National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy

Michaël Clairotte, Clovis Grinand, Ernest Kouakoua, Aurélie Thébault, Nicolas N. Saby, Martial Bernoux, Bernard G. Barthès

National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy

Michaël Clairotte [a,b], Clovis Grinand [c,d], Ernest Kouakoua [c], Aurélie Thébault [c,e], Nicolas P.A. Saby [f], Martial Bernoux [c], Bernard G. Barthès [c]*

[a] INRA, UMR Eco&Sols, 2 place Viala, 34060 Montpellier, France

[b] Present address: Joint Research Centre, Institute for Energy and Transport, Sustainable Transport Unit, Ispra 21027, Italy

[c] IRD, UMR Eco&Sols, 2 place Viala, 34060 Montpellier, France

[d] EtcTerra, 127 rue d'Avron, 75020 Paris, France

[e] Present address: ITK, Cap Alpha, avenue de l'Europe, 34830 Clapiers, France

[f] INRA, Unité Infosol, BP 20619, 45166 Olivet, France

*Corresponding author: bernard.barthes@ird.fr phone: +33(0)4.99.61.21.36

Abstract

This study presents the potential of infrared diffuse reflectance spectroscopy (DRS) to predict soil organic carbon (SOC) content. A large national soil library was used, including about 3800 samples collected at two soil depths (0-30 and 30-50 cm) using a 16 × 16 km plot grid over the French metropolitan territory (552,000 km^2). Reflectance spectra were collected in the laboratory using visible and near infrared (VNIR), near infrared (NIR) and mid infrared (MIR) spectrophotometers. The soil data library was broken down into calibration and validation sets through sample selection at random or based on spectral representativeness. The calibration intensity was investigated in order to assess the optimum number of calibration samples required to obtain accurate models. Predictions were achieved using global or local partial least square regression (PLSR) built using VNIR, NIR and MIR spectra separately or in combination. Local PLSR uses only calibration samples that are spectral neighbors of each validation sample, thus builds one model per validation sample. Model performance was evaluated on the validation set based on the standard error of prediction (SEP), the ratio of performance to deviation (RPD_v), and the ratio of performance to interquartile range ($RPIQ_v$).

Using all calibration samples, the global PLSR model provided the most precise predictions of SOC content with the MIR spectra, then with the NIR spectra, and less accurate predictions with the VNIR spectra (SEP = 2.6, 4.4 and 4.8 g kg^{-1}, RPD_v = 2.7, 2.3 and 1.5, and $RPIQ_v$ = 3.3, 2.2

and 1.9, respectively). The combination of spectral libraries did not improve model performance noticeably. Local PLSR provided better models than global PLSR, allowing accurate predictions with only 30% of the calibration set, whatever the spectral library (RPD_v and $RPIQ_v$ > 2.0). Optimum calibration intensity was estimated at about 60% for MIR spectra with both global and local PLSR, 30-40% for VNIR and NIR spectra with global PLSR, but 50% for VNIR spectra and 70% for NIR spectra with local PLSR. The most accurate models, which were obtained using the MIR spectra and local PLSR with calibration intensity higher than 50%, allowed very good SOC determination for the most frequent French soils (SEP < 2 g kg^{-1}). This highlights the potential of infrared DRS for national SOC monitoring, provided that calibration database is strengthened with samples from less frequent soil types.

Keywords: Soil organic carbon; Near infrared reflectance spectroscopy (NIRS); Mid infrared reflectance spectroscopy (MIRS); Global regression; Local regression.

Highlights

- Infrared spectroscopy was used to predict SOC content in a national soil database
- Spectral range and calibration sampling intensity and procedure were investigated
- For the samples and spectrometers studied, mid infrared was the most informative
- Optimum calibration intensity was 50% on average, regardless of the spectral range
- Local calibration provided the most accurate predictions, with error < 2 g C kg^{-1}

1. Introduction

Organic carbon is a key component in soil, where it plays a central role in essential functions. Soil organic carbon (SOC) quantification enables to assess soil quality through its structural stability, water retention, as well as chemical and biological fertility (Vaughan and Malcolm, 1985; Reeves, 1997). In addition, as the greatest terrestrial carbon pool, SOC is involved in global carbon cycling, thus in global warming (Batjes, 1996; Lal, 2004), with, for instance, an estimated amount of about 75 Mt SOC in the EU-27 (Schils et al., 2008). However, SOC declines across Europe to such an extent that it has become a dramatic threat (Lugato et al., 2014; de Brogniez et al., 2015).

With its ease to use, fast implementation and low cost, infrared diffuse reflectance spectroscopy (DRS) has become increasingly popular these last decades to estimate SOC (Gholizadeh et al., 2013; Soriano-Disla et al., 2014). In comparison with the traditional laboratory methods, infrared DRS does not require chemical reagents and tedious sample preparation, therefore, it can be quickly implemented in both laboratory and field conditions (Gras et al., 2014). The first studies demonstrating the capability of infrared DRS to determine SOC were based on visible and near infrared (VNIR), or only near infrared (NIR) spectral ranges (Ben-Dor and Banin, 1995; Shepherd and Walsh, 2002; Brown et al., 2005). Infrared DRS based on the mid infrared (MIR) spectral range has also demonstrated its ability to quantify SOC (Janik and Skjemstad, 1995; Grinand et al., 2012). Regarding information related to the chemical structure, the visible range is dominated by electronics transitions, the NIR range by the weak overtones and combinations of fundamental vibrations bands, and the MIR range by the fundamental vibrational bands for H-C, N-H and O-H bonds (Reeves, 2010).

Infrared DRS is generally based on calibrations, which require samples that have been characterized both conventionally (in the laboratory) and spectrally. Calibration models valid for large regions require large soil data libraries, and such libraries require demanding and expensive sampling and analysis campaigns (Nocita et al., 2015). Hence, calibration of soil properties has often been built from the scan of archived soil libraries (e.g., Genot et al., 2011; Viscarra Rossel and Webster, 2012; Shi et al., 2015). At the global scale, the ICRAF-ISRIC soil spectral library is composed of 4438 samples originating from Africa, Asia, Europe and America (Reeuwijk, 1992). At the continental scale, there are several large soil spectral libraries, regarding Australia (21,500 spectra from samples collected during a lot of surveys; Viscarra Rossel and Webster, 2012), the United States (144,833 spectra from samples collected under the Rapid Carbon Assessment project; USDA, 2013), and Europe (20,000 spectra from samples collected under the Land Use/Cover Area Frame Statistical Survey – so-called LUCAS; Stevens et al., 2013). These large global- and continental-scale soil libraries were scanned in the VNIR range. There is also a French national-scale soil library, which has been scanned in both the VNIR (Gogé et al., 2012) and MIR (Grinand et al., 2012). Several studies have addressed the question of the spectral range most appropriate for predicting SOC concentration, considering VNIR, NIR and/or MIR (Viscarra Rossel et al., 2006; Brunet et al., 2008; Rabenarivo et al., 2013). However, the spectral libraries used in these comparisons included rather limited numbers of samples, and had local or

regional application domains. Thus comparative studies based on large national soil libraries are scarce.

The scope of the present study was to assess the performances of SOC predictions made with a national soil spectral library, using VNIR, NIR and MIR spectra acquired with different spectrophotometers, individually or in combination. These instruments differed in terms of spectral range but also in terms of conception (e.g. more or less sophisticated dispersive elements); thus performances and comparisons did not refer strictly to spectral ranges, but rather, to spectral ranges and technologies. Moreover, different calibration strategies were tested:

- the proportion of calibration samples was varied for optimization purposes;
- calibration samples were either selected at random or according to their spectral representativeness;
- and calibration was either global (one unique model was built using all calibration samples) or local (one individual model was built for each validation sample, using calibration samples that were its spectral neighbors).

2. Materials and methods

2.1. The soil library

The studied soil samples belong to a large national soil library provided by the French national soil quality monitoring network (RMQS, Réseau de mesures de la qualité des sols). The RMQS aims at providing a national overview of soil quality, identifying gradients, monitoring the evolution of soil quality over time (with a frequency of a decade), and building a bank of soil samples (Arrouays et al., 2002). This soil library was built over a 10-year sampling campaign over the 552,000 km^2 of the French metropolitan territory. The sample design is based upon a square grid with spacing of 16 km (Figure 1). At the center of each square, 25 individual core samples were taken at 0-30 and 30-50 cm depth using an unaligned sampling design within a 20 × 20 m area, and were then bulked to obtain composite samples (Arrouays et al., 2002). In total, the RMQS national soil library is composed by more than 2200 sites and 3800 samples covering numerous soil types: Cambisols, Calcosols, Luvisols, Leptosols, Andosols, Albeluvisols, etc. (FAO, 2014).

2.2. Reference analyses

Reference determinations were achieved in a single ISO/CEI 17025:2005 accredited laboratory (INRA soil analysis laboratory, Arras, France), according to the standard French procedure NF ISO 10694 (ISO, 1995a). In brief, the method consists in a dry combustion of the air dried, 2-mm sieved then finely ground (< 0.25 mm) sample at circa 1200°C in an oxygen-rich atmosphere. Thus, carbon is converted into carbon dioxide, and quantified using a thermal conductivity detector (TCD). This was done using a CHN elemental analyzer (Flash 2000, Thermo Scientific, Bremen, Germany). This analysis leads to the determination of total carbon content (g kg^{-1}), which was then corrected for inorganic carbon possibly present in the sample as carbonates. This was done according to the French procedure NF ISO 10693 (ISO, 1995b), which consists of measuring the volume of carbon dioxide produced after the addition of chlorhydric acid (Pansu and Gautheyrou, 2006). In order to not unduly influence the prediction models, samples with SOC content higher than the mean plus 10 times the standard deviation were considered outliers and removed from the studied sample population. This led to the removal from the soil library of three samples with SOC > 200 g kg^{-1}. The SOC content of the library then ranged from 0.6 to 177 g kg^{-1}, averaged 18.9 g kg^{-1}, and had a skewness value close to 3 (Figure 2).

2.3. Spectral analysis

The soils were air dried, 2-mm sieved then finely ground (< 0.2 mm), and oven dried overnight at 40°C before spectral analysis. Three spectrophotometers were used in this study. Reflectance spectra in the VNIR region were acquired between 350 and 2500 nm (ca. 28,500 and 4000 cm^{-1}, respectively) at 1 nm interval using a portable system LabSpec 2500 (Analytical Spectral Devices, Boulder, CO, USA). This spectrophotometer is equipped with a halogen lamp source, a fixed diffraction grating, and a diode array detector. Soil samples were scanned manually with a contact probe (surface area scanned: 80 mm^2). Each VNIR spectrum resulted from the averaging of 32 co-added scans, and absorbance zeroing was carried out every hour using a reference standard (Spectralon, i.e. polytetrafluoroethylene). VNIR spectra were used in the range from 400 to 2500 nm (25,000 and 4000 cm^{-1}, respectively), due to frequent noise at the spectrum lower end. Reflectance spectra in the NIR region were acquired between 1100 and 2500 nm (ca. 9091 and 4000 cm^{-1}, respectively) at 2 nm interval using a Foss NIRSystems 5000 (Laurel, MD, USA). This spectrophotometer is equipped with a halogen lamp source, a scanning grating as dispersive

element and a PbS (lead sulfide) detector. Soil samples were placed in a ring cup with a quartz bottom, gently packed using a specific round cardboard, then scanned through the quartz window using a feeder for ring cups (soil surface area scanned: 42 mm^2). Each NIR spectrum resulted from 32 co-added scans, and internal reference standard was scanned automatically before each sample. Finally, reflectance spectra in the MIR region were acquired between 4000 and 400 cm^{-1} (2500 and 25,000 nm, respectively) at 3.86 cm^{-1} interval using a Fourier transform Nicolet 6700 (Thermo Fischer Scientific, Madison, WI, USA). This spectrophotometer is equipped with a silicon carbide source, a Michelson interferometer as dispersive element, and a DTGS (deuterated triglycine sulfate) detector. Soil samples were placed in a 17-well plate, where their surface was flattened with the flat section of a glass cylinder, and they were then scanned using an auto-sampler (soil surface area scanned: ca. 10 mm^2). Each MIR spectrum resulted from 32 co-added scans, and the body of the plate (beside wells) was used as reference standard and scanned once per plate (i.e. every 17 samples). MIR spectra were used in the range from 4000 to 450 cm^{-1} (2500 and ca. 22,000 nm, respectively) because the spectrum end was often noisy. According to suppliers' recommendations, spectrometer warm-up was generally 30 min for the VNIR instrument, 60 min for the NIR instrument, while the MIR instrument was always turned on. To a wider extent, the spectrum acquisition procedure was kept as constant as possible for a given instrument, to limit the nonsystematic noise effects (Pimstein et al., 2011). Moreover, during the full scanning process, the spectrometers and associated accessories (e.g. the contact probe of the VNIR instrument) were not substituted and were properly maintained, in order to minimize the systematic noise effects (Pimstein et al., 2011; Ben-Dor et al., 2015).The reflectance spectra were converted into absorbance spectra [$\log_{10}(1/R)$] and gathered to form one global library (GL) per spectral range. In addition to the three GLs created using the different spectrophotometers, the spectra were combined to form extended spectral libraries. The NIR spectra acquired with the Foss instrument and the MIR spectra were combined to create an extended library called hereinafter NIR+MIR. In the same way, another extended library composed by the so-called "visible" region of the VNIR (400 to 1100 nm), the NIR spectra acquired with the Foss instrument (1100 to 2500 nm) and the MIR region (2500 to 22,000 nm) was formed and called VIS+NIR+MIR.

Once the three outliers removed (cf. section 2.2), sample number varied from 3781 to 3844 according to the GL considered, depending on the possible removal of replicated samples used for controlling reference analyses or exhaustion of some samples between scanning campaigns.

2.4. Selection and number of calibration samples

2.4.1. Selection of a tuning set

Based on spectral representativeness, each GL was divided into calibration, validation and tuning sets, hereinafter referred to as CAL.set, VAL.set, and TUN.set, respectively. Sample selection was based on the Kennard-Stone algorithm (Kennard and Stone, 1969), which was implemented in R 3.1.0 (R Development Core Team, 2011) together with the package "prospectr" (Stevens and Ramirez-Lopez, 2013). CAL.set included the 80% most representative spectra, VAL.set the 10% best represented ones (i.e least representative), and TUN.set the 10% remaining ones (Figure 3). This first operation aimed at single out the TUN.set required for the tuning stage of the local regression.

2.4.2. Optimal calibration intensity

The aim of this stage was to determine the minimum number of calibration samples required to ensure correct predictions. This was done by varying the number of calibration samples, following two selection strategies: random selection and selection based on spectral representativeness.

- For random selection, CAL.set and VAL.set were merged to form a secondary library (SL; cf. Figure 3) and 380 samples were randomly selected (out of more than 3400) to create a new validation set, referred to as val.set. Then, the calibration intensity was investigated by selecting randomly from 10% to 90% of the samples within the remaining calibration set, referred to as cal.set; and finally, the highest calibration intensity was achieved using 100% of cal.set. Random selection was performed 10 times for each calibration intensity between 10% and 90%. In order to mitigate the influence of the initial val.set selection, the whole operation was carried out 10 times (i.e. with 10 different val.set).
- When based on spectral representativeness, sample selection involved the Kennard-Stone algorithm (Kennard and Stone, 1969). VAL.set was used for validation, while an increasing number of calibration samples was selected from CAL.set (from 10% to 100%).

Calibration intensity with random selection of calibration samples was investigated in Grinand et al. (2012), but with the substantial difference that samples not used for calibration were transferred to the validation set, making its composition different at each calibration intensity level. Here, whatever the calibration set selection and size, the size of the validation set was constant (380 samples). However, the VAL.set composition changed according to the spectral library considered, because composition depended on spectral representativeness, which depended on the spectral range.

The SOC content in CAL.set ranged from 0.6 g kg^{-1} to 177 g kg^{-1}, with an average (ca. 20 g kg^{-1}) and a standard deviation (SD; ca. 19.5 g kg^{-1}) nearly constant whatever the spectral library considered (Table 1). There were small differences in mean and SD between the different CAL.set because the set size was not exactly the same and because the most representative samples were not necessarily the same. In contrast, the mean and SD of TUN.set and, to a greater extent, VAL.set, varied noticeably according to the spectral library.

2.5. Regression procedures

Partial least squares regression (PLSR) was used to infer SOC content from the spectra. This was done using either global or local PLSR. Global PLSR, which is the common PLSR procedure, uses all calibration samples to be used (depending on the calibration intensity) for building a unique model that will be applied on all validation samples (Shenk and Westerhaus, 1991). The number of latent variables that minimized the root mean square error of cross-validation (RMSECV) was retained for the prediction model. The cross-validation was carried out by dividing the calibration set into six groups composed of a nearly equal number of randomly selected samples. Global PLSR was implemented in R together with the package “pls” (Mevik and Wehrens, 2007).

In contrast, local PLSR makes prediction for each sample individually, only using calibration samples that are its spectral neighbors (Shenk et al., 1997). These neighbors were selected according to two metrics: the correlation coefficient between spectra and the Mahalanobis distance H (i.e. in the principal component space; Mark and Tunnell, 1985). The distance H was calculated on the number of principal components that minimized the root mean square (RMS) deviation of SOC difference between every calibration sample and its closest spectral neighbor, so that spectral neighbors also tended to be "SOC neighbors", according to Ramirez-Lopez et al.

(2013a). Moreover, the number of latent variables was not determined through cross-validation, which would require long calculation. Instead, each prediction was calculated as the weighted average of the predicted values generated with 7 to 25 latent variables (different maximums were tested), each weight being calculated as the inverse of the product of the RMS of X-residuals (i.e. the difference between the actual spectrum and the spectrum approximated using the considered number of latent variables) and RMS of the regression coefficients (Shenk et al., 1997; Zhang et al., 2004).

In order to not over-fit local PLSR, the optimum number of neighbors and optimum distance metric were determined on TUN.set (Figure 3). Then, these optimum parameters were used to predict SOC content on VAL.set. As local PLSR requires a lot of computational time to build one model per sample to predict, it was only carried out when the calibration samples were selected based on their spectral representativeness. Local PLSR was implemented in R together with the spectrum-based learner algorithms from the package "resemble" (Ramirez-Lopez et al., 2013b).

For both global and local PLSR and each spectral library, different and common spectrum pre-processing methods were tested, in order to identify the most appropriate: mean centering, variance scaling, moving average, standard normal variate (SNV, i.e. mean centering and variance scaling), Savitzky-Golay smoothing and derivative, and continuum removal. The pre-processing yielding the lowest RMSECV was selected. Pre-processing methods were implemented in R together with the package "prospectr" (Stevens and Ramirez-Lopez, 2013).

The accuracy of the prediction models was estimated on the validation set (either val.set after random selection of calibration samples or VAL.set after Kennard-Stone selection) by computing the standard error of prediction (SEP), coefficient of determination (R^2_v), ratio of performance to deviation ($RPD_v = SD_v/SEP$, where SD_v is the SD of the validation set) and ratio of performance to interquartile range ($RPIQ_v = IQ_v/SEP$, where IQ_v is the interquartile range of the validation set), more appropriate than RPD_v for non-normal distributions (Bellon-Maurel et al., 2010).

Neither spectral outlier detection nor calibration outlier detection were achieved. The only samples removed from spectral analyses were the three with SOC content > 200 g kg^{-1} (cf. section 2.2).

3. Results and discussion

3.1. Spectrum pre-processing

The best pre-processing method was identified on each GL through cross-validation. Moving average on 10 bands followed by Savitzky-Golay first derivative gave the best cross-validation for the MIR and VIS+NIR+MIR libraries; but moving average on 10 bands was enough for VNIR and NIR libraries (data not shown). This contrasted with the results of Vasques et al. (2008), where Savitzky-Golay first derivative was the best pre-processing method for VNIR spectra, in agreement with the review of Gholizadeh et al. (2013) on infrared DRS for SOC content prediction. No pre-processing functions improved the results for the NIR+MIR library.

3.2. Global PLSR

The validation results are presented in Figure 4, for the different selection strategies, spectral libraries and calibration intensities. More detailed results for 30%, 50% and 100% calibration intensity are shown in Table 2.

3.2.1. Overall validation results after random selection of calibration samples

In that case SEP decreased smoothly with increasing calibration intensity, except at low calibration intensities, especially for NIR, MIR and NIR+MIR libraries (Figure 4A). Obviously, the more samples allocated in the calibration set, the more information available to build accurate models; however the benefit of 10% additional calibration samples was very limited at high calibration intensities (from 60-70%). In the same way, RPD_v and $RPIQ_v$ increased progressively with calibration intensity, with limited increase at high intensities (Figures 4C and 4E).

The libraries including the MIR spectra provided greater accuracy than the other spectral libraries: the best models were achieved with the MIR, NIR+MIR and VIS+NIR+MIR libraries (SEP = 6 g kg^{-1}, $RPD_v \geq 3$ and $RPIQ_v = 2.8$ at 100% calibration intensity), while NIR and to a greater extent VNIR libraries yielded noticeably poorer results (8-9 g kg^{-1}, 2.1-2.3 and 1.9-2.1, respectively). These results are consistent with RPD values reported by a regional-scale study (3585 km^2) carried out in Florida using VNIR spectra (Vasques et al., 2008).

Hierarchies in SEP and $RPIQ_v$ were similar across spectral libraries but RPD_v behaved sometimes differently (Table 2). Indeed, random segmentation of the secondary library (cf. section 2.4 and

Figure 3) caused small differences in SD_v and thus RPD_v, while IQ_v and thus $RPIQ_v$ were poorly affected.

3.2.2. Overall validation results after selection of spectrally representative calibration samples

Model accuracy increased more markedly with calibration intensity after Kennard-Stone selection than after random selection of calibration samples: from 10% to 100% calibration intensity, averaged over the five spectral libraries, SEP decreased twice more after Kennard-Stone selection than after random selection (3.4- vs. 1.7-g kg^{-1} decrease), RPD_v increased twice more (1.3 vs. 0.6 increase) and $RPIQ_v$ three times more (1.5 vs. 0.5 increase; Figures 4B, 4D and 4F). Moreover, adding calibration samples was useful even at high calibration intensities. This indicated that the calibration set contained more information useful for characterizing the validation samples when calibration samples were selected according to their spectral representativeness, which makes sense. Furthermore, increase in prediction accuracy with calibration intensity was not completely regular, and some disruptions occurred, for instance with the NIR spectra at 20-30% intensity. Indeed, increasing calibration intensity was carried out by increasing the number of calibration samples spectrally representative of CAL.set, which did not necessarily correspond to a better representation of VAL.set (cf. Figure 3).

The most accurate models were obtained with the MIR and NIR+MIR spectra (SEP = 2.3-2.6 g kg^{-1}, RPD_v = 2.7-2.8 and $RPIQ_v$ = 3.3 at 100% calibration intensity), the least accurate with the VNIR spectra (4.8 g kg^{-1},1.5 and 1.9, respectively), while intermediate results were achieved with the NIR and VIS+NIR+MIR spectra (RPD_v = 2.3-2.4 but SEP = 4.4 and 2.4 g kg^{-1}, and RPIQv = 2.2 and 3.1, respectively, due to noticeable differences in SD_v; Table 2). For the VNIR library, RPD_v increased relatively little with calibration intensity and always remained below the often cited threshold of 2, beyond which soil property prediction models have been considered suitable (Chang et al., 2001). Comparisons on the validation set between SOC content measured (reference) and predicted using the different spectral libraries and global PLSR built with 100% of the calibration samples selected by the Kennard-Stone procedure are presented in Figure 5A.

As seen for random selection but to a greater extent here, the way spectral ranges ranked was not the same according to SEP or RPD_v: for calibration intensity $\geq$ 30%, SEP was moderately higher with VNIR than with NIR spectra, and much lower with the libraries that included the MIR spectra; while RPD_v discriminated low values for the VNIR library, much higher values for NIR

and VIS+NIR+MIR libraries, and even higher values for MIR and NIR+MIR libraries. This discrepancy was due to differences in SD_v, (the SD_v of VAL.set ranged from 5.8 g kg^{-1} to 10.3 g kg^{-1} depending on the library). There was less variation in IQ_v (from 7.5 to 9.9 g kg^{-1}), and as a consequence, $RPIQ_v$ and SEP reflected more comparable trends, in particular, noticeably better results with VIS+NIR+MIR than with NIR spectra. These effect of spectral range and instrument are discussed more specifically in section 3.2.5

3.2.3. Respective variations of SEP, RPD_v and $RPIQ_v$ after Kennard-Stone vs. random selection of calibration samples

Surprisingly, though SEP was much lower after Kennard-Stone than after random selection of the calibration samples (-50% in average at 100% calibration intensity), in general RPD_v was lower too because SD_v was even much lower. In contrast, $RPIQ_v$ was often higher after Kennard-Stone than after random selection, especially at high calibration intensities, because SEP decreased more than IQ_v (reminder: $RPIQ_v = IQ_v/SEP$ while $RPD_v = SD_v/SEP$). Indeed, after Kennard-Stone selection, samples with extreme SOC values were more likely to be in the calibration set, and logically this affected SD_v more markedly than IQ_v. This underlines that RPIQ is a more appropriate performance parameter than RPD when considering very skewed variable distributions (Bellon-Maurel et al., 2010).

Moreover, after random selection of calibration samples, RPD_v was higher than $RPIQ_v$, but except for the NIR library, the opposite was seen after Kennard-Stone selection. This also related to SD_v and IQ_v variations according to the selection procedure. As a result of random selection of calibration samples, the total, calibration and validation sets indeed had similar SD, and similar IQ. In contrast, Kennard-Stone selection of calibration samples resulted in higher SD and IQ in the calibration set than in the total set, and to a greater extent, than in the validation set. Thus SD_v and IQ_v were smaller after Kennard-Stone than after random selection of calibration samples, and the decrease was higher for SD_v than for IQ_v, as mentioned above.

3.2.4. Calibration intensity

After random selection of calibration samples, the improvement of the model performance with calibration intensity was greater when the MIR spectra were involved: RPD_v (or $RPIQ_v$) increased by 30% over the full intensity range, compared to 15% for NIR or VNIR spectra

(Table 2, Figures 4C and 4E). Optimum calibration intensity (i.e. minimum number of calibration samples required to obtain a sufficiently accurate model) could be set at 50-60% of the calibration set when the MIR spectra were involved, but at only 30-40% for the NIR and VNIR libraries. The models with VNIR or NIR spectra reached their optimum more rapidly, probably because these spectra contained less useful information for predicting SOC, and less calibration samples were needed for extracting this information.

After Kennard-Stone selection of calibration samples, optimum calibration intensity was more difficult to draw up. This optimum was generally reached with larger calibration sets than after random selection. The best cost/benefit tradeoff could be set at 50% calibration intensity for the VNIR library, 60% for the MIR library, 70% for the NIR and NIR+MIR libraries, and 100% for the VIS+NIR+MIR library (Table 2, Figures 4D and 4F). It is worth noting that such optimum depends on the specific context of each given study, especially the financial means available (more funding allows higher calibration intensity) and the objective of the study (e.g. detecting small SOC changes requires high prediction accuracy thus high calibration intensity).

Considering a MIR database that included only the topsoil samples from the present study (i.e. 2086 samples ground at 0.2 mm), Grinand et al. (2012) performed global PLSR with random selection of calibration samples (10 replicates) and achieved optimal calibration intensity at 40% of the total population ($RPD_v = 3.2$). However, their validation set included all samples not used for calibration, thus its size decreased when calibration intensity increased, which did not allow easy comparison with the present study (where 100% calibration intensity actually represented 80% of the total set). Shepherd and Walsh (2002) used multivariate adaptive regression splines (MARS, a data mining approach) for predicting SOC content in a very diverse VNIR dataset of 2-mm sieved topsoil samples from eastern and southern Africa (1100 samples from seven countries representing 10 soil orders). Considering a constant validation set, they achieved optimum calibration intensity with 35% randomly selected samples (out of the total set). Brown et al. (2005) studied a more homogeneous set made of profile samples (ca. 300 samples, 2 mm sieved) from six sites in north central Montana (200-km wide area; USA). Using VNIR and global PLSR for predicting SOC content, they reached an optimum when using 50% randomly selected samples for calibration (out of the total set), but 33% only when randomly selecting samples grouped by profile. Actually, optimal calibration intensity depends on the soil library, on its diversity especially, which renders comparison between studies difficult.

3.2.5. Effects of spectral range and instrument on SOC prediction

On the whole, prediction was more accurate when the MIR spectra were involved. This is consistent with the literature review published by Bellon-Maurel and McBratney (2011), which stated that SOC content predictions made with spectra acquired on dried and ground sample were more accurate using MIR. This has been attributed to the fact that the MIR region displays much more details and is dominated by intensive vibration fundamentals, while the NIR region is dominated by weaker and broader signals from vibration overtones and combination bands (Janik et al., 1998; McCarty et al., 2002). However, most comparisons between SOC content predictions made with MIR vs. NIR spectra have been obtained on soils from temperate regions. Different results might be observed with soils from tropical regions, as indicated by works that have reported more accurate SOC predictions with NIR than MIR spectra in Brazil (Madari et al., 2006) or Madagascar (Rabenarivo et al., 2013). In addition, the results of the present study confirmed that VIS does not provide much additional information (Brunet et al., 2008), at least with the spectrometers used in this study. It is worth noting that most comparisons between SOC predictions made using different spectral regions were based on local or regional datasets, composed of a few soil types, which limited their significance when compared with the results of the present study.

Regardless of the selection strategy and the calibration intensity, prediction models were not improved notably by the combination of libraries (VIS+NIR+MIR and NIR+MIR), when compared with the MIR library alone. This is in accordance with Viscarra Rossel et al. (2006), who observed similarly accurate SOC predictions when using VIS+NIR+MIR and MIR spectra on a local-scale soil library (17.5 ha, 118 samples). MIR spectra contain the fundamental vibrational bands of the main chemical bonds found in organic matter, and information contained in the VIS and NIR ranges may be redundant, at least for soils from temperate regions.

The instrument technology might also contribute to the differences observed in prediction accuracy. Firstly, the different dispersive elements used in the VNIR and NIR spectrometers (fixed grating and scanning grating elements, respectively) could explain why the narrower NIR spectra provided better model accuracy than the broader VNIR spectra. The VNIR spectrometer used in the study is a portable instrument, dedicated to both laboratory and field measurements. Consequently, it is equipped with a robust fixed grating element (without mobile parts and

sophisticated optical system), which allows fast measurements, but as a tradeoff, is slightly less accurate than a scanning grating element. Secondly, there were differences in the scanning procedure. The acquisition of VNIR spectra was done manually using a contact probe, and zeroing was done by the operator, every hour. In contrast, NIR and MIR spectrum acquisitions were automated, with internal zeroing before every scan. Suggestions have been proposed in order to improve spectrum acquisition with the VNIR spectrometer considered, mainly through the control of indirect interferences with external factors (i.e. duration of the spectrometer warm-up, type and number of reference acquisitions prior sample acquisition, etc.) and online spectrum correction using an internal soil standard (Pimstein et al., 2011; Ben-Dor et al., 2015). Furthermore, with the spectrometer used, MIR spectra are acquired on a very small surface area (10 mm^2 vs. 42 and 80 mm^2 for the NIR and VNIR spectrometers, respectively), which makes it necessary to grind the samples finely (< 0.2 mm) to limit its heterogeneity. Grinding is not particularly required for the two other instruments though it often results in better SOC content predictions, as stated for NIR by Barthès et al. (2006) and Brunet et al. (2007).

3.2.6. Interpretation of regression coefficients

Regression coefficients of global PLSR models obtained using all calibrations samples selected by the Kennard-Stone procedure with NIR, MIR and NIR+MIR spectra were investigated, after SNV transformation of NIR and MIR spectra (Figure 6). The overall pattern expressed by the MIR regression coefficients was for the most part similar when the regression was performed using MIR alone and the combination NIR+MIR (R^2 = 0.71); nevertheless some regions had heavy contributions in one case only (e.g. 2780 nm positively for MIR alone and 2700 nm negatively for NIR+MIR), or their respective contribution might change (e.g. peaks at 3300 and 6000 nm). On the whole, similar heavily contributing bands were also observed in the NIR region when considering NIR alone and NIR+MIR (R^2 = 0.41), but their respective weights might vary (e.g. highest positive weight at 2350 nm for NIR alone and at 2100 nm for NIR+MIR), and some were even flattened dramatically (e.g. the region 1500-1800 nm from NIR alone to NIR+MIR). It is worth noting that for the combination NIR+MIR, the heaviest contributions were in the MIR, which might confirm that the MIR spectra contain more useful information for SOC content prediction than the NIR spectra, at least for the soils considered.

The absorption peak A around 1910 nm (5236 cm^{-1}; cf. Figure 6) has been assigned to the bending and stretching vibrations of the O-H bounds in free water (Viscarra Rossel et al., 2006). Viscarra Rossel and Webster (2012) found that absorbance at this wavelength was crucial in the model trees they used to predict SOC content from VNIR spectra. The wide absorption peak B around 2050-2150 nm (4878-4651 cm^{-1}) has been attributed to different organic molecules such as amides or proteins (2050-2060 nm especially), polysaccharides (2100 nm) and lipids (2140 nm; Workman and Weyer, 2008). The absorption peak C around 3460 nm (2890 cm^{-1}) has been assigned to aliphatic C-H bounds of methyl and methylene groups (Vohland et al., 2014), which is not particularly informative, and the peak D around 5747 nm (1740 cm^{-1}) to C=O bound from carboxyl acids, aldehydes and ketones (Janik et al., 1998). Both strong negative peaks E around 4000 nm (2500 cm^{-1}) and F around 5525 nm (1810 cm^{-1}) have been assigned to carbonates (Miller and Wilkins, 1952; Grinand et al., 2012).

3.3. Local PLSR

In order to limit computation time, local PLSR was carried out only for a few calibration intensities identified as significant using global PLSR: 30%, 50% and 100% (Table 3; Figures4B, 4D and 4F). Prediction of SOC content was greatly enhanced with local PLSR; however the performance ranking of spectral libraries remained globally similar in comparison with the global PLSR: the best models were obtained using the MIR spectra (or combinations including the MIR spectra) and the least accurate using the VNIR spectra only. When the MIR spectra and 100% of the calibration set were used, the best models had $SEP \leq 2$ g kg^{-1}, $R^2_v > 0.9$, $RPD_v > 3$ and $RPIQ_v > 4$, which has never been reported at that scale to date. This of course concerned the best represented samples of the total population; nevertheless it may be hypothesized that such very good results could be extended to a larger proportion of the French territory by enriching the calibration database with soil types poorly represented currently. It is worth noting that local PLSR produced good models (RPD_v and $RPIQ_v \geq 2$) with all spectral libraries even at only 30% calibration intensity. Moreover, the difference in performance between prediction models using VNIR and NIR spectra was less marked using local than global PLSR. Globally, this improvement through local PLSR tended to decrease when calibration intensity increased (e.g. at 30%, 50% and 100% calibration intensity, $RPIQ_v$ increased by 53%, 44% and 33% in average, respectively) and was higher for the VNIR library (+56% $RPIQ_v$ in average considering the three

intensities vs. +40% in average for the four other libraries). Noticeable performance increase from global to local PLSR was also observed by Shi et al. (2015) on a large Chinese dataset studied in the VNIR range (2732 topsoil samples, with 90% representative in terms of SOC content used for calibration): SEP decreased from 8.4 to 6.6 g kg^{-1}, RPD_v increased from 1.4 to 1.8, and $RPIQ_v$ from 1.5 to 2.3; performance was even better when local PLSR was carried out with calibration samples that were simultaneously spectral and geographical neighbors of validation samples (SEP, RPD_v and $RPIQ_v$ reached 6.0 g kg^{-1}, 2.0 and 2.6, respectively).

In the present study, optimal local PLSR parameters were estimated on the TUN.set. The optimal distance metric was systematically the Mahalanobis distance H for VNIR and NIR libraries, in accordance with Ramirez-Lopez et al. (2013a). However, the optimal distance metric was the correlation coefficient for the MIR library, while no systematic optimal distance metric could be found for combinations of spectral libraries. For the calibration intensities studied, the optimal number of neighbors ranged from 210 to 300 samples for VNIR, from 150 to 360 for NIR, from 180 to 540 for MIR, from 180 to 690 for NIR+MIR, and from 330 to 510 for VIS+NIR+MIR. It tended to increase with calibration intensity, especially when the MIR spectra were involved. Comparisons on the validation set between SOC content measured (reference) and predicted using VNIR, NIR, MIR, NIR+MIR or VIS+NIR+MIR spectra and global or local PLSR built with 100% of the calibration samples selected by the Kennard-Stone procedure are presented in Figure 5.

Using the best parameters defined above, it was possible to perform local PLSR on the whole spectral library to predict SOC on new samples originating from France. This was tested with the MIR library, for each sample of the current database successively, using all other samples of the database as potential calibration samples (depending on their distance to the sample to be predicted). In that manner, the Figure 7 presents maps of SOC content in France at 0-30 and 30-50 cm depth, using reference data vs. MIR predictions. For a given depth layer, both maps displayed a strong similarity, with a $RPIQ_v$ of 3.4 (RPD_v = 3.9) obtained for the full GL.

In the best conditions, the present study yielded an accuracy for SOC prediction (SEP) below 2 g kg^{-1}, not far from standard errors of laboratory, which is likely to render DRS more widely acceptable. Indeed, SEP reported in the literature for large database to date are > 3 g kg^{-1} (e.g. Shepherd and Walsh, 2002; Genot et al., 2011; Gogé et al., 2012; Grinand et al., 2012; Viscarra Rossel and Webster, 2012; Stevens et al., 2013; Shi et al., 2015). It must however be kept in mind

that the present study assessed infrared DRS prediction performance under the most favorable conditions, which are rarely met for large spectral libraries: spectra of a given range acquired with the same spectrometer, on finely ground samples (< 0.2 mm); against reference values measured by the same method and same ISO 17025 laboratory.

4. Conclusion

Over the French national soil library, much better predictions were achieved (i) using MIR than NIR or VNIR spectra, (ii) after Kennard-Stone than after random selection of calibration samples, and (iii) using local than global PLS regression. Information contained in each spectral range might explain this ranking, but instrument technology (dispersive system and acquisition mode) might also have an impact. The combination of spectra from different instruments did not improve the results, when compared with MIR spectra alone. Optimal calibration intensity ranged from 30% to 70% and was higher in general with MIR than with NIR or VNIR spectra and with local than global PLSR. However, local PLSR on MIR spectra yielded accurate predictions even using 30% of the calibration samples (SEP = 2.6 g kg^{-1}, RPD > 2.5, RPIQ > 3). The most accurate models, using MIR spectra, local PLSR and all calibration samples, allowed SOC content predictions with SEP $\leq$ 2 g kg^{-1}, RPD > 3 and RPIQ > 4. These results suggested that laboratory infrared DRS may become a standard method to assess SOC.

Acknowledgements

This work was supported by ADEME (Agence de l'environnement et de la maîtrise de l'énergie, which is a French government agency concerned with environmental protection and energy management; contract 0675C0102). RMQS soil sampling and physico-chemical analyses were supported by the GIS Sol, which is a scientific group of interest on soils involving the French Ministry for ecology and sustainable development and Ministry of agriculture, the French National forest inventory (IFN), ADEME, IRD (Institut de recherche pour le développement, which is a French public research organization dedicated to southern countries) and INRA (Institut national de la recherche agronomique, which is a French public research organization dedicated to agriculture s.l.). The work was also supported by the RIME–PAMPA project funded by AFD (Agence française de développement, which is a French public financial institution dedicated to development assistance to developing countries), the French Ministry for foreign

affairs, and the FFEM (Fonds français pour l'environnement mondial, which is a French public funding agency dedicated to environment protection in developing countries). Claudy Jolivet is thanked for his strong involvement in the RMQS monitoring network. Map production assistance was provided by Christina Corbane, who is warmly thanked. Manon Villeneuve, Emmanuel Bourdon, Didier Brunet, Jérôme Lourd, and Clément Robin (IRD) are thanked for their skilful technical assistance and strong investment in the tedious scanning process of the RMQS soil library. Finally, an anonymous reviewer and the editor-in-chief are warmly thanked for their fruitful comments.

References

Arrouays, D., Jolivet, C., Boulonne, L., Bodineau, G., Saby, N., Grolleau, E. 2002. A new initiative in France: a multi-institutional soil quality monitoring network. Comptes Rendus de l'Académie d'Agriculture de France 88, 93–105.

Barthès, B.G., Brunet, D., Ferrer, H., Chotte, J.-L., Feller, C., 2006. Determination of total carbon and nitrogen content in a range of tropical soils using near infrared spectroscopy: influence of replication and sample grinding and drying. J. Near Infrared Spectrosc. 14, 341–348.

Batjes, N.H., 1996. Total carbon and nitrogen in the soils of the world. Eur. J. Soil Sci. 47, 151–163.

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

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.

Ben-Dor, E., Banin, A., 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Sci. Soc. Am. J. 59, 364–372.

Ben-Dor, E., Ong, C., Lau, I.C., 2015. Reflectance measurements of soils in the laboratory: Standards and protocols. Geoderma 245–246, 112–124.

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.

Brunet, D., Barthès, B.G., Chotte, J.-L., Feller, C., 2007. Determination of carbon and nitrogen contents in Alfisols, Oxisols and Ultisols from Africa and Brazil using NIRS analysis: Effects of sample grinding and set heterogeneity. Geoderma 139, 106–117.

Brunet, D., Bernoux, M., Barthès, B.G., 2008. Comparison between predictions of C and N contents in tropical soils using a Vis–NIR spectrometer including a fibre-optic probe versus a NIR spectrometer including a sample transport module. Biosyst. Eng. 100, 448–452.

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 Sci. Soc. Am. J. 65, 480–490.

De Brogniez, D., Ballabio, C., Stevens, A., Jones, R.J.A., Montanarella, L., van Wesemael, B., 2015. A map of the topsoil organic carbon content of Europe generated by a generalized additive model. Eur. J. Soil Sci. 66, 121–134.

FAO (Food and Agriculture Organization), 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 106. FAO, Rome.

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.

Gholizadeh, A., Borůvka, L., Saberioon, M., Vašát, 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.

Gogé, F., Joffre, R., Jolivet, C., Ross, I., Ranjard, L., 2012. Optimization criteria in sample selection step of local regression for quantitative analysis of large soil NIRS database. Chemometr. Intell. Lab. Syst. 110, 168–176.

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.

Grinand, C., Barthès, B.G., Brunet, D., Kouakoua, E., Arrouays, D., Jolivet, C., Caria, G., Bernoux, M., 2012. Prediction of soil organic and inorganic carbon contents at a national scale (France) using mid-infrared reflectance spectroscopy (MIRS). Eur. J. Soil Sci. 63, 141–151.

ISO (International Organization for Standardisation), 1995a. ISO 10694:1995 - Soil Quality - Determination of Organic and Total Carbon after Dry Combustion (Elementary Analysis). ISO, Geneva.

ISO (International Organization for Standardisation), 1995b. ISO 10693:1995 – Determination of Carbonate Content – Volumetric Method. ISO, Geneva.

Janik, L., Skjemstad, J., 1995. Characterization and analysis of soils using mid-infrared partial least-squares .2. Correlations with some laboratory data. Aust. J. Soil Res. 33, 637–650.

Janik, L.J., Skjemstad, J.O., Merry, R.H., 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? Aust. J. Exp. Agric. 38, 681–696.

Kennard, R.W., Stone, L.A., 1969. Computer aided design of experiments. Technometrics 11, 137–148.

Lal, R., 2004. Soil carbon sequestration impacts on global climate change and food security. Science 304, 1623–1627.

Lugato, E., Panagos, P., Bampa, F., Jones, A., Montanarella, L., 2014. A new baseline of organic carbon stock in European agricultural soils using a modelling approach. Global Soil Biol. 20, 313–326.

Madari, B.E., Reeves, J.B., III, Machado, P.L.O.A., Guimarães, C.M., Torres, E., McCarty, G.W., 2006. Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. Geoderma 136, 245–259.

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

McCarty, G.W., Reeves, J.B., III, Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Sci. Soc. Am. J. 66, 640–646.

Mevik, B.-H., Wehrens, R., 2007. The pls package: Principal Component and Partial Least Squares Regression in R. J. Stat. Softw. 18.

Miller, F.A., Wilkins, C.H., 1952. Infrared spectra and characteristic frequencies of inorganic ions. Anal. Chem. 24, 1253–1294.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Ben Dor, E., Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J.A.M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H.,

Soriano-Disla, J.M., Shepherd, K.D., Stenberg, B., Towett, E.K., Vargas, R., Wetterlind, J., 2015. Soil spectroscopy: An alternative to wet chemistry for soil monitoring. Adv. Agron. 132, 139–159.

Pansu, M., Gautheyrou, J., 2006. Chapter 17 — Carbonates. In: Handbook of Soil Analysis. Mineral, Organic and Inorganic Methods. Springer, Berlin Heidelberg, Germany, pp. 593—604.

Pimstein, A., Notesco, G., Ben-Dor, E., 2011. Performance of three identical spectrometers in retrieving soil reflectance under laboratory conditions. Soil Sci. Soc. Am. J. 75, 746-759.

Rabenarivo, M., Chapuis-Lardy, L., Brunet, D., Chotte, J.-L., Rabeharisoa, L., Barthès, B.G., 2013. Comparing near and mid-infrared reflectance spectroscopy for determining properties of Malagasy soils, using global or LOCAL calibration. J. Near Infrared Spectrosc. 21, 495–509.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Viscarra Rossel, R.A., Demattê, J.A.M., Scholten, T., 2013a. Distance and similarity-search metrics for use with soil vis–NIR spectra. Geoderma 199, 43–53.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Stevens, A., Demattê, J.A.M., Scholten, T., 2013b. The spectrum-based learner: A new local approach for modeling soil vis–NIR spectra of complex datasets. Geoderma 195–196, 268–279.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. R Found. Stat. Comput., Vienna.

Reeuwijk, L.P. van., 1992. Procedures for Soil Analysis. International Soil Reference and Information Centre, Wageningen, the Netherlands.

Reeves, D.W., 1997. The role of soil organic matter in maintaining soil quality in continuous cropping systems. Soil Tillage Res. 43, 131–167.

Reeves III, J.B., 2010. Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? Geoderma 158, 3–14.

Schils, R., Kuikman, P., Liski, J., van Oijen, M., Smith, P., Webb, J., Alm, J., Somogyi, Z., van den Akker, J., Billett, M., Emmett, B., Evans, C., Lindner, M., Palosuo, T., Bellamy, P., Alm, J., Jandl, R., Hiederer R., 2008. Review of Existing Information on the Interrelations between Soil and Climate Change CLIMSOIL. Technical Report 2008-048, Wageningen UR, the Netherlands.

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

Shenk, J., Westerhaus, M., Berzaghi, P., 1997. Investigation of a LOCAL calibration procedure for near infrared instruments. J. Near Infrared Spectrosc. 5, 223–232.

Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Sci. Soc. Am. J. 66, 988–998.

Shi, Z., Ji, W., Viscarra Rossel, R.A., Chen, S., Zhou, Y., 2015. Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis–NIR spectral library. Eur. J Soil Sci. 66, 679–687.

Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., Macdonald, L.M., McLaughlin, M.J., 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Appl. Spectrosc. Rev. 49, 139–186.

Stevens, A., Nocita, M., Tóth, G., Montanarella, L., van Wesemael, B., 2013. Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy. PLoS ONE 8, e66409.

Stevens, A., Ramirez-Lopez, L., 2013. An introduction to the prospectr package. R Package Vignette.

USDA (United States Department of Agriculture), 2013. Rapid Carbon Assessment (RaCA) Methodology Sampling and Initial Summary. USDA, National Resources Conservation Service, National Soil Survey Center, Lincoln, NE, USA.

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.

Vaughan, D., Malcolm, R.E., 1985. Soil Organic Matter and Biological Activity. Springer Netherlands, Dordrecht.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75.

Viscarra Rossel, R.A., Webster, R., 2012. Predicting soil properties from the Australian soil visible-near infrared spectroscopic database. Eur. J. Soil Sci. 63, 848–860.

Vohland, M., Ludwig, M., Thiele-Bruhn, S., Ludwig, B., 2014. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. Geoderma 223–225, 88–96.

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

Zhang, M., Xu, Q., Massart, D., 2004. Averaged and weighted average partial least squares. Anal. Chim. Acta 504, 279–289.

Table 1. Descriptive statistics of the calibration, tuning and validation sets. N_s, M_s, SD_s ,IQ_s and Sk_s stand respectively for the number of samples, mean, standard deviation, interquartile range and skewness coefficient of the set s (with s = c for calibration, s = t for tuning and s = v for validation). M, SD, IQ and range are in g kg^{-1}.

Spectral	Calibration set (CAL.set)						Tuning set (TUN.set)						Validation set (VAL.set)					
library	N_c	M_c	SD_c	IQ_c	$Range_c$	Sk_c	N_t	M_t	SD_t	IQ_t	$Range_t$	Sk_t	N_v	M_v	SD_v	IQ_v	$Range_v$	Sk_v
VNIR	3048	20.2	19.3	17.5	0.6-177	2.7	380	13.9	9.4	11.8	0.6-73.2	1.7	380	12.1	7.4	9.3	1.1-45.3	1.3
NIR	3045	19.7	19.5	17.0	0.6-177	2.9	380	16.4	12.3	13.0	0.9-84.6	1.9	380	14.8	10.3	9.9	0.6-67.3	2.0
MIR	3084	20.5	19.6	17.8	0.6-177	2.8	380	14.1	9.8	8.7	1.7-62.8	2.0	380	11.3	6.9	8.5	1.5-55.8	1.9
NIR+MIR	3043	20.4	19.6	18.0	0.6-177	2.8	380	14.8	11.0	9.6	1.7-84.0	2.5	380	11.3	6.5	7.9	2.6-55.8	1.8
VIS+NIR+MIR	3021	20.2	19.2	17.9	0.6-177	2.7	380	14.8	11.3	10.4	0.9-82.0	2.4	380	11.0	5.8	7.5	2.6-35.6	1.1

Table 2. Predictions of SOC content on the validation sets using global PLSR models built with 30%, 50% or 100% of the calibration samples selected at random or according to the Kennard-Stone procedure.

Spectral library	N_c	Random selection (N_v = 380)							Kennard-Stone selection (N_v = 380)						
		SD_v	IQ_v	SEP	Bias	R^2_v	RPD_v	$RPIQ_v$	SD_v	IQ_v	SEP	Bias	R^2_v	RPD_v	$RPIQ_v$
		g kg^{-1}							g kg^{-1}						
Calibration intensity: 30%															
VNIR	900	18.6	16.4	9.2	-0.2	0.76	2.0	1.8	7.4	9.3	5.9	0.3	0.67	1.3	1.6
NIR	900	18.6	16.4	8.5	-0.2	0.80	2.2	2.0	10.3	9.9	6.5	0.2	0.79	1.6	1.5
MIR	900	18.9	16.6	6.7	0.3	0.87	2.8	2.5	6.9	8.5	3.3	-0.5	0.87	2.1	2.6
NIR+MIR	900	19.5	16.7	6.7	0.0	0.88	2.9	2.5	6.5	7.9	4.0	1.9	0.84	1.5	1.8
VIS+NIR+MIR	900	18.0	16.5	6.7	-0.2	0.86	2.7	2.5	5.8	7.5	3.8	1.1	0.82	1.5	1.9
Calibration intensity: 50%															
VNIR	1500	18.6	16.4	8.9	-0.2	0.77	2.1	1.8	7.4	9.3	5.4	0.4	0.69	1.4	1.7
NIR	1500	18.6	16.4	8.3	-0.2	0.81	2.2	2.0	10.3	9.9	5.0	0.1	0.85	2.0	2.0
MIR	1500	18.9	16.6	6.3	0.2	0.89	3.0	2.7	6.9	8.5	2.8	0.2	0.88	2.5	3.0
NIR+MIR	1500	19.5	16.7	6.3	0.1	0.89	3.1	2.7	6.5	7.9	3.0	1.5	0.87	2.0	2.4
VIS+NIR+MIR	1500	18.0	16.5	6.3	-0.2	0.88	2.9	2.7	5.8	7.5	3.2	0.6	0.84	1.8	2.3
Calibration intensity: 100%															
VNIR	3048	18.6	16.4	8.7	-0.2	0.78	2.1	1.9	7.4	9.3	4.8	0.3	0.73	1.5	1.9
NIR	3045	18.6	16.4	8.2	-0.2	0.81	2.3	2.1	10.3	9.9	4.4	0.3	0.87	2.3	2.2
MIR	3084	18.9	16.6	6.0	0.2	0.90	3.2	2.8	6.9	8.5	2.6	0.1	0.88	2.7	3.3
NIR+MIR	3043	19.5	16.7	6.0	0.0	0.90	3.2	2.8	6.5	7.9	2.3	0.4	0.89	2.8	3.3
VIS+NIR+MIR	3021	18.0	16.5	6.1	-0.2	0.89	3.0	2.8	5.8	7.5	2.4	0.4	0.87	2.4	3.1

N_c, N_v: calibration and validation sample number, respectively.
SD_v, IQ_v: standard deviation and interquartile range of the validation set, respectively.
SEP: standard error of prediction.
R^2_v: determination coefficient for validation.
RPD_v: ratio of SD_v to SEP.
$RPIQ_v$: ratio of the interquartile range to SEP.

Table 3. Predictions of SOC content on the tuning and validation sets using local PLSR models built at three calibration intensities, with calibration samples selected according to the Kennard-Stone procedure (using the pre-processing that minimized the standard error of tuning, SET).

Spectral library	N_c	Tuning set (N_t = 380)							Validation set (N_v = 380)						
		N_{neigh}	SD_t	IQ_t	SET	R^2_t	RPD_t	$RPIQ_t$	SD_v	IQ_v	SEP	Bias	R^2_v	RPD_v	$RPIQ_v$
		g kg^{-1}							g kg^{-1}						
Calibration intensity: 30%															
VNIR	900	210	9.4	11.8	4.6	0.81	2.1	2.6	7.4	9.3	3.7	0.1	0.77	2.0	2.5
NIR	900	150	12.3	13.0	4.9	0.85	2.5	2.7	10.3	9.9	4.3	0.0	0.85	2.4	2.3
MIR	900	180	9.8	8.7	2.7	0.93	3.6	3.1	6.9	8.5	2.6	-0.2	0.88	2.6	3.2
NIR+MIR	900	180	11.0	9.6	3.1	0.92	3.6	3.1	6.5	7.9	2.4	0.6	0.89	2.6	3.0
VIS+NIR+MIR	900	330	11.3	10.4	2.8	0.94	3.9	3.6	5.8	7.5	2.3	0.2	0.89	2.5	3.2
Calibration intensity: 50%															
VNIR	1500	300	9.4	11.8	3.8	0.84	2.5	3.1	7.4	9.3	3.5	0.2	0.79	2.1	2.7
NIR	1500	360	12.3	13.0	4.2	0.89	3.0	3.1	10.3	9.9	3.5	-0.1	0.89	3.0	2.9
MIR	1500	240	9.8	8.7	2.2	0.95	4.4	3.8	6.9	8.5	2.2	0.0	0.90	3.2	3.9
NIR+MIR	1500	180	11.0	9.6	2.4	0.96	4.3	3.8	6.5	7.9	2.1	-0.2	0.92	3.0	3.5
VIS+NIR+MIR	1500	480	11.3	10.4	2.9	0.94	3.8	3.5	5.8	7.5	2.4	0.3	0.87	2.4	3.2
Calibration intensity: 100%															
VNIR	3048	270	9.4	11.8	3.5	0.86	2.7	3.4	7.4	9.3	3.2	0.0	0.82	2.4	2.9
NIR	3045	180	12.3	13.0	3.6	0.92	3.4	3.6	10.3	9.9	3.5	-0.3	0.89	2.9	2.8
MIR	3084	540	9.8	8.7	2.0	0.96	4.9	4.3	6.9	8.5	2.0	0.0	0.92	3.5	4.3
NIR+MIR	3043	690	11.0	9.6	2.1	0.96	5.2	4.5	6.5	7.9	1.8	0.3	0.94	3.6	4.1
VIS+NIR+MIR	3021	510	11.3	10.4	2.0	0.97	5.7	5.2	5.8	7.5	1.8	0.3	0.91	3.2	4.1

N_c, N_t, N_v: calibration, tuning and validation sample number, respectively.
N_{neigh}: number of neighbors minimizing SET.
SD_t, SD_v, IQ_t and IQ_v: standard error and interquartile range of the tuning and validation sets, respectively.
SET, SEP: standard error of tuning and of prediction, respectively.
R^2_t, R^2_v, RPD_t, RPD_v, $RPIQ_t$ and $RPIQ_v$: determination coefficient, ratio of SD to standard error, and ratio of interquartile range to standard error, for tuning and validation, respectively.

Figure 1. The RMQS sampling grid.

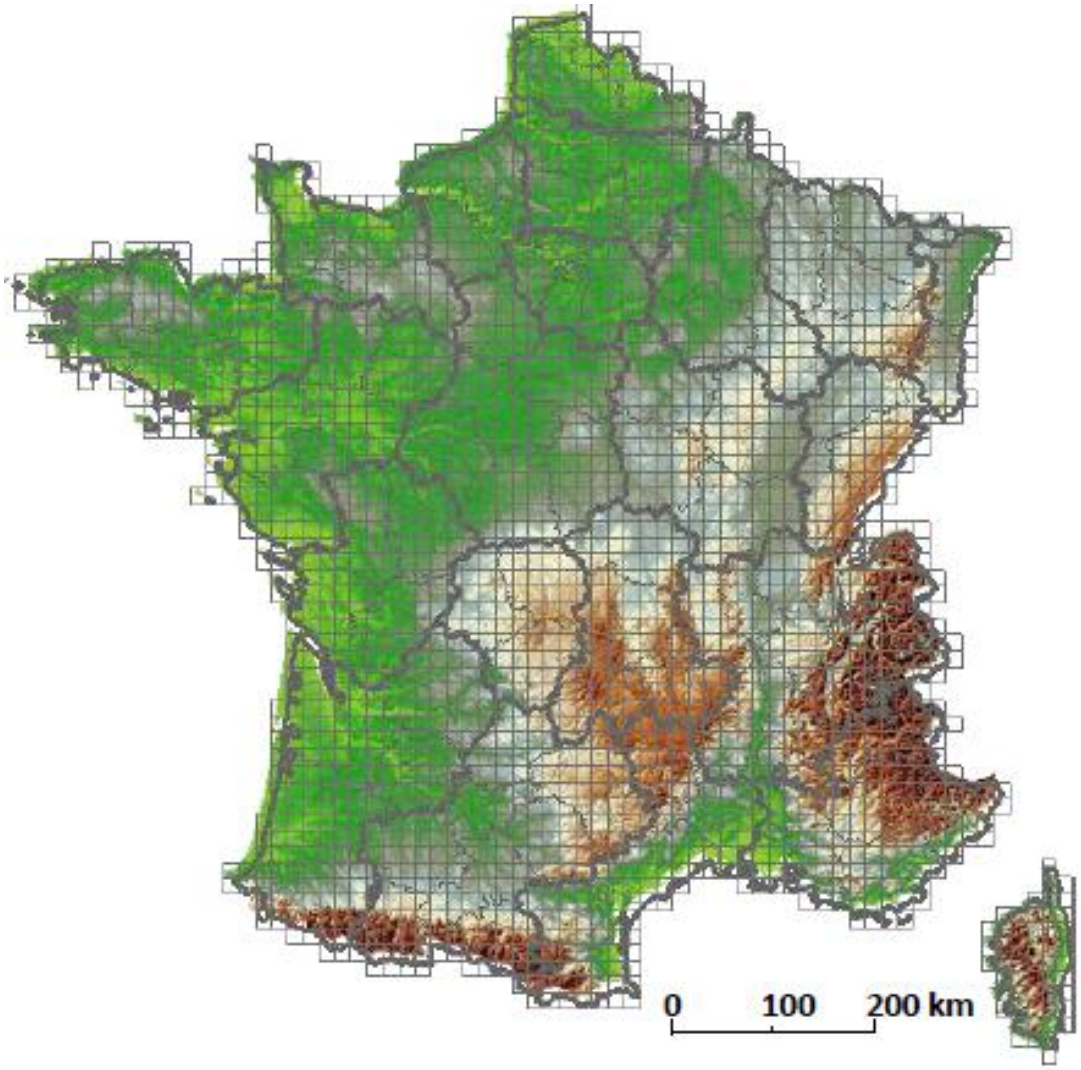

Figure 2. Descriptive statistics and frequency histogram for SOC concentration in the studied soil library (after the removal of three samples with SOC > 200 g kg^{-1}). Q1 and Q3 stand for the first and third quartile, respectively.

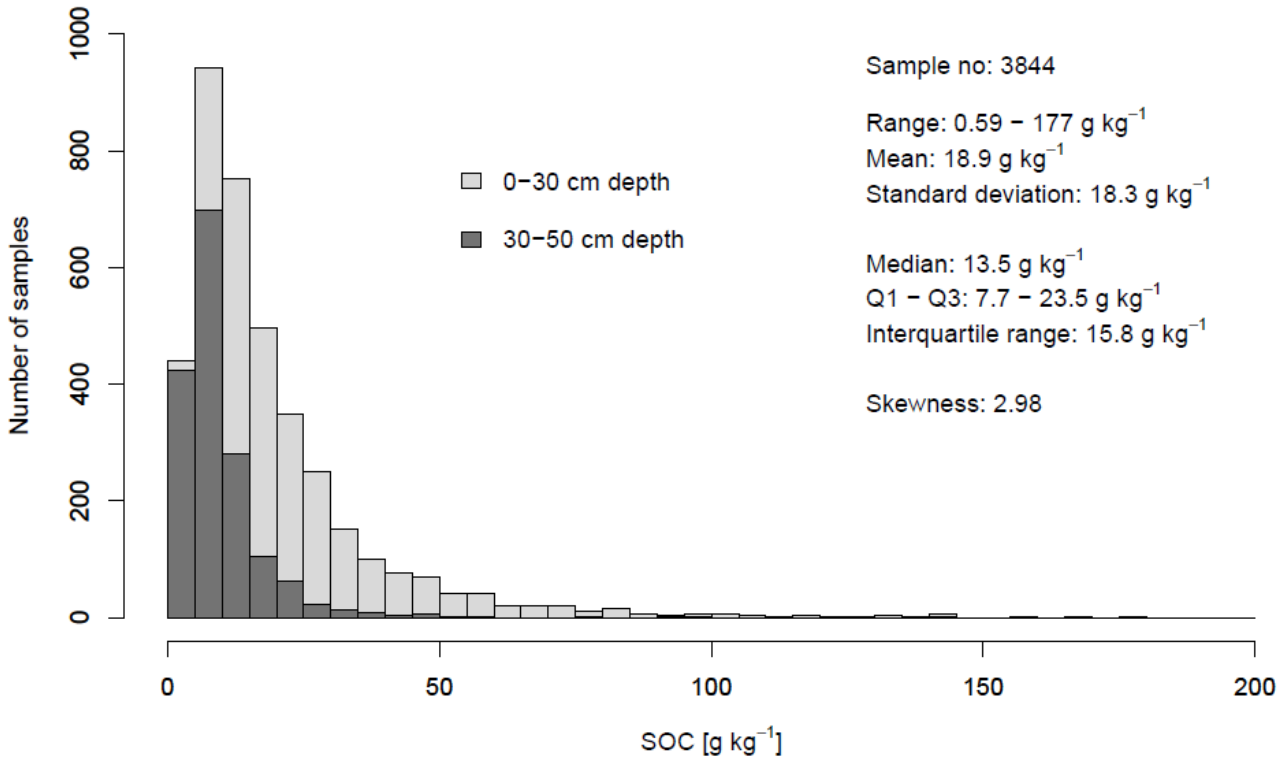

Figure 3. Schematic diagram of the sample selection procedure and calibration intensity study. N stands for the number of random selection iterations for val.set (10 in total). n stands for the number of random selection iterations for cal.set (10 in total for each N). i (in %) stands for the calibration intensity (from 10 to 100% with 10% steps).

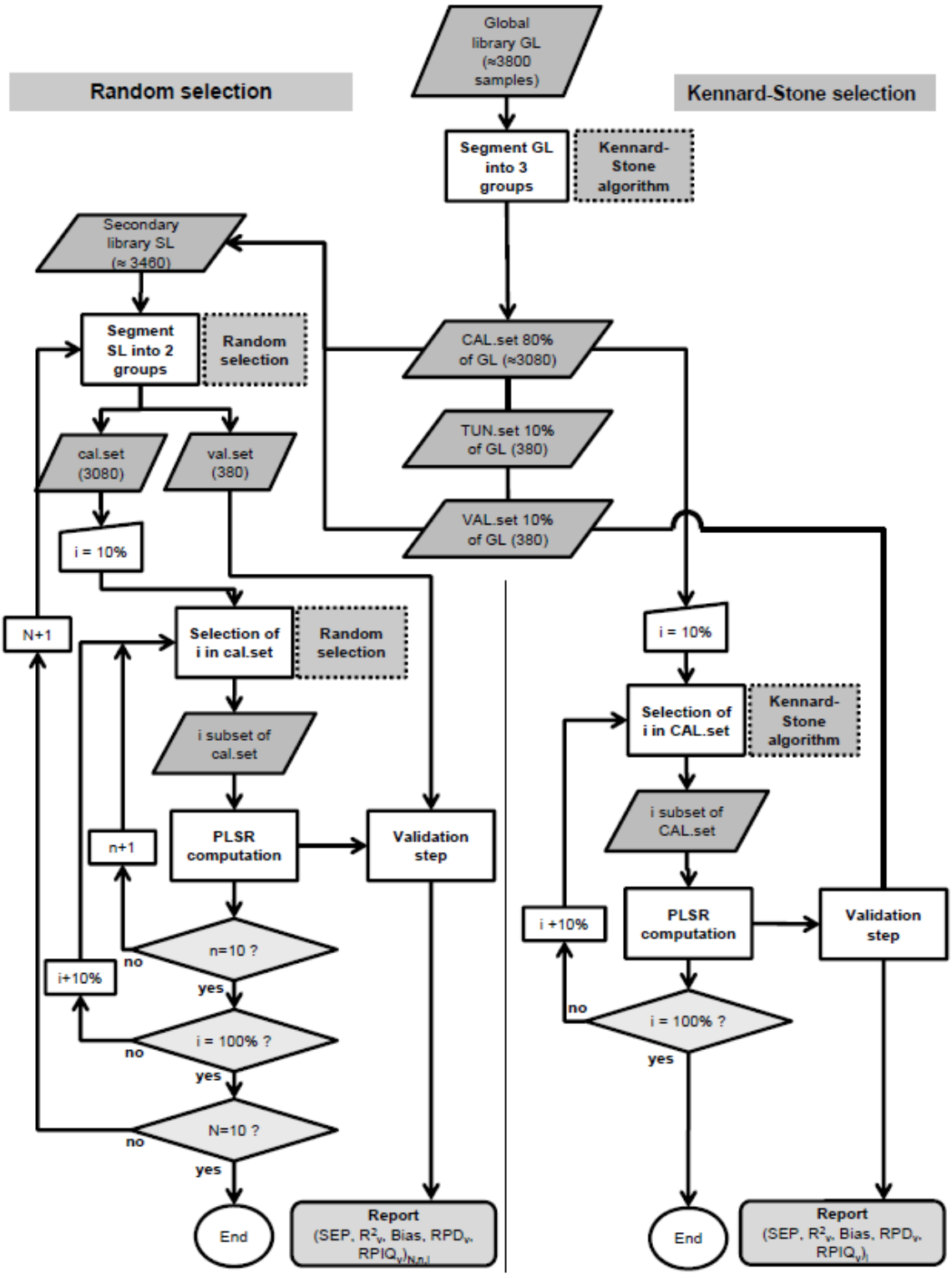

Figure 4. Performance of the SOC content predictions on the validation sets after calibration sample selection at random (left) or according to the Kennard-Stone procedure (right), using global or local PLSR models built at 10 calibration intensities (10% to 100%; local PLSR only with Kennard-Stone selection and for 30%, 50% and 100% intensities).

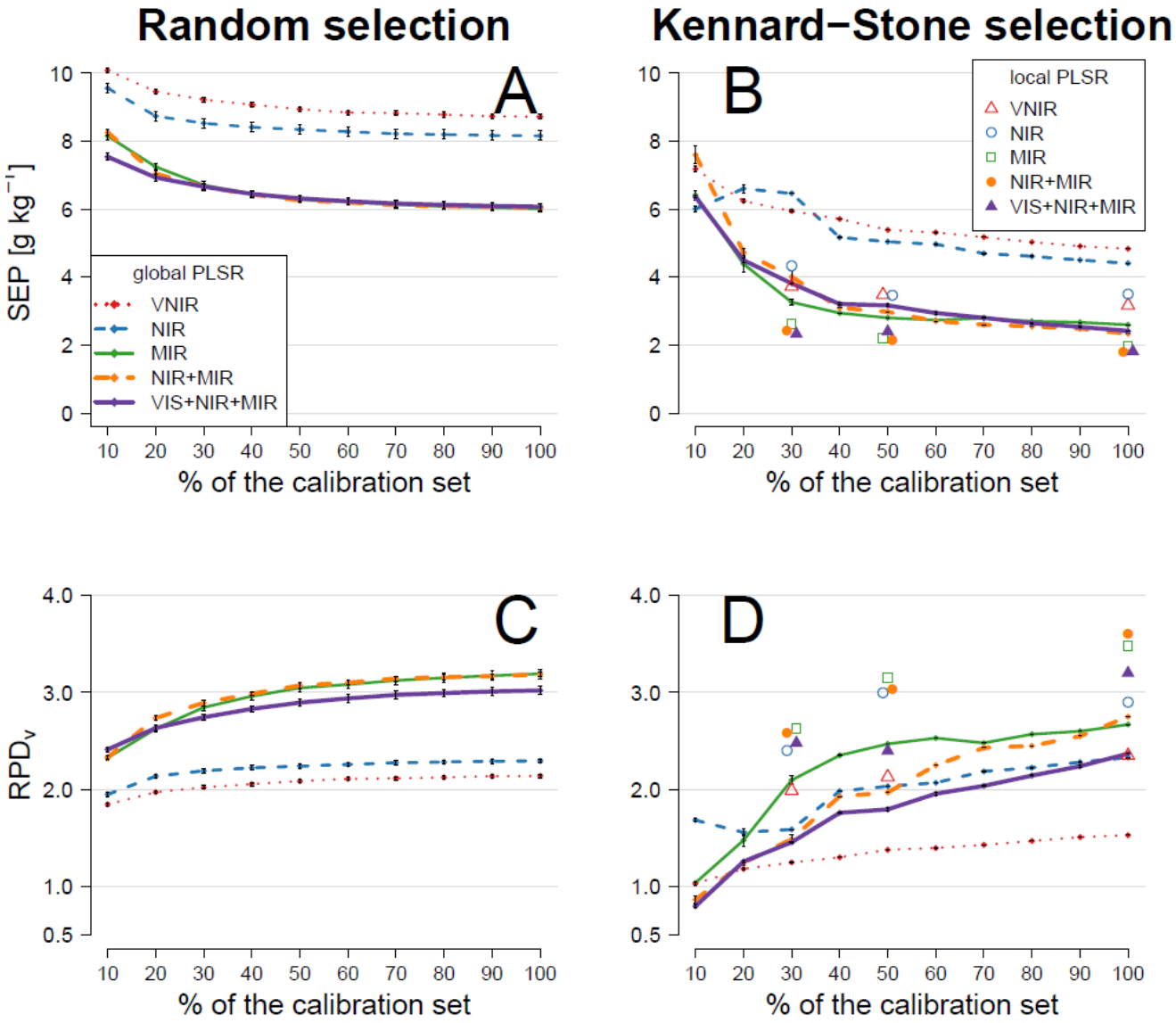

Figure 5. Comparisons between SOC contents measured (reference) and predicted on the validation set using VNIR, NIR, MIR, NIR+MR or VIS+NR+MIR spectra and global (5A) or local (5B) PLSR built with 100% of the calibration samples, selected by the Kennard-Stone procedure.

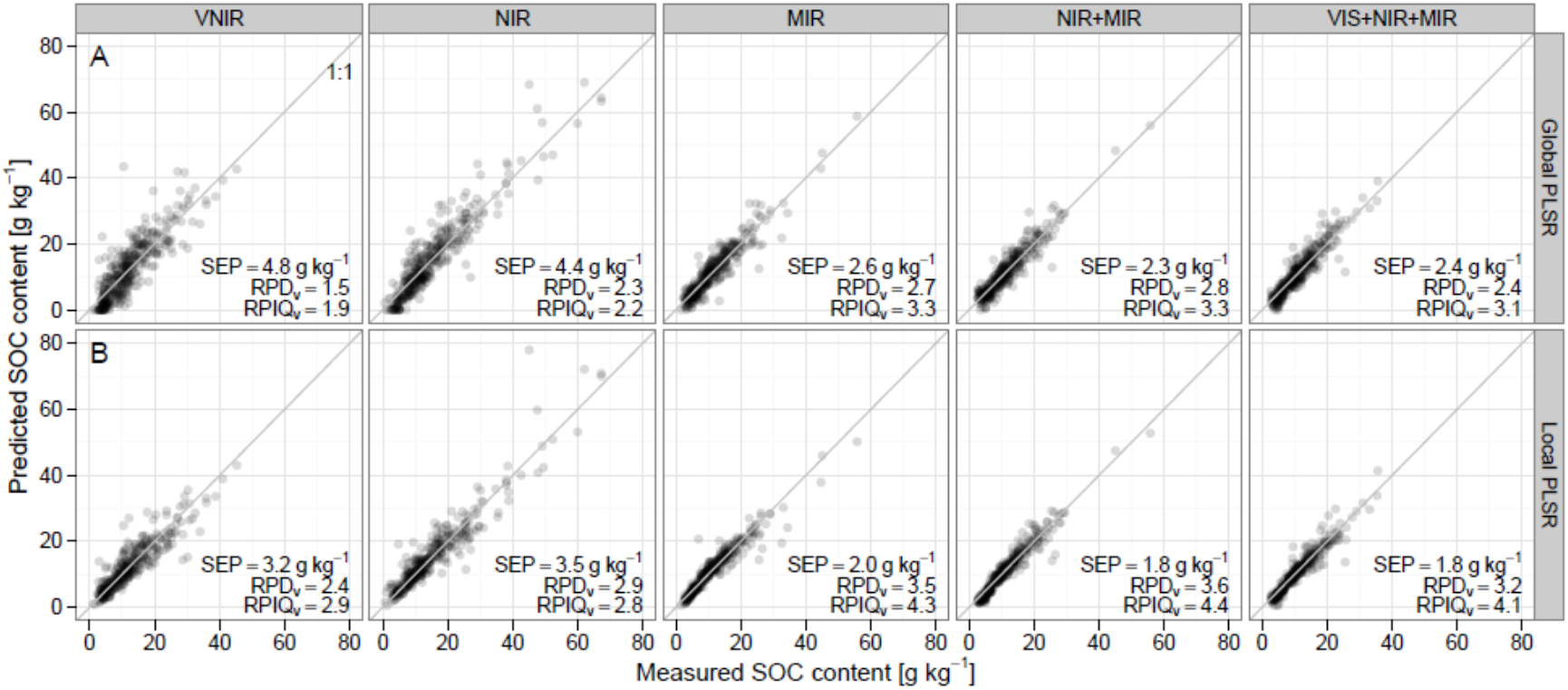

Figure 6. Regression coefficients of global PLSR using NIR, MIR or NIR+MIR spectra of 100% of the calibration set selected by the Kennard-Stone procedure (with SNV pre-processing).

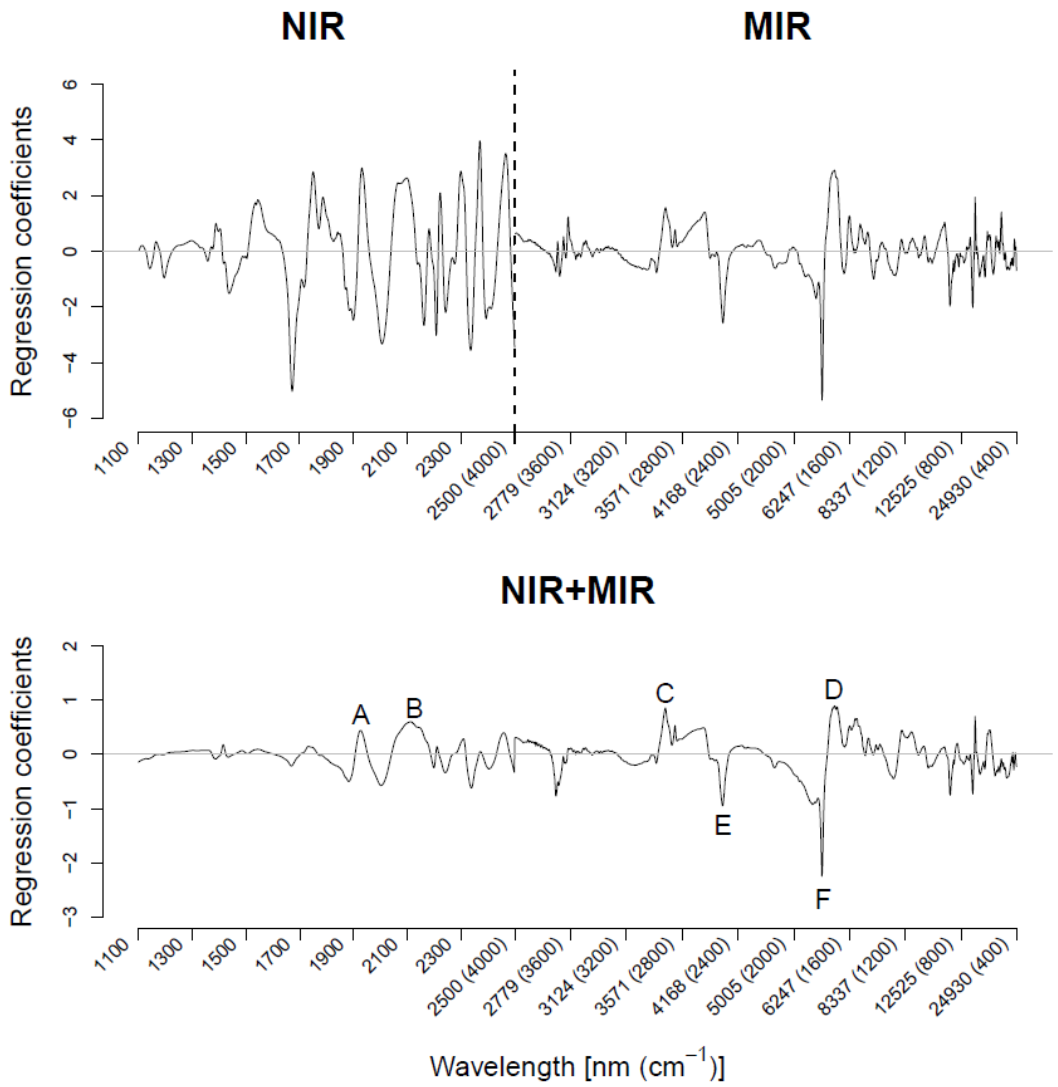

Figure 7. Maps of SOC content at 0-30 and 30-50 cm depths as measured conventionally (left) and predicted using the MIR spectra (right) and local PLSR leave-one-out cross-validation (thin soils, e.g. in mountain areas, had no 30-50 cm layer, hence blanks in the corresponding regions).