Geoderma 462 (2025) 117536

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

# Predicting the proportion of centennially stable soil organic carbon using mid-infrared spectroscopy

Lorenza Pacini [a,b], Marcus Schiedung [c], Marija Stojanova [a,d], Pierre Roudier [e], Pierre Arbelet [b], Pierre Barré [a,*], François Baudin [f], Aurélie Cambou [g], Lauric Cécillon [h], Jussi Heinonsalo [i], Kristiina Karhu [i], Sam McNally [e], Pascal Omondiagbe [e], Christopher Poeplau [c], Nicolas P.A. Saby [j]

[a] Laboratoire de Géologie, École Normale Supérieure, CNRS, PSL University, IPSL, Paris, France
[b] Greenback (commercial name: Genesis), Paris, France
[c] Thünen Institute of Climate-Smart Agriculture, Braunschweig, Germany
[d] ENS de Lyon, LGL-TPE, UMR 5276, Universite Claude Bernard Lyon1, UJM Saint-Etienne, CNRS, Lyon, France
[e] Manaaki Whenua – Landcare Research, Te Papaioea / Palmerston North, Aotearoa, New Zealand
[f] ISTeP, Sorbonne Université, CY Univ., CNRS, Paris, France
[g] Eco&Sols, Université de Montpellier, CIRAD, INRAE, IRD, Institut Agro Montpellier, France
[h] Ministère de l'Europe et des affaires étrangères, Ambassade de France au Kénya, Nairobi, Kenya
[i] Department of Forest Sciences, Faculty of Agriculture and Forestry, University of Helsinki, Helsinki, Finland
[j] INRAE, Info&Sols, 45075 Orléans, France

ARTICLE INFO

Handling Editor: Dr Budiman Minasny

*Keywords:*
SOC fractions
Thermal fractionation
Rock-Eval®
MIR spectroscopy
Chemometrics

ABSTRACT

Recent work has shown that it is possible to quantify the proportion of centennially stable soil organic carbon (SOC) by using Rock-Eval® thermal analysis results as input variables for PARTYsoc, a learning model calibrated on long term experiments data. This method of quantifying SOC biogeochemical stability holds promise for improving projections of SOC stock evolutions. Here, we assessed the potential of mid-infrared spectroscopy (MIR) as a lower-cost, higher-throughput technique to facilitate its wide-scale deployment.

We compiled a spectral library of over 1,800 records obtained through the scanning of samples from the French Réseau de Mesure de la Qualité des Sols to calibrate a model using MIR data to predict the proportion of centennially stable SOC. The model gave accurate predictions (RMSE = 0.06, RPD = 2.21, RPIQ = 3.26), suggesting that MIR spectra contain relevant information on SOC biogeochemical stability. We then tried to transfer this model directly to datasets acquired using another MIR spectrometer on German and Finnish soil samples. The accuracy of the predictions was degraded, even when using the CORAL data alignment method to harmonise the different spectral datasets.

Our results show that it is possible to predict the proportion of centennially stable carbon determined by the PARTYsoc model using MIR spectroscopy. However, we found that the transfer of such models to different soils, scanned with different instruments or different protocols, is difficult. Large-scale deployment of such models will require careful calibration transfer, probably associated to local calibration within a similar spectral space.

## 1. Introduction

Soil organic carbon (SOC) is essential for soil health and climate regulation (Lal, 2014). While the total quantity of SOC is an informative variable, it is also very important to evaluate SOC biogeochemical stability (tendency to resist microbial decay) to properly assess its contribution to both soil health and climate regulation. Indeed, from a soil life point of view, the benefits of organic matter depend on its decay that fuels microbial activity (Karlen et al., 1998; Wander and Bollero, 1999). Conversely, if SOC biogeochemical stability is low (high decay) higher C inputs would be needed to maintain or increase soil C stock. It is therefore crucial to evaluate SOC biogeochemical stability in order to assess its contribution to soil functioning vs. long-term carbon storage, predict the long-term evolution of SOC stocks, and better inform land

* Corresponding author.
*E-mail address:* pierre.barre@ens.fr (P. Barré).

https://doi.org/10.1016/j.geoderma.2025.117536
Received 18 June 2025; Received in revised form 29 September 2025; Accepted 1 October 2025
Available online 6 October 2025

management and planning (Janzen, 2006; Cécillon et al., 2021).

Different methods for assessing SOC stability have been proposed. The most widely-used methods are physical SOC fractionation procedures (e.g. Cambaradella & Elliott, 1992; Balesdent, 1996), which separate fractions of SOC, with contrasting biogeochemical stability, based on size or density. There are a multitude of physical SOC fractionation methods (Moni et al., 2012), however various studies have shown that a simple particle size fractionation separating particles larger than 50 μm (coarse SOC) from particles smaller than 50 μm (fine SOC) can result in two fractions of contrasting biogeochemical stability, on average (Poeplau et al., 2018), though each of these fractions remain widely heterogeneous (von Lützow et al. 2007).

More recently, an alternative method based on thermal analysis has been proposed (Cécillon et al., 2018, 2021), which distinguishes an active SOC fraction (with a mean residence time of 30–40 years) and a stable SOC fraction (with a mean residence time of 100 + years) (Delahaie et al., 2024). This fractionation uses Rock-Eval® thermal analysis (Disnar et al., 2003), the results of which are then used as input variables for the PARTYsoc, a machine learning model (Cécillon et al., 2021; Kanari et al., 2022) calibrated on observed SOC dynamics in long term agronomic experiments, and which estimates the proportion of stable SOC relative to total SOC ($C_{s,prop}$; unitless).

While current evidence suggests that the thermal SOC fractionation method better separates short and long residence time SOC pools compared to the physical fractionation method (von Lützow et al., 2007; Kanari et al., 2022), the use of the thermal SOC fractionation method requires expensive, specialist equipment, meaning it can be difficult to implement at scale. Infrared spectroscopy, both visible-near-infrared and mid-infrared (MIR), has been used as a rapid and inexpensive method for predicting soil properties for several decades (Cécillon et al., 2009; Stenberg et al., 2010; O'Rourke and Holden, 2011). The use of infrared spectroscopy in soil science relies on the energetic excitement of functional groups in the soil at different wavelengths, which includes the organic matter and the mineral phase. The absorbance information at different wavenumbers is then related to reference measurements of a range of soil properties (e.g. SOC), and an empirical statistical relationship can be built on the dataset for the chemometric prediction using a larger spectra dataset. Such approaches are currently developing for large soil spectral libraries (Safanelli et al. 2025).

In addition to the prediction of soil properties, the ability of infrared spectroscopy to quantify physical SOC fractions has already been evaluated several times across a range of pedoclimatic conditions (e.g. Barthès et al., 2008; Yang et al., 2012; Baldock et al., 2013; Jaconi et al., 2019; Sanderman et al., 2021; Ramifehiarivo et al., 2023; Cambou et al., 2024). However, very few studies have evaluated the ability of infrared spectroscopy to predict SOC thermal parameters or thermal SOC fractions (Cuarezma et al., 2019). In this study, we collate a large spectral dataset recorded from samples from the French soil monitoring network to evaluate the feasibility to predict the proportion of centennially stable SOC as determined using the PARTYsoc model using thermal analysis results as input variables from MIR spectroscopy. The transferability of the model was then assessed using data obtained on a further 412 German and 81 Finnish soil samples scanned with another MIR spectrometer.

## 2. Material and methods

### 2.1. Soil samples

#### 2.1.1. French « Réseau de mesure de la qualité des sols » (RMQS), calibration and test set

The Réseau de Mesures de la Qualité des Sols (RMQS) is the French soil monitoring network (Arrouays et al., 2003). Soils were sampled following a systematic random sampling design. The sites were located at the centre of each cell of a regular, square grid with a resolution of 16 km, resulting in a total of 2,144 distinct sites in mainland France, sampled between 2000 and 2009. A full description of the RMQS network and the soil sampling process of its first sampling campaign is available in Jolivet et al. (2006).

Due to operational reasons, A subset of 2,037 archived, finely ground topsoil samples (0–30 cm), collected during the first RMQS campaign in mainland France (2000–2010), were recovered and analyzed using Rock-Eval® thermal analysis (section 2.2). Only mineral soils (defined as having a total organic carbon concentration TOC $\leq$ 120 g/kg) were considered for this study. Further, the dataset was limited to soils for which MIR spectral data was readily available (see Section 2.3), and for which there was a good match between TOC measured using dry combustion and TOC measured using Rock-Eval® thermal analyser (i.e. a yield for TOC between 0.7 and 1.3, see section 2.2), for a final number of 1,876 soil samples.

The location of the sites provided a satisfactory spatial coverage of France (Supplementary Fig. 1). The distributions of the main pedological properties (TOC, pH, texture, and land cover) are reported in Supplementary Figs. 2 and 3.

#### 2.1.2. German and Finnish external validation and transfer sets

The German Agricultural Soil Inventory (BZE-LW) sampled between 2011 and 2017 a total of 3,104 sites under cropland, grassland or permanent crops to a depth of 100 cm, adding up to more than 16,000 oven-dried (40 °C) and sieved (2 mm) samples, which were archived (Poeplau et al. 2020). From this archive, a total of 440 topsoils (0–10 cm) were selected. These 440 samples are a combination of two data sources: 250 sampled were derived from the so called "core sites" of the BZE-LW, a subset of representative sites regarding land use, soil properties and geographical distribution (Jacobs et al. 2018, Vos et al. 2018), while another 190 were selected by Begill et al. (2023), to equally cover the whole range of SOC contents in German agricultural soils (from 0 to 120 g kg-1) in a previous study. Taking both sample sets together, the selection covered typical German agricultural soils and had an additional focus on SOC.

We selected 87 samples from Finnish croplands topsoils (0–20 cm). The samples were obtained from a network of 100 voluntary farmers that performed a carbon farming experiment in their farms (Mattila et al 2022, Salonen et al. 2024). The selected farms represented the variation in Finnish agriculture, including animal husbandry, grain and vegetable production in both organic (46 %) and conventional (54 %) farming. Soil types ranged from clay (55 %) to sand and silt (41 %), and organic soils (histosols, 4 %) (Mattila et al. 2022). Additional samples were obtained from a research site Yöni where the impacts of conventional and organic farming were investigated and compared to a natural meadow in a 24-year long experiment (Salonen et al. 2023).

As for the French dataset, only the mineral soil samples with Rock-Eval® yield for TOC between 0.7 and 1.3 (section 2.2) were considered for this study. It resulted in a selection of 412 samples out of 440 and 81 samples out of 87 for the German and Finnish datasets respectively. The geographical location of the samples is shown in Supplementary Fig. 1. The distributions of the main pedological properties (total organic carbon, pH, texture, and land cover) are reported in Supplementary Figs. 2 and 3.

### 2.2. Quantification of the proportion of the stable SOC thermal fractions

All soil samples were analysed by Rock-Eval® thermal analysis (Disnar et al., 2003; Barré et al., 2023). For each sample, ca. 60 mg of finely ground matter (< 250 μm) was placed in a special high-temperature-resistant stainless-steel pod, allowing the transport gas to pass through, and then placed inside a Rock-Eval® 6 (RE6) machine (Vinci Technologies).

There, each sample underwent a first phase of pyrolysis under an inert atmosphere ($N_2$) from ambient temperature to 650 °C (three-minute isotherm at 200 °C and then a temperature ramp of 30 °C $min^{-1}$), and a second phase of oxidation under the laboratory atmosphere

purged from water and $CO_2$, from 300 °C to 850 °C (one-minute isotherm at 300 °C and then a temperature ramp of 20 °C $min^{-1}$). During the pyrolysis phase, hydrocarbon effluents were monitored by a flame ionisation detector, and CO and $CO_2$ were monitored by infrared detectors. During the oxidation phase, CO and $CO_2$ were monitored by infrared detectors. The resulting thermograms were processed using the Geoworks software (Geoworks V1.6R2, Vinci Technologies, 2021) providing the 18 Rock-Eval® parameters used as input variables into the $PARTY_{SOC}$ model (Cécillon et al., 2018, 2021; a model trained on data from long-term agronomic experiments) to determine the proportion of the centennially-persistent SOC in the mineral topsoil samples. For each sample, the proportion of the centennially-stable SOC pool ($C_{s,prop}$) was then determined using the model $PARTY_{SOC}$ v2.0EU published in Cécillon et al. (2021).

The French (RMQS) samples were analyzed using a Rock-Eval® 6 Turbo (RE6Turbo) at Sorbonne University (France). The results were previously published in Delahaie et al. (2023; 2024). German (BZE) and Finnish samples were analyzed using a Rock-Eval® 6 Standard instrument at Vinci Technologies (France). Previous studies have shown that both machines provide very similar results (Pacini et al., 2023; Stojanova et al., 2024).

The Rock-Eval® thermal analysis provides a measure of the TOC, that reproduces very well the standard dry-combustion measure apart from a slight underestimation (Stojanova et al., 2024). The yield of a Rock-Eval® measure is defined as the ratio between the TOC measure obtained using Rock-Eval® and the TOC measure obtained using dry combustion. Rock-Eval® thermal analysis is highly reproducible. The samples for which the TOC yield was not between 0.7 and 1.3 were excluded from the analysis.

### 2.3. MIR spectroscopy measurements

MIR spectra were recorded for the different samples considered in this study on air-dried finely ground (< 250 μm) soil samples using Diffuse Reflectance Infrared Fourier-Transformed spectroscopy. For the French RMQS samples, MIR spectra were obtained between 4,000 and 400 cm − 1 (2,500 and 25,000 nm, respectively) at 3.86 cm − 1 resolution using a Nicolet 6700 (Thermo Fisher Scientific, Madison, WI, USA) as presented in Clairotte et al. (2016). Background reflectance was collected prior to each 17 scans. Concerning the Finnish and German samples (81 and 440 samples, respectively), all spectra were acquired in the same range (4,000 to 400 cm-1) at a resolution of 2 cm-1 using the same instrument, i.e. a Nicolet iS50 with a Collector II accessory (Thermo Fisher Scientific, Madison, WI, USA) similar to Schiedung et al. (2025). Before each set of 20 samples, background reflectance was collected using roughened gold mirrors (Plasmagold, Pleiger Laseroptik, Germany). Each reflectance spectrum of the French RMQS, German and Finnish datasets resulted from 32 co-added scans and was converted into absorbance spectrum as log10(1/reflectance). Both spectrometers considered (Nicolet 6700, Nicolet iS50) used deuterated triglycine sulfate (DTGS) detectors without additional cooling. All spectra were corrected using the background spectrum for atmospheric $CO_2$ and $H_2O$ interferences using the spectrometer internal software OMNIC.

The French RMQS, German and Finnish spectral datasets were compared using principal components analysis (PCA). The first two principal components were retained and used to plot the scores

### 2.4. Pre-processing of MIR spectra

The MIR absorbance spectra of all samples were pre-processed in order to increase model performance. The spectra of the German and Finnish datasets were resampled to match the resolution of the French dataset. The signal at wavelengths below 10,000 nm (wavenumber above 1,000 $cm^{-1}$) was discarded because it was highly noisy in the French dataset. Then, a Savitzky-Golay (SG) filter was applied for smoothing and derivation (the window size, polynomial order, and number of derivations for the SG filter were chosen following the same approach as for the predictive models hyper-parametrisation, presented at Section 2.5.3), followed by a Standard Normal Variate transformation and a linear interpolation to 1 nm resolution in wavelength. Finally, the signal was subsampled keeping only one band every 2 nm.

### 2.5. Regression model training and evaluation

#### 2.5.1. Training, test and validation sets definition

We trained a regression model that predicts the proportion of stable carbon ($C_{s,prop}$). $C_{s,prop}$ is unitless and provides the proportion of the TOC in the stable carbon fraction. The proportion of active carbon ($C_{a,prop}$) can be deduced from it using $C_{a,prop} = 1 - C_{s,prop}$. The concentration of stable and active carbon, expressed as gC kg $soil^{-1}$ can be calculated by multiplying the TOC by $C_{s,prop}$ and $C_{a,prop}$, respectively.

We used the French dataset for calibration and validation of the model, and the German and Finnish datasets for external validation of the model (section 2.7). The French dataset was randomly split into a training set (80 % of the data, n = 1,500) and a test set (20 % of the data, n = 376). The relative distributions of the training and test sets for each predicted variable was checked to ensure that they were consistent. The model training procedure is schematised in Fig. 1.

#### 2.5.2. Choice of a predictive model

For the prediction of the proportion of stable carbon ($C_{s,prop}$), using MIR spectral data, Partial Least Squares (PLS) regression was used to perform dimensionality reduction to $n_{components}$ latent variables, followed by a β−regression (Wood, 2017; Omondiagbe et al., 2024). Using a β−regression on the PLS latent variables has the advantage of ensuring predictions to be bounded to (0, 1), while also benefiting from the advantages of generalised additive models, such as capturing non-linear relationships between the spectral data and the properties modelled. Additionally, Bayesian implementations can allow posterior distributions to be obtained through resampling methods and provide a full description of the uncertainties associated with the predictions (Omondiagbe et al., 2024). In this study, we used the Python implementation of PLS provided in the Scikit-learn library (version 1.6.1) (Pedregosa et al., 2011) and the Python implementation of β−regression provided in the statsmodels library (version 0.14.4) (Seabold and Perktold, 2010).

Other, more complex models such as Cubist and Light Gradient Boosting Machine (LGBM) were tested, with or without dimensionality reduction through latent variables. However, these were not considered for final modelling since they performed similarly or worse than PLS and β−regression according to the studied variable, despite more complexity and compute time.

#### 2.5.3. Choice of the pre-processing and predictive model's hyperparameters

Hyper-parametrisation was required for both the pre-processing of the spectra and the calibration of the regression model used. The choice of the hyper-parameters was performed using grid search coupled with repeated five-fold cross-validation (20 repeats) on the training set (Fig. 1, top). The following pre-processing parameters were chosen following this procedure: the size of the averaging window *w*, the order of the polynomial *p*, and the order of derivation *m* of the SG filter. For the regression model, the number *l* of latent variables retained in the final model was also chosen using this procedure.

We tested the following values for the hyper parameters

- size of the averaging window *w: (5, 11, 16, 21)*
- order of the polynomial *p: (2, 3)*
- order of derivation *m: (0, 1, 2)*
- number of latent variables *l: (5, 10, 15, 20, 25, 30)*.

The optimal parameters were chosen using the one standard error method, which selects the simplest model (lowest number of

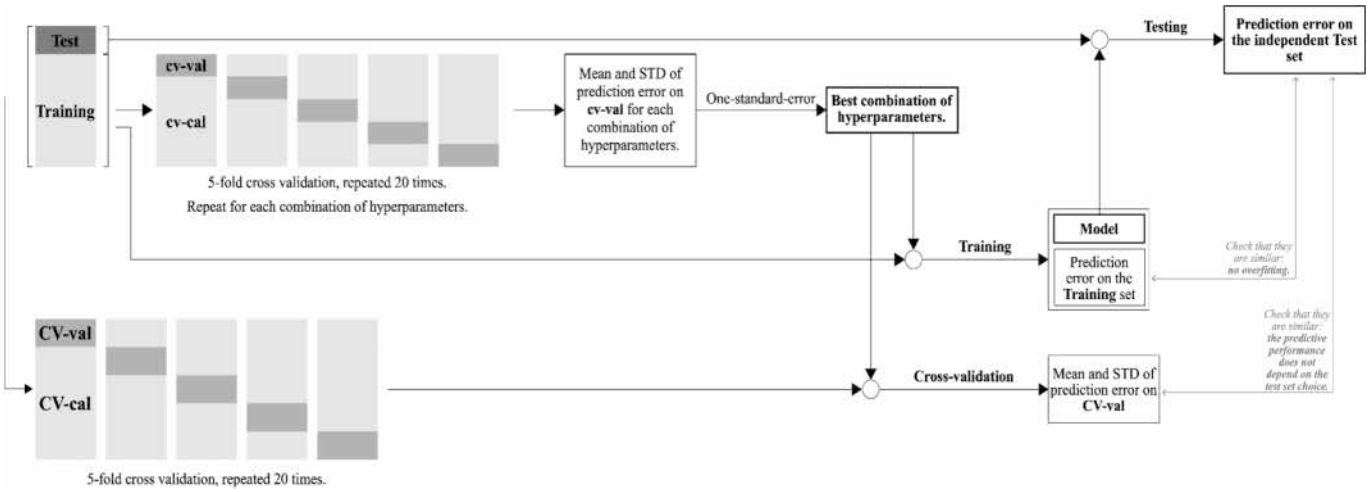


**Fig. 1.** Schematic representation of the model training and evaluation procedure applied to the RMQS dataset.

components) whose mean falls within one standard deviation of the minimum of a scoring function (Breiman et al., 1984). Here, we choose to maximise the cross-validated $R^2$, thus the scoring function was the $R^2$ with opposite sign.

### 2.6. MIR PLS model assessment

The predictive performance of each model was assessed on the test set using the following metrics: coefficient of determination $R^2$, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Ratio of Performance to Deviation (RPD), ratio of performance to interquartile range (RPIQ; Bellon-Maurel and McBratney, 2011) and bias. The $R^2$ is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (1)$$

Where $\hat{y}_i$ is the predicted value of the i-th sample, $y_i$ is the corresponding true value, and $\bar{y}$ is the mean of the observed data

The RMSE is calculated as:

$$RMSE = \left(\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2\right)^{1/2} \quad (2)$$

The MAPE is calculated as:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{|y_i|} \quad (3)$$

The RPD is calculated as:

$$RPD = \frac{\sigma}{RMSE} \quad (4)$$

With $\sigma$ the standard deviation of the observed data

The RPIQ is calculated as:

$$RPIQ = \frac{IQR}{RMSE} \quad (5)$$

With *IQR* the interquartile range of the observed data

The predictive performance metrics were also computed using repeated 5-fold cross-validation repeated 20 times on the training set to detect any sign of overfitting (Fig. 1, top right). Specifically, higher $R^2$ and lower RMSE in training compared to the independent test were interpreted as signs of overfitting.

Once the hyper-parameters were chosen, the whole training dataset (1,500 samples) was used to build a unique prediction model of the $C_{s,prop}$ that was applied to the test dataset (376 samples). The use of an independent test set ensures that the predictive performances are evaluated on data that have not influenced the model training nor the choice of the model hyperparameters. However, the prediction error obtained using this procedure may in principle be dependent on the choice of the test set. To make sure that the performance does not depend on the test set choice, we also calculated the average and standard deviation of the $R^2$ and RMSE obtained using repeated cross-validation (5-fold, repeated 20 times) on the whole dataset (training and test sets, 1,876 samples) and we verified that they were similar to the metrics obtained on the test set (Fig. 1, bottom).

The $C_{s,prop}$ values used for training, validating and testing the model are themselves output of a ML model ($PARTY_{SOC}$). As a consequence, $C_{s,prop}$ values are subject to prediction uncertainty. The RMSE for the $PARTY_{SOC}$ prediction of $C_{s,prop}$ was reported to be 0.08 (unitless) (Cécillon et al., 2021). We investigated the impact of the uncertainty of $C_{s,prop}$ on the model performances using the following perturbation analysis:

1. For each data point *i* of the training and test sets (RMQS data) associated to a $C_{s,prop,i}$ ground-truth value, replace $C_{s,prop,i}$ with a value randomly sampled from a normal distribution with mean equal to $C_{s,prop,i}$ and standard deviation equal to the RMSE for the $PARTY_{SOC}$ model (= 0.08).
2. Train a model using the perturbed training set. Evaluate the RMSE on the perturbed test set.
3. Repeat steps 2 and 3 for 100 times. Calculate the mean and standard deviation of the RMSE values obtained for the 100 repetitions.

### 2.7. Model transferability

We tested the transferability of the model to the prediction of $C_{s,prop}$ for the German and Finnish datasets using two strategies. The first approach consisted of directly applying the model to the novel dataset ("direct transfer"). The second approach consisted of using Correlation Alignement (CORAL) pre-processing to align the training set (French data; 1,876 samples) to each of the new datasets (German and Finnish datasets; 412 and 81 samples respectively) followed by re-training of the model using the aligned French data as training set. CORAL is a domain adaptation technique that reduces domain shift by aligning the second-order statistics (covariances) of the source and target domains, transforming the source domain data to match the covariance structure of the target domain and thereby minimizing distribution differences to improve model transferability across datasets (Sun et al., 2016). It does so by comparing the covariance matrices of the source and target domains $C_S$ and $C_T$, which are decomposed into lower triangular matrices

$L_S$ and $L_T$ using the Cholesky decomposition. The transformation matrix $\mathbf{A}_{CORAL}$ is then defined as:

$$\mathbf{A}_{CORAL} = \mathbf{L}_T \mathbf{L}_S^{-1} \quad (6)$$

This procedure provides a new, tailored model for each of the two datasets for which transfer was attempted (i.e. German and Finnish datasets). For comparison, two independent, local models were also trained and validated on the German and Finnish datasets, respectively, using the same procedure as for the RMQS dataset (section 2.5).

## 3. Results

### 3.1. Proportion of stable SOC thermal fractions

Table 1 shows the statistical characteristics of the proportion of the stable SOC thermal fractions ($C_{s,prop}$) measured on the different datasets considered in this study. The French dataset provided by far the largest number of samples. It presented a diversity of soils analysed, with $C_{s,prop}$ values ranging from 0.25 to 0.84, with mean equal to 0.45 and standard deviation equal to 0.12. The German dataset covered a similar range of $C_{s,prop}$ values (0.25 to 0.70) and had similar mean and standard deviation compared to the French dataset (mean = 0.42, standard deviation = 0.12). The Finnish cropland topsoils dataset provided the least number of samples and covered a narrower range of values (0.26 to 0.60, with standard deviation equal to 0.09). All datasets were moderately (<0.5) right-skewed, meaning that few samples with exceptionally high $C_{s,prop}$ are observed in each dataset.

### 3.2. Prediction of the stable SOC proportion on the French dataset using MIR spectra

The optimal pre-processing parameters for the SG filter were a window size of 5, a polynomial order of 3, and a derivation order of 1. The number of latent variables retained in the final model was 15. The results of the independent test of the predictive model are shown in Table 2 and Fig. 2. The performance metrics ($R^2$, RMSE, MAPE, RPD, RPIQ, Bias) were similar on the independent test set and in cross-validation (Table 2), showing that the performance of the model was independent of the choice of the test set.

We investigated the impact of the uncertainty of $C_{s,prop}$ on the model performances using a perturbation analysis on the RMQS dataset. We found that perturbing the ground-truth $C_{s,prop}$ values according to the $PARTY_{SOC}$ model uncertainty, the model RMSE increases to 0.08 (mean over 100 repetitions, standard deviation = 0.003).

### 3.3. Transferability of the predictive model to the German and Finnish datasets

We assessed the transferability of the model trained on the French dataset, on the German and Finnish datasets, on spectra obtained on different soils and using different instruments than the French dataset.

**Table 1**
Descriptive statistics of the proportion of stable carbon fraction (Cs, prop) measured on the datasets considered in this study (unitless). Min. = minimum; Max = maximum; Std. Dev = standard deviation; Skew. = skewness.

| Dataset | Proportion of the stable SOC thermal fractions ($C_{s,prop}$) | | | | | | |
|---|---|---|---|---|---|---|---|
| | n | Min. | Mean | Median | Max. | Std. Dev. | Skew. |
| French (RMQS) | 1,876 | 0.25 | 0.45 | 0.45 | 0.84 | 0.12 | 0.48 |
| German (BZE) | 412 | 0.25 | 0.42 | 0.42 | 0.79 | 0.12 | 0.40 |
| Finnish Cropland Topsoils | 81 | 0.26 | 0.39 | 0.36 | 0.60 | 0.09 | 0.48 |

**Table 2**
Performance metrics on the independent test set (Test) and in the 5-fold cross validation performed over the whole French dataset (training and test datasets; mean ± standard deviation, 20 repetitions).

| Dataset | $R^2$ | RMSE | MAPE | RPD | RPIQ | Bias |
|---|---|---|---|---|---|---|
| Test (N = 376) | 0.80 | 0.06 | 0.11 | 2.21 | 3.26 | −0.01 |
| Cross-validation (N = 376 in each fold) | 0.83 ± 0.01 | 0.05 ± 0.00 | −0.09 ± 0.00 | 2.42 ± 0.10 | 3.68 ± 0.22 | −0.00 ± 0.00 |

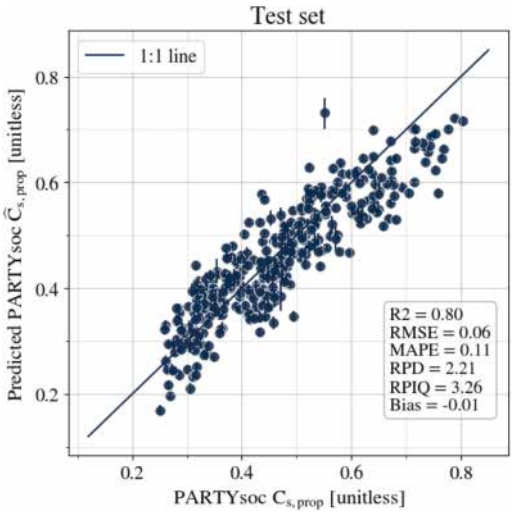


**Fig. 2.** Predicted vs. observed values of the stable carbon proportion ($C_{s,prop}$) in the independent test set (376 samples) of the French dataset. Vertical bars are 95% confidence intervals.

The prediction results without correction of the spectra (direct transfer, i.e. direct prediction of the German and Finnish spectra using the French model) are shown in Fig. 3. We found that the RMSE were doubled with respect to the performance of validation on the French test set (Fig. 3). Other metrics (RPD and RPIQ) also indicated a poor performance in prediction on the German and Finnish datasets. We then tested CORAL of the French training spectra to adapt to the German and Finnish spectra, and train two tailored prediction models. The results are shown in Fig. 4. We found that the predictions for the German dataset were slightly improved (RMSE dropping from 0.10 to 0.09; RPD and RPIQ increasing from 1.17 to 1.24 and 1.92 to 2.04 respectively). However, the differences in performances are not significant. Instead, the predictions for the Finnish dataset from CORAL-aligned data were significantly worse than without CORAL alignment of the spectra.

We investigated the difference between the three datasets by performing PCA, on the datasets without spectral correction and after CORAL alignment. The results are shown in Fig. 5. The distribution of the German spectra on the first two components overlapped to a large extent with that of the French samples. This overlap is even clearer after the use of CORAL. The Finnish samples, on the other hand, covered only a small part of the French sample distribution, and the CORAL alignment expanded the overlap.

Finally, we built two "local" models trained on 80 % of the German and Finnish datasets (German and Finnish training sets, respectively), and we tested them against the remaining 20 % of the datasets (German and Finnish test sets, respectively). The results for the German dataset are shown in Fig. 6. The model is satisfying with a RMSE of 0.07 and a RPD > 1.5. For the Finnish test, the model was largely overfitting ($R^2$ = 1.00 and RMSE = 0.01 for the Training set; $R^2 = 0.85 \pm 0.14$ and RMSE = 0.03 ± 0.01 in cross-validation), thus this model was not considered valid.

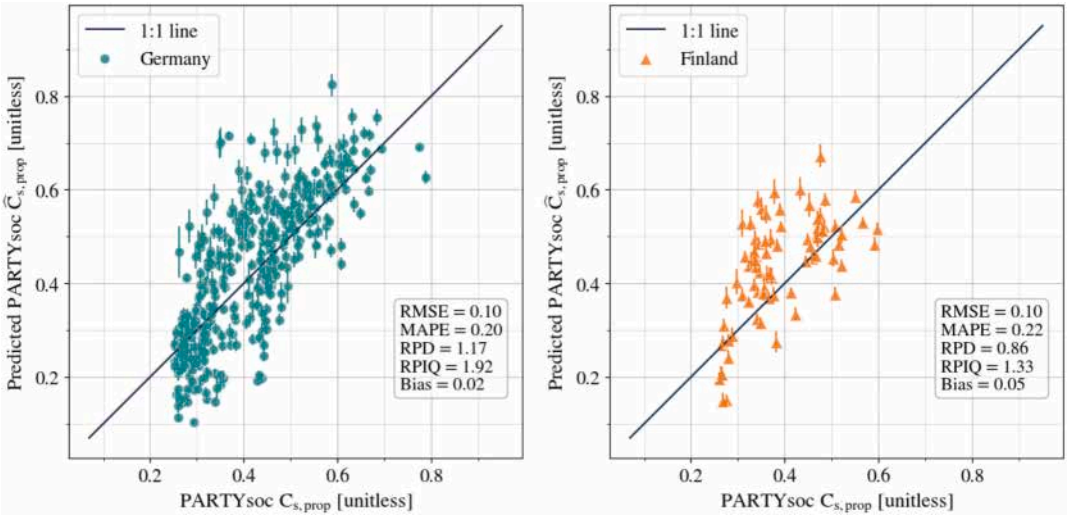


**Fig. 3.** Predicted versus observed stable carbon proportion (Cs, prop) for the German (412 samples) and Finnish (81 samples) datasets, using the regression model trained with the French spectra. Vertical bars are 95% confidence intervals.

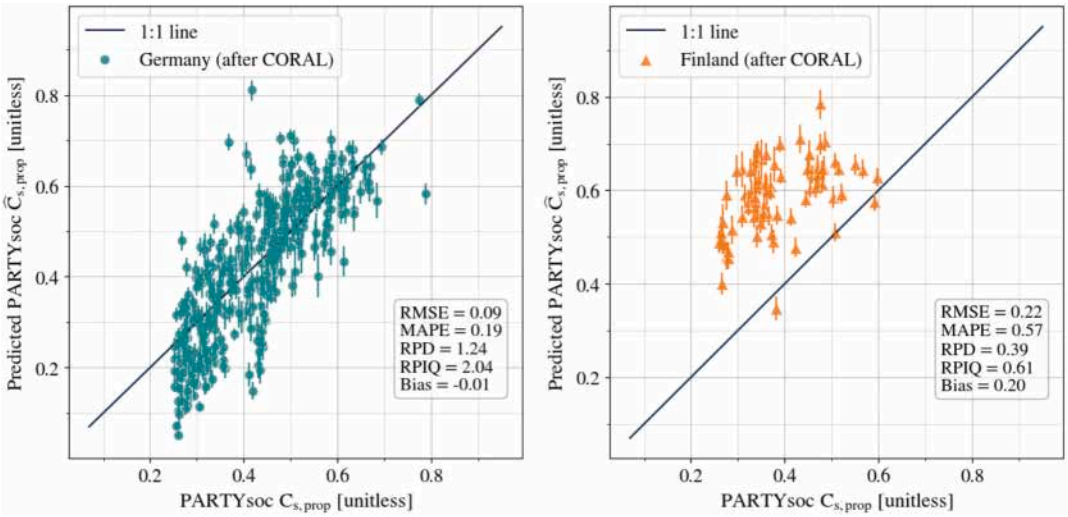


**Fig. 4.** Predicted versus observed stable carbon proportion (Cs, prop) for the German (412 samples) and Finnish datasets (81 samples), using the regression model trained with the French spectra after their CORAL alignment to the German and Finnish datasets, respectively. Vertical bars are 95% confidence intervals.

## 4. Discussion

### 4.1. Accuracy of the predictions

Over the past twenty years, a number of studies have sought to predict chemical or physical SOC fractions using MIR spectroscopy. Such studies have shown that it is possible to predict chemical SOC fractions extracted using NaOCl (Zimmerman et al., 2007) or HCl (Knox et al., 2015) and physical SOC fractions isolated by size, i.e. coarse SOC vs. fine SOC (Janik et al., 2007; Ramírez et al., 2021; Sanderman et al., 2021, Dobarco et al., 2023). However, fine SOC is generally more accurately predicted compared to coarse SOC (Ramirez et al., 2021; Remifehiarivo et al., 2023) with coarse SOC being possibly poorly predicted when the calibration is conducted on reduced sample sets or when the proportion of fine SOC is much higher than that of coarse SOC. Our study reveals that on top of chemical and physical SOC fractions, proportions of centennially stable SOC fractions as determined using Rock-Eval® thermal analyses and the PARTYsoc model can be satisfactorily predicted using MIR spectroscopy.

Specifically, we showed that the $C_{s,prop}$ can be accurately predicted using MIR spectral data for the French (Fig. 2; $R^2 = 0.80$ and RMSE = 0.06), a RMSE that is lower than the one of the PARTYsoc model calculating $C_{s,prop}$ (Cécillon et al., 2018, 2021). If the uncertainty on the PARTYsoc model is taken into account, we estimated that the RMSE of the model could increase to 0.08, that is the same as the RMSE of the PARTYsoc model itself. On the same French dataset, Clairotte et al. (2016) observed that TOC content can be predicted with MIR using global or local PLS regression models with a RPD values of 2.7 and 3.5 and a RMSE of 2.6 and 2.0 gC $kg^{-1}$ respectively. This means that stable and active SOC carbon content can be predicted using MIR data by multiplying the predicted $C_{s,prop}$ by the predicted TOC (or the measured TOC if available). For the French independent test set, we obtain a RMSE of 1.7 gC $kg^{-1}$ or 3.1 gC $kg^{-1}$ for the prediction of the active SOC quantity using a measured or a predicted value of TOC, respectively, and a RMSE of 1.7 gC $kg^{-1}$ or 2.8 gC $kg^{-1}$ for the prediction of the stable SOC quantity using a measured or a predicted value of TOC, respectively (data not shown).

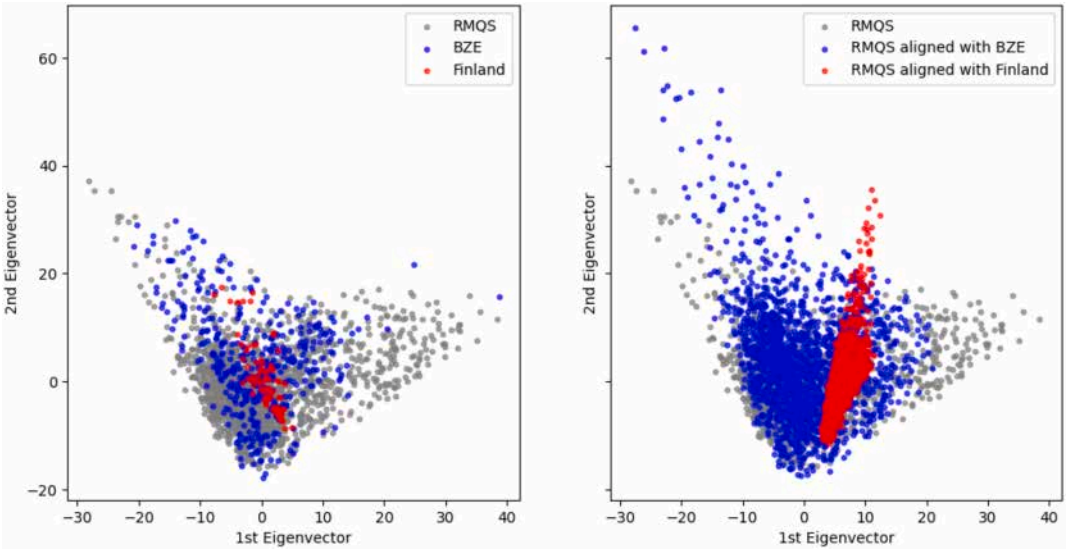


**Fig. 5.** Projection of the samples on the two first components of a principal component analysis built using the first derivative of the smoothened spectra of the French, German and Finnish datasets. Left panel: the French (BZE), German (BZE) and Finnish datasets before CORAL pre-processing. Right panel: the RMQS dataset without alignment (gray) and after alignment with the German and Finnish dataset s (blue and red, respectively). The first two components explain 28% and 24% of the variance of the combined datasets, respectively.

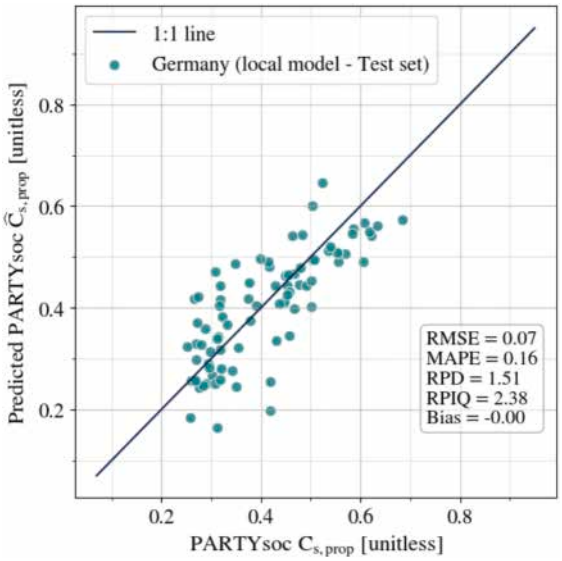


**Fig. 6.** Predicted versus observed stable carbon proportion (Cs, prop) for 20% of the German dataset (test set, 330 samples), using the regression model trained with the remaining 80% of the German dataset (training set, 82 samples).

### *4.2. Interpretation of the spectral regions contributing to the prediction of $C_{s,prop}$*

The loading values of the first PLS latent variable are shown alongside with the MIR spectra in Fig. 7. The first latent variable explained 46 % of the variance of the French spectral dataset (data not shown). It needs to be noted that while the first derivative of the spectra was used for the model predictions, the original spectra without derivation are shown in Fig. 7 for easier interpretability. The loading values of the first latent variable were highest in the regions between 1,800—-1,000 $cm^{-1}$, which includes organic and inorganic (i.e. mineral phase) bonds and their overlays (e.g. Gomez et al., 2020). The corresponding bonds and functional groups are indicated in Fig. 7a by numbers. Loading values with substantial contribution for the $C_{s,prop}$ prediction (positive and negative) can be partly assigned to organic components, which includes aliphatic (1), aromatic/carboxylic (7), aromatic bonds (8), phenolic bonds (9) and alcohol and ether (10) components (Mergenot et al., 2023). Here the region between 1,800—-1,400 cm+ (9–9) indicated mainly positive contributions, while regions between 1,400—1,000 $cm^{-1}$ (10) showed rather negative loading values. The loadings of the second latent variable (Supplementary Fig. 4) generally showed an opposite effect to those observed for the first latent variable, and explained an additional 9 % of the total variance of the spectral dataset (data not shown). Zhang et al. (2025) predicted fine SOC for 194 sites in Australia covering a range of land uses by applying deep learning on MIR spectra. The authors identified that phenolic (1400 $cm^{-1}$) bonds contributed positively to the fine (<50 μm) SOC predictions in croplands and pastures, while quartz related regions (1260 $cm^{-1}$) contributed negatively. This is partly in line with our own findings, and the loadings of the first latent variable showing substantial contributions for the $C_{s,prop}$ prediction (positive and negative) of regions assigned to phenolic bonds and silicate (Fig. 7a region 9). Additionally, in the present study, the region that can be related to aromatic and carboxylic bonds (1,680–1,580 $cm^{-1}$; Fig. 7a region 7 and 8) was identified to contribute positively to the prediction of $C_{s,prop}$. These regions are considered to represent processed and oxidized organic matter (Demyan et al., 2012) and represent a slow cycling pool of SOC (Laub et al., 2020; Schiedung et al., 2025). In addition, C=O stretching of amides in this region may indicate the importance of proteins (plant and/or microbial) to contribute to $C_s$. The abundance of these regions has also been found to be dominant in mineral associated organic carbon (Walden et al., 2025, Calderon et al 2011) and as predictors for thermally stable fractions extracted by ramped oxidation (Veeragathipillai et al. 2025). This can indicate the importance of microbial residue to contribute to $C_s$. However, the interpretation needs to be carefully considered given the organic matter in mineral phase interferences within the spectra over the wide range of soils.

Carbonate bands that are dominant at 2,515 $cm^{-1}$ and 1,790 $cm^{-1}$ (Gomez et al., 2020, Nguyen et al. 1991) also had strong impacts on the $C_{s,prop}$ with dominant positive loadings (Fig. 7 region 2 and 6), while quartz and Si-O related regions (Fig. region 3 and 5) showed rather negative loadings Across the French RMQS, Delahaie et al. (2024) identified that $C_{s,prop}$ was related to exchangeable cations and inorganic carbon, with significant positive correlations for pH, carbonates and

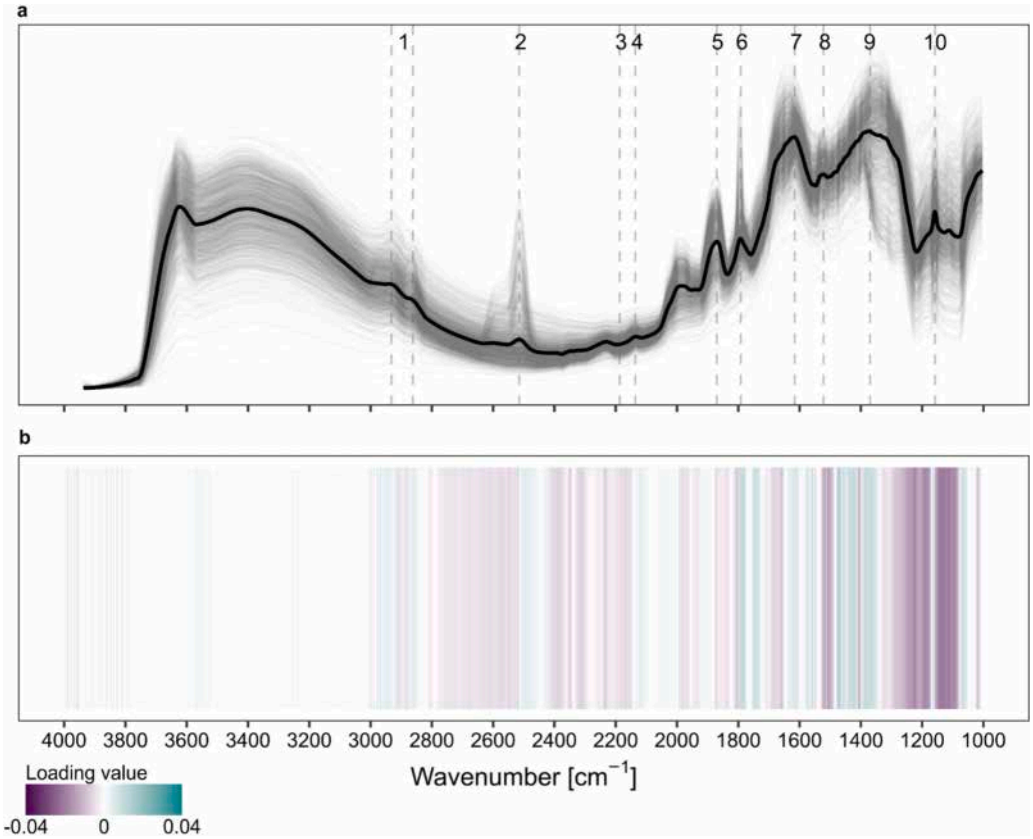


**Fig. 7. Absorbance** spectra of the French training dataset (n = 1,500) with the average spectrum in bold (Fig. 7a) and the loading values as positive and negative importance for each wavenumber of the first (Fig. 7b) latent variable of the final PLSR used for the prediction of $C_{s,prop}$. The first latent variable explained 46 % of the variance. The loadings of the second latent variable, which explained 9 % of the variance, is shown in supplementary Fig. 4. Major peaks (a) are indicated by numbers that correspond to: 1 = 2,930 and 2,866 $cm^{-1}$ for aliphatic (C-H) bonds, 2 = 2515 $cm^{-1}$ for carbonate and calcite (–$CO_3$), 3 = 2,187 for quartz and silicate (Si-O), 4 = 2,136 $cm^{-1}$ for carbohydrates (–COH), 5 = 1,870 $cm^{-1}$ for quartz and silicate (Si-O), 6 = 1,790 $cm^{-1}$ for calcite, 7 = 1,616 $cm^{-1}$ for aromatic and carboxylic (C=O and C=C), 8 = 1,522 $cm^{-1}$ for aromatic (C=C), 9 = 1,370 $cm^{-1}$ for phenolic (C-N) and 10 = 1160 $cm^{-1}$ for alcohol and ether (C-O) and Si-O bonds (Palrte et al., 2014; Parikh et al., 2014; Gomez et al., 2020; Mergenot et al., 2023, Nguyen et al. 1991). The spectra are visualized (a) as baseline off-set corrected after resampling to 2 $cm^{-1}$ and noise removal using Savitzky-Golay filtering (window size of 41 $cm^{-1}$ and fitting a polynomial fit of third order), which is different to the spectra processing for the modelling, for which the first derivative was used.

exchangeable $Ca^{2+}$. This assessment of the loading values and spectral importance for the prediction of $C_{s,prop}$ is thus in line with observed controlling factors observed in Delahaie et al. (2024). Overall, this underlines the contribution of mineral interactions and chemical composition of the SOC that contributes to the centennial stability of the $C_s$.

### *4.3. To what extent is the accuracy degraded when the model is used on another spectral dataset obtained with another spectrometer?*

Fig. 4 shows that the direct application (without correction) of the prediction model built using French data worked moderately well to predict $C_{s,prop}$ in German and Finnish samples. The $C_{s,prop}$ of the German and Finnish samples was determined using Rock-Eval® thermal analysis data produced by a different instrument to that used for the French samples. However, previous studies comparing these two instruments showed that they produced almost identical results (Pacini et al., 2023; Stojanova et al., 2024). This difference in Rock-Eval® equipment can therefore only very marginally explain the loss of precision when predicting $C_{s,prop}$ in German and Finnish samples using the prediction model built using French data.

The German and Finnish spectra from one side and the French spectra from the other side were obtained using different MIR spectrometers (different models but same manufacturer). CORAL preprocessing slightly improved the predictions for the German samples but worsened the predictions for the Finnish soil samples (Fig. 4). This is probably because, contrary to the German dataset, the Finnish dataset does not cover the same diversity as the French dataset (Fig. 5, left; Sup. Figs. 2, 3). The PCA results demonstrate that the Finnish samples occupy a more restricted area in the spectral feature space compared to the broader distribution of the French samples. When using domain adaptation techniques like CORAL, the transformation attempts to find a common subspace between source and target domains. In this case, aligning the information-rich French spectral data to match the more limited Finnish spectral variance results in a substantial loss of discriminative information. This information loss occurred because the alignment process essentially forced the French spectral features to conform to the narrower feature distribution of the Finnish dataset, eliminating spectral variations that were potentially valuable for predicting $C_{s,prop}$. The German dataset, by contrast, exhibited spectral characteristics that more broadly overlapped with the French dataset's diversity (probably due to the good overlap between the soil characteristics of the French and German sample sets, see Sup. Figs. 2 and 3), allowing the CORAL transformation to find a better common representation without significant information loss.

This observation highlights an important consideration when applying domain adaptation techniques like CORAL across soil datasets with different degrees of inherent variability. The effectiveness of such techniques depends not only on addressing instrument-specific variations but also on the underlying compositional diversity of the soil samples themselves. For datasets with limited sample size or skewed

distribution of the prediction variables (like the Finnish soil samples), alternative approaches may be more appropriate, such as selecting a subset from the source domain data (in the present study, the French dataset) that better matches the target domain characteristics (in the present study, the Finnish dataset), or developing local models specifically calibrated to the target spectral's range of variation.

It should be noted that the differences in the performances after CORAL alignment may also be influenced by the different pedological characteristics between the datasets. Indeed, the Finnish dataset includes soils with a narrower distribution of pH values compared to the other datasets (Supplementary Fig. 2) and it includes less sandy soils (Supplementary Fig. 3). It is not possible to estimate the importance of the pedological diversity of the datasets versus the spectral diversity.

Several studies have shown that calibration transfer between MIR spectrometers could improve the accuracy of predicted values (Sanderman et al., 2021; Safanelli et al., 2023). Such a calibration transfer would probably have been needed to improve the prediction of $C_{s,prop}$ on German and Finnish soil samples using the model developed on the French soil samples. Finally, if the quantity of data is sufficient, a local model may be better than a model built on a different larger dataset. Indeed, the model built on the German dataset and applied to the German dataset performs slightly better than the models trained on the French spectra, with and without CORAL alignment (see Fig. 6 compared with Figs. 3 and 4): we obtained lower RMSE and MAPE and higher RPD and RPIQ, even though they are very similar across the models (RPD = 1.17, 1.24, and 1.51 for the direct transfer of the model trained on the French spectra, after CORAL alignment of the French and German spectra, and of the model trained on the German spectra, respectively). However, this option is only possible when a sufficient dataset is available. For instance, in this study, the Finnish dataset was too small to build a robust model.

## 5. Conclusion

In this study, we assessed the potential for MIR spectroscopy to predict the stable and active SOC fractions estimated by RockEval + PARTYsoc. The main model was calibrated on 1,800 + samples from the French soil monitoring network RMQS, and our results showed that accurate predictions could be derived for $C_{s,prop}$ (RMSE = 0.06, RPD = 2.21, and RPIQ = 3.26). Those results reflect a good correlation between the MIR data and the proportion of stable SOC thermal fractions, also highlighting the potential of a direct use of MIR spectroscopy for estimating SOC biogeochemical stability. The model developed on French data was then applied to spectra recorded from German and Finnish soils. Results of this model transfer showed a degradation of the quality of the predictions, even after correlation alignment using the CORAL method was attempted. While our results suggest MIR spectroscopy can provide a way to significantly upscale the magnitude of data available for thermal SOC fractions, further work is still required to be able to transfer those calibrations across countries.

## 6. Code availability

The code associated with this paper is available on Zenodo (https://doi.org/10.5281/zenodo.17191300). Data will be shared upon reasonable request

## CRediT authorship contribution statement

**Lorenza Pacini:** Writing – original draft, Visualization, Validation, Software, Formal analysis, Data curation, Conceptualization. **Marcus Schiedung:** Writing – original draft, Visualization, Resources, Formal analysis. **Marija Stojanova:** Writing – review & editing, Visualization, Validation, Software, Formal analysis, Data curation. **Pierre Roudier:** Writing – review & editing, Methodology, Conceptualization. **Pierre Arbelet:** Writing – review & editing, Software, Conceptualization. **Pierre Barré:** Writing – original draft, Funding acquisition, Conceptualization. **François Baudin:** Writing – review & editing, Resources. **Aurélie Cambou:** Writing – review & editing, Resources. **Lauric Cécillon:** Writing – review & editing, Funding acquisition, Conceptualization. **Jussi Heinonsalo:** Writing – review & editing, Resources. **Kristiina Karhu:** Writing – review & editing, Resources. **Sam McNally:** Writing – review & editing. **Pascal Omondiagbe:** Writing – review & editing, Resources. **Christopher Poeplau:** Writing – review & editing, Resources. **Nicolas P.A. Saby:** Writing – review & editing, Methodology.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [Barre reports financial support was provided by French National Research Agency. Roudier reports financial support was provided by New Zealand Government. Barre reports a relationship with Vinci Technologies that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.].

## Acknowledgments

The FREACS project was funded by the external call of the EJP Soil (ANR-22-SOIL-0001). This work is part of project ALAMOD of the exploratory research program FairCarboN and received government funding managed by the Agence Nationale de la Recherche under the France 2030 program, ANR-22-PEXF-0002. This work was funded by the New Zealand Government to support the objectives of the Global Research Alliance on Agricultural Greenhouse Gases.The RMQS program is funded by the GIS Sol, a scientific interest group involving the French ministries in charge of the environment and of agriculture, INRAE (National Research Institute for Agriculture, Food and the Environment), ADEME (French Agency for Ecological Transition), IRD (French National Institute for Sustainable Development), IGN (National Institute of Geographic and Forest Information), OFB (French Biodiversity Agency) and BRGM (French geological survey). The authors thank the regional partners who collected soil samples and the staff of Info&Sols and of the European Conservatory for Soil Samples (CEES) involved in the RMQS program.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2025.117536.

## Data availability

Data will be made available on request.

## References

Arrouays, D., Jolivet, C., Boulonne, L., Bodineau, C., Ratié, N., Saby, N., Grolleau, E., 2003. Le réseau de Mesures de la Qualité des Sols (RMQS) de France. Etude et Gestion Des Sols 10, 241–250.

Baldock, J.A., Hawke, B., Sanderman, J., Macdonald, L.M., 2013. Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra. Soil Res. 51, 577–595. https://doi.org/10.1071/SR13077.

Begill, N., Don, A., Poeplau, C., 2023. No detectable upper limit of mineral-associated organic carbon in temperate agricultural soils. Glob. Chan. Biol. 29, 4662–4669. https://doi.org/10.1111/gcb.16804.

Balesdent, J., 1996. The significance of organic separates to carbon dynamics and its modelling in some cultivated soils. Eur. J. Soil Sci. 47, 485–493. https://doi.org/10.1111/j.1365-2389.1996.tb01848.x.

Barré P., Cécillon L., Kanari E., 2023. Charcaterization and evaluation of the stability of soil organic matter. In Baudin F. (coord). The Rock-Eval Method: Principles and application, Wiley, 181-207.

Barthès, B.G., Brunet, D., Hien, E., Enjalric, F., Conche, S., Freschet, G.T., d'Annunzio, R., Toucet-Louri, J., 2008. 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.

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.

Breiman, L., Friedman, J., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth.

Calderón, F.J., Reeves, J.B., 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.

Cambaradella, C.A., Elliott, E.T., 1992. Particulate soil organic matter changes across a grassland cultivation sequence. Soil Sci. Soc. Am. J. 56, 777–783.

Cambou, A., Houssoukpèvi, I.A., Chevallier, T., Moulin, P., Rakotondrazafy, N.M., Fonkeng, E.E., Harmand, J.-M., Aholoukpè, H.N.S., Amadji, G.L., Tabi, F.O., Chapuis-Lardy, L., Barthès, B.G., 2024. Quantification of soil organic carbon in particle size fractions using a near-infrared spectral library in West Africa. Geoderma 443, 116818. https://doi.org/10.1016/j.geoderma.2024.116818.

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. Eur. J. Soil Sci. 60, 770–784.

Cécillon, L., Baudin, F., Chenu, C., Houot, S., Jolivet, R., Kätterer, T., Lutfalla, S., Macdonald, A., van Oort, F., Plante, A.F., Savignac, F., Soucémarianadin, L.S., Barré, P., 2018. A model based on Rock-Eval thermal analysis to quantify the size of the centennial persistent organic carbon pool in temperate soils. Biogeosciences 15, 2835–2849. https://doi.org/10.5194/bg-15-2835-2018.

Cécillon, L., Baudin, F., Chenu, C., Christensen, B.T., Franko, U., Houot, S., Kanari, E., Kätterer, T., Merbach, I., van Oort, F., Poeplau, C., Quezada, J.C., Savignac, F., Soucémarianadin, L.S., Barré, P., 2021. Partioning soil organic carbon into its centennial stable and active fractions with machine-learning models based on Rock-Eval(r) thermal analysis. GMD 14, 3879–3898. https://doi.org/10.5194/gmd-14-3879-2021.

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.

Cuarezma, K., Barthès, B., Soucémarianadin, L., Baudin, F., De Danieli, S., Barrier, R., Nicolas, M., Perret, J., Barré, P., Cécillon, L., 2019. Prediction of soil organic carbon stability in French forest soils by near- and mid-infrared spectroscopy. Geophys. Res. Abstr.

Delahaie, A.A., Cécillon, L., Stojanova, M., Abiven, S., Arbelet, P., Baudin, F., Bispo, A., Boulonne, L., Chenu, C., Heinonsalo, J., Jolivet, C., Karhu, K., Martin, M., Pacini, L., Poeplau, C., Roudier, P., Saby, N.P.A., Savignac, F., Barré, P., 2024. Investigating the complementarity of thermal and physical soil organic carbon fractions. Soil 10, 795–812. https://doi.org/10.5194/soil-10-795-2024.

Demyan, M.S., Rasche, F., Schulz, E., Breulmann, M., Müller, T., Cadisch, G., 2012. Use of specific peaks obtained by diffuse reflectance Fourier transform mid-infrared spectroscopy to study the composition of organic matter in a Haplic Chernozem. Eur. Jour. Soil Sci. 63, 189–199. https://doi.org/10.1111/j.1365-2389.2011.01420.x.

Disnar, J.R., Guillet, B., Kéravis, D., Di-Giovanni, C., Sebag, D., 2003. Soil organic matter (SOM) characterization by Rock-Eval pyrolysis: scope and limitations. Org Geochem. 34, 327–343. https://doi.org/10.1016/S0146-6380(02)00239-5.

Dobarco, R.M., Wadoux, A.M.J., Malone, B., Minasny, B., McBratney, A.B., Searle, R., 2023. Mapping soil organic carbon fractions for Australia, their stocks and uncertainty. Biogeosciences 20, 1559–1586. https://doi.org/10.5194/bg-20-1559-2023.

Gomez, C., Chevallier, T., Moulin, P., Bouferra, I., Hmaidi, K., Arrouays, D., Joliver, C., Barthès, B.G., 2020. Prediction of soil organic and inorganic carbon concentrations in Tunisian samples by mid-infrared reflectance spectroscopy using a French national library. Geoderma 375, 114469. https://doi.org/10.1016/j.geoderma.2020.114469.

Jacobs, A., Flessa, H., Don, A., Heidkamp, A., Prietz, R., Dechow, R., Gensior, A., Poeplau, C., Riggers, C., Schneider, F., 2018. Landwirtschaftlich genutzte Böden in Deutschland: Ergebnisse der Bodenzustandserhebung. Thünen Report 64.

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.

Janik, L.J., Skjemstad, J.O., Shepherd, K.D., Spouncer, L.R., 2007. The prediction of soil carbon fractions using mid-infrared-partial least square analysis. Austr. J. Soil Res. 45, 73–81.

Janzen, H.H., 2006. The soil carbon dilemma:Shall we hoard it or use it ? Soil Biol. Biochem. 38, 419–424.https://doi.org/10.1016/j.soilbio.2005.10.008.

Jolivet, C., Boulonne, L., Ratié, C., 2006. Manuel du Réseau de Mesures de la Qualité des Sols, édition 2006, Unité InfoSol, INRA Orléans, France, 190 pp., ISBN 2-73-80-1235-3.

Kanari, E., Cécillon, L., Baudin, F., Clivot, H., Ferchaud, F., Houot, S., Levavasseur, F., Mary, B., Soucémarianadin, L., Chenu, C., Barré, P., 2022. A robust initialization method for accurate soil organic carbon simulations. Biogeosciences 19, 375–387. https://doi.org/10.5194/bg-19-375-2022.

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., 2014. Soil conservation and ecosystem services. Int. Soil Water Conserv. Res. 2, 36–47. https://doi.org/10.1016/S2095-6339(15)30021-6.

Laub, M., Demyan, M.S., Nkwain, Y.F., Blagodatsky, S., Kätterer, T., Piepho, H.-P., Cadisch, G., 2020. DRIFTS band areas as measured pool size proxy to reduce parameter uncertainty in soil organic matter models. Biogeosciences 17, 1393–1413. 10.5194/bg-17-1393-2020.

Mattila, T.J., Hagelberg, E., Söderlund, S., Joona, J., 2022. How farmers approach soil carbon sequestration? Lessons learned from 105 carbon-farming plans. Soil till. Res. 215, 105204.

Moni, C., Derrien, D., Hatton, P.J., Zeller, B., Kleber, M., 2012. Density fractions versus size separates: does physical fractionation isolate functional soil compartments? Biogeosciences 9, 5181–5197. https://doi:10.5194/bg-9-5181-2012.

Nguyen, T., Janik, L., Raupach, M., 1991. Diffuse reflectance infrared fourier transform (DRIFT) spectroscopy in soil studies. Soil Res. 29, 49. https://doi.org/10.1071/SR9910049.

Omondiagbe, O.P., Roudier, P., Lilburne, L., Ma, Y., McNeill, S., 2024. Quantifying uncertainty in the prediction of soil properties using mid-infrared spectra. Geoderma 448, 116954. https://doi.org/10.1016/j.geoderma.2024.116954.

O'Rourke, S.M., Holden, N.M., 2011. Optical sensing and chemometric analysis of soil organic carbon – a cost effective alternative to conventional laboratory methods? Soil Use Manag. 27, 143–155. https://doi.org/10.1111/j.1475-2743.2011.00337.x.

Pacini, L., Adatte, T., Barré, P., Boussafir, M., Bouton, N., Cécillon, L., Lamoureux-Var, V., Sebag, D., Verrecchia, E., Wattripont, A., Baudin, F., 2023. Reproducibility of Rock-Eval(r) thermal analysis for soil organic matter characterization. Org Geochem. 186, 104687. https://doi.org/10.1016/j.orggeochem.2023.104687.

Parikh, S.J., Goyne, K.W., Margenot, A.J., Mukome, F.N.D., Calderón, F.J., 2014. Soil Chemical Insights provided through Vibrational Spectroscopy. Adv. Agron. 126, 1–148. https://doi.org/10.1016/B978-0-12-800132-5.00001-8.

Pedregosa, F., et al., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

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.

Poeplau, C., Jacobs, A., Don, A., Vos, C., Schneider, F., Wittnebel, M., Tiemeyer, B., Heidkamp, A., Prietz, R., Flessa, H., 2020. Stocks of organic carbon in German agricultural soils—Key results of the first comprehensive inventory. J. Plant Nutr. Soil Sci. 1–17. https://doi.org/10.1002/jpln.202000113.

Ramifehiarivo, N., Barthès, B.G., Cambou, A., Chapuis-Lardy, L., Chevallier, T., Albrecht, A., Razafimbelo, T., 2023. Comparison of near and mid-infrared reflectance spectroscopy for the estimation of soil organic carbon fractions in Madagascar agricultural soils. Geoderma Reg. 33, e00638. https://doi.org/10.1016/j.geodrs.2023.e00638.

Ramírez, P.B., Calderón, F.J., Haddix, M., Lugato, E., Cotrufo, M.F., 2021. Using diffuse reflectance spectroscopy as a high throughput method for quantifying soil C and N and their distribution in particulate and mineral-associated organic matter fractions. Front. Environ. Sci. 9, 634472.

Safanelli, J.L., Sanderman, J., Bloom, D., Todd-Brown, K., Parente, L.L., Hengl, T., Adam, S., Albinet, F., Ben-Dor, E., Boot, C.M., Bridson, J.H., Chabrillat, S., Deiss, L., Demattê, J.A.M., Scott Demyan, M., Dercon, G., Doetterl, S., Van Egmond, F., Ferguson, R., Garrett, L.G., Haddix, M.L., Haefele, S.M., Heiling, M., Hernandez-Allica, J., Huang, J., Jastrow, J.D., Karyotis, K., Machmuller, M.B., Khesuoe, M., Margenot, A., Matamala, R., Miesel, J.R., Mouazen, A.M., Nagel, P., Patel, S., Qaswar, M., Ramakhanna, S., Resch, C., Robertson, J., Roudier, P., Sabetizade, M., Shabtai, I., Sherif, F., Sinha, N., Six, J., Summerauer, L., Thomas, C.L., Toloza, A., Tomczyk-Wójtowicz, B., Tsakiridis, N.L., Van Wesemael, B., Woodings, F., Zalidis, G. C., Żelazny, W.R., 2023. An interlaboratory comparison of mid-infrared spectra acquisition: Instruments and procedures matter. Geoderma 440, 116724. https://doi.org/10.1016/j.geoderma.2023.116724.

Salonen, A.R., Soinne, H., Creamer, R., Lemola, R., Ruoha, N., Uhlgren, O., de Goede, R., Heinonsalo, J., 2023. Assessing the effect of arable management practices on carbon storage and fractions after 24 years in boreal conditions of Finland. Geoderma Reg. 34, e00678.

Salonen, A.R., de Goede, R., Creamer, R., Heinonsalo, J., Soinne, H., 2024. Soil organic carbon fractions and storage potential in finnish arable soils. Eur. J. Soil Sci. 75, e13527. https://doi.org/10.1111/ejss.13527.

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. Biogeoch 156, 97–114. https://doi.org/10.1007/s10533-021-00755-1.

Schiedung, M., Barré, P., Peoplau, C., 2025. Separating fast from slow cycling soil organic carbon – a multi-method comparison on land use change sites. Geoderma 453, 117154. https://doi.org/10.1016/j.geoderma.2024.117154.

Seabold, S., Perktold, J., 2010. Statsmodels: econometric and statistical modeling with python. Proceedings of the 9th Python in Science Conference 7, 92-96.

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

Stojanova, M., Arbelet, P., Baudin, F., Bouton, N., Caria, G., Pacini, L., Proix, N., Quibel, E., Thin, A., Barré, P., 2024. Technical note: a validated correction method to quantify organic and inorganic carbon in soils using Rock-Eval thermal analysis. Biogeosciences 21, 4229–4237. https://doi.org/10.5194/bg-21-4229-2024.

Sun, B., Feng, J., Saenko, K., 2016. Return of frustratingly easy domain adaptation. Proceedings of the AAAI Conference on Artificial Intelligence 30, 2058–2065.
Veeragathipillai, M., Janik, L.J., Baldock, J., 2025. Can a Ramped High-Temperature Carbon Analyser with thermal Oxidation Be used to Quantify Soil Organic Carbon Pools? Talanta 295, 128358. https://doi.org/10.2139/ssrn.5175483.
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.
Vos, C., Jaconi, A., Jacobs, A., Don, A., 2018. Hot regions of labile and stable soil organic carbon in Germany–Spatial variability and driving factors, SOIL 4, 153-167.Vos, C., Don, a., Hobley, E.U., Prietz, R., Heidkamp, a., Freibauer, a., 2019. Factors controlling the variation in organic carbon stocks in agricultural soils of Germany. Eur. J. Soil Sci. 70, 550–564. https://doi.org/10.1111/ejss.12787.
Walden, L., Sepanta, F., Viscarra Rossel, R.A., 2025. FT-MIR Spectroscopic Analysis of the Organic Carbon Fractions in Australian Mineral Soils. Eur. J. Soil Sci. 76, e70084. https://doi.org/10.1111/ejss.70084.
Wander, M.M., Bollero, G.A., 1999. Soil Quality Assessment of Tillage Impacts in Illinois. Soil Sci. Soc. Am. J. 63, 961–971. https://doi.org/10.2136/sssaj1999.634961x.
Wood, S. N., 2017. Generalized additive models: an introduction with R. chapman and hall/CRC.
Yang, X.M., Xie, H.T., Drury, C.F., Reynolds, W.D., Yang, J.Y., Zhang, X.D., 2012. Determination of organic carbon and nitrogen in particulate organic matter and particle size fractions of Brookston clay loam soil using infrared spectroscopy. Eur. J. Soil Sci. 63, 177–188. https://doi.org/10.1111/j.1365-2389.2011.01421.x.
Zhang, M., Zefang, S., Walden, L., Sepanta, F., Luo, Z., Gao, L., Serrano, O., Rossel, V., R. a.,, 2025. Deep learning of the particulate and mineral-associated organic carbon fractions using a compositional transform and mid-infrared spectroscopy. Geoderma 455, 117207.
Zimmerman, M., Leifeld, J., Fuhrer, J., 2007. Quantifying soil organic carbon fractions by infra-red-spectroscopy. Soil Biol. Biochem. 39, 224–231. https://doi.org/10.1016/j.soilbio.2006.07.010.