Comparison of near and mid-infrared reflectance spectroscopy for the estimation of soil organic carbon fractions in Madagascar agricultural soils

Nandrianina Ramifehiarivo, Bernard G. Barthès, Aurélie Cambou, Lydie Chapuis Lardy, Tiphaine Chevallier, Alain Albrecht, Tantely Razafimbelo

Comparison of near and mid-infrared reflectance spectroscopy for the estimation of soil organic carbon fractions in Madagascar agricultural soils

Nandrianina Ramifehiarivo[1,2*], Bernard G. Barthès[3], Aurélie Cambou[3], Lydie Chapuis-Lardy[3], Tiphaine Chevallier[3], Alain Albrecht[3], Tantely Razafimbelo[1]

[1]*Laboratoire des RadioIsotopes, University of Antananarivo, BP3383 Route d'Andraisoro, 101 Antananarivo, Madagascar*

[2]*Ecole Supérieure des Sciences Agronomiques (ESSA), University of Antananarivo, Ambohitsaina, 101 Antananarivo, Madagascar*

[3]*Eco&sols, Université de Montpellier, Cirad, Inrae, IRD, Institut Agro, 34060 Montpellier, France*

* **Corresponding author:** Nandrianina Ramifehiarivo

Phone: +261341025147; email: ranandrianina@hotmail.fr; *Laboratoire des RadioIsotopes, University of Antananarivo, BP3383 Route d'Andraisoro, 101 Antananarivo, Madagascar*

Abstract

Assessing the different pools of soil organic carbon (SOC) improves our understanding of how and at what rate the different forms of carbon (C) are being formed or lost in soils. Physical fractionation of soil organic matter (SOM) has often been used to separate and quantify SOC pools, but this approach is very tedious and can rarely be performed on large sample sets. Infrared spectroscopy has proven useful for time- and cost-effective quantification of total SOC, which prompted us to study its ability to characterise the distribution of SOC in physical fractions. This study aimed to compare the potential of near- and mid-infrared reflectance spectroscopy (NIRS and MIRS, respectively) to predict the distribution of SOC in particle-size and particle-density fractions, using spectra of unfractionated soils. A set of 134 sieved (< 2 mm) soil samples originating from seven sites in contrasting pedoclimatic regions of Madagascar was studied. For each sample, five SOM fractions were separated: the particulate organic matter fraction (POM) and the particle-size fractions > 200 µm, 50-200-µm, 20-50-µm and < 20 µm. The mass (g fraction 100 g^{-1} soil), SOC concentration (gC kg^{-1} fraction) and SOC amount (gC fraction kg^{-1} soil) were determined for each fraction in the laboratory. The NIR and MIR spectra were acquired on finely ground (< 0.2 mm) aliquots of unfractionated soil. Then, spectra were used for the prediction of each variable (i.e., the mass, SOC concentration and amount of each fraction,

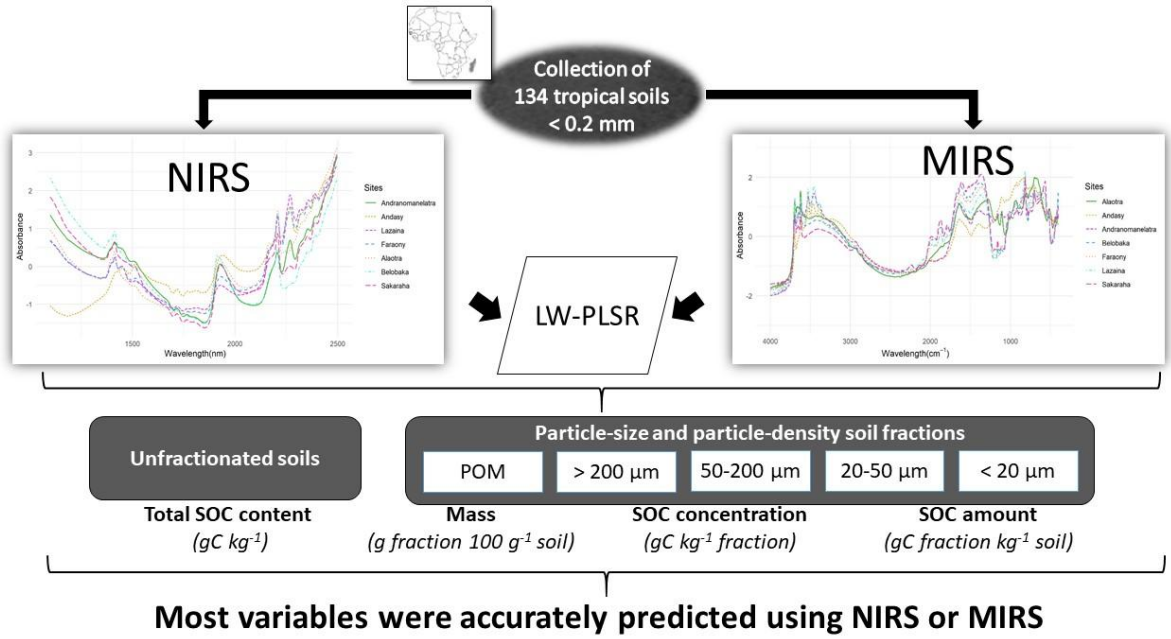
and SOC content of unfractionated soil), which was achieved with locally weighted partial least squares regression (LW-PLSR) on NIRS and MIRS data separately. For both spectral ranges, the same samples were used for calibration (n = 109, selected for spectral representativeness) and validation (n = 25).

Models based on NIRS and MIRS yielded excellent predictions for SOC content in unfractionated soil ($R^2 = 0.98$ and ratio of performance to interquartile range RPIQ ≥ 13) and accurate predictions ($R^2 \geq 0.75$ and RPIQ ≥ 2) for the mass, SOC concentration and SOC amount in most fractions. The predictions of SOC concentrations and SOC amounts were better in the fractions < 200 µm ($R^2 \geq 0.85$ and RPIQ > 3), especially in the fraction < 20 µm ($R^2 > 0.9$ and RPIQ > 4.5). In most cases, NIRS slightly outperformed MIRS. This result contradicts most previous studies performed on soils from temperate regions but confirms those performed on soils from tropical regions.

Infrared spectroscopy allowed accurate prediction of SOC distribution in particle-size fractions, which paves the way to high-throughput characterization of SOM.

Keywords: soil fractionation; locally weighted PLSR, infrared reflectance spectroscopy, soil organic matter, tropical soils, Ferralsols, Arenosols, Fluvisols

Graphical abstract



1. Introduction

Soil organic matter (SOM) plays multifunctional roles in ecosystem services provided by soils such as climate change regulation, biomass provision, water retention, agroecosystem resilience, and human prosperity (Prăvălie et al., 2021). SOM improves soil structure, supports soil biological activities that ensure plant nutrient recycling and is the largest stock of carbon (C) in the continental biosphere (Meurer et al., 2020; Swinton et al., 2007). The SOM content and the soil functions it supports are vulnerable to land use and management and to climate change (Osland et al., 2018; Paustian et al., 2016). The role of SOM in agronomic and environmental issues has been recognised as essential by several international initiatives, such as the Land Degradation Neutrality (Gnacadja and Wiese, 2016), the 4 per mille "soils for food security and climate" (Minasny et al., 2017), and the REDD+ (Vargas et al., 2013). These initiatives stress the importance of accurate and explicit information on soil, particularly on SOM and soil organic carbon (SOC) dynamics (Amelung et al., 2020).

The SOC stocks and dynamics depend on land use, soil type and/or crop management, vegetation and water resources (Lal, 2009; Wiesmeier et al., 2019). The SOC changes in agricultural soils after crop management seemed to be mostly controlled by the amount of organic inputs (Fujisaki et al. 2018), but questions remain about the persistence of the newly stored C in soils. Studying SOC dynamics is complex, as SOC persistence results from the continual transformation and movement of organic compounds throughout the soil matrix in interaction with decomposer activity (Dynarski et al. 2020). If scientists mostly agree that SOC dynamics should be understood through the concept of functional complexity derived from the spatial and temporal variation in diversity and composition of organic compounds in colocation with microbial communities (Lehmann et al. 2020), the way to easily observe and measure that complexity remains. For years, soil scientists have tried to experimentally approach this continuum and complex organic and mineral mixtures by separating meaningful pools with different turnovers (Chenu et al., 2015). The stability of SOC was studied by chemical extraction, physical fractionation, or more recently, thermal approaches (Malou et al., 2020, Soucémarianadin et al., 2019). Soil physical fractionation into SOM of several particle-sizes has been studied for 40 years (Balesdent, 1987; Christensen, 1992; Feller et al. 1979; Puget et al. 2000). The fine fractions are considered to be more decomposed and more stable with lower C:N ratios and higher turnover than the coarser fractions (Ding et al., 2014; Feller and Beare, 1997). The fine silt and clay-size fractions (<20 μm) have been recognised to be organo-mineral complexes and to highly contribute to SOC stabilization (Hassink et al. 1997). Therefore, the distribution of SOC in physical fractions has been largely used to

characterise SOC pools, with labile or stable, particulate or mineral-associated SOC, in various land uses and soil management practices in temperate or tropical areas (e.g., d'Annunzio et al., 2008; Poeplau et al., 2018; von Lützow et al., 2007). Improving organic inputs through a change in land use or management method can increase the total SOC content and modify the distribution of SOC into different fractions. The coarser fractions fed by the new organic inputs thus increase, but the finest fractions can also be C enriched (Feller and Beare, 1997; Feng et al. 2013; Fujisaki et al. 2018). This is particularly interesting since the SOC fine fractions have been recognised to stabilise SOC and to play a role in climate mitigation (Hassink et al. 1997; Wiesmeier et al. 2019), and the SOC coarse fractions have a positive effect on crop yields (Wood et al. 2016).

This explains why this experimental method of separation is still used, as it splits soil fractions that are easily conceptualised and understood with different turnovers and that can be integrated into SOC dynamic models (Lavallee et al. 2020). However, to better understand and model SOC dynamics using SOC fractions, more data are needed in different agricultural contexts. As the experimental steps to isolate each fraction are time-consuming and the analysis of C content in each soil fraction is expensive, such data are rarely obtained. Several methods exist and were recently reviewed by Poeplau et al. (2018). These methods mostly involve several steps: (i) completely dispersing soil samples in water with eventual chemical addition (e.g., sodium hexametaphosphate; NaHMP) and ultrasonication, (ii) sieving using different meshes, with eventual density fractionation (iii) recovering all fractions separately, (iv) drying and weighing the fractions, and (v) analysing the SOC concentration in every fraction. Applying a standard and simplified method would be helpful, but in highly aggregated soils, such as clayey tropical soils with Al and Fe oxyhydroxydes, disaggregation of the water-stable macroaggregates is difficult (Barthès et al. 2008b) and requires both chemical addition and ultrasonication (Gavinelli et al. 1995, Feller and Beare, 1997).

In contrast, alternative methods, such as mid- or near infrared reflectance spectroscopy (MIRS and NIRS, respectively), have been shown to allow time- and cost-effective determinations of total SOC content and (mineral) particle-size distribution. Some authors have studied the SOC content and particle-size distribution separately. For instance, Ge et al. (2014) showed that MIR spectra of finely ground samples allowed accurate predictions of clay, sand and SOC contents. Parent et al. (2021) achieved accurate prediction of the soil particle-size distribution using either NIRS or MIRS. Other authors used infrared (IR) spectra of bulk soils to directly predict different SOC pools. For instance, Knox et al. (2015) used MIR or visible and NIR (VNIR) spectra to predict recalcitrant SOC (ROC) and hydrolysable

SOC (HOC), while Madhavan et al. (2017) used MIR and NIR spectra to predict particulate SOC (POC), ROC and HOC. The ability of IR spectroscopy to predict SOC distribution in particle-size fractions based on spectra of bulk soil has also been reported, either using the NIR (Barthès et al., 2008a; Cozzolino and Morón, 2006, Jaconi et al., 2019) or MIR range (Sanderman et al., 2021, Zimmermann et al., 2007). Many studies have reported better MIRS than NIRS predictions of soil properties, especially organic properties (Bellon-Maurel and McBratney, 2011; Reeves, 2010), but some authors have suggested that this might be questionable for soils from tropical regions (Rabenarivo et al., 2013). A study that compared both spectral ranges for the prediction of SOC fractions in a temperate site did not show superiority of either range when spectra were acquired on ground dry samples (Greenberg et al., 2021), but such a comparison has not yet been carried out in tropical regions.

The objectives of the present work were (i) to test the ability of IR spectroscopy to predict the SOC distribution in different particle-size and -density fractions: the particulate organic matter (POM) and particle-size fractions > 200 µm, 50-200-µm, 20-50-µm and < 20 µm; and (ii) to compare NIRS and MIRS predictions of the mass (g fraction 100 g^{-1} soil), SOC concentration (gC kg^{-1} fraction) and SOC amount (gC fraction kg^{-1} soil) in each fraction, as well as SOC content in the unfractionated soil (gC kg^{-1}), for a collection of cultivated topsoil samples originating from a range of pedoclimatic conditions in Madagascar.

2. Materials and methods

2.1. Soil samples

The soil sample set (n = 134) gathered samples from previous studies on experimental plots and from a pot experiment (Razafimbelo et al., 2006, 2008, 2013). Samples originated from the 0-5-cm-depth layer at seven sites in contrasting pedoclimatic regions of Madagascar (the pot experiment also used the 0-5-cm soil layer). Altitude ranged from 12 to 1650 m above sea level, mean annual precipitation was from 800 to 2500 mm, and the mean annual temperature was from 16 to 28 °C (Table 1). The soils were Ferralsols, Arenosols and Fluvisols (IUSS Working Group WRB, 2015), which represent some major tropical soil types used for agriculture, and their texture ranged from sandy to clayey (Figure 1). Half of the samples considered in the present study had been collected from agronomic experiments on conservation agriculture, based on crop rotations, mulch-based and no-tillage systems, at five locations: Andranomanelatra (n = 32), Andasy (n = 8), Faraony (n = 6), Sakaraha (n = 8) and Alaotra Lake (n = 8; Table 1). The plot designs involved no-tillage (NT) and conventionally tilled (CT) systems with an association of upland (rainfed) rice (*Oryza sativa* L.), soybean

(*Glycine maxima* L.), and cowpea (*Vigna unguiculata*) or a grass (*Stylosanthes ruziziensis*) as cover crops. The second half of the samples was collected from a 3-month pot experiment set up for studying residue application effects. The pots were made of 1 kg of 2-mm sieved topsoils (0-5 cm) collected from fallows at Andranomanelatra (n = 24), Belobaka (n = 24) and Lazaina (n = 24; Table 1). Experimental treatments were bare soil without plants or residue, association of rice and soybean without residue application, rice-soybean association with residues of rice or soybean on the soil surface, or with residues of rice or soybean incorporated into the soil. At the end of the pot experiment, soils were air-dried at room temperature, gently crushed using a mortar and pestle when necessary, sieved through a 2-mm mesh, and stored for further analyses.

2.2. Particle-size fractionation

The particle-size fractionation of SOM was carried out in 2011 on 30 g of 2-mm sieved air-dried soil samples, using a protocol adapted from Gavinelli et al. (1995). This protocol combined chemical dispersion with NaHMP as well as physical dispersion with glass beads and ultrasonication (Grandière et al., 2007). Poeplau et al. (2018) indicated that dispersion performance was slightly better with hexametaphosphate (HMP) than with ultrasonication or beads and presented simpler methods with only one dispersion method. However, in clayey tropical soils rich in Fe and Al sesquioxides, which include soil aggregates that are especially water-stable (Barthes et al. 2008b), complete soil dispersion is difficult to achieve with only HMP. Thus, to check if the soil dispersion without SOM destruction was complete, the mass distribution of the soil fractions was compared to the particle-size analysis. Each sample was pre-soaked overnight at 4 °C in 200 mL deionised water with 0.8 g of NaHMP. Samples were then shaken with ten agate balls (diam. 1 cm) in a rotary shaker for 12 h at 45 rev min^{-1}. The soil suspension was wet-sieved through a 50-µm sieve. The fraction remaining on the 50-µm sieve was density-separated into a free light fraction by repeated floatation panning in water and a remaining fraction. The free light fraction was oven dried at 60 °C. This fraction corresponded to the POM fraction (>50 µm). The remaining soil suspension fraction (i.e., without POM) was wet-sieved at 200 µm to isolate the fraction > 200 µm. The remaining 0-200-µm soil suspension was then wet-sieved at 50-µm. The 50-200-µm soil suspension was ultrasonicated for 3 min at 195 $J.mL^{-1}$ with a probe-type ultrasound-generating unit to disperse highly stable macroaggregates. After ultrasonication, the soil suspension was wet-sieved through the 50-µm sieve again to separate the 50-200-µm soil fraction. The suspensions <50 µm were gathered and ultrasonicated for 10 min and passed through a 20-µm

mesh sieve. The three fractions > 200 µm, 50-200-µm and 20-50-µm were oven-dried at 60 °C and weighed to obtain their masses (g fraction 100 g^{-1} soil). The resulting suspension < 20 µm was transferred to a 1-L glass cylinder, where water was added to bring the volume to 1 L. The cylinder was shaken by hand (30 end-over-end tumblings), and a 200-mL aliquot of the suspension was withdrawn immediately after. This aliquot of the < 20-µm fraction was then oven-dried at 60 °C, weighed, and referred to the cylinder volume.

All the fractions dried at 60 °C were finely ground (< 200 µm) and analysed by dry combustion, using a CHN Fisons/Carlo Erba NA 2000 elemental analyser (Milan, Italy). In the absence of carbonates, all carbon was assumed to be organic. Fraction SOC was expressed as concentration, in gC kg^{-1} fraction, and as amount, in gC fraction kg^{-1} soil (amount calculated as the product of mass and concentration).

Soil particle-size analysis was also carried out on each soil sample using the Robinson pipette method, after pre-treatment with heat and hydrogen peroxide (35%) to remove organic matter and with NaOH to disperse soil aggregates (Gee and Bauder, 1986).

2.3. NIRS analysis

Reflectance in the near infrared was measured between 1100 and 2500 nm (i.e., 9091 and 4000 cm^{-1}, respectively) at 2-nm interval with a Foss NIRSystems 5000 spectrometer (Silver Spring, MD, USA). The area scanned with this instrument is 42 mm^2. Overnight oven-dried (40 °C) aliquots ~ 1 g of air-dried, 2-mm sieved and 0.2-mm ground unfractionated samples were placed in static ring cups with a micro sample insert with a 1.6-cm inner diameter. Each spectrum, automatically averaged from 32 scans, was recorded as the logarithm of the inverse of the reflectance (i.e., absorbance).

2.4. MIRS analysis

Reflectance in the mid-infrared region was measured between 4000 and 400 cm^{-1} (i.e., 2500 and 25,000 nm, respectively) at 3.86-cm^{-1} intervals with a Nicolet 6700 FT-IR spectrophotometer (Thermo Fisher Scientific, Madison, WI, USA). The area scanned with this instrument is 10 mm^2. Overnight oven-dried (40 °C) aliquots of ~ 1 g of air-dried, 2-mm sieved and 0.2-mm ground unfractionated samples were placed in a 17-well plate and then scanned using an auto-sampler. The average of 32 co-added scans was recorded as the absorbance, as for the NIR spectra.

2.5. Selection of calibration and validation subsets

The analyses were performed on full spectra. Each spectrum was centred then smoothed by second order Savitzky–Golay smoothing (Savitzky and Golay, 1964). Then, a principal component analysis (PCA), with four principal components, was performed on the centred, smoothed NIR spectra (Figure 2), and another PCA was performed on centred, smoothed MIR spectra (Figure 3), both using the *prospectr* package in the R environment (Stevens et al., 2020). The calibration subset was selected in three steps. The first step was the selection of the 121 most representative samples based on NIR spectra (i.e., 90% of the set made of 134 samples) using the Kennard Stone algorithm (Kennard and Stone, 1969; Stevens et al., 2013). The second step was the selection of the 121 most representative samples based on MIR spectra, also using the Kennard Stone algorithm. The final step was the selection of the soil samples that belonged to both above-mentioned selections based on the most representative NIR (Figures 2 and 4) and MIR spectra (Figures 3 and 4). This selection yielded 109 samples, which originated from all sites. Of note, these samples were not necessarily the most representative based on their NIR spectrum or based on their MIR spectrum. The validation subset was made of the 25 remaining samples, which originated from Andasy, Andranomanelatra, Belobaka and Lazaina.

2.6. Chemometric analyses

The regression models between spectra and reference values (mass, SOC concentrations and amounts in each particle-size fraction) were built using locally weighted partial least squares regression (LW-PLSR) from the *rnirs* package in the R environment (Lesnoff, 2020). As with the typical PLSR, LW-PLSR condenses the spectral information into a small number of latent variables, which are orthogonal combinations of absorbances that account for most spectral information and covary with reference values. However, unlike the usual PLSR, LW-PLSR models are built individually for each predicted sample, and weights are assigned to calibration samples based on their spectral similarity to each predicted sample (Boysworth and Booksh, 2008; Cambou et al., 2022; Gupta et al., 2018). For each validation sample, the similarity was calculated from Spearman's correlation coefficients between the spectrum of that sample and the spectra of calibration samples. Full cross-validation was performed on the calibration subset to determine the optimal number of latent variables to be included in the prediction model. The residuals of the predictions were pooled to calculate the root mean square error of cross-validation between the predicted and measured values (RMSEcv), and

the number of latent variables that minimised RMSEcv was selected. Then, the model with that number of latent variables was applied to the validation sample considered.

Several usual spectrum pretreatments were tested, such as first- or second-order derivation, smoothing, standard normal variate transform (SNV), and a combination of derivation and SNV. The most appropriate spectrum pretreatment was selected for NIRS and MIRS modelling.

2.7. Model evaluation

The performance of the calibration models was assessed using the coefficient of determination (R^2) in cross-validation (R^2cv), the value of RMSEcv, the ratio of the observed standard deviation (SD) of the calibration subset to RMSEcv (denoted RPDcv), and the ratio of the observed interquartile range (IQR) to RMSEcv (denoted RPIQcv), where IQR is the difference between the third quartile Q3 and the first quartile Q1.

The prediction accuracy of the models was also evaluated on the validation subset, using validation R^2 (R^2val), the root mean square error of prediction (RMSEP), the ratio of the observed SD of the validation subset to RMSEP (denoted RPDval), and the ratio of the observed IQR of the validation subset to RMSEP (denoted RPIQval). The lower the RMSE was and the higher the R^2, RPD and RPIQ were, the more accurate the prediction model was. According to Chang et al. (2001), an RPD below 1.4 corresponds to poor predictions of soil properties, and an RPD between 1.4 and 2 corresponds to acceptable predictions, and an RPD above 2 corresponds to accurate predictions. However, for variables that do not follow a normal distribution (such as many soil properties), RPIQ has been recommended instead of RPD (Bellon-Maurel et al., 2010). Therefore, model accuracy was evaluated according to the RPIQ but using the thresholds proposed by Chang et al. (2001) for the RPD.

3. Results

3.1. Reference data

Reference data were obtained using conventional procedures. The per-site SOC content in the unfractionated soil (gC kg^{-1}) is presented in Figure 5, while fraction masses (g fraction 100 g^{-1} soil), SOC concentrations (gC kg^{-1} fraction) and SOC amounts (gC fraction kg^{-1} soil) are presented in Figure 6 (details are presented in the supplementary material, Table SM1). Variable distributions were often multimodal, such as the SOC content in the unfractionated soil (Figure 5). The per-site average SOC content in unfractionated soil ranged from 17-

26 gC kg^{-1} in sandy soils (sand > 50 g 100 g^{-1}) to 45-70 gC kg^{-1} in clayey soils (sand = 15-20 g 100 g^{-1}; Table 1).

The cumulative yield of fraction masses ranged from 89.0 to 98.5 % of unfractionated soil, and averaged 94.8 %; when considering sites separately, this mass recovery averaged 99 % at Faraony, 97 % at Andasy, Belobaka and Alaotra Lake, 96 % at Andranomanelatra, 92 % at Sakaraha and 89 % at Lazaina. The fraction mass distribution varied between sites, which differed in texture (Figure 6A). The per-site average fraction mass ranged from 0.4 to 1.7 % for POM, from 1.6 to 74.2 % for the > 200-µm fraction, from 5.8 to 54.3 % for 50-200-µm fraction, from 2.5 to 14.1 % for 20-50-µm fraction and from 8.7 to 71.9 % for < 20-µm fraction.

The fraction SOC concentration varied between sites and between fractions. The per-site average fraction SOC concentration ranged from 186 to 419 gC kg^{-1} fraction for POM, from 0.8 to 14.5 gC kg^{-1} for > 200-µm fraction, from 4.9 to 83.8 gC kg^{-1} for 50-200-µm fraction, from 15.1 to 71.3 gC kg^{-1} for 20-50-µm fraction and from 24.1 to 58.9 gC kg^{-1} for < 20-µm fraction (Figure 6B). Among sites and regardless of the fraction, Lazaina had the lowest SOC concentration (except for the < 20-µm fraction, which was lower in Alaotra Lake), while among fractions and regardless of the site, the > 200-µm fraction had by far the lowest SOC concentration (except in Alaotra Lake, where it was lower in the 50-200-µm fraction), and the POM fraction had the highest SOC concentration.

The per-site average fraction SOC amount ranged from 1.0 to 5.5 gC fraction kg^{-1} soil for POM, from 0.1 to 1.6 gC fraction kg^{-1} soil for > 200-µm fraction, from 0.8 to 11.0 gC fraction kg^{-1} soil for 50-200-µm fraction, from 0.6 to 8.2 gC fraction kg^{-1} soil for 20-50-µm fraction, and from 4.9 to 38.2 gC fraction kg^{-1} soil for < 20-µm fraction (Figure 6C). The < 20-µm fraction had the highest SOC amount at each site, while the > 200-µm fraction had the lowest (except at Faraony, where the mass of POM was very small, and at Sakaraha, where SOC amounts in POM, 50-200-µm and 20-50-µm fractions were very poor). As a proportion of SOC content in unfractionated samples, the SOC recovery after fractionation averaged 40 % at Sakaraha, 69 % at Lazaina, 75 % at Belobaka and Faraony, 86 % at Alaotra Lake, 91 % at Andranomanelatra, 95 % at Andasy, and 85 % over all samples.

3.2. Results of NIRS predictions

The results are presented in Table 2. The SOC content prediction of the unfractionated soil was excellent ($R^2cv = R^2val = 0.98$, RPIQcv and RPIQval > 14).

Each fraction mass was accurately calibrated (R^2cv = 0.83-0.97, RPIQcv = 2.5-8.8) except for POM (R^2cv = 0.59 and RPIQcv = 1.1). In external validation, models had excellent predictive ability for the masses of the > 200-µm and < 20 µm fractions (R^2val = 0.95 and RPIQval = 7.0-7.3) and were acceptable for POM, 50-200-µm and 20-50-µm fractions (R^2val = 0.52-0.87, and RPIQval = 1.5-2.2).

Fraction SOC concentrations were excellently calibrated for all fractions < 200 µm (R^2cv = 0.91-0.96, RPIQcv = 4.1-6.8) but not for POM and the > 200-µm fraction (R^2cv= 0.43-0.49, RPIQcv = 1.4-1.7). The same trend was observed in external validation, with accurate predictions for the 50-200-µm, 20-50-µm and < 20-µm fractions (R^2val = 0.92-0.99 and RPIQval = 4.9-13.7), acceptable accuracy for POM (R^2val = 0.50 and RPIQval = 1.9) and poor results for the > 200-µm fraction (R^2val = 0.58 and RPIQval = 0.7).

The prediction of SOC amount in the POM fraction was poor in calibration (R^2cv = 0.74 and RPIQcv = 1.3) but accurate in validation (R^2val = 0.93 and RPIQval = 3.4), while the prediction of SOC amount in the > 200-µm fraction was poor in both calibration and validation (R^2cv and R^2val < 0.5, RPIQcv and RPIQval ≤ 0.8). The SOC amounts in the other fractions (< 200 µm) were accurately calibrated (R^2cv = 0.87-0.95, RPIQcv= 3.4-8.5) and validated (R^2val =0.78-0.95, RPIQval = 2.3-8.4).

3.3. Results of MIRS predictions

The results are presented in Table 2. The predictions of SOC content of unfractionated samples were excellent as they were also for NIRS (R^2cv = R^2val = 0.98, RPIQcv and RPIQval > 13). Calibration results of fraction masses based on MIRS data showed the same trend as for NIRS data. Fraction masses were accurately calibrated (R^2cv = 0.81-0.96, RPIQcv = 2.0-7.3) except for POM (R^2cv = 0.55 and RPIQcv = 1.0). External validation results for fraction masses also showed the same trend as for NIRS: excellent predictive ability for the 200-µm and < 20-µm fractions (R^2val = 0.95 and RPIQval = 6.6-7.3), but lower accuracy for the POM, 50-200-µm and 20-50-µm fractions (R^2val = 0.50-0.74 and RPIQval = 1.5-2.1).

Calibration results for fraction SOC concentrations based on MIRS data showed the same trend as for NIRS: excellent for all fractions < 200 µm (R^2cv = 0.85-0.94, RPIQcv = 3.2-5.7) but not for POM and the fraction > 200 µm (R^2cv = 0.24-0.40, RPIQcv = 1.1-1.6). External validation results for fraction SOC concentrations also showed the same trend as for NIRS data except for POM: excellent predictions for 50-200-µm, 20-50-µm and < 20-µm fractions (R^2val = 0.91-0.97, and RPIQval = 4.6-8.3), lower but still accurate for POM (R^2val = 0.60

and RPIQval = 2.2), and poor prediction for the > 200-µm fraction (R^2val = 0.59 and RPIQval = 0.7).

Calibration and validation results for fraction SOC amounts based on MIRS data showed the same trend as for NIRS data. Prediction of the SOC amount in POM was poor in calibration (R^2cv = 0.59 and RPIQcv = 1.1) but accurate in validation (R^2val = 0.77 and RPIQval = 2.2), while prediction of the SOC amount in the fraction > 200 µm was poor for both calibration and validation (R^2cv and R^2val < 0.4, RPIQcv and RPIQval ≤ 0.7). The other fractions (< 200 µm) were excellently calibrated (R^2cv = 0.85-0.95, RPIQcv = 3.4-8.2) and validated (R^2val = 0.75-0.95, RPIQval = 2.3-8.1).

4. Discussion

4.1. Prediction performance depending on the variable and possible explanations

The predictions were very good for SOC content in the unfractionated soil (RPIQ ≥ 13.9). Excellent NIRS, MIRS or visible and NIRS (VNIRS) predictions of total SOC content have already been reported for Madagascar by Rabenarivo et al. (2013; R^2val = 0.92-0.99 and RPDval = 3.3-8.6 for a multilocal set of topsoil samples, mostly from Ferralsols), Andriamananjara et al. (2016; R^2cv = 0.97 for a regional set of Ferralsols sampled at different depths) and Kawamura et al. (2017; R^2cv = 0.97 and RPDcv = 6.0 for a regional set of topsoils under rice). In the present study, the very good performances could be attributed to the spectral representativeness of calibration samples and the regression procedure. Indeed, calibrations based on spectral neighbours, such as LW-PLSR, have shown their ability to accurately predict soil properties, especially SOC content (Dangal et al., 2019; Fernández Pierna and Dardenne, 2008; Rabenarivo et al., 2013; Vasava and Das, 2022).

The predictions of SOC concentration and amount in all fractions < 200 µm were accurate using either NIRS or MIRS (RPIQ > 3 in most cases and always ≥ 2.0). This confirmed the results of other authors on NIRS prediction of particle-size fraction SOC from unfractionated soil spectra for samples originating from three sites in Burkina Faso and one region in Congo (Barthès et al., 2008a; R^2val = 0.70-0.95 after calibration on most spectrally representative samples; RPDval and RPIQval were not specified) or from 10 sites in Uruguay (Cozzolino and Morón, 2006; R^2cv = 0.85-0.90 and RPDcv = 2.0-3.2). This result also confirmed the result of Zimmermann et al. (2007) on a range of Swiss soils, where MIR spectra of bulk soil and PLS regression were appropriate to quantify SOC in different size and density fractions (RPDval = 2.0-4.1 after calibration on representative samples). In the present study, regardless of the variable considered, predictions were particularly accurate for the fraction

< 20 µm (RPIQ ≥ 4.6). The high proportion of the fraction < 20 µm in the samples, almost half of the soil mass on average over all sites, and the large amount of SOC it accounted for (22 gC fraction kg^{-1} soil on average) probably facilitated the predictions (Peng et al., 2014; Wight et al., 2016). In contrast, predictions were less accurate for SOC in the POM fraction (RPIQ = 1.1-3.4) and poor for SOC in the > 200 µm fraction (RPIQ = 0.4-1.4). Our results confirmed those of Jaconi et al. (2019) on a wide range of German soils, who also observed that NIRS and PLS regression yielded accurate predictions of the fine fraction SOC (0-63 µm; RPDval = 2.6 and RPIQval = 3.1 in independent validation) but less accurate results for the labile SOC fraction (sum of dissolved, particulate and coarse-aggregate included SOC; RPDval = 1.8, RPIQval = 1.0). Accurate predictions might be attributed to the presence of strong and well-defined absorption features of soil organic compounds in the NIR and MIR regions (Briedis et al., 2020; Reeves et al., 2006; Viscarra Rossel and Hicks, 2015). Calderón et al. (2011) studied MIR spectra of soil particle-size and -density fractions at four sites in the USA and observed that each fraction had distinct spectral features regardless of the site: PCA discriminated POM and the fraction < 2 µm clearly, and 2-20-µm and 20-53-µm fractions to a lesser extent.

In the present study, fractionation was carried out on 2-mm sieved samples, but NIR and MIR spectra were collected on 0.2-mm ground aliquots. Indeed, MIRS requires fine grinding (Barthès et al., 2016; Le Guillou et al., 2015); thus, unbiased comparison between NIRS and MIRS also requires fine grinding before NIR spectrum acquisition because NIRS prediction accuracy tends to decrease when samples are less finely prepared, especially for fine-textured soils (Brunet et al., 2007). Therefore, MIR and NIR spectra were acquired on samples where particles > 200 µm had been crushed; hence, poor predictions were often observed for fractions > 200 µm and POM. Nevertheless, the mass of the fraction > 200 µm was accurately predicted (RPIQ > 6 for both spectral ranges), possibly because the quartz particles it is mainly made of still had a specific spectral signature after 0.2-mm grinding. Indeed, Nduwamungu et al. (2009) reported that NIRS prediction of sand content on a 15-ha site was more accurate when spectra were acquired on more finely ground samples (from 2 to 0.2 mm; RPD > 6). On a wide range of Australian soils, Le Guillou et al. (2015) also observed better MIRS prediction of sand content when sample preparation was finer (from 2 to 0.106 mm; RPD > 2), and they explained that this might be due to the breaking of clay coatings on the larger particles due to fine grinding, so that quartz was more exposed.

In the present study, the fraction > 200 µm included little SOM (5.2 gC kg^{-1} fraction and 0.4 gC fraction kg^{-1} soil, on average), as most coarse organic particles were in the POM

fraction. POM is made of poorly decomposed plant residues, so its biochemical composition differs from that of denser fractions (Christensen, 1992), and its spectral signature also differs (Calderón et al., 2011). Indeed, although coarse-size organic particles were affected by 0.2-mm grinding before spectrum acquisition, the prediction of SOC concentration in POM was acceptable using both NIRS and MIRS (RPIQ = 1.6-2.2). However, the prediction of SOC amount in POM was sometimes poor , and more variable (RPIQ = 1.1-3.4), probably because the POM mass was not well predicted (RPIQ = 1.0-1.8) due to its low weight (1% of the sample mass on average) and POM heterogeneity (Angst et al., 2018; Soucémarianadin et al., 2019). It might be hypothesised that SOC had the same nature in the fraction > 200 µm as in POM, but for some reason, it had not been extracted by flotation. In addition to issues due to 0.2-mm grinding, very low SOC content in the fraction > 200 µm could explain poor prediction (Gupta et al., 2018). Indeed, scanning an area of a few tens of mm² could rarely capture the spectral signature of scarce and unevenly distributed organic particles.

The conventional determination of fraction SOC could also explain the poor predictions of SOC in the fraction > 200 µm: of the seven sites, the sum of the fraction SOC amounts was ≤ 75 % of the unfractionated soil SOC for the four sites with > 30 % coarse sands on average (Sakaraha, Lazaina, Belobaka and Faraony), whereas it was ≥ 86 % for the remaining three sites with < 5 % coarse sands on average (Alaotra Lake, Andranomanelatra and Andasy; cf. subsection 3.1 and Table SM1). Therefore, lower SOC recovery occurred in soils richer in coarse sands, where SOC in the fraction > 200 µm was higher and contributed more to SOC in the unfractionated soil: on average per site, 0.3 to 1.6 gC fraction kg^{-1} soil (representing 1.8 to 6.2 % of SOC in the unfractionated soil) in the four sites with high coarse sand content vs. 0.1 to 0.3 gC fraction kg^{-1} soil (representing 0.3 to 0.7 % of SOC in the unfractionated soil) in the three sites with low coarse sand content. This strongly suggested that the fraction > 200 µm was particularly affected by low SOC recovery, so conventional SOC data were moderately reliable for this fraction, which was an obstacle for accurate NIRS or MIRS predictions. Furthermore, accurate predictions of the mass of the fraction > 200 µm suggested that reliability issues for conventional SOC data in this fraction resulted from SOC analysis rather than from the physical fractionation procedure. Indeed, sampling representative aliquots of a few tens of mg (for dry combustion analysis of C) is difficult in fractions that include scarce and unevenly distributed organic particles.

In short, predictions of SOC concentrations and amounts were more accurate for fractions < 200 µm, especially for the fraction < 20 µm due to its high contribution to total SOC. Predictions were less accurate for POM and that fraction > 200 µm, as 0.2-mm grinding

before spectrum acquisition complicated predictions on these coarse fractions, and possibly due to the poor reliability of conventional data.

4.2. Was NIRS or MIRS better suited for predicting SOC particle-size and -density distribution?

When considering all fractions and variables (i.e., masses, SOC concentrations and amounts, and SOC content in unfractionated soil), NIRS outperformed MIRS in all but one case, according to RPIQcv and RPIQval. Better MIRS than NIRS prediction was only achieved for the SOC concentration in POM. However, the benefit of NIRS over MIRS for predicting fraction masses and SOC amounts was limited in general, except for the mass of the fraction > 200 µm and the SOC amount of POM (two fractions affected by 0.2-mm grinding before spectrum acquisition). The benefit of NIRS was larger for predicting fraction SOC concentrations, especially for the three fractions < 200 µm (50-200-µm, 20-50-µm and < 20-µm fractions).

Most studies that compared NIRS and MIRS predictions of soil properties using spectra that were obtained on dry, ground samples reported better MIRS than NIRS predictions, especially for SOC content. This has been attributed to better spectral information in the MIR, especially because MIR peaks are better resolved (Reeves, 2010; Soriano-Disla et al., 2014). The fact that NIR would be more sensitive than MIR to instrumental errors has also been mentioned (Bellon-Maurel and McBratney, 2011). However, some studies, especially on tropical soils, have reported better NIRS than MIRS predictions of soil organic properties. Rabenarivo et al. (2013), who reported such results for Malagasy soils, suggested that the possible superiority of NIRS for predicting organic properties in tropical soils could result from the abundance of some minerals and the noise it might cause in the MIR spectra. Madari et al. (2006), who studied Brazilian soils, also reported better MIRS than NIRS prediction of SOC but did not explain it. Greenberg et al. (2021), for a German site, studied physicochemical SOC fractions and compared MIRS vs. VNIRS predictions using spectra acquired in field (on the soil surface) or laboratory conditions (ground dry samples): laboratory spectra yielded better MIRS than VNIRS predictions for the total and stabilised SOC, but the opposite was observed for labile and resistant SOC and for clay, silt and sand fractions; Field spectra always yielded better VNIRS than MIRS predictions. Therefore, generally, it cannot be considered that MIRS allows better predictions of soil properties than NIRS and VNIRS, especially in soils from tropical regions but even in soils from temperate regions.

In short, NIRS slightly outperformed MIRS in all but one case probably due to the abundance of some minerals and the noise it might cause in the MIR spectra.

5. Conclusions

The use of NIRS and MIRS was tested for quantifying the distribution of SOC in particle-size and -density fractions of Malagasy topsoils, using spectra from bulk soil (< 2 mm) that was ground to 0.2 mm. The results showed that both spectral ranges yielded accurate predictions of the mass (g 100 g^{-1} soil), SOC concentration (gC kg^{-1} fraction) and SOC amount (gC fraction kg^{-1} soil), especially for the fractions < 200 µm, and among them, for the fraction < 20 µm (which included more SOC than other fractions). Moreover, NIRS tended to outperform MIRS, as already reported for sesquioxide-rich soils.

Such results open exciting perspectives, because characterizing the SOC particle-size and -density distribution is a very valuable approach for studying SOM and its dynamics, but could be very tedious to experimentally achieve in such well aggregated tropical soils. Therefore, IR spectroscopy offers the possibility of developing a time- and cost-effective approach for greatly improving studies on SOM. Moreover, including existing results on SOM fractionation into calibration databases (along with the spectra of corresponding bulk soils) represents an important valorisation: indeed, the fractionation performed on a given sample no longer provides information on this sample only but allows characterizing new samples based on their spectrum and calibration. Therefore, infrared spectroscopy allows leveraging the value of conventional analyses.

The present study paves the way to high-throughput quantification of SOM particle-size and-density, especially in Madagascar, where the data already available represent the emergence of a calibration database. To become more representative of Malagasy soils, this database must be enriched with data from other sites, especially coarse-textured soils. Then, the Malagasy spectral database on SOM particle-size and -density distribution could be integrated into larger spectral databases, for instance, at the initiative of the Global Soil Laboratory Network (GLOSOLAN, established under the FAO Global Soil Partnership; https://www.fao.org/global-soil-partnership/glosolan/en/) or the SoilSpec4GG network (Soil Spectroscopy for Global Good, funded by the USDA; https://soilspectroscopy.org/)

Acknowledgements

This work was supported by the French Facility for Global Environment (FFEM) [RIME-PAMPA; grant number: CZZ 3076], the International Foundation of Science; and the BNP

Paribas Foundation [SoCa; Climate Initiative program]. The authors are deeply grateful to the NGO TAFA [TAny sy FAmpandrosoana] for access to experimental designs and permission for soil collection, and to Fela-Nirina Rakotondrasolo and Jocelin Rakotondramanana for technical assistance in soil fractionation. Finally, the helpful comments by anonymous referees are gratefully acknowledged.

References

Amelung, W., Bossio, D., de Vries, W., Kögel-Knabner, I., Lehmann, J., Amundson, R., Bol, R., Collins, C., Lal, R., Leifeld, J., Minasny, B., Pan, G., Paustian, K., Rumpel, C., Sanderman, J., van Groenigen, J.W., Mooney, S., van Wesemael, B., Wander, M., Chabbi, A., 2020. Towards a global-scale soil climate mitigation strategy. Nat. Commun. 11, 5427. https://doi.org/10.1038/s41467-020-18887-7

Andriamananjara, A., Hewson, J., Razakamanarivo, H., Andrisoa, R.H., Ranaivoson, N., Ramboatiana, N., Razafindrakoto, M., Ramifehiarivo, N., Razafimanantsoa, M.-P., Rabeharisoa, L., Ramananantoandro, T., Rasolohery, A., Rabetokotany, N., Razafimbelo, T., 2016. Land cover impacts on aboveground and soil carbon stocks in Malagasy rainforest. Agric. Ecosyst. Environ. 233, 1–15. https://doi.org/10.1016/j.agee.2016.08.030

Angst, G., Messinger, J., Greiner, M., Häusler, W., Hertel, D., Kirfel, K., Kögel-Knabner, I., Leuschner, C., Rethemeyer, J., Mueller, C.W., 2018. Soil organic carbon stocks in topsoil and subsoil controlled by parent material, carbon input in the rhizosphere, and microbial-derived compounds. Soil Biol. Biochem. 122, 19–30. https://doi.org/10.1016/j.soilbio.2018.03.026

Balesdent, J., 1987. The turnover of soil organic fractions estimated by radiocarbon dating. Science of The Total Environment 62, 405-408.

Barthès, B.G., Brunet, D., Hien, E., Enjalric, F., Conche, S., Freschet, G.T., d'Annunzio, R., Toucet-Louri, J., 2008a. Determining the distributions of soil carbon and nitrogen in particle size fractions using near-infrared reflectance spectrum of bulk soil samples. Soil Biol. Biochem. 40, 1533–1537. https://doi.org/10.1016/j.soilbio.2007.12.023

Barthès, B.G., Kouakoua, E., Larré-Larrouy, M.-C., Razafimbelo, T.M., Luca, E.F. de, Azontonde, A., Neves, C.S.V.J., Freitas, P.L. de, Feller, C.L., 2008b. Texture and sesquioxide effects on water-stable aggregates and organic matter in some tropical soils. Geoderma 1–2, 14–25. https://doi.org/10.1016/j.geoderma.2007.10.003

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. J. Near Infrared Spectrosc. 24, 199–214. https://doi.org/10.1255/jnirs.1232

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. https://doi.org/10.1016/j.trac.2010.05.006

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

Boysworth, M.K., Booksh, K.S., 2008. Aspects of multivariate calibration applied to near-infrared spectroscopy, in: Burns, D.A., Ciurczak, E.W. (Eds.), Handbook of Near-Infrared Analysis (third ed.). CRC Press, Boca Raton, Florida, pp. 207–229.

Briedis, C., Baldock, J., de Moraes Sá, J.C., dos Santos, J.B., Milori, D.M.B.P., 2020. Strategies to improve the prediction of bulk soil and fraction organic carbon in Brazilian samples by using an Australian national mid-infrared spectral library. Geoderma 373, 114401. https://doi.org/10.1016/j.geoderma.2020.114401

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. https://doi.org/10.1016/j.geoderma.2007.01.007

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. https://doi.org/10.1016/j.biosystemseng.2008.04.008

Calderón, F.J., Reeves, J.B. III, Collins, H.P., Paul, E.A. 2011. Chemical differences in soil organic matter fractions determined by diffuse-reflectance mid-infrared spectroscopy. Soil Sci. Soc. Am. J. 75, 568–579. https://doi.org/10.2136/sssaj2009.0375

Cambou, A., Barthès, B.G., Moulin, P., Chauvin, L., Faye, E.H., Masse, D., Chevallier, T., Chapuis-Lardy, L., 2022. Prediction of soil carbon and nitrogen contents using visible and near infrared diffuse reflectance spectroscopy in varying salt-affected soils in Sine Saloum (Senegal). Catena 212, 106075. https://doi.org/10.1016/j.catena.2022.106075

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. https://doi.org/10.2136/sssaj2001.652480x

Chenu, C., Rumpel, C., Lehmann, J., 2015. Methods for studying soil organic matter: nature, dynamics, spatial accessibility, and interactions with minerals, in: Paul, E.A. (Ed.), Soil Microbiology, Ecology and Biochemistry (fourth ed.), Academic Press, Cambridge pp. 383–419. https://doi.org/10.1016/B978-0-12-415955-6.00013-X

Christensen, B.T., 1992. Physical fractionation of soil and organic matter in primary particle size and density separates, Stewart, B.A. (Eds.), Advances in Soil Science, volume 20. Springer, New York, pp. 1–90. https://doi.org/10.1007/978-1-4612-2930-8_1

Cozzolino, D., Morón, A., 2006. Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions. Soil Tillage Res. 85, 78–85. https://doi.org/10.1016/j.still.2004.12.006

d'Annunzio, R., Conche, S., Landais, D., Saint-André, L., Joffre, R., Barthès, B.G., 2008. Pairwise comparison of soil organic particle-size distributions in native savannas and Eucalyptus plantations in Congo. For. Ecol. Manag. 255, 1050–1056. https://doi.org/10.1016/j.foreco.2007.10.027

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

Ding, F., Huang, Y., Sun, W., Jiang, G., Chen, Y., 2014. Decomposition of organic carbon in fine soil particles is likely more sensitive to warming than in coarse particles: an incubation study with temperate grassland and forest soils in northern China. PLoS ONE 9, e95348. https://doi.org/10.1371/journal.pone.0095348

Dynarski, K.A., Bossio, D.A., Scow, K.M., 2020. Dynamic stability of soil carbon: reassessing the "permanence" of soil carbon sequestration. Front. Environ. Sci. 8, 514701. https://doi.org/10.3389/fenvs.2020.514701

Feller, C., 1979. Une méthode de fractionnement granulométrique de la matière organique des sols. Application aux sols tropicaux, à textures grossières, très pauvres en humus. Cahiers ORSTOM série Pédologie 17, 339-346.

Feller, C., Beare, M.H., 1997. Physical control of soil organic matter dynamics in the tropics. Geoderma 79, 69–116. https://doi.org/10.1016/S0016-7061(97)00039-6

Feng, W., Plante, A., Six, J., 2013. Improving estimates of maximal organic carbon stabilization by fine soil particles. Biogeochemistry 112, 81-93. https://doi.org/10.1007/s10533-011-9679-7

Fernández Pierna, J., Dardenne, P., 2008. Soil parameter quantification by NIRS as a Chemometric challenge at '*Chimiométrie 2006*'. Chemometr. Intell. Lab. Syst. 91, 94–98. https://doi.org/ doi:10.1016/j.chemolab.2007.06.007

Fujisaki, K., Chapuis-Lardy, L., Albrecht, A., Razafimbelo, T., Chotte, J.-L., Chevallier, T., 2018. Data synthesis of carbon distribution in particle size fractions of tropical soils: Implications for soil carbon storage potential in croplands. Geoderma 313, 41–51. https://doi.org/10.1016/j.geoderma.2017.10.010

Gavinelli, E., Feller, C., Larré-Larrouy, M.C., Bacye, B., Djegui, N., Nzila, J. de D., 1995. A routine method to study soil organic matter by particle-size fractionation: examples for tropical soils. Commun. Soil Sci. Plant Anal. 26, 1749–1760. https://doi.org/10.1080/00103629509369406

Ge, Y., Thomasson, J.A., Morgan, C.L.S., 2014. Mid-infrared attenuated total reflectance spectroscopy for soil carbon and particle size determination. Geoderma 213, 57–63. https://doi.org/10.1016/j.geoderma.2013.07.017

Gee, G.W., Bauder, J.W. 1986. Particle-size analysis, in: Klute A. (Ed.) methods of soil analysis Part 1. Soil science society of America book series 5, Madison, Wisconsin, pp 383-411

Gnacadja, L., Wiese, L., 2016. Land degradation neutrality: will Africa achieve it? Institutional solutions to land degradation and restoration in Africa, in: Lal, R., Kraybill, D., Hansen, D.O., Singh, B.R., Mosogoya, T., Eik, L.O. (Eds.), Climate change and multi-dimensional sustainability in African agriculture: climate change and sustainability in agriculture. Springer International Publishing, Cham, pp. 61–95. https://doi.org/10.1007/978-3-319-41238-2_5

Grandière, I., Razafimbelo, T., Barthès, B., Blanchart, E., Louri, J., Ferrer, H., Chenu, C., Wolf, N., Albrecht, A., & Feller, C. (2007). Distribution granulo-densimétrique de la matière organique dans un sol argileux sous semis direct avec couverture végétale des Hautes Terres malgaches. Etude et Gestion Des Sols, 14(2), 117–133. http://afes.fr/afes/egs/EGS_14_2_grandiere.pdf

Greenberg, I., Seidel, M., Vohland, M., Ludwig, B., 2021. Performance of field-scale lab vs in situ visible/near- and mid-infrared spectroscopy for estimation of soil properties. Eur. J. Soil Sci. 73, e13180. https://doi.org/10.1111/ejss.13180

Gupta, A., Vasava, H.B., Das, B.S., Choubey, A.K., 2018. Local modeling approaches for estimating soil properties in selected Indian soils using diffuse reflectance data over visible to near-infrared region. Geoderma 325, 59–71. https://doi.org/10.1016/j.geoderma.2018.03.025

Hassink, J., Whitmore, A.P., Kubát, J., 1997. Size and density fractionation of soil organic matter and the physical capacity of soils to protect organic matter. Eur. J. Agron. 1–3, 189–199. https://doi.org/10.1016/S1161-0301(97)00045-2

IUSS Working Group WRB, 2015. World Reference Base for Soil Resources 2014, Update 2015. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106, Rome: FAO.

Jaconi, A., Poeplau, C., Ramirez-Lopez, L., Van Wesemael, B., Don, A., 2019. Log-ratio transformation is the key to determining soil organic carbon fractions with near-infrared spectroscopy. Eur. J. Soil Sci. 70, 127–139. https://doi.org/10.1111/ejss.12761

Kawamura, K., Tsujimoto, Y., Rabenarivo, M., Asai, H., Andriamananjara, A., Rakotoson, T., 2017. Vis-NIR Spectroscopy and PLS regression with waveband selection for estimating the total C and N of paddy soils in Madagascar. Remote Sens. 9, 1081. https://doi.org/10.3390/rs9101081

Kennard, R.W., Stone, L.A., 1969. Computer aided design of experiments. Technometrics 11, 137–148. https://doi.org/10.1080/00401706.1969.10490666

Knox, N.M., Grunwald, S., McDowell, M.L., Bruland, G.L., Myers, D.B., Harris, W.G., 2015. Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. Geoderma 239–240, 229–239. https://doi.org/10.1016/j.geoderma.2014.10.019

Lal, R., 2009. Soil degradation as a reason for inadequate human nutrition. Food Secur. 1, 45–57. https://doi.org/10.1007/s12571-009-0009-z

Lavallee, J.M., Soong, J.L., Cotrufo, M.F., 2020. Conceptualizing soil organic matter into particulate and mineral-associated forms to address global change in the 21st century. Glob. Change Biol. 26, 261–273. https://doi.org/10.1111/gcb.14859

Le Guillou, F., Wetterlind, W., Viscarra Rossel, R., Hicks, W., Grundy, M., Tuomi, S., 2015. How does grinding affect the mid-infrared spectra of soil and their multivariate calibrations to texture and organic carbon? Soil Res. 53, 913–921. https://doi.org/10.1071/SR15019

Lesnoff, M., Metz, M., Roger, J.-M., 2020. Comparison of locally weighted PLS strategies for regression and discrimination on agronomic NIR data. J. Chemom. 34, e3209. https://doi.org/10.1002/cem.3209

Lehmann, J., Hansel, C.M., Kaiser, C., Kleber, M., Maher, K., Manzoni, S., Nunan, N., Reichstein, M., Schimel, J.P., Torn, M.S., Wieder, W.R., Kögel-Knabner, I., 2020. Persistence of soil organic carbon caused by functional complexity. Nat. Geosci. 13, 529–534. https://doi.org/10.1038/s41561-020-0612-3

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. https://doi.org/10.1016/j.geoderma.2006.03.026

Madhavan, D.B., Baldock, J.A., Read, Z.J., Murphy, S.C., Cunningham, S.C., Perring, M.P., Herrmann, T., Lewis, T., Cavagnaro, T.R., England, J.R., Paul, K.I., Weston, C.J., Baker, T.G., 2017. Rapid prediction of particulate, humus and resistant fractions of soil organic carbon in reforested lands using infrared spectroscopy. J. Environ. Manage. 193, 290–299. https://doi.org/10.1016/j.jenvman.2017.02.013

Malou, O.P., Sebag, D., Moulin, P., Chevallier, T., Badiane-Ndour, N.Y., Thiam, A., Chapuis-Lardy, L., 2020. The Rock-Eval® signature of soil organic carbon in arenosols of the Senegalese groundnut basin. How do agricultural practices matter? Agric. Ecosyst. Environ. 301, 107030. https://doi.org/10.1016/j.agee.2020.107030

Meurer, K.H.E., Chenu, C., Coucheney, E., Herrmann, A.M., Keller, T., Kätterer, T., Nimblad Svensson, D., Jarvis, N., 2020. Modelling dynamic interactions between soil structure and the storage and turnover of soil organic matter. Biogeosciences 17, 5025–5042. https://doi.org/10.5194/bg-17-5025-2020

Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.-S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.-C., Vågen, T.-G., van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. Geoderma 292, 59–86. https://doi.org/10.1016/j.geoderma.2017.01.002

Nduwamungu, C., Ziadi, N., Tremblay, G.F., Parent, L.-É., 2009. Near-infrared reflectance spectroscopy prediction of soil properties: effects of sample cups and preparation. Soil Sci. Soc. Am. J. 73, 1896–1903. https://doi.org/10.2136/sssaj2008.0213

Osland, M.J., Gabler, C.A., Grace, J.B., Day, R.H., McCoy, M.L., McLeod, J.L., From, A.S., Enwright, N.M., Feher, L.C., Stagg, C.L., Hartley, S.B., 2018. Climate and plant controls on soil organic matter in coastal wetlands. Glob. Change Biol. 24, 5361–5379. https://doi.org/10.1111/gcb.14376

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G.P., Smith, P., 2016. Climate-smart soils. Nature 532, 49–57. https://doi.org/10.1038/nature17174

Parent, E.J., Parent, S.-É., Parent, L.E., 2021. Determining soil particle-size distribution from infrared spectra using machine learning predictions: Methodology and modeling. PLoS ONE 16, e0233242. https://doi.org/10.1371/journal.pone.0233242

Peng, Y., Knadel, M., Gislum, R., Schelde, K., Thomsen, A., Greve, M.H., 2014. Quantification of SOC and clay content using visible near-infrared reflectance–mid-infrared reflectance spectroscopy with jack-knifing partial least squares regression. Soil Sci. 179, 325–332. https://doi.org/10.1097/SS.0000000000000074

Poeplau, C., Don, A., Six, J., Kaiser, M., Benbi, D., Chenu, C., Cotrufo, M.F., Derrien, D., Gioacchini, P., Grand, S., Gregorich, E., Griepentrog, M., Gunina, A., Haddix, M., Kuzyakov, Y., Kühnel, A., Macdonald, L.M., Soong, J., Trigalet, S., Vermeire, M.-L., Rovira, P., van Wesemael, B., Wiesmeier, M., Yeasmin, S., Yevdokimov, I., Nieder, R., 2018. Isolating organic carbon fractions with varying turnover rates in temperate agricultural soils – A comprehensive method comparison. Soil Biol. Biochem. 125, 10–26. https://doi.org/10.1016/j.soilbio.2018.06.025

Prăvălie, R., Nita, I.-A., Patriche, C., Niculiță, M., Birsan, M.-V., Roșca, B., Bandoc, G., 2021. Global changes in soil organic carbon and implications for land degradation neutrality and climate stability. Environ. Res. 201, 111580. https://doi.org/10.1016/j.envres.2021.111580

Puget, P., Chenu, C., Balesdent, J., 2000. Dynamics of soil organic matter associated with particle-size fractions of water-stable aggregates. Eur. J. Soil Sci. 51, 595–605. https://doi.org/10.1111/j.1365-2389.2000.00353.x

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. https://doi.org/10.1255/jnirs.1080

Razafimbelo, T., Barthès, B.G., Larré-Larrouy, M.-C., De Luca, E.F.., Laurent, J.-Y., Cerri, C.C., Feller, C., 2006. Effect of sugarcane residue management (mulching versus burning)

on organic matter in a clayey Oxisol from southern Brazil. Agric. Ecosyst. Environ. 115, 285–289. https://doi.org/10.1016/j.agee.2005.12.014

Razafimbelo, T., Chevallier, T., Albrecht, A., Chapuis-Lardy, L., Rakotondrasolo, F.N., Michellon, R., Rabeharisoa, L., Bernoux, M., 2013. Texture and organic carbon contents do not impact amount of carbon protected in Malagasy soils. Sci. Agric. 70, 204–208. https://doi.org/10.1590/S0103-90162013000300009

Razafimbelo, T., Albrecht, A., Oliver, R., Chevallier, T., Chapuis-Lardy, L., Feller, C., 2008. Aggregate associated-C and physical protection in a tropical clayey soil under Malagasy conventional and no-tillage systems. Soil Tillage Res. 98, 140–149. https://doi.org/10.1016/j.still.2007.10.012

Reeves, J.B. III, Follett, R.F., McCarty, G.W., Kimble, J.M., 2006. Can near or mid-infrared diffuse reflectance spectroscopy be used to determine soil carbon pools? Commun. Soil Sci. Plant Anal. 37, 2307–2325. https://doi.org/10.1080/00103620600819461

Reeves, J.B. III, 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. https://doi.org/10.1016/j.geoderma.2009.04.005

Sanderman, J., Baldock, J.A., Dangal, S.R.S., Ludwig, S., Potter, S., Rivard, C., Savage, K., 2021. Soil organic carbon fractions in the Great Plains of the United States: an application of mid-infrared spectroscopy. Biogeochemistry 156, 97–114. https://doi.org/10.1007/s10533-021-00755-1

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639. https://doi.org/10.1021/ac60214a047

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. https://doi.org/10.1080/05704928.2013.811081

Soucémarianadin, L., Cécillon, L., Chenu, C., Baudin, F., Nicolas, M., Girardin, C., Delahaie, A., Barré, P., 2019. Heterogeneity of the chemical composition and thermal stability of particulate organic matter in French forest soils. Geoderma 342, 65–74. https://doi.org/10.1016/j.geoderma.2019.02.008

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. https://doi.org/10.1371/journal.pone.0066409

Stevens, A.; Ramirez-Lopez,2020.Package 'Prospectr'. R Package Version 2020. Available online: https://cran.r-project.org/web/packages/prospectr/prospectr.pdf

Swinton, S.M., Lupi, F., Robertson, G.P., Hamilton, S.K., 2007. Ecosystem services and agriculture: Cultivating agricultural ecosystems for diverse benefits. Ecol. Econ. 64, 245–252. https://doi.org/10.1016/j.ecolecon.2007.09.020

Vargas, R., Paz, F., de Jong, B., 2013. Quantification of forest degradation and belowground carbon dynamics: ongoing challenges for monitoring, reporting and verification activities for REDD+. Carbon Manag. 4, 579–582. https://doi.org/10.4155/cmt.13.63

Vasava, H.B., Das, B.S., 2022. Assessment of soil properties using spectral signatures of bulk soils and their aggregate size fractions. Geoderma 417, 115837. https://doi.org/10.1016/j.geoderma.2022.115837

Viscarra Rossel, R., Hicks, W.S., 2015. Soil organic carbon and its fractions estimated by visible–near infrared transfer functions. Eur. J. Soil Sci. 66, 438–450. https://doi.org/10.1111/ejss.12237

Von Lützow, M., Kögel-Knabner, I., Ekschmitt, K., Flessa, H., Guggenberger, G., Matzner, E., Marschner, B., 2007. SOM fractionation methods: relevance to functional pools and to stabilization mechanisms. Soil Biol. Biochem. 39, 2183–2207. https://doi.org/10.1016/j.soilbio.2007.03.007

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H.-J., Kögel-Knabner, I., 2019. Soil organic carbon storage as a key function of soils - A review of drivers and indicators at various scales. Geoderma 333, 149–162. https://doi.org/10.1016/j.geoderma.2018.07.026

Wight, J.P., Ashworth, A.J., Allen, F.L., 2016. Organic substrate, clay type, texture, and water influence on NIR carbon measurements. Geoderma 261, 36–43. https://doi.org/10.1016/j.geoderma.2015.06.021

Wood, S.A., Sokol, N., Bell, C.W., Bradford, M.A., Naeem, S., Wallenstein, M.D., Palm, C.A., 2016. Opposing effects of different soil organic matter fractions on crop yields. Ecol. Appl. 26, 2072–2085. https://doi.org/10.1890/16-0024.1

Zimmermann, M., Leifeld, J., Fuhrer, J., 2007. Quantifying soil organic carbon fractions by infrared-spectroscopy. Soil Biol. Biochem. 39, 224–231. https://doi.org/10.1016/j.soilbio.2006.07.010

Figure 1. Map of site locations, where (i) numbers in brackets are the average per-site percentages of clay-silt-sand, (ii) the underlined number is the number of samples for a given site, and (iii) sites marked with * correspond to pot experiments, while sites marked with " correspond to experimental plots.

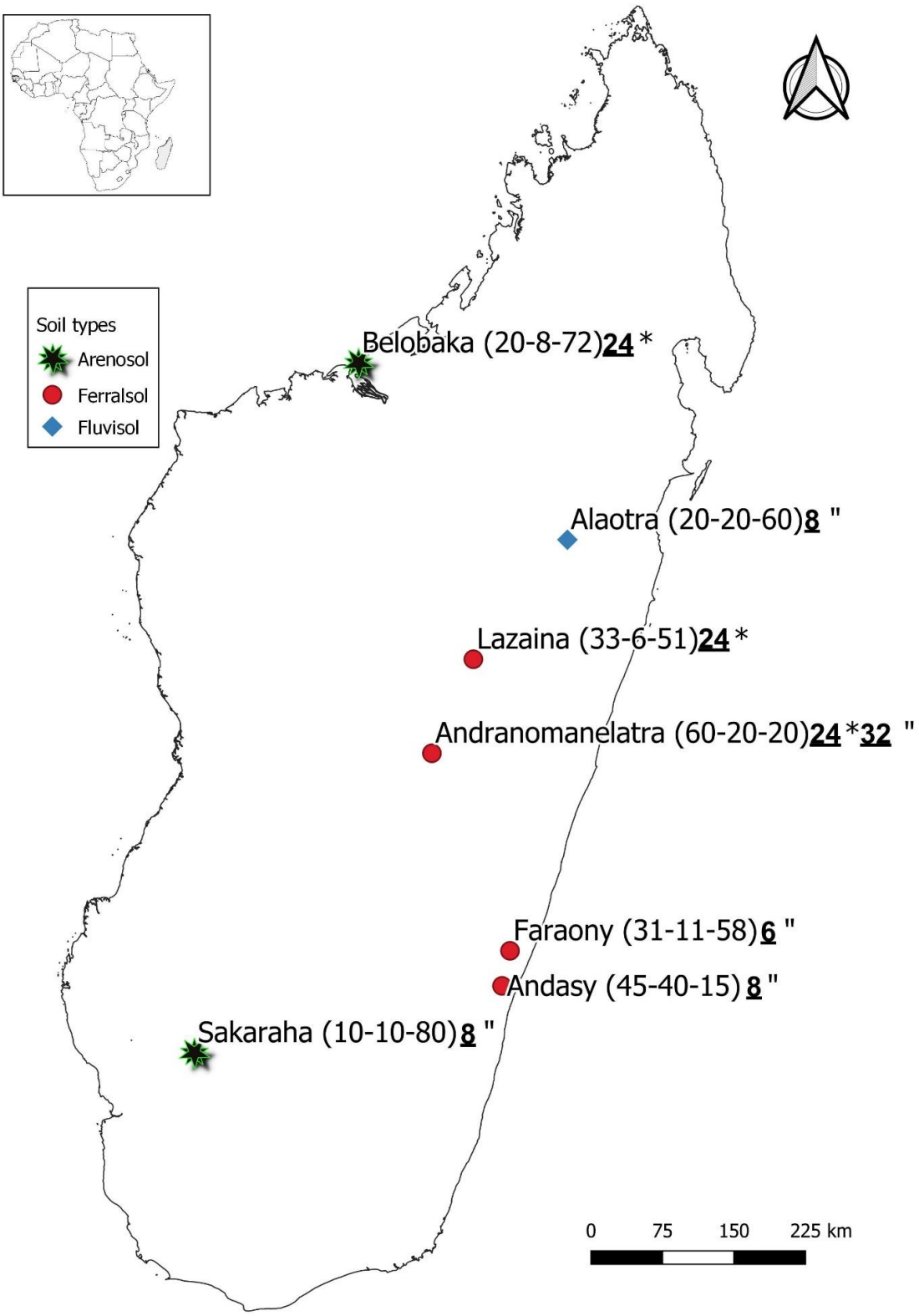


Figure 2. Principal component analysis (two first components) of NIR spectra (after spectrum centring and smoothing).

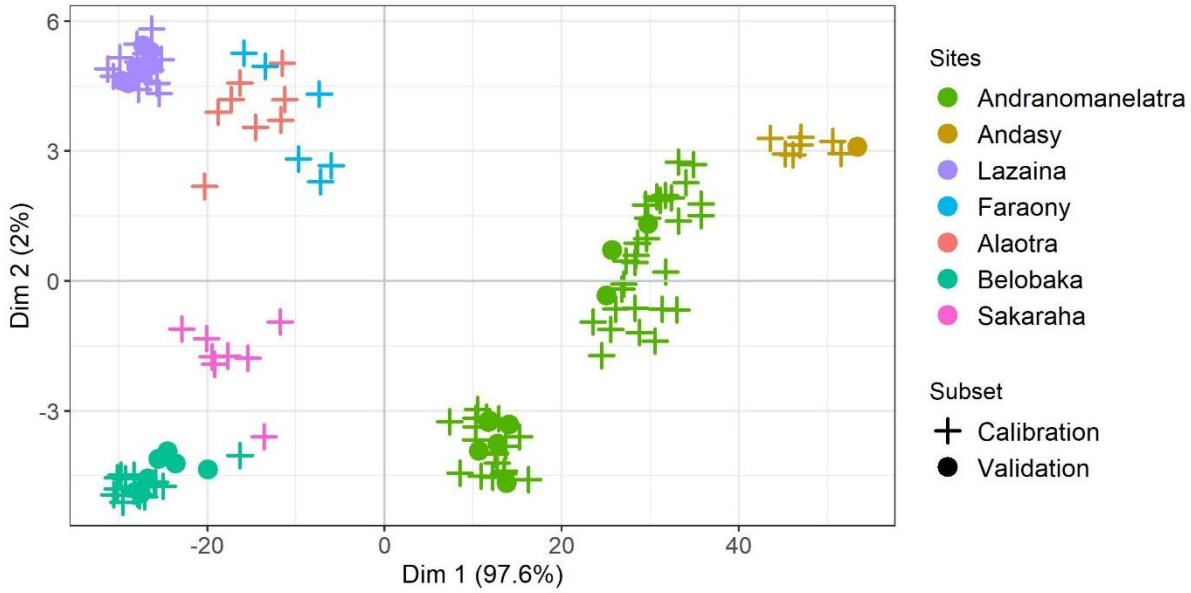


Figure 3. Principal component analysis (two first components) of MIR spectra (after spectrum centring and smoothing).

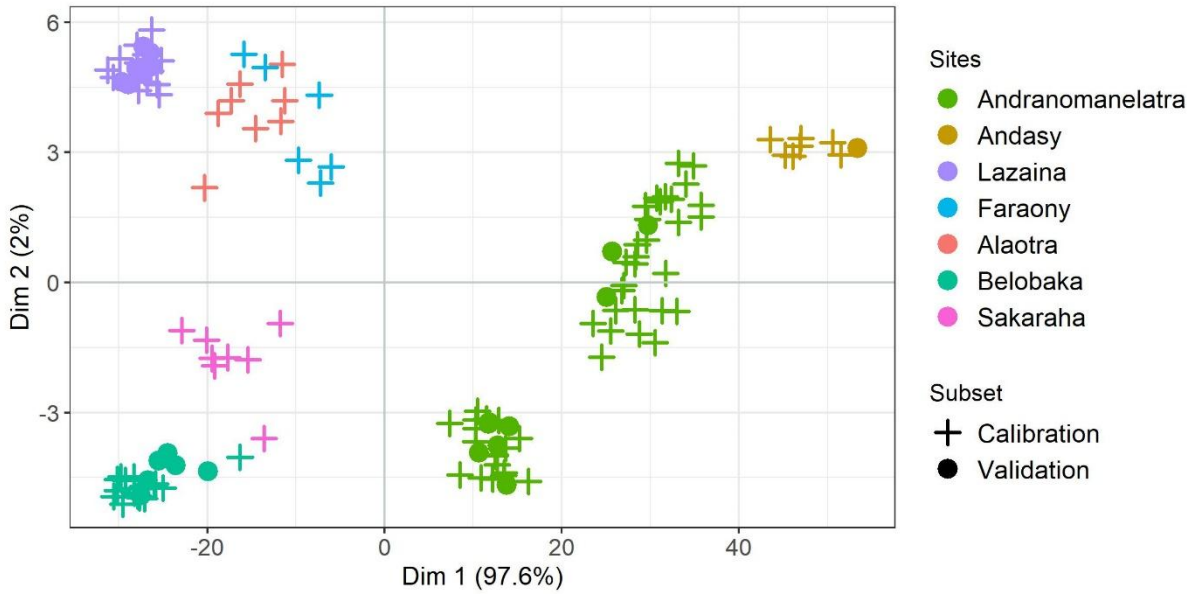


Figure 4. Average per-site NIR and MIR spectra after SNV pretreatment.

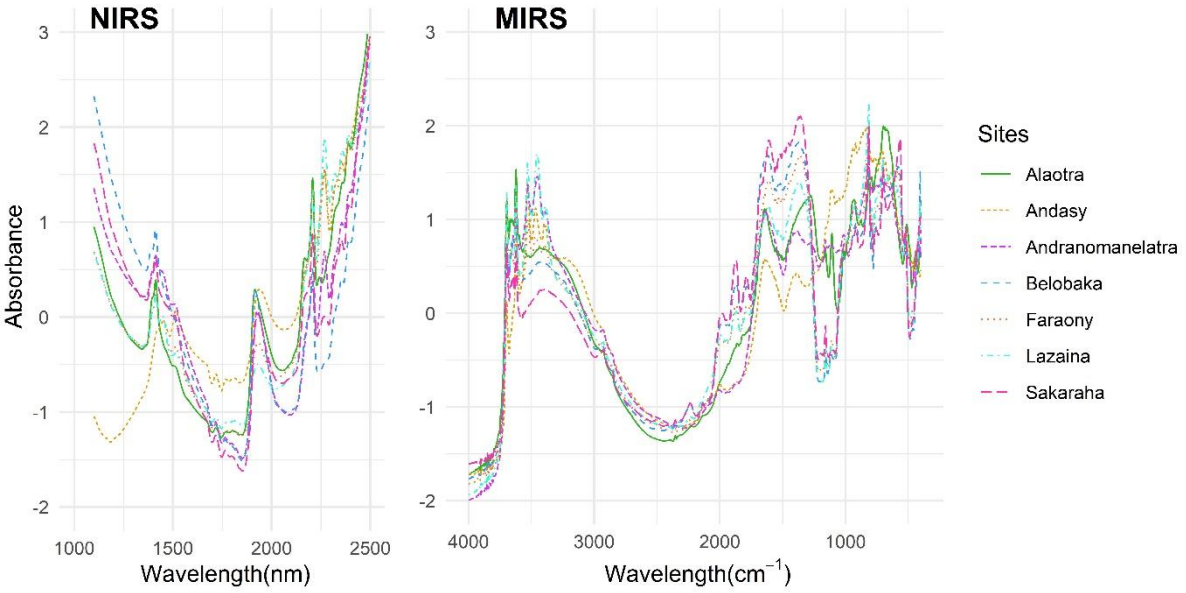


Figure 5. Distribution of SOC content in the unfractionated soil (< 2 mm).

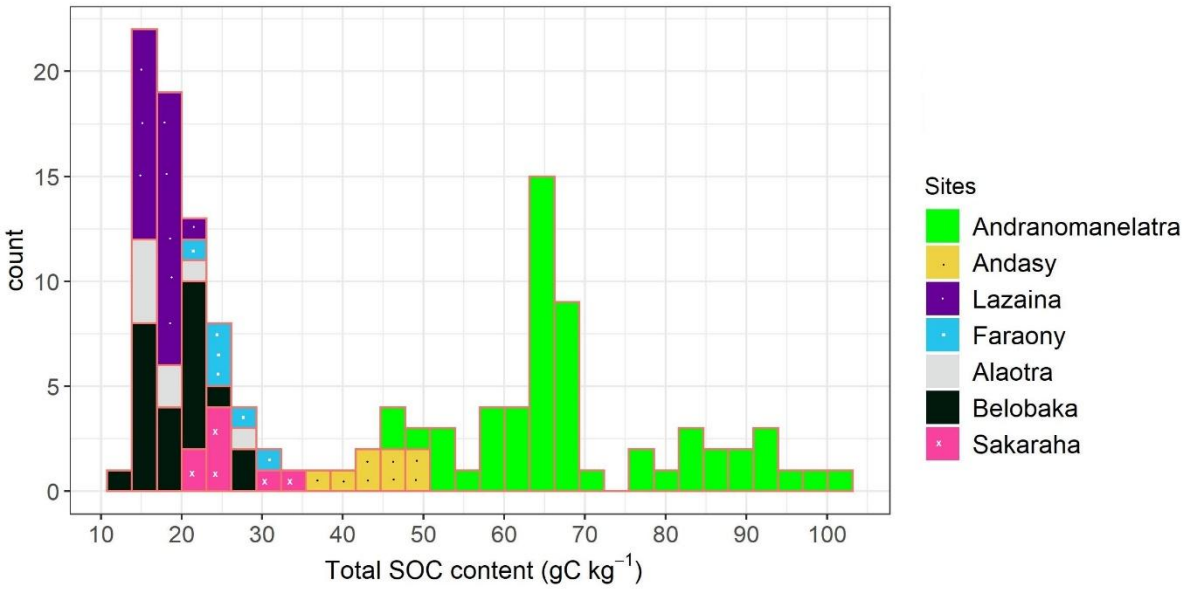


Figure 6. Per-site distribution of (A) average fraction mass, (B) average fraction SOC concentration, and (C) average fraction SOC amount.

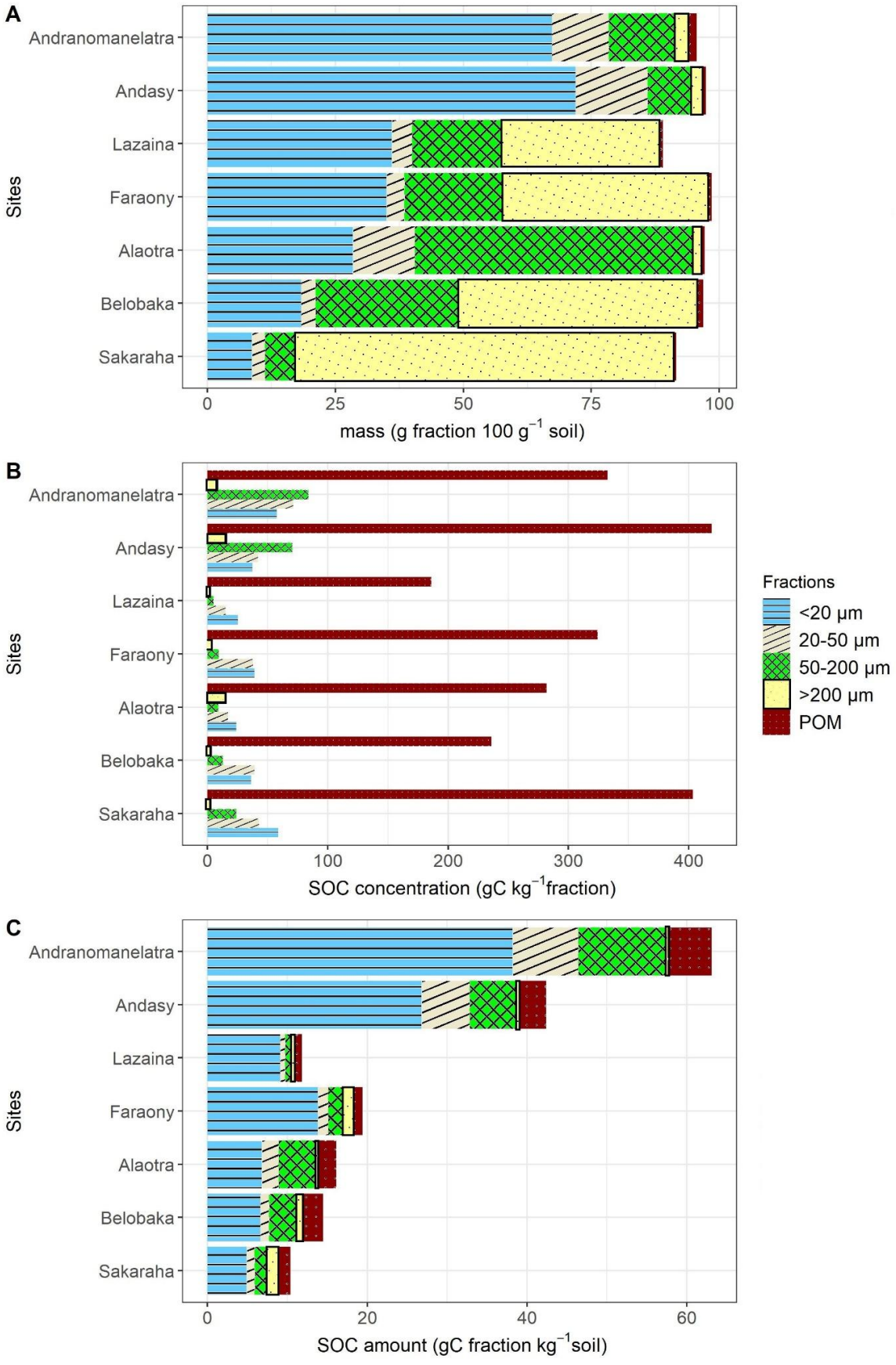


Table 1. Origin and characteristics of the soil sample set (0-5 cm topsoil).

Site No.	Site location (region & place)	Elevation (m)	Mean annual rainfall and temperature	Soil type[a]	Average clay-silt-sand (%)	Experiment modalities[b]	Number of samples
1	Antsirabe - Andranomanelatra 19°46' S, 47°06' E	1650	1600 mm 16°C	Ferralsol	60-20-20	Plot design Crops (soybean or rice) in CT and NT treatments	32
						Pot experiment Six modalities of residue restitutions: Control, Roots, Mulch (Rice, Soybean), Mixed (Rice, Soybean)	24
2	Manakara - Andasy 22°12' S, 47°50' E	50	2500 mm 23°C	Ferralsol	45-40-15	Plot design Crops (rice with S. *ruziziensis*) in NT treatments	8
3	Antananarivo - Lazaina 18°47' S, 47°32' E		1300 mm 18°C	Ferralsol	33-6-51	Pot experiment Six modalities of residue restitutions: Control, Roots, Mulch (Rice, Soybean), Mixed (Rice, Soybean)	24
4	Manakara – Faraony 21°50' S, 47°55' E	42	2500 mm 23°C	Ferralsol	31-11-58	Plot design Crops (rice with *S. ruziziensis*) in NT treatments	6
5	– Marololo- Alaotra Lake 17°32' S, 48°31' E	770	1200 mm 20°C	Fluvisol	20-20-60	Plot design Crops (rice with *S. ruziziensis*) in NT treatments	8
6	Mahajanga - Belobaka 15°41' S 46°20' E	12	1500 mm 27°C	Arenosol	20-8-72	Pot experiment Six modalities of residue restitutions: Control, Roots, Mulch (Rice, Soybean), Mixed (Rice, Soybean)	24
7	Tulear - Sakaraha 22°54' S, 44°37' E	640	800 mm 28°C	Arenosol	10-10-80	Plot design Crops (rice with *Vigna unguiculata*) in NT treatments	8

[a] IUSS Working Group WRB (2015).
[b] conventional tillage (CT), no-tillage (NT) .

Table 2. Calibration and validation results of NIRS and MIRS predictions for fraction masses (g fraction $100g^{-1}$ soil), SOC concentrations (gC kg^{-1} fraction) and SOC amounts (gC fraction kg^{-1} soil), and for SOC content in unfractionated soil (gC kg^{-1}). The numbers of calibration and validation samples were 109 and 25, respectively.

Variable and fraction (µm)	NIR														MIR							
	Calibration							Validation							Calibration				validation			
	Mean	SD	IQR	R^2cv	RMSEcv	RPDcv	RPIQcv	Mean	SD	IQR	R^2val	RMSEP	RPDval	RPIQval	R^2cv	RMSEcv	RPDcv	RPIQcv	R^2val	RMSEP	RPDval	RPIQval
Mass (g 100 g^{-1})																						
POM	1.2	1.2	0.8	0.59	0.7	1.6	1.1	1.0	0.8	0.7	0.87	0.4	2.1	1.8	0.55	0.8	1.5	1.0	0.74	0.4	1.9	1.6
> 200	21.3	23.9	36.6	0.97	4.1	5.8	8.8	22.5	18.2	29.5	0.95	4.2	4.3	7.0	0.96	5.0	4.7	7.3	0.95	4.5	4.1	6.6
50-200	18.4	12.6	11.0	0.87	4.4	2.8	2.5	18.2	8.7	8.8	0.52	5.9	1.5	1.5	0.81	5.5	2.3	2.0	0.50	6.1	1.4	1.5
20-50	7.9	4.7	8.4	0.83	2.0	2.4	4.3	6.9	5.1	6.8	0.65	3.0	1.7	2.2	0.83	1.9	2.5	4.3	0.60	3.2	1.6	2.1
< 20	46.0	23.4	42.1	0.93	6.3	3.7	6.7	45.9	22	37.6	0.95	5.2	4.3	7.3	0.92	6.7	3.5	6.3	0.95	5.1	4.3	7.3
SOC concentration (gC kg^{-1} fraction)																						
POM	298.7	96.8	118.0	0.49	68.9	1.4	1.7	279.5	95.9	138.9	0.50	71.4	1.3	1.9	0.40	75.8	1.3	1.6	0.60	63.7	1.5	2.2
> 200	5.7	6.5	6.9	0.43	4.9	1.3	1.4	3.1	3.9	1.7	0.58	2.5	1.6	0.7	0.24	6.0	1.1	1.1	0.59	2.5	1.6	0.7
50-200	47.0	44.7	64.7	0.95	10.0	4.5	6.5	35.4	37.7	63.5	0.99	4.6	8.2	13.7	0.91	13.1	3.4	5.0	0.97	7.7	4.9	8.3
20-50	48.7	27.6	34.1	0.91	8.3	3.3	4.1	41.3	25	39.4	0.95	5.4	4.6	7.3	0.85	10.5	2.6	3.2	0.95	5.8	4.3	6.7
< 20	45.0	16.4	22.6	0.96	3.3	4.9	6.8	39.3	14.7	20.2	0.92	4.1	3.6	4.9	0.94	4.0	4.1	5.7	0.91	4.4	3.4	4.6
SOC amount (gC fraction kg^{-1} soil)																						
POM	3.5	3.7	2.5	0.74	1.9	2.0	1.3	2.9	2.9	3.1	0.93	0.9	3.2	3.4	0.59	2.4	1.6	1.1	0.77	1.4	2.0	2.2
> 200	0.4	0.7	0.3	0.21	0.7	1.1	0.4	0.3	0.3	0.2	0.46	0.2	1.4	0.8	0.11	0.7	1.0	0.4	0.37	0.3	1.2	0.7
50-200	6.4	5.9	7.2	0.87	2.1	2.8	3.4	5.1	5.5	5.9	0.92	1.7	3.2	3.4	0.87	2.1	2.8	3.4	0.90	1.8	3.0	3.2
20-50	4.5	4.4	5.7	0.87	1.6	2.8	3.6	3.6	4.4	4.9	0.78	2.1	2.1	2.3	0.85	1.7	2.6	3.4	0.75	2.2	2.0	2.3
< 20	22.2	15.8	29.6	0.95	3.5	4.5	8.5	19.7	14.6	27.3	0.95	3.3	4.5	8.4	0.95	3.6	4.4	8.2	0.95	3.4	4.3	8.1
SOC content (gC kg^{-1})																						
Unfractio-nated soil	43.2	26.1	46.0	0.98	3.3	8.0	14.1	37.4	24.7	46.8	0.98	3.0	8.3	15.7	0.98	3.3	7.9	13.9	0.98	3.2	7.8	14.7

Mean and SD are the mean and standard deviation of reference values, respectively, in the calibration and validation subsets.
IQR is the interquartile range (Q3-Q1 difference) of the reference values in the calibration and validation subsets.
RMSEcv and RMSEP are the root mean squared error of cross-validation and of prediction, respectively.
RPDcv is the ratio of the SD of the calibration set to RMSEcv, RPDval is the ratio of the SD of the validation set to RMSEP.
RPIQcv is the ratio of the IQR of the calibration set to RMSEcv; RPIQval is the ratio of the IQR of the validation subset to RMSE.

Table SM1. Per-site reference data on fraction masses (g fraction 100 g^{-1} soil), SOC concentrations (gC kg^{-1} fraction) SOC amounts (gC fraction kg^{-1} soil), and unfractionated soil C content (gC kg^{-1}).

Variable and fraction	Total set								Andranomanelatra								Andasy							
(µm)	Mean	SD	Min	Q1	Med	Q3	Max	Skew	Mean	SD	Min	Q1	Med	Q3	Max	Skew	Mean	SD	Min	Q1	Med	Q3	Max	Skew
Mass (g 100 g^{-1})																								
POM	1.2	1.1	0.1	0.5	0.8	1.3	5.6	2.2	1.7	1.4	0.4	0.8	1.1	1.9	5.6	1.4	0.8	0.8	0.1	0.4	0.7	0.8	2.6	2.3
> 200	21.5	22.8	0.2	2.5	3.5	37.6	77.7	0.8	2.7	0.6	0.5	2.5	3.0	39.5	4.9	0.2	2.0	0.6	1.1	1.6	2.0	2.4	2.8	-0.1
50-200	18.4	11.9	3.6	10.6	15.2	21.0	65.6	1.8	12.8	3.4	6.8	11.0	14.5	21.2	21.3	0.3	8.6	2.4	5.8	7.6	8.0	9.1	13.7	1.6
20-50	7.8	4.8	1.6	3.3	7.4	11.6	17.8	0.4	11.1	3.3	3.1	3.3	9.3	11.8	17.8	0.0	14.1	1.7	12.0	12.7	14.1	14.9	16.9	0.4
< 20	45.9	23.1	4.3	27.1	41.3	68.9	78.1	-0.1	67.3	7.5	42.9	21.9	64.8	71.4	78.1	-1.0	71.9	1.5	69.7	70.8	71.9	73.1	73.9	-0.1
SOC concentration (gC kg^{-1} fraction)																								
POM	295	97	38	240	318	365	450	2.4	333	58	85	271	330	356	441	1.5	419	15	394	411	422	428	438	2.2
> 200	5.2	6.2	0.4	1.0	2.8	7.5	37.2	5.6	6.6	5.4	1.5	1.6	3.8	7.3	37.2	3.2	14.5	2.6	11.4	11.7	14.9	16.9	17.6	-0.1
50-200	44.8	43.6	1.3	8.5	27.0	71.0	183.0	1.5	83.8	36.9	34.4	15.9	63.5	77.7	182.9	0.9	70.4	16.9	54.9	57.5	64.5	79.0	103.0	0.4
20-50	47.3	27.2	7.8	24.2	44.6	62.2	127.0	1.3	71.3	21.7	42.0	44.3	56.3	70.5	126.7	0.8	42.0	7.5	31.7	38.6	41.7	45.3	56.1	1.6
< 20	44.0	16.2	21.3	31.7	40.1	54.9	84.3	0.4	57.4	12.4	36.9	38.3	50.7	57.9	84.3	0.2	37.2	2.3	33.6	36.1	37.7	39.0	40.1	-0.4
SOC amount (gC fraction kg^{-1} soil)																								
POM	3.4	3.6	0.1	1.3	2.4	3.9	20.7	-0.7	5.5	4.5	0.1	2.3	3.1	4.7	20.7	-1.9	3.4	3.2	0.4	1.7	2.9	3.2	10.9	-0.5
> 200	0.4	0.7	0.03	0.1	0.2	0.4	6.0	2.2	0.2	0.1	0.03	0.1	0.2	0.4	0.9	3.6	0.3	0.1	0.2	0.3	0.3	0.3	0.4	-0.1
50-200	6.1	5.8	0.2	1.7	4.4	8.9	28.0	1.1	11.0	5.9	2.8	4.1	6.7	12.0	28.0	1.1	5.9	1.3	4.1	5.2	5.7	6.4	7.8	1.1
20-50	4.3	4.4	0.3	0.9	2.0	6.3	17.9	0.8	8.2	4.1	1.8	1.4	5.3	8.8	17.9	0.9	6.0	1.6	4.0	5.3	5.6	6.3	9.5	0.6
< 20	21.7	15.6	2.5	7.4	13.8	36.6	54.2	0.6	38.2	7.1	20.9	7.9	35.0	39.7	54.2	0.5	26.8	1.5	24.6	25.8	27.0	27.9	28.4	-0.7
SOC content (gC kg^{-1})																								
Unfraction-ated soil	42.1	25.8	12.3	18.2	29.1	64.7	102.0	0.5	69.6	13.6	46.3	22.2	63.8	67.4	101.6	0.7	44.6	4.3	38.0	42.9	45.3	47.0	50.0	-0.5

Table SM1 (continued1). Per-site reference data on fraction masses (g fraction 100 g^{-1} soil), SOC concentrations (gC kg^{-1} fraction) SOC amounts (gC fraction kg^{-1} soil), and unfractionated soil C content (gC kg^{-1}).

Variable and fraction	Lazaina								Faraony								Alaotra Lake							
(µm)	Mean	SD	Min	Q1	Med	Q3	Max	Skew	Mean	SD	Min	Q1	Med	Q3	Max	Skew	Mean	SD	Min	Q1	Med	Q3	Max	Skew
Mass (g 100 g^{-1})																								
POM	0.6	0.5	0.1	0.3	0.5	0.7	2.6	2.7	0.4	0.1	0.3	0.4	0.4	0.4	0.4	-2.4	0.8	0.4	0.5	0.5	0.7	1.1	1.3	0.4
> 200	31.0	3.5	17.3	29.7	32.2	32.7	34.8	-2.6	40.4	3.8	36.7	38.2	39.8	40.6	47.6	1.6	1.6	2.1	0.2	0.5	0.6	1.6	6.3	2.1
50-200	17.4	5.3	3.6	15.2	16.4	19.2	35.2	1.0	19.2	2.4	15.3	18.2	19.6	20.9	21.8	-0.9	54.3	6.4	43.4	52.2	53.2	57.3	65.6	0.1
20-50	4.0	1.3	2.7	3.4	3.8	4.2	9.5	3.5	3.5	0.2	3.2	3.4	3.5	3.7	3.8	-0.3	12.1	2.2	9.0	10.7	11.8	13.9	14.9	0.1
< 20	36.0	5.0	25.8	32.7	36.8	38.6	46.9	0.0	35.0	2.2	31.9	34.5	34.7	35.3	38.8	0.7	28.4	5.3	17.8	27.2	28.5	30.5	36.9	-0.7
SOC concentration (gC kg^{-1} fraction)																								
POM	186	94	38	114	183	235	441	4.1	324	52	254	294	320	365	388	0.0	282	34	265	267	269	274	364	0.9
> 200	0.8	0.3	0.4	0.6	0.8	1.0	1.2	0.5	3.5	6.4	0.6	0.9	0.9	1.1	16.5	2.4	14.9	10.4	1.9	5.0	18.2	21.2	29.6	-0.2
50-200	4.9	1.5	1.3	4.0	4.3	5.5	9.0	0.8	9.3	2.8	5.8	7.3	9.0	11.7	12.6	0.2	9.0	2.7	5.8	7.2	8.0	10.6	13.7	1.5
20-50	15.1	3.7	7.8	13.0	15.2	17.9	21.6	3.2	38.0	11.9	21.3	29.8	39.9	46.5	51.7	-0.6	17.1	3.1	13.2	15.1	16.9	18.1	23.4	1.6
< 20	25.1	0.7	23.6	24.6	25.2	25.7	26.3	0.1	39.2	4.0	34.3	35.8	39.7	42.4	43.6	0.2	24.1	1.7	21.3	23.0	24.1	25.3	26.7	-0.1
SOC amount (gC fraction kg^{-1} soil)																								
POM	1.0	1.0	0.3	0.5	0.8	1.1	5.3	0.7	1.2	0.2	0.9	1.0	1.2	1.3	1.4	-0.1	2.4	1.2	1.3	1.5	1.9	3.1	4.6	2.7
> 200	0.3	0.1	0.1	0.2	0.2	0.3	0.4	0.1	1.3	2.3	0.2	0.4	0.4	0.4	6.0	2.4	0.1	0.0	0.1	0.1	0.1	0.1	0.1	-0.2
50-200	0.8	0.4	0.2	0.6	0.8	1.0	1.8	0.6	1.8	0.7	1.0	1.2	1.8	2.4	2.7	0.2	4.7	1.1	3.8	3.9	4.4	5.2	7.0	0.8
20-50	0.6	0.3	0.3	0.4	0.6	0.7	2.0	-0.2	1.3	0.4	0.7	1.1	1.4	1.7	1.7	-0.3	2.1	0.6	1.5	1.8	1.9	2.1	3.5	1.1
< 20	9.1	1.3	6.5	8.3	9.1	9.8	12.0	-0.2	13.8	2.2	10.9	12.3	13.8	15.0	16.9	-0.3	6.8	1.3	4.5	6.4	6.9	7.2	9.1	-0.1
SOC content (gC kg^{-1})																								
Unfraction-ated soil	17.1	1.6	14.5	16.3	17.1	17.7	21.6	0.7	26.0	2.7	21.8	25.1	25.8	27.8	29.6	-0.3	18.8	3.9	15.5	16.3	17.2	20.1	26.5	1.3

Table SM1 (continued2). Per-site reference data on fraction masses (g fraction 100 g^{-1} soil), SOC concentrations (gC kg^{-1} fraction) SOC amounts (gC fraction kg^{-1} soil), and unfractionated soil C content (gC kg^{-1}).

Variable and fraction	Belobaka								Sakaraha							
(µm)	Mean	SD	Min	Q1	Med	Q3	Max	Skew	Mean	SD	Min	Q1	Med	Q3	Max	Skew
Mass (g 100 g^{-1})																
POM	1.2	0.5	0.5	1.0	1.1	1.3	3.0	2.5	0.4	0.3	0.1	0.3	0.3	0.4	0.9	1.6
> 200	46.9	6.6	34.7	42.4	47.6	52.0	57.3	-0.3	74.2	2.0	72.2	72.8	73.5	75.0	77.7	0.8
50-200	27.7	6.6	16.5	21.8	28.3	31.6	39.0	0.2	5.8	1.1	4.6	4.9	5.7	6.4	7.8	0.4
20-50	2.8	0.6	1.6	2.3	2.8	3.1	3.9	0.0	2.5	0.4	1.7	2.4	2.5	2.6	2.9	-0.7
< 20	18.3	4.7	13.1	14.9	16.6	20.6	34.4	1.9	8.7	5.5	4.3	4.7	5.1	14.7	16.4	1.2
SOC concentration (gC kg^{-1} fraction)																
POM	236	82	70	187	249	291	364	0.1	403	37	340	385	410	426	450	1.0
> 200	1.4	0.9	0.7	0.9	1.1	1.4	5.0	2.5	2.1	1.4	0.7	1.1	1.8	2.4	5.1	1.6
50-200	13.0	5.7	4.0	9.5	12.1	14.5	28.9	0.2	24.1	12.9	13.7	17.3	20.0	24.6	52.8	1.8
20-50	39.0	10.4	16.5	33.3	38.5	44.2	72.6	2.1	42.9	16.8	26.6	32.1	35.4	50.6	76.4	1.1
< 20	36.4	4.1	31.3	32.7	36.4	38.3	48.8	2.1	58.9	8.3	49.7	53.7	54.7	65.8	70.7	1.2
SOC amount (gC fraction kg^{-1} soil)																
POM	2.7	1.1	0.5	2.3	2.4	3.2	4.7	-0.4	1.5	0.9	0.3	1.0	1.4	1.8	3.1	-0.6
> 200	0.7	0.4	0.3	0.4	0.5	0.7	2.3	3.0	1.6	1.1	0.5	0.8	1.4	1.8	3.8	1.6
50-200	3.4	1.1	1.2	2.6	3.4	4.2	5.6	1.3	1.4	0.7	0.7	1.1	1.2	1.3	2.9	1.8
20-50	1.1	0.5	0.5	0.9	1.0	1.3	2.8	1.1	1.0	0.4	0.6	0.9	0.9	1.2	1.9	1.1
< 20	6.6	1.8	4.5	5.7	6.3	7.1	13.1	1.1	4.9	2.7	2.5	2.7	3.6	7.6	8.8	0.5
SOC content (gC kg^{-1})																
Unfraction-ated soil	19.3	4.1	12.3	15.9	18.9	21.1	28.7	0.8	25.7	4.6	20.6	23.0	24.7	26.5	34.5	1.2