

# Optimizing acquisition parameters in diffuse reflectance infrared Fourier transform spectroscopy of soils

Leonardo Deiss<sup>1</sup>  | Andrew J. Margenot<sup>2</sup>  | Steve W. Culman<sup>3</sup>  | M. Scott Demyan<sup>4</sup> 

<sup>1</sup>School of Environment and Natural Resources, Ohio State Univ., 414A, Kottman Hall, 2021 Coffey Road, Columbus, OH 43210

<sup>2</sup>Crop Sciences Dep., Univ. of Illinois Urbana-Champaign, 1201 S Dorner Dr, Urbana, IL 61801

<sup>3</sup>School of Environment and Natural Resources, Ohio State Univ., 1680, Madison Ave, Wooster, Ohio 44691

<sup>4</sup>School of Environment and Natural Resources, Ohio State Univ., 408B, Kottman Hall, 2021 Coffey Road, Columbus, OH 43210.

## Correspondence

Leonardo Deiss, School of Environment and Natural Resources, Ohio State Univ., 414A, Kottman Hall, 2021 Coffey Road, Columbus, OH 43210.

Email: leonardodeiss@gmail.com

## Funding information

The School of Environment and Natural Resources at Ohio State University; Foundation for Food and Agricultural Research

## Abstract

The number of co-added scans and spectral resolution are two fundamental acquisition parameters in diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS). However, systematic studies evaluating these parameters in soil science applications are lacking, especially when using mid-infrared frequencies (mid-DRIFTS). The objective of this study was to evaluate how numbers of co-added scans and spectral resolution affect both qualitative and predictive applications of high-throughput mid-DRIFTS of soils. We first explored the literature with a systematic review to better understand variability in acquisition parameters and their hypothesized relationships with mid-DRIFTS predictive performance (accuracy) of soil organic and/or total carbon (as a reference variable), but no evident relationship could be established across studies. Second, we experimentally evaluated how spectral resolutions (4, 8, 16, and 32  $\text{cm}^{-1}$ ), and number of co-added scans (8, 16, and 24 scans) affected (i) specific spectral peaks representing mineral and organic functional groups, and (ii) predictive performance of soil variables clay, sand, pH, total organic carbon, and permanganate oxidizable carbon (POXC). Decreasing the number of co-added scans from 24 to 8 increased wavenumber-specific spectral variability and decreased both the predictive performance and the ability to characterize smaller peaks of mineral and organic functional groups. In contrast, broadening spectral resolution from 4 to 32  $\text{cm}^{-1}$  reduced multivariate scores dispersion and had a positive effect on the predictive performance, even though some smaller peaks disappeared in resolutions wider than 8  $\text{cm}^{-1}$ . Acquisition parameters can be set to reduce both scanning time and computational demand while maintaining qualitative and predictive applications of mid-DRIFTS in soil analysis.

## 1 | INTRODUCTION

Diffuse reflectance infrared Fourier transform spectroscopy in the mid-infrared region (mid-DRIFTS) offers two possi-

bilities for the study of soils: (1) Qualitative, that is, compositional analysis by functional groups for organic matter and mineralogical compositions; and (2) Predictive, that is, calibrating and validating datasets to predict a wide range of soil physical, chemical, and biological properties. The utility of mid-DRIFTS is that specific polar bonds and their associated vibrations absorb infrared electromagnetic energy at characteristic frequencies, enabling identification (and

**Abbreviations:** DRIFTS, diffuse reflectance infrared Fourier transform spectroscopy; PCA, principal component analysis; POXC, permanganate oxidizable carbon; TC, total carbon; TOC, total organic carbon.

ideally, quantification) of specific molecular bond types, known as functional groups (Parikh, Goyne, Margenot, Mukome, & Calderón, 2014). Thus, an absorbance spectrum exhibits peaks that represent absorption of infrared electromagnetic energy at frequencies (commonly expressed as wavenumbers,  $\text{cm}^{-1}$ ) being specific to the type and vibrational mode(s) of functional groups. Absorbances in the mid-infrared region ( $4000\text{--}400\text{ cm}^{-1}$ ) involve fundamental vibrations of mineral and organic functional groups, and therefore mid-DRIFTS has an immense potential to be used in soil sciences (Nocita et al., 2015). This technique has been applied to soil samples for decades, but there is still a notable lack of studies that evaluated specific acquisition parameters in mid-DRIFTS using a wide range of soils types and soil properties.

Choices of acquisition parameters such as number of co-added scans (number of scans acquired and averaged to produce a spectrum) and resolution can affect the spectral quality (Azambre, Heintz, Schneider, Krzton, & Weber, 1999; Blitz & Klarup, 2002; Griffiths, 1972; Griffiths & Haseth, 1986; Hanssen, 1993), and therefore outcomes from mid-DRIFTS. Spectral quality can be a highly restrictive factor in Fourier transform infrared spectroscopy of soils that limits the utility of this technique (Coûteaux, Berg, & Rovira, 2003); however, only a few studies have evaluated the impact of selecting acquisition parameters in soil samples (e.g., Adeline, Gomez, Gorretta, & Roger, 2017; Knadel, Stenberg, Deng, Thomsen, & Greve, 2013; Peng, Shi, Song, & Gao, 2014; Robin, Petit, Beaufort, & Prêt, 2013; Yang, Kuang, & Mouazen, 2012). Notably, little information is available on how acquisition parameters affect spectral quality and thus qualitative and predictive outcomes from mid-DRIFTS of soils. Multiple metrics can be used to assess spectral quality, such as spatial resolution, signal-to-noise ratio, and presence of spectral artifacts (Kimber & Kazarian, 2017). These metrics can significantly affect the limit of detection of particular chemical signatures, particularly in narrow regions of interest in the spectrum (Kimber & Kazarian, 2017). Spectra quality can be evaluated by the wavenumber-specific variability and principal component analysis (PCA) scores dispersion and loading vectors (Baldock, Hawke, Sanderman, & MacDonald, 2013; Le Guilou et al., 2015; Stumpe, Weihermüller, & Marschner, 2011). To effectively address the impacts of experimental acquisition parameters on spectral quality, mid-DRIFTS qualitative and predictive outcomes should be evaluated. Moreover, optimizing these parameters while maintaining relevant spectral features can increase laboratorial efficiency when measuring soil properties. However, this optimization is based on trade-offs between acquisition time and computational demand and the ability to properly represent spectral features of a sample.

Spectral resolution refers to the ability of an instrument to distinguish spectral features in close proximity. The broader the resolution, the less sharp the spectral features will be, and consequently a spectrum is less distinguishable for

### Core Ideas

- DRIFTS has transformed soil analyses as a high-throughput technique.
- Co-added scan number and resolution are core parameters in mid-DRIFTS.
- Both parameters affected spectral variability and chemometric outcomes from mid-DRIFTS.
- Vetting acquisition parameters allows optimizing soil analysis with mid-DRIFTS.

specific features that are close together. In a high-resolution spectrum, features are well resolved because the data point spacing is sufficiently small to allow these features to be distinguished (Smith, 2011). As resolution increases ( $<1\text{ cm}^{-1}$ ), the number of data points, i.e., specific wavenumbers (Workman, 2016), the time for spectra acquisition, and the computational demand to process data also increase. Generally, a sample with many polar bonds of diverse functional groups has broad spectral features, and a sample with fewer functional groups (e.g., pure compounds) has narrow spectral features. Due to the large number of distinct chemical molecules in heterogeneous samples such as soils, their absorbance values are typically  $10\text{ cm}^{-1}$  or greater, so a spectrum with a resolution of  $4\text{ cm}^{-1}$  will easily resolve most bands (Smith, 2011). Using higher resolutions than needed would result in longer measurement time and the appearance of peaks would be similar (Smith, 2011). Experimental acquisition parameters therefore require optimization to maximize analytical efficiency by broadening resolution up to the limit where there is no loss of relevant spectral information.

A potential drawback of obtaining high-resolution mid-infrared spectra is presence of greater noise than low-resolution spectra (Smith, 2011). Noise is defined as random fluctuations below and above the baseline and it can be generated by several sources from the instrument or sample. Detector noise and other instrumental sources of electronic and mechanical noise can affect measurements in mid-DRIFTS (Schwartz, Eshel, & Ben-Dor, 2011), and these are independent of the sample characteristics. In soil samples for which very weak spectral features are used for chemometric modeling, these noise factors can alter the use of a selected spectral model for a wide range of spectrometers and users (Gholizadeh, Luboš, Saberioon, & Vašát, 2013). In a hypothetical optimal condition, the random (detector) noise in the spectrum is reduced by the square root of the number of co-added scans (Cazes, 2004), so four scans would have half the noise and twice the signal to noise ratio as compared to a single scan (Workman, 2016). This situation only occurs for highly reproducible data, where there is an exact in-phase

registration between interferograms (Cazes, 2004). The number of co-added scans should be estimated for each type of sample and/or spectrum (Workman, 2016), and in less reproducible samples such as soil spectra, one should expect reductions in random (detector) noise by increasing the number of co-added scans. As a tradeoff, increasing the number of co-added scans increase the time of spectra acquisition, and therefore, finding the smallest number of co-added scans that do not compromise the spectra usage can optimize analytical efficiency. Quantifying the impact of the number of co-added scans on the qualitative analysis (e.g., specific peak areas), and prediction performance (uncertainty/error of predictions), stands to inform decision making on tradeoffs of acquisition parameters for mid-DRIFTS of soils.

Given the intrinsic soil matrix complexity and variability, mid-DRIFTS of soils may require a clearer understanding of how experimental acquisition parameters affect spectral quality, and how these parameters affect spectra qualitative characterization, and predictive performance of soil properties. The objective of this study is to determine more suitable experimental parameters to optimize qualitative and predictive applications of high-throughput mid-DRIFTS in soil sciences. We first performed a systematic literature review to evaluate how experimental parameters (wavenumber range and resolution, number of co-added scans, spectra replicates, and soil grinding) have affected prediction performances of mid-DRIFTS using total soil carbon or organic carbon as reference variables. We then experimentally tested how the qualitative characterization of soil mineral and organic composition, and predictive (quantitative) performance of different soil variables (total organic carbon, clay, pH, permanganate oxidizable carbon) are affected by the wavenumber resolution, and number of co-added scans, using diverse soil sets representative of the US Midwest.

## 2 | MATERIALS AND METHODS

### 2.1 | Systematic review

A systematic literature review was conducted to evaluate how acquisition parameters (wavenumber range, wavenumber resolution, and number of co-added scans), replicates, and sample preparation have affected the prediction performance in mid-DRIFTS of soils. Soil total carbon (TC) or total organic carbon (TOC) were used as reference variables given the widespread availability of studies that measured these soil properties and evaluated the predictive applications of mid-DRIFTS.

A database search was conducted to identify published papers that predicted soil TC or TOC with mid-DRIFTS. We used the Web of Knowledge platform and the following search terms were included to locate the studies: (dif-

fuse reflectance infrared Fourier transform or mid-infrared or mid-IR or middle-infrared) AND (soil) AND (carbon). The search was constrained for studies published since 2013 to 2018 from which 265 results were obtained (excluding patents). We followed pre-defined eligibility criteria to select studies that included: mid-infrared spectra region; predictive applications of mid-DRIFTS (non-qualitative/compositional analysis); soil TC or TOC values (TOC was used when both were available); neat soil samples (non-sediments and non-liquid samples); sieved/ground samples (non-intact cores); and laboratory-based measurements. Field based measurements or laboratory-based measurements of intact soil cores were excluded. When studies evaluated separately or together more than one geographic region, the validation results including all regions were used, or ranges from the different regions were presented when overall prediction was not available. When studies compared different grinding sizes, all data available was extracted, and ranges were presented. When studies combined mid-infrared with other infrared regions (i.e., visible/near-infrared), results from those other regions were excluded. Different infrared spectroscopic approaches such as transmission, photoacoustic (PA) or attenuated total reflection (ATR) were excluded from the survey. Given that the different spectroscopy techniques may be sample- and purpose-specific, it can potentially generate different spectral outcomes (Smith, 2011), and as the DRIFTS application is currently a more widely used technique in soil sciences, we focused solely on DRIFTS.

From selected papers, we compiled data regarding the following information: wavenumber range, wavenumber resolution, number of co-added scans, sample/spectra replication, grinding size, number of calibration and validation samples, root mean squared error (RMSE), coefficient of determination ( $R^2$ ), and the measured range of soil TC or TOC (Table 1). If the spectra wavenumber range was further reduced (e.g., to eliminate noisy regions) for chemometric reasons, this reduced wavenumber region is presented here. Root mean squared errors and  $R^2$  were extracted from validation results (independent datasets), or cross-validation when independent validation coefficients were not available.

### 2.2 | Laboratory experiments

#### 2.2.1 | Soils and study area

We selected 400 soil samples from the National Cooperative Soil Survey (NCSS) distributed over four representative physiographic regions of the US Midwest: Glaciated Allegheny Plateau, Unglaciated Allegheny Plateau, Till Plains, and Huron-Erie Lake Plains, and covered a wide variety of land-uses including natural and anthropically modified ecosystems. Within each physiographic region, 100 soil

**TABLE 1** Summary of the experimental parameters used to predict soil total carbon or total organic carbon in diffuse reflectance infrared Fourier transform spectroscopy of soils in the mid-infrared region. Literature review was targeted from 2013 to 2018

Reference	Wave number		Number of co-added scans	Spectra replicates		Grinding		Samples ( <i>n</i> )		Soil total carbon or organic carbon <sup>a</sup>		RMSE <sup>a</sup> g kg <sup>-1</sup>	R <sup>2b</sup>
	range	resolution cm <sup>-1</sup>		Different subsamples	Within a subsample	size	Cal.	Val.	g kg <sup>-1</sup>				
Baldock, Beare, Curtin, & Hawke, 2018	6000–1030 <sup>d</sup>	8	60	–	–	–	628	325	628	325	10.2–38.4	0.073 <sup>d</sup>	0.89
Hutengs et al., 2018	4000–650 <sup>d</sup>	2	200	–	2	–	–	40	–	40	6.2–27.0	2.2	0.73
Wang, Sanderman, & Yoo, 2018	6000–180	4	60	1	–	–	–	109	–	109	–	–	0.97
Riedel, Denk, Müller, Barth, & Gläßer, 2018	4000–666	16	–	–	12	–	103	100	103	100	1–95	8.9	0.71
Zhang et al., 2018	4000–400	2	64	–	3	–	193	83	193	83	10.3–162	6.9	0.92
Madhavan et al., 2017	7800–450	4	64	–	–	–	153	–	153	–	15.1–59.9	3.0	0.96
Mirzaeitalarposhti et al., 2017	3900–700 <sup>c</sup>	4	16	3	–	–	1656	93	1656	93	0.5–5.5	–	0.95–0.97
Chen et al., 2017	3800–800	2	–	–	–	–	52	25	52	25	10.40–36.3	3.58	0.75
Wilson et al., 2017	7800–400	8	–	–	–	–	258	80	258	80	10–39.6	–	0.99
Jia et al., 2017	4000–650	4	10	–	3	–	16	8	16	8	0.39–130	9.2	0.82
Takoutsing, Weber, Tchoundjeu, & Shepherd, 2016	–	–	–	–	–	–	54	–	54	–	~12.2–18.3	1.3	0.94
Clairotte et al., 2016	4000–400	3.86	32	–	–	–	3084	380	3084	380	0.59–177	–	0.92
O'Rourke et al., 2016	4000–400	2	–	–	–	–	242	80	242	80	1.4–35.0	3.25	0.78
Madhavan, Kitching, Mendham, Weston, & Baker, 2016	4000–450	4	64	–	–	–	177	62	177	62	10.2–126	0.4 <sup>e</sup>	0.92
Hobley et al., 2016	6000–1030	8	60	–	–	–	612	–	612	–	~0–30	–	0.95
Sila, Shepherd, & Pokhariyal, 2016	–	–	32	4	–	–	1329	575	1329	575	1.4–112.9	–	0.8
Ludwig et al., 2016	4000–370	2	–	–	2	–	51/84	–	51/84	–	7.5–22.2	–	0.87–0.94

(Continues)

TABLE 1 (Continued)

Reference	Wave number		Number of co-added scans	Spectra replicates		Grinding		Soil total carbon or organic carbon <sup>a</sup>		RMSE <sup>a</sup> g kg <sup>-1</sup>	R <sup>2b</sup>
	range	resolution cm <sup>-1</sup>		Different subsamples	Within a subsample	size	mm	Cal.	Val.		
Henaka Arachchi, Field, & McBratney, 2016	4000–600	4	60	–	–	Bulk soil	37	13	2.53–29.8	2.6	0.88
Winowiecki, Vägen, & Huising, 2016	4000–601	4	–	–	–	0.1	5600	32 (7 sites)	~0–60	4.3	>0.95
Barthès et al., 2016	4000–400	3.86	32	2–1	–	2 to 0.2	97	–	2–121	–	0.84–0.95
Towett, Shepherd, Sila, Aynekulu, & Cadisch, 2015	4000–600	4	32	4	–	0.1	700	–	1–110	2.0	0.9
Ludwig, Sawallisch, Heinze, Joergensen, & Vohland, 2015	4000–370	2	–	–	2	0.2	120/160	–	3.4–15	–	0.9–0.95
Mirzaeitarpouhi et al., 2015	3900–700 <sup>c</sup>	4	16	3	–	Ball-milled	126	84	7.0–49	0.8	0.99
Knox et al., 2015	6000–400	4	32	–	–	Ball-milled	696	296	1.33–523	0.23 <sup>d</sup>	0.96
Le Guillou et al., 2015	7500–600	4	64	3	–	2 to 0.1	227	–	0.3–201	8.0–9.6	0.66–0.77
Hicks et al., 2015	4000–600	4	64	4	–	0.5	346	148	0–207	4.0	0.88
Araújo et al., 2015	4000–400	–	64	–	–	0.149	45	103	4.4–56	1.1	0.94
Vohland et al., 2014	4000–800	4	100	–	5	Finely-ground	60	–	10.9–37.7	2.0–2.4	0.78–0.91
Veum, Goynes, Kremer, Miles, & Sudduth, 2014	3000–1025	2	64	–	–	Finely-ground	–	24	9.4–37.1	2.2–3.4	0.76–0.89
Giacometti et al., 2013	3900–700	4	–	3	–	Ball-milled	48	48	6.74–8.97	0.23	0.89
Rabenarivo et al., 2013	3961–439	3.86	32	–	–	0.2	150	209	6.4–59.1	–	0.92
Ballock et al., 2013	6000–1030	8	60	–	–	Finely-ground 180 s	13100	6626	0–147.6	5.1	0.85

<sup>a</sup>Total organic carbon was used when both total carbon and organic carbon were available.

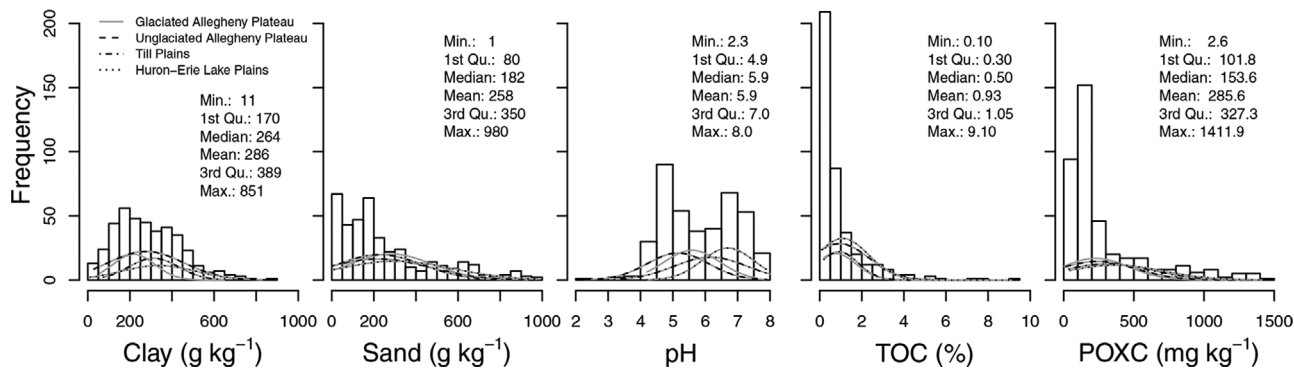
<sup>b</sup>The coefficient of determination (R<sup>2</sup>) and root mean squared error (RMSE) were extracted from independent validation datasets or cross-validation when independent validation was not available.

<sup>c</sup>The CO<sup>2</sup> region (2400–2300 cm<sup>-1</sup>) was also excluded from the analysis.

<sup>d</sup>log-transformed data.

<sup>e</sup>square-root-transformed data.





**FIGURE 1** Summary statistics of soil variables covering four physiographic regions of the US Midwest ( $n = 100$  for each region). The bars represent the actual frequency across all regions ( $n = 400$ ) and lines represent the estimated normal distribution for each region. TOC, total organic carbon; POXC, permanganate oxidizable carbon

samples were systematically selected to cover a broad range of measured values (Figure 1). Soils were originally sampled from different genetic horizons over the period from 1950–2012 as part of the NCSS in Ohio. Soil samples were analyzed and archived at the Ohio State University in collaboration with the NCSS. Data was digitally available at <https://ncsslabsdatamart.sc.egov.usda.gov/>. Samples were assigned to a physiographic region based on their county location. Samples from counties containing two or more physiographic regions were not included.

## 2.2.2 | Soil analysis

Laboratory soil analysis for soil texture, pH, and TOC followed the methods described in the Soil Survey Laboratory Information Manual (Burt, 2011; SSIR no. 42, Soil Survey Staff, 2014). Briefly, soil particle size distribution (soil texture) was determined using the pipet method (method 3A1), and the variables used in this study were total sand (<2.0 mm and >0.05 mm) and total clay (<0.002 mm). Soil pH was measured in a 1:1 (v/v) soil/water mixture (method 4C1a2a), and TC was measured by dry combustion (method 6A2a). Total carbon was measured in a Lindberg tube furnace heated to 900°C, with evolved CO<sub>2</sub> being swept by an oxygen carrier gas to an Ascarite filled Nesbitt absorption bulb. Weight change of the absorption bulb was recorded with a microbalance and converted to TC. In soils without carbonates, TC was taken as TOC, while in soils with carbonates, inorganic carbon was determined separately by the gasometric method (Dreimanis, 1962) and inorganic carbon content subtracted from TC to yield TOC. Permanganate oxidizable carbon (POXC, mg kg soil<sup>-1</sup>) was measured based on the methods of Weil, Islam, Stine, Gruver, and Samson-Liebig (2003) adapted by Culman, Freeman, and Snapp (2012). In brief, 20 mL of 0.02 mol L<sup>-1</sup> KMnO<sub>4</sub> was added to 50 mL tubes containing 2.5 g air-dried soil. The tubes were shaken for

2 min at 240 oscillations min<sup>-1</sup> then allowed to settle for 10 min. After settling, 0.5 mL of the supernatant was diluted with 49.5 mL of deionized water and sample absorbance was quantified at 550 nm on a spectrophotometer.

## 2.2.3 | Sample preparation for DRIFTS

The samples from the soil survey were originally crushed and sieved to <2.0 mm and stored in an air-dried state. Before acquiring spectra, <2.0 mm soils were dried for >48 h at 40°C and at 12–14% relative humidity. This temperature (40°C) is suitable to evaluate SOM composition (e.g., Demyan et al., 2013). To analyze samples in the mid-DRIFTS instrument, 24-well anodized aluminum plates were used. These plates fit 24 removable polystyrene sample cups with a top circular opening area of 10-mm diameter and 5.5-mL volume (with no cap). The sample cups were loaded by over-filling the cups with soil, then tapping the cup side gently three times to settle the soil into the cup, and finally smoothing the surface by scraping excess soil with the narrow edge of a stainless steel spatula. The soil was not packed or compressed into the well other than by tapping and scraping to avoid artifacts of matrix density (Terhoeven-Urselmans, Vagen, Spaargaren, & Shepherd, 2010).

## 2.2.4 | DRIFTS instrument set-up

Spectra were obtained using an X,Y Autosampler (Pike Technologies Inc., Madison, WI) coupled with a Nicolet iS50 spectrometer equipped with a diffuse reflectance accessory (Thermo Fisher Scientific Inc., Waltham, MA). We used FT-IR grade (≥99% trace metals basis) potassium bromide (KBr) (Sigma-Aldrich Inc., St. Louis, MO) for background spectrum collected at the beginning of each plate reading (i.e., every 23 samples). All measurements were conducted from 4000 to

400  $\text{cm}^{-1}$ , and we further reduced the spectral data to 4000 to 700  $\text{cm}^{-1}$  to eliminate increased noise at the spectrum edge, to conduct spectral analysis and predictions. Each soil subsample (one sample cup) was measured four times to generate the spectral replicates that were further averaged prior to qualitative analysis and predictions. The four readings were done using the random oversampling motion function of the X,Y Autosampler (within a 3-mm diameter of the sample cup's centroid), configured in AutoPro software (Pike Technologies Inc., Madison, WI).

### 2.2.5 | Number of co-added scans and wavenumber resolution

To test how different acquisition parameters affect the spectral quality and outcomes from mid-DRIFTS, soil samples ( $n = 400$ ) were analyzed using: (a) 8, 16, and 24 co-added scans, and (b) 32  $\text{cm}^{-1}$ , 16  $\text{cm}^{-1}$ , 8  $\text{cm}^{-1}$ , and 4  $\text{cm}^{-1}$  wavenumber resolutions. The same cups/plates loaded with the same soil samples were used for all analysis. Additional details including the number of spectral bands, time of spectra acquisition, and file size are summarized in Table 2. The number of co-added scans was obtained using the greatest resolution (4  $\text{cm}^{-1}$  resolution), and the wavenumber resolution was collected using the greatest number of co-added scans (24 co-added scans). Data was at narrower spacing for zero filling/interpolation purposes.

### 2.2.6 | Spectral properties assessment

Spectral characterization of absorbance ( $\log R^{-1}$ , where  $R$  is reflectance) spectra was summarized using principal components analysis (PCA) with spectra centered to zero mean and unit variance. We used the iterative NIPALS algorithm (Martens & Naes, 1990) to derive the principal components (R package 'chemometrics', Varmuza & Filzmoser, 2009). We plotted the first two principal component scores and loadings to evaluate the effects of treatments on the spectra dispersion and wavenumber-specific loadings distribution. These PCA loadings are useful to identify key wavenumbers that explain variability in the spectra.

To compare the spectral dispersion among the sources of variation, we first evaluated the spectra for heterogeneity using the Levene's test for homogeneity of variances, and then used a permutational multivariate analysis of variance to test for the partitioning of the Euclidean distance matrices (R package 'Vegan', Oksanen, 2019). In both tests, the number of co-added scans and resolution were included as sources of variance, and the first two principal components scores (PC-1 and PC-2) from PCA-NIPALS were evaluated. The spectral scores were dimensioned using

the Mahalanobis distances (Mardia, Kent, & Bibby, 1979; Oksanen, 2019).

Variability in the spectra was assessed by computing the wavenumber-specific standard deviations of the apparent absorbances (Le Guillou et al., 2015) in the four analytical replicates from each sample, and then deriving average standard deviation spectra for each treatment (R package 'stats', R Core Team, 2016).

### 2.2.7 | Qualitative evaluation

Qualitative characterization of mineral and organic functional groups was assessed by integrating peak areas using the local baseline technique, as described by Demyan et al. (2012). The local baseline is a virtual straight line added to the base of peak connecting the peak left and right limits. Local peak areas were determined in the absorbance spectra using the triangle method (R package 'geometry', Sterratt, 2019). Ten functional groups were selected to evaluate the impact of number of co-added scans and resolution. The selected mineral components were the O-H functional group of hydroxyl stretching (kaolinite and others) (peak ranging from/to 3723–3686  $\text{cm}^{-1}$ , Russell, 1987), Si-O functional group of 2:1 layer alumino-silicates (3686–3565  $\text{cm}^{-1}$ , Nguyen, Janik, & Raupach, 1991),  $\text{CO}_3$  functional group of calcite (2650–2420  $\text{cm}^{-1}$ , Nguyen et al., 1991), Si-O functional group of quartz (three peaks between 2080 and 1754  $\text{cm}^{-1}$ , Nguyen et al., 1991; these peaks can potentially overlap with organic functional groups as described in Janik, Skemstad, Shepherd, and Spouncer (2007) and Spaccini & Piccolo, 2007),  $\text{CO}_3$  functional group of carbonates (890–860  $\text{cm}^{-1}$ , Tatzber et al., 2007), fundamental Si-O functional group stretching (854–780  $\text{cm}^{-1}$ , Soda, 1961), and  $\text{CO}_3$  functional group of dolomite (735–723  $\text{cm}^{-1}$ , Tatzber et al., 2007) and calcite (723–707  $\text{cm}^{-1}$ , Tatzber et al., 2007). The selected organic functional groups were the aliphatic C-H of methyl and methylene groups (3010–2800, Orlov, 1986; potentially overlapping with the mineral component calcite, Nguyen et al., 1991), and aromatic C=C functional group stretch and/or asymmetric-COO-stretch (humic and fulvic acid) (1660–1580  $\text{cm}^{-1}$ , Baes & Bloom, 1989).

### 2.2.8 | Prediction model calibration and independent validation

Several spectral treatments were evaluated for ability to extract vibrational information from the spectra, and increase model robustness, accuracy, repeatability, and reproducibility (Stevens & Ramirez-Lopez, 2015). Treatments tested were Savitzky-Golay smoothing and derivative, GapSegment derivative, continuum-removal, detrend normalization,

**TABLE 2** Description of acquisition parameters used on diffuse reflectance infrared Fourier transform spectroscopy (mid-DRIFTS) of soils. The wavenumber range was from 4000 to 700  $\text{cm}^{-1}$

Acquisitions parameters	Co-added scans	Resolution	Number of spectral bands <sup>a</sup>	Time, in seconds per spectrum	File size, in kilobites per spectrum <sup>b</sup>
Co-added scans					
8 scans	8 scans	4 $\text{cm}^{-1}$	825	13.2	217
16 scans	16 scans	4 $\text{cm}^{-1}$	825	27.0	217
24 scans	24 scans	4 $\text{cm}^{-1}$	825	40.2	217
Resolution					
4 $\text{cm}^{-1}$	24 scans	4 $\text{cm}^{-1}$	825	40.2	217
8 $\text{cm}^{-1}$	24 scans	8 $\text{cm}^{-1}$	413	24	109
16 $\text{cm}^{-1}$	24 scans	16 $\text{cm}^{-1}$	206	16.8	54
32 $\text{cm}^{-1}$	24 scans	32 $\text{cm}^{-1}$	103	12.6	27

<sup>a</sup>Data were at narrower spacing for zero filling/interpolation purposes.

<sup>b</sup>File size in comma separated values (.csv) format.

standard normal variate (SNV), block scaling, sum of squares block weighting, standard normal variate transformation (Fearn, 2008), and detrend normalization after filtering (Savitzky–Golay and Gap-Segment) (R package ‘prospectr’, Stevens & Ramirez-Lopez, 2015).

Calibration models were developed on a representative training set (75% of the dataset) selected using the Kennard–Stone sampling algorithm (Kennard & Stone, 1969) to explain  $\geq 95\%$  of the total variance and validated on the remaining samples (25% of the dataset) (R package ‘prospectr’, Stevens & Ramirez-Lopez, 2015). This selection process was separately performed for each one of the four physiographic locations ( $n = 100$  each) for a final calibration set of  $n = 300$  and validation set of  $n = 100$ .

We trained the support vector machine regression (SVM) models with different algorithms on the calibration sets; these were subsequently tested on the independent validation set (R package ‘e1071’, Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2015). Four kernels (classes of algorithms in SVM) were tested, including linear, polynomial (second and third degrees), radial basis, and sigmoid (Karatzoglou, Meyer, & Hornik, 2006). These kernels were tested with all spectral treatments previously mentioned, and both spectra and predictor were scaled to zero mean and unit variance prior to calibration (Meyer et al., 2015). Best models were selected for each variable and treatment based on the lowest root mean squared error ( $\text{RMSE}_v$ ), greatest residual prediction deviation ( $\text{RPD}_v$ ), and greatest coefficient of determination ( $R^2$ ) of the independent validation datasets.

To compare the prediction outcomes obtained with different resolutions and numbers of co-added scans, the  $\text{RMSE}$  of the validation data-sets ( $\text{RMSE}_v$ ) ( $n = 100$ ) were bootstrapped using the ordinary resampling method and 99 replicates (Davison & Hinkley, 1997; R package ‘boot’, Canty, 2017). Multiple means comparison was done by the estimated

marginal means (least-squares means) (R package ‘emmeans’, Lenth, 2019).

## 2.2.9 | Data processing and statistical analyses

Data was processed and analyzed with R version 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria) using the Ohio Supercomputer Center (OSC, 1987) computing resources. The packages used were ‘boot’ (Canty, 2017), ‘chemometrics’ (Varmuza & Filzmoser, 2009), ‘ChemometricsWithR’ (Wehrens, 2011), ‘emmeans’ (Lenth, 2019), ‘e1071’ (Meyer et al., 2015), ‘geometry’ (Sterratt, 2019), ‘prospectr’ (Stevens & Ramirez-Lopez, 2015), ‘Vegan’ (Oksanen, 2019), and ‘stats’ (R Core Team, 2016).

## 3 | RESULTS

### 3.1 | Systematic review

The systematic review identified a wide range of acquisition parameters being used across studies employing mid-DRIFTS for soil analyses (Table 1). These include variations in wavenumber range (ranging 180–7800  $\text{cm}^{-1}$ ), resolution (2–16  $\text{cm}^{-1}$ ), number of co-added scans (16–200 scans), types of spectra replication (in the same or different soil subsamples), numbers of spectral replicates, and different grinding sizes of soils (from 0.1–2 mm). The number of studies which did not report information regarding these parameters was considerable, in particular the type and number of replicates. Out of 32 studies, there were 11 studies that omitted either wavenumber resolution or number of co-added scans and 16 studies which gave no information on spectral replication. Moreover, no clear relationship between



acquisition parameters and predictive performance of soil TC or TOC could be established. This response was anticipated given the many variables that could potentially affect accuracy of predictions from mid-DRIFTS. These include the variability of acquisition parameters (Table 1), but also other factors such as geographical scale, intrinsic variability of soils, sample pre-treatments (e.g., grinding), chemometric methods including multivariate regression models and spectral pretreatments, number of samples, and quality of the conventional measurements. Comparing prediction performances across studies, there were results with great prediction accuracy ( $r^2 = 0.95\text{--}0.99$ ) using relatively wider resolutions ( $8\text{ cm}^{-1}$ ) (Hobley, Baldock, & Wilson, 2016; Wilson, King, Grown, & Veeragathipillai, 2017), while other studies obtained less satisfactory accuracies ( $r^2 = 0.75\text{--}0.78$ ) with high-resolution spectra ( $2\text{ cm}^{-1}$ ) (Chen, Dong, Li, & Wang, 2017; O'Rourke, Minasny, Holden, & McBratney, 2016). Other studies that used similar resolutions (e.g.,  $4\text{ cm}^{-1}$ ) but different number of co-added scans (16 or 64 scans) also obtained improved prediction accuracy ( $r^2 = 0.96\text{--}0.99$ ) (Madhavan et al., 2017; Mirzaeitalarposhti et al., 2015). Finally, studies that used a relatively higher number of co-added scans (100 or 200 scans), did not necessarily improve prediction accuracy ( $r^2 = 0.73\text{--}0.78$ ) (Hutengs, Ludwig, Eisele, & Vohland, 2018; Vohland, Ludwig, Thiele-Bruhn, & Ludwig, 2014).

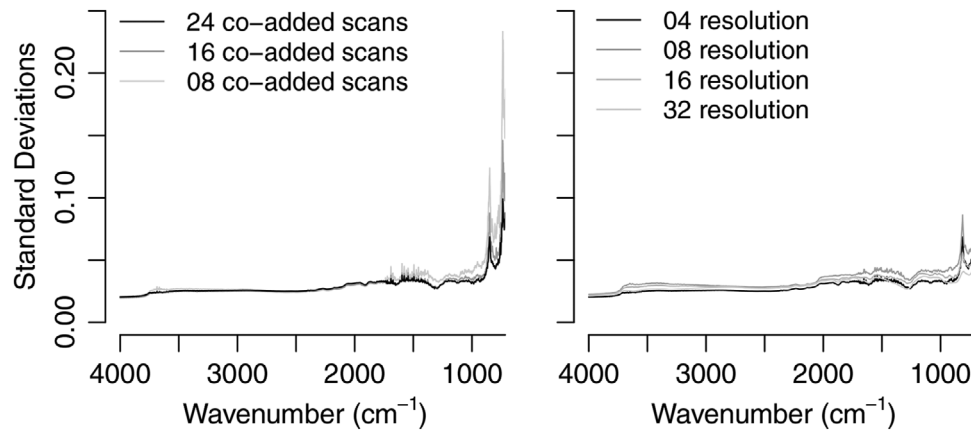
### 3.2 | Spectral properties

Evaluating our experimental results of soils from diverse physiographic regions of the US Midwest, the absorbance spectra and spectral variability were affected by the number of co-added scans and resolution. Decreasing the number of co-added scans increased the presence of spectral artifacts and decreased the mid-DRIFTS sensitivity (resolution) to certain functional groups (Supplemental Fig. S1). Presence of spectral artifacts, that is, intense and sharp absorbance values that do not characterize a peak, increased mostly in wavenumbers  $<1750\text{ cm}^{-1}$ . Moreover, decreasing the number of co-added scans changed the sensitivity to certain functional groups. For instance, the peak attributed to  $\text{CO}_3$  functional group of calcite ( $2650\text{--}2420\text{ cm}^{-1}$ , Nguyen et al., 1991) was not as clearly resolved in the spectra maximum (greatest absorbances in spectra across all wavenumbers) using 8 co-added scans, but the same peak was present in the spectra maximum of 16 and 24 co-added scans (Supplemental Fig. S1). The PCA scores dispersion and loading vectors were not affected by the different numbers of co-added scans (Supplemental Fig. S2). There was no significant segregation among scores of different numbers of co-added scans in the permuta-

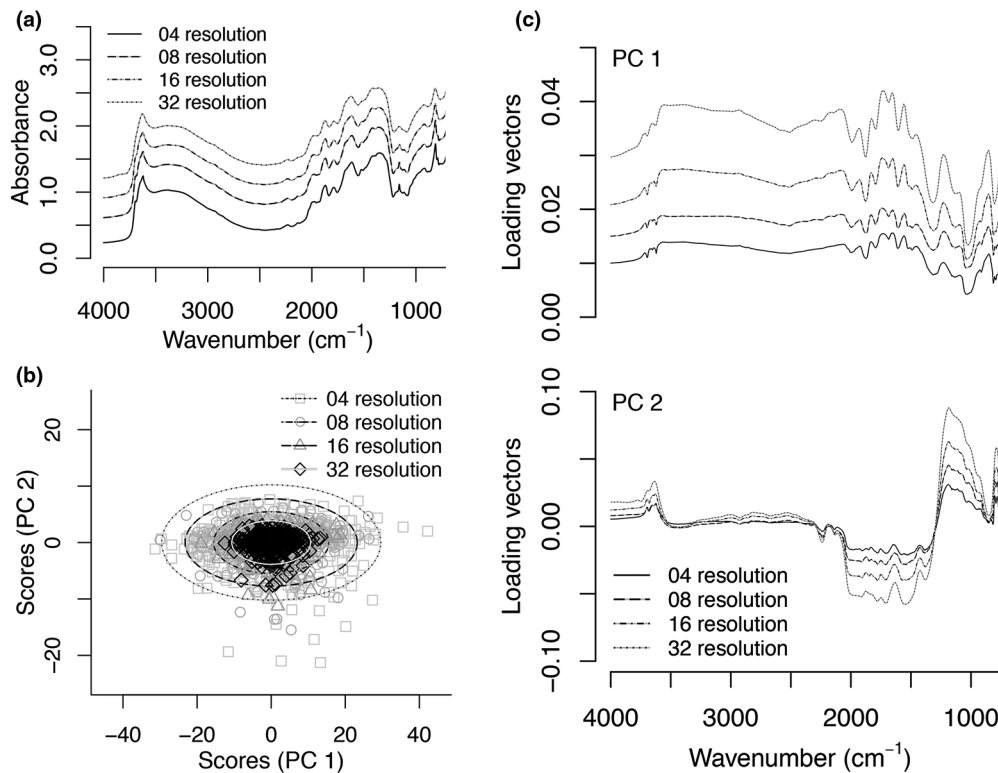
tional multivariate analysis of variance (Heterogeneity Test:  $p = 0.74$ ; MANOVA:  $R^2 = 0.004$ ,  $F = 2.52$ , 1197 degrees of freedom,  $p = 0.07$ ). The loading vectors of the PCA first two components (PC-1 and PC-2) had a similar distribution across the spectra comparing the different numbers of co-added scans. There were subtle differences in loading intensity in specific regions of the spectrum, but loadings mostly overlapped across the mid-infrared using different number of co-added scans.

The spectral variability as given by the specific-wavenumber standard deviations increased by decreasing the number of co-added scans, and it was markedly pronounced in specific regions of the spectrum (Figure 2). The main peaks affected by the increased variability were between wavenumbers  $1750$  and  $1250\text{ cm}^{-1}$ , and wavenumbers  $<900\text{ cm}^{-1}$ . There was composition of several high standard deviations between  $1750$  and  $1250\text{ cm}^{-1}$ , a region of peaks that include C=O functional group vibrations of carboxyl, aldehydes, ketones, and esters groups (Hesse, Meier, & Zeeh, 2005), aromatic C=C stretch and/or asymmetric-COO-stretch of humic and fulvic acid (Baes & Bloom, 1989), and symmetric-COO- stretch and/or-CH bending of aliphatics of fulvic acid. Higher standard deviations were also observed between  $900$  and  $700\text{ cm}^{-1}$ , a region that includes peaks of  $\text{CO}_3$  functional group of carbonates ( $890\text{--}860\text{ cm}^{-1}$ , Tatzber et al., 2007), fundamental Si-O functional group stretching ( $854\text{--}780\text{ cm}^{-1}$ , Soda, 1961), and  $\text{CO}_3$  functional group of both dolomite ( $735\text{--}723\text{ cm}^{-1}$ , Tatzber et al., 2007) and calcite ( $723\text{--}707\text{ cm}^{-1}$ , Tatzber et al., 2007).

Broadening resolution had a smoothing effect on the spectra and consequently a strong influence on specific regions of interest for soil analysis. Notably, smaller peaks were smoothed or flattened when insufficient spatial resolution was used, and peaks in general had a decrease in size, especially in lower resolutions of  $16$  and  $32\text{ cm}^{-1}$ . Additional details about local peak area characterization are in the following section under qualitative characterization of spectra (see *Predictive and Qualitative Outcomes*). There was significant heterogeneity of variances among PCA scores of different resolutions (Heterogeneity Test:  $p < 0.001$ ); however, the segregation among resolutions in the permutational multivariate analysis of variance was not significant, possibly because PCA scores were completely overlapped (MANOVA:  $R^2 < 0.001$ ,  $F < 0.001$  on 1596 degrees of freedom,  $p > 0.99$ ) (Figure 3). The PCA loading vectors were more strongly affected by resolution than the number of co-added scans. Following the previously described characteristics of absorbance spectra, loading vector smoothness increased with broadening resolution, but loadings distribution across the wavenumbers followed a similar trend/shape using different resolutions. There was a clear separation of



**FIGURE 2** Specific-wavenumber standard deviation as affected by the number of co-added scans and spectral resolution in diffuse reflectance infrared Fourier transform spectroscopy in the mid-infrared region (mid-DRIFTS). The standard deviations were determined from four spectral replicates of each soil subsample and later averaged across the 400 soil samples



**FIGURE 3** Absorbance spectra (offset by 0.3 units) (a) and principal component analysis (PCA, NIPALS algorithm) multivariate scores (b) and loading vectors (c) of the first two principal components (PC-1, PC-2) as affected by spectral resolution ( $\text{cm}^{-1}$ ) in diffuse reflectance infrared Fourier transform spectroscopy in the mid-infrared region (mid-DRIFTS)

loading intensity in the first two components of PCA. This was characterized by a complete separation of loadings in PC-1, and more negative or positive intensities in PC-2 as resolutions broadened. Finally, the specific-wavenumber standard deviations of different resolutions followed a similar behavior to the one described for the 24 co-added scans, but with increased smoothness as resolution increased.

### 3.3 | Predictive and qualitative outcomes

The described spectral changes affected the qualitative characterization of the spectra and prediction outcomes from SVM models. The qualitative characterization of mineral and organic functional groups was affected by the number of co-added scans and resolution (Table 3). The Pearson

**TABLE 3** Qualitative characterization of soil mineral and organic functional groups as affected by the number of co-added scans and spectral resolution in diffuse reflectance infrared Fourier transform spectroscopy in the mid-infrared region (mid-DRIFTS). Pearson correlation coefficients are between peak areas of the reference parameters (24 co-added scans or 4 cm<sup>-1</sup> resolution) and different acquisition parameters

Functional groups (peak assignment)	Peak limits		Number of co-added scans			Resolution			
	Left cm <sup>-1</sup>	Right cm <sup>-1</sup>	24	16	8	4 cm <sup>-1</sup>	8	16	32
O-H (3697 cm <sup>-1</sup> )	3723	3686	Ref. <sup>a</sup>	0.99	0.99	Ref.	>0.99	0.97	0.97
Si-O (3630 cm <sup>-1</sup> )	3686	3565	Ref.	>0.99	0.99	Ref.	>0.99	0.98	0.94
Aliphatic C-H (2930 cm <sup>-1</sup> )	3010	2800	Ref.	0.99	0.99	Ref.	>0.99	0.97	0.97
CO <sub>3</sub> (2512 cm <sup>-1</sup> )	2650	2420	Ref.	>0.99	>0.99	Ref.	>0.99	>0.99	>0.99
Si-O/Quartz (1995/1870/1796 cm <sup>-1</sup> )	2080	1754	Ref.	0.98	0.97	Ref.	>0.99	0.96	0.96
Aromatic C = C (1620 cm <sup>-1</sup> )	1660	1580	Ref.	0.93	0.89	Ref.	>0.99	0.91	0.91
CO <sub>3</sub> (875 cm <sup>-1</sup> )	890	860	Ref.	0.90	0.82	Ref.	>0.99	0.86	0.63
Si-O (810 cm <sup>-1</sup> )	854	780	Ref.	0.80	0.70	Ref.	>0.99	0.80	0.78
CO <sub>3</sub> /dolomite (730 cm <sup>-1</sup> )	735	723	Ref.	0.10	0.39	Ref.	>0.99	0.04	-0.13
CO <sub>3</sub> /calcite (713 cm <sup>-1</sup> )	723	707	Ref.	0.20	0.10	Ref.	>0.99	0.17	0.37

<sup>a</sup>Reference parameter.

correlation coefficients showed that as the number of co-added scans was reduced, peaks with smaller areas located at wavenumbers <1660 cm<sup>-1</sup> were less correlated to the peak areas obtained with 24 co-added scans. These less correlated peaks from spectra with fewer co-added scans (16 or 8 scans) included the selected aromatic C=C functional group (1660–1580 cm<sup>-1</sup>, Baes & Bloom, 1989) and carbonate CO<sub>3</sub> functional group (890–860 cm<sup>-1</sup>, Tatzber et al., 2007), mineral silicate Si-O functional group (854–780 cm<sup>-1</sup>, Soda, 1961), and CO<sub>3</sub> functional of both dolomite (735–723 cm<sup>-1</sup>, Tatzber et al., 2007) and calcite (723–707 cm<sup>-1</sup>, Tatzber et al., 2007).

The qualitative characterization of mineral and organic functional groups was also affected by broadening resolutions. No difference was found across all evaluated peaks between resolutions 4 and 8 cm<sup>-1</sup>, but as resolution decreased to 16 and 32 cm<sup>-1</sup>, peaks located at wavenumbers <1660 cm<sup>-1</sup> were less correlated to peaks obtained with 4 cm<sup>-1</sup> resolution. These affected peaks were from the CO<sub>3</sub> functional group of carbonates (890–860 cm<sup>-1</sup>, Tatzber et al., 2007), fundamental Si-O functional group (854–780 cm<sup>-1</sup>, Soda, 1961), and CO<sub>3</sub> functional group of both dolomite (735–723 cm<sup>-1</sup>, Tatzber et al., 2007) and calcite (723–707 cm<sup>-1</sup>, Tatzber et al., 2007). Moreover, most local peak areas were reduced in broader resolutions when compared to high-resolution spectra (i.e., 4 cm<sup>-1</sup>), with the degree of reduction depending on each measured peak. The degree of reductions in peak areas was determined from the slope of the linear relation (null intercept) between peak areas with 4 cm<sup>-1</sup> resolution versus peak areas with the broader resolution treatments. For example, Si-O (3686–3565 cm<sup>-1</sup>, Nguyen et al., 1991), one of the most severely affected peaks in terms of local peak area size reduction when compared to the resolution of 4 cm<sup>-1</sup>, was reduced by 36.3% with 16 cm<sup>-1</sup> resolu-

tion and 98.3% with 32 cm<sup>-1</sup> resolution, but did not change (slope = 1) when compared to 8 cm<sup>-1</sup> resolution.

Increasing the number of co-added scans, from 8 to 24, improved prediction accuracy for most soil variables in independent validation sets (Table 4; Figure 4). Based on validation coefficients, greatest prediction accuracies were obtained for soil pH, TOC, and POXC with 24 co-added scans, while at least 16 co-added scans were needed for clay and sand (Figure 4). Calibration coefficients did not necessarily follow the same magnitudes of change as the validation sets, but better accuracies were found with at least 16 co-added scans for sand, TOC, and POXC.

In contrast, broadening resolution from 4 to 32 cm<sup>-1</sup> improved prediction accuracy for all soil variables in independent validation sets (Table 5; Figure 4). Calibration coefficients had a more skewed distribution as affected by resolution, and better accuracies were not necessarily obtained from spectra with broader resolution. These differences between calibration and validation sets, as affected by either the number of co-added scans or resolution, could be due to the criteria used to select best performing models, which was done based on validation statistical coefficients. Moreover, differences in distribution of soil variables between those two sets, characterized by wider ranges and variability in calibration sets, could have contributed to differences between calibration and validation coefficients in response to mid-DRIFTS acquisition parameters.

## 4 | DISCUSSION

Diffuse reflectance infrared Fourier transform spectroscopy is an effective method for rapid soil analysis for both

**TABLE 4** Soil properties prediction performance using different number of co-added scans in diffuse reflectance infrared Fourier transform spectroscopy in the mid-infrared region (mid-DRIFTS). Spectra were acquired with 4 cm<sup>-1</sup> resolution and four spectral replicates in each soil sample

Soil property	Co-added scans	Calibration (training set $n = 300$ )			Validation (test set $n = 100$ )		
		RMSE <sup>a</sup>	RPD	$R^2$	RMSE <sub>V</sub>	RPD <sub>V</sub>	$R^2_{V}$
Clay (g kg <sup>-1</sup> )	8 scans	32.5	4.46	0.96	38.4	2.98	0.87
	16 scans	51.2	2.81	0.90	35.1	3.25	0.91
	24 scans	32.4	4.45	0.96	40.8	2.08	0.82
Sand (g kg <sup>-1</sup> )	8 scans	80.2	2.64	0.89	86.5	1.97	0.80
	16 scans	39.0	5.82	0.97	82.2	2.05	0.79
	24 scans	50.3	4.26	0.95	79.6	2.34	0.87
pH	8 scans	0.30	3.14	0.92	0.49	1.77	0.78
	16 scans	0.42	2.20	0.85	0.48	1.98	0.79
	24 scans	0.44	2.16	0.85	0.44	2.12	0.81
Total organic C (%)	8 scans	0.72	1.00	0.67	0.20	2.60	0.89
	16 scans	0.56	1.72	0.80	0.20	3.07	0.89
	24 scans	0.55	1.70	0.80	0.16	4.04	0.93
Permanganate oxidizable C (mg kg <sup>-1</sup> )	8 scans	177	1.11	0.70	93	1.55	0.79
	16 scans	101	2.63	0.90	80	2.78	0.89
	24 scans	105	2.45	0.89	72	3.07	0.92

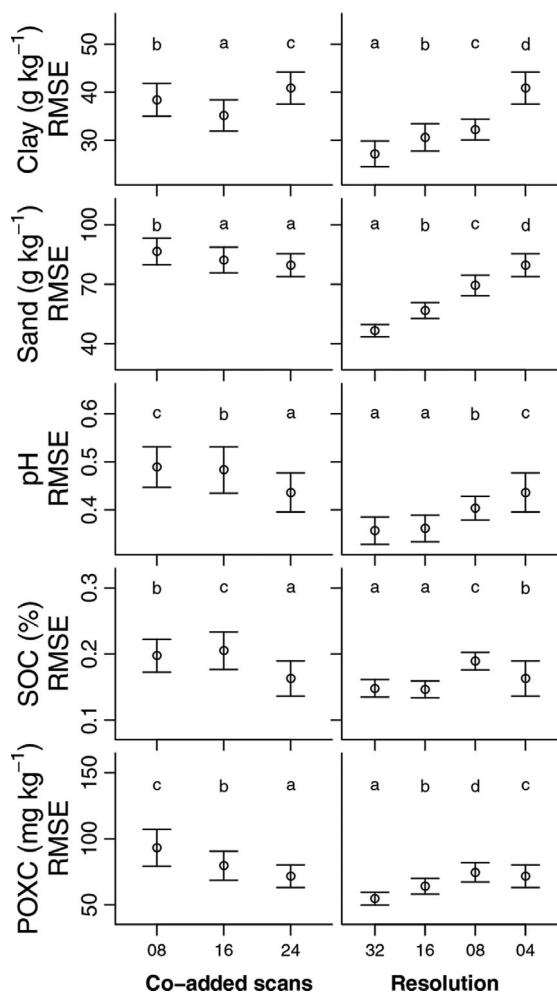
<sup>a</sup>RMSE, root mean squared error; RPD, residual prediction deviation;  $R^2$ , coefficient of determination.

qualitative and quantitative purposes. This is mostly due to the potential of obtaining accurate results along with the often-reduced time, labor, and costs when estimating soil properties (Gholizadeh et al., 2013; Soriano-Disla, Janik, Viscarra Rossel, MacDonald, & McLaughlin, 2014). These factors have enabled scientists and other stakeholders to work with larger sets of soil samples, in successive samplings, ultimately leading to a better understanding of ecosystems functioning and services as driven by soil processes (Adhikari & Hartemink, 2016; Vereecken et al., 2016). These factors have promoted mid-DRIFTS usage across the globe, but at the same time, methods associated with the technique have been used indiscriminately. This can be verified by the wide range of experimental approaches that have been used in mid-DRIFTS of soils during the last years, but also to the lack of information in some specific studies regarding these experimental parameters (Table 1). Though this does not mean that a universal set of acquisition parameters should be used in all studies, but rather there should be clarity on the effects of such parameters, reporting of parameter details, and justification for their selection. This is especially the case when parameters are further away from the current literature based norms of spectra acquisition, such as lower number of co-added scans, broader resolutions, lower number of replicates or lack thereof, and coarser grinding sizes.

The comparison across different studies showed that there is no clear effect of the experimental parameters in mid-DRIFTS prediction performance (Table 1), likely due to the

potential influence of many factors that can affect predictive outcomes. These factors are soil sample properties, sample preparation, instrument utilization, and chemometrics, all of which can potentially affect outcomes from mid-DRIFTS. Yet even with such variability, there is no clarity on what experimental acquisition parameters should be used to produce accurate outcomes from mid-DRIFTS of soils, and moreover, what parameters can be optimized to make the technique even more efficient when measuring soil samples. If there is no difference in using broader resolutions and/or lower number of co-added scans, that means that those parameters could be optimized to reduce time and computation demand while acquiring spectra to improve efficiency of measuring soil properties in mid-DRIFTS. However, such optimization should be understood and quantified, a priori, on how adjusting acquisition parameters leads to changes on spectral quality, qualitative characterization of spectra, and prediction performance.

Our experimental findings indicated that both the number of co-added scans and resolution affected spectral quality and outcomes from mid-DRIFTS of soils. Increasing the number of co-added scans from 8 to 24 decreased both the variability and presence of spectral artifacts (Figure 2; Supplemental Fig. S1 and S2). This effect increased the ability to characterize peaks at specific regions of the spectrum (Table 3) and improved multivariate model prediction performance for most soil variables (Table 4; Figure 4). The spectral resolution had a contrasting effect to the number of co-added scans, indicating possibilities for increasing laboratorial efficiency. This



**FIGURE 4** Bootstrapped root mean squared error ( $RMSE_V$ ) of independent validation sets ( $n = 100$ ) as affected by the number of co-added scans and spectral resolutions ( $cm^{-1}$ ) in diffuse reflectance infrared Fourier transform spectroscopy in the mid-infrared region (mid-DRIFTS)

improvement was verified for the predictive applications of mid-DRIFTS, up to the broad resolution of  $32\text{ cm}^{-1}$  (Table 5; Figure 4). However, broadening spectral resolution reduced the ability to identify mineral and organic functional groups based on specific spectral peaks (Table 3). Local peak area quantification was limited by broadening spectral resolution to more than  $8\text{ cm}^{-1}$  (i.e., 16 and  $32\text{ cm}^{-1}$ ).

The number of co-added scans had a direct impact over spectral noise, presence of artifacts, and resolving peaks (Figure 2; Supplemental Fig. S1). The peak resolve is particularly important for peaks with smaller local areas from functional groups occurring in lower concentration in soils (Table 3). In general, soils are predominantly composed of minerals, and to lesser extent organic functional groups. This means when analyzing a soil sample in mid-DRIFTS, mineral absorbances dominate the spectrum. When an insufficient number of co-added scans is used, either organic or mineral

functional groups that are present in smaller proportions in soils can potentially be overlapped by those predominant components. Similarly, instrumental noise can limit detection of specific functional groups if noise is greater than spectral signals. Therefore, better representing mid-infrared active spectral features requires a certain number of co-added scans that comprehensively resolves peaks and overcomes instrument noise. According to our experimental results, fewer numbers of co-added scans negatively impacted spectral quality (Figure 2; Supplemental Fig. S1 and S2) and reduced the ability to generate more accurate qualitative (Table 3) and predictive outcomes from mid-DRIFTS (Table 4; Figure 4). This predictive deterioration with less co-added scans was verified for all soil variables in independent validation sets and most variables in calibration sets.

Broadening resolution improved predictions for all evaluated soil properties (Table 5; Figure 4), but at the same time compromised characterization of peaks with smaller local peak areas (Table 3). Predictive improvements with broader resolutions were verified for all soil variables in independent validation sets, while calibration coefficients had a less defined variation as affected by resolution. The prediction performance could be compromised when resolution was broadened to the point where there was degradation of specific peaks in the spectra. However, our results showed that even with significant changes in peak areas with broadening spectral resolution (Table 3), the prediction performance for all tested soil variables improved with broader resolutions (Table 5; Figure 4). This suggests models are using information from different parts of the spectrum that are not necessarily related to those peaks/functional groups (surrogate calibrations), or models are artificially learning to interpret and be resilient to changes in peak sizes and, or, absence of peaks. Some properties can be totally or partially predicted in multivariate models because of their correlation or covariation with other soil properties (Chang, Laird, Mausbach, & Hurburgh, 2001; Stenberg, Viscarra Rossel, Mouazen, & Wetterlind, 2010). Following this rationale, TOC concentrations are generally related to soil texture and mineralogy, and therefore TOC could be indirectly predicted by variations in peaks attributed to, for example, O-H functional group of hydroxyl stretching (kaolinite and others) ( $3723\text{--}3686\text{ cm}^{-1}$ , Russell, 1987), Si-O functional group of 2:1 layer aluminosilicates ( $3686\text{--}3565\text{ cm}^{-1}$ , Nguyen et al., 1991). Moreover, if the peak areas change in magnitude as resolution broadened, but there is still a correlation between specific peak areas in narrow resolution and broad resolution spectra (Table 3), it is expected that those models will still rely on those same peaks to interpret spectra and predict soil properties.

In addition, reductions in redundancy of information from different wavenumbers as well as in spectral dispersion and variability may have enabled models to predict soil variables more accurately as resolution broadened.



**TABLE 5** Soil properties prediction performance using different spectral resolutions in diffuse reflectance infrared Fourier transform spectroscopy in the mid-infrared region (mid-DRIFTS). Spectra were acquired with 24 co-added scans and four spectral replicates in each soil sample

Soil property	Resolution, $\text{cm}^{-1}$	Calibration (training set $n = 300$ )			Validation (test set $n = 100$ )		
		RMSE <sup>a</sup>	RPD	$R^2$	RMSE	RPD	$R^2$
Clay ( $\text{g kg}^{-1}$ )	4	32.4	4.45	0.96	40.8	2.08	0.82
	8	55.6	2.52	0.88	32.2	3.36	0.92
	16	35.8	4.15	0.95	30.6	3.48	0.92
	32	41.1	3.62	0.94	27.1	4.41	0.95
Sand ( $\text{g kg}^{-1}$ )	4	50.3	4.26	0.95	79.6	2.34	0.87
	8	55.7	3.92	0.94	69.3	2.70	0.88
	16	59.7	3.81	0.94	56.6	2.20	0.85
	32	63.5	3.62	0.93	46.6	3.22	0.91
pH	4	0.44	2.16	0.85	0.44	2.12	0.81
	8	0.30	3.65	0.93	0.40	2.58	0.84
	16	0.43	2.33	0.85	0.33	2.55	0.88
	32	0.38	2.52	0.88	0.35	2.47	0.87
Total organic C (%)	4	0.55	1.70	0.80	0.16	4.04	0.93
	8	0.57	1.62	0.79	0.19	3.51	0.92
	16	0.56	1.70	0.80	0.15	4.22	0.94
	32	0.58	1.64	0.78	0.15	4.47	0.94
Permanganate oxidizable C ( $\text{mg kg}^{-1}$ )	4	105	2.45	0.89	72	3.07	0.92
	8	84	3.00	0.93	74	2.82	0.87
	16	88	3.25	0.92	64	3.44	0.92
	32	92	3.08	0.91	54	4.59	0.95

<sup>a</sup>RMSE, root mean squared error; RPD, residual prediction deviation;  $R^2$ , coefficient of determination.

Peng et al. (2014) reported that spectral data with narrow, numerous wavenumbers are generally redundant and can decrease efficiency of variable selection and modeling when predicting soil properties. According to our results, this redundancy effect was more clearly expressed in the PCA loading vectors as affected by resolution (Figure c). Particularly in the PC-2, loadings peaks were more intense, either negative or positive, and smoother in broader resolutions, and explained a greater extent of spectral variability. Spectral redundancy can also be reduced by selecting/eliminating specific frequencies of the spectrum (e.g., Calderón et al., 2017; Giacometti et al., 2013), but this was not evaluated in our study.

Most studies that have evaluated effects of spectral resolution on analysis of soil samples have employed the near-infrared range of spectrum, and using spectral resampling techniques to test different resolutions (e.g., Adeline et al., 2017; Knadel et al., 2013; Peng et al., 2014), rather than acquiring spectra with different resolutions with the instrument. This pseudo-resolution approach can move the experiment away from actual analytical evaluation (Blitz & Klarup, 2002), if not properly calibrated or adjusted according to the results obtained from the instrument. Despite issues with resampling methods as opposed to actual instrument measurements, most evaluations showed no or positive effects

of broadening resolutions when predicting soil properties using the near-infrared region ( $12000\text{--}4000\text{ cm}^{-1}$ ). Peng et al. (2014) evaluated spectral sampling intervals from 1 to 10 nm spectra and established that 9 nm best predicted soil TOC content using both SVM and partial least squares. Knadel et al. (2013) and Mouazen, Saeys, Xing, de Baerdemaeker, and Ramon (2005) found slight differences in prediction performance for clay content, TOC, and soil moisture by comparing spectrometers with different resolutions (1 to 10 nm) among other spectral specifications and instrument variations. Yang et al. (2012) showed that reduction to 21 resampled spectral bands with 100 nm uniform spectral interval did not alter the prediction performance of soil nitrogen and carbon. Lastly, Adeline et al. (2017) showed that broadening resolution up to 60 nm did not affect prediction performances of soil physical and chemical properties. These authors argued that predictions could be distinguished according to their sensitivity of soil variables to the spectral degradation, in which clay, iron oxides and  $\text{CaCO}_3$  were driven by specific changes in absorption features that are directly related to those properties, whereas pH—a variable without direct presence of functional groups and thus specific peaks—relied on variations of spectral features of other functional groups (Adeline et al., 2017).

Comparison of prediction results (Tables 4 and 5) to the other results identified in the systematic review (Table 1)

demonstrates that we achieved comparable or better prediction performances than most studies previously reported. Prediction accuracy can be maintained or increased while reducing time of spectra acquisitions and lowering computing demand (Table 2). Based on this evaluation, we recommend optimizing experimental settings to at least 24 co-added scans and  $8\text{ cm}^{-1}$  or higher resolution for quantitative and qualitative applications to minimize analysis costs and maximize prediction accuracy of soil properties in mid-DRIFTS.

## 5 | CONCLUSION

There is a wide variation of approaches being used for spectra acquisition in mid-DRIFTS of soils, but the lack of a priori determination of acquisition parameters means that hypothesized trade-offs may not be optimized. Establishing best practices for mid-DRIFTS of soils is a high priority to enable comparability of mid-DRIFTS results in soil testing. While a single set of parameters may not be appropriate for all soils and applications, more systematic approaches to identify practices that maximize precision, accuracy, and reproducibility of mid-DRIFTS are needed in soil science. There should be more scrutiny during the peer-review process on reporting information of experimental acquisition parameters. Our results suggest the number of co-added scans and resolution can markedly affect spectral quality from mid-DRIFTS of soils, and that different acquisition parameters can be used for qualitative or predictive applications. At least 24 co-added scans may be necessary for both qualitative and predictive applications of mid-DRIFTS. A resolution of  $8\text{ cm}^{-1}$  allows using mid-DRIFTS for quantitative and qualitative applications, but resolutions can be broadened to up to  $32\text{ cm}^{-1}$  for predictive applications only. The number of co-added scans and spectral resolution dictate total spectra acquisition time and computational needs for data processing, and therefore can significantly affect laboratory capabilities to measure soil properties, especially for large sample sets. Analytical efficiency can be optimized by setting acquisition parameters to maximize qualitative and predictive applications of mid-DRIFTS in soil analysis. Below we provide recommendations for optimizing mid-DRIFTS for the purposes of measuring soil properties.

### 5.1 | Recommendations for optimizing mid-DRIFTS to measure soil properties

#### (I) Sample preparation

- a. **Sieving/grinding:** Soil should be homogenized and made to pass a  $<2.0\text{-mm}$  sieve (Deiss, Demyan, & Culman, 2020; Janik, Soriano-Disla, Forrester, & McLaughlin, 2016). Finely grinding soils (e.g.,  $<0.5\text{ mm}$ ) often improves prediction accuracy of

soil properties (Le Guillou et al., 2015; Deiss et al., 2020). Recommendations on how much grinding is needed should consider the time and effort needed to prepare soil samples as well as purpose of analysis (Le Guillou et al., 2015).

- b. **Drying:** Oven-drying samples until constant dry mass immediately prior to analysis to minimize water interference on spectra.
- c. **Loading process:** Do not pack or compress soil into the well other than by gentle tapping and surface smoothing by scrapping to avoid artifacts of matrix density (Terhoeven-Urselmans et al., 2010).

#### (II) Instrument setup for spectra acquisition

- a. **Types and number of analytical replicates** (soil subsamples/spectra replicates): We recommend at least three replicates when using different soil subsamples (i.e., different wells) and at least four spectral replicates when using the same soil subsample (Deiss et al., 2020; Peng et al., 2014).
- b. **Background:** potassium bromide (KBr) is a highly hygroscopic material, and therefore KBr should be dried beforehand and cooled/stored in a desiccator, and KBr background spectrum should be collected often, e.g., at every  $\sim 20$  to 25 samples.
- c. **Wavenumber range:** Spectra are generally acquired within the wavenumber range  $4000$  to  $400\text{ cm}^{-1}$ . Further reduction of this range may be necessary to eliminate noise at the boundaries of the MIR spectrum (edge of detector limit), e.g.,  $4000$  to  $700\text{ cm}^{-1}$ .
- d. **Number of co-added scans:** At least 24 co-added scans may be necessary for both qualitative and predictive applications of mid-DRIFTS.
- e. **Spectral resolution:** Spectra resolution choices may depend on purpose of spectra utilization. A resolution of at least  $8\text{ cm}^{-1}$  allows using mid-DRIFTS for quantitative and qualitative applications, but resolutions can be broadened to up to  $32\text{ cm}^{-1}$  for predictive applications only.

#### (III) Data processing (chemometrics)

- a. **Spectral treatment:** Several mathematical spectral treatments should be tested to determine which treatment(s) most enhances spectral features and increase multivariate regression models robustness, accuracy, repeatability, and reproducibility (Gholizadeh et al., 2013; Stevens & Ramirez-Lopez, 2015). Spectral treatment may be soil dataset- and variable-specific.
- b. **Multivariate model:** There are two classes of multivariate models: linear and the nonlinear (Wehrens, 2011) and their accuracies may depend on targeted soil set and soil variable. Model choice should be based on the accuracy of predictions on independent validation sets.

- c. **Calibration (training set):** Calibration models must be developed on a representative portion of samples based on location, soil variable distribution, spectral properties, and any other source of variation in a dataset.
- d. **Validation (test set):** A representative and independent validation set must be used. Though common, cross-validation—using the same samples in the calibration set for the validation set—is not recommended for predicting unknown future samples as cross-validation likely overestimates calibration stability (Stenberg et al., 2010; Stumpe et al., 2011).
- e. **Prediction of variables of new samples** (unknown measured values): New samples must be spectrally similar to the ones used in the spectral library to train models (e.g., based on Euclidean distances of a spectral matrix) (Hicks, Viscarra Rossel, & Tuomi, 2015; Mirzaeitalarposhti, Demyan, Rasche, Cadisch, & Müller, 2017). A continuous monitoring scheme to validate predicted results with measured traditional methods (e.g., wet chemistry) should be established.

## ACKNOWLEDGMENTS

We thank Bethany Herman, Nicole C. Hoekstra, Thomas Doohan, Mason Gingery, Madison Campbell, and Meredith Mann for their help in the laboratory. This research was supported by the Foundation for Food and Agricultural Research and the School of Environment and Natural Resources at Ohio State University.

## ORCID

Leonardo Deiss  <https://orcid.org/0000-0003-2001-9238>

Andrew J. Margenot  <https://orcid.org/0000-0003-0185-8650>

Steve W. Culman  <https://orcid.org/0000-0003-3985-257X>

M. Scott Demyan  <https://orcid.org/0000-0001-6198-3774>

## REFERENCES

- Adeline, K. R. M., Gomez, C., Gorretta, N., & Roger, J. M. (2017). Predictive ability of soil properties to spectral degradation from laboratory vis-NIR spectroscopy data. *Geoderma*, 288, 143–153. <https://doi.org/10.1016/j.geoderma.2016.11.010>
- Adhikari, K., & Hartemink, A. E. (2016). Linking soils to ecosystem services—A global review. *Geoderma*, 26, 101–111. <https://doi.org/10.1016/j.geoderma.2015.08.009>
- Araújo, S. R., Söderström, M., Eriksson, J., Isendahl, C., Stenberg, P., & Demattê, J. A. M. (2015). Determining soil properties in Amazonian Dark Earths by reflectance spectroscopy. *Geoderma*, 237–238, 308–317. <https://doi.org/10.1016/j.geoderma.2014.09.014>
- Azambre, B., Heintz, O., Schneider, M., Krzton, A., & Weber, J. V. (1999). Optimization of some instrumental factors in diffuse reflectance infrared Fourier transform spectroscopy. *Talanta*, 50, 359–365. [https://doi.org/10.1016/S0039-9140\(99\)00026-0](https://doi.org/10.1016/S0039-9140(99)00026-0)
- Baes, A. U., & Bloom, P. R. (1989). Diffuse reflectance Fourier transform infrared (DRIFT) of humic and fulvic acids. *Soil Science Society of America Journal*, 53, 695–700. <https://doi.org/10.2136/sssaj1989.03615995005300030008x>
- Baldock, J. A., Hawke, B., Sanderman, J., & MacDonald, L. M. (2013). Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra. *Soil Research*, 51, 577–595. <https://doi.org/10.1071/SR13077>
- 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, 468–480. <https://doi.org/10.1071/SR17221>
- Barthès, B. G., Kouakoua, E., Moulin, P., Hmaid, K., Gallali, T., Clairotte, M., ... Chevallier, T. (2016). Studying the physical protection of soil carbon with quantitative infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 24, 199–214. <https://doi.org/10.1255/jnirs.1232>
- Blitz, J. P., & Klarup, D. G. (2002). Information in the instrumental analysis laboratory. *Journal of Chemical Education*, 79, 1358–1360. <https://doi.org/10.1021/ed079p1358>
- Burt, R. (2011). Soil survey laboratory information manual. Soil Survey Investigations Report No. 45 (Version 2.0). USDA, Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE.
- Calderón, F. J., Culman, S., Six, J., Franzluebbers, A. J., Schipanski, M., Beniston, J., ... Kong, A. Y. Y. (2017). Quantification of soil permanganate oxidizable carbon (POXC) using infrared spectroscopy. *Soil Science Society of America Journal*, 81, 277–288. <https://doi.org/10.2136/sssaj2016.07.0216>
- Canty, A. (2017). The Package ‘boot’: Reference manual. The Comprehensive R Archive Network, R Development Core Team, Vienna. <https://cran.r-project.org/web/packages/boot/boot.pdf> (accessed 24 Jan. 2019).
- Cazes, J. (2004). *Analytical instrumentation handbook*. Boca Raton: CRC Press. <https://doi.org/10.1201/9780849390395>
- 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, 480–490. <https://doi.org/10.2136/sssaj2001.652480x>
- Chen, C., Dong, D., Li, Z., & Wang, X. (2017). A novel soil nutrient detection method based on combined ATR and DRIFT mid-infrared spectra. *Analytical Methods*, 9, 528–533. <https://doi.org/10.1039/C6AY02904C>
- Clairotte, M., Grinand, C., Kouakoua, E., Thébault, A., Saby, N. P. A., Bernoux, M., & Barthès, B. G. (2016). National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy. *Geoderma*, 276, 41–52. <https://doi.org/10.1016/j.geoderma.2016.04.021>
- Coûteaux, M. M., Berg, B., & Rovira, P. (2003). Near-infrared reflectance spectroscopy for determination of organic matter fractions including microbial biomass in coniferous forest soils. *Soil Biology and Biochemistry*, 35, 1587–1600. <https://doi.org/10.1016/j.soilbio.2003.08.003>
- Culman, S. W., Freeman, M., & Snapp, S. S. (2012). *Procedure for the determination of permanganate oxidizable carbon*. Hickory

- Corners, MI: Long Term Ecological Research Protocols, Kellogg Biological Station.
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge Univ. Press. <https://doi.org/10.1017/CBO9780511802843>
- Deiss, L., Demyan, M. S., & Culman, S. W. (2020). Grinding and spectra replication often improves mid-DRIFTS predictions of soil properties. *Soil Sci Soc Am J*, 2019, 1–16. <https://doi.org/10.1002/saj2.20021>
- 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, 189–199. <https://doi.org/10.1111/j.1365-2389.2011.01420.x>
- Demyan, M. S., Rasche, F., Schütt, M., Smirnova, N., Schulz, E., & Cadisch, G. (2013). Combining a coupled FTIR-EGA system and in situ DRIFTS for studying soil organic matter in arable soils. *Biogeochemistry*, 10, 2897–2913. <https://doi.org/10.5194/bg-10-2897-2013>
- Dreimanis, A. (1962). Quantitative gasometric determination of calcite and dolomite by using Chittick apparatus. *Journal of Sedimentary Research*, 32, 520–529. <https://doi.org/10.1306/74D70D08-2B21-11D7-8648000102C1865D>
- Fearn, T. (2008). The interaction between standard normal variate and derivatives. *NIR News*, 19, 16–17. <https://doi.org/10.1255/nirn.1098>
- Gholizadeh, A., Luboš, B., Saberioon, M., & Vašát, 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, 1349–1362. <https://doi.org/10.1366/13-07288>
- Giacometti, C., Demyan, M. S., Cavani, L., Marzadori, C., Ciavatta, C., & Kandeler, E. (2013). Chemical and microbiological soil quality indicators and their potential to differentiate fertilization regimes in temperate agroecosystems. *Applied Soil Ecology*, 64, 32–48. <https://doi.org/10.1016/j.apsoil.2012.10.002>
- Griffiths, P. R. (1972). “Trading rules” in infrared Fourier transform spectroscopy. *Analytical Chemistry*, 44, 1909–1913. <https://doi.org/10.1021/ac60319a015>
- Griffiths, P. R., & Haseth, J. A. (1986). *Fourier transform infrared spectrometry*. New York: John Wiley & Sons. <https://doi.org/10.1002/bbpc.19860901224>
- Hanssen, L. M. (1993). Parameters for an infrared diffuse reflectance standard. *Optical Engineering*, 32, 877–879. <https://doi.org/10.1117/12.61216>
- Henaka Arachchi, M. P. N. K., Field, D. J., & McBratney, A. B. (2016). Quantification of soil carbon from bulk soil samples to predict the aggregate-carbon fractions within using near- and mid-infrared spectroscopic techniques. *Geoderma*, 267, 207–214. <https://doi.org/10.1016/j.geoderma.2015.12.030>
- Hesse, M., Meier, H., & Zeeh, B. (2005). Spe ktroskopische Methoden in der Organischen Chemie. (In German.) Georg Thieme Verlag, Stuttgart. <https://doi.org/10.1002/pauz.19960250417>
- Hicks, W., Viscarra Rossel, R. A., & Tuomi, S. (2015). Developing the Australian mid-infrared spectroscopic database using data from the Australian Soil Resource Information System. *Soil Research*, 53, 922–931. <https://doi.org/10.1071/SR15171>
- Hobley, E. U., Baldock, J., & Wilson, B. (2016). Environmental and human influences on organic carbon fractions down the soil profile. *Agriculture, Ecosystems & Environment*, 223, 152–166. <https://doi.org/10.1016/j.agee.2016.03.004>
- Hutengs, C., Ludwig, B., Eisele, A., & Vohland, M. (2018). Comparison of portable and bench-top spectrometers for mid-infrared diffuse reflectance measurements of soils. *Sensors (Basel)*, 18, 993. <https://doi.org/10.3390/s18040993>
- Janik, L. J., Skemstad, J. O., Shepherd, K. D., & Spouncer, L. R. (2007). The prediction of soil carbon fractions using mid-infrared-partial least square analysis. *Australian Journal of Soil Research*, 45, 73–81. <https://doi.org/10.1071/SR06083>
- 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, 889–904. <https://doi.org/10.1071/SR16011>
- Jia, X., Chen, S., Yang, Y., Zhou, L., Yu, W., & Shi, Z. (2017). Organic carbon prediction in soil cores using VNIR and MIR techniques in an alpine landscape. *Scientific Reports*, 7, 2144. <https://doi.org/10.1038/s41598-017-02061-z>
- Karatzoglou, A., Meyer, D., & Hornik, K. (2006). Support vector machines in R. *Journal of Statistical Software*, 15, 1–28. <https://doi.org/10.18637/jss.v015.i09>
- Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics*, 11, 137–148. <https://doi.org/10.1080/00401706.1969.10490666>
- Kimber, J. A., & Kazarian, S. G. (2017). Spectroscopic imaging of biomaterials and biological systems with FTIR microscopy or with quantum cascade lasers. *Analytical and Bioanalytical Chemistry*, 409, 5813–5820. <https://doi.org/10.1007/s00216-017-0574-5>
- Knadel, M., Stenberg, B., Deng, F., Thomsen, A. G., & Greve, M. H. (2013). Comparing predictive abilities of three visible-near infrared spectrophotometers for soil organic carbon and clay determination. *Journal of Near Infrared Spectroscopy*, 21, 67–80. <https://doi.org/10.1255/jnirs.1035>
- Knox, N. M., Grunwald, S., McDowell, M. L., Bruland, G. L., Myers, D. B., & Harris, W. G. (2015). Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. *Geoderma*, 239–240, 229–239. <https://doi.org/10.1016/j.geoderma.2014.10.019>
- Le Guillou, F., Wetterlind, W., Viscarra Rossel, R. A., Hicks, W., Grundy, M., & Tuomi, S. (2015). How does grinding affect the mid-infrared spectra of soil and their multivariate calibrations to texture and organic carbon? *Soil Research*, 53, 913–921. <https://doi.org/10.1071/SR15019>
- Lenth, R. (2019). The Package ‘emmeans’: Reference manual. The Comprehensive R Archive Network, R Development Core Team, Vienna. <https://cran.r-project.org/web/packages/emmeans/emmeans.pdf> (accessed 25 Jan. 2019).
- Ludwig, B., Sawallisch, A., Heinze, S., Joergensen, R. G., & Vohland, M. (2015). Usefulness of middle infrared spectroscopy for an estimation of chemical and biological soil properties: Underlying principles and comparison of different software packages. *Soil Biology and Biochemistry*, 86, 116–125. <https://doi.org/10.1016/j.soilbio.2015.03.015>
- Ludwig, B., Linsler, D., Höper, H., Schmidt, H., Piepho, H. P., & Vohland, M. (2016). Pitfalls in the use of middle-infrared spectroscopy: Representativeness and ranking criteria



- for the estimation of soil properties. *Geoderma*, 268, 165–175. <https://doi.org/10.1016/j.geoderma.2016.01.010>
- Madhavan, D. B., Baldock, J. A., Read, Z. J., Murphy, S. C., Cunningham, S. C., Perring, M. P., ... Baker, T. G. (2017). Rapid prediction of particulate, humus and resistant fractions of soil organic carbon in reforested lands using infrared spectroscopy. *Journal of Environmental Management*, 193, 290–299. <https://doi.org/10.1016/j.jenvman.2017.02.013>
- 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>
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis.*, New York: Academic press. <https://doi.org/10.1002/bimj.4710240520>
- Martens, H., & Naes, T. (1990). *Multivariate calibration.*, Chichester: John Wiley & Sons. <https://doi.org/10.1002/cem.1180040607>
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2015). e1071: Misc functions of the Department of Statistics. R Foundation for Statistical Computing, Vienna. <https://CRAN.R-project.org/package=e1071> (accessed 21 June 2017).
- Mirzaeitalarposhti, R., Demyan, M. S., Rasche, F., Poltoradnev, M., Cadisch, G., & Müller, T. (2015). MidDRIFTS-PLSR-based quantification of physico-chemical soil properties across two agroecological zones in Southwest Germany: Generic independent validation surpasses region specific cross-validation. *Nutrient Cycling in Agroecosystems*, 102, 265–283. <https://doi.org/10.1007/s10705-015-9698-1>
- Mirzaeitalarposhti, R., Demyan, M. S., Rasche, F., Cadisch, G., & Müller, T. (2017). Mid-infrared spectroscopy to support regional-scale digital soil mapping on selected croplands of South-West Germany. *CATENA*, 149, 283–293. <https://doi.org/10.1016/j.catena.2016.10.001>
- Mouazen, A., Saeys, W., Xing, J., de Baerdemaeker, J., & Ramon, H. (2005). Near-infrared spectroscopy for agricultural materials: An instrument comparison. *Journal of Near Infrared Spectroscopy*, 13, 87–97. <https://doi.org/10.1255/jnirs.461>
- Nguyen, T., Janik, L. J., & Raupach, M. (1991). Diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy in soil studies. *Soil Research*, 29, 49–67. <https://doi.org/10.1071/SR9910049>
- Nocita, M., Stevens, A., Van Wesemael, B., Aitkenhead, M., Bachmann, M., Barth, 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>
- Ohio Supercomputer Center. (1987). *Ohio supercomputer center.* Columbus, OH: Ohio Supercomputer Center.
- O'Rourke, S. M., Minasny, B., Holden, N. M., & McBratney, A. B. (2016). Synergistic use of Vis NIR, MIR, and XRF spectroscopy for the determination of soil geochemistry. *Soil Science Society of America Journal*, 80, 888–899. <https://doi.org/10.2136/sssaj2015.10.0361>
- Oksanen, J. (2019). The Package 'vegan': Reference manual. The Comprehensive R Archive Network, R Development Core Team, Vienna. <https://cran.r-project.org/web/packages/vegan/vegan.pdf> (accessed 15 Jan. 2019).
- Orlov, D. S. (1986). *Humus acids of soil.*, Rotterdam: Balkema. <https://doi.org/10.1002/jpln.19871500116>
- Parikh, S. J., Goynes, K. W., Margenot, A. J., Mukome, F. N. D., & Calderón, F. J. (2014). Soil chemical insights provided through vibrational spectroscopy. *Advances in Agronomy*, 126, 1–148. <https://doi.org/10.1016/B978-0-12-800132-5.00001-8>
- Peng, X., Shi, T., Song, A., & Gao, W. (2014). Estimating soil organic carbon contents from visible/near-infrared spectroscopy using the combination of support vector machine regression and successive projection algorithm. *Remote Sensing*, 6, 2699–2717. <https://doi.org/10.3390/rs6042699>
- R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/> (accessed 21 June 2017).
- Rabenaarivo, M., Chapuis-Lardy, L., Brunet, D., Chotte, J. L., Rabeharisoa, L., & Barthès, B. G. (2013). Comparing near and mid-infrared reflectance spectroscopy for determining properties of Malagasy soils, using global or LOCAL calibration. *Journal of Near Infrared Spectroscopy*, 21, 495–509. <https://doi.org/10.1255/jnirs.1080>
- Riedel, F., Denk, M., Müller, I., Barth, N., & Gläßer, C. (2018). Prediction of soil parameters using the spectral range between 350 and 15,000 nm: A case study based on the Permanent Soil Monitoring Program in Saxony, Germany. *Geoderma*, 315, 188–198. <https://doi.org/10.1016/j.geoderma.2017.11.027>
- Robin, V., Petit, S., Beaufort, D., & Prêt, D. (2013). Mapping kaolinite and dickite in sandstone thin sections using infrared microspectroscopy. *Clays and Clay Minerals*, 61, 141–151. <https://doi.org/10.1346/CCMN.2013.0610211>
- Russell, J. D. (1987). Infrared spectroscopy of inorganic compounds. In: H. Willis (Ed.), *Laboratory methods in infrared spectroscopy.* New York: John Wiley & Sons.
- Schwartz, G., Eshel, G., & Ben-Dor, E. (2011). Reflectance spectroscopy as a tool for monitoring contaminated soils. In: S. Pascucci (Ed.), *Soil contamination.*, New York: InTech. p. 67–90. <https://doi.org/10.5772/23661>
- Sila, A. M., Shepherd, K. D., & Pokhariyal, G. P. (2016). Evaluating the utility of mid-infrared spectral subspaces for predicting soil properties. *Chemometrics and Intelligent Laboratory Systems*, 153, 92–105. <https://doi.org/10.1016/j.chemolab.2016.02.013>
- Smith, B. C. (2011). *Fundamentals of Fourier transform infrared spectroscopy.*, Boca Raton: CRC Press. <https://doi.org/10.1201/b10777>
- Soda, R. (1961). Infrared absorption spectra of quartz and some other silica modification. *Bulletin of the Chemical Society of Japan*, 34, 1491–1495. <https://doi.org/10.1246/bcsj.34.1491>
- Soil Survey Staff. (2014). Kellogg soil survey laboratory methods manual. Soil Survey Investigations Report No. 42, Version 5.0. USDA, Natural Resources Conservation Service, Washington, DC.
- Soriano-Disla, J. M., Janik, L. J., Viscarra Rossel, R. A., MacDonald, L. M., & McLaughlin, M. J. (2014). The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Applied Spectroscopy Reviews*, 49, 139–186. <https://doi.org/10.1080/05704928.2013.811081>
- Spaccini, R., & Piccolo, A. (2007). Molecular characterization of compost at increasing stages of maturity. I. Chemical fractionation and infrared spectroscopy. *Journal of Agricultural and Food Chemistry*, 55, 2293–2302. <https://doi.org/10.1021/jf0625398>
- Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., & Wetterlind, J. (2010). Visible and near-infrared spectroscopy in soil



- science. *Advances in Agronomy*, 107, 163–215. [https://doi.org/10.1016/S0065-2113\(10\)07005-7](https://doi.org/10.1016/S0065-2113(10)07005-7)
- Sterratt, D. C. (2019). The Package ‘geometry’: Reference manual. The Comprehensive R Archive Network. R Development Core Team, Vienna. <https://cran.r-project.org/web/packages/geometry/geometry.pdf> (accessed 1 April 2019).
- Stevens, A., & Ramirez-Lopez, L. (2015). An introduction to the prospectr package. GitHub, Inc., San Francisco, CA. <http://antoinestevens.github.io/prospectr/> (accessed 21 June 2017).
- 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, 849–862. <https://doi.org/10.1111/j.1365-2389.2011.01401.x>
- Takoutsing, B., Weber, J. C., Tchoundjeu, Z., & Shepherd, K. (2016). Soil chemical properties dynamics as affected by land use change in the humid forest zone of Cameroon. *Agroforestry Systems*, 90, 1089–1102. <https://doi.org/10.1007/s10457-015-9885-8>
- Tatzber, M., Stemmer, M., Spiegel, H., Katzlberger, C., Haberhauer, G., & Gerzabek, M. (2007). An alternative method to measure carbonate in soils by FT-IR spectroscopy. *Environmental Chemistry Letters*, 5, 9–12. <https://doi.org/10.1007/s10311-006-0079-5>
- Terhoeven-Urselmans, T., Vagen, T.-G., Spaargaren, O., & Shepherd, K. D. (2010). Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library. *Soil Science Society of America Journal*, 74, 1792–1799. <https://doi.org/10.2136/sssaj2009.0218>
- Towett, E. K., Shepherd, K. D., Sila, A., Aynekulu, E., & Cadisch, G. (2015). Mid-infrared and total x-ray fluorescence spectroscopy complementarity for assessment of soil properties. *Soil Science Society of America Journal*, 79, 1375–1385. <https://doi.org/10.2136/sssaj2014.11.0458>
- Varmuza, K., & Filzmoser, P. (2009). *Introduction to multivariate statistical analysis in chemometrics.*, Boca Raton: CRC Press. <https://doi.org/10.1201/9781420059496>
- Veum, K. S., Goyne, K. W., Kremer, R. J., Miles, R. J., & Suduth, K. A. (2014). Biological indicators of soil quality and soil organic matter characteristics in an agricultural management continuum. *Biogeochemistry*, 117, 81–99. <https://doi.org/10.1007/s10533-013-9868-7>
- Vereecken, H., Schnepf, A., Hopmans, J. W., Javaux, M., Or, D., Roose, T., ... Young, I. M. (2016). Modeling soil processes: Review, key challenges, and new perspectives. *Vadose Zone Journal*, 15, 1–57. <https://doi.org/10.2136/vzj2015.09.0131>
- 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>
- Wang, X., Sanderman, J., & Yoo, K. (2018). Climate-dependent topographic effects on pyrogenic soil carbon in southeastern Australia. *Geoderma*, 322, 121–130. <https://doi.org/10.1016/j.geoderma.2018.02.025>
- Wehrens, R. (2011). *Chemometrics with R: Multivariate data analysis in the natural sciences and life sciences.*, Berlin: Springer. <https://doi.org/10.1007/978-3-642-17841-2>
- Weil, R. R., Islam, K. R., Stine, M. A., Gruver, J. B., & Samson-Liebig, S. E. (2003). Estimating active carbon for soil quality assessment: A simplified method for laboratory and field use. *American Journal of Alternative Agriculture*, 18, 3–17. <https://doi.org/10.1079/AJAA2003003>
- Wilson, B. R., King, D., Grown, I., & Veeragathipillai, M. (2017). Climatically driven change in soil carbon across a basalt landscape is restricted to non-agricultural land use systems. *Soil Research*, 55, 376–388. <https://doi.org/10.1071/SR16205>
- Winowiecki, L., Vågen, T. G., & Huising, J. (2016). Effects of land cover on ecosystem services in Tanzania: A spatial assessment of soil organic carbon. *Geoderma*, 263, 274–283. <https://doi.org/10.1016/j.geoderma.2015.03.010>
- Workman, J. J. (2016). *Concise handbook of analytical spectroscopy: Theory, applications, and reference materials.*, Singapore: World Scientific. <https://doi.org/10.1142/8800-vol3>
- Yang, H., Kuang, B., & Mouazen, A. M. (2012). Quantitative analysis of soil nitrogen and carbon at a farm scale using visible and near infrared spectroscopy coupled with wavelength reduction. *European Journal of Soil Science*, 63, 410–420. <https://doi.org/10.1111/j.1365-2389.2012.01443.x>
- Zhang, L., Yang, X., Drury, C., Chantigny, M., Gregorich, E., Miller, J. ... Yang, J. (2018). Infrared spectroscopy prediction of organic carbon and total nitrogen in soil and particulate organic matter from diverse Canadian agricultural regions. *Canadian Journal of Soil Science*, 90, 1–23. <https://doi.org/10.1139/cjss-2017-0070>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Deiss L, Margenot AJ, Culman SW, Demyan MS. Optimizing acquisition parameters in diffuse reflectance infrared Fourier transform spectroscopy of soils. *Soil Sci Soc Am J*. 2020;1–19. <https://doi.org/10.1002/saj2.20028>