Soil Security 5 (2021) 100024

Contents lists available at ScienceDirect

Soil Security

journal homepage: www.sciencedirect.com/journal/soil-security

Comparison of soil organic carbon stocks predicted using visible and near infrared reflectance (VNIR) spectra acquired *in situ* vs. on sieved dried samples: Synthesis of different studies

Aurélie Cambou [a], Victor Allory [b], Rémi Cardinael [c,d,e], Lola Carvalho Vieira [a], Bernard G. Barthès [a,*]

[a] Eco&Sols, Université de Montpellier, Cirad, Inrae, IRD, Institut Agro, 34060 Montpellier, France
[b] Laboratoire Sols et Environnement, Université de Lorraine, Inrae, 54505 Vandoeuvre-lès-Nancy, France
[c] Cirad, UPR Aida, Harare, Zimbabwe
[d] Aida, Université de Montpellier, Cirad, Montpellier, France
[e] Department of Plant Production Sciences and Technologies, University of Zimbabwe, Harare, Zimbabwe

ARTICLE INFO

Keywords:
Diffuse reflectance spectroscopy
Global PLS regression
Local PLS regression

ABSTRACT

There is increasing demand for data on soil organic carbon (SOC) stock (SSOC; kgC m^{-2}), but the acquisition of such data, which relies on the determination of volumetric SOC content (SOCv; gC dm^{-3}), is often tedious or complex. Visible and near infrared reflectance spectroscopy (VNIRS) has proven useful for soil characterization, but has rarely been used for direct prediction of SOCv. The objectives of this work were: (i) to compare SOCv predictions using VNIR spectra collected in situ vs. on 2-mm sieved air-dried soil (laboratory conditions), on three sample sets separately (with in situ spectra collected differently for each set); and (ii) to assess SOCv prediction in independent validation using laboratory spectra from all sets.

Predictions of SOCv were more accurate using laboratory than in situ spectra for two sets, but not for the third set, where coarse particles content was rather high and variable. Considering the total set of laboratory spectra, predictions in independent validation (leave-one-site-out) yielded accurate SOCv and SSOC predictions (standard errors of prediction were 1.9 gC dm^{-3} and 0.36 kgC m^{-2} at 0–30 cm depth, respectively). This result was achieved using local partial least squares regression (PLSR), based on spectral neighbors, which noticeably outperformed global PLSR (which uses all calibration samples equally), as often reported when using large soil spectral libraries for independent validation.

Finally, this work demonstrated that SSOC could be quantified accurately using a VNIRS library built on archive soil samples, which offers important perspectives for SSOC accounting.

1. Introduction

Soil organic carbon (SOC) is the main component of soil organic matter, which has a crucial role in soil physical, chemical and biological functioning: for instance, it improves soil aggregation and aeration, is a reserve of nutrients for plants, and stimulates microbial activity (Lal, 2014). Moreover, SOC storage has been presented as a solution to help mitigate climate change. Indeed, soils represent the largest terrestrial pool of carbon, but according to different factors (e.g. land use, management practices), they can behave as a sink or source for atmospheric carbon (Metz et al., 2007; Eglin et al., 2010; Dignac et al., 2017). Moreover, a small change in SOC stocks (SSOC) can have major effect on reducing or enhancing greenhouse gas concentrations in the atmosphere (Baldock et al., 2012). In this context, the 4 per 1000 initiative (launched during the COP21 in 2015) aims to support soil management practices that contribute to long-term SOC storage (Dignac et al., 2017; Minasny et al., 2017; "4 per 1000″ Initiative, 2018).

However, some progress is required to be able to quantify SOC efficiently. Indeed, quantification of SOC at field, region or country scale requires data on SSOC (in kgC m^{-2}), which is calculated from volumetric SOC content (SOCv, in gC dm^{-3}) in the different depth layers of the soil profiles considered. Sample SOCv is not determined directly but calculated from gravimetric SOC content (SOCg, in gC kg^{-1}) and bulk density (Db, in kg dm^{-3}). Conventional determination of SOCg based on Dumas

* Corresponding author.
E-mail address: bernard.barthes@ird.fr (B.G. Barthès).

https://doi.org/10.1016/j.soisec.2021.100024
Received 15 April 2021; Received in revised form 18 November 2021; Accepted 28 November 2021
Available online 30 November 2021

combustion or sulfochromic oxidation requires reagents, generates wastes, and might be tedious for carbonated samples (Pansu and Gautheyrou, 2006). Conventional Db determination requires sampling a known volume of intact soil, which is labor-intensive; while alternative approaches based on radiation transmitting or scattering are often complex and expensive (e.g. X-ray tomography or gamma-ray attenuation; Helliwell et al., 2013; Casanova et al., 2015; Lobsey and Viscarra Rossel, 2016; Al-Shammary et al., 2018). Thus SOCv determination is not trivial task.

Near infrared, visible and near infrared, and mid infrared reflectance spectroscopy (NIRS, VNIRS and MIRS, respectively) have received much attention in recent decades for their ability, coupled with chemometrics, to characterize a wide range of soil properties, SOCg especially, in laboratory conditions or even in situ (Stenberg et al., 2010; Gholizadeh et al., 2013; Angelopoulo et al., 2020; Barthès and Chotte, 2020). However, attempts to determine Db using these approaches have often been disappointing, because Db is mostly related to structural pore-space condition and would not have strong absorbance features in the infrared range (Minasny et al., 2008; Moreira et al., 2009; Bellon--Maurel and McBratney, 2011; Veum et al., 2015). Moreover, despite the recommendation of Bellon-Maurel and McBratney (2011), few studies have attempted to use infrared spectroscopy for direct SOCv quantification: this has been tested mostly at local scale, with VNIR spectra collected in situ on intact or disturbed soil cores (Roudier et al., 2015; Cambou et al., 2016; Allo et al., 2020) or on pit walls (Allory et al., 2019), without independent validation. Interestingly, Allory et al. (2019) and Allo et al. (2020) also achieved good SOCv prediction using spectra collected on 2-mm sieved air-dried samples.

Though a range of multivariate regression procedures have been used for fitting infrared spectra to soil properties, partial least squares regression (PLSR) is by far the most common currently (Angelopoulo et al., 2020; Barthès and Chotte, 2020; Barra et al., 2021). It has most often been used in global calibration, where all calibration samples contribute equally to model development. Yet, several studies have demonstrated the interest of local calibration, based on spectral neighbors, especially when using large calibration databases (Genot et al., 2011; Nocita et al., 2014; Barthès et al., 2020).

This work used samples and data from three previous studies, and its objectives were:

- to compare VNIRS predictions of SOCv using spectra collected on 2-mm sieved air-dried samples (laboratory conditions) vs. in situ according to different acquisition procedures (on intact or disturbed soil cores, or on pit walls), through global calibration on representative samples (as in initial studies);
- to optimize VNIRS prediction of SOCv (then SSOC) in independent validation within the total set of sieved dried samples, using global or local PLSR;
- to examine whether SOCv predictions could be indirect, as a result of SOCg (or Db) predictions and correlations between SOCv and SOCg (or Db).

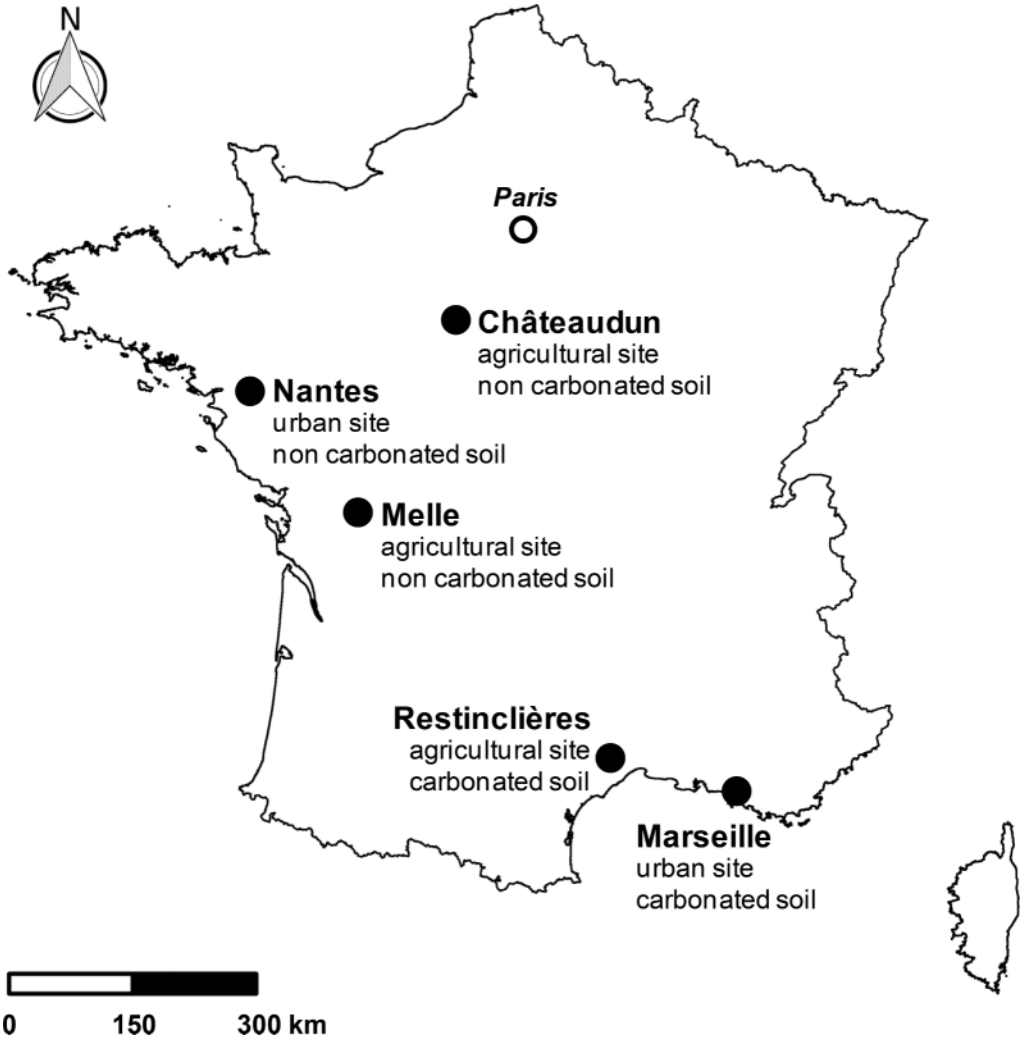

Fig. 1. Location of the studied sites, in France.

2. Materials and methods

2.1. Sites

The samples were collected during three different studies: the first one in Restinclières (RES) in late May 2013; the second one in Châteaudun (CHA) and Melle (MEL) in March and April 2014, respectively; and the last one in Marseille (MAR) and Nantes (NAN) in February and May 2017, respectively (Fig. 1). Mean annual temperature is 11–12°C in CHA, MEL and NAN, and 14–15°C in RES and MAR. Mean annual rainfall is 500–600 mm in CHA and MAR and 700–900 mm in RES, MEL and NAN, respectively.

The RES experimental site is located in Prades-le-Lez (43°43′N, 04°01′E), in southern France (15 km north of the city of Montpellier). The soil is a carbonated silty Fluvisol (IUSS Working Group WRB, 2014). The site comprised (i) an agroforestry system (3.2 ha) that associated 2-m wide lines of 18-yr-old walnut trees (110 trees ha^{-1}) covered by spontaneous vegetation, and 11-m wide alleys generally cultivated with ploughed and fertilized durum wheat; and (ii) an adjacent treeless cropland used as control (1.4 ha), with the same annual crop and management as in the agroforestry alleys. More information on this site can be found in Cardinael et al. (2015).

Both CHA and MEL sites were famers' fields, which also included an agroforestry part (ca. 4–5 ha), with 2-m wide lines of 6-yr-old walnut trees (35 trees ha^{-1}) covered with sown grasses and cultivated alleys 26- or 29-m wide; and an adjacent control plot with the same annual crop but without tree lines (ca. 1–2 ha). Soil at both sites is a silty loam Luvisol (IUSS Working Group WRB, 2014). The CHA site (48°06′N, 01°18′E) is close to the little city of Châteaudun (130 km southwest of Paris), and the MEL site (46°12′N, 00°10.5′W) close to the little city of Melle (250 km southwest of Châteaudun). In CHA, alleys were under wheat/rapeseed rotation ploughed one year out of three and harrowed the other years, with crop residue return. In MEL, alleys were under wheat/rapeseed/sunflower rotation ploughed every year, with crop residue removal but manure application, and winter cover crop (radish or mustard) before sunflower cultivation. More information on these sites can be found in Cambou et al. (2016).

Marseille (MAR) is a large city located in the south of France, along the Mediterranean Sea (43°18′N, 05°23′E). Two subsites were studied in MAR, both on calcareous Anthrosols (IUSS Working Group WRB, 2014), with variable texture but often rich in coarse particles (> 2 mm): the Borély Park (MAR1), 17-ha large, 140-yr old, covered by lawn and managed tree groves; and the Ste Marthe wilderness (MAR2), 4.4-ha large, which has long been fallow (> 20 years), with the largest part covered by grass and some trees, beside a grove. Nantes (NAN) is another large city, located in the west of France, about 50 km from the Atlantic Ocean (47°13′N, 01°33′W). Two subsites were also studied: the Cemetery Park (NAN1) in the north of the city, 50-ha large, 60-yr old, with a part covered by well managed vegetation, mainly lawn, trees and hedges, and the other by spontaneous grassland and urban woodland, on Anthrosol (IUSS Working Group WRB, 2014) developed on loess and mica-schist; and an urbanized area close to the railway station in the city center (NAN2), 0.3-ha large, mainly covered by car parks and pavement, on Technosol (IUSS Working Group WRB, 2014) developed on sandy backfills and sealed for more than 50 years. More information on MAR and NAN sites can be found in Allory et al. (2019).

2.2. Soil sampling

At agroforestry sites (RES, CHA and MEL), sampling followed a protocol designed to catch SOC variations according to distance to tree lines and trees: replicated transects were defined perpendicular to tree lines (in front of trees and between trees) and on tree lines between trees. On adjacent control plots without trees, sampling points were located at the ends of replicated transects of the same dimensions. In this way, 40, 60 and 93 sampling points were defined on tree lines, in alleys and in adjacent control plots in RES, 12, 24 and 12 in CHA, and 12, 18 and 6 in MEL, respectively (sampling density varied according to each study).

In RES, intact soil cores were collected down to 2 m at each sampling point using a motor-driven micro caterpillar driller (8.5-cm diameter probe), from which representative core segments originating from 0 to 10, 10–30, 30–50, 50–70, 70–100 and 160–180 cm depths were selected (according to a procedure presented in Section 2.5.1). They were analyzed for both volumetric SOC content (SOCv, in gC dm^{-3}) and VNIRS (Cardinael et al., 2015). For the present work, 167 samples from this selection were available.

In CHA and MEL, intact soil cores were collected at each sampling point at 0–10, 10–20 and 20–30 cm depth in small pits using 0.5-L cylinders, for SOCv determination and VNIR spectrum acquisition in laboratory conditions, on a total of 144 samples per site. Moreover, disturbed soil cores were collected at the same depths at three locations about 0.4 m around each cylinder sampling location, using a handheld auger, for in situ VNIR spectrum acquisition (initially the study aimed at predicting SOCv from in situ spectra acquired on handheld auger cores, which are less informative on SOCv than cylinder cores but much less tedious to collect; Cambou et al., 2016).

In MAR and NAN, three or four pits were dug down to 70 to 160 cm depth in each of the four studied subsites, using an excavator, and soil profiles were divided into horizons according to macromorphology (color, structure, etc.). Soil samples were collected by horizon: in total, 137 samples were collected using a knife, for laboratory analyses, and 49 using 0.25-L cylinders, for Db determination (Allory et al., 2019).

2.3. Conventional determinations

Intact soil cores collected with cylinders or motorized driller were air-dried, gently broken up and sieved to 2 mm. Then, one aliquot was weighed, oven-dried at 105°C for 48 h and weighed again, to determine the dry mass of sample fine earth (< 2 mm), which was divided by sample volume to calculate Db (kg fine earth dm^{-3} sample). Sample volume was cylinder volume, or core segment volume in the case of motorized driller (i.e. corer tube internal section multiplied by segment length). And another aliquot was finely ground (< 0.2 mm) for SOC analysis, except for MAR and NAN, where SOC was analyzed on 0.2-mm ground aliquots of 2-mm sieved, air-dried samples collected using a knife (cf. 2.2).

For RES samples, carbonates were removed by HCl fumigation; then decarbonated samples were analyzed for gravimetric SOC (SOCg, in gC kg^{-1} fine earth) with a CHN elemental analyzer (Carlo Erba NA 2000, Milan, Italy; Cardinael et al., 2015). Samples from CHA and MEL were carbonate-free, so were directly analyzed for SOCg using the same elemental analyzer (Cambou et al., 2016). For MAR and NAN samples, SOCg was analyzed according to two procedures: for carbonate-rich samples, by difference between total carbon determined with a CHN elemental analyzer (Flash EA 1112, CE Instruments, Rhodano, Italy) and soil inorganic carbon determined by volumetric calcimetry using a Bernard calcimeter; for carbonate-poor samples, using the same CHN elemental analyzer on samples that had been decarbonated after successive HCl additions (Allory et al., 2019).

Volumetric SOC content (SOCv, in gC dm^{-3}) was calculated as the product of SOCg content of fine earth and Db, weighted by the 105°C oven-dry mass proportion of fine earth (Hobley et al., 2018). In all rigor, units for SOCg and Db are gC kg^{-1} soil < 2 mm and kg soil < 2 mm dm^{-3} soil, but throughout the text they have been simplified as gC kg^{-1} and kg dm^{-3}, respectively. Then SOC stock (SSOC, in kgC m^{-2} for a given soil depth) was calculated by summing up SOCv, weighted by the thickness of respective soil layers, over the profile.

2.4. Spectrum acquisitions

On the one hand, VNIR spectrum acquisition was carried out in situ,

according to different procedures for (i) RES, (ii) CHA and MEL, and (iii) MAR and NAN; and on the other hand, in laboratory conditions, with the same procedure for all samples. All spectra were acquired in diffuse reflectance from 350 to 2500 nm at 1 nm interval using a portable spectrophotometer ASD LabSpec 2500 (Analytical Spectral Devices, Boulder, CO, USA). This instrument is equipped with a contact probe, which delivers the light to the sample then collects the reflected signal and transmits it to the spectrometer. After every spectral acquisition, the window of the contact probe was cleaned with lens paper and ethanol. A disk made of Spectralon (compressed polytetrafluoroethylene powder) was used as white reference standard, with zero absorbance, and its reflectance was measured about every 10 acquisitions. Each reflectance spectrum provided by the spectrometer resulted from the averaging of 32 co-added scans. Spectra were recorded as absorbance, which is the decimal logarithm of the inverse of reflectance [$\log_{10}(1/\text{reflectance})$].

In RES, in situ spectra were acquired on intact soil cores, after their surface was refreshed using a knife, at four locations on each sample (which actually was a core segment); and the four spectra collected on each sample were averaged (Cardinael et al., 2015). So spectra were collected on the samples used for SOCv determination. In CHA and MEL sites, in situ spectra were acquired on the outer side of disturbed soil cores collected using a handheld auger, at two depths per 10-cm soil layer and at three locations around each cylinder sampling location; then the six spectra collected at each 10-cm depth layer per cylinder sampling location were averaged (Cambou et al., 2016). So in situ spectra were not collected on the cylinder cores used for SOCv determination, but on three disturbed soil samples collected 0.4 m around. In MAR and NAN, two in situ spectra were collected on the refreshed surface of pit walls, close to the place where each of the 137 sample had be collected with a knife, and they were then averaged (Allory et al., 2019). So spectra were not collected on the samples used for SOCv determination, but on samples collected just around.

Then spectra were acquired in laboratory conditions, on 2-mm sieved air-dried samples, after oven-drying at 40°C for 24 to 48 h. The samples were those collected with cylinders in RES and in CHA and MEL, and using a knife in MAR and NAN. Spectra were acquired on two aliquots per samples then averaged. So every SOCv value was associated to one average spectrum acquired in situ and to another acquired on sieved dried soil.

2.5. Spectrum sets and selection of calibration and validation subsets

In situ spectra had been collected in different sampling conditions depending on the study, so could hardly be analyzed together; while laboratory spectra had been collected in similar conditions whatever the study and could be analyzed together. So the three separate sample sets considered in previous studies (RES; CHA and MEL; MAR and NAN) led to seven spectrum sets: in situ and laboratory spectra from RES; in situ and laboratory spectra from CHA and MEL; in situ and laboratory spectra from MAR and NAN; and laboratory spectra from all sites. Each set was divided into a calibration subset, used for building the prediction model, and an external validation subset, for validating it.

2.5.1. Sets of in situ spectra and corresponding laboratory spectra

In RES, in situ spectra were collected on all samples (core segments), but conventional determination of SOCv was only carried out on spectrally representative samples (which were used for predicting SOCv on the other samples, based on spectra; Cardinael et al., 2015). These representative samples were selected using the Kennard-Stone algorithm, based on distances between in situ spectra in the principal component space (Kennard and Stone, 1969), and this was done with the R package prospectr (Stevens and Ramirez-Lopez, 2013). For the present work, 167 samples from this selection (with observed SOCv values) were available. They were divided into a calibration subset, which included 100 spectrally representative samples according to the Kennard-Stone algorithm, and a validation subset, which included the 67 other samples.

The 288 samples from CHA and MEL were studied together. Three samples were removed as spectral outliers, according to principal component analysis (PCA) built on all 288 in situ spectra (these three samples were far from the others and had Mahalanobis distance > 3; Cambou et al., 2016). The 285 remaining samples were divided into a calibration subset, which included 200 spectrally representative samples according to the Kennard-Stone algorithm, and a validation subset with the 85 other samples, as in the initial study (Cambou et al., 2016).

The 137 samples from MAR and NAN were studied together. Based on PCA on all 137 in situ spectra, five samples were removed as spectral outliers (Hotelling's $T^2_{\alpha=0.05}$ distance > 2; Allory et al., 2019). One out of the three or four pits opened in each of the four subsites (MAR1, MAR2, NAN1 and NAN2) was used for validation, based on PCA built with in situ spectra of each subsite separately: the sample spectra of the validation pit had to be in intermediate position, not too scattered or atypical. The four validation pits corresponded to 37 validation samples, and the 95 other samples were used for calibration, as in the initial study (Allory et al., 2019).

For each of the three separate sample sets, same calibration and validation subsets were used for predictions based on laboratory spectra (collected on sieved dried samples), to allow proper comparisons with predictions using in situ spectra.

2.5.2. Total set of laboratory spectra

The laboratory spectra from all five sites (RES, CHA, MEL, MAR and NAN) had been collected in similar conditions, on 2-mm sieved air-dried soil, and so could be studied together. Samples that had been considered outliers in the comparisons between predictions using in situ and laboratory spectra (cf. 2.5.1) were again removed from the analysis. In order to achieve independent validation, four sites were used for calibration and the fifth site for validation. It was considered important that the validation site was well represented spectrally by the calibration sites; otherwise validation, though independent, would have limited interest (cf. McCarty et al., 2002; Brown et al., 2005; Barthès et al., 2020). To evaluate how each site was represented spectrally by the four other sites, PCA was performed on the total spectrum set (584 samples). Spectrum scores on the first and second principal components showed that most samples from MAR, NAN and RES were in peripheral positions; while scores on the third principal component showed that samples from MEL were also peripheral (Fig. 2). So it was decided to use CHA as validation site, and the four other sites as library, for calibration. The validation site included 48 profiles made of 144 samples collected at 0–10, 10–20 and 20–30 cm depth.

2.6. Spectral analysis

Models were built on the seven spectrum sets (cf. 2.5) to predict SOCv, SOCg and Db from VNIR spectra using partial least squares regression (PLSR; Wold et al., 1983). This multivariate regression procedure reduces complex matrices of explanatory variables (VNIR spectra) into a few orthogonal variables, called latent variables (LV), which are linear combinations of explanatory variables built to maximize covariance with the response variable (SOCv, SOCg or Db). Calculations were made using the WinISI4 software (Foss NIRSystems/Tecator Infrasoft International, State College, PA, USA).

2.6.1. Global calibration

Common PLSR procedure, which has often been called global calibration as opposed to local calibration presented below, was performed on the seven spectrum sets separately. In this procedure, all calibration spectra are used for building a unique model that is then applied uniformly on all validation spectra. The optimal number of LV was determined by cross-validation on the calibration subset, which was ranked according to sampling time then divided into six groups cyclically (i.e. the 1st, 7th, etc. samples in the first group, the 2nd, 8th, etc. in the second group, etc., the 6th, 12th, etc. in the sixth group). In six-group

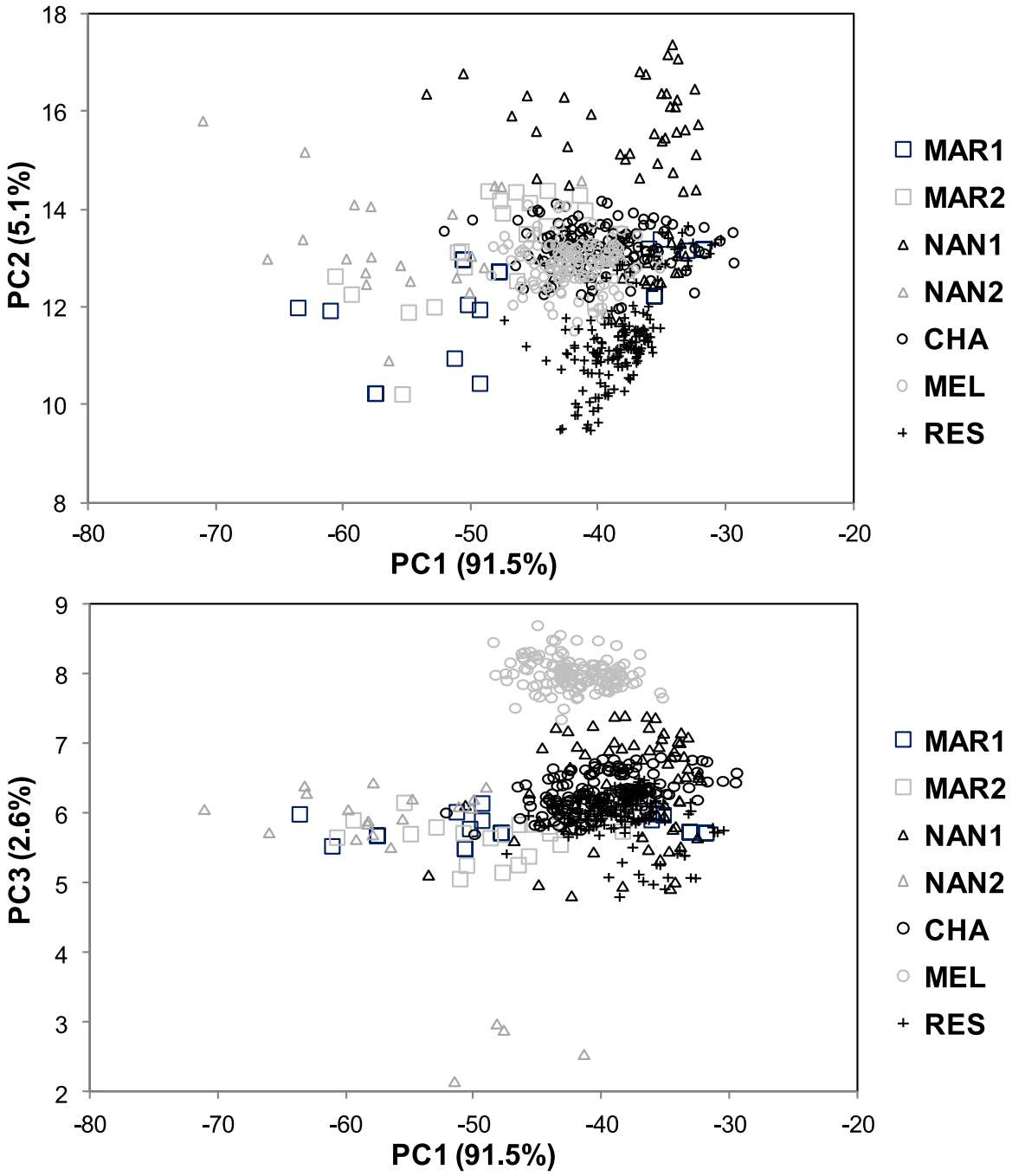

Fig. 2. Projection of the VNIR spectra collected on 2-mm sieved, air-dried samples (laboratory conditions) on the first and second principal components (PC1 and PC2), and on the first and third principal components (PC1 and PC3), respectively.

cross-validation, five groups are used for developing the model and the last one for testing it, and the procedure is repeated six times to use all groups for both model development and prediction; then the predictions are pooled. The optimal number of LV was that after which the standard error of cross-validation (SECV; see Section 2.6.4 for definition) no longer decreased meaningfully (Bjørsvik and Martens, 2001).

2.6.2. *Local calibration*

Local PLSR was also developed on the total set of laboratory spectra from all sites (with independent validation, cf. Section 2.5.2). In that case, prediction on each validation sample is made individually, using calibration samples that are its spectral neighbors (Shenk et al., 1997; Genot et al., 2011; Barthès et al., 2020). These calibration neighbors were selected according to the correlation coefficient R between calibration spectra and each validation spectrum, with a cut-off value $R = 0.95$ below which samples were not considered neighbors; and the minimum number of calibration neighbors was fixed to 4 (otherwise prediction would not be made on the validation sample considered; but this never happened in the present case). Minimum number of calibration neighbors and R cut-off value were fixed according to preliminary tests (Barthès et al., 2020). Here the number of LV was not optimized, because cross-validation would require much computational time. Instead, each prediction was calculated as the weighted mean of predicted values achieved with 3 to 16 LV: each weight was calculated as the inverse of the product of the root mean square (RMS) of spectral residuals (i.e. difference between the actual spectrum and the spectrum

approximated using the considered number of LV) and RMS of regression coefficients using the considered number of LV (Shenk et al., 1997).

2.6.3. Spectrum pretreatment

Models were developed with raw absorbance spectra but also with pretreated spectra, in order to optimize spectral information (Gholizadeh et al., 2013). Common spectrum pretreatments were used, and possibly combined: smoothing, standard normal variate (SNV; i.e. mean-centering and variance-scaling), first-order detrend (D; i.e. linear trend removal), both SNV and D (SNVD), and first-order derivation (Barthès et al., 2020). In total, 16 spectrum types were considered: raw spectra (denoted None001), SNV spectra (SNV001), D spectra (D001), SNVD spectra (SNVD001); the same with 5-point smoothing (None005, SNV005, D005 and SNVD005); the same with first-order derivation over 5-point gap and 5-point smoothing (None155, SNV155, D155 and SNVD155); and with first-order derivation over 15-point gap and 5-point smoothing (None1155, SNV1155, D155 and SNVD1155).

2.6.4. Evaluation of prediction results

The goodness of fit of predictions was evaluated using the following parameters:

- the coefficient of determination (R^2), which was computed as 1−ESS/TSS, where ESS is the error sum of squares and TSS the total sum of squares, and was calculated for both cross-validation (on the calibration subset; R^2cv) and external validation (R^2val);
- the standard error of cross-validation (SECV), defined as the root mean square error (RMSE) of cross-validation, and the standard error of prediction (SEP), as the RMS of external validation;
- the ratio of performance to deviation (RPD) in cross-validation (RPDcv) and in external validation (RPDval), calculated as the ratio of standard deviation (SD) of the calibration subset (SDcal) to SECV, and as the ratio of SD of the validation subset (SDval) to SEP, respectively;
- the ratio of performance to interquartile range (RPIQ) in cross-validation (RPIQcv) and in validation (RPIQval), calculated as the ratio of the interquartile range (IQ) of the calibration subset (IQcal) to SECV, and as the ratio of IQ of the validation subset (IQval) to SEP; IQ is the difference between the third and first quartiles; RPIQ has been recommended for variables with non-normal distributions (Bellon-Maurel et al., 2010).

These parameters are presented in the tables, but the text mainly focuses on RPIQval.

2.7. Comparisons between SOCv, SOCg and Db prediction models

The question arose whether SOCv predictions could be indirect, due to SOCg or Db predictions and correlations between SOCv and SOCg or Db. This was examined by comparing regression coefficients of prediction models of SOCv, SOCg and Db, which allowed identifying spectral regions that contributed most to predictions. This was done for the seven spectrum sets considered (cf. 2.5). For each set, comparison of prediction models was carried out using the spectrum type without derivative that yielded best validation results in average over SOCv and SOCg, according to RPIQval, with similar or almost similar optimal number of LV for both models (first derivative complicates chemical interpretation of regression coefficients, and difference in LV complicates comparison of regression coefficients). For the total set of laboratory spectra, comparison was made for global calibration models, because local calibration did not produce regression at the set scale (as prediction was made on each validation sample individually, cf. 2.6.2).

3. Results

The presentation will focus on SOCv (gC dm^{-3}) predictions, while SOCg (gC kg^{-1}) and Db (kg dm^{-3}) predictions will be considered to a lesser extent, as they have already been addressed in many publications.

3.1. Distributions of observed SOCv, SOCg and Db

Firstly the RES set, CHA and MEL set, and MAR and NAN set were studied separately, with the aim to compare SOCv predictions using VNIR spectra obtained in situ vs. in laboratory conditions. In RES, where samples had been collected regularly from 0 to 10 to 160–180 cm depth, observed SOCv ranged from 7.2 to 39.2 gC dm^{-3}, mean and SD were 14.9 and 6.0 gC dm^{-3}, respectively, coefficient of variation (CV, i.e. SD/mean) was 40% and skewness coefficient was 1.6; CV and skewness were 55% and 1.5 for SOCg, and 19% and −0.5 for Db, respectively. In CHA and MEL, which were studied together, samples had been collected at 0–10, 10–20 and 20–30 cm depth, and observed SOCv ranged from 7.9 to 27.7 gC dm^{-3}, mean and SD were 16.2 and 3.3 gC dm^{-3}, CV was 20% and skewness coefficient was 0.2; CV and skewness were 25% and 0.7 for SOCg, and 11% and −0.1 for Db, respectively. In MAR and NAN, which were studied together, samples had been collected from 0 to 160 cm depth according to horizons, and observed SOCv ranged from 0.3 to 54.6 gC dm^{-3}, mean and SD were 16.3 and 13.7 gC dm^{-3}, CV was 84% and skewness coefficient was 0.6; CV and skewness were 85% and 0.9 for SOCg, and 11% and −0.3 for Db, respectively. For each set, distributions of observed SOCv in calibration and validation subsets are presented in Fig. 3. The distribution of SOCv was wide and multimodal for the MAR and NAN set, which included SOC-poor samples collected under pavement and SOC-rich samples collected in parks or fallows (Allory et al., 2019). Distributions were much less diverse for the two sets of agricultural soil samples studied in RES and in CHA and MEL, especially in the latter, where tree lines were more recent thus less enriched in SOC than in the former. The positive effect of tree lines on SOC has been studied in Cardinael et al. (2017). In addition, CV was comparable for SOCg and SOCv in a given set, and two to eight times higher than for Db; and according to skewness coefficient, Db had more balanced distribution than SOCg and SOCv. Moreover, the correlation coefficient between SOCg and SOCv ranged from 0.90 (CHA and MEL) to 0.95 (MAR and NAN); while correlations with Db were negative and much weaker: from −0.17 to −0.69 between SOCg and Db, and from 0.00 to −0.60 between SOCv and Db, depending on the set.

Then all sets were pooled to study SOCv predictions using VNIR spectra collected on sieved dried samples, with independent validation. For SOCv, mean and SD were 15.6 and 7.6 gC dm^{-3}, respectively, CV was 48% and skewness coefficient was 1.1; CV and skewness were 56% and 1.6 for SOCg, and 18% and 0.2 for Db, respectively. Over this total set, correlation coefficient was 0.93 between SOCv and SOCg, −0.12 between SOCv and Db, and −0.27 between SOCg and Db. Fig. 4 presents SOCv distributions for CHA (validation site) and the other sites (library, i.e. RES, MEL, MAR and NAN): SOCv distribution was wider but less balanced over the library (440 samples) than over the validation site (144 samples); it ranged from 0.3 to 54.6 vs. 9.5 to 27.7 gC dm^{-3} and skewness coefficient was 1.1 vs. 0.5, respectively.

3.2. Comparisons between predictions using in situ vs. laboratory VNIR spectra

For the set of RES samples, predictions of SOCv were more accurate using VNIR spectra collected on sieved dried samples than in situ (Table 1): this was the case either using raw absorbance spectra (RPIQval = 1.9 vs. 1.6), most appropriate spectrum types (RPIQval = 1.9 vs. 1.7), or in average over the 16 spectrum types tested (mean RPIQval = 1.9 vs. 1.6). Few LV were used for these predictions (3 or 4 in average). The results were little affected by spectrum type (SD of RPIQval over 16 spectrum types were < 0.1).

For the set of CHA and MEL, SOCv predictions were even more accurate using laboratory than in situ spectra (Table 1): RPIQval = 2.5 vs. 1.9 with raw spectra, 2.9 vs. 1.9 with most appropriate spectrum types,

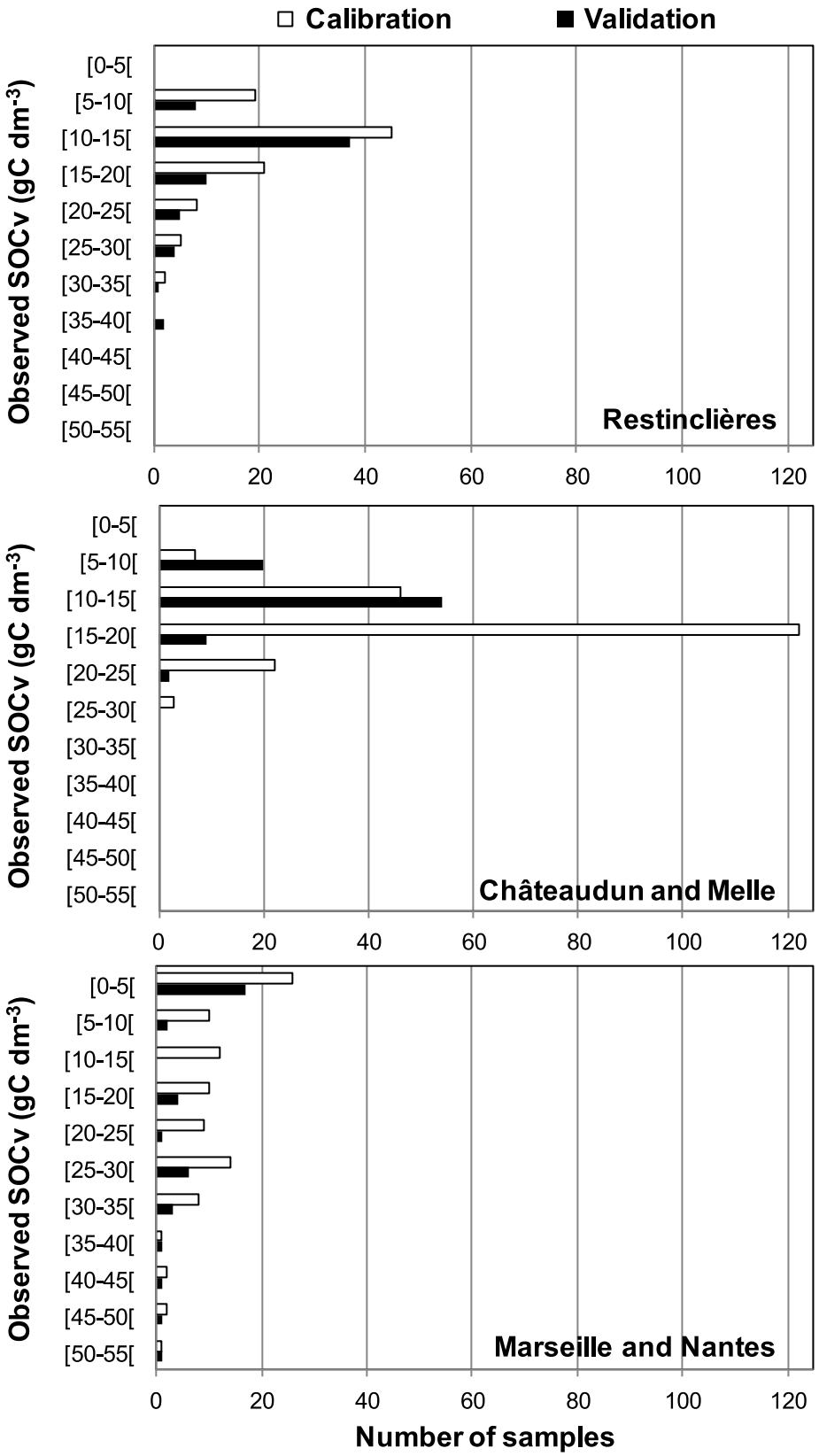

Fig. 3. Distribution of observed volumetric SOC content (SOCv; in gC dm^{-3}) over the calibration and validation subsets of the three separate sets: Restinclières (RES); Châteaudun and Melle (CHA and MEL); and Marseille and Nantes (MAR and NAN).

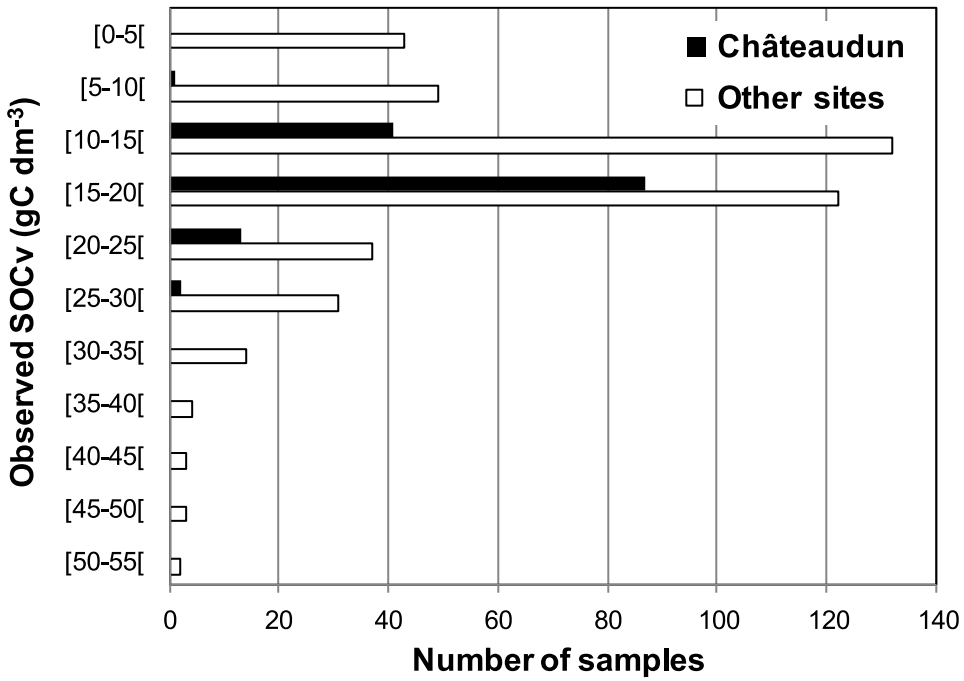

Fig. 4. Distribution of observed volumetric SOC content (SOCv; gC dm^{-3}) over the Châteaudun site (validation site, 144 samples) and over other sites (library, 440 samples).

Table 1

. Predictions of volumetric SOC content (SOCv; gC dm^{-3}) using in situ vs. laboratory VNIR spectra in three separate sample sets (RES; CHA and MEL; MAR and NAN), with global PLSR calibration on representative samples. Results are presented for raw absorbance spectra (None001), for the most appropriate spectrum types (i.e. highest RPIQval; None001 sometimes), then in average (with standard deviation, SD) over 16 spectrum types (including raw spectra and 15 pretreatments, cf. section 2.6.3).

	Calibration					Validation					
Restinclières (RES): Ncal=100, Nval=67; mean, SD and IQ were 14.6, 5.5 and 5.8 gC dm^{-3} in calibration, and 15.7, 6.7 and 7.0 gC dm^{-3} in validation, respectively											
In situ	LV	SECV	R^2cv	RPDcv	RPIQcv	SEP	Bias	Slope	R^2val	RPDval	RPIQval
None001	5	2.7	0.76	2.07	2.18	4.4	-0.5	1.00	0.58	1.54	1.61
D001	4	2.7	0.76	2.03	2.14	4.1	-0.2	0.90	0.62	1.62	1.69
Mean	3.8	2.8	0.73	1.95	2.05	4.3	-0.4	0.95	0.60	1.57	1.64
SD	1.3	0.2	0.04	0.12	0.13	0.1	0.1	0.05	0.02	0.04	0.04
Laboratory	LV	SECV	R^2cv	RPDcv	RPIQcv	SEP	Bias	Slope	R^2val	RPDval	RPIQval
None001	4	2.6	0.77	2.11	2.23	3.8	0.3	1.01	0.68	1.78	1.86
SNVD155	2	2.7	0.76	2.04	2.15	3.7	0.0	0.69	0.69	1.81	1.90
Mean	2.8	2.6	0.77	2.08	2.19	3.8	0.1	1.01	0.68	1.77	1.85
SD	0.9	0.1	0.01	0.04	0.04	0.1	0.1	0.02	0.01	0.03	0.03
Châteaudun (CHA) and Melle (MEL): Ncal=200, Nval=85; mean, SD and IQ were 16.6, 3.3 and 3.6 gC dm^{-3} in calibration, and 15.5, 3.3 and 4.3 gC dm^{-3} in validation, respectively											
In situ	LV	SECV	R^2cv	RPDcv	RPIQcv	SEP	Bias	Slope	R^2val	RPDval	RPIQval
None001	9	2.1	0.58	1.55	1.71	2.3	0.4	0.97	0.52	1.43	1.88
Mean	6.9	2.1	0.59	1.56	1.71	2.5	0.4	1.09	0.45	1.33	1.76
SD	2.3	0.0	0.01	0.02	0.02	0.1	0.1	0.06	0.05	0.06	0.07
Laboratory	LV	SECV	R^2cv	RPDcv	RPIQcv	SEP	Bias	Slope	R^2val	RPDval	RPIQval
None001	12	1.6	0.76	2.04	2.23	1.7	0.3	1.08	0.74	1.93	2.54
SNVD1155	12	1.7	0.74	1.98	2.18	1.5	0.1	1.02	0.80	2.22	2.92
Mean	10.8	1.7	0.74	1.98	2.17	1.7	0.3	1.04	0.75	1.99	2.63
SD	1.6	0.1	0.01	0.06	0.06	0.1	0.1	0.03	0.02	0.09	0.11
Marseille (MAR) and Nantes (NAN): Ncal=95, Nval=37; mean, SD and IQ were 16.6, 13.0 and 20.3 gC dm^{-3} in calibration, and 15.6, 15.5 and 26.6 gC dm^{-3} in validation, respectively											
In situ	LV	SECV	R^2cv	RPDcv	RPIQcv	SEP	Bias	Slope	R^2val	RPDval	RPIQval
None001	14	6.1	0.78	2.13	3.33	5.9	0.6	1.10	0.86	2.62	4.49
SNV001	13	5.8	0.80	2.22	3.48	5.4	0.2	0.98	0.88	2.87	4.92
Mean	9.5	5.9	0.79	2.18	3.41	6.4	0.0	0.97	0.83	2.43	4.16
SD	2.9	0.1	0.01	0.05	0.08	0.5	0.8	0.07	0.03	0.22	0.38
Laboratory	LV	SECV	R^2cv	RPDcv	RPIQcv	SEP	Bias	Slope	R^2val	RPDval	RPIQval
None001	13	5.1	0.85	2.56	4.01	8.0	-2.4	0.96	0.76	1.95	3.34
SNV005	10	4.8	0.86	2.68	4.19	6.1	-1.0	0.80	0.88	2.55	4.37
Mean	10.8	4.8	0.86	2.73	4.27	7.7	-2.5	0.85	0.79	2.04	3.49
SD	1.3	0.3	0.02	0.17	0.27	0.9	0.9	0.07	0.05	0.26	0.44

Ncal and Nval are the size of calibration and validation subsets; IQ is the interquartile range; LV is the number of PLS latent variables; SECV and SEP are standard error of cross-validation and of prediction (external validation), respectively (in gC dm^{-3}); R^2cv, RPDcv and RPIQcv are coefficient of determination, ratio of SDcal to SECV, and ratio of IQcal to SECV, respectively (unitless); R^2val, RPDval and RPIQval are their counterparts for the validation.

and 2.6 vs. 1.8 in average over 16 spectrum types, respectively. More LV were required, in laboratory conditions especially (11 in average, vs. 7 using in situ spectra), and results were more affected by spectrum type, in laboratory conditions especially (SD of RPIQval were ca. 0.1).

In contrast, for the MAR and NAN set, less accurate SOCv predictions were achieved using laboratory than in situ spectra (Table 1): RPIQval = 3.3 vs. 4.5 with raw spectra, 4.4 vs. 4.9 with most appropriate spectrum types, and 3.5 vs. 4.2 in average, respectively. The optimal number of LV averaged 10–11. Moreover, results were noticeably affected by spectrum type (SD of RPIQval were 0.4).

As regarded SOCg, in any case predictions were more accurate with laboratory than in situ spectra, either using most appropriate spectrum types or in average over all spectrum types (RPIQval = 2.2 vs. 1.6 or 2.1 vs. 1.5 for RES, 4.0 vs. 1.8 or 3.4 vs. 1.7 for CHA and MEL, and 5.9 vs. 4.4 or 3.8 vs. 3.6 in MAR and NAN, respectively; data not shown). Moreover, whatever the set, according to RPIQval, predictions tended to be more accurate for SOCv than SOCg with in situ spectra, but more accurate for SOCg than SOCv with laboratory spectra.

As regarded Db, predictions were more accurate with laboratory than in situ spectra in CHA and MEL (RPIQval = 2.8 vs. 2.3 at best and 2.7 vs. 2.3 in average, respectively), but the opposite was observed for MAR and NAN (1.4 vs. 1.8 and 1.3 vs. 1.5, respectively), while predictions with both spectrum acquisition conditions were similar for RES (2.4 vs. 2.3 and 2.2 vs. 2.3, respectively; data not shown). Moreover, predictions were more accurate for Db than SOCg and SOCv with both spectrum acquisition conditions in RES, but the opposite was observed in MAR and NAN; while in CHA and MEL, Db prediction was more accurate than SOCv and SOCg predictions in situ, but intermediate in laboratory conditions.

3.3. Inter-site predictions in laboratory conditions

3.3.1. SOCv, SOCg and Db

All spectra collected on sieved dried samples were considered here. Calibrations were built on the library, which included all samples from RES, MEL, MAR and NAN (440 samples); and CHA was used as independent validation site (144 samples; cf. 2.5.2). Results are presented for SOCv, SOCg and Db in global and local calibrations using raw absorbance spectra (None001), the most appropriate spectrum type (highest RPIQval), then in average (with SD) over 16 spectrum types (including raw spectra and 15 spectrum pretreatments; Table 2).

In global calibration, all validation results for SOCv, SOCg and Db were poor (RPIQval $\leq$ 0.5), due in particular to large biases (bias represented 95% of SEP in average over the three variables and 16 spectrum types). Cross-validation results were much better, with RPIQcv ranging from 2.2 to 2.5 for SOCv, 2.7 to 3.1 for SOCg, and 2.0 to 2.2 for Db. In local calibration, validation results were also poor in general, similarly due to large biases, except for SOCv and SOCg with appropriate pretreatments:

- for SOCv, RPIQval reached 2.0 with D001 and D005 (detrend possibly with smoothing) and 1.6 with SNVD155, but ranged from 0.3 to 1.4 otherwise;
- for SOCg, RPIQval reached 1.7 with SNVD005 and D005, 1.5–1.6 with D001 and SNVD001, but ranged from 0.3 to 1.4 otherwise;
- for Db, RPIQval was always $\leq$ 0.4.

So accurate SOCv predictions and rather accurate SOCg predictions could be achieved in independent validation when using the library of sieved dried samples for local calibration. Comparison on the validation subset between observations and predictions of SOCv using local calibration with pretreatment D005 is presented Fig. 5. Local calibration yielded better SOCv predictions than global calibration, either using raw spectra (RPIQval = 0.6 vs. 0.5, respectively), most appropriate spectrum types (2.0 vs. 0.5) or in average over all spectrum types (0.9 vs. 0.3); and the same observation was made for SOCg (RPIQval = 0.5 vs. 0.4 with raw spectra, 1.7 vs. 0.4 at best, and 0.9 vs. 0.3 in average, respectively). But local calibration had not benefit for Db (RPIQval = 0.3 vs. 0.4, 0.4 vs. 0.5, and 0.3 vs. 0.3, respectively).

3.3.2. SSOC

Best SOCv predictions, at sample level (gC dm^{-3}), were used for SSOC predictions, at profile level (kgC m^{-2} at 0–30 cm depth). So SOCv was predicted using local calibration with pretreatment D005, then summed up at profile level (0–30 cm). This yielded SEP = 0.36 kgC m^{-2} at 0–30 cm (i.e. 7.4% of the mean), bias = 0.03 kgC m^{-2} and RPIQval = 2.1 (Fig. 5). Mean (± SD) observed and predicted SSOC over the 48 validation profiles were 4.93 (± 0.53) and 4.89 (± 0.40) kgC m^{-2}, respectively.

Five random selections of 10 validation profiles yielded the following means (± SD) of observed vs. predicted SSOC at 0–30 cm: 4.87 (± 0.62) vs. 5.08 (± 0.36), 4.76 (± 0.53) vs. 4.85 (± 0.38), 5.02 (± 0.52) vs. 4.93 (± 0.38), 5.05 (± 0.34) vs. 4.88 (± 0.30), and 4.75 (± 0.50) vs. 4.84 (± 0.46) kgC m^{-2}, respectively; so over the five replicates, mean observed profile SSOC per replicate averaged (± SD) 4.89 ± 0.14 kgC m^{-2} and mean predicted SSOC per replicate averaged 4.92 ± 0.10 kgC m^{-2}. When the number of profiles was decreased to five per random selection, mean (± SD) observed vs. predicted SSOC were 4.73 (± 0.61) vs. 4.99 (± 0.44), 4.71 (± 0.60) vs. 4.65 (± 0.32), 4.91 (± 0.53) vs. 4.90 (± 0.36), 5.02 (± 0.37) vs. 4.91 (± 0.15), and 4.64 (± 0.53) vs. 4.86 (± 0.35) kgC m^{-2}, respectively; so over the five replicates, mean observed and predicted SSOC per replicate averaged 4.80 ± 0.16 and 4.86 ± 0.13 kgC m^{-2}, respectively.

3.4. Comparisons between SOCv, SOCg and Db prediction models

There were close correlations between observed SOCv and SOCg (correlation coefficients ranged from 0.90 to 0.95), but weaker and negative correlations between SOCv or SOCg and Db (from 0.00 to −0.69; cf. 3.1). Fig. 6 presents the regression coefficients of SOCv and SOCg prediction models per spectrum set with most appropriate spectrum type (without derivative) in average over SOCv and SOCg, which was: in situ, SNV001 for RES, for CHA and MEL and for MAR and NAN; with laboratory spectra, None001 for RES, D005 for CHA and MEL, None005 for MAR and NAN, and None001 for the total set. For a given spectrum set, the ranges of regression coefficients were the same order of magnitude for SOCv and SOCg, and even similar sometimes (e.g. for RES in situ), but 30 to 100 times smaller for Db (data not shown for Db). Some spectral regions contributed strongly to predictions of both SOCv and SOCg (and Db sometimes):

- for RES in situ, 710–780, 1400, 1720–1740 and 2140–2150 nm (1400 nm also for Db);
- for RES in laboratory conditions, 1420 and 1830 nm (1420 nm also for Db);
- for CHA and MEL in situ, 720–730, 820 and 970 nm;
- for CHA and MEL in laboratory conditions, 720–730, 810–830, 1360–1390, 1880–1890, 2160–2170, 2240–2250, 2310 and 2420–2430 nm (1360 nm also for Db);
- for MAR and NAN in situ, 660–680, 750–770, 1370–1380, 1520–1590, 1720–1740, 2240–2260 and 2340 nm (2340 nm also for Db);
- for MAR and NAN in laboratory conditions, 670–680, 740–750, 1310–1320, 1450–1480, 1740, 1820–1830 and 2000 nm (1830–1840 nm also for Db);
- for the total set of laboratory spectra, 670–680, 1300–1310, 1660–1670, 1730, 2230, 2260 and 2350–2360 nm.

However, other regions contributed more strongly to the prediction of one variable, such as:

- for RES in situ, 620–660 nm for SOCv and 1740–1750 nm for SOCg;

Table 2
. Predictions of SOCv (gC dm^{-3}), SOCg (gC kg^{-1}) and Db (kg dm^{-3}) in global and local PLSR calibration over the total set of VNIR laboratory spectra, with independent validation. Results are presented for raw absorbance spectra (None001), for the most appropriate spectrum types (highest RPIQval; None001 sometimes), then in average (with standard deviation, SD) over 16 spectrum types (cf. section 2.6.3).

Pre-treatment	Calibration									Validation									
	Ncal	Meancal	SDcal	IQcal	LV	SECV	R^2cv	RPDcv	RPIQcv	Nval	Meanval	SDval	IQval	SEP	Bias	Slope	R^2val	RPDval	RPIQval
SOCv, global calibration																			
None001	440	15.7	8.6	8.1	14	3.7	0.82	2.34	2.21	144	16.4	3.0	3.9	8.5	8.0	1.81	0.22	0.35	0.45
Mean	440	15.7	8.6	8.1	13.9	3.5	0.84	2.50	2.36	144	16.4	3.0	3.9	11.5	10.8	3.45	0.13	0.27	0.35
SD	0	0.0	0.0	0.0	0.9	0.2	0.02	0.13	0.13	0	0.0	0.0	0.0	2.1	2.0	1.24	0.04	0.05	0.07
SOCv, local calibration																			
None001	440	15.7	8.6	8.1	nd	nd	nd	nd	nd	144	16.4	3.0	3.9	6.7	6.4	1.28	0.47	0.44	0.57
D005	440	15.7	8.6	8.1	nd	nd	nd	nd	nd	144	16.4	3.0	3.9	1.9	0.1	0.84	0.60	1.56	2.03
Mean	440	15.7	8.6	8.1	nd	nd	nd	nd	nd	144	16.4	3.0	3.9	5.9	5.0	1.40	0.44	0.67	0.88
SD	0	0.0	0.0	0.0	nd	nd	nd	nd	nd	0	0.0	0.0	0.0	2.8	3.4	0.51	0.08	0.43	0.56
SOCg, global calibration																			
None001	440	12.6	8.1	7.0	15	2.5	0.90	3.23	2.77	144	14.3	2.9	3.5	9.6	9.4	1.34	0.51	0.30	0.36
D001	440	12.6	8.1	7.0	13	2.6	0.90	3.15	2.71	144	14.3	2.9	3.5	8.6	8.2	1.65	0.35	0.34	0.40
Mean	440	12.6	8.1	7.0	14.2	2.4	0.91	3.36	2.88	144	14.3	2.9	3.5	10.9	10.5	2.07	0.31	0.27	0.32
SD	0	0.0	0.0	0.0	1.1	0.1	0.01	0.17	0.14	0	0.0	0.0	0.0	1.4	1.3	0.55	0.10	0.04	0.04
SOCg, local calibration																			
None001	440	12.6	8.1	7.0	nd	nd	nd	nd	nd	144	14.3	2.9	3.5	6.4	6.1	1.43	0.57	0.45	0.54
SNVD005	440	12.6	8.1	7.0	nd	nd	nd	nd	nd	144	14.3	2.9	3.5	2.0	0.6	0.81	0.58	1.43	1.71
Mean	440	12.6	8.1	7.0	nd	nd	nd	nd	nd	144	14.3	2.9	3.5	5.5	4.3	1.39	0.53	0.71	0.85
SD	0	0.0	0.0	0.0	nd	nd	nd	nd	nd	0	0.0	0.0	0.0	2.6	3.6	0.47	0.09	0.43	0.52
Db, global calibration																			
None001	440	1.46	0.25	0.35	12	0.16	0.58	1.54	2.16	144	1.16	0.09	0.12	0.35	0.32	8.70	0.03	0.26	0.36
SNVD1155	440	1.46	0.25	0.35	11	0.16	0.59	1.55	2.18	144	1.16	0.09	0.12	0.28	0.24	9.52	0.02	0.33	0.45
Mean	440	1.46	0.25	0.35	7.9	0.16	0.55	1.50	2.11	144	1.16	0.09	0.12	0.36	0.30	4.11	0.12	0.25	0.35
SD	0	0.00	0.00	0.00	3.0	0.01	0.03	0.05	0.07	0	0.00	0.00	0.00	0.04	0.04	2.35	0.07	0.03	0.04
Db, local calibration																			
None001	440	1.46	0.25	0.35	nd	nd	nd	nd	nd	144	1.16	0.09	0.12	0.39	0.38	1.36	0.30	0.23	0.32
SNVD001	440	1.46	0.25	0.35	nd	nd	nd	nd	nd	144	1.16	0.09	0.12	0.35	0.34	1.30	0.29	0.26	0.35
Mean	440	1.46	0.25	0.35	nd	nd	nd	nd	nd	144	1.16	0.09	0.12	0.38	0.34	1.19	0.30	0.24	0.33
SD	0	0.00	0.00	0.00	nd	nd	nd	nd	nd	0	0.00	0.00	0.00	0.02	0.02	0.30	0.01	0.01	0.01

Ncal and Nval are the size of calibration and validation subsets, respectively. Meancal, Meanval, SDcal, SDval, IQcal and IQval are the mean, standard deviation and interquartile range for calibration and validation subsets, respectively (in gC kg^{-1} for SOCg and in gC dm^{-3} for SOCv). LV is the number of PLS latent variables. SECV and SEP are standard error of cross-validation and of prediction (external validation), respectively (in gC kg^{-1} for SOCg and in gC dm^{-3} for SOCv). R^2cv, RPDcv and RPIQcv are coefficient of determination, ratio of SDcal to SECV, and ratio of IQcal to SECV, respectively (unitless); R^2val, RPDval and RPIQval are their counterparts for the validation.

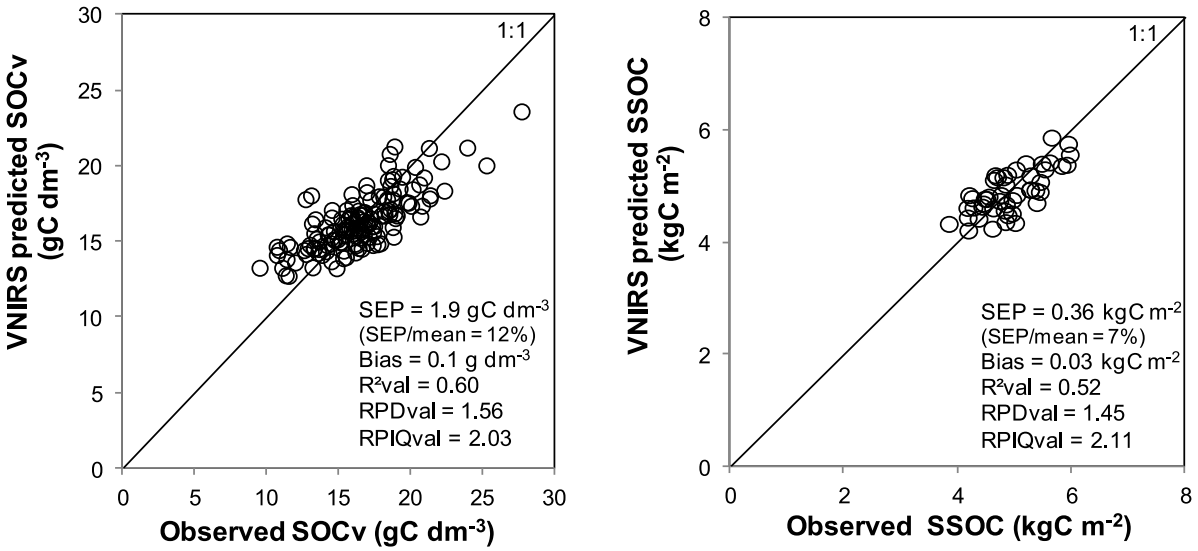

Fig. 5. Comparison between observed and VNIRS-predicted volumetric SOC content (SOCv; gC dm^{-3}) and SOC stock (SSOC; kgC m^{-2} at 0–30 depth). The validation subset included 144 samples from Châteaudun (48 profiles) and the calibration subset the samples from all other sites (440 samples). Regression procedure was local PLSR, with pretreatment D005 (smoothing then detrend).

- for RES in laboratory conditions, 630–800 nm for SOCv and 1300–1320 nm for SOCg;
- for CHA and MEL in situ, 2130–2150 nm for SOCv and 1000–1010 nm for SOCg;
- for CHA and MEL in laboratory conditions, 1730 nm for SOCv and 2350 nm for SOCg;
- for MAR and NAN in situ, 830–840 nm for SOCv and 1430–1440 nm for SOCg;
- for MAR and NAN in laboratory conditions, 1920–1930 nm for SOCv;
- for the total set of laboratory spectra, 1360 nm for SOCv and 1140 nm for SOCg.

Moreover, some regions contributed strongly to SOCv and/or SOCg predictions using both in situ and laboratory spectra:

- for RES, 630–780 nm and around 1400–1420 nm for SOCv, and around 2130–2140 nm for SOCg;
- for CHA and MEL, 720–730 and 820 nm for both SOCv and SOCg;
- for MAR and NAN, 670–680, 750 and 1740 nm for both SOCv and SOCg.

However, no spectral region contributed strongly to SOCv or SOCg prediction over all spectrum sets, or even all in situ spectrum sets or all laboratory spectrum sets. Similarities between spectral regions that contributed strongly to SOCv and/or SOCg prediction models were mainly observed within a given sample set (i.e. RES; CHA and MEL; MAR and NAN; and all samples). For instance, considering spectral regions that contributed strongly to predictions, there were more similarities between SOCv prediction using in situ spectra and SOCg prediction using laboratory spectra both from CHA and MEL (720–730, 820–830, and 960–970 nm to a lesser extent; same sample set but different variables and spectrum acquisitions) than between SOCv predictions using in situ spectra from RES and from CHA and MEL (2130–2140 nm; same variable and spectrum acquisition but different sample sets).

4. Discussion

4.1. Predictions with in situ vs. laboratory VNIR spectra

Better VNIRS predictions of SOCv using spectra collected on sieved dried samples than in situ, as observed for RES and for CHA and MEL, is consistent with what has generally been reported in the literature for other soil properties, SOCg especially (cf. review by Barthès and Chotte, 2020). And indeed, better SOCg prediction in laboratory than in situ conditions was also observed for the three separate sample sets studied here. Such difference has firstly been attributed to the variability of soil moisture content during field campaigns, which represents noise in the spectra collected in situ thus complicates calibrations; while laboratory spectra are collected in homogeneous moisture conditions (Morgan et al., 2009; Hutengs et al., 2019). Sample structure and roughness, contamination by dust and temperature variations have also been cited to explain poorer in situ VNIRS prediction of soil content in various compounds (Stevens et al., 2008; Stenberg et al., 2010). As regarded SOCv specifically, it might be expected that spectra acquired on cohesive material (e.g. cores) would be suitable; but the present result indicated that air-dried, gently crushed then 2-mm-sieved soil material still included useful spectral information on SOCv.

For MAR and NAN, better SOCv prediction was however achieved using VNIR spectra collected in situ than on sieved dried samples (accurate predictions were nevertheless achieved using laboratory spectra). Though rare, better VNIRS prediction of soil properties in situ than in laboratory conditions has been reported sometimes (Gras et al., 2014; Barthès and Chotte, 2020): this has been attributed, in particular, to better light transmission due to high cohesion of samples scanned in situ, or possibly to rather high and homogeneous moisture content. Soil material scanned in situ in MAR and NAN had variable moisture content, but high cohesion (pit walls; Allory et al., 2019). Moreover, coarse particles (> 2 mm) content was rather high and variable (mean ca. 25±15 g 100 g^{-1}), hence variable discrepancies between bulk soil used for determining SOCv conventionally and 2-mm sieved samples scanned in the laboratory; and this probably complicated SOCv prediction from sieved samples. In contrast, proportions of coarse particles were much lower in RES and in CHA and MEL samples (< 1 and 8 g 100 g^{-1} in average, Cardinael et al., 2015, and Cambou et al., 2016, respectively). On a small range of tropical soils, Allo et al. (2020) achieved more accurate SOCv predictions with VNIR spectra collected on fresh intact cores than on 2-mm sieved air-dried samples; and better prediction in situ could also be attributed to coarse-particle content, which was not specified but could be estimated at 24 g 100 g^{-1} in average (according to mean SOCg, SOCv and Db).

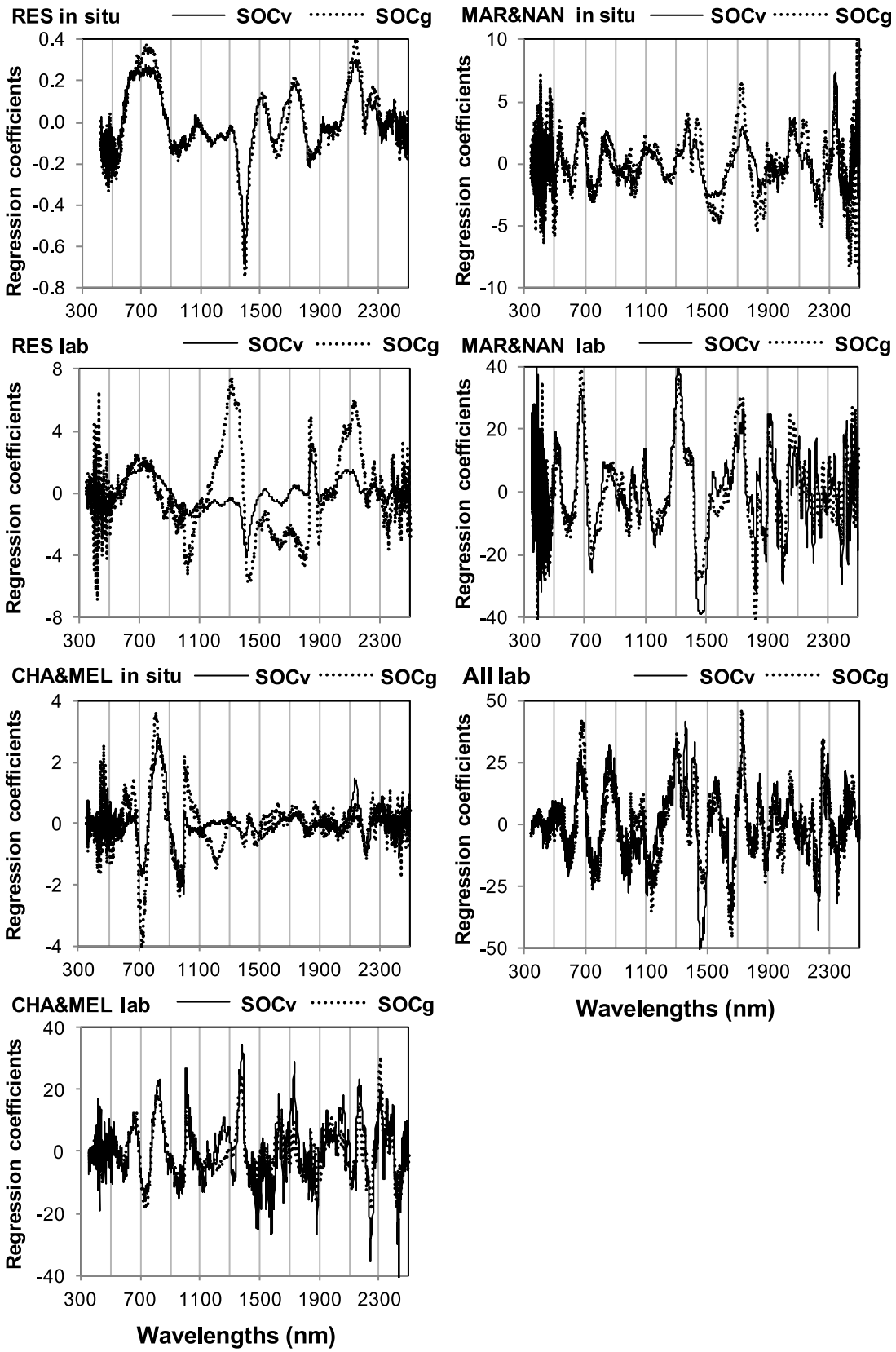

Fig. 6. Regression coefficients of SOCv (gC dm^{-3}) and SOCg (gC kg^{-1}) prediction models for the seven spectrum sets considered: RES, CHA and MEL, and MAR and NAN in situ (all with pretreatment SNV001) and in laboratory conditions (with None001, D005 and None005, respectively), and all laboratory spectra (with None001).

4.2. Inter-site prediction of SOCv and SSOC in laboratory conditions

When the total set of laboratory spectra was studied (five sites with different mineralogies, textures and land uses), accurate SOCv and SSOC predictions could be achieved in independent validation (leave-one-site-out) using spectra collected on sieved dried samples (RPIQval = 2.0–2.1; Fig. 5). Allory et al. (2019) and Allo et al. (2020) also used VNIR spectra of sieved dried samples for SOCv predictions, but set size and diversity were smaller, and validations were not independent. Roudier et al. (2015) and Cambou et al. (2016) used in situ VNIR spectra for SOCv prediction in one or two sites, with no independent validation either. So the present work represents the first attempt to use a library of VNIR spectra collected on sieved dried soils (i.e. archive soil samples) for quantifying SSOC. And such time- and cost-effective quantification of SSOC offers important perspectives, considering the need for wider quantification of SSOC (Bellon-Maurel and McBratney, 2011; Paustian et al., 2019; Smith et al., 2019); and also considering that other approaches are either tedious, when they involve intact core sampling, or complex, when based on radiation transmitting or scattering for Db prediction in addition to VNIRS prediction of SOCg (Lobsey and Viscarra Rossel, 2016; Priori et al., 2016; Viscarra Rossel et al., 2016, 2017).

The result was achieved using local PLSR, based on spectral neighbors, which seems much more appropriate than global PLSR for (V)NIRS or MIRS predictions of soil properties when using large spectral libraries for independent validation (Gogé et al., 2014; Gupta et al., 2018; Dangal et al., 2019; Barthès et al., 2020); but in such conditions, comparable performances would probably be achieved with machine learning approaches such as classification and regression trees (e.g. Dangal et al., 2019).

Predictions at profile level, calculated as the sum of predictions made at 0–10, 10–20 and 20–30 cm, were slightly more accurate than predictions at sample level (RPIQval = 2.1 vs. 2.0, respectively; Fig. 5). This confirmed that under-predictions at some depth levels and over-predictions at other depth levels could offset each other at profile level (Allory et al., 2019). Moreover bias was small (0.03 kgC m^{-2} at 0–30 cm), which indicated that under-predictions for some profiles were offset by over-predictions for other profiles, so that average SSOC prediction at the site scale was accurate. For this CHA site, used as independent validation subset, mean (± SD) observed SSOC over 48 profiles was 4.93 (± 0.53) kgC m^{-2} at 0–30 cm. This could be considered a valuable SSOC estimation at the scale of the site, which covered 5 ha, with limited effect of land use and management on SSOC (tree plantation was recent; data not shown). Mean (± SD) predicted SSOC over the 48 profiles allowed accurate approximation of this estimation: 4.89 (± 0.40) kgC m^{-2}. But accurate approximations could also be achieved based on predicted SSOC on much fewer validation profiles: over five replicates of random profile selections, mean (± SD) predicted SSOC per selection ranged from 4.84 (± 0.46) to 5.08 (± 0.36) kgC m^{-2} depending on the replicate when selections included 10 profiles, and from 4.65 (± 0.32) to 4.99 (± 0.44) kgC m^{-2} when selections included five profiles. Approximations were more accurate with 10- than five-profile selections, which was expected, nevertheless the benefit of using twice more profiles was rather limited. But the important point is that SOCv predictions using the library and spectra of sieved dried samples from a few validation profiles led to accurate approximations of average SSOC at the scale of the independent validation field. This offers important perspectives for SSOC accounting, for instance for SSOC inventories, though the level of prediction error hardly allowed SSOC monitoring (i.e. detecting SSOC differences between dates).

4.3. Prediction models of SOCv vs. SOCg and Db

There were close correlations between SOCv (gC dm^{-3}) and SOCg (gC kg^{-1}), and their variabilities were similar, according to coefficients of variation (SD/mean; cf. 3.1). According to RPIQval, SOCv tended to be more accurately predicted than SOCg when using in situ spectra in the three separate sample sets (cf. 3.2), and using laboratory spectra in the total set (cf. 3.3); but the opposite was seen when using laboratory spectra in the three separate sample sets (cf. 3.2). Discrepancies were due to respective variations of IQval and SEP for SOCg vs. SOCv depending on the spectrum set considered. The interesting point is that VNIRS prediction could not be considered as systematically less accurate for SOCv than SOCg. Moreover, according to regression coefficients, all spectral regions did not contribute similarly to SOCv and SOCg prediction models within a given spectrum set, though some contributed strongly to both predictions (cf. 3.4 and Fig. 6). So VNIRS predictions of SOCv could hardly be considered as indirect due to VNIRS predictions of SOCg and correlations between SOCv and SOCg, as might have been hypothesized. Indeed, besides close correlation between SOCv and SOCg, this would have required better SOCg than SOCv prediction and similar contributions of spectral regions to both prediction models. Furthermore, correlations between SOCv and Db were negative and not close in general (cf. 3.1), and most spectral regions contributed differently to SOCv and Db prediction models (cf. 3.4); so indirect SOCv predictions due to Db predictions and correlations between SOCv and Db did not seem possible. There was thus no evidence of indirect VNIRS predictions of SOCv as a result from VNIRS predictions of SOCg or Db.

4.4. Chemical compounds involved in prediction models

When considering the seven spectrum sets, some spectral regions often contributed strongly to SOCv or SOCg prediction models, using in situ or laboratory spectra (cf. 3.4 and Fig. 6). According to (Workman and Weyer, 2008), except when otherwise specified, these regions have been assigned to aliphatic organic compounds (740–760, 820–830, 1360–1400, 1440, 1730, 2310 and 2330–2360 nm), aromatic organic compounds (670–680 nm, Manoj and Kunjomana, 2011; 1140, 1420–1430, 1670, 1740, 2150 and 2420–2440 nm), N organic compounds such as amines and amides (1000–1010, 1450–1480, 2010, 2160–2180 nm), lipids (1820–1840 nm, Williams and Norris, 2001; 2140 nm), carboxylic acids (1870 nm, Dyer, 1965; 1890 and 2130 nm), silicates (1300–1320 nm, McDowell et al., 2012), iron oxides (720–730 nm, Sherman et al., 1982; 2240–2260 nm, Bishop et al., 2008) and water (970 and 1920–1940 nm).

However, it was difficult to identify chemical compounds that would contribute more clearly to predictions of SOCv than SOCg, or using in situ than laboratory spectra, and vice-versa. For instance, in CHA and MEL, N organic compounds seemed more involved in SOCv than SOCg prediction with laboratory spectra (1000–1010, 1450–1480 and 2160–2170 nm were strongly involved in SOCv prediction vs. 2160–2170 nm only for SOCg); but the opposite was observed with in situ spectra (no corresponding region strongly involved for SOCv vs. 1000–1010 nm for SOCg; Fig. 6). Similarly, aliphatic compounds seemed more involved in SOCv prediction using laboratory than in situ spectra in CHA and MEL (810–830, 1360–1390, 1730 and 2310 nm with laboratory spectra vs. 820 nm only in situ), but the opposite was seen in MAR and NAN (740–750 nm vs. 740–760, 830–840, 1370–1380 and 2240 nm, respectively). As already mentioned (cf. 3.4), specific contributions of spectral regions to prediction models seemed to depend firstly on the sample set; so specific contributions of chemical compounds (that these regions have been assigned to) depended probably on the sample set too.

5. Conclusion

Few studies have used VNIRS for direct SOC stock prediction. The present work showed that accurate predictions could be achieved in independent validation using VNIR spectra of sieved dried samples, at sample level (SOCv, in gC dm^{-3}) and at profile level (SSOC, in kgC m^{-2}; RPIQval ≥ 2.0). The level of prediction error (SEP = 0.36 kgC m^{-2} at 0–30 cm) hardly allowed SSOC monitoring (i.e. detecting SSOC differences between dates). Nevertheless, average SSOC at the level of the

independent validation site (5 ha), as estimated from 48 SSOC observations, could be approximated accurately with predictions using the library for calibration and spectra from only five to 10 randomly selected profiles from this site. This offers important perspectives for SOC accounting, for instance for SOC inventories, considering the time- and cost-effectiveness of VNIRS, when compared with other SOC stock determination procedures. However, for building spectral libraries, there is a need for collections of sieved dried samples analyzed for SOC stock, which are not necessarily available for wide areas yet. Moreover, local calibration (based on spectral neighbors) yielded more accurate predictions than global calibration, which confirmed studies that used large soil spectral libraries for prediction on independent samples.

For different methods of in situ spectrum acquisition, more accurate SOCv predictions were achieved with spectra acquired on sieved dried samples than in situ. This was consistent with most publications that compared in situ vs. laboratory VNIRS prediction of soil properties, due in particular to varying soil moisture conditions. However, for one sample set studied in the present work (urban soils), more accurate SOCv prediction was achieved with in situ spectra (accurate prediction was nevertheless achieved with laboratory spectra). This was attributed to rather high and variable soil content in coarse particles, which probably complicated SOCv prediction using spectra of sieved samples. With the possible exception of soils rich in coarse particles, sieved dried samples could thus be recommended for VNIRS prediction of SOCv.

There were close correlations between observed SOCv and SOCg (gC kg^{-1}), but in several instances VNIRS predictions were more accurate for SOCv than SOCg (in situ and for the total set of laboratory spectra). Moreover, according to regression coefficients of prediction models, all spectral regions did not contribute similarly to SOCv and SOCg predictions. Correlations between SOCv and Db were weaker, and contributions of spectral regions to SOCv and Db prediction models were not similar either. So there was no evidence of indirect VNIRS prediction of SOCv as a result of VNIRS prediction of SOCg or Db. Moreover, when examining regression coefficients, there were no clear trends about spectral regions (and chemical compounds they have been assigned to) which would be more involved in predictions of SOCv than SOCg, or using in situ than laboratory spectra, and vice versa. Specific contributions of spectral regions to prediction models seemed to depend firstly on the sample set considered.

Declaration of Competing Interests

The authors have no competing interest to declare.

Acknowledgements

Ernest Kouakoua, Manon Villeneuve, Patricia Moulin, Céline Durand and Dominique Renard are thanked for their help in field work, and Tiphaine Chevallier, Claire Chenu, Laure Vidal-Beaudet, Patrice Cannavo and Christophe Schwartz, for their contribution to research project management. The study in Restinclières (RES) was part of the AGRIPSOL project (Agroforestry for soil protection; contract 1260C0042) financed by ADEME (French agency for ecological transition) through the call for proposals REACCTIF (Research on climate change mitigation in agriculture and forestry). The study in Châteaudun and Melle (CHA and MEL) was part of the INCA project (In-field spectroscopy for carbon accounting; contract 1060C0093) supported by the GESSOL program of the French Ministry of ecology, sustainable development and energy, and was funded by ADEME. The study in Marseille and Nantes (MAR and NAN) was part of the SUPRA project (Sols urbains et projets d'aménagement; contract 1772C0021) financed by ADEME, which also funded the Ph.D. grant of Aurélie Cambou, along with the Région Pays de la Loire, France.

References

Workman Jr., J., Weyer, L., 2008. Pratical Guide to Interpretative Near-Infrared Spectroscopy. CRC Press, Boca Raton, FL, USA.

"4 per 1000" Initiative, 2018. http://4p1000.org (accessed 25 Feb. 2021).

Allo, M., Todoroff, P., Jameux, M., Stern, M., Paulin, L., Albrecht, A., 2020. Prediction of tropical volcanic soil organic carbon stocks by visible-near- and mid-infrared spectroscopy. Catena 189, 104452 https://doi.org/10.1016/j.catena.2020.104452.

Allory, V., Cambou, A., Moulin, P., Schwartz, C., Cannavo, P., Vidal-Beaudet, L., Barthès, B.G., 2019. Quantification of soil organic carbon stock in urban soils using visible and near infrared reflectance spectroscopy (VNIRS) in situ or in laboratory conditions. Sci. Total Environ. 686, 764–773 https://doi.org/10.1016/j.scitotenv.2019.05.192.

Al-Shammary, A.A.G., Kouzani, A.Z., Kaynak, A., Khoo, S.Y., Norton, M., Gates, W., 2018. Soil bulk density estimation methods: a review. Pedosphere 28, 581–596 https://doi.org/10.1016/S1002-0160(18)60034-7.

Angelopoulo, T., Balafoutis, A., Zalidis, G., Bochtis, D., 2020. From laboratory to proximal sensing spectroscopy for soil organic carbon estimation—A review. Sustainability 12, 443 https://doi.org/10.3390/su12020443.

Baldock, J.A., Wheeler, I., McKenzie, N., McBratney, A., 2012. Soils and climate change: potential impacts on carbon stocks and greenhouse gas emissions, and future research for Australian agriculture. Crop Pasture Sci 63, 269–283 https://doi.org/10.1071/CP11170.

Barra, I., Haefele, S.M., Sakrabani, R., Kebede, F., 2021. Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: recent advances–A review. Trends Analyt. Chem. 135, 116166 https://doi.org/10.1016/j.trac.2020.116166.

Barthès, B.G., Chotte, J.L., 2020. Infrared spectroscopy approaches support soil organic carbon estimations to evaluate land degradation. Land Degrad. Dev. 32, 310–322 https://doi.org/10.1002/ldr.3718.

Barthès, B.G., Kouakoua, E., Coll, P., Clairotte, M., Moulin, P., Saby, N.P.A., Le Cadre, E., Etayo, A., Chevallier, T., 2020. Improvement in spectral library-based quantification of soil properties using representative spiking and local calibration – The case of soil inorganic carbon prediction by mid-infrared spectroscopy. Geoderma 369, 114272 https://doi.org/10.1016/j.geoderma.2020.114272.

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. Trends Analyt. Chem. 29, 1073–1081 https://doi.org/10.1016/j.trac.2010.05.006.

Bellon-Maurel, V., McBratney, A., 2011. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. Soil Biol. Biochem. 43, 1398–1410 https://doi.org/10.1016/j.soilbio.2011.02.019.

Bishop, J.L., Lane, M.D., Dyar, M.D., Brown, A.J., 2008. Reflectance and emission spectroscopy study of four groups of phyllosilicates: smectites, kaolinite-serpentines, chlorites and micas. Clay Miner 43, 35–54 https://doi.org/10.1180/claymin.2008.043.1.03.

Bjørsvik, H.R., Martens, H., 2001. Data analysis: calibration of NIR instruments by PLS regression. In: Burns, D.A., Ciurczak, E.W. (Eds.), Handbook of Near-Infrared Analysis, 2nd ed. Taylor and Francis, Boca Raton, FL, pp. 185–207.

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 https://doi.org/10.1016/j.geoderma.2005.01.001.

Cambou, A., Cardinael, R., Kouakoua, E., Villeneuve, M., Durand, C., Barthès, B.G., 2016. Prediction of soil organic carbon stock using visible and near infrared reflectance spectroscopy (VNIRS) in the field. Geoderma 261, 151–159 https://doi.org/10.1016/j.geoderma.2015.07.007.

Cardinael, R., Chevallier, T., Barthès, B.G., Saby, N.P., Parent, T., Dupraz, C., Bernoux, M., Chenu, C., 2015. Impact of alley cropping agroforestry on stocks, forms and spatial distribution of soil organic carbon – A case study in a Mediterranean context. Geoderma. 259–260, 288–299. https://doi.org/10.1016/j.geoderma.2015.06.015.

Cardinael, R., Chevallier, T., Cambou, A., Béral, C., Barthès, B.G., Dupraz, C., Durand, C., Kouakoua, E., Chenu, C., 2017. Increased soil organic carbon stocks under agroforestry: a survey of six different sites in France. Agric. Ecosyst. Environ. 236, 243–255 https://doi.org/10.1016/j.agee.2016.12.011.

Casanova, M., Tapia, E., Seguel, E., Salazar, O., 2015. Direct measurement and prediction of bulk density on alluvial soils of central Chile. Chil. J. Agric. Res. 76, 105–113 https://doi.org/10.4067/S0718-58392016000100015.

Dangal, S.R.S., Sanderman, J., Wills, S., Ramirez-Lopez, L., 2019. Accurate and precise prediction of soil properties from a large mid-infrared spectral library. Soil Syst 3, 11 https://doi.org/10.3390/soilsystems3010011.

Dignac, M.-F., Derrien, D., Barré, P., Barot, S., Cécillon, L., Chenu, C., Chevallier, T., Freschet, G.T., Garnier, P., Guenet, B., Hedde, M., Klumpp, K., Lashermes, G., Maron, P.-A., Nunan, N., Roumet, C., Basile-Doelsch, I., 2017. Increasing soil carbon storage: mechanisms, effects of agricultural practices and proxies. A review. Agron. Sustain. Dev. 37, 14 https://doi.org/10.1007/s13593-017-0421-2.

Dyer, J.R., 1965. Applications of Absorption Spectroscopy of Organic Compounds. Prentice-Hall, Englewood Cliffs, NJ, USA.

Eglin, T., Ciais, P., Piao, S.L., Barre, P., Bellassen, V., Cadule, P., Chenu, C., Gasser, T., Koven, C., Reichstein, M., Smith, P., 2010. Historical and future perspectives of global soil carbon response to climate and land-use changes. Tellus B: Chem. Phys. Meteorol. 62, 700–718 https://doi.org/10.1111/j.1600-0889.2010.00499.x.

Genot, V., Colinet, G., Bock, L., Vanvyve, D., Reusen, Y., Dardenne, P., 2011. Near infrared reflectance spectroscopy for estimating soil characteristics valuable in the

diagnosis of soil fertility. J. Near Infrared Spectrosc. 19, 117–138 https://doi.org/10.1255/jnirs.923.
Gholizadeh, A., Boruvka, L., Saberioon, M., Vasat, R., 2013. Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: state-of-the-art and key issues. Appl. Spectrosc. 67, 1349–1362 https://doi.org/10.1366/13-07288.
Gogé, F., Gomez, C., Jolivet, C., Joffre, R., 2014. Which strategy is best to predict soil properties of a local site from a national Vis-NIR database? Geoderma 213, 1–9 https://doi.org/10.1016/j.geoderma.2013.07.016.
Gras, J.P., Barthès, B.G., Mahaut, B., Trupin, S., 2014. Best practices for obtaining and processing field visible and near infrared (VNIR) spectra of topsoils. Geoderma 214–215, 126–134 https://doi.org/10.1016/j.geoderma.2013.09.021.
Gupta, A., Vasava, H.B., Das, B.S., Choubey, A.K., 2018. Local modeling approaches for estimating soil properties in selected Indian soils using diffuse reflectance data over visible to near-infrared region. Geoderma 325, 59–71 https://doi.org/10.1016/j.geoderma.2018.03.025.
Helliwell, J.R., Sturrock, C.J., Grayling, K.M., Tracy, S.R., Flavel, R.J., Young, I.M., Whalley, W.R., Mooney, S.J., 2013. Applications of X-ray computed tomography for examining biophysical interactions and structural development in soil systems: a review. Eur. J. Soil Sci. 64, 279–297 https://doi.org/10.1111/ejss.12028.
Hobley, E.U., Murphy, B., Simmons, A., 2018. Comment on "Soil organic stocks are systematically overestimated by misuse of the parameters bulk density and rock fragment content" by Poeplau et al. (2017). Soil. 4, 169–171 https://doi.org/10.5194/soil-4-169-2018.
Hutengs, C., Seidel, M., Oertel, F., Ludwig, B., Vohland, M., 2019. In situ and laboratory soil spectroscopy with portable visible-to-near infrared and mid-infrared instruments for the assessment of organic carbon in soils. Geoderma 355, 113900 https://doi.org/10.1016/j.geoderma.2019.113900.
IUSS Working Group WRB, 2014. World Reference Base For Soil Resources 2014. International Soil Classification System For Naming Soils and Creating Legends For Soil Maps. World Soil Resources Reports No. 106. FAO, Rome.
Kennard, R.W., Stone, L.A., 1969. Computer aided design of experiments. Technometrics 11, 137–148 https://doi.org/10.2307/1266770.
Lal, R., 2014. Soil conservation and ecosystem services. Int. Soil Water Conserv. Res. 2, 36–47 https://doi.org/10.1016/S2095-6339(15)30021-6.
Lobsey, C.R., Viscarra Rossel, R.A., 2016. Sensing of soil bulk density for more accurate carbon accounting. Eur. J. Soil Sci. 67, 504–513 https://doi.org/10.1111/ejss.12355.
Manoj, B., Kunjomana, A.G., 2011. Analytical study of two differently ranked coals using UV-VIS-NIR spectroscopy. J. Minerals Materials Characterization Eng. 10, 905–911 https://doi.org/10.4236/jmmce.2011.1010070.
McCarty, G.W., Reeves III, J.B., Reeves, V.B., Follett, R.F., Kimble, J., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Sci. Soc. Am. J. 66, 640–646 https://doi.org/10.2136/sssaj2002.0640.
McDowell, M.L., Bruland, G.L., Deenik, J.L., Grunwald, S., Knox, N.M., 2012. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. Geoderma 189-190, 312–320 https://doi.org/10.1016/j.geoderma.2012.06.009.
Metz, B., Davidson, O.R., Bosch, P.R., Dave, R., Meyer, L.A. (Eds.), 2007. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel On Climate Change. Cambridge University Press, Cambridge, UK, and New York.
Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.-S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.-C., Vågen, T.-G., van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. Geoderma 292, 59–86 https://doi.org/10.1016/j.geoderma.2017.01.002.
Minasny, B., McBratney, A.B., Tranter, G., Murphy, B.W., 2008. Using soil knowledge for the evaluation of mid-infrared diffuse reflectance spectroscopy for predicting soil physical and mechanical properties. Eur. J. Soil Sci. 59, 960–971 https://doi.org/10.1111/j.1365-2389.2008.01058.x.
Moreira, C.S., Brunet, D., Verneyre, L., Sá, S.M.O., Galdos, M.V., Cerri, C.C., Bernoux, M., 2009. Near infrared spectroscopy for soil bulk density assessment. Eur. J. Soil Sci. 60, 785–791 https://doi.org/10.1111/j.1365-2389.2009.01170.x.
Morgan, C.L.S., Waiser, T.H., Brown, D.J., Hallmark, C.T., 2009. Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy. Geoderma 151, 249–256 https://doi.org/10.1016/j.geoderma.2009.04.010.
Nocita, M., Stevens, A., Toth, G., Panagos, P., van Wesemael, B., Montanarella, L., 2014. Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach. Soil Biol. Biochem. 68, 337–347 https://doi.org/10.1016/j.soilbio.2013.10.022.
Pansu, M., Gautheyrou, J., 2006. Handbook of Soil Analysis: Mineralogical, Organic and Inorganic Methods. Springer, Heidelberg, Germany.
Paustian, K., Collier, S., Baldock, J., Burgess, R., Creque, J., DeLonge, M., Dungait, J., Ellert, B., Frank, S., Goddard, T., Govaerts, B., Grundy, M., Henning, M., Izaurralde, R.C., Madaras, M., McConkey, B., Porzig, E., Rice, C., Searle, R., Seavy, N., Skalsky, R., Mulhern, W., Jahn, M., 2019. Quantifying carbon for agricultural soil management: from the current status toward a global soil information system. Carbon Manag 10, 567–587 https://doi.org/10.1080/17583004.2019.1633231.
Priori, S., Fantappie, M., Bianconi, N., Ferrigno, G., Pellegrini, S., Costantini, E.A.C., 2016. Field-scale mapping of soil carbon stock with limited sampling by coupling gamma-ray and Vis-NIR spectroscopy. Soil Sci. Soc. Am. J. 80, 954–964 https://doi.org/10.2136/sssaj2016.01.0018.
Roudier, P., Hedley, C.B., Ross, C.W., 2015. Prediction of volumetric soil organic carbon from field-moist intact soil cores. Eur. J. Soil Sci. 66, 651–660 https://doi.org/10.1111/ejss.12259.
Shenk, J., Westerhaus, M., Berzaghi, P., 1997. Investigation of a LOCAL calibration procedure for near infrared instruments. J. Near Infrared Spectrosc. 5, 223–232 https://doi.org/10.1255/jnirs.115.
Sherman, D.M., Burns, R.G., Burns, V.M., 1982. Spectral characteristics of the iron oxides with application to the Martian bright region mineralogy. J. Geophys. Res. Solid Earth. 87 (B12), 10,169–10,180https://doi.org/10.1029/JB087iB12p10169.
Smith, P., Soussana, J.F., Angers, D., Schipper, L., Chenu, C., Rasse, D.P., Batjes, N.H., van Egmond, F., McNeill, S., Kuhnert, M., Arias-Navarro, C., Olesen, J.E., Chirinda, N., Fornara, D., Wollenberg, E., Álvaro-Fuentes, J., Sanz-Cobena, A., Klumpp, K., 2019. How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. Glob. Change Biol. 26, 219–241 https://doi.org/10.1111/gcb.14815.
Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and near infrared spectroscopy in soil science. Adv. Agron. 107, 163–215 https://doi.org/10.1016/S0065-2113(10)07005-7.
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 https://doi.org/10.1016/j.geoderma.2007.12.009.
Stevens, A., Ramirez-Lopez, L., 2013. An Introduction to the Prospectr Package. R Package Version 0.1.3. R Foundation for Statistical Computing, Vienna.
Veum, K.S., Sudduth, K.A., Kremer, R.J., Kitchen, N.R., 2015. Estimating soil quality index with VNIR reflectance spectroscopy. Soil Sci. Soc. Am. J. 79, 637–649 https://doi.org/10.2136/sssaj2014.09.0390.
Viscarra Rossel, R.A., Brus, D.J., Lobsey, C., Shi, Z., McLachlan, G., 2016. Baseline estimates of soil organic carbon by proximal sensing: comparing design-based, model-assisted and model-based inference. Geoderma 265, 152–163 https://doi.org/10.1016/j.geoderma.2015.11.016.
Viscarra Rossel, R.A., Lobsey, C.R., Sharman, C., Flick, P., McLachlan, G., 2017. Novel environmental sensing for monitoring soil organic C stocks and condition. Environ. Sci. Technol. 51, 5630–5641 https://doi.org/10.1021/acs.est.7b00889.
Williams, P., Norris, K.H., 2001. Near-Infrared Technology in the Agricultural and Food Industries, 2nd ed. American Association of Cereal Chemists, St. Paul, MN, USA.
Wold, S., Martens, H., Wold, H., 1983. The multivariate calibration problem in chemistry solved by the PLS method, in: Kågström, B., Ruhe, A. (Eds.), Matrix Pencils. Lecture Notes in Mathematics, vol. 973. Springer, Berlin, Heidelberg, pp. 286–293. https://doi.org/10.1007/BFb0062108.