



HAL
open science

Portable mid-infrared spectroscopy to predict parameters related to carbon storage in vineyard soils: Model calibrations under varying geopedological conditions

Ralf Wehrle, Guillaume Coulouma, Stefan Pätzold

► To cite this version:

Ralf Wehrle, Guillaume Coulouma, Stefan Pätzold. Portable mid-infrared spectroscopy to predict parameters related to carbon storage in vineyard soils: Model calibrations under varying geopedological conditions. *Biosystems Engineering*, 2022, 222, pp.1-14. 10.1016/j.biosystemseng.2022.07.012 . hal-03776179

HAL Id: hal-03776179

<https://hal.inrae.fr/hal-03776179>

Submitted on 13 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

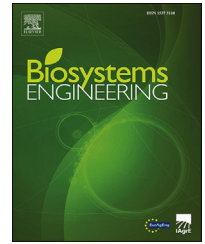
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Portable mid-infrared spectroscopy to predict parameters related to carbon storage in vineyard soils: Model calibrations under varying geopedological conditions

Ralf Wehrle ^{a,*}, Guillaume Coulouma ^b, Stefan Pätzold ^a^a University of Bonn, Institute for Crop Science and Resource Conservation (INRES) – Soil Science and Soil Ecology, Nussallee 13, 53115 Bonn, Germany^b LISAH, University Montpellier, INRAE, IRD, Montpellier Supagro, 34060 Montpellier, France

ARTICLE INFO

Article history:

Received 10 November 2021

Received in revised form

13 July 2022

Accepted 20 July 2022

Published online 14 August 2022

Keywords:

Precision agriculture

Precision viticulture

Machine learning

Support Vector Machines

Proximal soil sensing

Labile carbon

Portable mid-infrared spectroscopy (pMIRS) combined with machine learning was used to predict selected parameters for soil organic carbon (SOC) storage. In particular, SOC, soil inorganic C (SIC), hot-water extractable C (hwC), clay and sand content were predicted for ten vineyards with varying geopedological settings. As a pre-test, spectra were collected from sieved and pressed tablets with 30 and 90 kPa respectively and compared to powdery samples in order to optimise sample preparation. Further, spectra from 30 kPa tablets were used to calibrate prediction models for a sample set ($n = 540$) of 10 vineyards with pronounced geopedological variation using Support Vector Machines (SVM). The calibrated SVM models performed well with $R^2 = 0.81–0.98$ and RPIQ = 5.20–13.0 for all investigated parameters. Third, two years after the calibration samples, follow-up samples were collected from four of the vineyards. While the models performed excellent for hwC ($R^2 = 0.93$), prediction accuracy for SOC was lower. Segmentation of the total dataset into SIC-free and SIC-containing samples resulted in better predictions of SOC of the first sampling period. For the prediction of the follow-up sampling dates, model performance could not be maintained. We conclude that pMIRS-SVM calibrations are suited for the prediction of parameters related to soil C storage under varying geopedological conditions and may provide potential for future C monitoring. Extending the database with additional samples from geopedological scenarios not included in this dataset may strengthen model robustness and help to evaluate effects of SIC content on model performance.

© 2022 The Author(s). Published by Elsevier Ltd on behalf of IAGrE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Corresponding author.

E-mail address: r.wehrle@uni-bonn.de (R. Wehrle).<https://doi.org/10.1016/j.biosystemseng.2022.07.012>1537-5110/© 2022 The Author(s). Published by Elsevier Ltd on behalf of IAGrE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Soil forms the largest terrestrial carbon (C) pool, acting as a potential long-term sink for C. The storage of soil organic C (SOC) is influenced by various parameters. Thus, in the context of climate change, the evaluation of these parameters is of crucial importance (Minasny et al., 2017). Especially subsoils can provide higher SOC storage potential, because factors affecting SOC mineralisation such as aeration and microbial activity are reduced (Wordell-Dietrich, 2021). Vineyards are of special interest as they potentially provide higher SOC contents in subsoils as a result of enhanced organic fertilisation and deep tillage before planting (Alcántara, Don, Well, & Nieder, 2016). By this measure, organic materials may be mixed into deeper soil layers where C mineralisation is slower (Jakšić et al., 2021). On the other hand, topsoils of vineyards are often exposed to frequent tillage and erosion, which potentially leads to enhanced SOC loss over time (Costantini et al., 2018; Napoli, Marta, Zanchi, & Orlandini, 2017; Novara et al., 2020).

The SOC storage potential of soils depends on various chemical and physical soil properties. Besides actual SOC concentration, soil inorganic C (SIC) is important, because it regulates soil pH. Soil texture, is of outstanding relevance because SOC bound to clay minerals may reveal a high recalcitrance against microbial degradation (Hassink, 1997; Lal, 2004). Further, the composition of SOC and the proportion of labile SOC pools are important parameters to whether SOC can be stored or is predominantly mineralised. Hot-water extractable C (hWC) is a sensitive indicator for such a labile C-pool (Ghani, Dexter, & Perrott, 2003; Körschens, 2010; Körschens, Weigel, & Schulz, 1998; Zmora-Nahum, Markovitch, Tarchitzky, & Chen, 2005). As a consequence, hWC is a promising indicator for estimating C stability in soils.

Soil heterogeneity at field and landscape scale may lead to a significant spatial variation the above mentioned soil parameters (Vos et al., 2019). Novel sensor technologies allow to take infield heterogeneity into account and enable rapid estimation of soil parameters in large sample numbers (Gholizadeh, Boruvka, Saberioon, & Vašát, 2013). For this purpose, diffuse reflectance mid-infrared spectroscopy (MIRS; 4000–400 cm^{-1}) and especially novel portable MIRS devices (pMIRS) may be beneficial (Viscarra Rossel & Bouma, 2016). The application of pMIRS has recently been tested for calibrations on single fields or in homogeneous geopedological scenarios with promising results for SOC with $R^2 = 0.78$, $\text{RMSE} = 1.9 \text{ g kg}^{-1}$ (Hutengs, Seidel, Oertel, Ludwig, & Vohland, 2019), and $R^2 = 0.86$ and $\text{RMSE} = 6.5 \text{ g kg}^{-1}$ (Ji et al., 2016). Yet, studies of multiple sites from variable geopedological backgrounds in one prediction model are still scarce. Therefore, more research is needed to develop pMIRS towards a feasible and universally applicable tool for the assessment of parameters related to soil C storage.

The quality of MIRS measurements and associated models is highly dependent on sample surface condition (Janik, Soriano-Disla, & Forrester, 2020; Janik, Soriano-Disla, Forrester, & McLaughlin, 2016; Stumpe, Weihermüller, &

Marschner, 2011). As samples usually need to be dried, sieved, and ground, the sample preparation process remains time demanding. Manufacturing of compressed tablets from dried and sieved soil samples could create a sample surface that is smooth enough for pMIRS measurements. This approach would also facilitate the workflow by making sample grinding superfluous. This method was for the first time tested by Marmette and Adamchuk (2020) on a smaller dataset with results suggesting suitability for the prediction of SOC ($R^2 = 0.77$; $\text{RMSE} = 2.7 \text{ g kg}^{-1}$).

In the MIR region, absorbance spectra are strongly influenced by SIC content and reveal specific peaks at e.g. 700, 880 and 1450 cm^{-1} or $2600\text{--}2500 \text{ cm}^{-1}$ (Nguyen, Janik, & Raupach, 1991; Gomez, Moulin, Barthès, 2022; Leenen, Welp, Gebbers, & Pätzold, 2019). As a consequence, SIC content can have strong impact on MIRS calibrations as its spectral signature is strong relative to SOC. For this study on vineyard soils, the SIC influence on SOC prediction is of particular interest, because many vineyards are planted on calcareous soils, namely in the Mediterranean, but also in German vine-growing regions (Backes, Böhm, Gröber, Jung, & Spies, 2015; Gomez & Coulouma, 2018).

With this background, non-linear machine learning approaches such as Support Vector Machines (SVM) are becoming more appropriate to calibrate MIRS prediction models (Wadoux, Minasny, & McBratney, 2020). Support Vector Machines are a non-parametric, non-linear statistical learning method which does not assume a known statistical distribution of the data (Mountrakis, Im, & Ogole, 2011). In general, SVM are capable of maintaining a high generalisation potential for unknown (test) data and are particularly powerful for large datasets containing complex data (Ludwig, Murugan, Parama, & Vohland, 2019). Therefore, a non-linear SVM regression approach may be well suited for the prediction of soil properties especially for a complex dataset, where multiple sites of different geopedological regions are to be combined.

Against the background of diverse challenges in rapid evaluation of parameters related to SOC storage in vineyard soils, the aims of the study were (i) to develop site-independent SVM prediction models for SOC, hWC, SIC, clay and sand content. To achieve site-independent model validity, 10 geopedological variable vineyards including top and subsoils were investigated. Further aims were (ii) to test the model performance for monitoring purposes, on follow-up samples and (iii) to evaluate the alteration of model performance by arbitrarily occurring SIC. In order to simplify and optimise sample preparation during the study, pre-tests were conducted to evaluate the feasibility of compressing soil tablets prior to recording pMIRS spectra.

2. Materials and methods

2.1. Characterization of the study sites

Ten vineyards with a wide target value range, i.e., from variable geopedological scenarios and with pronounced in-field soil heterogeneity were investigated in this study. Seven

were located in three German wine growing regions (Palatinate, Rhine-Hesse, and Mosel): Two vineyards were located at the Federal Institute for Grapevine Breeding Geilweilerhof, Siebeldingen (Palatinate). Here, one newly planted vineyard in 2018 (Sieb N) and an approximately 15 years old plantation (Sieb B) were investigated. Two vineyards were located in Sprendlingen (Rhine-Hesse): A new plantation, established in 2019 (Spre N) and an older vineyard established in 2013 (Spre B). In the Mosel region, one vineyard was located in Leiwien (Leiw) and two in Kanzem (Saar district; Kan JB and Kan St). Further, three vineyards in southern France were investigated, described in more detail by Coulouma, Caner, Loonstra, and Lagacherie (2016). These sites were located in the Languedoc region, near Alignan-du-Vent (Alig), Narbonne (Narb) and Colombiers (Colo). A rough location of the investigated

vineyards is presented in Fig. S1 within the supplementary material.

In total, the dataset comprised 540 samples of topsoils (0–20 cm) and subsoils (20–60 cm). For the vineyards Narb and Colo, only top soils were investigated. Descriptive statistics of conventionally measured data and geopedological characteristics of the individual vineyards are given in Table 1. For overall comprehension, the dataset from Table 1 will be referred to as “dataset 1”. The overall ground truth data used for this study varied from <5.00 to 683 g kg⁻¹ SIC, 0.8–36 g kg⁻¹ SOC, 0.02–1.55 g kg⁻¹ hWC, 52.4–708 g kg⁻¹ clay and 28.3–732 g kg⁻¹ sand over all examined vineyards, reflecting the large variability of soil characteristics under study. The desired in-field heterogeneity of the individual vineyards was ensured (see SD values for the single sites, Table 1).

Table 1 – Soil inorganic carbon (SIC), soil organic carbon (SOC), hot-water extractable carbon (hWC), Clay and sand content in dataset 1 of the examined vineyards determined by conventional laboratory analysis (n = 540). All samples were taken during the initial sampling campaign.

Site and geopedological setting	Statistical parameter	SIC	SOC	hWC	Clay	Sand
		(g kg ⁻¹)				
Alig: (Pleistocene fluvial deposits and Miocene marine deposits)	Min	<5	3.12	0.06	149	338
	Max	371	22.1	1.27	366	603
	Mean	116	10.2	0.63	274	475
	SD	117	6.17	0.32	50.6	52.1
Colo: (Miocene marine deposits)	Min	162	9.64	0.12	67.8	195
	Max	308	15.4	0.49	261	343
	Mean	209	11.8	0.31	151	282
	SD	47.3	1.74	0.10	54.2	47.7
Sieb B1: (Pleistocene loess loam)	Min	<5	4.11	0.08	52.4	224
	Max	90.4	15.5	0.72	316	516
	Mean	13.6	9.63	0.34	146	293
	SD	19.8	2.69	0.13	54.9	52.8
Sieb N1: (Keuper/Bunter Sandstone rock)	Min	<5	0.84	0.02	101	130
	Max	182	31.4	1.45	510	387
	Mean	28.3	15.4	0.52	3256	272
	SD	45.8	8.23	0.32	102	48.8
Kan JB: (Devonian shale)	Min	<5	5.06	0.10	63.0	471
	Max	<5	34.3	1.24	238	657
	Mean	<5	15.6	0.60	152	549
	SD	<5	7.68	0.35	27.3	54.5
Kan St: (Pleistocene fluvial deposits/Devonian shale)	Min	<5	3.41	0.16	103	500
	Max	<5	28.1	1.50	263	732
	Mean	<5	14.8	0.83	140	672
	SD	<5	7.78	0.44	43.6	70.9
Leiw: (Pleistocene fluvial deposits)	Min	<5	4.07	0.03	76.6	253
	Max	<5	36.0	1.50	263	554
	Mean	<5	17.6	0.75	171	452
	SD	<5	7.63	0.36	39.0	75.7
Narb: (Holocene marine sand deposits)	Min	204	6.66	0.05	104	328
	Max	683	14.9	0.74	199	453
	Mean	440	10.3	0.35	166	378
	SD	117	2.17	0.18	28.2	41.1
Spre B1: (Oligocene marl)	Min	263	2.44	0.24	113	63.4
	Max	471	29.3	1.16	597	251
	Mean	420	16.2	0.75	467	150
	SD	30.0	7.33	0.19	111	42.6
Spre N1: (Oligocene marl)	Min	191	5.93	0.06	226	28.3
	Max	486	27.9	1.55	708	191
	Mean	276	16.3	0.57	507	99.9
	SD	52.2	5.14	0.31	91.8	29.2

Further, for the vineyards Sieb B, Sieb N, Spre B and Spre N, sampling was repeated two years after the initial sampling as listed in Table 1. These follow-up samples formed an independent test-set to evaluate the models' feasibility for a future pMIRS-based monitoring program for SOC and hwC (in the following denoted as "dataset 2"). Samples for dataset 2 were taken in immediate vicinity (<0.5 m) of the sampling point from dataset 1. The descriptive statistics of the ground truth data, for the follow-up samples, are given in Table 2. In total, the dataset 2 comprised 243 topsoil and subsoil samples of four vineyards located in Palatinate (Sieb B and Sieb N) and Rhine-Hesse (Spre B and Spre N), Germany.

For an overview, the composition of the different datasets can be seen in Table 3. Further on, to test if SIC content affects model performance the datasets were divided into samples with and without SIC.

2.2. Sample preparation and determination of ground truth data

Prior to further processing, all soil samples were air-dried, pestled and sieved to <2 mm. Soil organic carbon was determined from the difference between total C and SIC. Total C was determined by dry combustion and elemental analysis (ISO 10694, 1995). If present, SIC content was calculated as 0.12 x the calcium carbonate content, which was determined by the gas-volumetric Scheibler Method (ISO 10693). Otherwise, if no SIC was present, total C was rated as SOC for further analyses. Determination of hwC followed the method of Körschens et al. (1998) which is based on a 1-h extraction of 5 g air dried and sieved (<2 mm) soil with 25 ml distilled water at 100 °C under reflux. After extraction, cooling and centrifugation at 2600 rpm for 10 min, the C concentration in the supernatant was analysed

Table 2 – Dataset 2, follow-up sampling to test model feasibility for monitoring purposes: Soil organic carbon (SOC) and hot-water extractable carbon (hwC) contents at the study sites in Siebeldingen and Sprendlingen as determined by conventional laboratory analysis (n = 243).

Site	Statistical parameter	SOC (g kg ⁻¹)	hwC
Sieb B2 2nd	Min	2.81	0.02
	Max	21.4	0.69
	Mean	10.4	0.33
	SD	4.73	0.16
Sieb N2 2nd	Min	4.76	0.01
	Max	18.4	0.96
	Mean	11.2	0.50
	SD	3.82	0.18
Spre B2 2nd	Min	14.9	0.29
	Max	29.2	1.57
	Mean	21.3	0.88
	SD	3.67	0.28
Spre N2 2nd	Min	6.88	0.07
	Max	20.1	1.13
	Mean	15.8	0.63
	SD	2.82	0.22

Table 3 – Composition of the datasets 1 and 2 (total, with and without).

Dataset	Site	n total	without SIC	with SIC
Dataset 1 (Calibration & Validation)	Alig	62	42	20
	Colo	16	0	16
	Sieb B1	80	48	32
	Sieb N1	79	46	33
	Kan JB	46	46	0
	Kan St	18	18	0
	Leiw	53	53	0
	Narb	20	20	0
	Spre B1	103	0	103
	Spre N1	63	0	63
Total		540	273	267
Dataset 2 (Monitoring; follow-up)	Sieb B2	55	29	26
	Sieb N2	54	13	41
	Spre B2	87	0	87
	Spre N2	47	0	47
	Total	243	42	201

with a TOC-analyser (Shimadzu TOC-VCPA; Shimadzu Deutschland GmbH, Duisburg, Germany). Soil texture analysis was carried out via the combined sieve and pipette method (ISO 11277, 2002). All analyses were carried out in duplicate.

2.3. Sample preparation and acquisition of portable MIR spectra

As a pre-test, a simplified sample preparation approach that relies on compressing soil tablets instead of grinding samples was tested. For this purpose, spectra acquisition via pMIRS of three sample preparation methods was evaluated. First, sieved samples were ground in a ball mill and 2 g of each sample were placed in a Petri dish and smoothed by gentle pressing (in the following denoted as powdery samples). Second and third, soil tablets were pressed at two different pressures in order to avoid the laborious grinding step. Therefore, sieved soil samples were compressed to tablets without previously grinding or pestling. For manufacturing the soil tablets (Fig. 1b) a hand-driven hydraulic press was used. To test if the pressure level had an effect on spectra quality and model calibration, tablets with 30 kPa and 90 kPa pressure were created. Portable MIRS models were calibrated and compared on both tablet variants as well as on the powdery samples.

For pMIRS measurements, a handheld FTIR Agilent 4300 (Agilent Technologies, Santa Clara, CA, USA) equipped with a deuterated triglycine sulphate detector and a zinc selenide beam splitter, a DRIFT interface, and a golden reference cap was used (Fig. 1a).

The acquisition of pMIRS spectra was carried out as described by Wehrle, Welp, and Pätzold (2021) where three repeated measurements of each sample were carried out after slightly rotating the Petri dish/compressed sample between the measurements while each spectrum was recorded with 80 repeated scans. Spectra acquisition with pMIRS was carried out on an instrument stand provided by the manufacturer

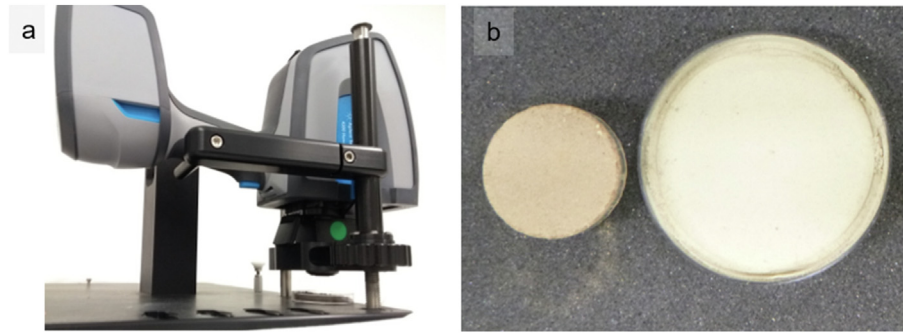


Fig. 1 – Mid-infrared spectrometer (a) portable MIRS Agilent 4300 and (b) powdery (ground) soil sample (right), sieved and compressed sample tablet (left).

(Fig. 1a). The portable device recorded spectra in the $4000\text{--}650\text{ cm}^{-1}$ range at a spectral resolution of 4 cm^{-1} . For compensation of instrument drift and variation in the environment of the measuring chamber, a background spectrum was taken every 10 min using a golden reference cap.

2.4. Spectra pre-processing and model calibration

For further analysis, all spectra of each sample were averaged in order to reduce noise. From pMIRS spectra, spectral range of $3800\text{--}650\text{ cm}^{-1}$ was selected for further investigations. To visualise spectral differences of the studied vineyards, a Principal Component Analysis (PCA) was carried out. For this purpose, centred values and a correlation matrix were used. The SVM model calibrations as well as PCA analysis were performed with the statistic software R (2013) using the packages “e1701” (Meyer et al., 2020), “prospectr” (Stevens, Ramirez-Lopez, & Hans, 2020), and “ggplot2” (Wickham et al., 2020), the latter for visualization. Six pre-processing techniques for the absorbance spectra were tested to remove light scattering effects, to correct baseline offset and to improve model performance: no pre-processing, multiplicative scatter correction (MSC), Savitzky Golay Filter (SG), Standard Normal Variate-Detrend algorithm (SNV), first derivation (1st Der) and first derivation + SG (1st Der SG). These pre-processing approaches were evaluated by the associated cross-validation results and the best model was finally chosen (for model performance criteria see below). Prior to model calibration, dataset 1 was divided into independent calibration (70%) and validation (30%) samples by using the *k*-means sampling algorithm with 100 iterations.

For the non-linear SVM approach, the radial basis function kernel was used. The SVM prediction models were trained using repeated 10-fold cross validation for all spectral pre-processes in order to find the optimal prediction model for each investigated parameter. Cross-validation was optimised by an automated grid search for the SVM hyperparameters gamma and cost. The range for both hyperparameters was set to 0.001, 0.01, 0.1, 1, 10 and 100. Then, a test-set validation was performed to test the model performance on unknown samples.

For the preliminary study on testing the feasibility of pressing soil tablets for pMIRS measurements, linear PLSR

calibrations using the “Kernel PLS” algorithm combined with leave-one-out cross validation (LOOCV) were programmed. To avoid over-fitting of the PLS model, the maximum number of latent variables was limited to ten. To determine the quality of the predictive models, the coefficient of determination (R^2), the root mean squared error (RMSE) and the ratio of performance to interquartile (RPIQ) according to (Bellon-Maurel, Fernandez-Ahumada, Palagos, Roger, & McBratney, 2010), were used and calculated as follows:

$$R^2 = \frac{\sum_{i=1}^n (f_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2} \quad (2)$$

$$\text{RPIQ} = \frac{\text{IQ}}{\text{RMSE}} \quad (3)$$

where f_i is the predicted, and y_i the respective observed value and IQ is the interquartile distance that gives the range that accounts for 50% of the population around the median. For RPIQ values, the threshold for an insufficient model performance was defined by $\text{RPIQ} < 2.5$ according to (Ludwig et al., 2019).

To determine the wavenumber importance for the examined soil parameters, a recursive feature elimination was used (R “caret” package; Kuhn, 2022). This method implements backwards selection of predictors (wavenumbers) based on importance ranking from the first to the last important

Table 4 – Dataset for the evaluation of sample preparation and instrument comparison. Soil inorganic carbon (SIC), soil organic carbon (SOC) and hot-water extractable carbon (hwC) contents in samples from the vineyard Spre B1 determined by conventional laboratory analysis ($n = 42$).

Property (g kg^{-1})	Min	Max	Mean	SD
SIC	292	533	366	52.2
SOC	13.3	28.4	19.5	3.83
hwC	0.27	1.13	0.59	0.24

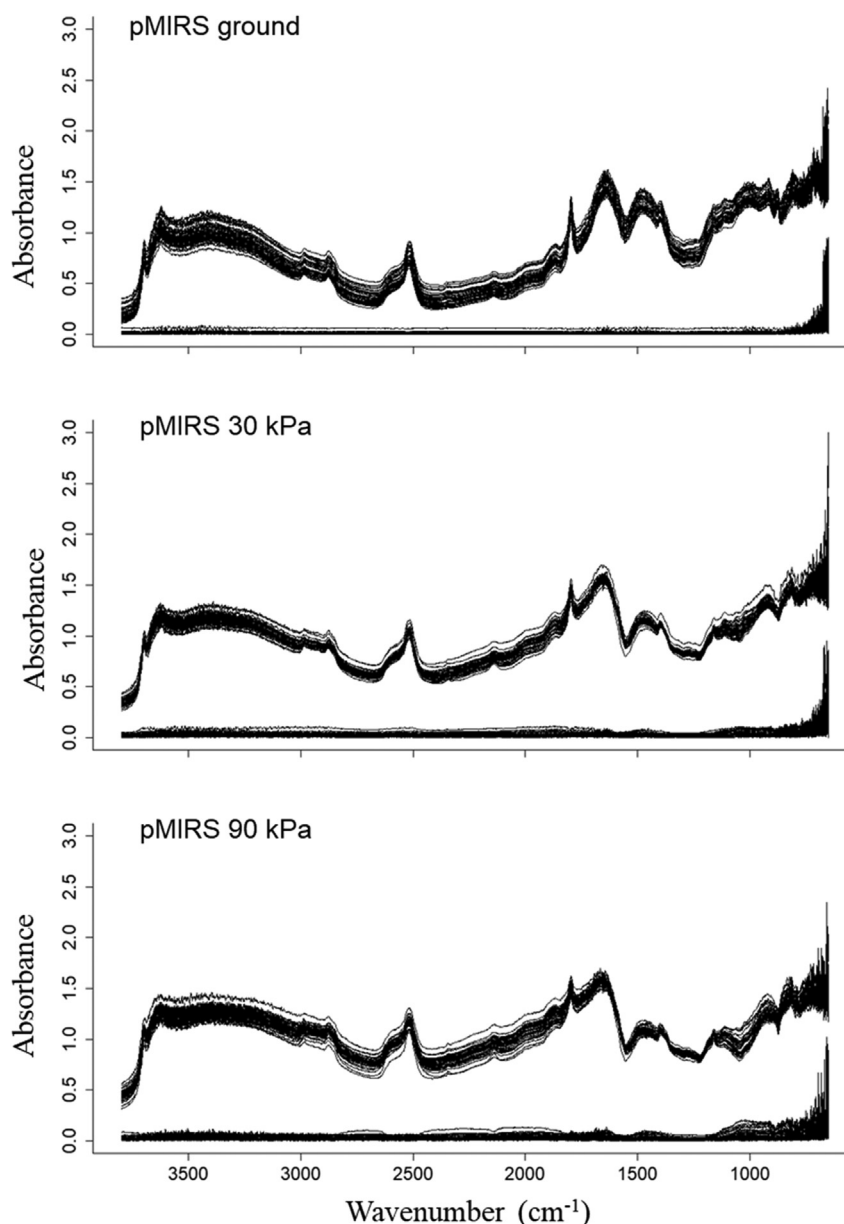


Fig. 2 – Spectra and standard deviations for portable MIRS with powdery (ground) samples, portable MIRS with 30 kPa soil tablets and portable MIRS with 90 kPa soil tablets at the site Spremlingen B1 ($n = 42$).

predictor. The recursive feature elimination was processed via 10-fold CV.

3. Results and discussion

3.1. Applicability of compressed soil tablets for pMIRS model calibration

In a pre-test the pMIRS models for the 30 and 90 kPa tablets were compared to those from powdery samples on a smaller

dataset from the vineyard Spre B1. The ground truth data are displayed in Table 4 and the spectra as well as standard deviations of each measuring approach are displayed in Fig. 2. For this dataset in-field heterogeneity is evident with large data ranges for SIC, SOC and hWC (Table 4).

Overall spectral standard deviation (SD) for all pMIRS methods was small except for the spectral region between 800 and 650 cm^{-1} . The same was previously reported by Wehrle et al. (2021) for powdery pMIRS spectra of organic soil amendments. Yet larger noise in this spectral region did not negatively affect the model calibration process in their study. While noise

and spectral SD did not differ between pMIRS with powdery and 30 kPa samples, spectral SD of 90 kPa compressed samples was larger for the region of 122–650 cm⁻¹. For all sample preparation approaches, the distinct spectral alteration between 2400 and 2600 cm⁻¹ caused by SIC is clearly visible.

For this site-specific pre-test on the smaller dataset a PLSR approach and LOOCV was used for calibration. Here, the LOOCV calibration for SIC reveals similar results for all sample preparations; all variants performed well with similar high R² values (0.94) and RPIQ (5.96–6.44; Table 5). For SOC and hwc, the models for tablets tended to better LOOCV results than for powdery calibrations with high R² and lowest RMSE values.

Stumpe et al. (2011) and Barthès et al. (2016) showed that sample surface conditions strongly impact MIR spectral quality, where grinding significantly improved model calibrations via PLSR. When sieved samples were compared to milled ones using the same pMIRS as in this study, Breure et al. (2022) found lower model performance (reduced RPIQ by up to 76%) for the prediction of SOC, clay and pH on sieved samples. In this study, simply sieved and compressed tablets (30 kPa pressure), revealed a smooth surface without pestling, grinding and milling. Obviously, this surface quality was sufficient for reliable pMIRS spectra acquisition at minimised sample preparation time. The feasibility of this simplified sample preparation enables perspectives for pMIRS device application in field campaigns after further optimization (e.g. concerning soil moisture). To date, soils need to be air-dried, irrespective of pressing or not, because of strong water effects on spectral information. Therefore, future field and calibration studies should take into account the level of soil moisture e.g. Rodionov et al. (2014) did for vis-NIR.

The applied pressure did not have an effect on LOOCV results, but as spectral SD was smaller for 30 kPa, this variant was used for further spectra acquisition and model calibration.

3.2. Calibrating prediction models for parameters related to carbon storage in vineyard soils with varying geopedological background via SVM

In Fig. 3a the PCA of top- and subsoil pMIRS spectra of compressed 30 kPa-tablets and ten vineyards are displayed. The dataset revealed high spectral variability within and between the study sites where PC1 and PC2 explained 90% of the occurring variability. Soil depth did not affect spectral

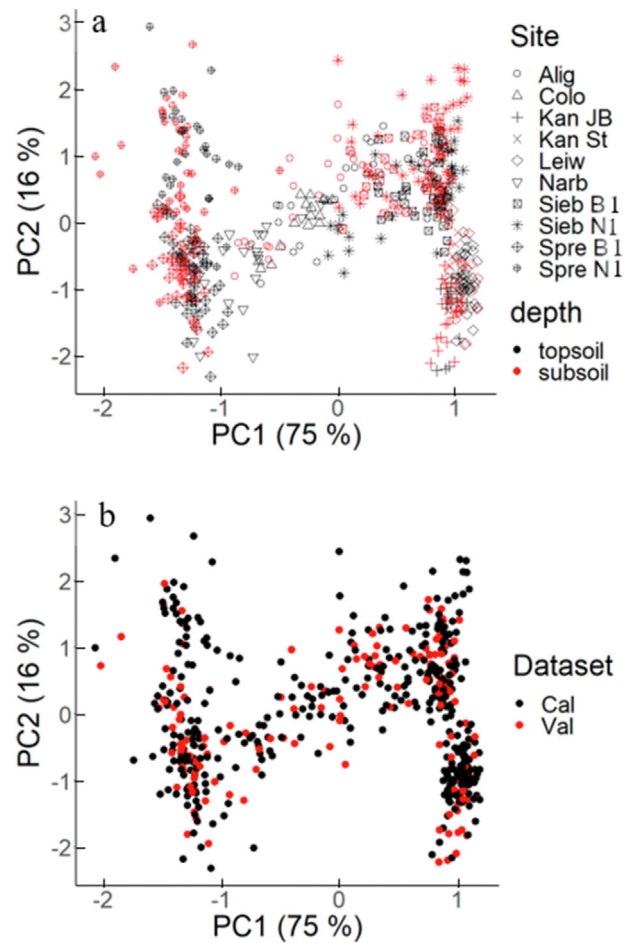


Fig. 3 – Principal components (PC1 and PC2 scores) of multiplicative scatter corrected (MSC) spectra as influenced by study site (a) and calibration (black; n = 378) and validation (red; n = 162) samples as selected by k-means sampling (b). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

information to a large extent. This may be because of deep tillage that is common, sometimes down to 1 m working depth, before planting vines. This leads to mixing of the soil over the entire working depth, to suppression of soil horizons and thus to homogenization of spectral information.

Table 5 – PLSR-LOOCV results for a pre-test comparing pMIRS with ground powdery samples (gr), 30 kPa and 90 kPa compressed soil tablets for soil carbon parameters at Sprendlingen B1 (n = 42).

Parameter	sample preparation	spec pre-processing	RMSE _{CV}	R ² _{CV}	RPIQ _{CV}	LV
SIC (g kg ⁻¹)	port. gr	SNV	12.1	0.94	6.44	2
	port. 30 kPa	No treat	13	0.94	6.02	3
	port. 90 kPa	1st Der	13.1	0.94	5.96	2
SOC (g kg ⁻¹)	port. gr	SNV	1.25	0.90	4.95	6
	port. 30 kPa	1st Der	0.92	0.95	6.96	6
	port. 90 kPa	1st Der	0.92	0.95	6.78	4
hwc (g kg ⁻¹)	port. gr	1st Der SG	3.89	0.96	6.61	8
	port. 30 kPa	1st Der	3	0.98	9.32	7
	port. 90 kPa	SNV	2	0.99	10.4	8

Table 6 – Calibration (n = 378) and test-set validation (TSV, n = 162) results and model hyperparameters for site-independent SVM models for the prediction of organic carbon storage parameters of top and subsoils using pMIRS and soil tablets (Dataset 1).

Property (g kg ⁻¹)	Calibration (CV)			TSV			Hyperparameters*			
	RMSE	R ²	RPIQ	RMSE	R ²	RPIQ	s.p.	γ	C	SV
SIC	14	0.99	21.5	25.2	0.98	13	1st Der	0.1	100	350
SOC	1.83	0.93	5.04	2.81	0.81	5.2	SNV	0.01	10	243
hwC	1.02	0.91	4.63	1.32	0.8	5.2	1st Der	0.1	10	129
Clay	46.1	0.92	6.07	53.8	0.91	5.76	MSC	0.1	100	369
Sand	26.9	0.98	10.8	43.9	0.92	6.43	No treat	0.1	100	376

*Hyperparameters: s.p. = spectral pre-processing; LV = Latent variables (PLS); γ = Gamma (SVM), C = Cost (SVM); SV = Number of support vectors (SVM).

Sites located in the same geopedological region were more similar to each other (e.g. Spre N and Spre B, Sieb N and Sieb B as well as Kan JB and Kan St, Fig. 3a). Nevertheless, the PC scores of the regions were not separately clustered by regions but overlapped. As a consequence, the k-means sampling algorithm resulted in an evenly distributed calibration and validation dataset within the PCA space, where top and subsoil samples of the individual study sites are represented in both datasets (Fig. 3b).

For calibrating a universal prediction model of parameters related to SOC storage, a non-linear SVM approach was used. Calibrations via PLSR were also tested but results were worse compared to SVM (data not shown). Therefore, only the SVM calibration results are presented. Calibration and independent test-set validation results as well as algorithmic hyperparameters and spectral pre-processing for this approach are displayed in Table 6. Spectral pre-processing is generally used to reduce noise or eliminate sources of variation which disturb the prediction of target values. Several studies focused on the identification of the best pre-processing approach but there is no general agreement about which pre-processing technique is most effective (Barra, Haefele, Sakrabani, & Kebede, 2021). This assumption is underlined by this study, because there was no uniform pre-processing technique which resulted in best prediction for the tested calibration algorithms and soil parameters. In general, the SVM method resulted in satisfactory model calibrations combining the ten vineyards and two sampling depths. Calibration and validation of SIC content had excellent model robustness and accuracy, where observed and predicted values were consistent and close to the 1:1 line (Fig. 4). While predicted and observed values for hwC were more scattered, the model parameters revealed satisfactory calibration results with excellent R²_{TSV} and RMSE_{TSV} and RPIQ_{TSV}. (Table 6, Fig. 4). For hwC and a benchtop instrument, Vohland, Ludwig, Thiele-Bruhn, and Ludwig (2014) predicted hwC of 60 topsoil samples with a PLSR approach combined with competitive adaptive reweighted sampling with good model performance (R² = 0.85; RMSE = 89 mg kg⁻¹). Wehrle et al. (2021) for the first time used pMIRS to successfully predict hwC content of organic amendments. This study revealed that similar results can be achieved via pMIRS for topsoil and subsoil samples on a dataset with high geopedological variability.

The SVM approach also yielded convenient model robustness and accuracy for model validation of SOC (RPIQ_{TSV}: 5.20; R²_{TSV}: 0.81), which is also visible in the predicted vs. observed plot (Fig. 4). These results are in line with Deiss, Margenot, Culman, and Demyan (2020) where tuning of SVM hyperparameters resulted in highest model performance for SOC prediction at two study sites.

For the prediction of SOC, the SVM model considered the wavebands from 2900 to 2940 cm⁻¹ most important (Fig. 5). These bands are associated with aliphatic C-H stretching and aliphatic OH functional groups (Haberhauer, Rafferty, Strebl, & Gerzabek, 1998; Rumpel, Janik, Skjemstad, & Kögel-Knabner, 2001).

For SOC storage, soil clay content is a parameter of outstanding interest (Hassink, 1997; Wiesmeier et al., 2019). For the prediction of clay and sand content SVM also performed well on the calibration and validation set with excellent accuracy (R²_{TSV}: clay = 0.91; sand = 0.92), robustness (RPIQ_{TSV} clay = 5.76; sand = 6.43) and low prediction error (RMSE_{TSV}: clay = 53.8 g kg⁻¹; sand = 43.9 g kg⁻¹ (Table 6, Fig. 4). Overall, the SVM approach showed reliable prediction results of important C storage parameters for the complex dataset under study.

3.3. Applicability of calibrated models for the monitoring of SOC and hwC

Following the calibration results (section 3.2), we continued with the SVM approach to test the feasibility of our pMIRS models for future monitoring purposes of SOC and hwC. This part of the study was conducted with independent follow-up samples that were taken at the vineyards in Sieb and Spre two years after the initial sampling campaign for the calibration study (dataset 2, Table 2). For the monitoring samples, a PCA revealed similar spectral information (Fig. S2, supplementary material) to the calibration dataset underlining the reproducibility of the MIRS measuring process using compressed soil tablets. These results further underline a good reliability of the sampling method used.

While model performance for SOC (Table 7) showed moderate R² (0.75) but good RPIQ (3.03), the predictions for hwC were excellent with high R² (0.93) and RPIQ (4.79) and low RMSE (0.08 g kg⁻¹).

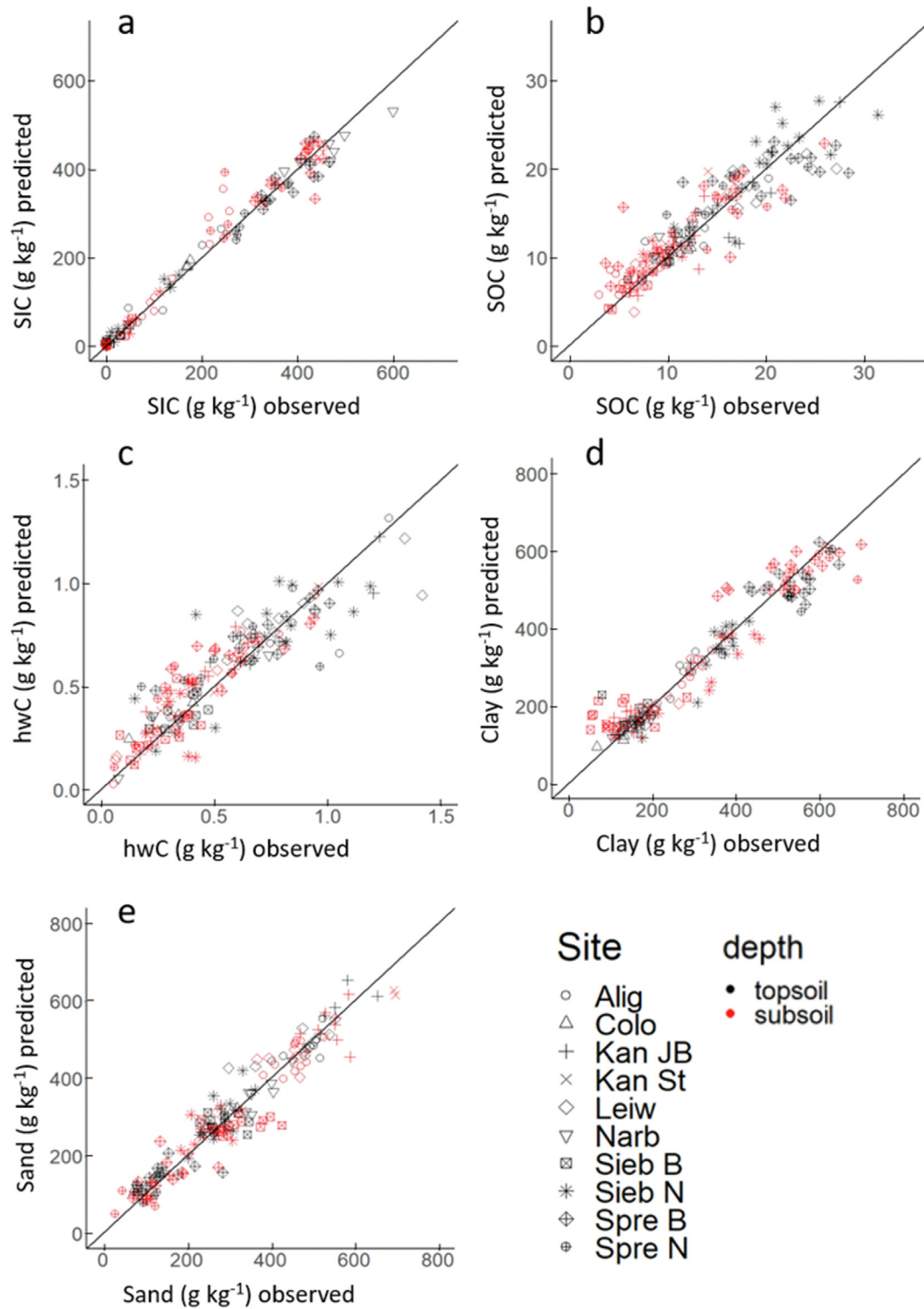


Fig. 4 – Test-set validation of SVM-pMIRS prediction models for soil inorganic carbon (SIC), soil organic carbon (SOC) and hot-water extractable carbon (hwC), clay and sand content for top and subsoils of ten vineyards (dataset 1).

Predicted values for hwC were close to the 1:1 line (Fig. 6), while those for SOC had lower accuracy. Figure 6 also shows no noticeable change in the amount of the labile hwC fraction. The samples for this study originate from

numerous sites with regular (in very most cases annual) application of organic amendments such as compost or pomace. Due to these regular amendments, a kind of flow equilibrium has possibly established, because in vineyards

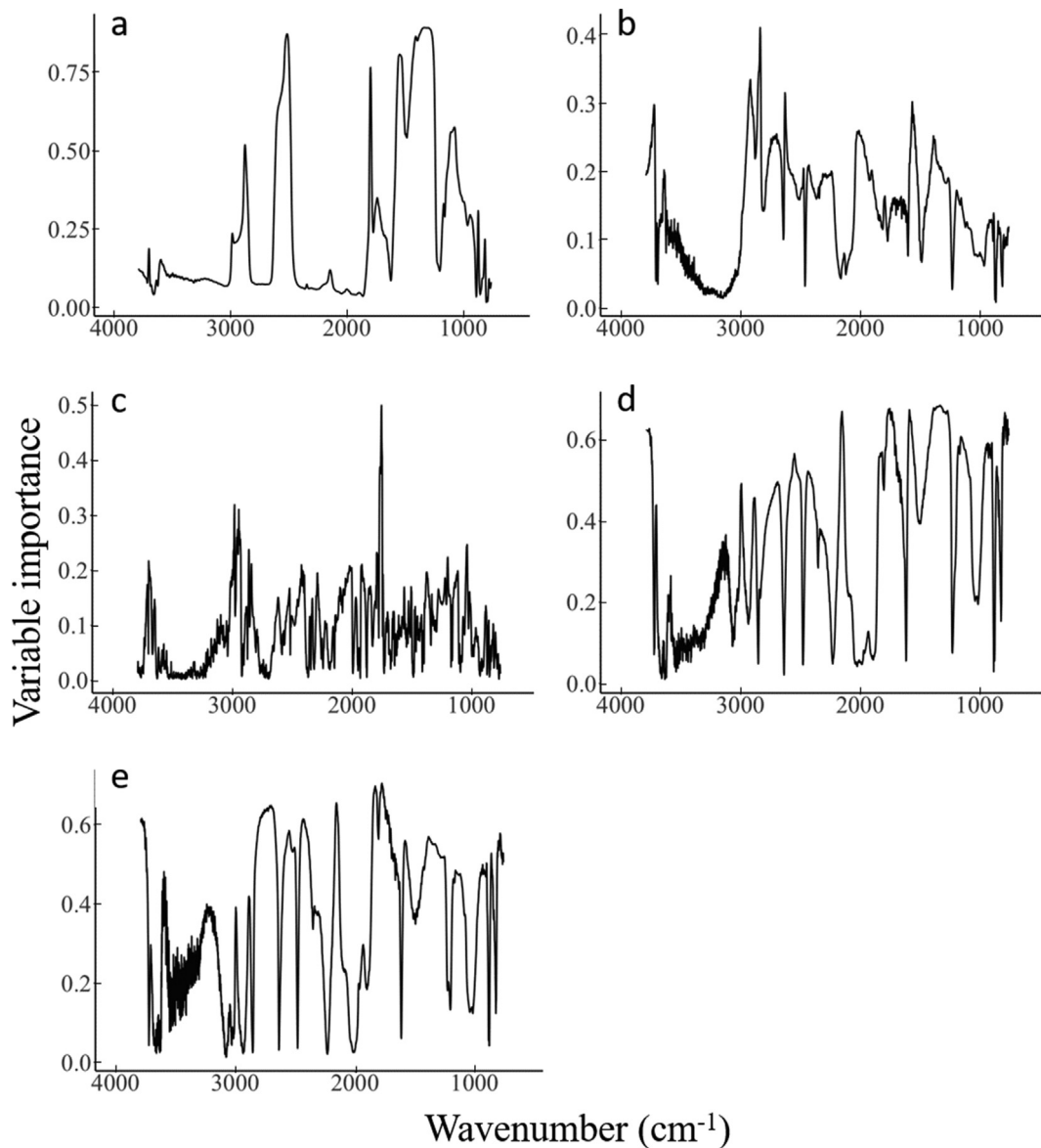


Fig. 5 – Wavenumber importance of support vector machines (SVM) regression determined by recursive feature selection of a) Soil inorganic carbon (SIC), b) soil organic carbon (SOC), c) hot-water extractable carbon (hwc), d) sand and e) clay content.

the mean residence time of labile/active SOC tends to be longer than the application frequency (Dignac et al., 2017; Ramesh et al., 2019).

Despite the constant organic matter input, the prediction of the follow-up samples for dataset 2 performed well. This demonstrates, to our knowledge for the first time, the potential of pMIRS-SVM for C monitoring approaches in future. Yet, the model transferability to independent sites is to be validated, but at least for defined reference sites, the approach is feasible. In a previous project, transferability of MIR prediction models for phosphate was not satisfactory with respect to differences in SOC dynamics at different sites and sampling dates (Pätzold et al., 2020).

Further, SIC is an important factor for C storage, because it influences pH value and aggregation dynamics (Paradelo, Virto, & Chenu, 2015). At the same time, SIC is known to strongly affect spectral soil characteristics. Therefore, we investigated the potential SIC influence on our models (see next section).

3.4. Effects of SIC content on SVM model performance

Because higher SIC content can suppress spectral information regarding SOC content the formation of sample subsets with regard to SIC content could possibly enhance SVM model performance for SOC and hwc. In this section, this influence is

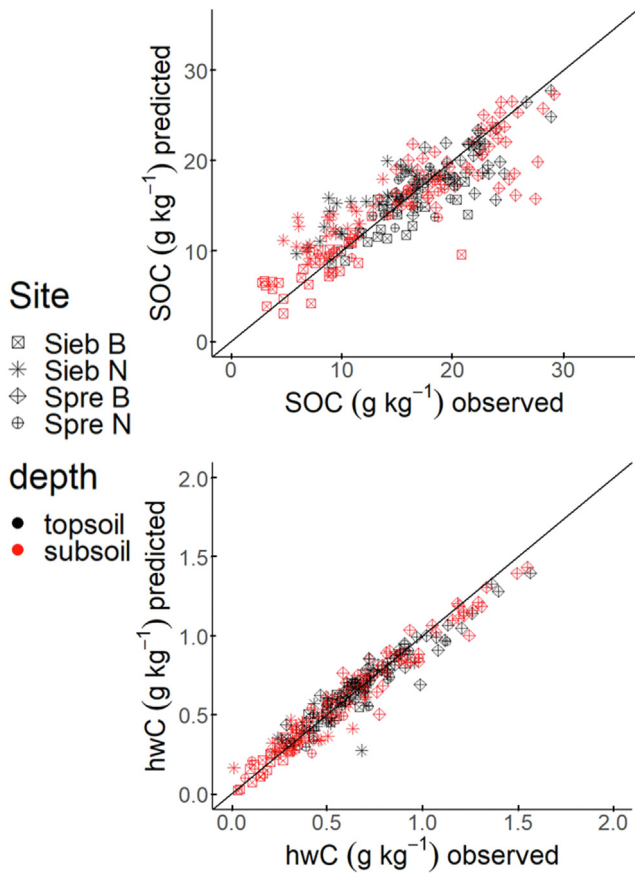


Fig. 6 – Predicted and observed values (n = 243) and 1:1-line of pMIRS SVM models for the monitoring of SOC and hWC for top and subsoils of four vineyards (Dataset 2).

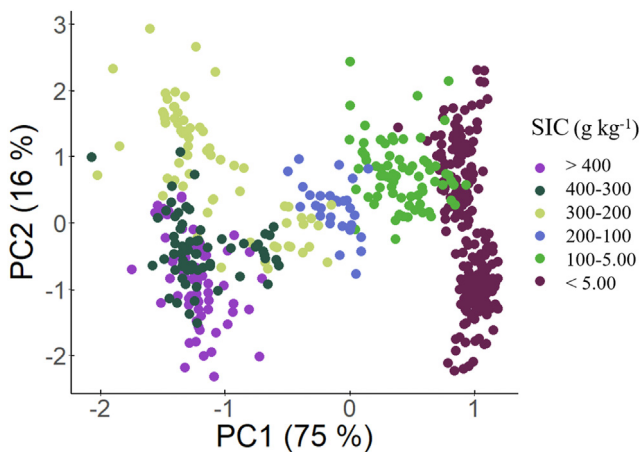


Fig. 7 – Principal components (PC1 and PC2 scores) of multiplicative scatter corrected (MSC) spectra as influenced by SIC content (dataset 1; n = 540).

Table 7 – Results of pMIRS SVM prediction models for the follow-up sampling dataset 2 (n = 243) of top and subsoils in four vineyards for SOC and hWC.

Property (g kg ⁻¹)	RMSE	R ²	RPIQ
SOC	2.98	0.75	3.03
hWC	0.08	0.93	4.79

tested with both datasets 1 (CV and TSV) and 2 (follow-up samples), respectively. Figure 7 shows a strong influence of SIC content on spectral information within the PCA space of dataset 1.

Therefore, the division of dataset with respect to SIC content and irrespective of the sampling site was reasonable for this part of the study. To minimize exaggerated segmentation of data, two groups were formed according to common field practice (>5 g kg⁻¹ SIC and <5 g kg⁻¹ SIC). As a result of this dataset segmentation, sample size and target value range as well as distribution varied between the different subsets (Table 3). Thus, in this section we preferred RPIQ instead of RMSE as a model quality parameter, because RPIQ takes the interquartile distance into consideration and therefore represents the spread of the population (Bellon-Maurel et al., 2010). As expected, the SIC content had a clear effect on model performance (Table 8). Models for SOC and hWC calibrated on samples with SIC content <5 g kg⁻¹ performed better than those for samples with SIC >5 g kg⁻¹, especially for the test-set validation of SOC and dataset 1 (Table 8). Further, models for the segmented dataset with respect to samples with SIC <5 g kg⁻¹ had higher R²_{TSV} for SOC than the uniform model for the entire calibration dataset 1 (R²_{TSV} = 0.81, see Table 6). Nevertheless, the TSV for the entire dataset 1 had higher RPIQ for SOC and therefore has higher model robustness for unknown data. This is confirmed when predicting the later sampling points of dataset 2 (Table 9).

Here model performance for the entire undivided follow-up sampling (dataset 1) was better for both parameters, when comparing RPIQ values of Tables 7 and 9. Therefore, dividing the dataset with respect to SIC is neither beneficial nor necessary with respect to the monitoring of SOC and hWC via pMIRS-SVM models.

Lower model performance for samples of dataset 2 not containing SIC are explained by the limited number of observations. This is in line with Ludwig et al. (2019) who pointed to the importance of sample numbers in SVM modelling. In accordance, Gomez and Coulouma (2018) showed that predicting soil properties that rely on a wide range of spectral features (such as SOC and hWC in our study) performs better at regional than at field scale, i.e. when the model comprises different geopedological settings. Here, the undivided dataset contained most vineyards and a larger number of samples and thus spectral information which made the calibrated models more robust for the unknown monitoring data.

Table 8 – Calibration and test-set validation (TSV) results as well as model hyperparameters for SVM models predicting SOC and hWC of top and subsoils using pMIRS and soil tablets for datasets with SIC <5 (n = 273) and >5 g kg⁻¹ (n = 267) (dataset 1).

Parameter	Dataset	Calibration (CV)		TSV			Hyperparameters*		
		R ²	RPIQ	R ²	RPIQ	s.p.	γ	C	SV
SOC (g kg ⁻¹)	<5 SIC	0.94	6.51	0.90	4.34	SG	0.001	100	108
	>5 SIC	0.82	3.09	0.64	2.34	SG	0.01	100	144
hWC (g kg ⁻¹)	<5 SIC	0.93	5.07	0.91	3.78	1st Der	0.01	100	71
	>5 SIC	0.86	3.95	0.84	3.68	SG	0.001	100	92

*Hyperparameters: s.p. = spectral pre-processing; γ = Gamma, C= Cost; SV= Number of support vectors.

Table 9 – Results of pMIRS SVM models predicting SOC and hWC in monitoring samples (dataset 2, divided into with SIC <5 (n = 42) and >5 g kg⁻¹ (n = 201).

Parameter	Dataset	follow-up sampling	
		R ²	RPIQ
SOC (g kg ⁻¹)	<5 g kg ⁻¹	0.60	2.20
	>5 g kg ⁻¹	0.78	2.92
hWC (g kg ⁻¹)	<5 g kg ⁻¹	0.84	3.58
	>5 g kg ⁻¹	0.86	2.98

4. Conclusions

Variability of geopedological background of the studied vineyards in Western Germany and Southern France and the variation of parameters for SOC storage within and between the fields underline the necessity of a rapid and precise determination method for site-specific evaluation and future monitoring purposes. The chosen approach with combination of portable MIRS instruments, simplified sample preparation, i.e., compressed soil tablets, and SVM as calibration method yielded precise and robust models. The approach is even feasible to reliably predict the rather dynamic hWC content, thus bearing potential for timely resolved studies on C dynamics. With the background of climate change mitigation strategies, this approach provides a valuable foundation for future on-site SOC monitoring programs using pMIRS devices. Soil inorganic carbon content clearly influenced spectral information and model calibrations. Models for SIC <5 g kg⁻¹ samples performed better for the prediction of SOC and hWC than those containing SIC contents >5 g kg⁻¹. Yet, the better model performance could not be maintained for the prediction of follow-up data of later sampling dates. Therefore, a calibration from larger datasets is more appropriate for a future MIRS monitoring program, because the SVM approach can cope with high variation in large and complex datasets. Nevertheless, future research should focus on the role of SIC content on MIR spectral information. Further, the procedure of compressing soil samples may facilitate in-field pMIRS applications, which bears great potential for application in precision agriculture and viticulture.

Funding

This work was funded by the German Federal Ministry of Food and Agriculture (BMEL) within the project "VitiSoil" (grant number: 281B302016).

Conflicts of interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank Florian Ruhnau for hot-water extractions, Henry Dittmer for pMIRS measurements and Andreas Veller for texture analysis. The authors acknowledge Nicolas Saurin and the experimental french INRAE Station of "Pech Rouge" for the help during the field works in "Narbonne" site. The authors would like to thank the German Federal Ministry of Food and Agriculture (BMEL) for the funding of this study and further acknowledge Germany's Excellence Strategy, EXC-2070-390732324-PhenoRob and Wulf Amelung for their support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.biosystemseng.2022.07.012>.

REFERENCES

- Alcántara, V., Don, A., Well, R., & Nieder, R. (2016). Deep ploughing increases agricultural soil organic matter stocks. *Global Change Biology*, 22(8), 2939–2956. <https://doi.org/10.1111/gcb.13289>
- Backes, Böhm, Gröber, Jung, & Spies. (2015). *Vineyard soil of Rhineland-Palatinate. Ministry of Economic Affairs, Climate Protection, Energy and spatial planning, Rhineland-Palatinate.*
- Barra, I., Haefele, S. M., Sakrabani, R., & Kebede, F. (2021). Soil spectroscopy with the use of chemometrics, machine learning

- and pre-processing techniques in soil diagnosis: Recent advances—A review. *TrAC, Trends in Analytical Chemistry*, 135, Article 116166. <https://doi.org/10.1016/j.trac.2020.116166>
- Barthès, B. G., Kouakoua, E., Moulin, P., Hmaid, K., Gallali, T., Clairotte, M., et al. (2016). Studying the physical protection of soil carbon with quantitative infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 24(3), 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. *TrAC, Trends in Analytical Chemistry*, 29(9), 1073–1081. <https://doi.org/10.1016/j.trac.2010.05.006>
- Breure, T. S., Prout, J. M., Haefele, S. M., Milne, A. E., Hannam, J. A., Moreno-Rojas, S., et al. (2022). Comparing the effect of different sample conditions and spectral libraries on the prediction accuracy of soil properties from near- and mid-infrared spectra at the field-scale. *Soil and Tillage Research*, 215, Article 105196. <https://doi.org/10.1016/j.still.2021.105196>
- Costantini, E. A. C., Castaldini, M., Diago, M. P., Giffard, B., Lagomarsino, A., Schroers, H.-J., et al. (2018). Effects of soil erosion on agro-ecosystem services and soil functions: A multidisciplinary study in nineteen organically farmed European and Turkish vineyards. *Journal of Environmental Management*, 223, 614–624. <https://doi.org/10.1016/j.jenvman.2018.06.065>
- Coulouma, G., Caner, L., Loonstra, E. H., & Lagacherie, P. (2016). Analysing the proximal gamma radiometry in contrasting Mediterranean landscapes: Towards a regional prediction of clay content. *Geoderma*, 266, 127–135. <https://doi.org/10.1016/j.geoderma.2015.12.006>
- Deiss, L., Margenot, A. J., Culman, S. W., & Demyan, M. S. (2020). Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. *Geoderma*, 365, Article 114227. <https://doi.org/10.1016/j.geoderma.2020.114227>
- Dignac, M.-F., Derrien, D., Barré, P., Barot, S., Cécillon, L., Chenu, C., et al. (2017). Increasing soil carbon storage: Mechanisms, effects of agricultural practices and proxies. A review. *Agronomy for Sustainable Development*, 37(2). <https://doi.org/10.1007/s13593-017-0421-2>
- Ghani, A., Dexter, M., & Perrott, K. (2003). Hot-water extractable carbon in soils: A sensitive measurement for determining impacts of fertilisation, grazing and cultivation. *Soil Biology and Biochemistry*, 35(9), 1231–1243. [https://doi.org/10.1016/S0038-0717\(03\)00186-X](https://doi.org/10.1016/S0038-0717(03)00186-X)
- Gholizadeh, A., Borůvka, L., Saberioon, M., & Vašat, R. (2013). Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: State-of-the-art and key issues. *Applied Spectroscopy*, 67(12), 1349–1362. <https://doi.org/10.1366/13-07288>
- Gomez, C., Chevallier, T., Moulin, P., Arrouays, D., & Barthès, B. G. (2022). Using carbonate absorbance peak to select the most suitable regression model before predicting soil inorganic carbon concentration by mid-infrared reflectance spectroscopy. *Geoderma*, 405, Article 115403. <https://doi.org/10.1016/j.geoderma.2021.115403>
- Gomez, C., & Coulouma, G. (2018). Importance of the spatial extent for using soil properties estimated by laboratory VNIR/SWIR spectroscopy: Examples of the clay and calcium carbonate content. *Geoderma*, 330, 244–253. <https://doi.org/10.1016/j.geoderma.2018.06.006>
- Haberhauer, G., Rafferty, B., Strebl, F., & Gerzabek, M. H. (1998). Comparison of the composition of forest soil litter derived from three different sites at various decompositional stages using FTIR spectroscopy. *Geoderma*, 83(3–4), 331–342. [https://doi.org/10.1016/S0016-7061\(98\)00008-1](https://doi.org/10.1016/S0016-7061(98)00008-1)
- Hassink, J. (1997). *Plant and Soil*, 191(1), 77–87. <https://doi.org/10.1023/A:1004213929699>
- Hutengs, C., Seidel, M., Oertel, F., Ludwig, B., & Vohland, M. (2019). In situ and laboratory soil spectroscopy with portable visible-to-near-infrared and mid-infrared instruments for the assessment of organic carbon in soils. *Geoderma*, 355, Article 113900. <https://doi.org/10.1016/j.geoderma.2019.113900>
- Jakšić, S., Ninkov, J., Milić, S., Vasin, J., Banjac, D., Milić, S., & Jakšić, D. (2021). The state of soil organic carbon in vineyards as affected by soil types and fertilization strategies (Tri Morave region, Serbia). *Agronomy*, 11(1), 9. <https://doi.org/10.3390/agronomy11010009>
- Janik, L. J., Soriano-Disla, J. M., & Forrester, S. T. (2020). Feasibility of handheld mid-infrared spectroscopy to predict particle size distribution: Influence of soil field condition and utilisation of existing spectral libraries. *Soil Research*, 58(6), 528. <https://doi.org/10.1071/SR20097>
- Janik, L. J., Soriano-Disla, J. M., Forrester, S. T., & McLaughlin, M. J. (2016). Effects of soil composition and preparation on the prediction of particle size distribution using mid-infrared spectroscopy and partial least-squares regression. *Soil Research*, 54(8), 889. <https://doi.org/10.1071/SR16011>
- Ji, Adamchuk, Biswas, Dhawale, Sudarsan, Zhang, ... Shi. (2016). Assessment of soil properties in situ using a prototype portable MIR spectrometer in two agricultural fields. *Biosystems Engineering*, 152, 14–27. <https://doi.org/10.1016/j.biosystemseng.2016.06.005>. In this issue.
- Körschens, M. (2010). Der organische Kohlenstoff im Boden (C org) – bedeutung, Bestimmung, Bewertung. Soil organic carbon (C org) – importance, determination, evaluation. *Archives of Agronomy and Soil Science*, 56(4), 375–392. <https://doi.org/10.1080/03650340903410246>
- Körschens, M., Weigel, A., & Schulz, E. (1998). Turnover of soil organic matter (SOM) and long-term balances - tools for evaluating sustainable productivity of soils. *Zeitschrift für Pflanzenernährung und Bodenkunde*, 161(4), 409–424. <https://doi.org/10.1002/jpln.1998.3581610409>
- Kuhn, M. (2022). *The package "caret". Reference manual. The comprehensive R Archive Network*. <https://cran.r-project.org/web/packages/caret/caret.pdf>.
- Lal, R. (2004). Soil carbon sequestration impacts on global climate change and food security. *Science (New York, N.Y.)*, 304(5677), 1623–1627. <https://doi.org/10.1126/science.1097396>
- Leenen, M., Welp, G., Gebbers, R., & Pätzold, S. (2019). Rapid determination of lime requirement by mid-infrared spectroscopy: A promising approach for precision agriculture. *Journal of Plant Nutrition and Soil Science*, 182(6), 953–963. <https://doi.org/10.1002/jpln.201800670>
- Ludwig, B., Murugan, R., Parama, V. R. R., & Vohland, M. (2019). Accuracy of estimating soil properties with mid-infrared spectroscopy: Implications of different chemometric approaches and software packages related to calibration sample size. *Soil Science Society of America Journal*, 83(5), 1542–1552. <https://doi.org/10.2136/sssaj2018.11.0413>
- Marmette, M.-C., & Adamchuk, V. (2020). *Comparison of the performance of laser-induced breakdown spectroscopy and color, visible, near-infrared and mid-infrared spectroscopy in the prediction of various soil properties*. McGill University, Department of Biosource Engineering.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingassel, A., Leich, F., Chang, C.-C., et al. (2020). *R package e1071: Misc functions of the Departem of Statistics. Probability Theory Group. TU Wien: Version 1.7-4*. Retrieved from <https://CRAN.R-project.org/package=e1071>.
- Minasny, B., Malone, B. P., McBratney, A. B., Angers, D. A., Arrouays, D., Chambers, A., et al. (2017). Soil carbon 4 per mille. *Geoderma*, 292, 59–86. <https://doi.org/10.1016/j.geoderma.2017.01.002>

- Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>
- Napoli, M., Marta, A. D., Zanchi, C. A., & Orlandini, S. (2017). Assessment of soil and nutrient losses by runoff under different soil management practices in an Italian hilly vineyard. *Soil and Tillage Research*, 168, 71–80. <https://doi.org/10.1016/j.still.2016.12.011>
- Nguyen, Janik, & Raupach. (1991). Diffuse reflectance infrared fourier transform (DRIFT) spectroscopy in soil studies. *Australian Journal of Soil Research*, 29(1), 49–67. <https://doi.org/10.1071/SR9910049>. In this issue.
- Novara, A., Favara, V., Novara, A., Francesca, N., Santangelo, T., Columba, P., et al. (2020). Soil carbon budget account for the sustainability improvement of a mediterranean vineyard area. *Agronomy*, 10(3), 336. <https://doi.org/10.3390/agronomy10030336>
- Paradelo, R., Virto, I., & Chenu, C. (2015). Net effect of liming on soil organic carbon stocks: A review. *Agriculture, Ecosystems & Environment*, 202, 98–107. <https://doi.org/10.1016/j.agee.2015.01.005>
- Pätzold, S., Leenen, M., Frizen, P., Heggemann, T., Wagner, P., & Rodionov, A. (2020). Predicting plant available phosphorus using infrared spectroscopy with consideration for future mobile sensing applications in precision farming. *Precision Agriculture*, 21(4), 737–761. <https://doi.org/10.1007/s11119-019-09693-3>
- Ramesh, T., Bolan, N. S., Kirkham, M. B., Wijesekara, H., Kanchikerimath, M., Srinivasa, R. C., et al. (2019). Soil organic carbon dynamics: Impact of land use changes and management practices: A review. *Advances in Agronomy*, 156, 1–107. <https://doi.org/10.1016/bs.agron.2019.02.001>
- Rodionov, A., Pätzold, S., Welp, G., Pallares, R. C., Damerow, L., & Amelung, W. (2014). Sensing of soil organic carbon using visible and near-infrared spectroscopy at variable moisture and surface roughness. *Soil Science Society of America Journal*, 78(3), 949–957. <https://doi.org/10.2136/sssaj2013.07.0264>
- Rumpel, C., Janik, L. J., Skjemstad, J. O., & Kögel-Knabner, I. (2001). Quantification of carbon derived from lignite in soils using mid-infrared spectroscopy and partial least squares. *Organic Geochemistry*, 32(6), 831–839. [https://doi.org/10.1016/s0146-6380\(01\)00029-8](https://doi.org/10.1016/s0146-6380(01)00029-8)
- Stevens, A., Ramirez-Lopez, L., & Hans, G. (2020). R package *prospectr*: Miscellaneous functions for processing and sample selection of spectroscopic data. Retrieved from Version 0.2.1. <https://CRAN.R-project.org/package=prospectr>.
- Stumpe, B., Weihermüller, L., & Marschner, B. (2011). Sample preparation and selection for qualitative and quantitative analyses of soil organic carbon with mid-infrared reflectance spectroscopy. *European Journal of Soil Science*, 62(6), 849–862. <https://doi.org/10.1111/j.1365-2389.2011.01401.x>
- Viscarra Rossel, R. A., & Bouma, J. (2016). Soil sensing: A new paradigm for agriculture. *Agricultural Systems*, 148, 71–74. <https://doi.org/10.1016/j.agsy.2016.07.001>
- Vohland, M., Ludwig, M., Thiele-Bruhn, S., & Ludwig, B. (2014). Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. *Geoderma*, 223–225, 88–96. <https://doi.org/10.1016/j.geoderma.2014.01.013>
- 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. *European Journal of Soil Science*, 70(3), 550–564. <https://doi.org/10.1111/ejss.12787>
- Wadoux, A. M.-C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, 210, Article 103359. <https://doi.org/10.1016/j.earscirev.2020.103359>
- Wehrle, R., Welp, G., & Pätzold, S. (2021). Total and hot-water extractable organic carbon and nitrogen in organic soil amendments: Their prediction using portable mid-infrared spectroscopy with support vector machines. *Agronomy*, 11(4), 659. <https://doi.org/10.3390/agronomy11040659>
- Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., et al. (2020). R package *ggplot2*: Create elegant data visualisations using the grammar of graphics. Retrieved from Version 3.3.3. <https://CRAN.R-project.org/package=ggplot2>.
- Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., Lützw, M. von, Marin-Spiotta, E., et al. (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>
- Wordell-Dietrich, P. (2021). *Stability of soil organic carbon in the subsoil*.
- Zmora-Nahum, S., Markovitch, O., Tarchitzky, J., & Chen, Y. (2005). Dissolved organic carbon (DOC) as a parameter of compost maturity. *Soil Biology and Biochemistry*, 37(11), 2109–2116. <https://doi.org/10.1016/j.soilbio.2005.03.013>