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

Bernard Barthès, Ernest Kouakoua, Patrice Coll, Michael Clairotte, Patricia Moulin, Nicolas P. A. Saby, Edith Le Cadre, Amandine Etayo, Tiphaine Chevallier

To cite this version:

Bernard Barthès, Ernest Kouakoua, Patrice Coll, Michael Clairotte, Patricia Moulin, et al.. 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, 2020, 369, pp.114272. 10.1016/j.geoderma.2020.114272 . hal-02534172

HAL Id: hal-02534172
https://hal.inrae.fr/hal-02534172v1

Version of Record: https://www.sciencedirect.com/science/article/pii/S0016706119320749
Manuscript_878032e9e7ddc39eb979b1d1755ee345

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

Bernard G. Barthès[a,*], Ernest Kouakoua[a], Patrice Coll[a,b], Michaël Clairotte[a,c], Patricia Moulin[d], Nicolas P.A. Saby[e], Edith Le Cadre[a,f], Amandine Etayo[g], Tiphaine Chevallier[a]

[a] *Eco&Sols, Université de Montpellier, CIRAD, INRA, IRD, Montpellier SupAgro, 34060 Montpellier, France*
[b] *present address: Viti-Oeno Conseil, 66000 Perpignan, France*
[c] *present address: EC JRC, Directorate Energy, Transport and Climate, Sustainable Transport Unit, 21027 Ispra, Italy*
[d] *Imago, IRD, Dakar, Senegal*
[e] *US 1106 Infosol, INRA, 45166 Olivet, France*
[f] *present address: SAS, Agrocampus Ouest, 35042 Rennes, France*
[g] *Laboratoire d'analyses des sols, INRA, 62000 Arras, France*
*Corresponding author: UMR Eco&Sols, Montpellier SupAgro, bâtiment 12, 2 place Viala, 34060 Montpellier Cedex 2, France; email: bernard.barthes@ird.fr
Other email adresses: ernest.kouakoua@free.fr, p.coll@vo-conseil.com, michael.clairotte@ec.europa.eu, patricia.moulin@ird.fr, nicolas.saby@inra.fr, edith.lecadre@agrocampus-ouest.fr, amandine.etayo@inra.fr, tiphaine.chevallier@ird.fr

Abstract

Mid-infrared reflectance spectroscopy (MIRS) is time- and cost-effective. It was used for quantifying soil inorganic carbon (SIC) concentration in France based on a national library, and performances were evaluated on an independent regional set. Our objective was to improve the accuracy of MIRS predictions based on common multivariate regression, through spiking (enrichment of the national library with some representative target samples) with possible extra-weighting (replication of spiking samples) and local calibration (only using calibration samples that are spectral neighbours of each target samples), which have not been fully explored yet, in combination especially.

Global (i.e. common) calibration yielded accurate prediction (standard error of prediction, SEP, was ≈ 5 g kg^{-1}), which could be improved when the library was completed with spiking samples (optimally 10 samples extra-weighted 40 times; SEP = 3.3 g kg^{-1}). Using spiking samples only (without the library) yielded slightly less accurate results (SEP = 3.6 g kg^{-1}). Prediction was more accurate using local calibration without spiking, but on a validation set that was reduced because some validation samples lacked calibration neighbours (SEP = 2.5-2.7 g kg^{-1}). Local calibration with spiking (optimally 10 samples without extra-weight) yielded somewhat less accurate prediction but for the full validation set when few calibration neighbours were required (SEP = 2.7 g kg^{-1}), or higher accuracy on the reduced validation set when many neighbours were required (SEP = 2.3 g kg^{-1}).

These accurate predictions demonstrated the usefulness of representative spiking and local calibration for rendering large soil spectral libraries fully operational, while extra-weighting had no additional benefit. Along with more exhaustive spectral libraries, this paves the way for extensive use of MIRS for SIC determination.

Highlights

- We used a national MIRS library for quantifying soil inorganic C in a French region
- Combining representative spiking and local calibration improved prediction markedly
- Requiring more calibration neighbours increased accuracy but on less samples
- MIRS could increasingly replace calcimetry for SIC analysis

Keywords: Diffuse reflectance spectroscopy, national spectral library, carbonates, France.

1. Introduction

In the context of increasing carbon dioxide (CO_2) content in the atmosphere, there is growing interest in the potential of soils to sequester carbon (Lal, 2004; Lal et al., 2015; Dignac et al., 2017). Comprehensive understanding of sinks and sources of soil carbon and their relationships with global change is a priority research topic, in order to produce accurate estimates of future atmospheric CO_2 concentrations. The soil carbon pool comprises two distinct components, soil organic carbon (SOC) and soil inorganic carbon (SIC), which roughly contribute two-thirds and one-third, respectively (Batjes, 2014). Attention is paid firstly to SOC, which is larger and considered more dynamic than SIC and a key component in soil functioning. Nevertheless SIC is a pool of carbon that matters and may evolve with time, land use and climate change (Emmerich, 2003; Sanderman, 2012; Chevallier et al., 2016), thus is worth quantifying. Moreover, to date, quantifying SOC in calcareous soils often requires determining SIC, SOC being consequently calculated by difference between total carbon determined by dry combustion and SIC determined by calcimetry. Direct measurement of SOC in calcareous soils is possible, through wet oxidation (Walkley and Black, 1984) or dry combustion after acid fumigation (Harris et al., 2001), but these methods use hazardous and polluting reactants, and have been regularly criticized for inaccurate results; thus alternative approaches are being looked for (Apesteguia et al., 2018).

The usefulness of infrared diffuse reflectance spectroscopy coupled with chemometrics for quantifying soil properties cost- and time-effectively has been demonstrated extensively (Viscarra Rossel et al., 2006; Cécillon et al., 2009; Soriano-Disla et al., 2014; Gredilla et al., 2016). Much attention has been paid to the prediction of SOC concentration by near infrared reflectance spectroscopy (NIRS), but better predictions have often been achieved for SIC concentration and/or using mid-infrared reflectance spectroscopy (MIRS; McCarty et al., 2002; Bellon-Maurel and McBratney, 2011; Barthès et al., 2016). Quantification by NIRS and MIRS requires calibrations, using samples that have been characterized both spectrally and conventionally (e.g. dry combustion for total carbon and calcimetry for SIC). Large soil databases that include conventionally- and spectrally-analysed samples representative at a country or even wider scale are now becoming available (Brown et al., 2006; Genot et al., 2011; Grinand et al., 2012; Viscarra Rossel and Webster, 2012; Stevens et al., 2013; Viscarra Rossel et al., 2016). Such spectral libraries pave the way for the development of many applications, as indicated by several works that used large-scale NIRS databases for making predictions on target sample sets, for instance at local scale (Brown, 2007; Sankey et al., 2008; Guerrero et al., 2010, 2014; Gogé et al., 2014). Some studies tested specific approaches to improve NIRS or MIRS predictions when using large soil spectral libraries, for instance local calibration: instead of using all library samples for building a prediction model, local calibration selects those most appropriate considering the target set, and in some cases, even builds a prediction model for each target sample separately, using calibration samples that are its spectral neighbours (Shenk et al., 1997; Ramirez-Lopez et al., 2013; Lobsey et al., 2017). Other studies tested spiking, which consists of enriching the calibration set with some samples originating from the target set (Sankey et al., 2008; Guerrero et al., 2010); these spiking samples can even be extra-weighted, to increase their influence in the calibration database (Guerrero et al., 2014). Noticeable improvements in predictions have been reported when applying local calibration (Genot et al., 2011; Rabenarivo et al., 2013; Gogé et al., 2014; Clairotte et al., 2016; Lobsey et al., 2017) or spiking, without extra-weighting (Guerrero et al., 2010; Gogé et al., 2014), and even more, with extra-weighting (Guerrero et al., 2014, 2016). But local calibration and spiking have rarely been used in conjunction (Gogé et al., 2014), and have never been used with spectrally representative spiking samples or extra-weighting.

The French soil quality monitoring network (Réseau de mesures de la qualité des sols, RMQS; Arrouays et al., 2002) represents a collection of soil samples originating from more than 2100 sites located regularly over the whole French metropolitan territory. A range of soil properties have been analysed on these samples, and their NIR and MIR spectra have been collected (Gogé et al., 2012; Grinand et al., 2012; Clairotte et al., 2016).

The objective of this work was to improve MIRS prediction of SIC concentration in France based on the RMQS spectral library, through spiking with representative target samples (possibly extra-weighted) and local calibration, which have not been fully explored yet, in combination especially (i.e. local calibration after spiking).

2. Materials and methods

2.1. National soil library

The soil samples used as calibration database belong to a large national soil library provided by the French national soil quality monitoring network (RMQS; Arrouays et al., 2002). 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. This soil library was built during a 10-year sampling campaign over the 552 000 km² of the French metropolitan territory (Corsica included), which latitude ranges from 41 to 51°N and longitude from 5.0°W to 9.5°E. The sampling design was based upon a square grid with 16-km spacing. At the centre of each square, 25 individual core samples were taken at 0–30 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). Samples were also collected at 30–50 cm, but these were not considered for this study. In total, 2178 samples were considered here, representing numerous soil types: Cambisols, Calcosols, Luvisols, Leptosols, Andosols, Albeluvisols, etc. (IUSS Working Group WRB, 2014).

2.2. Target set

The target set was made of 164 composite topsoil samples originating from commercial vineyard plots (one sample per plot) located in nine villages or small towns of the Languedoc-Roussillon region, in southern France. Vine represents the main agricultural production in Languedoc-Roussillon, which is the largest vine-growing region in France, and the studied vineyard plots were considered representative of Languedoc-Roussillon (Coll, 2011). The nine sites are presented in Table 1. Briefly, they have latitude ranging from 42.5 to 44.0°N, longitude from 2.5 to 4.0°E, elevation from 5 to 358 m a.s.l., with either Calcisols, Arenosols, Cambisols or Luvisols (IUSS Working Group WRB, 2014). Samples were taken from plots under different vineyard management systems, which varied according to the types of pesticides and fertilizers used, and possible inter-row weeding and grass cover. Samples were collected in inter-row centre at 0–15 cm depth using a hand-held gouge auger. Ten soil subsamples were taken in each plot and carefully mixed to form a composite sample. More information on the sites and samples has been provided by Coll (2011) and by Salomé et al. (2014, 2016). This set was independent from the national library.

2.3. Soil conventional analysis

All determinations of carbonate content were made in the Laboratoire d'analyses des sols from INRA (Institut national de la recherche agronomique) in Arras, France, which is ISO/CEI 17025:2005 accredited. They were carried out on finely ground (< 0.25 mm) aliquots of 2-mm sieved air-dried soils samples using a Bernard calcimeter, according to the standard French procedure NF ISO 10693 (ISO, 1995), which consists of measuring the volume of CO_2 produced after the addition of chlorhydric acid (Pansu and Gautheyrou, 2006). The carbonate content was calculated after calibration with a pure calcium carbonate ($CaCO_3$) standard and was expressed as equivalent $CaCO_3$ content. Soil inorganic carbon content was calculated as 0.12 × soil $CaCO_3$content (because the molar mass of carbon is 12 g mol^{-1} and that of $CaCO_3$ is 100 g mol^{-1}).

According to the NF ISO 10693 procedure (ISO, 1995), acceptable repeatability for this analysis is 0.4 g kg^{-1} when SIC ≤ 6 g kg^{-1}, 6% between 6 and 18 g kg^{-1}, 1.1 g kg^{-1} between 18 and 21.6 g kg^{-1}, and 5% when SIC > 21.6 g kg^{-1}. Replicated measurements allowed calculating the expanded uncertainty associated with SIC analysis in the above-mentioned laboratory. Expanded uncertainty increases the standard measurement uncertainty, calculated as the quadratic sum of measurement repeatability and intermediate precision, by multiplying it by a coverage factor (> 1) related to the confidence level requested (JCGM, 2012). Measurement repeatability was calculated based on triplicate SIC analysis of 20 samples from an inter-laboratory network; intermediate precision, based on the analysis of control samples over time; and the coverage factor was fixed at 2 for achieving a confidence level of 0.95. In such conditions, and using a modelling approach, the expanded uncertainty was empirically estimated according to Equation 1:

$$\text{Expanded uncertainty} = 0.025\ \text{SIC} + 0.13\ (\text{SIC in g kg}^{-1}) \qquad \text{Equation 1}$$

Expanded uncertainty was considered as standard error of laboratory (SEL). On the target site, where SIC averaged 19 g kg^{-1} (Table 1), SEL could thus be estimated at 0.6 g kg^{-1} [i.e. (0.025×19)+0.13].

2.4. Measurement of mid-infrared reflectance (MIR)

Air-dried, 2-mm sieved then finely ground samples were oven-dried overnight at 40°C before spectral analysis. Reflectance spectra in the mid-infrared region were acquired at 934 wavenumbers between 4000 and 400 cm^{-1} (i.e. 2500 and 25 000nm, 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). Reflectance was converted into apparent absorbance, which is the decimal logarithm of the inverse of reflectance. Twenty wavenumbers were removed due to often noisy spectrum end, and MIR spectra were used in the range from 4000 to 478 cm^{-1} (2500 and 20 909 nm, respectively).

2.5. Data analysis

2.5.1. Principles

Mid-infrared absorbance spectra and SIC measurements of the national library samples were used to build a calibration model that expressed SIC content as a function of sample spectrum. The model was then applied on validation samples in order to predict their SIC content from their MIR spectrum. Modified Partial least squares regression (PLSR) was the linear multivariate regression procedure used to infer SIC content from spectra. To date, PLSR is the most common procedure for analysing infrared spectral data, in soil science in particular (Stenberg et al., 2010). It reduces a complex spectral matrix into a few orthogonal components (or terms, or latent variables LV), which are built in order to maximise their covariance with the variable of interest (here SIC; Bjørsvik and Martens, 2001). The modification proposed by Shenk and Westerhaus (1991) consisted of scaling the conventional data and the absorbance data at each wavelength to have a standard deviation of 1.0 before each PLS term. All calculations were done using the WinISI 4 software (Foss NIRSystems/Tecator Infrasoft International, State College, PA, USA). The accuracy of the prediction models was estimated on the validation set by computing the standard error of prediction (SEP, calculated according to Equation 2), bias (mean residual), coefficient of determination (R^2_{val}) and RPD_{val} ratio (ratio of SD_{val} to SEP, where SD_{val} is the standard deviation of the validation set).

$$\mathrm{SEP} = \sqrt{\frac{\sum_1^n (y_i - \hat{y}_i)^2}{n}} \qquad \text{Equation 2}$$

where y_i and $\hat{y}_i$ are the observed and MIRS-predicted values for sample i, and n the total number of samples in the validation set.

2.5.2. The validation set

Out of the 164 samples of the target set, 30 were not used for validation but were kept for spiking (cf. 2.5.5). The validation set then included 134 samples. The 30 samples kept for spiking were the most representative spectrally, selected according to a procedure proposed by Shenk and Westerhaus (1991):

- a principal component analysis (PCA) was performed on the spectra of the 164 target samples;
- distances between samples in the PCA space were calculated using the Mahalanobis distance H (Mark and Tunnell, 1985);
- the sample that had the most neighbours closer than a given distance d was selected while its neighbours were discarded;
- the process was continued until no samples remained with neighbours closer than d;

- the distance d was set so that 30 samples could be selected.

A PCA was also performed on the spectra of the national library. The spectra of the target samples were then projected onto this PCA space to calculate their Mahalanobis distance H and evaluate how they were represented spectrally by the library (Mark and Tunnell, 1985).

2.5.3. Global calibration

Global PLSR, which is the common PLSR procedure, uses all calibration samples for building a unique model that is then applied uniformly to all validation samples. The number of PLSR latent variables that minimized the standard error of cross-validation (SECV) was retained for the prediction model, with an upper limit set to 16 (SECV was calculated using Equation 2, with n being the number of calibration samples). The cross-validation was carried out by dividing the calibration set into four groups composed cyclically (i.e. the 1st, 5th, 9th samples in the first group, the 2nd, 6th, 10th samples in the second group, etc.), after the samples had been ranked by increasing SIC values. Three groups were used to develop the model and one to test it, and the procedure was performed four times to use all samples for both model development and prediction. The residuals of the four predictions were pooled to calculate SECV.

2.5.4. Local calibration

In contrast, local PLSR makes prediction for each validation sample individually, only using calibration samples that are its spectral neighbours (Shenk et al., 1997). These neighbours were selected according to the correlation coefficient R between the spectra of calibration samples and each validation sample, with an R cut-off value below which samples were not considered neighbours. According to preliminary tests, this R cut-off value was set at 0.95.

The minimum number of calibration neighbours requested was varied from 4 to 50. It is important to note that prediction was not made for validation samples that did not have enough calibration neighbours. There was firstly no requirement on the maximum number of calibration neighbours; then this maximum was set to 300 and 100, but this had limited interest and the corresponding results will not be presented.

Actually local calibration differs from global calibration on two points: (i) prediction is made on each validation sample individually, only using calibration neighbours, not all calibration samples; and (ii) prediction is not made for validation samples that do not have enough calibration neighbours (i.e. that are poorly represented by the calibration set). To separate these both aspects, prediction was also made using all calibration samples (i.e. global calibration), but only on the validation samples that had enough calibration neighbours. In that case, predictions by local and global calibration were made on the same validation samples, and the difference between local and global calibrations was only due to calibration by neighbours.

In local calibration the number of latent variables was not determined through cross-validation. Instead, each prediction was calculated as the weighted average of the predicted values generated with 3 to 16 latent variables, each weight being calculated as the inverse of the product of the root mean square (RMS; cf. Equation 3) of spectral 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 using the considered number of latent variables (Shenk et al., 1997).

$$\mathrm{RMS} = \sqrt{\frac{\sum_1^n x_i^2}{n}} \qquad \text{Equation 3}$$

where x_i are the values considered (i.e. spectral residuals or regression coefficients) and n their number.

2.5.5. Spiking

Spiking consists of adding a few target samples to the calibration set, so that the prediction model better suits the validation set (Guerrero et al., 2010). This was achieved using the 30 representative target samples removed from the validation set (cf. 2.5.2), which will thereafter be called "spiking samples". Spiking was performed for both global and local calibrations and different sizes were tested for the spiking subset: 7, 10, 15, 20 and 30 samples. When 7 to 20 samples were used for spiking, they were selected as the most spectrally representative in the set of 30 samples removed from the target set (in which they were the most spectrally representative, cf. 2.5.2).

Moreover, the spiking subset could be extra-weighted, which consisted of adding several copies of this subset to the calibration set, in order to increase its influence in the calibration (Guerrero et al., 2014). Different weights were tested, from 10 to 200, which means that each spiking sample was represented by 10 to 200 replicates in the calibration set, respectively. Of course spiking uses (some) target samples for calibration, but the improvement in model performance really seems worthwhile (Guerrero et al., 2010, 2014, 2016). Spiking without or with extra-weighting was performed for both global and local calibration.

2.5.6. Pre-processing

In addition, different common spectrum pre-processing methods, or pretreatments, were tested: none (no scatter correction); standard normal variate transformation (SNV), which consists of mean-centring and variance-scaling the spectrum; detrending (D), which consists of removing a linear trend from the spectrum; SNVD (i.e. both SNV and D); and multiplicative scatter correction (MSC), which consists of centering and scaling the spectrum with, respectively, the intercept and slope calculated when regressing linearly this spectrum against the average calibration spectrum (Geladi et al., 1985; Barnes et al., 1989). These transformations were possibly followed by first derivation. Derivation aims at reducing baseline variation and enhancing spectral features, and was calculated over a 5-, 15- or 20-point gap, with 5-point smoothing, in order to reduce signal random noise (Bertrand, 2000). The conditions of derivation were denoted 155, 1155 and 1205 for first derivation with 5-, 15- and 20-point gap, respectively, and 5-point smoothing. No derivation and no smoothing was denoted 001.

3. Results

3.1. Distributions of SIC content

In the national library and target set, minimum SIC content was ≈ 0 g kg^{-1} (under the detection limit), maximum was 104 and 86 g kg^{-1}, mean 6.4 and 18.8 g kg^{-1}, median ≈ 0 and 13.3 g kg^{-1}, and standard deviation 16.0 and 21.4 g kg^{-1}, respectively. The national library was dominated by samples with no SIC (53%) or very low SIC content (22% with $0 < \text{SIC} < 1$ g kg^{-1}), while soils with SIC > 20 and > 50 g kg^{-1} accounted for 12% and 4% of the library, respectively (Fig. 1). Indeed, the French territory has large mountain regions with acid rocks and soils (Arrouays et al., 2011). In the target set, samples with no SIC accounted for a noticeable proportion of the set (31%), but SIC-rich soils were more frequent: samples with SIC > 20 and > 50 g kg^{-1} accounted for 43% and 11% of the set, respectively (Fig. 1). Carbonated rocks and soils cover an important proportion of the region considered, in southern France.

3.2. Prediction of SIC content using global calibration without spiking

Global calibration without spiking represents usual calibration, built with the whole national library. In these conditions, the best MIRS predictions were achieved using the pretreatment None1155 (i.e. first derivation with 15-point gap and 5-point smoothing) and yielded SEP = 5.2 g kg^{-1} and $\text{RPD}_{\text{val}} = 3.7$ (Table 2 and Fig. 2a).

Predictions were particularly poor for the samples that originated from Aigues-Mortes. When projected onto the PCA space of the national library, most Aigues-Mortes samples were not close to national library samples (Fig. 3), though only two were spectral outliers formally (i.e. their Mahalanobis distance H with the national library, i.e. its centre, was > 3). Salomé et al. (2014), who studied the same sample set, also observed that the soils from Aigues-Mortes were particular, being quite rich in SIC though very sandy.

3.3. Prediction of SIC content using global calibration with spiking

3.3.1. Global calibration using spiking samples only (and not the national library)

Firstly calibrations were built using only spiking samples, to address their importance, and the national library was not used. This led to surprisingly accurate predictions on the 134 validation samples (SEP ranged from 4.5 to 3.6 g kg^{-1} and RPD_{val} from 4.3 to 5.3; Table 2 for 10 and 30 spiking-calibration samples, and Fig. 4). Actually global calibration with 10 representative samples from the target site led to more accurate predictions than global calibration using the national library (RPD_{val} = 5.3 vs. 3.7, respectively).

3.3.2. Global calibration using the national library; effect of the number of spiking samples

Then calibrations were built using the national library enriched with representative spiking samples. Adding one copy of the spiking samples had little effect on prediction, due to the large size of the library (SEP = 4.9 g kg^{-1} and RPD_{val} = 3.9 with 10 to 30 spiking samples; Table 2 and Fig. 4), as also observed by Guerrero et al. (2014). So the spiking samples were extra-weighted, meaning that each was replicated 10 to 200 times (Guerrero et al., 2014).

As could be expected, prediction accuracy tended to increase with the number of spiking samples; but rather unexpectedly, the additional benefit was limited beyond 10 spiking samples (Fig. 4). Predictions using 7 spiking samples were rather poor (SEP ≈ 5 g kg^{-1} and RPD_{val} < 4 with extra-weight ≥ 10), probably because 7 spiking samples were not sufficient to represent the validation set correctly.

3.3.3. Global calibration using the national library; effect of the weight of spiking samples

As could also be expected, prediction accuracy increased with the weight of spiking samples, but the additional benefit was limited beyond 40 replicates (Fig. 4): with 10 spiking samples replicated 40 times vs. 30 spiking samples replicated 200 times, SEP was 3.3 vs. 3.0 g kg^{-1} and RPD_{val} 5.9 vs. 6.4, respectively (Table 2). Thus prediction accuracy was little improved when multiplying the number of spiking samples by three and their weight by five, which required three times more conventional analyses, noticeably more computational time, thus did not seem relevant. Computational time was not measured. However 10 spiking samples replicated 40 times vs. 30 spiking samples replicated 200 times represented additions of 400 vs. 6000 samples to a library already including 2178 samples, resulting in calibration datasets of 2578 vs. 8178 samples, respectively.

It seemed that SEP continued to decrease slightly when the weight of spiking samples increased beyond 200, but this was not explored because it required much computational time. It would be expected that, with ever increasing weight of spiking samples, prediction performance would approach that achieved when using spiking samples only (SEP = 3.6-4.0 g kg^{-1} for ≥ 10 spiking samples; Fig. 4).

3.4. Prediction of SIC content using local calibration without spiking

3.4.1. Number of calibration neighbours

Local calibration makes prediction on each validation sample separately, only using calibration samples that are its spectral neighbours according to correlation between spectra, with an R cut-off value that was set to 0.95; and there is no prediction for validation samples that do not have enough calibration neighbours. Fig. 5a presents the distribution of validation samples according to their number of spectral neighbours in the national library; in particular, 1 validation sample had no calibration neighbour, 5 samples had 4 to 9 calibration neighbours, and all 6 samples with < 10 calibration neighbours originated from Aigues-Mortes (cf. 3.2). Most validation samples had more than 100 calibration neighbours, except those from Aigues-Mortes, which had 22 in average (data not shown).

3.4.2. Local calibration without spiking; effect of the number of calibration neighbours required

When the minimum number of calibration neighbours was set to 4, local calibration had limited benefit when compared with global calibration (SEP = 4.4-4.8 g kg^{-1} and RPD_{val} = 4.1-4.4 on 133 validation samples; Table 3). When the minimum number of required neighbours was increased from 4 to 10, 20, 30 then 50, progressively the number of validation samples that had enough neighbours decreased to 115 samples, while prediction accuracy increased (SEP decreased to 2.7 g kg^{-1} and RPD_{val} increased to 7.6; Table 3 for ≥ 50 neighbours, other data not shown).

3.4.3. Reasons for differences between local and global calibration without spiking

Actually local calibration did not yield much better prediction than global calibration on the same validation samples ($RPD_{val} = 4.1$ vs. 4.0 on 133 samples with ≥ 4 neighbours required and 7.6 vs. 6.4 on 115 samples with ≥ 50 neighbours, respectively, i.e. an increase ≤ 1.2; Table 3). By contrast, global calibration on these validation samples that had enough calibration neighbours yielded noticeably more accurate prediction in general than global calibration on the whole validation set ($RPD_{val} = 6.4$ vs. 3.7 on 115 vs. 134 samples, respectively, i.e. an increase of 2.6 when ≥ 50 neighbours were required), except when few calibration neighbours were required ($RPD_{val} = 4.0$ vs. 3.7 on 133 vs. 134 samples, respectively, with ≥ 4 neighbours; Tables 3 and 2). These comparisons show that without spiking, most improvement in prediction accuracy from global to local calibration resulted in general from removing validation samples with too few calibration neighbours; while the benefit of calibration by neighbours was limited. This was however not the case when few calibration neighbours were required for local calibration: in that case, improvement from global to local calibration resulted mainly from calibration by neighbours, because almost all validation samples had enough neighbours thus were predicted. But this improvement from global to local calibration, due to calibration by neighbours, was limited.

3.5. Prediction of SIC content using local calibration with spiking

Local calibration was carried out using the national library enriched with 7 to 30 spiking samples originating from the target sites, and extra-weighted 10 to 50 times, or not extra-weighted. When extra-weighted, a spiking sample neighbouring a validation sample was represented by several copies, among the calibration samples used for making a prediction on this validation sample.

3.5.1. Local calibration with spiking; effect of the number of spiking samples

Local calibration with 7 spiking samples had little effect on prediction results when compared with local calibration without spiking (e.g. with ≥ 50 calibration neighbours required, $RPD_{val} = 7.6$ without spiking or with 7 spiking samples extra-weighted 10 times, on 133 samples; Table 3 without spiking, data not shown for 7 spiking samples).

Using 10 spiking samples was more useful, but the effects varied, depending firstly on the number of calibration neighbours required (Table 3):

- with few neighbours required (i.e. 4), prediction was much better with than without spiking, on all validation samples ($RPD_{val} = 7.2\text{-}7.3$ on 134 samples vs. 4.1 on

133 samples, respectively; Fig. 2b for extra-weight 1, i.e. no extra-weighting), because spiking increased the number of calibration neighbours thus improved prediction accuracy;

- with many neighbours required (i.e. 50), the effect was less clear and depended on the extra-weight of spiking samples:
  - with low extra-weight (×1 or ×10), prediction was moderately better with than without spiking, on the same (reduced) validation set ($RPD_{val} = 8.9\text{-}9.1$ vs. 7.6 on 115 samples, respectively; Fig. 2c for extra-weight 10), because the small number of replicates from spiking samples did not provide enough calibration neighbours;
  - with high extra-weight (×50), prediction was moderately less accurate with than without spiking, but on all validation samples ($RPD_{val} = 6.8$ on 134 samples vs. 7.6 on 115 samples, respectively; Table 3), because the large number of replicates from spiking samples provided enough calibration neighbours.

Using more than 10 spiking samples did not improve prediction thus had no interest (with either 10 or 30 spiking samples ×1 or ×10, $RPD_{val} = 7.2\text{-}7.3$ on 134 samples with ≥ 4 neighbours required; and $RPD_{val} = 8.8\text{-}9.2$ on 115-116 samples with ≥ 50 neighbours; Table 3). Surprisingly, the number of spiking samples had little effect on the number of validation samples that had enough calibration neighbours (using either 10 or 30 spiking samples, this number was 134 with ≥ 4 neighbours requested, and 115 or 116 with ≥ 50 neighbours; Table 3). Actually, the distribution of validation samples according to their number of calibration neighbours was not much affected by the number of spiking samples (Fig. 5b vs. 5e, 5c vs. 5f, and 5d vs. 5g).

3.5.2. Local calibration with spiking; effect of the weight of spiking samples

Extra-weighting the spiking samples had little effect on prediction results, with either few calibration neighbours required (i.e. 4; with 10 spiking samples ×1 to ×50, $RPD_{val} = 7.2\text{-}7.3$ on 134 samples) or many neighbours required (i.e. 50; with 10 spiking samples ×1 or ×10, $RPD_{val} = 8.8\text{-}9.2$ on 115-116 samples; Table 3). Thus extra-weighting had limited interest in local calibration, which contrasted with global calibration. High extra-weight of spiking samples allowed making prediction on all 134 validation samples when many neighbours were required (i.e. 50; for 10 spiking samples ×50, $RPD_{val} = 6.8\text{-}7.2$); but similar result was achieved when few neighbours were required (for 10 spiking samples ×1, $RPD_{val} = 7.3$; Table 3).

3.5.3. Local calibration with spiking; effect of the number of calibration neighbours required

As observed for local calibration without spiking (cf. 3.4.2), increasing the minimum number of calibration neighbours improved prediction accuracy on a validation set that was reduced (with 10 spiking samples×1 and ≥ 4 vs. ≥ 50 neighbours required, SEP = 2.7 vs. 2.3 g kg^{-1} and RPD_{val} = 7.3 vs. 8.9 on 134 vs. 115 samples, respectively); but this effect vanished at high extra-weight (with 10 spiking samples ×50 and ≥ 4 vs. ≥ 50 neighbours required, SEP = 2.7 vs. 2.7 g kg^{-1} and RPD_{val} = 7.2 vs. 6.8 on 134 samples, respectively; Table 3).

3.5.4. Reasons for differences between local and global calibration with spiking

To compare local and global calibration with spiking, it was useful to consider whether prediction was possible for all validation samples or not, as was also the case without spiking. Prediction was possible for all validation samples when all had enough calibration neighbours, which was achieved (i) when few calibration neighbours were required, or (ii) when many were required but spiking samples were highly extra-weighted. In that case where prediction was possible for all validation samples, the difference with global calibration was not due to the removal of validation samples with too few calibration neighbours. The difference was only due to calibration by neighbours, and the benefit of local calibration was noticeable, but it decreased when extra-weighting increased (with 10 spiking samples ×1, ×10 and ×50, local calibration with ≥ 4 neighbours vs. global calibration yielded RPD_{val} = 7.3 vs. 3.9, 7.2 vs. 5.0 and 7.2 vs. 5.9 on 134 samples, respectively; Table 3). Indeed, extra-weighting the spiking samples improved prediction using global calibration but not using local calibration. By contrast, when prediction was not possible for some validation samples with too few calibration neighbours, which in general were poorly predicted otherwise, the difference between local and global calibration was larger and had two causes: the removal of poorly predicted samples from the validation set, which was the dominant cause in general, and calibration by neighbours (with 10 spiking samples ×1 and ×10, local calibration with ≥ 50 neighbours vs. global calibration on the same validation samples vs. global calibration on the complete validation set yielded RPD_{val} = 8.9-9.1 vs. 7.5-7.8 vs. 3.9-5.0 on 115 vs. 115 vs. 134 samples, respectively; Tables 3 and 2, and Fig. 2c and 2d for local and global calibration with same calibration and validation samples).

3.6. Complementary considerations

The absolute value of bias was large (< -1 g kg^{-1}) with global calibration and < 10 spiking samples, or ≥ 10 spiking samples extra-weighted < 20 times. With ≥ 10 spiking samples or, in local calibration, with ≥ 50 neighbours required, prediction improvement was mainly an effect of bias reduction. Guerrero et al. (2014) also attributed to large bias the poor prediction of SOC content of target samples when using a large unspiked NIRS library.

Mean SIC was 17.8 g kg^{-1} on the validation set, thus SEL could be estimated at 0.6 g kg^{-1} according to Equation 1 (with SEL = expanded uncertainty), and its contribution to SEP was limited.

4. Discussion

4.1. Global calibration without spiking

Studying almost the same national library than in the present work, also with MIRS, Grinand et al. (2012) obtained SEP = 2.9 g kg^{-1} and $RPD_{val} = 5.6$ in average when calibrating randomly (with five replicates) on 20% of the library and validating on 80%, and SEP = 2.4 g kg^{-1} and $RPD_{val} = 6.8$ when calibrating on 80% and validating on 20%. This seems better than in the present study, but actually, mean SIC was smaller in the national library than in our target set, and SEP represented 38% to 45% of the mean vs. 29% here. McCarty et al. (2002) made global MIRS calibrations on profile samples from 14 locations in nine US states, with RPD_{val} that could be estimated at 8.0 in external random validation (i.e. possibly non-independent) and at 2.4 in independent validation (one-site-out). This pointed out that, even for soil attributes linked to molecular functional groups that absorb in the MIR range, such as SIC content, global calibration without spiking may lead to poor independent validation when validation samples are poorly represented by calibration samples, as noticed by Guerrero et al. (2014) from their study and several other works. Similar performances ($RPD_{val} = 2.4$) were achieved in independent validation after global visible-and-NIRS (VNIRS) calibration by Brown et al. (2006) on a very large and diverse set of > 4100 independent samples from four continents, and by Morgan et al. (2009) on a set of core samples from six central Texas fields. Brown et al. (2006) used boosted regression trees, which outperformed PLS regression, and achieved $RPD_{val} = 2.7$ when using sand content as auxiliary predictor. Comstock et al. (2019) used a national US library made of 1268 MIR spectra and PLS global calibration for predicting SIC on 209 independent samples in two US states, and achieved SEP = 3.1 g kg^{-1} and $RPD_{val} = 6.4$. This suggests that the target samples were better represented by the national library than in the present study, though this aspect was not specified, from spectral viewpoint especially. These references and our work underline that prediction accuracy is greatly affected by sample set diversity, especially by the distribution of the variable of interest and by the ability of the calibration set to represent the validation samples, which vary across studies thus render comparisons difficult.

4.2. Global calibration with spiking

The results achieved when using spiking samples only (and not the national library) for calibration were not very different from those reported by Grinand et al. (2012). These authors used MIRS for predicting SIC within the same library than in the present study: random calibration on 10% of the library yielded $RPD_{val} = 5.0$ on the remaining samples, close to $RPD_{val} = 5.3$ achieved here when using 10 representative target samples to make prediction on 134 (i.e. 7%).

Moreover, several literature papers have reported NIRS or VNIRS predictions at local scale through global calibration on a large soil library completed with spiking samples. Sankey et al. (2008) performed global VNIRS calibration at local scale in north-western USA and reported better SIC predictions when enriching a national soil library with spiking samples, but predictions were not necessarily better than when using these spiking samples only (i.e. without the national library). These authors used more spiking samples than in the present study (50% of the target set), which could explained the limited benefit of the national library; but we also achieved comparably accurate predictions when using local samples only than when using them for spiking the national library (Table 2). Overall, Sankey et al. (2008) achieved poor predictions ($RPD < 1.6$, $SEP > 80\%$ of the mean), probably due to less informative spectral range (VNIR vs. MIR) and to random selection of spiking samples, which did not optimize their representativeness. Guerrero et al. (2010) also observed that global NIRS-PLSR calibration of soil nitrogen (N) in south-eastern Spain using only a few samples from the target site could yield similarly and even more accurate validations than large regional libraries spiked with these samples; thus additional information provided by regional libraries was not particularly useful. The spiking samples, which were not replicated, had more influence in general when the library size decreased (different sizes being considered). This led to the idea of extra-weighting the spiking samples, the interest of which was demonstrated next by Guerrero et al. (2014). These authors used a national NIRS library from Spain to build global PLS calibrations of SOC content for target sites in Spain, UK and Sweden. They got better predictions when the spiking subset was selected according to spectral representativeness, which was our choice. Moreover, they observed for large libraries that increasing the weight of spiking samples was more useful than increasing their number.

They did not fully explore the effect of varying extra-weighting but tested a weight determined as the ratio of library size to target set size, which would yield 13 in the present study. We also observed that extra-weighting of spiking samples was more fruitful than spiking our national library with more than 10 spiking samples, because these represented the target set correctly. Guerrero et al. (2014) reported accurate predictions when using spiking samples only, but more accurate predictions when using libraries enriched with extra-weighted spiking samples, which was also our observation. In contrast, predicting SIC in a 24-km² area using the same national library than in the present study but in the VNIR range, Gogé et al. (2014) found that spiking the national library had limited benefit. They attributed this result to the strong spectral features of carbonates, which would render the national library sufficient, while spiking improved prediction accuracy for soil properties supposedly less strongly featured (SOC, clay, etc.). In this work, the spiking samples were selected at random, thus not particularly representative of the target site, moreover they were not extra-weighted, and we can assume this also limited the benefit of spiking.

4.3. Local calibration without spiking

We observed that some validation samples had few calibration neighbours, thus were poorly represented by the national library, though most of them were not spectral outliers (according to the Mahalanobis distance H calculated between each validation sample and the library centre, cf. 3.2). Brown et al. (2005) developed and applied global VNIRS calibrations on six sites of north-western USA having similar features, and similarly observed, in one-site-out cross-validations, that SIC predictions might be poor in some individual sites though they were well represented spectrally by the other sites. This underlines that target samples may be well represented globally by a calibration set according to usual tests (e.g. H distance between target samples and the calibration set, i.e. its centre), but nevertheless have few spectral neighbours in this set, which often leads to poor predictions.

In their above-cited paper that dealt with VNIRS prediction of SIC in a 24-km² area, based on the same national library than in the present study, Gogé et al. (2014) compared global calibration with locally weighted calibration involving 300 neighbours (in this local calibration approach, neighbours' contribution to prediction increases with their spectral similarity to the target sample). They found little difference in prediction accuracy between both approaches. This again was attributed to the strong spectral features of carbonates, which would be correctly represented by the whole library (i.e. would render specific representation by spectral neighbours useless), while local calibration was more useful for predicting soil properties supposedly less strongly featured (e.g. SOC or clay). Moreover, the authors underlined that the optimal number of calibration neighbours, which was fixed according to preliminary work, depended on the soil property considered and on the library.

Local calibration involves two aspects: not predicting validation samples with too few calibration neighbours, on the one hand, and calibration by neighbours, on the other hand. In the present study, calibration by neighbours had limited benefit without spiking: when local calibration outperformed global calibration clearly, this was mainly due to not predicting validation samples with too few calibration neighbours; and this occurred when many calibration neighbours were required. This suggests that strongly improving prediction accuracy at the level of the validation set (i.e. SEP, RPD_{val}, etc.) was only possible when removing and making no prediction on validation samples poorly represented by the calibration database. It seems relevant to not use a spectral library for making prediction on samples it poorly represents; but having to make prediction on less samples to improve overall prediction accuracy may be somewhat disturbing. Actually, making no prediction on target samples poorly represented by the calibration set is not specific to local calibration, and may occur in global calibration too: a cut-off value of a distance parameter between each target sample and the calibration set can also be used for removing target samples that are considered outliers (e.g. Castaldi et al., 2018, who used the H distance). We must admit the incomplete ability of MIRS to quantify SIC and other soil properties on certain samples as long as soil spectral libraries are incomplete. It is however predictable that the representativeness of soil spectral libraries will become more and more complete with time.

4.4. Local calibration with spiking

Few authors have performed local calibration with spiking. Gogé et al. (2014) performed locally weighted calibration with spiking for VNIRS prediction of SIC in a 24-km² target area using the French national library. This approach had little benefit when compared to locally weighted calibration without spiking and to global calibration with and even without spiking, which again was attributed to the strong spectral features of carbonates. Limited effect of spiking might also be attributed to random selection of spiking samples. Seidel et al. (2019) used VNIRS and PLS for predicting SOC in two fields in Germany. Spectral neighbours of target samples (50 per sample) were selected from a national library then all used for building global calibration. Selecting spectral neighbours in the national library reminds of local calibration, though prediction was not made for each target sample individually. However, there was no threshold below which a calibration sample was not considered as neighbour;

and as target samples were poorly represented by the national library (as specified by PCA), predictions were poor without spiking ($\mathrm{RPD_{val}} \leq 1.3$ vs. ≤ 1.2 with the full national library). This underlines the interest of such threshold (e.g. cut-off R), which ensures that samples used for calibration are really neighbours, and that if neighbours are lacking, there will be no prediction for the target samples considered, thus no risk of poor prediction. Then Seidel et al. (2019) observed that spiking improved predictions markedly ($\mathrm{RPD_{val}} \geq 2.5$, ≥ 2.9 and ≥ 3.3 with 10, 15 and 20 spiking samples, respectively).

In the present study, prediction using local calibration became possible for all validation samples through spiking, when either few calibration neighbours were required or when spiking samples were heavily extra-weighted. Prediction accuracy was similar in both cases, indicating that extra-weighting had little interest in local calibration, which contrasted with global calibration. When compared with global calibration with spiking, the benefit of local calibration with spiking resulted often from not predicting validation samples poorly represented by the calibration database, as already observed without spiking. But in contrast to what was observed without spiking, noticeable benefit could also result from calibration by neighbours, when the spiking samples were moderately extra-weighted (i.e. global calibration was little improved by spiking) and few calibration neighbours were required (i.e. prediction using local calibration was possible for all validation samples).

The optimisation of parameters proposed here could not be generalised, as it clearly depended on the size, spectral diversity and spectral proximity of the calibration database and validation set. Values proposed for number and extra-weight of spiking samples and number of calibration neighbours could be seen as orders of magnitude.

5. Conclusion

Global calibration yielded accurate prediction ($\mathrm{RPD_{val}} \approx 4$), which could be noticeably improved when the library was completed with extra-weighted spiking samples (optimally 10 samples × 40 times; $\mathrm{RPD_{val}} \approx 6$). Prediction was more accurate using local calibration without spiking, but on a validation set that was reduced from 134 to 115 samples ($\mathrm{RPD_{val}} \approx 8$; on the same reduced set, global calibration without spiking yielded $\mathrm{RPD_{val}} \approx 6$). Local calibration with spiking (optimally 10 samples without extra-weight) yielded moderately less accurate prediction but for the full validation set when ≥ 4 calibration neighbours were required ($\mathrm{RPD_{val}} \approx 7$), or higher accuracy on 115-116 samples when ≥ 50 neighbours were required ($\mathrm{RPD_{val}} \approx 9$).

Of course, parameter optimisation depended on the size, diversity and proximity of the calibration database and validation set, and the optimums proposed here could not be generalised (number and extra-weight of spiking samples, number of calibration neighbours). Nevertheless, the results of the present study confirmed that MIRS prediction of SIC is accurate, and demonstrated that the usefulness of large soil spectral libraries can be improved by the combination of representative spiking and local calibration, while extra-weighting of spiking samples has no additional effect.

Local calibration raised the question of either making prediction, or not, on validation samples poorly represented by the calibration database, as long as the database is incomplete; and the choice depends on user's priority (e.g. most accurate prediction or prediction on all samples). Anyway, identifying such samples is crucial, at least to know that prediction is not fully reliable for them. Moreover, it is predictable that spectral libraries will become more and more complete with time, and that accurate prediction will become possible for all samples. In such conditions, calcimetry could be increasingly replaced by MIRS for SIC analysis.

Acknowledgements

Soil sampling and SIC analyses regarding the RMQS (national library) 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(Agence de l'environnement et de la maîtrise de l'énergie, which is a French government agency concerned with environmental protection and energy management), IRD (Institut de recherche pour le développement, which is a French public research organisation dedicated to southern countries) and INRA (Institut national de la recherche agronomique, which is a French public research organisation dedicated to agriculture s.l.).

Spectrum acquisition on RMQS samples was supported by the MIRSol project, which was funded by ADEME (contract 0675C0102).

Soil sampling and SIC analyses regarding the target set were carried out during Patrice Coll's PhD, which was funded by ADEME and Montpellier SupAgro (which is a French public institute of higher education in agricultural science).

In addition, we thank two anonymous reviewers for their helpful recommendations.

References

Apesteguia, M., Plante, A.F., Virto, I., 2018. Methods assessment for organic and inorganic carbon quantification in calcareous soils of the Mediterranean region. Geoderma Regional 12, 39–48. DOI: 10.1016/j.geodrs.2017.12.001

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.

Arrouays, D., Saby, N.P.A., Thioulouse, J., Jolivet, C., Boulonne, L., Ratié, C., 2011. Large trends in French topsoil characteristics are revealed by spatially constrained multivariate analysis. Geoderma 161, 107-114. DOI: 10.1016/j.geoderma.2010.12.002

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied Spectroscopy 43, 772–777. DOI: 10.1366/0003702894202201

Barthès, B.G., Kouakoua, E., Moulin, P., Hmaidi, K., Gallali, T., Clairotte, M., Bernoux, M., Bourdon, E., Toucet, J., Chevallier, T., 2016. Studying the physical protection of soil carbon with quantitative infrared spectroscopy. Journal of Near Infrared Spectroscopy 24, 199–214. doi: 10.1255/jnirs.1232

Batjes, N. H., 2014. Total carbon and nitrogen in the soils of the world. European Journal of Soil Science 65, 4–21. DOI: 10.1111/ejss.12114_2

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 Biology & Biochemistry 43, 1398–1410. doi: 10.1016/j.soilbio.2011.02.019

Bertrand, D., 2000. Prétraitement des données spectrales. In: Bertrand, D., Dufour, E. (Eds.), La Spectroscopie Infrarouge et ses Applications Analytiques, 2nd Ed. Tec & Doc, Paris, pp. 351–393.

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 & Francis, Boca Raton, FL, USA, pp. 185–207.

Brown, D.J., 2007. Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. Geoderma 140, 444–453. doi:10.1016/j.geoderma.2007.04.021

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. doi:10.1016/j.geoderma.2005.01.001

Brown, D.J., Shepherd, K.D., Walsh, M.G., Mays, M.D., Reinsch, T.G., 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma 132, 273–290. doi: 10.1016/j.geoderma.2005.04.025

Castaldi, F., Chabrillat, S., Chartin, C., Genot, V., Jones, A.R., van Wesemael, B., 2018. Estimation of soil organic carbon in arable soil in Belgium and Luxembourg with the LUCAS topsoil database. European Journal of Soil Science 69, 592–603. doi: 10.1111/ejss.12553

Cécillon, L., Barthès, B.G., Gomez, C., Ertlen, D., Genot, V., Hedde, M., Stevens, A., Brun, J.J., 2009. Assessment and monitoring of soil quality using near infrared reflectance spectroscopy (NIRS). European Journal of Soil Science 60, 770–784. doi: 10.1111/j.1365-2389.2009.01178.x

Chang, C.W., Laird, D.A., Mausbach, M.J., Hurburgh, C.R.J., 2001. Near-infrared reflectance spectroscopy—principal components regression analyses of soil properties. Soil Science Society of America Journal 65, 480–490. doi: 10.2136/sssaj2001.652480x

Chevallier, T., Cournac, L., Hamdi, S., Gallali, T., Bernoux, M., 2016. Temperature dependence of CO_2 emissions rates and isotopic signature from a calcareous soil. Journal of Arid Environments 135, 132–139. DOI: 10.1016/j.jaridenv.2016.08.002

Clairotte, M., Grinand, C., Kouakoua, E., Thébault, A., Saby, N.P.A., Bernoux, M., Barthès, B.G., 2016. National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy. Geoderma 276, 41–52. doi: 10.1016/j.geoderma.2016.04.021

Coll, P., 2011. Vineyard soil quality in Languedoc-Roussillon. Effects of agricultural practices. PhD Thesis, Montpellier SupAgro, Montpellier, France. (At: http://www.supagro.fr/theses/extranet/11-0030_Coll.pdf. Accessed: 03/09/2019)

Comstock, J.P., Sherpa, S.R., Ferguson, R., Bailey, S., Beem-Miller, J.P., Lin, F., Lehmann, J., Wolfe, D.W., 2019. Carbonate determination in soils by mid-IR spectroscopy with regional and continental scale models. PLoS ONE 14(2), e0210235. https://doi.org/10.1371/journal.pone.0210235

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. Agronomy for Sustainable Development 37, 14. DOI: 10.1007/s13593-017-0421-2

Emmerich, W.E., 2003. Carbon dioxide fluxes in a semiarid environment with high carbonate soils. Agricultural and Forest Meteorology 116, 91–102. DOI: 10.1016/S0168-1923(02)00231-9

Geladi, P., MacDougall, D., Martens, H., 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. Applied Spectroscopy 39, 491–500. DOI: 10.1366/0003702854248656

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. Journal of Near Infrared Spectroscopy 19, 117–138. DOI: 10.1366/0003702854248656

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. http://doi.org/10.1016/j.geoderma.2013.07.016

Gredilla, A., Fdez-Ortiz de Vallejuelo, S., Elejoste, N., de Diego, A., Madariaga, J.M., 2016. Non-destructive spectroscopy combined with chemometrics as a tool for green chemical analysis of environmental samples: A review. Trends in Analytical Chemistry 76, 30–39. DOI: 10.1016/j.trac.2015.11.011

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). European Journal of Soil Science 63, 141–151. doi: 10.1111/j.1365-2389.2012.01429.x

Guerrero, C., Zornoza, R., Gómez, I., Mataix-Beneyto, J., 2010. Spiking of NIR regional models using samples from target sites: Effect of model size on prediction accuracy. Geoderma 158, 66–77. doi:10.1016/j.geoderma.2009.12.021

Guerrero, C., Stenberg, B., Wetterlind, J., Viscarra Rossel, R.A., Maestre, F.T., Mouazen, A.M., Zornoza, R., Ruiz-Sinoga, J.D., Kuang, B., 2014. Assessment of soil organic carbon at local scale with spiked NIR calibrations: effects of selection and extra-weighting on the spiking subset. European Journal of Soil Science 65, 248–263. doi: 10.1111/ejss.12129

Guerrero, C., Wetterlind, J., Stenberg, B., Mouazen, A.M., Gabarrón-Galeote, M.A., Ruiz-Sinoga, J.D., Zornoza, R., Viscarra Rossel, R.A., 2016. Do we really need large spectral libraries for local scale SOC assessment with NIR spectroscopy? Soil & Tillage Research 155, 501–509. http://doi.org/10.1016/j.still.2015.07.008

Harris, D., Horwath, W.R., van Kessel, C., 2001. Acid fumigation of soils to remove carbonates prior to total organic carbon or carbon-13 isotopic analysis. Soil Science Society of America Journal 65, 1853–1856. DOI: 10.2136/sssaj2001.1853

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

IUSS (International Union of Soil Science) Working Group WRB (World Reference Base), 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.

JCGM (Joint Committee for Guides in Metrology), 2012. International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM), 3rd Ed. ISO, Geneva.

Lal, R., 2004. Soil carbon sequestration to mitigate climate change. Geoderma, 123, 1–22. DOI: 10.1016/j.geoderma.2004.01.032

Lal, R., Negassa, W., Lorenz, K. 2015. Carbon sequestration in soil. Current Opinion in Environmental Sustainability 15, 79–86. DOI: 10.1016/j.cosust.2015.09.002

Lobsey, C.R., Viscarra Rossel, R.A., Roudier, P., Hedley, C.B., 2017. RS-LOCAL data-mines information from spectral libraries to improve local calibrations. European Journal of Soil Science 68, 840-852. doi: 10.1111/ejss.12490

Mark, H.L., Tunnell, D., 1985. Qualitative near-infrared reflectance analysis using Mahalanobis distances. Analytical Chemistry 57, 1449–1456. DOI: 10.1021/ac00284a061

McCarty, G.W., Reeves III, J.B., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Science Society of America Journal 6, 640–646. DOI: 10.2136/sssaj2002.0640

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. doi: 10.1016/j.geoderma.2009.04.010

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

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. Journal of Near Infrared Spectroscopy 21, 495–509. doi: 10.1255/jnirs.1080

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Stevens, A., Demattê, J.A.M., Scholten, T., 2013. The spectrum-based learner: A new local approach for modeling soil vis-NIR spectra of complex datasets. Geoderma 195, 268–279. doi: 10.1016/j.geoderma.2012.12.014

Salomé, C., Coll, P., Lardo, E., Villenave, C., Blanchart, E., Hinsinger, P., Marsden, C., Le Cadre, E., 2014. Relevance of use-invariant soil properties to assess soil quality of vulnerable ecosystems: the case of Mediterranean vineyards. Ecological Indicators 43, 83–93. DOI: 10.1016/j.ecolind.2014.02.016

Salomé, C., Coll, P., Lardo, E., Metay, A., Villenave, C., Marsden, C., Blanchart, E., Hinsinger, P., Le Cadre, E., 2016. The soil quality concept as a framework to assess management practices in vulnerable agroecosystems: A case study in Mediterranean vineyards. Ecological Indicators 61, 456–465. DOI: 10.1016/j.ecolind.2015.09.047

Sanderman, J., 2012. Can management induced changes in the carbonate system drive soil carbon sequestration? A review with particular focus on Australia. Agriculture, Ecosystems & Environment 155, 70–77. DOI: 10.1016/j.agee.2012.04.015

Sankey, J.B., Brown, D.J., Bernard, M.L., Lawrence, R.L., 2008. Comparing local vs. global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C. Geoderma 148, 149-158. doi:10.1016/j.geoderma.2008.09.019

Seidel, M., Hutengs, C., Ludwig, B., Thiele-Bruhn, S., Vohland, M., 2019. Strategies for the efficient estimation of soil organic carbon at the field scale with vis-NIR spectroscopy: Spectral libraries and spiking vs. local calibrations. Geoderma 354, 113856. https://doi.org/10.1016/j.geoderma.2019.07.014

Shenk, J.S., Westerhaus, M.O., 1991. Population definition, sample selection and calibration procedures for near infrared reflectance spectroscopy. Crop Science 31, 469–474. DOI: 10.2135/cropsci1991.0011183X003100020049x

Shenk, J., Westerhaus, M., Berzaghi, P., 1997. Investigation of a LOCAL calibration procedure for near infrared instruments. Journal of Near Infrared Spectroscopy 5, 223–232. DOI: 10.1255/jnirs.115

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. Applied Spectroscopy Reviews 49, 139–186. http://dx.doi.org/10.1080/05704928.2013.811081

Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and near infrared spectroscopy in soil science. Advances in Agronomy 107, 163–215. doi: 10.1016/S0065-2113(10)07005-7

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. doi:10.1371/journal.pone.0066409

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. doi: 10.1016/j.geoderma.2005.03.007

Viscarra Rossel, R.A.,Webster, R., 2012. Predicting soil properties from the Australian soil visible-near infrared spectroscopic database. European Journal of Soil Science 63, 848–860. doi: 10.1111/j.1365-2389.2012.01495.x

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodský, L., Changwen, D., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero Maestre, C., Hedley, C.B., Knadel, M., Morrás, H.J.M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Rufasto Campos, E.M., Sanborn, P., Sellitto, V.M., Sudduth, K.A., Rawlins, B.G., Walter, C., Winowiecki, L.A., Young Hong, S., Ji, W., 2016. A global spectral library to characterize the world's soil. Earth-Science Reviews 155, 198–230. doi: 10.1016/j.earscirev.2016.01.012

Walkley, A., Black, I.A., 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modificaton of the chromic acid titration method. Soil Science 37, 29–38.

Table 1. Presentation of the sites of the target set.

Town or village	Number of plots	Longitude and latitude	Elevation (m a.s.l.)	Soil type[a]	Mean clay content (%)	Mean sand content (%)	SIC range and mean (g kg^{-1})
Terrats	11	42° 36’ 27” N, 02° 46’ 14” E	135	Luvisol and Cambisol	27	41	0-3, 0
Lesquerde	19	42° 48’ 01” N, 02° 31’ 47” E	358	Arenosol	7	81	0-1, 0
Montagnac	21	43° 28’ 50” N, 03° 29’ 02” E	54	Calcisol	38	32	23-84, 50
Faugères	21	43° 33’ 57” N, 03° 11’ 19” E	284	Cambisol	16	52	0-12, 1
Aigues-Mortes	18	43° 34’ 02” N, 04° 11’ 33” E	5	Arenosol	7	88	17-25, 23
Vergèze	17	43° 44’ 37” N, 04° 13’ 14” E	32	Cambisol	36	18	2-50, 28
Jonquières Saint-Vincent	19	43° 49’ 38” N, 04° 33’ 48” E	37	Rhodic Luvisol	18	43	0-6, 1
Saint-Hippolyte-du-Fort	15	43° 57’ 56” N, 03° 51’ 28” E	170	Calcisol	42	22	7-86, 43
Saint-Victor-la-Coste	23	44° 03’ 38” N, 04° 38’ 29” E	143	Calcisol	16	54	1-43, 20
Whole target set	164	-	-	-	-	-	0-86, 19
National library	2178	-	-	-	-	-	0-104, 6

[a] IUSS Working Group WRB (2014).

Table 2. Prediction results after global calibration without or with spiking.

Spiking samples and weight	Best pre-treatment	Calibration							Validation							
		N_{cal}[a]	$Mean_{cal}$[b]	SD_{cal}[b]	LV[c]	SECV[d]	R^2_{cv}[e]	RPD_{cv}[f]	N_{val}[a]	$Mean_{val}$[b]	SD_{val}[b]	SEP[d]	bias	slope	R^2_{val}[e]	RPD_{val}[f]
0	None1155	2178	6.4	16.0	11	2.7	0.97	6.0	134	17.8	19.3	5.2	-2.8	1.10	0.96	3.7
10×1	D1155	10 spiking samples only	22.7	30.2	3	5.3	0.97	5.7	134	17.8	19.3	3.6	0.4	1.04	0.97	5.3
30×1	D1205	30 spiking samples only	23.3	29.0	3	5.0	0.97	5.8	134	17.8	19.3	3.8	-0.6	1.10	0.97	5.1
10×1	D1155	2188	6.5	16.1	11	2.7	0.97	6.0	134	17.8	19.3	4.9	-2.5	1.09	0.96	3.9
10×10	D1205	2278	7.1	17.1	13	2.9	0.97	5.9	134	17.8	19.3	3.9	-1.2	1.07	0.97	5.0
10×40	D1205	2578	8.9	19.5	14	3.2	0.97	6.1	134	17.8	19.3	3.3	-0.1	1.04	0.97	5.9
10×50	D1155	2678	9.4	20.0	14	3.2	0.97	6.3	134	17.8	19.3	3.3	0.2	1.03	0.97	5.9
30×10	None1155	2478	8.4	18.8	11	2.8	0.98	6.7	134	17.8	19.3	4.0	-1.6	1.08	0.97	4.8
30×200	D1155	8178	18.8	26.8	14	3.1	0.99	8.7	134	17.8	19.3	3.0	-0.2	1.04	0.98	6.4

[a] N_{cal} and N_{val} are the number of calibration samples and of validation samples predicted, respectively.

[b] $Mean_{cal}$, SD_{cal}, $Mean_{val}$ and SD_{val} are the mean and standard deviation over the calibration and validation set, respectively (in g kg^{-1}).

[c] LV is the number of PLS latent variables (or terms).

[d] SECV and SEP are the standard error of cross-validation and prediction, respectively (in g kg^{-1}).

[e] R^2_{cv} and R^2_{val} are the determination coefficient for cross-validation and validation, respectively.

[f] RPD_{cv} is the ratio of SD_{cal} to SECV, RPD_{val} is the ratio of SD_{val} to SEP.

Table 3. Prediction results after local calibration without or with spiking (and comparison with global calibration on the same validation samples).

Spiking samples and weight	Procedure	Best pre-treatment	Validation							
			N_{val}[a]	$Mean_{val}$[b]	SD_{val}[b]	SEP[c]	bias	slope	R^2_{val}[d]	RPD_{val}[e]
0	Local calibration using 4 to possibly 2178 calibration neighbours	None001	133	17.8	19.4	4.8	-2.0	1.07	0.95	4.1
0	Global calibration with 2178 samples and the same 133 validation samples	None1155	133	17.8	19.4	4.9	-2.7	1.09	0.96	4.0
0	Local calibration using 50 to possibly 2178 calibration neighbours	None001	115	17.0	20.7	2.7	-0.8	1.02	0.99	7.6
0	Global calibration with 2178 samples and the same 115 validation samples	None1155	115	17.0	20.7	3.3	-1.6	1.05	0.98	6.4
10×1	Local calibration using 4 to possibly 2188 calibration neighbours	SNV001	134	17.8	19.3	2.7	-0.0	1.01	0.98	7.3
10×1	Local calibration using 50 to possibly 2188 calibration neighbours	SNVD001	115	17.0	20.7	2.3	-0.5	1.03	0.99	8.9
10×1	Global calibration with 2188 samples and the same 115 validation samples	SNV1155	115	17.0	20.7	2.7	-0.2	1.00	0.98	7.8
10×10	Local calibration using 4 to possibly 2278 calibration neighbours	SNV001	134	17.8	19.3	2.7	-0.5	1.01	0.98	7.2
10×10	Local calibration using 50 to possibly 2278 calibration neighbours	SNVD001	115	17.0	20.7	2.3	-0.4	1.03	0.99	9.1
10×10	Global calibration with 2278 samples and the same 115 validation samples	D1205	115	17.0	20.7	2.7	-0.5	1.04	0.98	7.5
10×50	Local calibration using 4 to possibly 2678 calibration neighbours	SNV001	134	17.8	19.3	2.7	-0.5	1.01	0.98	7.2
10×50	Local calibration using 50 to possibly 2678 calibration neighbours	D001	134	17.8	19.3	2.7	-0.8	1.03	0.98	6.8
10×50	Global calibration with 2678 samples and all 134 validation samples	D1155	134	17.8	19.3	3.3	0.2	1.03	0.97	5.9
30×1	Local calibration using 4 to possibly 2208 calibration neighbours	SNV001	134	17.8	19.3	2.7	-0.5	1.03	0.98	7.2
30×1	Local calibration using 50 to possibly 2208 calibration neighbours	SNVD001	115	17.0	20.7	2.3	-0.5	1.03	0.99	9.2
30×1	Global calibration with 2208 samples and the same 115 validation samples	SNV1155	115	17.0	20.7	2.6	-0.4	1.00	0.98	7.9
30×10	Local calibration using 4 to possibly 2478 calibration neighbours	SNV001	134	17.8	19.3	2.7	-0.5	1.04	0.98	7.2
30×10	Local calibration using 50 to possibly 2478 calibration neighbours	SNV001	116	17.0	20.6	2.3	-0.5	1.04	0.99	8.8
30×10	Global calibration with 2478 samples and the same 116 validation samples	SNVD1155	116	17.0	20.6	2.5	-0.2	1.03	0.99	8.2

[a] N_{val} is the number of validation samples predicted.

[b] $Mean_{val}$ and SD_{val} are the mean and standard deviation over the validation set (in g kg^{-1}).

[c] SEP is the standard error of prediction (in g kg^{-1}).

[d] R^2_{val} is the determination coefficient for validation.

[e] RPD_{val} is the ratio of SD_{val} to SEP.

Figure captions

Fig. 1. Distribution of soil inorganic carbon (SIC) content in the national library (2178 samples, 0–30 cm depth) and in the target set (164 samples, 0–15 cm depth).

Fig. 2. Comparison between reference measurements and MIRS predictions of SIC content on the validation set using (a) global calibration without spiking, (b) local calibration using 4 to 2188 spectral neighbours from the national library enriched with 10 spiking samples without extra-weighting, (c) local calibration using 50 to 2278 neighbours from the national library enriched with 10 spiking samples extra-weighted 10 times, and (d) global calibration with similar spiking conditions and the same validation samples.

Fig. 3. Projection of the validation samples on the PCA built with the national library.

Fig. 4. Standard error of prediction (SEP) of SIC content when performing global MIRS calibration with the national library and spiking samples, as affected by the number and weight of spiking samples, or with spiking samples only (i.e. without the national library).

Fig. 5 Distribution of validation samples according to their number of spectral neighbours in the calibration set (a) without spiking; (b) with 10 spiking samples replicated 1 time, (c) replicated 10 times, and (d) replicated 50 times; and (e) with 30 spiking samples replicated 1 time, (f) replicated 10 times, and (g) replicated 50 times.

Fig. 1. Distribution of soil inorganic carbon (SIC) content in the national library (2178 samples, 0–30 cm depth) and in the target set (164 samples, 0–15 cm depth).

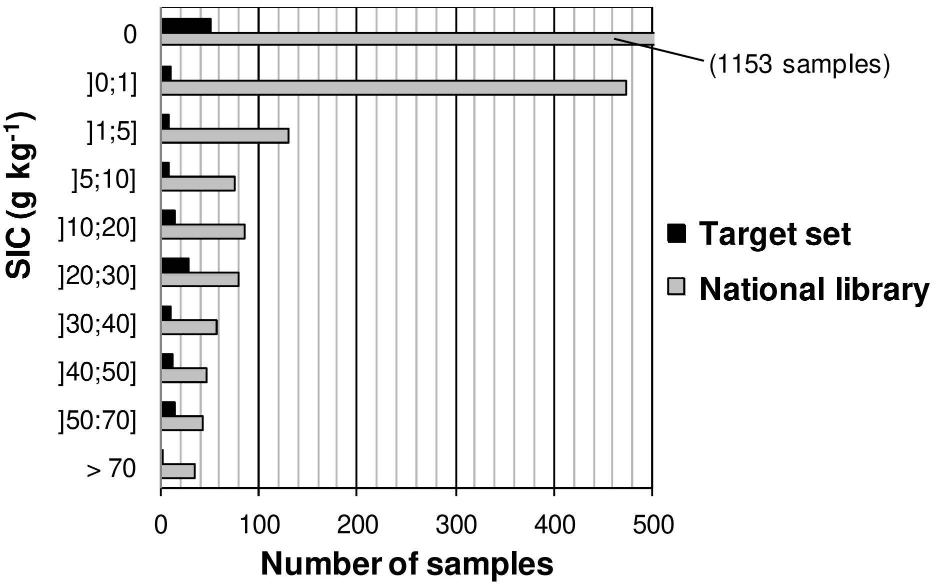

Fig. 2. Comparison between reference measurements and MIRS predictions of SIC content on the validation set using (a) global calibration without spiking, (b) local calibration using 4 to 2188 spectral neighbours from the national library enriched with 10 spiking samples without extra-weighting, (c) local calibration using 50 to 2278 neighbours from the national library enriched with 10 spiking samples extra-weighted 10 times, and (d) global calibration with similar spiking conditions and the same validation samples.

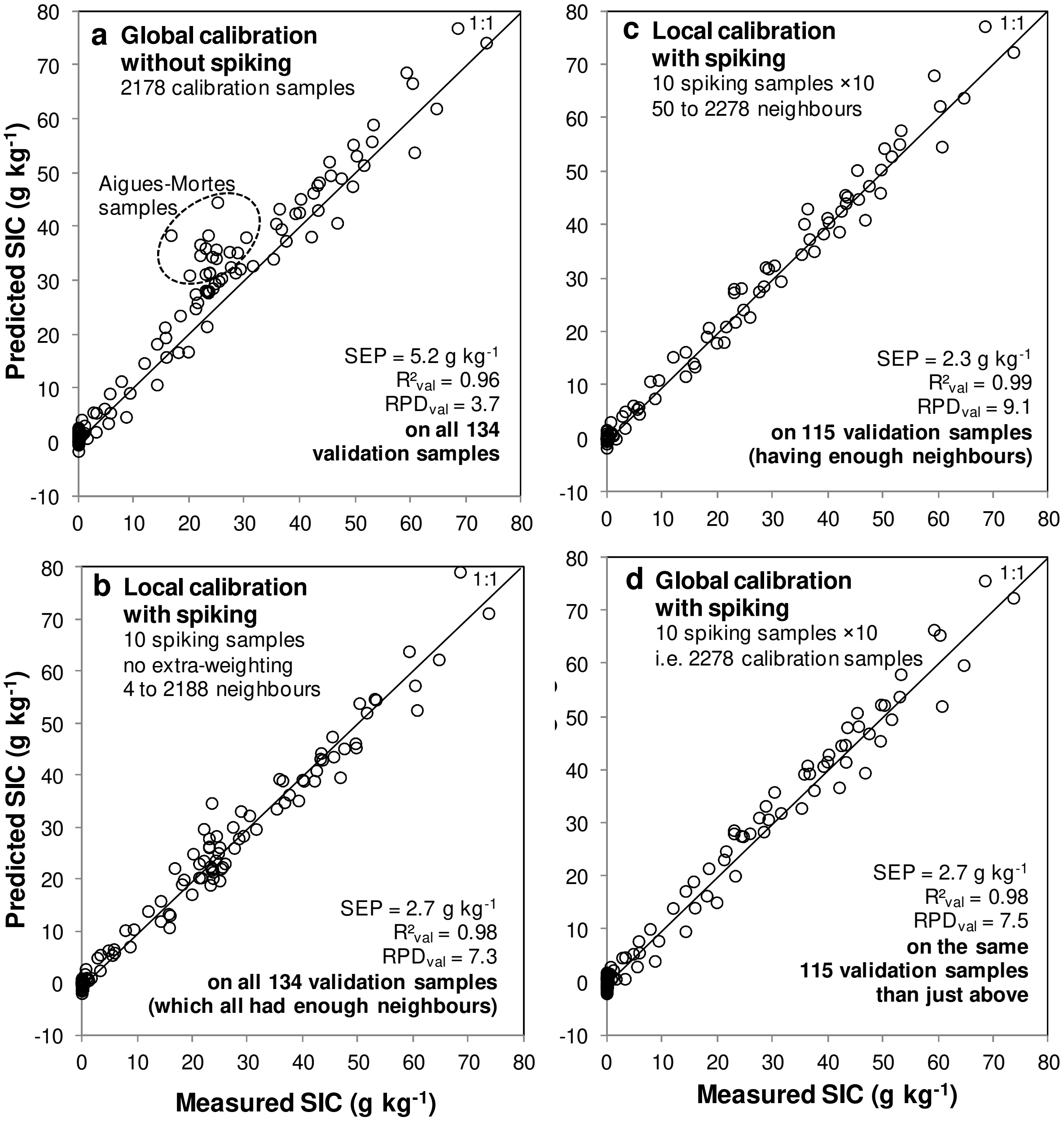

Fig.3. Projection of the validation samples on the PCA built with the national library.

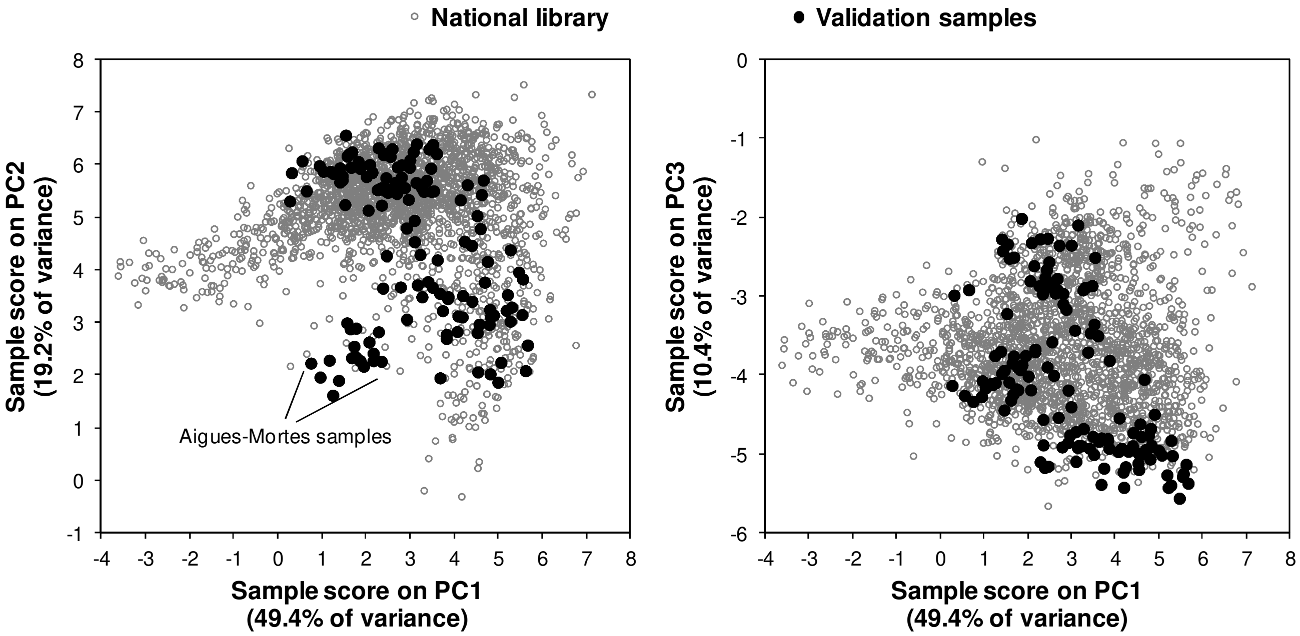

Fig. 4. Standard error of prediction (SEP) of SIC content when performing global MIRS calibration with the national library and spiking samples, as affected by the number and weight of spiking samples, or with spiking samples only (i.e. without the national library).

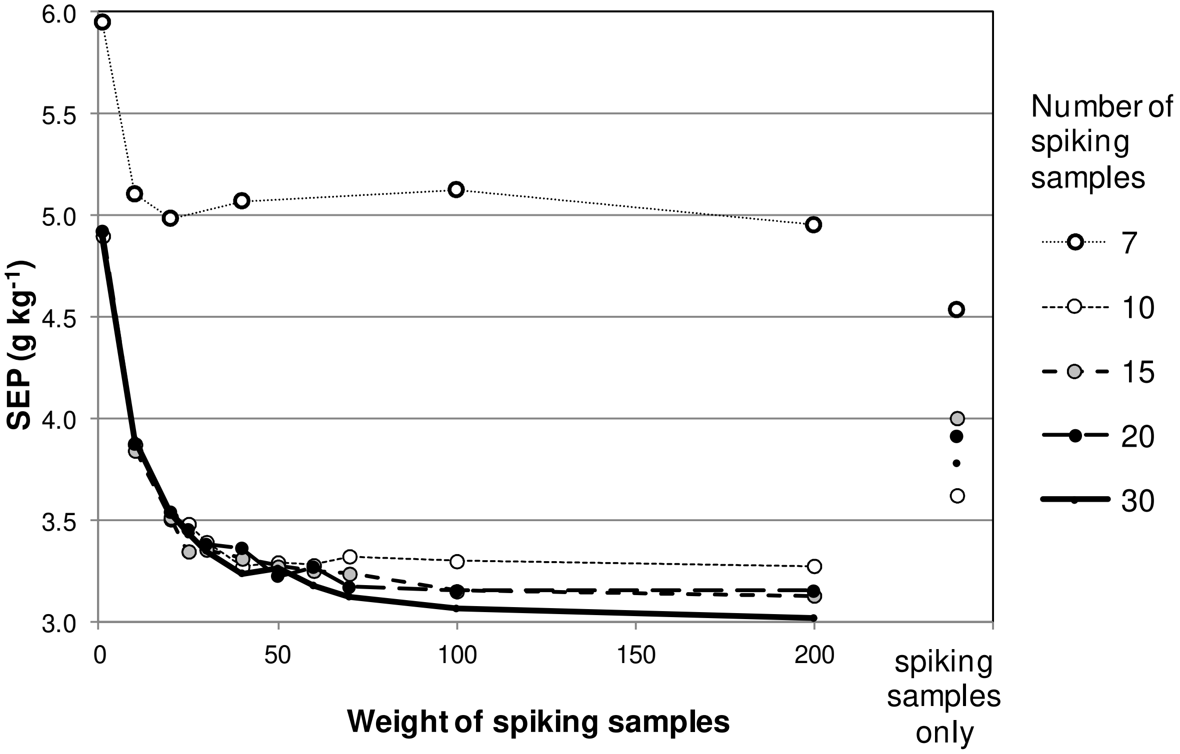

Fig. 5. Distribution of validation samples according to their number of spectral neighbours in the calibration set (a) without spiking; (b) with 10 spiking samples replicated 1 time, (c) replicated 10 times, and (d) replicated 50 times; and (e) with 30 spiking samples replicated 1 time, (f) replicated 10 times, and (g) replicated 50 times.