing a free regularization parameter C beforehand that confines the influence of critical training patterns. As kernel we selected the Gaussian radial basis function (RBF) due to computational convenience. The RBF kernel requires only to select one free parameter (σ) beforehand that controls the smoothness properties of the interpolating function. In order to obtain a reliable set of the free parameters C, σ and ε and to identify outliers in the training data set a leave-one-out cross-validation strategy was
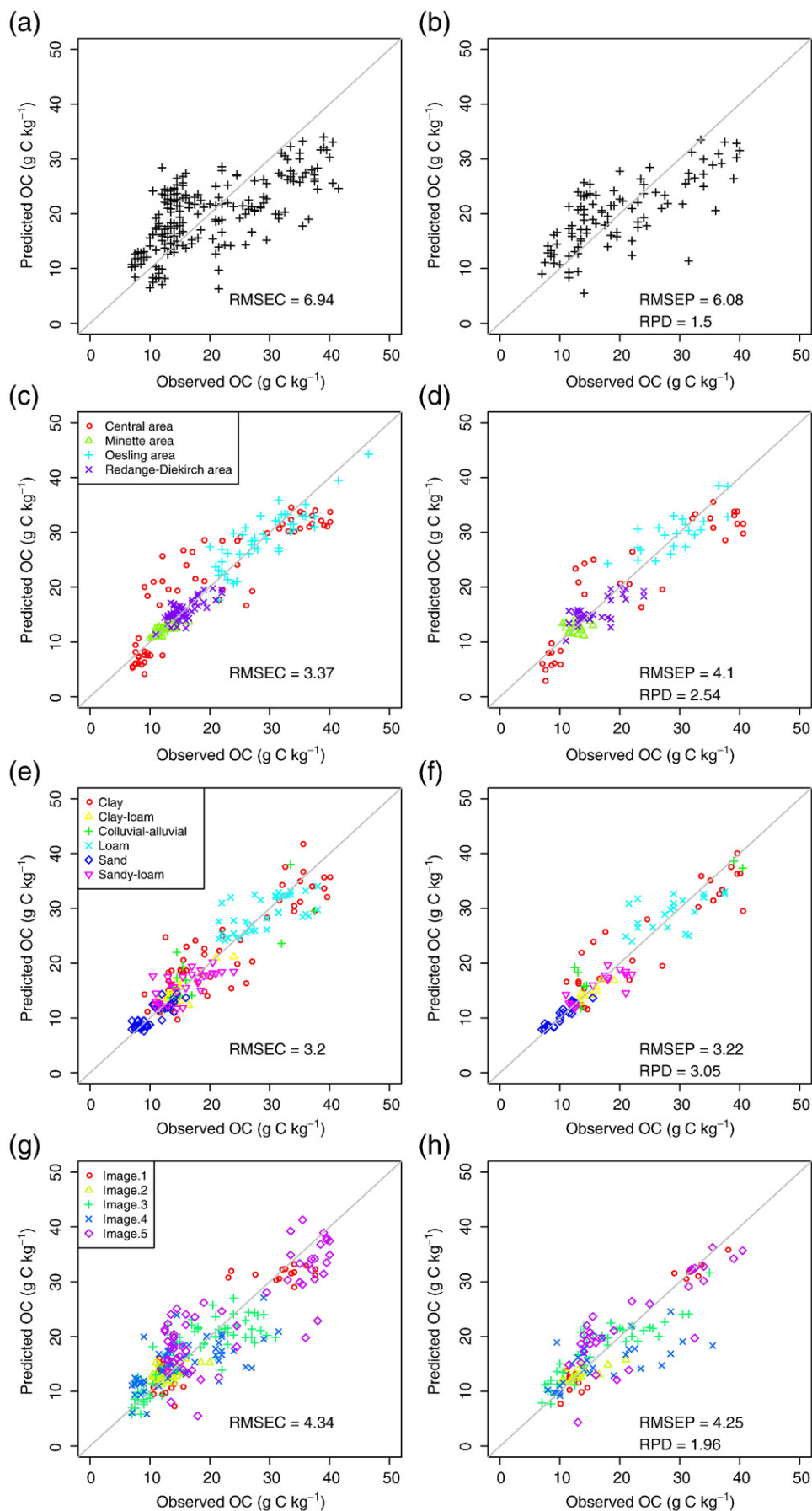
followed. A systematic grid search in the 2D-space spanned by σ and C was carried out. For Y-outlier detection, a confidence level of three standard deviations of the mean was considered. This outlier detection procedure is different from the one used in PLSR and PSR due to the higher computing time compared to other multivariate techniques. For the same reason, it was not possible to test every pre-treatments, so that calibrations were run only on raw reflectance. This may introduce inconsistencies when comparing calibrations from the different techniques. However, validation results (RMSEP, RPD) are comparable because potential outliers are not removed.

3. Results and discussion

3.1. Soil samples descriptive statistics

SOC contents varied from 7 to 61 g C kg^{-1} and differed markedly between soil types and agro-geological regions (Fig. 2a). For instance, soils of the Oesling area (North) contained, on average, more than twice the SOC content observed in the Minette area (South). Except for the Minette area, SOC contents showed also a high variability within the same region, emphasizing the need for techniques able to measure SOC with a high spatial resolution. Large disparities could also be noticed when looking at SOC content as a function of soil type. Sand, sandy-loam and clay-loam soils exhibited a relatively low variation and contained less than 25 g C kg^{-1} while clay, colluvial–alluvial and loam soils may contain up to 40 g C kg^{-1} and present a large range of SOC contents.

Moisture content, as measured during the field campaign, was relatively low (median: 5.9%, range: 0.9–19.1%) and varied greatly according to soil type (Fig. 2b). Higher moisture contents were found in clay and colluvial–alluvial soils than in sand, sandy-loam and loam soils. While this relationship between texture and moisture content is likely to be valid, this can also partly result from the sampling protocol. Theoretically, only the first few mm of the soil should have been sampled. However, this protocol is particularly difficult to follow for clay soils due to the higher degree of aggregate cohesion. Therefore, since water content increases rapidly with soil depth, this textural class may exhibit higher moisture than expected.

A strong variation in ferrous oxides was also found between agro-geological regions, with soils of the Minette and Oesling areas having much higher contents than the ones in Central and Redange–Diekirch areas (Fig. 2c). No correlation between Fe and SOC content was observed ($R = -0.17$).

3.2. Spectral data quality and analysis

The Signal-to-Noise Ratio (SNR) of AHS images varied from 10:1 to 70:1 in the VNIR–SWIR1 part of the spectrum, depending on the target area (white, gray or black surfaces, Fig. 3) while a SNR lower than 20:1 was observed for the SWIR2 part of the spectrum (bands from 22 to 63). Such pattern was also observed by Stevens et al (2008) with the same sensor. Bands from 22 to 25 and 56 to 63 showed even lower SNR due to water absorption and low atmospheric transmission and were further removed before the statistical analysis. Although the SWIR2 is characterized generally by low SNR, it was decided to keep spectral information at these wavelengths to investigate to which extent PLSR, PSR and SVMR are able to deal with noisy features. Therefore, multivariate calibrations were applied either only over the VNIR–SWIR1 part of the spectrum (430–1600 nm) or over the entire spectral range of the sensor (VNIR–SWIR1–SWIR2 430–2411 nm) for comparison.

Visual analysis of spectral data revealed that reflectance varied with soil type (Fig. 4a). Average reflectance decreased from sand, sandy-loam and loam soils to clay, clay-loam and colluvial–alluvial soils. Most of the variation occurred in the SWIR part of the spectrum rather than in the VNIR. This was also mentioned by Ben-Dor et al. (1999) after analyzing 6 representative Israeli soil samples. The observed decrease of reflectance may be partly related to mineralogy but also to differences in SOC or soil moisture content. In general, soil reflectance decreases with organic matter concentration (Stoner and Baumgardner, 1981). There is also a strong negative exponential relationship between moisture content and soil reflectance (Lobell and Asner, 2002). The influence of SOC content is highlighted in Fig. 4b showing mean reflectance according to four SOC classes. A small absorption can be distinguished at 600–750 nm as SOC content increases. Such feature is difficult to assign to any specific soil component related to SOC. However, the visible part of the spectrum (400–700 nm) is known to be a spectral region showing high absolute correlations with organic matter. For instance, based on 40 soil samples originating from several climatic zones, Bartholomeus et al. (2008) found the highest correlation between inverse of reflectance and SOC at wavelengths between 640 nm and 690 nm.

3.3. Model calibration and validation

Best pre-treatments of spectral data for each multivariate calibration technique were identified based on highest RPD values. Model performance statistics are gathered in Table 1. When considering global calibrations, RPD values oscillated between ~1.5 and ~2 depending on multivariate technique and spectral range. PLSR gave better results when restricted to the VNIR–SWIR1 spectral range while the reverse was true with PSR. No differences were observed between SVMR models using the VNIR–SWIR1 region or the entire spectral range available. According to the classification of Chang et al. (2001), the performance of the models fall in an intermediate category (1.4 < RPD < 2). This was confirmed by R^2 values in the validation set reaching 0.53–0.74 with a relatively low bias in the validation set (0.01–0.7 g C kg^{-1} in absolute value). PSR performed clearly better in global mode than PLSR, regardless the spectral range considered. The increase in model accuracy was not spectacular (difference of 0.72–1.02 g C kg^{-1} for RMSECs and 0.33–0.65 g C kg^{-1} for RMSEPs). However, the analysis of predicted vs. observed values (Figs. 5a–b and 6a–b) showed that PLSR tends to strongly underestimate high SOC values (>30 g C kg^{-1}) while this phenomenon is less severe for PSR, resulting in a lower bias (Table 1). This increased prediction error at higher SOC content was also observed by Ben-Dor and Banin (1995) and Sorensen and Dalsgaard (2005). Since PLSR and PSR are both linear techniques, this indicates that the relationship between SOC and reflectance is probably non-linear and therefore a good global calibration is difficult to obtain. Such problem can be overcome by SVMR models, which are able to approximate non-linear function mappings between multidimensional spaces (Schölkopf and Smola, 2002). While SVMR models yielded the highest RPD values in global mode (RPD = 1.97), a substantial spread in observed vs predicted SOC values above 30 g C kg^{-1} in the validation set indicates that the problem has not been entirely fixed (Fig. 7b). This phenomenon does not appear in the calibration set (Fig. 7a), which suggests an overfitting problem and emphasizes the need to validate flexible tools such as SVMR.

These latter observations strengthened the approach consisting of subsetting the spectral data based on auxiliary information (soil type, region and image number) and run local calibrations over each subset separately. This strategy allowed, depending on the multivariate calibration type, spectral range and attribute used to split the dataset, to improve considerably the accuracy of the models. The same kind of

Fig. 5. Plots of predicted versus observed OC (g C kg^{-1}) in the calibration (left column) and the validation (right column) sets as obtained by PLSR using only VNIR–SWIR1 spectral data after (a–b) global calibration, (c–d) local calibrations based on agro-geological regions, (e–f) local calibrations based on soil type and (g–h) local calibrations based on image number.
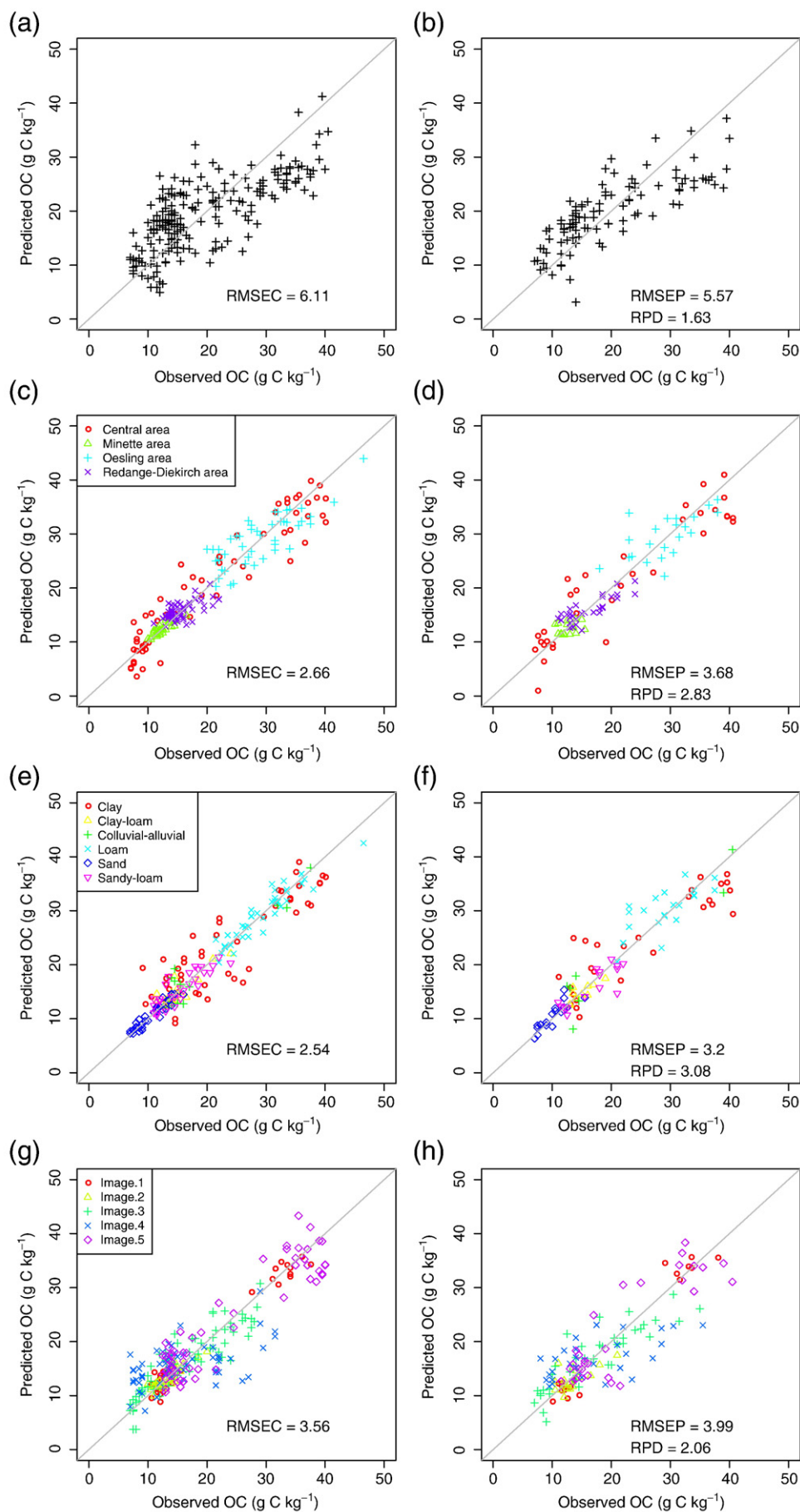
results was obtained by Van Waes et al. (2005) under laboratory conditions. They showed that dividing samples into groups according to agricultural practices or texture improved the prediction of soil organic carbon in grassland soils by 7–16%. In our case, local calibrations were able to grow RPD values by a factor 1.01–1.9, increase R^2 up to 0.7–0.89 and decrease RMSEC by 2.6–6.5 g C kg^{-1} and RMSEP by 0.8–2.9 g C kg^{-1} (Table 1). Biases in the validation sets were equivalent to the figures observed under global calibrations.

Calibrations based on soil type and agro-geological regions had RPD values above 2, indicating very good predictive models. For PLSR and PSR models, calibrations based on regions (Figs. 5c–d and 6c–d) performed a little less than those based on soil type (Figs. 5e–f and 6e–f). These differences were mainly due to the relatively poor predictions on samples located in the Central region compared to other regions (Figs. 5c–d and 6c–d). A good local calibration seems rather difficult to achieve in this area where both sandy and clayey soils – which may have very different spectral characteristics as shown in Fig. 4a – are present. On the other hand, local calibrations are of little benefit when the range in SOC content is small within a given group (e.g. Minette area or sandy soils, see Figs. 2a and 5–6). Indeed, RMSEC/RMSEP were often very large in comparison with variability in SOC content, resulting in a poor predictive power (low RPD, Figs. 5c–d and 6c–d). On the other hand, locally calibrated SVMR models were less accurate compared to PLSR/PSR models (Fig. 7c–d and e–f). In particular, Fig. 7e–f illustrate that loamy soils could not be satisfactorily modeled using soil related models. This is due to the fact that the SVMR is local in nature and requires, apart from a reliable choice of the free model parameters, a sufficient number of trainings data to achieve good generalization properties. The linear structures visible in all calibration data set in Fig. 7 and the associated very high coefficients of determination in Table 1 can be attributed to the concept of support vectors that uses a subset of training samples to define the interpolation function in a high dimensional feature space.

The improvement of local calibrations compared to global calibrations is less significant when image number was used (RPD < 2). The quality of the models is clearly degraded due to poor predictions in images numbers 4–5 for PLSR/PSR models (Figs. 5g–h and 6g–h) and image number 3 for SVMR models (Fig. 7g–h). Moreover, the total number of outliers under image-based calibrations was much larger than in other calibration types (Table 1). The number of field samples per image (56–116 per image) may have been not sufficient to produce robust calibrations considering the large variation in SOC content and soil types. This may be also explained by the positions of the plots in relation to the nadir of the sensor. Most of the outliers were found at the edges of the image with high view zenith angles, which are affected by strong back-scattering or forward-scattering effects due to the large field of view of the AHS (90°).

One can also notice that the problem of non-linearity of PLSR and PSR models at high SOC content no longer occurs when applying local calibrations (Figs. 5–6). This improvement cannot be attributed to the restricted range of SOC content within each soil type or region. For instance, clay soils as well as the Central and Oesling area showed a variability in SOC comparable to the one found in the entire data set (Fig. 2a). Nevertheless, local PLSR and PSR calibrations were both able to correctly predict high SOC content within these groups of samples (Figs. 5–6). It appears therefore that non-linearity may be related to indirect correlations of SOC with other soil physical or chemical constituents. Fig. 2b–c indicate that a relationship exists between moisture content and soil type and between ferrous oxide content and agro-geological region. Water and iron oxides are considered, along with organic matter, as the main soil chromophores (Ben-Dor et al., 1999). Moreover, absorption bands of ferrous oxides are located at about 450–700 nm (Ben-Dor et al., 1999), which may interact with

the absorption feature observed at 600–750 nm in Fig. 4b, possibly related to SOC content. Thus, the spectral variability induced by these two constituents is likely to be reduced when considering the different soil types and regions separately. Palacios-Orueta and Ustin (1999) have for instance shown that the detection of iron content and SOM with AVIRIS data depends on the relative proportion of one property to the other and on the sand fraction. Under laboratory conditions, Chang et al. (2005) produced better predictions when analyzing air-dried samples rather than moist samples.
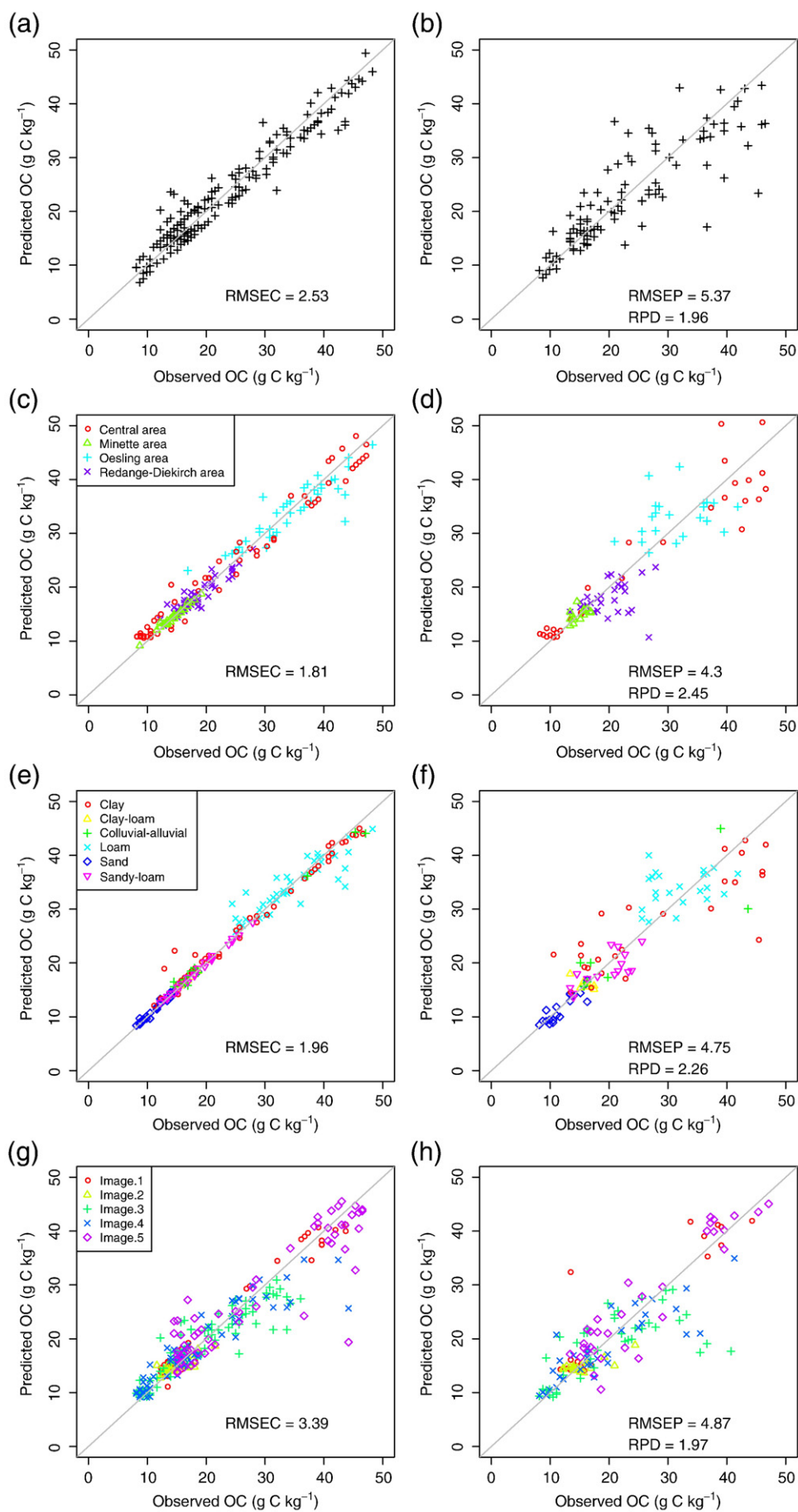
When comparing regression coefficients after processing of raw reflectance values by PLSR and PSR in global mode, it is evident that PSR computes much smoother regression coefficients than PLSR (Fig. 8). This is likely to result in a greater stability and accuracy in the predictions at locations which were under different environmental conditions as the ones represented in the existing calibration/validation set. The divergence between coefficients is minor in the VNIR region, both techniques using high coefficients (in absolute values) near 600–750 nm, suggesting that the absorption feature observed in Fig. 4b is actually related to SOC. The difference in the regression coefficient is much more pronounced in the SWIR2 part of the spectrum (2080–2411 nm). Coefficients of PLSR at these wavelengths display chaotic features indicating that noisy components have been included in the prediction model (Fig. 8). This explains why, unlike for PSR, PLSR calibrations based on the VNIR–SWIR1 region only were generally more accurate in predicting the validation set than those based on the entire spectral range. Also, despite a low SNR (Fig. 3), PSR seemed able to exploit the information hidden in the SWIR2 region to predict SOC, which may explain higher RPD values compared to the ones obtained by PLSR (Table 1). Henderson et al. (1992) found indeed that SWIR bands (1955–1965, 2215, 2265, 2285–2295, and 2315–2495 nm) gave high correlation with SOC content ($R^2 > 0.96$). These bands have the advantage to be uncorrelated with Fe content and therefore may demonstrate a higher predictive power for different parent materials (Henderson et al., 1992).

The accuracy of SOC predictions in terms of RMSEP (~3–6 g C kg^{-1}) was in the same range as those found in the literature (Uno et al., 2005; Selige et al., 2006; Stevens et al., 2006, 2008). However, RPD values were often higher in our study due to the large variation in SOC contents measured (see Fig. 2). Uno et al. (2005) obtained a RMSEP of 4.9–5.92 g OM kg^{-1} (i.e. ±2.8–3.4 g OC kg^{-1}). Using the AHS sensor, over an area in Belgian Lorraine (geologically similar to the Central area), Stevens et al. (2008) reached a RMSEP as low as 1.7 g C kg^{-1}. However, both studies were conducted within areas of small SOC variability, resulting finally in relatively low RPD values (Uno et al., 2005: 1.3–1.57; Stevens et al., 2008: 1.47). Stevens et al. (2006) achieved a RMSEP of 5.1 g C kg^{-1} and RPD of 1.86 in a contrasted environment similar to the Oesling (Ardennes massif) and Central area (Belgian Lorraine). Our best models for the Oesling and Central areas gave better results, with a RMSEP of 4.5–4.7 g C kg^{-1} and RPD of 1.8–2.8 (results not shown). Selige et al. (2006) were able to produce models with RMSEPs of 2.2–2.9 g C kg^{-1} (no RPD values mentioned).

3.4. Soil mapping

The local calibrations developed here demonstrated sufficient accuracy to be applied on a pixel-by-pixel basis over the image's soil mask. However, for illustration purposes, only two agricultural fields with spatial patterns of SOC are shown (Fig. 9). Their location within the study area is represented by a star symbol in Fig. 1. The SOC content in a single field of ~2.5 ha varied by a factor two (22 g C kg^{-1} to 46 g C kg^{-1}) and was found to correspond more or less with topography (Fig. 9a). High SOC content was mainly found in the

Fig. 6. Plots of predicted versus observed OC (g C kg^{-1}) in the calibration (left column) and the validation (right column) sets as obtained by PSR using the entire spectral range available (VNIR–SWIR1–SWIR2) after (a–b) global calibration, (c–d) local calibrations based on agro-geological regions, (e–f) local calibrations based on soil type and (g–h) local calibrations based on image number.
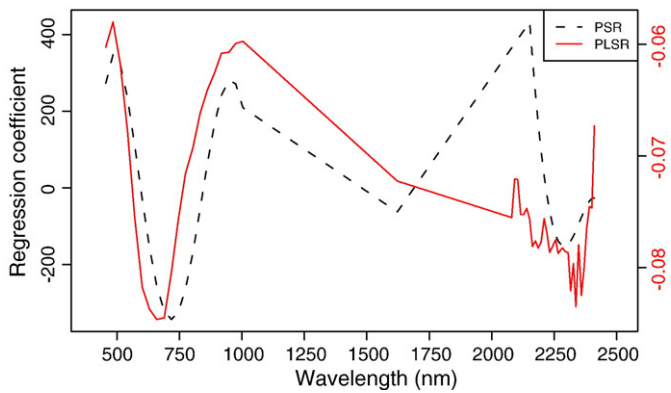
(a) RMSEC = 2.53

(b) RMSEP = 5.37, RPD = 1.96

(c) Central area, Minette area, Oesling area, Redange-Diekirch area. RMSEC = 1.81

(d) RMSEP = 4.3, RPD = 2.45

(e) Clay, Clay-loam, Colluvial-alluvial, Loam, Sand, Sandy-loam. RMSEC = 1.96

(f) RMSEP = 4.75, RPD = 2.26

(g) Image.1, Image.2, Image.3, Image.4, Image.5. RMSEC = 3.39

(h) RMSEP = 4.87, RPD = 1.97

Fig. 8. Regression coefficients obtained after processing of raw reflectance by PLSR (right axis, plain line) and PSR (left axis, dashed line).

concavities and lower SOC content on the shoulder and the plateau. Such patterns have been studied by De Gryze et al. (2008), who established a strong relationship between SOC content and transport of sediments induced by erosional processes under conventional tillage (deposition in concavities, erosion on convexities). Fig. 8b shows large differences in SOC content between two adjacent agricultural fields (up to 5–10 g C kg^{-1}). No information was available to explain this inter-field variability in SOC. However, the patterns may be related to differences in land management or land use history. The same kind of spatial pattern was detected by Stevens et al. (2006) and attributed to previous land use (former grassland). A possible effect of parcel limits can also be observed in Fig. 9b. Higher SOC contents were indeed found uphill of the field borders (south-west of Fig. 9b). This may again be controlled by soil erosion, which favours deposition of sediments at uphill position of parcel limits. Follain et al. (2006) mapped the soil thickness in a landscape dominated by croplands in Normandy (France). They showed a significantly deeper soil upslope from hedgerows that had been removed during land consolidation several decades before. Van Wesemael et al. (2006) studied a small catchment in south-east Spain that had been converted to almond groves in the 1970 s. Field borders and access roads across the slopes had created lynchets in the landscape resulting in a sedimentation rate upslope of these field borders of 21 tons ha^{-1} y^{-1}.

As a matter of fact, these two small examples show that airborne imaging spectroscopy can represent a tool to improve regional SOC inventory assessments. As shown also by Odlare et al. (2005), spatial variability of SOC is often at a scale too detailed to be captured by a coarse sampling grid. This is true at the field level, but it becomes even more critical at a landscape/regional scale. At these scales, SOC variability induced by local transfer processes (e.g. erosion) interacts with variability related to broader pedogenetic factors (climate, soil) and generates often high uncertainties in regional SOC assessments (Goidts and van Wesemael, 2007). RMSEP observed in this study (~3 g C kg^{-1}) seem still unsatisfactory in the context of SOC monitoring relative to traditional method of analysis (standard error of replicates < 1 g C kg^{-1}; Genot, pers. comm.). The accuracy still has to be improved before the technique can be exploited for SOC monitoring. Nevertheless, imaging spectroscopy allows to unravel the spatial structure of SOC at the field and regional scales and therefore will be very useful for the design of efficient sampling strategies (i.e. by avoiding pseudo-replication and selecting truly representative locations). Hence, relatively few soil cores should be taken to accurately measure SOC stocks and detect changes.
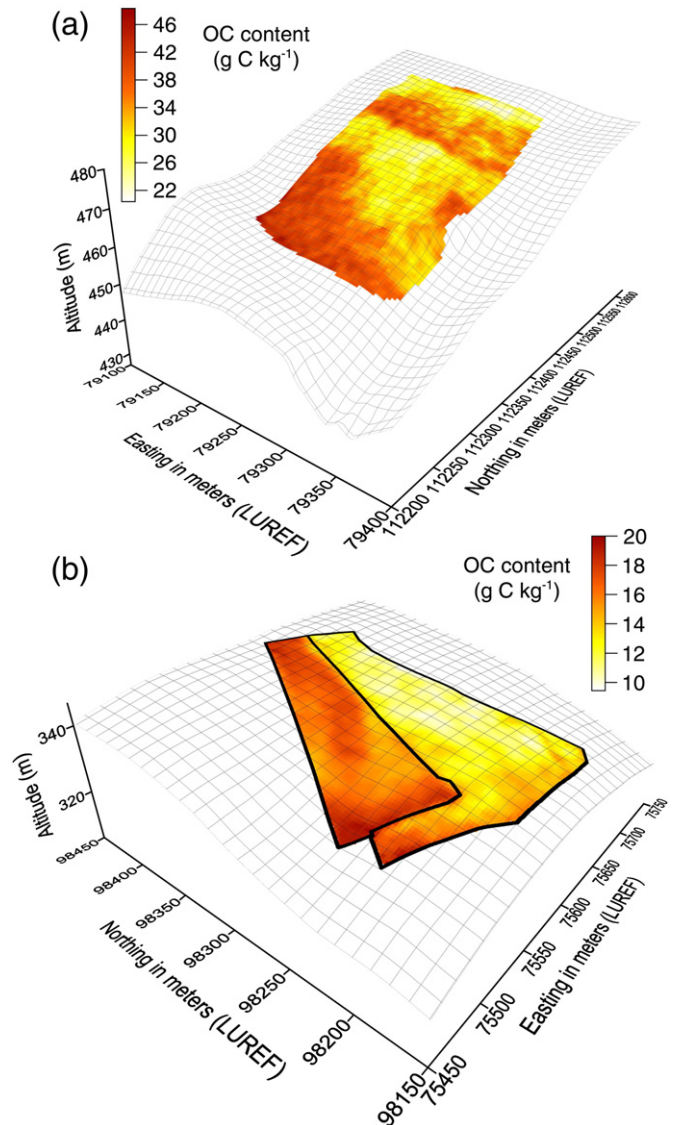

Fig. 9. Maps of freshly ploughed fields, showing (a) intra-field variability of OC (g C kg^{-1}) related to topography and (b) inter-field variability of OC (g C kg^{-1}). Coordinates are in the LUREF projection system. Their geographical positions within the study area are indicated by star symbols in Fig. 1. SOC contents have been predicted with PSR using calibrations based on agro-geological regions (see Table 1 for statistics on model accuracy).

4. Conclusion

Considering the large variation in SOC content and soil types encountered in our study area, the results presented here showed that airborne imaging spectroscopy can be an alternative to conventional analytical techniques. In general, a lack of fit at high SOC content was observed under global calibrations. PSR showed a greater ability to handle noisy spectral features, resulting in more robust calibrations than PLSR. Local calibrations based on soil types and agro-geological regions appeared to be more efficient than global calibrations, due to the correlation of the strata with important chromophores like soil moisture or ferrous oxide content. The main difference between the SVMR models and the PLSR/PSR related approaches is that the former perform better for the global data set. On the other hand, SVMR validation results are of minor quality for the soil and region related

Fig. 7. Plots of predicted versus observed OC (g C kg^{-1}) in the calibration (left column) and the validation (right column) sets as obtained by SVMR using only VNIR–SWIR1 spectral data after (a–b) global calibration, (c–d) local calibrations based on agro-geological regions, (e–f) local calibrations based on soil type and (g–h) local calibrations based on image number.

sub-models compared to the PLSR/PSR models. The analysis of field-scale SOC maps revealed spatial patterns of SOC related to topographic and management variables, which confirmed the importance of both inter- and intra-field variability in the assessment of SOC contents at larger scale.

Despite these encouraging results, the exploitation for monitoring purposes of SOC maps produced by airborne imaging spectroscopy should still be considered with caution. In particular, a post-validation over fields not covered by the existing calibration/validation sets would be necessary to assess the actual predictive ability of airborne imaging spectroscopy for SOC determination at large scales. Another critical issue is the representativeness of the models and whether a given pixel can be predicted by our models or not due e.g. to differences in surface conditions. The results of this study suggest that further research is needed on how to classify pixels according to general soil surface conditions to facilitate the development of meaningful local calibrations.

Acknowledgments

References

Bajwa, S.G., Tian, L.F., 2005. Soil fertility characterization in agricultural fields using hyperspectral remote sensing. Transactions of the ASAE 48, 2399–2406.

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variation transformation and de-trending of near-infrared diffuse reflectance spectra. Applied Spectroscopy 43, 772–777.

Bartholomeus, H.M., Schaepman, E.M., Kooistra, L., Stevens, A., Hoogmoed, B.W., Spaargaren, O.S.P., 2008. Spectral reflectance based indices for soil organic carbon quantification. Geoderma 145, 28–36.

Ben-Dor, E., Banin, A., 1995. Near-infrared analysis (NIRA) as a method to simultaneously evaluate spectral featureless constituents in soils. Soil Science 159, 259–270.

Ben-Dor, E., Irons, J.R., Epema, J.F., 1999. Soil reflectance. In: Rencz, N. (Ed.), Remote sensing for the Earth Science. Wiley and Sons Inc., New-York, USA, pp. 111–188.

Ben-Dor, E., Levin, N., 2000. Determination of surface reflectance from raw hyperspectral data without simultaneous ground data measurements: a case study of the GER 63-channel sensor data acquired over Naan, Israel. International Journal of Remote Sensing 21, 2053–2074.

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.

Ben-Dor, E., Kindel, B., Goetz, A.F.H., 2004. Quality assessment of several methods to recover surface reflectance using synthetic imaging spectroscopy data. Remote Sensing of Environment 90, 389–404.

Ben-Dor, E., Taylor, R.G., Hill, J., Dematte, J.A.M., Whiting, M.L., Chabrillat, S., Sommer, S., 2008. Imaging spectrometry for soil applications, p. 321+.

Berk, A., Anderson, G.P., Acharya, P.K., Chetwind, J.H., Bernstein, L.S., Shettle, E.P., Matthew, M.W., Alder-Golden, S.M., 1999. Modtran4 User's Manual. Air Force Research Laboratory, Hanscom, USA, p. 93.

Brown, D.J., Bricklemyer, R.S., Miller, P.R., 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. Geoderma 129, 251–267.

Cécillon, L., Cassagne, N., Czarnes, S., Gros, S., Vennetier, M., Brun, J.J., 2008. Predicting soil quality indices with near infrared analysis in a wildfire chronosequence. Science of the Total Environment 407, 1200–1205.

Canu, S., Grandvalet, Y., Guigue, V., Rakotomamonjy, A., 2005. SVM and Kernel Methods Matlab Toolbox. Perception Systèmes et Information. INSA de Rouen, Rouen, France.

Chang, C.W., Laird, D.A., Mausbach, M.J., Hurburgh, C.R., 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. Soil Science Society of America Journal 65, 480–490.

Chang, G.W., Laird, D.A., Hurburgh, C.R., 2005. Influence of soil moisture on near-infrared reflectance spectroscopic measurement of soil properties. Soil Science 170, 244–255.

Clark, R.N., Gallagher, A.J., Swayze, G.A., 1990. Material absorption band depth mapping of imaging spectrometer data using a complete band shape least-squares fit with library reference spectra. Proceedings of the Second Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop. : JPL Publication, 90–54. Pasadena, USA, pp. 176–186.

Cohen, M., Mylavarapu, R.S., Bogrekci, I., Lee, W.S., Clark, M.W., 2007. Reflectance spectroscopy for routine agronomic soil analyses. Soil Science 172, 469–485.

Cristianini, N., Shawe-Taylor, J., 2003. An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, New-York.

Dalal, R.C., Henry, R.J., 1986. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. Soil Science Society of America Journal 50, 120–123.

De Gryze, S., Six, J., Bossuyt, H., Kristof, V.O., Merckx, R., 2008. The relationship between landform and the distribution of soil C, N and P under conventional and minimum tillage. Geoderma 144, 180–188.

De Tar, W.R., Chesson, J.H., Penner, J.V., Ojala, J.C., 2008. Detection of soil properties with airborne hyperspectral measurements of bare fields. Transactions of the ASABE 51, 463–470.

Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. Statistical Science 11, 89–102.

Eilers, P.H.C., 2003. A perfect smoother. Analytical Chemistry 75, 3631–3636.

Faber, N.M., Rajko, R., 2007. How to avoid over-fitting in multivariate calibration − The conventional validation approach and an alternative. Analytica Chimica Acta 595, 98–106.

FAO, 1998. World reference base for soil resources. World Soil Resources Report 84. FAO, Rome, Italy.

Follain, S., Minasny, B., McBratney, A.B., Walter, C., 2006. Simulation of soil thickness evolution in a complex agricultural landscape at fine spatial and temporal scales. Geoderma 133, 71–86.

Goidts, E., van Wesemael, B., 2007. Regional assessment of soil organic carbon changes under agriculture in Southern Belgium (1955–2005). Geoderma 141, 341–354.

Gomez, C., Rossel, R.A.V., McBratney, A.B., 2008. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. Geoderma 146, 403–411.

Henderson, T.L., Szilagyi, A., Baumgardner, M.F., Chen, C.C.T., Landgrebe, D.A., 1989. Spectral band selection for classification of soil organic matter content. Soil Science Society of America Journal 53, 1778–1784.

Henderson, T.L., Baumgardner, M.F., Franzmeier, D.P., Stott, D.E., Coster, D.C., 1992. High dimensional reflectance analysis of soil organic-matter. Soil Science Society of America Journal 56, 865–872.

Idowu, O.J., van Es, H.M., Abawi, G.S., W., W. D., Ball, J.I., Gugino, B.K., Moebius, B.N.a., Schindelbeck, R.R., Bilgili, A.V., 2008. Farmer-oriented assessment of soil quality using field, laboratory, and VNIR spectroscopy methods. Plant and Soil 307, 243–253.

Koshoubu, J., Iwata, T., Minami, S., 2001. Elimination of the uninformative calibration sample subset in the modified UVE (Uninformative Variable Elimination)-PLS (Partial Least Squares) method. Analytical Sciences 17, 319–322.

Lagacherie, P., Baret, F., Feret, J., Netto, J.M., Robbez-Masson, J.M., 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. Remote Sensing of Environment 112, 825–835.

Lobell, D.B., Asner, G.P., 2002. Moisture effects on soil reflectance. Soil Science Society of America Journal 66, 722–727.

Malley, D., Martin, P., Ben-Dor E., 2004. Application in analysis of soils. In: Roberts, C.A., Workman, J., ReevesIII III, J.B. (Eds.), Near-Infrared Spectroscopy in Agriculture. A Three Society Monograph (ASA, SSSA, CSSA), Madison, USA, pp. 729–784.

Marx, B.D., Eilers, P.H.C., 1999. Generalized linear regression on sampled signals and curves: A P-spline approach. Technometrics 41, 1–13.

Marx, B.D., Eilers, P.H.C., 2002. Multivariate calibration stability: a comparison of methods. Journal of Chemometrics 16, 129–140.

Norris, K.H., Williams, P.C., 1984. Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in hard red spring wheat. 1. Influence of particle-size. Cereal Chemistry 61, 158–165.

Odlare, M., Svensson, K., Pell, M., 2005. Near infrared reflectance spectroscopy for assessment of spatial soil variation in an agricultural field. Geoderma 126, 193–202.

Palacios-Orueta, A., Ustin, S.L., 1999. Remote sensing of soil properties in the Santa Monica Mountains. I. Spectral analysis. Remote Sensing of Environment 65, 170–183.

Patzold, S., Mertens, F.M., Bornemann, L., Koleczek, B., Franke, J., Feilhauer, H., Welp, G., 2008. Soil heterogeneity at the field scale: a challenge for precision crop protection. Precision Agriculture 9, 367–390.

Powlson, D.S., Smith, P., Coleman, K., Smith, J.U., M. J. G., Korschens, M., Franko, U., 1998. A European network of long-term sites for studies on soil organic matter. Soil and Tillage Research 47, 263–274.

R Development Core Team, 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria3-900051-07-0. URL: http://www.R-project.org.

Richter, R., Schlapfer, D., Muller, A., 2006. An automatic atmospheric correction algorithm for visible/NIR imagery. International Journal of Remote Sensing 27, 2077–2085.

Rodger, A., Lynch, M.J., 2001. Determining atmospheric column water vapour in the 0.4–2.5 μm spectral region. Proceedings of the AVIRIS Workshop 2001. Pasadena, California, USA.

Rossel, R.A.V., 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.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least-square procedures. Analytical Chemistry 36, 1627–1638.

Schölkopf, B., Smola, A., 2002. Learning with Kernels. Massachusetts Institute of Technology. MIT Press.

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

Shenk, J.S., Westerhaus, M.O., 1991. Population structuring of near-infrared spectra and modified partial least-squares regression. Crop Science 31, 1548–1555.

Smith, P., 2004. Monitoring and verification of soil carbon changes under Article 3.4 of the Kyoto Protocol. Soil Use and Management 20, 264–270.

Sorensen, L.K., Dalsgaard, S., 2005. Determination of clay and other soil properties by near infrared spectroscopy. Soil Science Society of America Journal 69, 159–167.

Stevens, A., Van Wesemael, B., Vandenschrick, G., Toure, S., Tychon, B., 2006. Detection of carbon stock change in agricultural soils using spectroscopic techniques. Soil Science Society of America Journal 70, 844–850.

Stevens, A., van Wesemael, B., Bartholomeus, H., Rosillon, D., Tychon, B., Ben-Dor, E., 2008. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. Geoderma 144, 395–404.

Stoner, E.R., Baumgardner, M.F., 1981. Characteristic variations in reflectance of surface soils. Soil Science Society of America Journal 45, 1161–1165.

Udelhoven, T., Emmerling, C., Jarmer, T., 2003. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: a feasibility study. Plant and Soil 251, 319–329.

Uno, Y., Prasher, S.O., Patel, R.M., Strachan, I.B., Pattey, E., Karimi, Y., 2005. Development of field-scale soil organic matter content estimation models in Eastern Canada using airborne hyperspectral imagery. Canadian Biosystems Engineering 47, 9–14.

Vapnik, V.N., 1995. The nature of statistical learning theory. Information Science and Statistics. Springer, New York.

Van Waes, C., Mestdagh, I., Lootens, P., Carlier, L., 2005. Possibilities of near infrared reflectance spectroscopy for the prediction of organic carbon concentrations in grassland soils. Journal of Agriculture Science 143, 487–492.

van Wesemael, B., Rambaud, X., Poesen, J., Mark, M., Cammeraat, E., Stevens, Stevens, A., 2006. Spatial patterns of land degradation and their impacts on the water balance of rainfed treecrops: a case study in South East Spain. Geoderma 133, 43–56.

Vasques, G.M., Grunwald, S., Sickman, J.O., 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma 146, 14–25.

Venkoba Rao, B., Gopalakrishna, S.J., 2009. Hardgrove grindability index prediction using support vector regression. International Journal of Mineral Processing 9, 55–59.

Wehrens, R. and Mevik, B.-H., 2007. pls: Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR). R package version 2.1–0, URL: http://mevik.net/work/software/pls.html.

Wold, S., Sjostrom, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58, 109–130.