

Robustness of visible near-infrared and mid-infrared spectroscopic models to changes in the quantity and quality of crop residues in soil

Isabel Greenberg¹  | Deborah Linsler¹ | Michael Vohland²  | Bernard Ludwig¹ 

¹Department of Environmental Chemistry, University of Kassel, Nordbahnhofstr. 1a, Witzenhausen, 37213, Germany

²Geoinformatics and Remote Sensing, Institute for Geography, Leipzig University, Johannisallee 19a, Leipzig, 04103, Germany

Correspondence

Isabel Greenberg, Department of Environmental Chemistry, University of Kassel, Nordbahnhofstr. 1a, 37213 Witzenhausen, Germany
Email: isabel.greenberg@uni-kassel.de

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Numbers: LU 583/19-1, VO 1509/7-1

Abstract

The robustness of soil organic carbon (SOC) and total nitrogen (TN) content prediction accuracy by visible near-infrared spectroscopy (visNIRS) and mid-infrared spectroscopy (MIRS) models after a change in the quantity or quality of crop residues requires investigation. Arable soils (0–20 cm) from 20 locations across Germany were collected, and 0, 2, 4, or 8 g C kg soil⁻¹ of wheat straw (C/N ratio, 54) or clover (C/N ratio, 13) were added. Before and after a 56-d incubation, dried and ground samples were measured for SOC and TN content and by visNIRS and MIRS. The complete dataset ($n = 280$) was subdivided into calibration and validation datasets to test the robustness of partial least squares regression models to changes in crop residue quantity and quality in soil. Noise-reducing data pretreatments included region selection, moving averages, resampling every second data point, and the Savitzky-Golay algorithm. The MIRS estimates for SOC (7.4–33 g kg⁻¹) had lower root mean squared error of validation (RMSE_v = 0.9–2.9 g kg⁻¹) compared with visNIRS (RMSE_v = 1.6–7.1 g kg⁻¹). Total N estimates (0.7–2.8 g kg⁻¹) were more comparable for MIRS (RMSE_v = 0.1–0.3 g kg⁻¹) and visNIRS (RMSE_v = 0.1–1.0 g kg⁻¹). Loadings of partial least squares regression components suggested the predictive mechanisms for SOC and TN were more similar for visNIRS than for MIRS. Differing crop residue quantity or quality in calibration versus validation resulted in biased SOC and TN estimates by visNIRS and MIRS models. However, calibration with a global residue model containing all soils and crop residue quantities and qualities lowered RMSE_v for SOC and TN prediction with visNIRS and MIRS, demonstrating the usefulness of this approach.

Abbreviations: MIR, mid-infrared; MIRS, mid-infrared spectroscopy; RPIQ, ratio of performance to interquartile distance; RPIQ_v, ratio of performance to interquartile distance of validation; SOC, soil organic carbon; TN, total nitrogen; PLSR, partial least squares regression; RMSE_v, root mean squared error of validation; RPD, ratio of performance to deviation; RPD_v, ratio of prediction to deviation of validation; visNIR, visible near-infrared; visNIRS, visible near-infrared spectroscopy.

1 | INTRODUCTION

Infrared spectroscopy is a method to rapidly and cost-effectively gain information about soil properties (Kuang et al., 2012). Although estimates of soil properties may be less accurate than traditional laboratory methods, the relative ease of acquiring high-density information regarding

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. Soil Science Society of America Journal © 2020 Soil Science Society of America

spatially and temporally heterogeneous soil properties based on a calibration model is a clear advantage in certain contexts, such as for soil mapping and long-term monitoring studies (Stevens et al., 2008).

Key chemical, biological, and physical soil properties have been estimated by both visible near-infrared spectroscopy (visNIRS) and mid-infrared spectroscopy (MIRS) reliably and often with excellent accuracy in independent validations when the calibration model, containing both spectra and values measured in the laboratory, adequately represents the validation dataset (Soriano-Disla, Janik, Viscarra Rossel, Macdonald, & McLaughlin, 2014). This can be attributed to several soil components of interest being spectrally active in the visible near-infrared (visNIR) and mid-infrared (MIR) regions (e.g., the interaction of C-H, C-O, and C-N groups enables the prediction of soil organic carbon [SOC] and nitrogen [N] content).

Calibration stability (i.e., the robustness of models to application in altered contexts) determines the frequency with which models must be recalibrated (Stevens, van Wesemael, Vandenschrick, Touré, & Tychon, 2006) and thus the efficiency of spectroscopy in comparison to traditional laboratory methods. Soil type and mineralogy, but also management aspects, such as land use, tillage, fertilization, and handling of crop residues, may need to be considered in creating a suitable calibration model (Araújo, Wetterlind, Demattê, & Stenberg, 2014; Castaldi et al., 2018; Zeng et al., 2016). In this context, several studies have been conducted to determine whether visNIRS and MIRS can be used to estimate SOC content with sufficient accuracy to distinguish between the effects of various management strategies and document change in C stocks, with implications for both soil productivity and climate change. The success of these models depends largely on the site-specificity of the calibrations (Bellon-Maurel & McBratney, 2011). For instance, Baldock, Beare, Curtin, and Hawke (2018) reported for a split-plot field trial on a silt-loam Udic Dystrocrept in New Zealand with cultivation as a main-plot factor and presence of a winter cover crop as a sub-plot factor that when samples collected from the same field and sampling date were included in both model calibration and validation, SOC stocks across diverse land-use and management systems could be predicted by an MIR partial least squares regression (PLSR) model with a root mean squared error of validation ($RMSE_V$) of 1.08 g kg^{-1} for soils with SOC contents ranging from 10.2 to 38.4 g kg^{-1} and a ratio of prediction to deviation of validation (RPD_V) of 3.02. Two-way ANOVA investigating the effect of measurement method (laboratory measured vs. MIR-PLSR predicted) and agricultural management on SOC stocks found no significant effect of the measurement method, whereas a significant management effect was detected. Thus, MIR-PLSR models were recommended as a means for analysis of land management effects on SOC stocks.

Core Ideas

- MIR-PLSR models were more robust than visNIR-PLSR models for the prediction of SOC content with altered quantities and qualities of crop residues in soil.
- visNIR and MIR models had more comparable accuracy for TN estimates.
- Loadings of PLSR components suggested the predictive mechanisms for SOC and TN were more similar for visNIRS than MIRS.
- Differences in residue quantity and quality between calibration and validation datasets caused systematic bias in SOC and TN estimates.
- Use of a global residue calibration model containing soils with all crop residue quantities and qualities lowered validation RMSE of SOC and TN estimates.

However, validation results were more mixed in cases where a broad calibration dataset was used but no samples from the same field and date were present in both calibration and validation datasets. Madhavan, Kitching, Mendham, Weston, and Baker (2016) investigated SOC and total N (TN) content prediction accuracy using MIR-PLSR models for 31 paired pasture and eucalyptus plantation sites across southwestern Australia (total $n = 177$ from 23 paired sites for calibration, total $n = 62$ from eight paired sites for validation) and found that SOC and TN could be estimated in validation with very good accuracy (SOC: $RPD_V = 3.21$ and $RMSE_V = 0.4 \text{ g kg soil}^{-1}$ for soils with contents ranging from 16.8 to 89.9 g kg^{-1} ; TN: $RPD_V = 2.88$ and $RMSE_V = 0.12 \text{ g kg}^{-1}$ for soils with TN contents ranging from 0.70 to 5.78 g kg^{-1}). However, Page, Dalal, and Dang (2013) found that a MIR-PLSR model calibrated with 20,195 samples across 4000 sites in Australia and validated with a Vertosol at an independent site with various tillage, residue, and fertilization treatments was able to detect losses in SOC stocks at that site over a 28-yr period, but the magnitude was greatly underestimated: losses of 7.9 Mg C ha^{-1} in the 0- to 30-cm depth (from 63.5 to $55.5 \text{ Mg C ha}^{-1}$) were underestimated by 27%. Furthermore, there were disagreements between the laboratory measurements and model predictions as to the effects of various management strategies on SOC stocks: laboratory measurements found stubble retention and N fertilization decreased SOC loss, whereas MIR-PLSR models predicted burning stubble and no N fertilization decreased SOC loss to the greatest extent. Finally, in the case of Stevens et al. (2008), who used portable visNIR measurements for the prediction of SOC by PLSR, an independent validation ($n = 37$) based

on a calibration ($n = 99$) with samples collected just 2 yr apart at the same site delivered poor results ($RPD_V = 1.09$ and $RMSE_V = 4.6 \text{ g kg}^{-1}$ for SOC ranging from 5.7 to 22.8 g kg^{-1}), in part due to the presence of vegetation residues lying on the soil surface at the time field spectral measurements were taken. They recommended performing a calibration before each measurement campaign due to the sensitivity of the method to small changes in the field conditions at sampling (e.g., moisture content, roughness, and vegetation). It has also been emphasized that bias should be calculated to evaluate model performance due to the common occurrence of a high correlation between measured and predicted values but consistent over- or underestimation (Bellon-Maurel & McBratney, 2011; Seidel, Hutengs, Ludwig, Thiele-Bruhn, & Vohland, 2019).

Therefore, more information about visNIRS and MIRS model robustness to specific differences between calibration and validation datasets would support the selection of appropriate calibration datasets for the estimation of soil properties. This is especially relevant because in situ visNIR and MIR spectrometers are increasingly used due to their efficiency gains, but the results can be greatly affected by conditions at the time of sampling (Reeves, McCarty, & Hively, 2010; Stevens et al., 2008). Thus, the objective of this study was to isolate the effect of a common source of heterogeneity in agricultural soils (i.e., changes in the quantity and quality of crop residue) on the robustness of visNIR and MIR-PLSR models used to predict SOC and TN content. In addition, the robustness of a global residue calibration model, featuring diverse residue quantities and qualities, was tested.

2 | MATERIALS AND METHODS

2.1 | Soils

Twenty surface soils (0–20 cm) were sampled in March and April 2012 from 20 arable, long-term monitoring fields across Lower Saxony, Germany with different soil types and textures. Fourteen of these sites were under conventional management with two- to four-season crop rotations including sugar beet, cereals, canola, potatoes, maize, and/or legumes; four sites were under organic management with longer, diverse crop rotations; and two sites rotated fallow with flax or pasture, respectively. Prior to soil analysis, visible crop residues were removed from the samples, but small crop fragments of diverse origin may have remained. The SOC content of the soils ranged from 7.5 to 26 g C kg soil^{-1} . Sand, silt, and clay contents ranged from 2.4 to 90.9%, from 3.8 to 87.2%, and from 0.5 to 35.0%, respectively, and the soil textures were sand ($n = 1$), loamy sand ($n = 6$), sandy loam ($n = 3$), silt ($n = 1$), silt loam ($n = 5$), loam ($n = 1$), silty clay loam ($n = 2$), and clay ($n = 1$). The sampled soil horizons were

plow horizons ($n = 13$), eluvial plow horizons ($n = 5$), a plaggic plow horizon ($n = 1$), and a plow horizon composed of material transported by fluvial and colluvial processes ($n = 1$) (Soil Survey Staff, 2014).

2.2 | Residue treatments

Each of the 20 soils was sieved to 2 mm and subdivided into seven 70-g subsamples, creating a total of 140 samples. One of the seven subsamples of each soil was a control treatment with no residue additions. Three subsamples were mixed with wheat straw (*Triticum aestivum* L.; 45% C; C/N ratio, 53.7) at rates of 2, 4, and 8 g C kg^{-1} soil, and another three subsamples were mixed with white clover (*Trifolium repens* L.; 42% C; C/N ratio, 12.5) at rates of 2, 4, and 8 g C kg^{-1} soil. The residues were just the aboveground plant material and did not include flowers or seed heads. The residues were dried for 48 h at 40 °C and milled to <2 mm using a Fritsch Pulverisette 19 Cutting Mill prior to weighing and incorporation in the soil. Given the C/N ratios of the crop residues, N additions were 4.3 times higher for clover versus wheat straw treatments. The selected residue application densities are equivalent to approximately 3–12 t clover or straw ha^{-1} on a dry matter basis (assuming a 0- to 5-cm depth of incorporation and 1.3 g cm^{-3} soil density). These application rates were therefore selected in order to span the range of residue amounts that might reasonably be expected under agricultural management (Agriculture and Horticulture Development Board, 2016; European Commission, 2007; FAO, 2017; Widdup et al., 2015). Given the range in initial SOC contents of the 20 soils prior to residue additions (7.4–26 g C kg soil^{-1}), 8 g C kg soil^{-1} residue additions approximately doubled the SOC content of some soils (three loamy sands and a sandy loam) but only represented about 25% of the SOC content of other soils (two loamy sands).

2.3 | Laboratory analysis

Laboratory incubations were conducted by placing 70 g of the field-moist subsamples in 120-ml plastic beakers. Distilled water was added to bring the soil moisture to 60% of the soil water-holding capacity (Linn & Doran, 1984). The plastic beakers were then placed into 1.5-L glass jars, and the soil was incubated in a climate cabinet (Kühlbrutschrank ICP 800, Memmert) for 56 d at 20 °C. Cumulative $\text{CO}_2\text{-C}$ emissions (and thus the loss of labile C) ranged from 0.06 (control treatment of a loamy sand) to 4.69 g C kg soil^{-1} (8 g clover-C addition kg soil^{-1} treatment of a sand; data not shown).

Before and after the incubation, the SOC and TN contents of the soil were determined. First, all the soil samples were dried at 40 °C for 48 h and ball-milled. The total C (TC)

and TN concentrations were measured by dry combustion with a CN elemental analyzer (Elementar Vario El, Heraeus). Contents of inorganic C were determined with the Scheibler-Method. Briefly, 10% HCl was added to the dried, ball-milled soil, and the evolved CO₂ was measured volumetrically. Soil organic C content was calculated by subtracting the inorganic C from TC.

2.4 | Spectral measurements

First, each of the soils and soil–residue mixtures before and after incubation ($n = 280$) were freeze-dried for 3 d (Alpha 2–4, Christ GmbH) and subsequently ball-milled to a particle size <0.2 mm using a Retsch MM 400 with 10 zirconium oxide balls at 30 Hertz for 5 min. The visNIR and MIR spectral measurements were made on two laboratory replicates of each soil and soil–residue mixture both before and after incubation as well as on two laboratory replicates of the wheat straw and white clover residues (dried, milled) in the absence of soil. The reflectance values of laboratory replicates were averaged and converted to absorbance ($\log [1/\text{reflectance}]$).

2.4.1 | Visible near-infrared spectral measurements

The visNIR spectra in the range of 400–2500 nm (25,000–4000 cm⁻¹) were measured at 2-nm intervals with 32 co-added scans on ~10 g of sample filled into a cell (diameter, 5 cm) with a quartz window using a Foss XDS Rapid Content Analyzer. Due to the presence of instrumental artifacts from 400 to 500 nm (Stevens, Nocita, Tóth, Montanarella, & van Wesemael, 2013) and a detector change at 1100 nm (Si detector from 400 to 1100 nm, PbS detector from 1100 to 2500 nm), only the wavelengths from 502 to 1092 and from 1108 to 2498 nm were used in the analysis, resulting in 992 data points per spectra. Figure 1 shows the average spectra of the crop residues, soils, and soil–residue mixtures before and after incubation.

2.4.2 | Mid-infrared spectral measurements

Diffuse reflectance infrared Fourier transform spectra of the samples (~1.5 g) in the range of 7000–370 to cm⁻¹ were recorded with a Bruker-TENSOR 27 MIR spectrometer with an A562 integrating sphere detector and the diffuse-reflectance accessory (Ulbricht-Kugel). The range from 4000 to 370 cm⁻¹ (2500–27,027 nm) is the MIR region; 7000–4000 cm⁻¹ (1429–2500 nm) is longwave NIR and was therefore excluded from the analysis. Potassium bromide was not added to the soils. The spectra were measured with 200 scans

at approximately 2 cm⁻¹ intervals, resulting in 1883 data points per spectra.

2.5 | Formation of calibration and validation datasets

To investigate the robustness of spectral models to changes in the quantity and quality of crop residues in the soil, three analysis approaches were carried out using various subdivisions of the complete dataset, as outlined in Table 1 and described below.

2.5.1 | Analysis 1: Residue quantity experiment

To investigate model robustness to the incorporation of crop residues in soil, a dataset containing all samples with a low quantity of crop residues (0 and 2 g C kg soil⁻¹ residue treatments, $n = 120$) and another containing all samples with a high quantity of crop residues (4 and 8 g C kg soil⁻¹ residue treatments, $n = 160$) were separated (including both wheat straw and clover residues as well as pre- and postincubation treatments). Analysis was carried out using the low–residue quantity dataset for model calibration and the high–residue quantity dataset for model validation.

To investigate model robustness to decomposition of crop residues in soil, a dataset containing all preincubation samples ($n = 140$) and another containing all postincubation samples ($n = 140$) were separated (including soils with both clover and wheat straw residues of all residue application rates as well as control soils without residue addition). Analysis was carried out using the preincubation dataset for model calibration and the postincubation dataset for model validation.

2.5.2 | Analysis 2: Residue quality experiment

To investigate model robustness to changes in the quality of crop residues in the soil, a dataset containing all samples with white clover residues ($n = 120$) and another containing all samples with wheat straw residues ($n = 120$) were separated (including both pre- and postincubation treatments). Analysis was carried out using the clover residues dataset for model calibration and the wheat straw dataset for model validation, and vice versa.

2.5.3 | Analysis 3: Global residue model experiment

To investigate the robustness of a global calibration dataset, two random subdivisions of the complete dataset were created: (a) selection of a random two-thirds of the complete

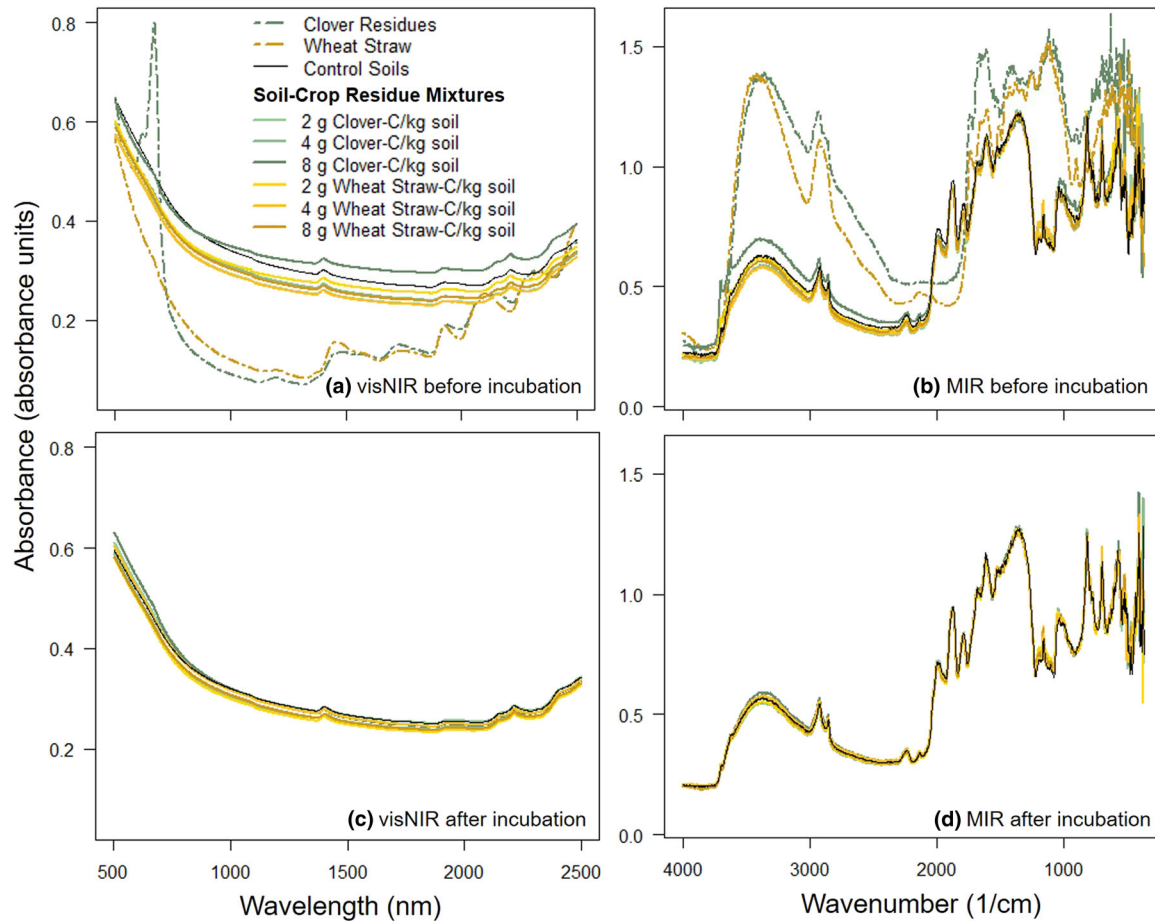


FIGURE 1 Average visible near-infrared (visNIR) and mid-infrared (MIR) spectra of crop residues, soils, and soil–crop residue mixtures before and after incubation.

dataset for calibration ($n = 187$) and the remaining one-third of the dataset for validation ($n = 93$) and (b) selection of a random two-thirds of the soils collected from 20 sites for calibration (i.e., 13 soils, including all associated residue and incubation treatments; $n = 182$) and the remaining one-third of the soils collected from 20 sites for validation (i.e., seven soils, including all associated residue and incubation treatments, $n = 98$). Thus, in the former case, all 20 soils were present in both calibration and validation datasets, whereas in the latter, the 13 soils in the calibration dataset differed from the seven soils in the validation dataset.

2.6 | Chemometric approach

Separate PLSR analyses were performed for the visNIR and MIR spectral ranges with the statistical software R (version 3.4.4; R Core Team, 2018). Due to its ease of implementation, PLSR is regarded as the standard modeling approach in spectroscopy (Soriano-Disla et al., 2014). Data pretreatment was carried out with the `prospectr` package, and PLSR was performed using the `pls` package.

Data pretreatment began with an automatic selection of all possible combinations of six regions making up the complete spectra. For visNIRS, these regions ranged from 500 to 834 nm (Region 1), from 834 to 1167 (Region 2), from 1167 to 1500 (Region 3), from 1500 to 1833 (Region 4), from 1833 to 2166 (Region 5), and from 2166 to 2500 (Region 6). For MIRS, these regions ranged from 4000 to 3682.1 cm^{-1} (Region 1), from 3682.1 to 3020.5 (Region 2), from 3020.5 to 2358.9 (Region 3), from 2358.9 to 1693.5 (Region 4), from 1693.5 to 1030 (Region 5), and from 1030 to 368.4 (Region 6). This approach, implemented in the statistical software R to imitate data pretreatment carried out by the OPUS Quant 2 software (Bruker Optik GmbH), divides the spectra into regions of equal size rather than on the basis of spectral features. The latter may be preferable due to an exclusion of irrelevant information. However, the approach using an automatic selection of different regions of equal sizes has been shown to improve SOC estimation accuracy by MIR-PLSR compared with use of the full spectral region (Ludwig, Murugan, Parama, & Vohland, 2019).

Each of the 63 possible region combinations was then tested in conjunction with the following mathematical

TABLE 1 Descriptive statistics of the datasets used for soil organic carbon (SOC) and total nitrogen (TN) content estimation according to the three analysis approaches

Analysis	Datasets	Sample size	Property	– g kg ⁻¹ –					Shapiro-Wilk test <i>p</i> value
				Minimum	Maximum	Median	Mean	SD	
1. Residue quantity	low: 3 (Control, WS ₂ , CR ₂) × 2 (pre-, post-inc) × 20 soils	120	SOC	7.44	28.30	13.85	15.20	5.38	2.4 × 10 ⁻⁶
			TN	0.67	2.43	1.35	1.38	0.45	1.2 × 10 ⁻⁴
	high: 4 (WS _{4,8} , CR _{4,8}) × 2 (pre-, post-inc) × 20 soils	160	SOC	8.67	33.27	17.16	17.93	5.58	1.6 × 10 ⁻⁵
			TN	0.75	2.84	1.51	1.59	0.48	4.3 × 10 ⁻⁴
	preincubation: 7 (control, WS _{2,4,8} , CR _{2,4,8}) × 20 soils	140	SOC	7.50	33.27	16.58	17.64	5.71	2.9 × 10 ⁻⁴
			TN	0.67	2.82	1.45	1.50	0.48	7.3 × 10 ⁻⁴
postincubation: 7 (Control, WS _{2,4,8} , CR _{2,4,8}) × 20 soils	140	SOC	7.44	29.74	14.61	15.91	5.46	3.0 × 10 ⁻⁶	
		TN	0.68	2.84	1.45	1.50	0.48	5.7 × 10 ⁻⁴	
2. Residue quality	clover residues: 3 (CR _{2,4,8}) × 2 (pre-, post-inc) × 20 soils	120	SOC	8.25	33.27	15.73	16.87	5.58	5.6 × 10 ⁻⁵
			TN	0.81	2.84	1.57	1.66	0.48	2.3 × 10 ⁻³
	wheat straw: 3 (WS _{2,4,8}) × 2 (pre-, post-inc) × 20 soils	120	SOC	8.48	33.23	16.60	17.46	5.64	2.3 × 10 ⁻⁴
3. Global residue model	calibration, random 2/3 of dataset	187	SOC	7.50	33.27	16.05	16.90	5.58	3.0 × 10 ⁻⁵
			TN	0.67	2.84	1.46	1.52	0.47	1.5 × 10 ⁻⁴
	validation, random 1/3 of dataset	93	SOC	7.44	29.74	14.82	16.52	5.80	2.5 × 10 ⁻⁵
			TN	0.68	2.82	1.38	1.47	0.49	1.1 × 10 ⁻³
	calibration, random 2/3 of 20 soils (i.e. 13 soils) and associated treatments	182	SOC	7.44	33.27	14.91	16.15	5.58	1.2 × 10 ⁻⁷
			TN	0.67	2.84	1.45	1.51	0.50	1.1 × 10 ⁻⁴
validation, random 1/3 of 20 soils (i.e., 7 soils) and associated treatments	98	SOC	8.72	29.25	17.59	17.95	5.61	5.2 × 10 ⁻⁴	
		TN	0.84	2.67	1.45	1.48	0.42	2.3 × 10 ⁻⁴	

Note. The 14 soil treatments applied to arable soils from 20 sites ($n = 280$) included a control (no crop residues), application of wheat straw (WS) or clover residues (CR) at rates of 2, 4, or 8 g C kg soil⁻¹ (rates given as subscripts), and measurement before (pre-inc) or after (post-inc) a 56-d incubation.

pretreatments: (a) with or without moving averages (calculated over 17 datapoints), (b) with or without resampling to keep every second data point due to collinearity in the absorbance data, and/or (c) with or without application of the Savitzky-Golay algorithm for the reduction of noise. The latter was applied with the polynomial degree (PD) set to 2, the order of the derivative (DER) ranging from 1 to 2 (with PD-DER: 2-1 or 2-2), and a window smoothing size of 5, 11, 17, or 23.

The maximum number of latent variables was set to 15, and the optimal number was determined in calibration, whereby the model with the smallest RMSE was selected. Calibration included leave-one-out cross-validation. Independent valida-

tions were then carried out using the visNIR and MIR models created in calibration resulting in the highest ratio of performance to interquartile distance in (defined below).

In pre-tests, we also performed all PLSR analysis with the full spectra and no data pretreatment for comparison. We found that MIR-PLSR models using data pretreatment always outperformed MIR-PLSR models using the full spectra and no data pretreatment in validation, whereas only 50% of the tested visNIR-PLSR models with data pretreatment outperformed visNIR-PLSR models using the full spectra and no data pretreatment in validation (data not shown). Therefore, future studies should consider that the validation performance of MIRS seems to benefit consistently from

noise-reducing data pretreatments, whereas untreated visNIR models are sometimes more robust in validation.

2.7 | Statistical analysis and performance measures

The statistical software R (version 3.4.4; R Core Team, 2018) was used for all statistical analyses. Descriptive statistics were calculated, and the Shapiro-Wilk test of normality was performed for each dataset. To provide insights into the predictive processes for TN, the loadings of the first three PLSR regression components (latent variables) were plotted for SOC and TN models, and correlation analysis was performed with SOC using the nonparametric Spearman rank correlation coefficient due to non-normal distributions of SOC and TN within the datasets. To evaluate the performance of the calibration and validation models, RMSE was calculated as follows:

$$\text{RMSE} = \sqrt{\frac{\sum (\hat{y}_i - y_i)^2}{n}}$$

where \hat{y}_i is the modeled soil parameter, y_i is the measured soil parameter, and n is the sample size. In addition, the ratio of performance to interquartile distance (RPIQ) was calculated as the interquartile distance of the measured values divided by the RMSE of estimation. We calculated RPIQ rather than ratio of prediction to deviation (RPD) due to the non-normality of parameters of interest within all datasets. Our results were evaluated according to the classification system of Chang, Laird, Mausbach, and Hurburgh (2001). For this, the RPD classification system was converted to RPIQ values by multiplying RPD by 1.349 because this is the ratio of the interquartile distance to the standard deviation in a normally distributed dataset. Thus, a model with RPIQ < 1.89 is considered poor, 1.89–2.70 is satisfactory, and > 2.70 is very good. However, one has to keep in mind that the usefulness of a model must always be judged based on the context in which it is applied. Bias of the model estimates was calculated as the sum of the differences between measured and estimated values divided by the sample size. Finally, R^2 was calculated as follows:

$$R^2 = 1 - \frac{\sum (\hat{y}_i - y_i)^2}{\sum (y_i - \bar{y}_i)^2}$$

where \hat{y}_i are the modelled values, y_i are the measured values, and \bar{y}_i is the mean of the measured values.

3 | RESULTS AND DISCUSSION

3.1 | Descriptive statistics

Descriptive statistics of the 10 datasets used in the analysis are given in Table 1. Values for SOC and TN were non-normally distributed within the datasets ($p < .05$ for the Shapiro-Wilk test). Due to the low C/N ratio of soil compared with crop residues, the low-residue quantity dataset had a lower median C/N ratio (10.3) compared with the high-residue quantity dataset (11.4). Incubation resulted in a loss of SOC, whereas TN stayed constant, leading to a median soil C/N ratio of 11.4 versus 10.1 before and after incubation, respectively. Due to the lower C/N ratio of white clover residues (12.5) compared with wheat straw (53.7), the dataset of soils amended with clover residues had 18% higher median TN content compared with the soils amended with straw residues.

3.2 | Soil organic carbon estimation accuracy

Specifications and performance of the SOC estimation models are given in Table 2, and measured versus validation estimated SOC contents are plotted in Figure 2. For SOC predictions using MIRS, application of moving averages resulted in the optimal model in all cases, whereas the Savitzky-Golay algorithm was only useful in one of the six models (Table 2). In contrast, the optimal visNIR model only used moving averages and resampling in one of six models and always used the Savitzky-Golay algorithm. Although visNIR models always used the maximum number of model factors ($n = 15$), MIR models often were optimized with fewer factors (range, 8–15). The optimal MIRS models always used Region 1 (4000–3682.1 cm^{-1}) and Regions 3 and 4 (3020.5–1693.5 cm^{-1}) but did not use Region 6 (<1030 cm^{-1}). The MIR region below 1000 cm^{-1} has been shown to contain more noise (Hutengs, Ludwig, Jung, Eisele, & Vohland, 2018) and overlapping peaks of organic and mineral compounds (Nocita et al., 2015; Soriano-Disla et al., 2014). Our data pretreatment therefore affirmed these findings by eliminating this noisy region from all models in calibration. The optimal visNIRS models always used Region 2 (834–1167 nm) and Region 6 (2166–2500 nm), but all regions were used by at least one of the optimal models.

Across the three analysis approaches, the MIR models always resulted in the lowest RMSE_V and the highest RPIQ_V compared with the visNIR models. Thus, MIR-PLSR models can be recommended as the most robust alternative for SOC modeling given changes in the quantity and quality of crop residues in soil. However, these experimental findings are specific to the spectrometers used (i.e., Foss XDS Rapid Content Analyzer and Bruker-TENSOR 27 MIR spectrometer) and therefore reflect not only differences in the usefulness of visNIRS versus MIRS for the prediction of SOC content but

TABLE 2 Visible near-infrared (visNIR) and mid-infrared (MIR) spectroscopy partial least squares regression (PLSR) models for the prediction of soil organic carbon (SOC) content

SOC g kg ⁻¹	Calibration							Validation		
	Spectral region	Selected regions ^a	Math treatments ^b	Factors	RPIQ _{CV}	RMSE _{CV}	Bias _{CV}	RPIQ _V	RMSE _V	Bias _V
						- g kg ⁻¹ -			- g kg ⁻¹ -	
Analysis 1: Residue quantity										
Dataset: Low quantity						High quantity				
	visNIR	1-3, 5, 6	n-n-2-1-5-822	15	9.12	0.96	-0.044	5.21	1.66	0.62
	MIR	1-4	17-y-0-0-0-594	10	11.98	0.73	0.003	5.75	1.50	0.93
Dataset: Before incubation						After incubation				
	visNIR	1, 2, 4-6	n-n-2-1-5-821	15	7.43	1.15	0.015	3.96	2.19	0.68
	MIR	1, 3-5	17-n-0-0-0-1187	15	10.14	0.84	0.005	6.68	1.30	-0.43
Analysis 2: Residue quality										
Dataset: Clover residues						Wheat straw				
	visNIR	2, 4, 6	n-n-2-1-17-483	15	7.74	1.13	-0.018	5.08	1.73	0.69
	MIR	1-4	17-n-0-0-0-1187	10	10.69	0.82	-0.001	7.53	1.17	0.23
Dataset: Wheat straw						Clover residues				
	visNIR	1, 2, 6	n-y-2-1-5-242	15	7.46	1.18	-0.006	1.24	7.05	-5.02
	MIR	1-4	17-n-0-0-0-1187	10	8.28	1.06	-0.015	7.96	1.10	-0.19
Analysis 3: Global residue model										
Dataset: 2/3 of sample						1/3 of sample				
	visNIR	2, 4, 6	n-n-2-1-17-483	15	7.11	1.19	-0.013	5.71	1.57	-0.07
	MIR	1-5	17-y-0-0-0-766	15	9.29	0.91	0.004	10.07	0.89	-0.03
Dataset: 13 soils & associated residue treatments						7 soils & associated residue treatments				
	visNIR	1, 2, 4, 6	17-n-2-2-11-632	15	7.03	1.10	-0.013	1.49	6.97	3.93
	MIR	1, 3, 4	17-y-2-1-23-411	8	8.82	0.88	-0.009	3.58	2.90	1.28

^avisNIR: 500–834 nm (Region 1), 834–1167 (Region 2), 1167–1500 (Region 3), 1500–1833 (Region 4), 1833–2166 (Region 5), 2166–2500 (Region 6); MIR: 4000–3682.1 cm⁻¹ (Region 1), 3682.1–3020.5 (Region 2), 3020.5–2358.9 (Region 3), 2358.9–1693.5 (Region 4), 1693.5–1030 (Region 5), 1030–368.4 (Region 6).

^bNo use of moving averages (n) or averaging over 17 data points (17); no resampling (n) or resampling (y); polynomial degree (0, 2); derivative (0–2); smoothing (0, 5, 11, 17, 23); number of data points.

also differences in the sample presentation and measurement specifications of these spectrometers.

The SOC predictions of all MIR validation models can be classified as very good according to the RPIQ_V values, ranging from 3.58 (for the global residue model using 13 of the 20 soils and associated residue treatments for calibration and 7 of the 20 soils and associated residue treatments for validation) to 10.07 (for the global residue model using a random two-thirds of the sample for model calibration and one-third of the sample for validation). These models also produced the highest (2.90 g kg⁻¹) and lowest (0.89 g kg⁻¹) RMSE_V, respectively.

For visNIR, the SOC estimates of validation models were more variable, ranging from poor (RPIQ_V = 1.24, RMSE_V = 7.05 g kg⁻¹, and bias_V = -5.02 g kg⁻¹ for the model calibrated with wheat straw-amended soil and validated with clover-amended soil) to very good (RPIQ_V = 5.71, RMSE_V = 1.57 g kg⁻¹, and bias_V = -0.07 g kg⁻¹ for the global residue model using a random two-thirds of the com-

plete dataset for calibration and one-third of the complete dataset for validation).

To put the scale of the estimation error in context, the RMSE_V as a percent of median SOC content ranged from 6.0 to 16.5% for MIRS models and from 8.4 to 44.8% for visNIRS models. In comparison, the precision of laboratory determination of C content by dry combustion has been found to range from 1.3 to 7.3% (Goidts, van Wesemael, & Crucifix, 2009). Therefore, the SOC estimation accuracy of some MIRS models approached the accuracy of traditional laboratory methods, whereas the accuracy of visNIRS models was worse and more variable.

Bias in SOC content estimates was very low in cross-validation (ranging from -0.001 to -0.044 g kg⁻¹) but increased in validation (ranging from -0.03 to -5.02 g kg⁻¹). The visNIR models were generally more biased than MIR models, but the direction of bias_V was often matching for both spectral regions within each analysis approach: (a) calibration with low residue quantity resulted in SOC underestimations

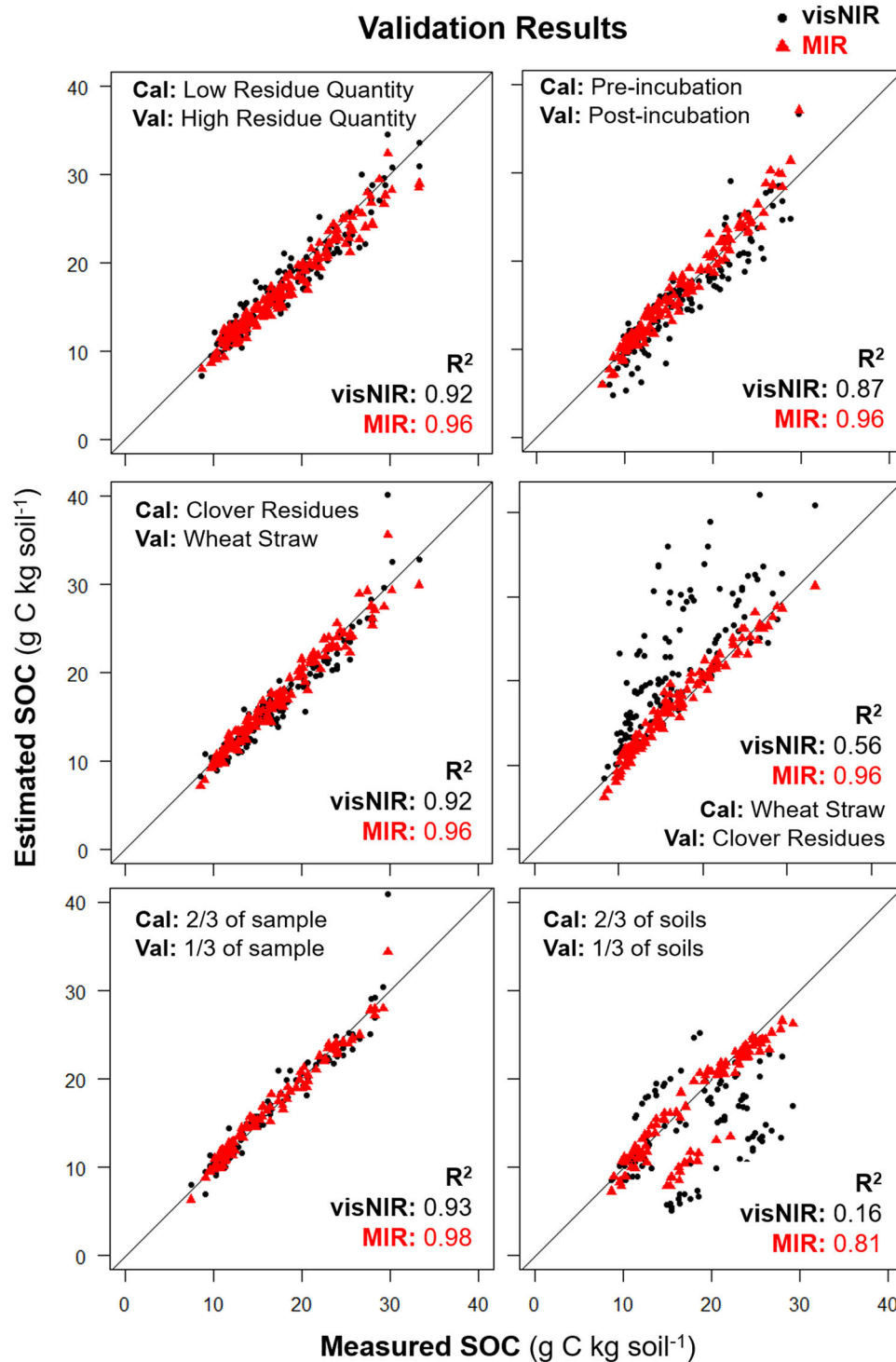


FIGURE 2 Measured versus validation estimated soil organic carbon (SOC) content using specified calibration (Cal) and validation (Val) datasets (described in Table 1). Estimates were based on visible near-infrared (visNIR) and mid-infrared (MIR) spectroscopy partial least squares regression models.

for validation with high residue quantity, whereas calibration with preincubation samples resulted in SOC under- and overestimation in validation with postincubation samples for visNIRS and MIRS, respectively; (b) calibration with clover residues resulted in SOC underestimations for validation with

wheat straw, whereas calibration with wheat straw resulted in SOC overestimations for validation with clover residues; (c) calibration with a random two-thirds of the sample resulted in unbiased SOC estimations in validation, whereas calibration with 13 of the 20 soils and associated residue treatments

resulted in distinctly biased validation estimates for all seven treatments of certain soil types.

The bias_V following calibration with clover-amended soils and validation with wheat straw-amended soils, and vice versa, can be attributed to differences in the chemical makeup of these residues, which are apparent from the crop residue spectra (Figure 1). In the visNIR region, there was pronounced absorbance of red light (670 nm) by clover residues. In the MIR region, differences in the shapes of several absorbance peaks for key C compounds are evident, such as the aliphatic peak at 2930 cm⁻¹ representing labile C compounds, the protein amide peak at 1670 cm⁻¹, peaks between 1600 and 1500 cm⁻¹ representing primarily aromatic C compounds, and peaks between 1450 and 1400 cm⁻¹ associated with primarily aliphatic C compounds (Demyan et al., 2012; Nocita et al., 2015; Soriano-Disla et al., 2014). Underlying these differences are the higher cellulose, hemicellulose, and lignin contents and the lower protein content of wheat straw compared with white clover (Haider, 1996). Quantification of lignin, in particular, by spectroscopy is challenging due to its complex structure that varies between and even within plant species (Jung, Buxton, Hatfield, & Ralph, 1993). Stenberg et al. (2004) found that, whereas cellulose (a homogeneous substance) could be predicted accurately, visNIRS models had a poor ability to predict lignin despite calibration using a diverse range of plant species.

However, in our case, use of global residue calibration models with a random two-thirds of the sample effectively lowered RMSE_V of visNIRS and MIRS models compared with all other analysis approaches. This demonstrates that creation of a calibration dataset with mixed quantities and qualities of crop residues is an effective strategy to improve SOC prediction accuracy. However, although the MIRS global residue calibration model with a random two-thirds of the 20 soils and associated residue treatments was also able to predict of SOC in validation with very good accuracy (RPIQ_V = 3.58), the equivalent visNIRS model was poor (RPIQ_V = 1.49). This highlights the greater robustness of MIRS SOC prediction models to changes in soil mineralogy compared with visNIRS models.

3.3 | Total nitrogen estimation accuracy

Specifications and performance of the TN estimation models are given in Table 3, and measured versus validation estimated TN contents are plotted in Figure 3. For TN predictions using MIRS, moving averages, resampling, and the Savitzky-Golay algorithm were used in the optimal model in four, five, and five of six validations, respectively. For TN predictions with visNIRS, moving averages and resampling were less useful (each selected in only one of six models), and the Savitzky-Golay algorithm was used to produce the optimal model in

five of six validations. The PD-DER was 2-1 in all cases when the algorithm was used for visNIRS and MIRS. All models used 15 model factors, with the exception of one MIR model optimized with nine factors. The optimal MIRS models always used Regions 3 and 4 (3020.5–1693.5 cm⁻¹), and again Region 6 (<1030 cm⁻¹) was not selected. The optimal visNIRS model always used Region 6 (2166–2500 nm), but all other regions were selected in at least two of the six models.

The optimal model regarding RPIQ_V was the MIRS model in five of six cases, but RPIQ_V performance classifications for visNIRS and MIRS were the same in each analysis approach. In validation, MIRS models ranged from very good (RPIQ_V = 5.18, RMSE_V = 0.13 g kg⁻¹, bias_V = -0.007 g kg⁻¹ for the model calibrated with a random two-thirds of the complete dataset) to poor (RPIQ_V = 1.88, RMSE_V = 0.26 g kg⁻¹, and bias_V = -0.164 g kg⁻¹ for the model calibrated with a random 13 of the 20 soils and associated residue treatments). In validation, visNIRS models ranged from very good (RPIQ_V = 4.74, RMSE_V = 0.14 g kg⁻¹, and bias_V = 0.003 g kg⁻¹ for the model calibrated with a random two-thirds of the complete dataset) to poor (RPIQ_V = 0.46, RMSE_V = 1.04 g kg⁻¹, bias_V = -0.202 g kg⁻¹ for the model calibrated with 13 of 20 soils and associated residue treatments). Thus, as with SOC, the performance of visNIRS in validation was more variable than that of MIRS, but the analysis approaches producing the best and worst validations were similar for SOC and TN estimates.

Bias was low in cross-validation estimates of TN (range, -0.0002 to 0.0029 g kg⁻¹) but increased in validation (range, 0.003 to -0.202 g kg⁻¹). Neither spectral region was consistently less biased in validation. The direction of bias_V was often matching for both spectral regions within each analysis approach: (a) as for SOC, calibration with low residue quantity resulted in TN underestimations for validation with high residue quantity, whereas calibration with preincubation samples resulted in TN underestimation in validation with postincubation samples; (b) in contrast to SOC, calibration with clover residues resulted in TN overestimations for validation with wheat straw, whereas calibration with wheat straw resulted in TN underestimations for validation with clover residues; (c) once again, calibration with a random two-thirds of the sample resulted in unbiased TN estimates in validation, whereas calibration with 13 of the 20 soils and associated residue treatments resulted in distinctly biased estimates for all seven treatments of certain soils.

Given the correlation between SOC and TN in the calibration datasets (Spearman's $r = 0.78$ – 0.85), it is plausible that TN estimates were to some extent estimated indirectly based on SOC content. In this case, the loss of SOC during incubation while TN remained constant would lead to a TN underestimation in validation with postincubation datasets,

TABLE 3 Visible near-infrared (visNIR) and mid-infrared (MIR) spectroscopy partial least squares regression (PLSR) models for the prediction of soil total nitrogen (TN) content

TN g kg ⁻¹	Spectral region	Calibration			Validation					
		Selected regions ^a	Math treatments ^b	Factors	RPIQ _{CV}	RMSE _{CV}	Bias _{CV}	RPIQ _V	RMSE _V	Bias _V
						– g kg ⁻¹ –			– g kg ⁻¹ –	
Analysis 1: Residue quantity										
Dataset: Low quantity						High quantity				
	visNIR	1, 3–6	n-n-2-1-5-829	15	8.72	0.07	–0.0014	4.11	0.17	0.065
	MIR	1-5	17-y-2-1-23-755	15	8.28	0.07	–0.0007	4.39	0.16	0.043
Dataset: Before incubation						After incubation				
	visNIR	1, 4–6	n-n-2-1-5-662	15	8.99	0.08	0.0029	2.94	0.24	0.099
	MIR	1–5	n-y-2-1-23-759	15	6.98	0.10	–0.0008	4.09	0.17	0.037
Analysis 2: Residue quality										
Dataset: Clover residues						Wheat straw				
	visNIR	1–3, 6	17-n-0-0-0-643	15	5.05	0.14	–0.0023	2.74	0.22	–0.166
	MIR	3, 4	17-y-2-1-23-344	9	5.20	0.14	0.0008	2.83	0.22	–0.176
Dataset: Wheat straw						Clover residues				
	visNIR	1, 3, 4, 6	n-n-2-1-5-662	15	8.22	0.07	0.0004	3.42	0.21	0.047
	MIR	1–4	17-y-2-1-17-586	15	9.22	0.07	–0.0006	2.97	0.24	0.178
Analysis 3: Global residue model										
Dataset: 2/3 of sample						1/3 of sample				
	visNIR	1, 4–6	n-n-2-1-5-662	15	6.09	0.11	–0.0012	4.74	0.14	0.003
	MIR	1–5	n-y-2-1-23-759	15	5.72	0.12	–0.0002	5.18	0.13	–0.007
Dataset: 13 soils & associated residue treatments						7 soils & associated residue treatments				
	visNIR	2–6	n-y-2-1-5-410	15	7.17	0.11	–0.0004	0.46	1.04	–0.202
	MIR	1–5	17-n-0-0-0-1531	15	6.13	0.13	–0.0012	1.88	0.26	–0.164

Note: Datasets used for cross-validation (CV) and validation (V) according to the three analysis approaches are described in Table 1. Root mean squared error (RMSE), the ratio of prediction to interquartile range (RPIQ), and bias are given for PLSR performed with the optimal data pretreatment from CV.

^avisNIR: 500–834 nm (Region 1), 834–1167 (Region 2), 1167–1500 (Region 3), 1500–1833 (Region 4), 1833–2166 (Region 5), 2166–2500 (Region 6); MIR: 4000–3682.1 cm⁻¹ (Region 1), 3682.1–3020.5 (Region 2), 3020.5–2358.9 (Region 3), 2358.9–1693.5 (Region 4), 1693.5–1030 (Region 5), 1030–368.4 (Region 6).

^bNo use of moving averages (n) or averaging over 17 data points (17); no resampling (n) or resampling (y); polynomial degree (0, 2); derivative (0–2); smoothing (0, 5, 11, 17, 23); number of data points.

which was indeed observed for both visNIRS and MIRS. The patterns of bias_V for the models testing the effect of residue quality also suggest an indirect estimation of TN based on SOC: the lower C/N ratio of clover compared with wheat straw (13 vs. 54) combined with indirect estimation of TN based on SOC would explain the observed overestimation of TN in straw-amended soils when models were calibrated with clover-amended soils. The same logic explains the underestimation of TN in validation with clover-amended soils when the model was calibrated with straw-amended soils.

To verify these hypotheses regarding the predictive mechanisms for SOC and TN, the loadings of the first three components of the visNIR-PLSR and MIR-PLSR models testing the effect of residue quantity and quality were plotted (Figure 4). The loadings of the regression components (latent variables) of the PLSR equation indicate the contributions of particular

wavelengths to that component (Wehrens, 2011). It is difficult to attribute specific peaks to the vibrations of particular bonds for visNIR because this region shows sometimes overlapping overtones and/or combinations of fundamental vibrations occurring in the MIR range (Kuang et al., 2012). However, it is clear from Figure 4 that all important peaks for TN models (536, 570, 652, 686, 1402, 1420, 1898, 2200, 2300, and 2430 nm) were also important peaks in all or some of the SOC models.

For MIR prediction of SOC, the loadings for the first three PLSR components had broad peaks, signifying a wide distribution of important wavelengths for SOC prediction. The broad peaks were centered on 3630 cm⁻¹ (absorption by clay minerals due to OH-stretching between 3700 and 3600 cm⁻¹ [Nguyen, Janik, & Raupach, 1991], which could be correlated with SOC storage due to physical stabilization processes) and

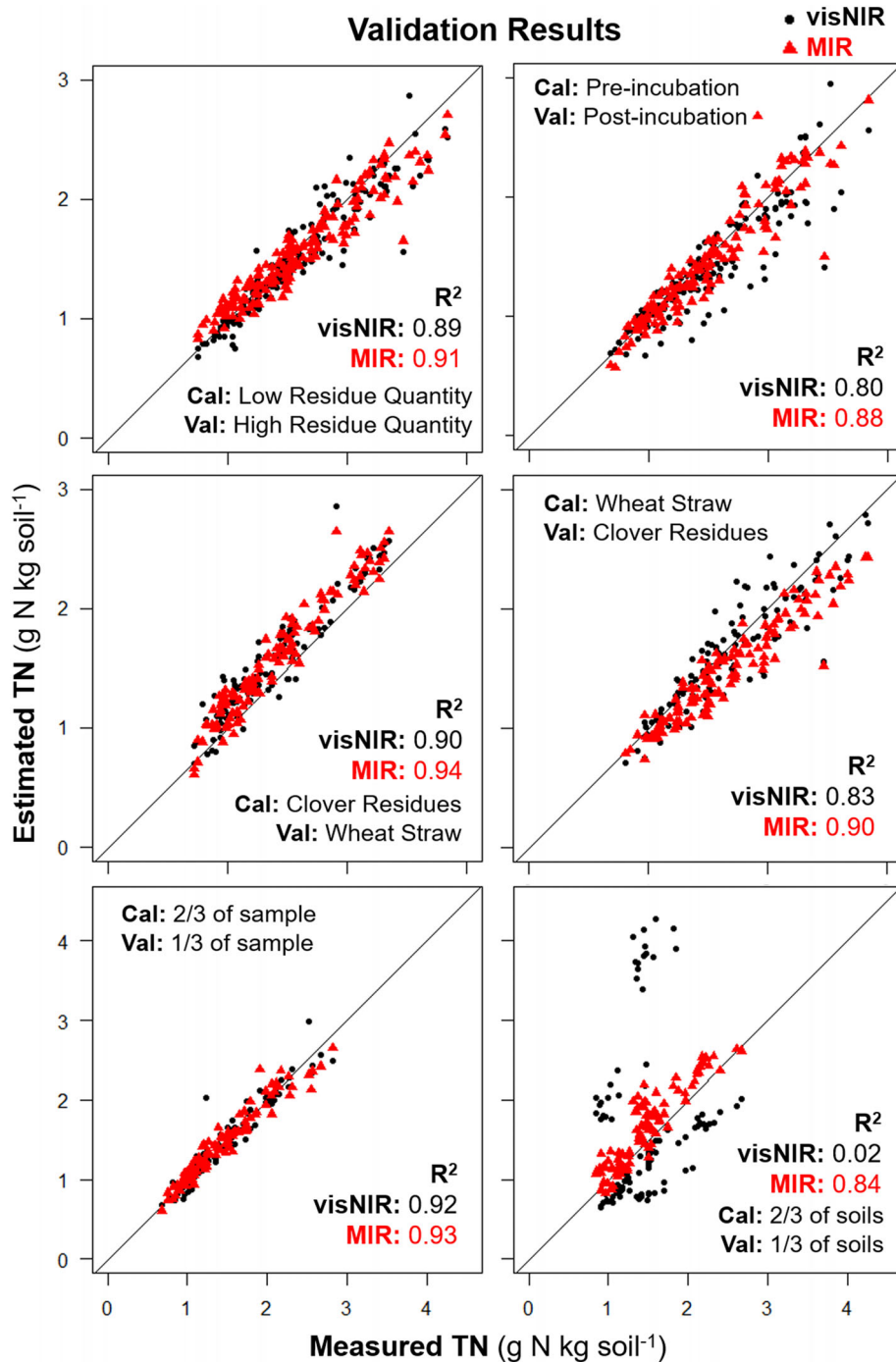


FIGURE 3 Measured versus validation estimated soil total nitrogen (TN) content using specified calibration (Cal) and validation (Val) datasets (described in Table 1). Estimates were based on visible near-infrared (visNIR) and mid-infrared (MIR) spectroscopy partial least squares regression models.

2922 and 2856 cm^{-1} (aliphatic C-H stretching between 3000 and 2800 cm^{-1} [Zimmermann, Leifeld, & Fuhrer, 2007]). In contrast, TN had more distinct wavelengths, with large loadings values at 3700, 3591, 2960, 2208, 2035, 1907, 1853, 1400, and 1280 cm^{-1} . These peaks were not directly aligned with SOC peaks but sometimes fell within areas of the spectra with relatively high loading values for SOC. Therefore,

although the predictive mechanisms for SOC and TN appear to be quite similar for visNIRS, important wavelengths for the prediction of SOC and TN were less aligned for MIRS.

Finally, visNIRS and MIRS global residue calibration models with a random two-thirds of the sample produced the lowest RMSE_V of all the tested analysis approaches, demonstrating that creating a calibration dataset with mixed

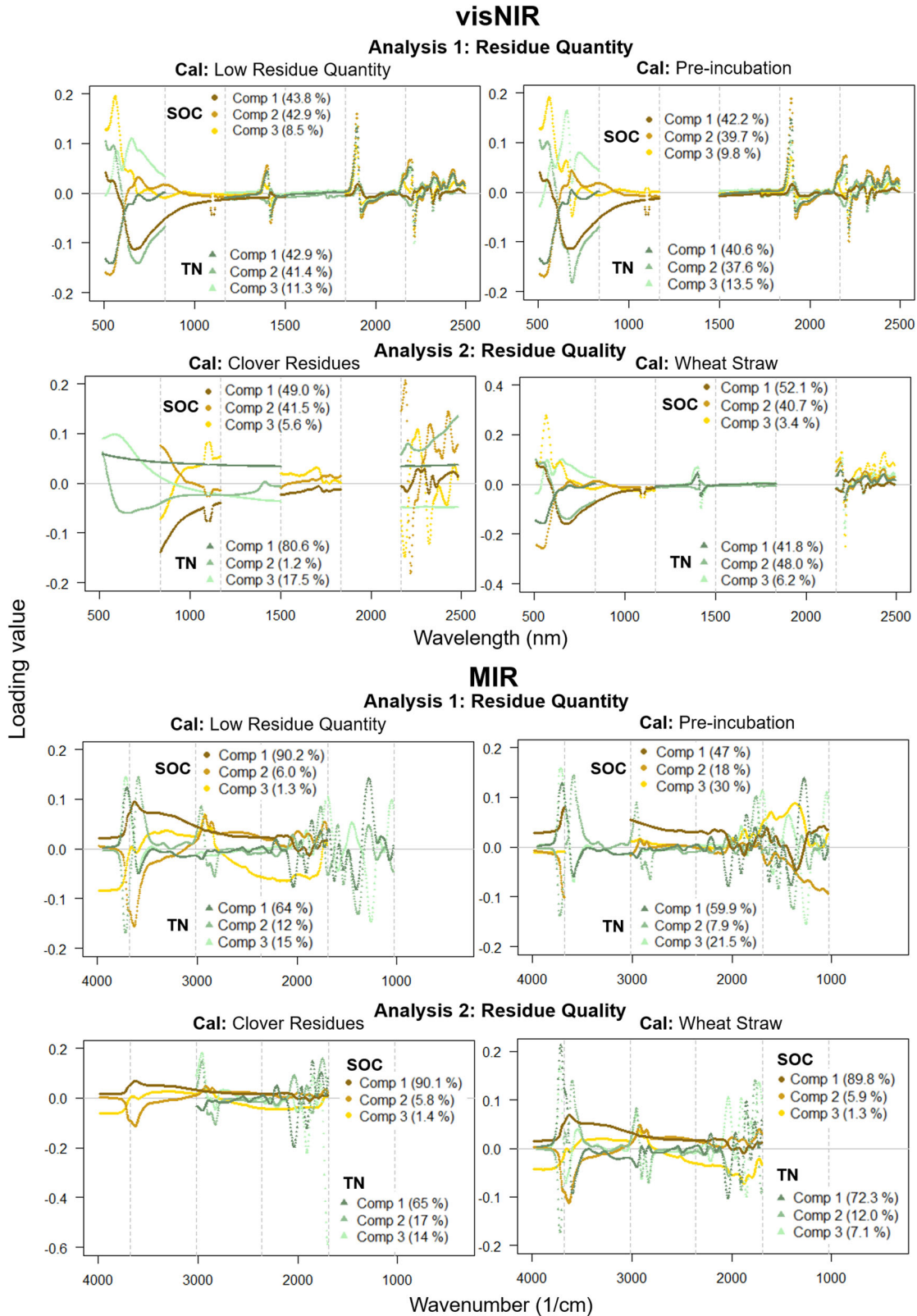


FIGURE 4 Loadings of the first three components (Comp) of visible near-infrared (visNIR) and mid-infrared (MIR) partial least squares regression models developed using the specified calibration datasets (Cal; described in Table 1) for the prediction of soil organic carbon (SOC) and total nitrogen (TN) content. The percentages correspond to the variation in the spectra explained by each regression component. Dashed grey lines divide the six spectral regions.

quantities and qualities of crop residues is an effective strategy to improve TN prediction accuracy. However, visNIR and MIR global residue calibration models with 13 of the 20 soils and associated residue treatments were both poor for the prediction of TN in validation ($RPIQ_V = 0.46$ and 1.88 , respectively). This highlights the necessity of having soils with similar mineralogy in the calibration and validation datasets for accurate TN estimations.

4 | CONCLUSION

This experiment investigated the robustness of visNIRS and MIRS model predictions of SOC and TN content to changes in the quantity and quality of crop residues present in the soil. The MIRS models produced more accurate and less biased estimates of SOC content compared with visNIRS models, whereas TN estimates were more comparable for MIRS and visNIRS. Loadings of PLSR components indicated that the predictive mechanisms for SOC and TN were more similar for visNIRS than MIRS.

Differing crop residue quantity or quality in calibration versus validation datasets resulted in systematically biased SOC and TN estimates by visNIRS and MIRS models. However, the global residue calibration model containing all soils and soil mixtures with crop residues of all quantities and qualities effectively lowered $RMSE_V$ for SOC and TN prediction with visNIRS and MIRS compared with all other models. We therefore recommend that long-term agricultural monitoring studies using soil spectroscopy create a calibration dataset consisting of all soil–residue mixtures present in the crop rotation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

The authors thank Anja Sawallisch and her team for technical assistance. This project was supported by a grant from the German Research Foundation (DFG, LU 583/19-1, VO 1509/7-1).

ORCID

Isabel Greenberg  <https://orcid.org/0000-0002-4762-8474>

Michael Vohland  <https://orcid.org/0000-0002-6048-1163>

Bernard Ludwig  <https://orcid.org/0000-0001-8900-6190>

REFERENCES

Agriculture and Horticulture Development Board. (2016). *Managing clover for better returns*. Warwickshire, U.K.: Better Returns Programme.

Araújo, S. R., Wetterlind, J., Demattê, J. A. M., & Stenberg, B. (2014). Improving the prediction performance of a large tropical vis-NIR

spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. *European Journal of Soil Science*, *65*(5), 718–729. <https://doi.org/10.1111/ejss.12165>

Baldock, J. A., Beare, M. H., Curtin, D., & Hawke, B. (2018). Stocks, composition and vulnerability to loss of soil organic carbon predicted using mid-infrared spectroscopy. *Soil Research*, *56*(5), 468. <https://doi.org/10.1071/SR17221>

Bellon-Maurel, V., & McBratney, A. (2011). Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils: Critical review and research perspectives. *Soil Biology and Biochemistry*, *43*(7), 1398–1410. <https://doi.org/10.1016/j.soilbio.2011.02.019>

Castaldi, F., Chabrilat, S., Chartin, C., Genot, V., Jones, A. R., & van Wesemael, B. (2018). Estimation of soil organic carbon in arable soil in Belgium and Luxembourg with the LUCAS topsoil database. *European Journal of Soil Science*, *69*(4), 592–603. <https://doi.org/10.1111/ejss.12553>

Chang, C.-W., Laird, D. A., Mausbach, M. J., & Hurburgh, C. R. (2001). Near-infrared reflectance spectroscopy: Principal components regression analyses of soil properties. *Soil Science Society of America Journal*, *65*(2), 480. <https://doi.org/10.2136/sssaj2001.652480x>

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. *European Journal of Soil Science*, *63*(2), 189–199. <https://doi.org/10.1111/j.1365-2389.2011.01420.x>

European Commission. (2007). *Cereals straw resources for bioenergy in the European Union* (EUR 22626 EN). Brussels: European Commission Directorate General Joint Research Centre, Institute for Environment and Sustainability.

Food and Agriculture Organization (FAO). (2017). FAOSTAT database. Retrieved from <http://www.fao.org/home/en>

Goidts, E., van Wesemael, B., & Crucifix, M. (2009). Magnitude and sources of uncertainties in soil organic carbon (SOC) stock assessments at various scales. *European Journal of Soil Science*, *60*(5), 723–739. <https://doi.org/10.1111/j.1365-2389.2009.01157.x>

Haider, K. (1996). *Biochemie des bodens*. Stuttgart: F. Enke.

Hutengs, C., Ludwig, B., Jung, A., Eisele, A., & Vohland, M. (2018). Comparison of portable and bench-top spectrometers for mid-infrared diffuse reflectance measurements of soils. *Sensors*, *18*(4), E993. <https://doi.org/10.3390/s18040993>

Jung, H. G., Buxton, D. R., Hatfield, R. D., & Ralph, J. (1993). *Forage cell wall structure and digestibility*. Madison, WI: ASA, CSSA, SSSA.

Kuang, B., Mahmood, H. S., Quraishi, M. Z., Hoogmoed, W. B., Mouazen, A. M., & van Henten, E. J. (2012). Sensing soil properties in the laboratory, in situ, and on-Line: A review. *Advances in Agronomy*, *114*, 155–223. <https://doi.org/10.1016/B978-0-12-394275-3.00003-1>

Linn, D. M., & Doran, J. W. (1984). Effect of water-filled pore space on carbon dioxide and nitrous oxide production in tilled and non-tilled soils. *Soil Science Society of America Journal*, *48*, 1267–1272. <https://doi.org/10.2136/sssaj1984.03615995004800060013x>

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>
- Madhavan, D. B., Kitching, M., Mendham, D. S., Weston, C. J., & Baker, T. G. (2016). Mid-infrared spectroscopy for rapid assessment of soil properties after land use change from pastures to Eucalyptus globulus plantations. *Journal of Environmental Management*, 175, 67–75. <https://doi.org/10.1016/j.jenvman.2016.03.032>
- Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., ... Wetterlind, J. (2015). Soil spectroscopy: An alternative to wet chemistry for soil monitoring. *Advances in Agronomy*, 132, 139–159. <https://doi.org/10.1016/bs.agron.2015.02.002>
- Nguyen, T. T., Janik, L. J., & Raupach, M. (1991). Diffuse reflectance infrared fourier transform (DRIFT) spectroscopy in soil studies. *Australian Journal of Soil Research*, 29(1), 49–67.
- Page, K. L., Dalal, R. C., & Dang, Y. P. (2013). How useful are MIR predictions of total, particulate, humus, and resistant organic carbon for examining changes in soil carbon stocks in response to different crop management? A case study. *Soil Research*, 51(8), 719. <https://doi.org/10.1071/SR13064>.
- R Core Team. (2018). *R: A language and environment for statistical computing. Version 3.4.4*. Vienna: R Foundation for Statistical Computing.
- Reeves, J. B., McCarty, G. W., and Hively, W. D. (2010). Mid- versus near-infrared spectroscopy for on-site analysis of soil. *Proximal Soil Sensing*, 38, 133–142.
- Seidel, M., Hutengs, C., Ludwig, B., Thiele-Bruhn, S., & Vohland, M. (2019). Strategies for the efficient estimation of soil organic carbon at the field scale with vis-NIR spectroscopy: Spectral libraries and spiking vs. local calibrations. *Geoderma*, 354, 113856. <https://doi.org/10.1016/j.geoderma.2019.07.014>
- Soil Survey Staff. (2014). *Keys to soil taxonomy* (12th ed.). Washington, DC: USDA-Natural Resources Conservation Service.
- Soriano-Disla, J. M., Janik, L. J., Viscarra Rossel, R. A., Macdonald, L. M., & McLaughlin, M. J. (2014). The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Applied Spectroscopy Reviews*, 49(2), 139–186. <https://doi.org/10.1080/05704928.2013.811081>
- Stenberg, B., Jensen, L. S., Nordkvist, E., Breland, T. A., Pedersen, A., Guðmundsson, J., ... Korsæth, A. (2004). Near infrared reflectance spectroscopy for quantification of crop residue, green manure and catch crop C and N fractions governing decomposition dynamics in soil. *Journal of Near Infrared Spectroscopy*, 12(5), 331–346. <https://doi.org/10.1255/jnirs.442>
- Stevens, A., Nocita, M., Tóth, G., Montanarella, L., & van Wesemael, B. (2013). Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy. *PLOS ONE*, 8(6), e66409. <https://doi.org/10.1371/journal.pone.0066409>
- Stevens, A., van Wesemael, B., Bartholomeus, H., Rosillon, D., Tychon, B., & Ben-Dor, E. (2008). Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. *Geoderma*, 144(1-2), 395–404. <https://doi.org/10.1016/j.geoderma.2007.12.009>
- Stevens, A., van Wesemael, B., Vandenschrck, G., Touré, S., & Tychon, B. (2006). Detection of carbon stock change in agricultural soils using spectroscopic techniques. *Soil Science Society of America Journal*, 70(3), 844–850. <https://doi.org/10.2136/sssaj2005.0025>
- Wehrens, R. (2011). *Chemometrics with R*. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-17841-2>.
- Widdup, K. H., Ford, J. L., Cousins, G. R., Woodfield, D. R., Caradus, J., R., & Barrett, B. A. (2015). A comparison of New Zealand and overseas white clover cultivars under grazing in New Zealand. *Journal of New Zealand Grasslands*, 77, 51–56.
- Zeng, R., Zhao, Y.-G., Li, D.-C., Wu, D.-W., Wei, C.-L., & Zhang, G.-L. (2016). Selection of “Local” models for prediction of soil organic matter using a regional soil Vis-NIR spectral library. *Soil Science*, 181(1), 13–19. <https://doi.org/10.1097/SS.0000000000000132>
- Zimmermann, M., Leifeld, J., & Fuhrer, J. (2007). Quantifying soil organic carbon fractions by infrared-spectroscopy. *Soil Biology and Biochemistry*, 39(1), 224–231. <https://doi.org/10.1016/j.soilbio.2006.07.010>

How to cite this article: Greenberg I, Linsler D, Vohland M, Ludwig B. Robustness of visible-near infrared and mid-infrared spectroscopic models to changes in the quantity and quality of crop residues in soil. *Soil Sci. Soc. Am. J.* 2020;84:963–977. <https://doi.org/10.1002/saj2.20067>