Received: 17 May 2019 | Accepted: 12 October 2019

DOI: 10.1002/saj2.20021

Soil Science Society of America Journal

NUTRIENT MANAGEMENT & SOIL & PLANT ANALYSIS

Grinding and spectra replication often improves mid-DRIFTS predictions of soil properties

Leonardo Deiss[1] | **Steve W. Culman**[2] | **M. Scott Demyan**[3]

[1]School of Environment and Natural Resources, Ohio State Univ., 2021 Coffey Road, Columbus, OH, 43210

[2]School of Environment and Natural Resources, Ohio State Univ., 1680, Madison Ave, Wooster, OH, 44691

[3]School of Environment and Natural Resources, Ohio State Univ., 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.

Abstract

There is an increased interest in using diffuse reflectance infrared Fourier transform spectroscopy in the mid-infrared region (mid-DRIFTS) for high-throughput prediction of soil properties, but basic methodological factors toward this end have yet to be thoroughly vetted. This study aimed to determine how the combined effects of soil grinding (sieved to <2.0 mm and finely ground to <0.5 mm) and sample replication (single or multiple soil subsamples, using one-to-four replicates) affect the spectral quality and predictive performance (accuracy) of automated plate-based mid-DRIFTS for soil analysis. We evaluated chemometric prediction performance of soil physical, chemical, and biological variables (clay, sand, pH, total organic C, and permanganate-oxidizable C [POXC]) in 397 soils from the U.S. Midwest. Sieved soils (<2.0 mm) increased the overall spectral variability compared to finely ground soils (<0.5 mm) and led to a distinct wavenumber importance allocation in support vector machine models. These spectral changes degraded prediction performance of <2.0 mm samples when compared to <0.5 mm samples. The number of spectral replicates had a smaller effect on spectral properties, but impacted prediction accuracies of soil properties. In general, prediction outcomes improved with four spectral replicates either within a single soil subsample or across different soil subsamples. Our data collectively suggest that soil particle-size reduction to <0.5 mm and collecting multiple spectra improve mid-DRIFTS predictions. Recommendations to optimize high-throughput mid-DRIFTS should consider the tradeoffs between prediction accuracy and the effort needed to prepare soil samples and acquire spectra.

1 | INTRODUCTION

Methodological decisions in diffuse reflectance infrared Fourier transform spectroscopy in the mid-infrared region (mid-DRIFTS) can introduce error, compromising its analytic and predictive applications in soil sciences (Gholizadeh, Luboš, Saberioon, & Vašát, 2013; Soriano-Disla, Janik, Viscarra Rossel, MacDonald, & McLaughlin, 2014). Error can originate from sample preparation (e.g., grinding and drying), instrument set-up and parameterization (e.g., sample replication, resolution, and number of co-added scans), and/or data processing (chemometrics). Yet in many cases, commonly employed experimental settings for mid-DRIFTS of soils have been used without systematic evaluation and optimization. These experimental parameters can affect the spectral quality and consequently the prediction performance

Abbreviations: mid-DRIFTS, diffuse reflectance infrared Fourier transform spectroscopy in the mid-infrared region; POXC, permanganate-oxidizable carbon; TOC, total organic carbon.

of mid-DRIFTS. Spectral quality (defined here as the ability to represent molecular bond's vibrational information from a sample when stimulated by infrared electromagnetic energy) is actually one of the most restrictive factors in infrared spectroscopy of soils (Coûteaux, Berg, & Rovira, 2003). Objective evaluations of these parameters are necessary to understand how they impact spectral properties and prediction performance, ultimately providing guidelines for best practices.

Even though soil grinding has a solid theoretical base on how it can affect the spectral quality in mid-DRIFTS, previously reported effects on prediction performance have been inconsistent (Table 1). Some studies have achieved better prediction performances for soil total organic carbon (TOC) with smaller (finely ground) particle size distributions (Barthès et al., 2016; Le Guillou et al., 2015), but the response can vary by soil texture (Baldock, Hawke, Sanderman, & MacDonald, 2013). Stumpe, Weihermüller, and Marschner (2011) showed that finely grinding soils for 2 min in a mill when compared to <2.0 mm samples improved mid-DRIFTS predictions for either soil TOC or pH; however, an additional time in the mill (total of 4 min) decreased the prediction performance for both soil variables (Stumpe et al., 2011). Evaluating soil texture, Janik, Soriano-Disla, Forrester, and McLaughlin (2016) found a positive effect of grinding when predicting soil clay and silt, but obtained a negative effect for sand. Le Guillou et al. (2015) obtained more accurate predictions with soil grinding for clay and sand, but found no differences for silt. In general, TOC and clay prediction accuracies improved with grinding while coarser particle sizes had a less consistent response.

Finely grinding soil samples (<0.25 mm) is a common recommendation due to the 1–2 mm diameter beam size in mid-IR spectrometers (Reeves & Smith, 2009), which is smaller than larger soil particles (e.g., coarse sand and aggregates). Positive effects of grinding include prevention of specular reflections from large particles in the soil sample (e.g., large quartz particles) (Le Guillou et al., 2015) and reduction of both light scattering and presence of spectral artifacts, thereby improving spectral quality (Kimber & Kazarian, 2017). As a result, spectra from finely ground soils generally have lower spectral variability among soil analytical replicates and lower multivariate scores dispersions across different soil samples (Baldock et al., 2013; Le Guillou et al., 2015; Stumpe et al., 2011). The intense peaks produced from specular reflections can completely mask specific peaks in the diffuse reflectance spectra of soil (Nguyen, Janik, & Raupach, 1991). On the other hand, excessive grinding could destroy chemical bonds of soil organic or mineral constituents (Stumpe et al., 2011), change soil particle size distribution (Stenberg, Jonsson, & Börjesson, 2000), and increase overall spectral reflectance (Stenberg, Viscarra-Rossel, Mouazen, & Wetterlind, 2010); all of which can result in less specific light reflections in the spectra. Finally, besides these contrasting positive and negative effects, Le Guillou et al. (2015) concluded that recommendations on how much grinding is needed for mid-DRIFTS should also consider the time and effort needed to prepare the soil samples as well as the purpose of the analysis. Reeves (2010) stated that while spectral distortions do occur as a function of non-optimal experimental conditions (e.g., specular reflections), multivariate models that are used to predict properties could at least partially help ameliorate the problem.

Core Ideas

- There is no consensus for soil grinding and sample replication in mid-DRIFTS.
- Soil grinding to <0.5 mm compared to <2.0 mm decreased spectral variability.
- Finely grinding soils (<0.5 mm) with four replicates improved prediction performance.
- Spectra replicated in the same or different subsamples had similar spectral properties.
- High-throughput mid-DRIFTS yields robust predictions of soil properties.

The effect of replicating spectra within a soil sample or analyzing multiple soil subsamples (i.e., analytic replicates) on the spectral quality and prediction performance in mid-DRIFTS is less understood relative to the effect of soil grinding. There is a lack of studies that have explicitly evaluated these effects in the mid-infrared region. Spectra replication in a single soil subsample has been facilitated by new models of bench-top DRIFTS equipped with autosamplers that enable spectral readings in multiple places within a single sample cup or well. The replication method *Within* a well is done by moving the beam (or the plate) in a random fashion, reducing the number of wells (different soil subsamples) needed to obtain a representative spectrum. Many studies have used analytical replicates as a way to counteract spectral variability, increase sample representativeness, and account for differences in particle size and packing density (e.g., Mirzaeitalarposhti, Scott, Rasche, Cadisch, & Müller, 2017; Riedel, Denk, Müller, Barth, & Gläßer, 2018; Terhoeven-Urselmans, Vagen, Spaargaren, & Shepherd, 2010; Zhang et al., 2018), but most studies have not explicitly shown how replicating spectra affect spectral characteristics and prediction performance. In one study that evaluated the number of replicates in mid-DRIFTS, Peng et al. (2014) showed that cross-validated prediction errors for TOC and clay decreased markedly (by about 20%) from one to three analytical replicates (different soil subsamples); and the improvement was minor thereafter up to 10 replicates. Given the soil intrinsic bio-physical-chemical heterogeneity, one might expect that

TABLE 1 Studies reporting the effect of soil grinding on diffuse reflectance infrared Fourier transform spectroscopy model prediction performance in the mid-infrared region (mid-DRIFTS)

Reference	Variable (range)[a]	Grinding (<mm)	RMSE[c]	R^{2c}	Samples (*n*)	
					Cal. or CV	Val.
Stumpe et al., 2011	pH (3.5–7.6)	2	0.28	0.86	180	–
		(+2 min)[b]	0.22	0.96	180	–
		(+4 min)[b]	0.24	0.92	180	–
	Organic C	2	0.09[d]	0.68	180	–
	(0.5–4.5%)	(+2 min)[b]	0.05[d]	0.94	180	–
		(+4 min)[b]	0.07[d]	0.91	180	–
Le Guillou et al., 2015	Organic C	2	0.96	0.66	227	–
	(0.03–20.1%)	1	0.88	0.72	227	–
		0.5	0.86	0.73	227	–
		0.25	0.82	0.75	227	–
		0.106	0.8	0.77	227	–
	Clay (0–860 g kg^{-1})	2	102	0.75	227	–
		1	98	0.77	227	–
		0.5	101	0.75	227	–
		0.25	95	0.78	227	–
		0.106	86	0.82	227	–
	Silt (0–620 g kg^{-1})	2	63	0.66	227	–
		1	61	0.65	227	–
		0.5	63	0.69	227	–
		0.25	66	0.68	227	–
		0.106	65	0.68	227	–
	Sand (22–1000 g kg^{-1})	2	123	0.78	227	–
		1	117	0.8	227	–
		0.5	110	0.82	227	–
		0.25	103	0.84	227	–
		0.106	92	0.88	227	–
Barthès et al., 2016	Organic C	2	–	0.84	97	–
	(0.2–12.1%)	0.2	–	0.95	97	–
Janik et al., 2016	Clay (10–890 g kg^{-1})	2	90	0.8	1313	664
	Clay (0–830 g kg^{-1})	0.1	70	0.88	792	406
	Silt (30–980 g kg^{-1})	2	120	0.78	1313	664
	Silt (10–510 g kg^{-1})	0.1	50	0.59	797	330
	Sand (0–670 g kg^{-1})	2	70	0.63	1313	664
	Sand (50–980 g kg^{-1})	0.1	90	0.84	793	408

[a]Stumpe et al. (2011) measured soil pH in a 1:2.5 soil to 0.01 m $CaCl_2$ solution ratio (m/v), and soil organic carbon was determined by the dry combustion method. Le Guillou et al. (2015) used samples from a database where soil organic carbon and soil texture were measured by various methods (Viscarra-Rossel & Webster, 2012). Barthès et al. (2016) measured soil organic carbon by the dry combustion method. Janik et al. (2016) determined soil texture using the pipette method.

[b] <2 mm samples plus the given grinding time in minutes in a mill.

[c]The coefficient of determination (R^2) and root mean squared error (RMSE) were extracted from independent validation datasets or cross-validation when validation sets were not available.

[d]Log-transformed data.

replicating spectra in the same soil subsample would produce less variable spectra than different subsamples.

We hypothesize that with coarsely sieved soils, more spectral variability can be expected due to the additive effects of particle size heterogeneity and intrinsic variability among different soil subsamples. With finely ground soils, more variability would be expected due to the intrinsic variability among different subsamples than to the soil particle size heterogeneity. However, the interactive effect of grinding and methods to replicate spectra has yet to be explored.

It remains unclear if soil grinding and spectra replication improve prediction performances in mid-DRIFTS.

The objective of this study was to determine suitable experimental parameters to optimize predictive applications of automated plate-based high-throughput mid-DRIFTS for soils. Specifically, we aim to determine how the combined effects of soil grinding (soils crushed to pass <2 mm or crushed and then finely ground to <0.5 mm) and spectral replication (one-to-four spectra replicates *Within* a subsample or *Across* different subsamples) affect the spectral variability and prediction accuracy of mid-DRIFTS for soil analysis. We evaluated the prediction performance of five soil variables (clay, sand, pH, TOC, and permanganate-oxidizable C) covering a diversity of soils from Ohio, USA.

2 | MATERIALS AND METHODS

2.1 | Soils and study area

We selected 397 legacy soil samples from the National Cooperative Soil Survey (NCSS, https://www.nrcs.usda.gov/wps/portal/nrcs/main/soils/survey/) distributed over four different physiographic regions of the U.S. Midwest: Glaciated Allegheny Plateau (AP), Unglaciated Allegheny Plateau (GP), Till Plains (TP), and Huron-Erie Lake Plains (LP). These physiographic regions (ODGS, 1998) corresponded to the following Major Land Resource Areas (MLRA) (USDA-NRCS, 2006; MLRA names are given in italics with the corresponding MLRA number in parentheses): AP, *Western* (124) and *Central* (126) *Allegheny Plateau*; GP, *Lake Erie Glaciated Plateau* (139); TP, *Indian and Ohio Till Plain*, *Central* (111A), *Northeastern* (111B), *Western* (111D), and *Eastern* (111E) *Parts*; and LP, *Erie and Huron Lake Plain* (99).

For each region, soil samples were systematically selected to cover a broad range of soil variables values (Table 2). 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. Legacy data was digitally available at https://ncsslabdatamart.sc.egov.usda.gov/. Samples were assigned to a physiographic region based on their county location, and samples from counties containing two or more physiographic regions were not included. Additionally, only samples were considered after 1966, as prior to this year wet combustion was used to determine TOC.

2.2 | Soil analysis

Soil texture, pH, and TOC were part of the legacy data while permanganate oxidizable carbon (POXC) was determined specifically for this study. These soil properties are soil health indicators that enable evaluating how well soil is performing its ecological functions (https://www.nrcs.usda.gov/wps/portal/nrcs/main/soils/health/). Laboratory soil analysis of the legacy data 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 soil/water mixture (v/v) (method 4C1a2a), and total C (TC) was measured by dry combustion (method 6A2a). Total carbon was measured in a Lindberg tube furnace heated to 900 °C, with evolved CO_2 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 measurements were 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^{-1} $KMnO_4$ was added to 50-ml tubes containing 2.5 g air-dried soil. The tubes were shaken for 2 min at 240 oscillations min^{-1} 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.

2.3 | Sample preparation for DRIFTS

A diagram of the procedures used to prepare the samples, setup the instrument, and data processing (chemometrics) is presented in Figure 1. The samples from the soil survey were originally stored as air-dried, crushed, and sieved to <2.0 mm. Before acquiring spectra for each grinding size, soils were dried for > 48 h at 40 °C and at 12–14% relative humidity. After acquiring the spectra for the <2.0 mm samples (spectra acquisition process is described below), the same soil subsamples were ground to a <0.5 mm size by placing into 20-ml cylindrical glass vials with three stainless steel bars (0.63 cm diameter and 3.81 cm length), and loaded onto a roller grinder. The vials were rolled for 16 h at 60 rpm and the stainless steel bars reduced sample sizes by crushing soil between falling bars as the vials rotated. We determined how effective this grinding method was at reducing soils to <0.5 mm in a pretest. Ten soils were selected to represent a wide range of soil textures (clay: 140–437 g kg^{-1}; silt: 39–531 g kg^{-1}; sand: 32–782 g kg^{-1}). These samples were ground as described above, weighed, passed through a 0.5 mm sieve and the fraction

TABLE 2 Summary statistics of measured soil variables covering four physiographic regions of the U.S. Midwest

	Clay	Sand	pH	TOC[a]	POXC[a]	Depth, upper	Depth, lower
	g kg^{-1}	g kg^{-1}		%	mg kg^{-1}	cm	
Unglaciated Allegheny Plateau (A)							
Min.	30	1	2.3	0.10	14	0	1.3
Median	238	131	4.9	0.34	134	26.0	38.1
Mean	268	228	5.2	0.94	247	41.2	55.6
Max.	851	930	7.9	7.2	1330	312.4	355.6
n	100	100	100	100	97	97	97
Glaciated Allegheny Plateau (GA)							
Min.	57	5	4.2	0.20	3	0	3
Median	204	253	5.4	0.40	136	29.0	38.6
Mean	219	281	5.6	0.76	229	47.3	62.1
Max.	648	853	7.6	6.0	1233	203.2	228.6
n	100	100	100	100	100	100	100
Huron-Erie Lake Plains (LP)							
Min.	11	11	4.2	0.13	45	0	5.1
Median	368	191	6.9	0.60	196	35.6	50.8
Mean	335	306	6.7	1.12	371	41.1	57.6
Max.	710	980	7.9	9.1	1412	134.6	147.3
n	99	99	99	99	98	98	98
Till Plains (T)							
Min.	74	8	3.2	0.20	37	0	3
Median	318	165	6.1	0.6	181	30.5	45.7
Mean	308	230	6.1	0.9	311	40.3	55.4
Max.	645	820	8	4.3	1377	165.1	180.3
n	98	98	98	98	97	97	97

[a]TOC, total organic carbon; POXC, permanganate oxidizable carbon.

that passed through the sieve was weighed again. Averaged across all soils, grinding effectively reduced 95.8% of the soil to <0.5 mm. Most remaining particles that did not pass thought the <0.5 mm sieve were small rocks. These remaining particles did not present risk of increasing particle heterogeneity once they were taken out by a trained technician if detected in the sample cup's top surface. The two grinding sizes used in our study were primarily defined based on the results found by Le Guillou et al. (2015) that did find significant increases in model predictions errors with decreased particle size going from <2.0 mm to <0.25 mm, but there was no difference going from <1.0 mm to <0.25 mm (including <0.5 mm).

2.4 | DRIFTS instrument set-up

Absorbance spectra (log R^{-1}, where R is reflectance) 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). The autosampler enables plate-based measurements of multiple samples across wells, but it also enables reading multiple locations within each single well, locations which can be specifically configured in AutoPro software (Pike Technologies Inc., Madison, WI). We used anodized aluminum plates that fit 24 polystyrene sample cups. We used oven dried (>16 h at 100 °C) potassium bromide as a background (KBr), and background measurements were done at the beginning of each plate. All measurements were conducted from 4000 to 400 cm^{-1}, 4 cm^{-1} wavenumber resolution and with 24 co-added scans. We further reduced the spectral data to 4000 to 700 cm^{-1} to conduct spectral analysis and predictions.

The polystyrene sample cups (also referred to plate wells) had an internal volume of 5.5 ml and a top circular opening area of 10 mm diameter. The sample cups were loaded with soil subsamples by over-filling the cups with soil, then tapping the cup side gently three times to settle the soil into the cup, and finally excess soil on the top was evenly scrapped off using the narrow edge of a stainless steel spatula. The soil was not packed or compressed into the well other than by tapping and scrapping.

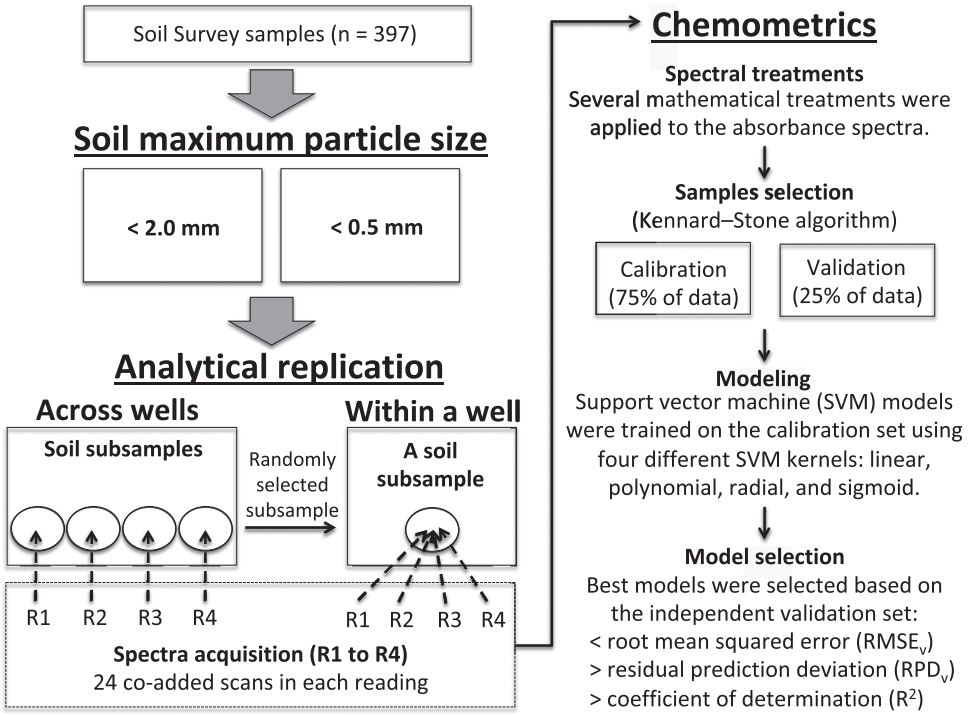

FIGURE 1 Diagram of the procedures used to pretreat soil samples with grinding, set-up for different methods of spectra replication, and process the spectral data (chemometrics) in diffuse reflectance infrared Fourier transform spectroscopy in the mid-infrared region (mid-DRIFTS). Instrument settings: wavenumber window 4000–700 cm^{-1}, resolution 4 cm^{-1}, and 24 co-added scans

To evaluate the two methods of sample replication, namely *Across* or *Within* soil subsamples, the following approaches were used (Figure 1). (I) In the replication *Across* wells, one spectral reading per well was done in four different wells (four sample cups loaded with different subsamples of the same soil). The spectral readings were done in the central position of each well. (II) In the replication *Within* a well, four spectral readings were done in a randomly selected subsample (one sample cup) from the four wells described in the first step. The four readings *Within* a well were done using the random oversampling motion function of the X,Y Autosampler (within a 3 mm diameter of the sample cup's centroid to avoid edge effects). Each spectrum consisted of 24 co-added scans. These measurements were done for all 397 soils. Afterward, to evaluate the effect of number of analytical replicates; one-to-four replicates of both replication methods (*Across* and *Within*), were randomly selected and averaged.

2.5 | Spectral quality assessment

Absorbance spectra were characterized by principal components analysis (PCA) using mean-centered spectra. We used the iterative NIPALS algorithm (Martens & Naes, 1989) to derive the PCA (R package 'chemometrics'; Varmuza & Filzmoser, 2009). We plotted the first two principal components scores and loadings to evaluate the effects of treatments on the absorbance data variability (PCA scores dispersion) and wavenumber-specific loadings distribution.

To compare the PCA scores dispersion, we first evaluated the spectra for homogeneity using Levene's test for homogeneity of variances, and then used a permutational multivariate analysis to test for the mentioned sources of variance (R package 'Vegan'; Oksanen, 2018). The first two principal components scores from PCA-NIPALS were evaluated for the spectral dissimilarity (homogeneity of multivariate scores dispersion) using the Mahalanobis distances. Mahalanobis distances are Euclidean distances of a spectral matrix where wavenumbers are mean-centered, have unit variance, and are uncorrelated (Mardia, Kent, & Bibby, 1979; Oksanen, 2018). The permutational multivariate analysis of variance was used to test for the partitioning of the Euclidean distance matrices among the sources of variation (Oksanen, 2018).

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

To determine the wavenumber importance in the support vector machine (SVM) models, a recursive feature elimination algorithm was used (R package 'caret'; Kuhn, 2018). This approach implements backward selection of predictors (wavenumbers) based on predictor importance ranking from the first to the least important wavenumbers. This analysis

considered all wavenumbers across the targeted mid-infrared spectrum (4000 to 700 cm^{-1}). The recursive feature elimination was processed using a 10-fold cross-validation with a 75% training set to 25% leave-group out cross-validation. The same spectral pretreatment method selected to develop the prediction model of each variable (described below) was used to determine the wavenumber importance.

2.6 | Prediction model calibration and independent validation

Several spectral treatments were applied to test which mathematical treatment had greater ability to enhance spectral features from 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, 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 portion of samples (75% of the dataset) selected using the Kennard–Stone sampling algorithm (Kennard & Stone, 1969) to explain ≥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 done for each one of the four physiographic locations. Prior to modeling, outliers were verified in the absorbance spectra considering orthogonal distance and score distance. Orthogonal distance was between the true position of each data point and its projection in space of the first few principal components to explain ≥80% of the total variance. Score distance was the projection of a sample to the center of all sample projections (Wehrens, 2011). The final dataset was constrained to a sample set with orthogonal distance <25 and score distance <6.

We first trained SVM regression models with different algorithms on the calibration data set; and subsequently tested them 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 in all spectral treatments, and a common configuration tested in each kernel was that both spectra (all spectra) and predictor were scaled or not scaled to zero mean and unit variance prior to calibration (Meyer et al., 2015). Best models were selected for each variable and treatment (combinations of grinding size, replication method, and number of analytical replicates) based on sequential criteria looking first at the lowest root mean squared error ($RMSE_v$), then greatest residual prediction deviation (RPD_v), and then greatest coefficient of determination ($R^2{}_V$) of the independent validation datasets. The RMSE compute the difference between observed values and the predicted values. The RPD is defined as the standard deviation of observed values divided by the RMSE. The RDP takes both the prediction error and the variation of observed values into account, providing a metric of model validity that is more objective for comparisons across variables and studies. The R^2 provides a measure of how well observed outcomes are reproduced by the model, based on the proportion of total variation explained by the model. The R^2 also allow comparison across variables and studies, but is highly dependent on variable's range of values.

To compare the prediction outcomes obtained with different numbers of analytical replicates, the measured and predicted values of the independent validation data-sets were bootstrapped using the RMSE ($RMSE_v$) function, ordinary resampling method and 99 replicates (Davison & Hinkley, 1997; R package 'boot', Canty, 2017). Multiple means comparison was done including both replication methods (*Across* and *Within*, one-to-four analytical replicates) and grinding sizes (<2.0 mm and <0.5 mm), by the Tukey test ($p <. 05$) with estimated marginal means (least-squares means) (R package 'emmeans', Lenth, 2019).

2.7 | Data processing and statistical analyses

Data was processed and analyzed in R version 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria) using the Ohio Supercomputer Center (OSC, 1987) computing resources. The specific statistical methods are described in the Methods sections *Spectral quality assessment* and *Prediction model calibration and independent validation.*

3 | RESULTS

The four physiographic regions had distinctive spectral characteristics, as indicated by the PCA scores dispersion of the first two principal components (PC1, PC2) in principal component analysis (NIPALS algorithm) (Supplemental Figure S1). The permutational multivariate analysis of variance indicated that this distinction occurred for both grinding sizes and replication methods (<2.0 mm | *Across*: $p <$ .001; <2.0 mm | *Within*: $p <$ .001; <0.5 mm | *Across*: $p <$ .001; and <0.5 mm | *Within*: $p <$.001). In general, when comparing the different physiographic regions, Huron-Erie Lake Plains region had the widest PCA scores dispersion while Glaciated Allegheny Plateau had a less variable one. Both Unglaciated Allegheny Plateau and Till Plains had intermediate PCA scores dispersion when compared to those previous regions.

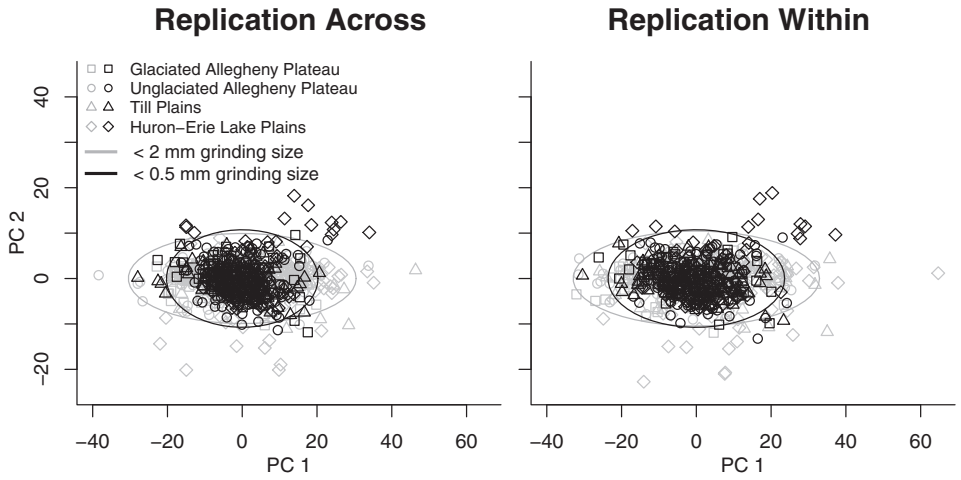

FIGURE 2 Soil grinding and sample replication effects on multivariate scores dispersion of the first two principal components (PC1, PC2) in the principal component analysis (PCA, NIPALS algorithm) of diffuse reflectance infrared Fourier transform spectra in the mid-infrared region (mid-DRIFTS) as affected by soil grinding (sieved to <2 mm or ground to <0.5 mm) and spectra replication. Replication of the mid-infrared spectra was done *Across* different subsamples or *Within* a randomly selected subsample. Spectra were averaged from four soil analytical replicates in both methods of soil replication (*Across* and *Within*). Soils were from four physiographic regions of the USA Midwest (total $n = 397$). The PCA was separated for each grinding size to evaluate the effect of replication methods, but results were presented in different panels for each replication method. The first component (PC1) explained 81.5 and 68.9%, and the second component (PC2) 8.32 and 16.6% of the variance in the spectra for the <2 mm and <0.5 mm grinding sizes, respectively. Spectra were scaled by mean centering prior to PCA. Instrument settings: wavenumber window 4000–700 cm^{-1}, resolution 4 cm^{-1}, and 24 co-added scans

Following a similar trend, Huron-Erie Lake Plains had the widest range of TOC and sand contents, while Glaciated Allegheny Plateau had more constrained ranges for most soil variables (Table 2).

The PCA scores dispersion (Figure 2; Supplemental Figure S2) and the loading vectors distribution (Figure 3) were different for the two grinding sizes (2.0 mm and 0.5 mm), but these distinctions were not as clear between the replication methods (*Across* and *Within*) (Figure 2) nor for the number of analytical replicates (Supplemental Figure S2). The permutational multivariate analysis of variance indicated that the soil grinding affected PCA scores dispersion (<2.0 mm vs. <0.5 mm | *Across*: $p <$ .001 and <2.0 mm vs. <0.5 mm | *Within*: $p <$ .001), but there was no distinction of PCA scores dispersion between replication methods into each grinding size (*Across* vs. *Within* | <2.0 mm size: $p = 0.79$ and *Across* vs. *Within* | <0.5 mm: $p = 0.99$) (Figure 2). In each grinding size and replication method, there was no difference in the PCA scores dispersion among the different number of analytical replicates (for all $p > 0.99$) (Supplemental Figure S2).

Following a similar trend of PCA scores dispersion, PCA loading vectors distribution across the wavenumbers was more affected by soil grinding than by either replication methods or number of analytical replicates (Figure 3). In general, the <2.0 mm samples had a higher diversity of more intense loadings on the PC1 at the 2000 to 700 cm^{-1}, while the <0.5 mm samples had a similar response in the same region but on loadings of the PC2. Another expressive distinction between the two grinding sizes was the intensity of the loadings at specific wavenumbers. For example, the loadings between 3723 and 3686 cm^{-1} (wavenumbers attributed to the Si-O functional group of 2:1 layer alumino-silicates, Nguyen et al., 1991) were more intense for the <0.5 mm samples when compared to <2.0 mm samples. These distinctions were not evident when comparing the methods of replication within each grinding size nor the number of analytical replicates into each replication method.

Soil grinding, replication method, and number of analytical replicates affected the absorbance spectra and its variability (Figure 4; Supplemental Figure S3). Finely ground soils (<0.5 mm) had lower averaged absorbance values (averaged over all wavenumbers) than sieved soils (<2.0 mm); however, averaged absorbance values were less variable between replication methods or among number of analytical replicates (Supplemental Figure S3). The spectral variability (wavenumber-specific standard deviation) was affected by the tested treatments in terms of the averaged variability (across all wavenumbers) and intensity of variability in specific regions of the spectra (Figure 4). The averaged spectral variability was generally lower for the <0.5 mm samples than the <2.0 mm samples (within each replication method). For either grinding size, replication *Across* wells (different subsamples) generated more variability when compared to replication *Within* a well (same soil subsample). More specifically, smaller particle size distributions (<0.5 mm) combined

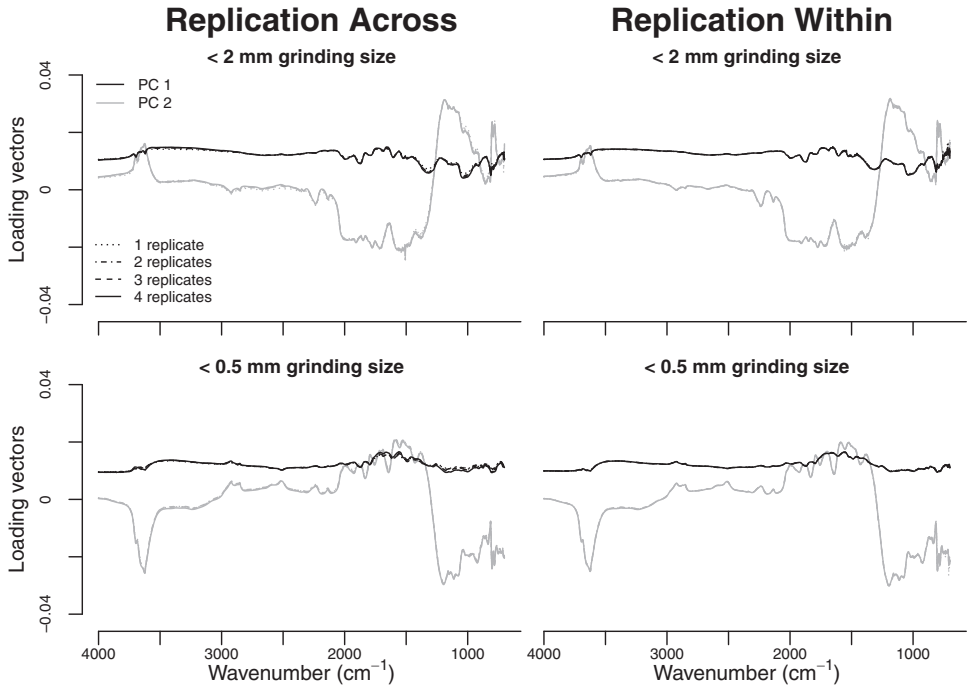

FIGURE 3 Loading vectors of the first two principal components (PC1, PC2) of principal component analysis (PCA, NIPALS algorithm) of diffuse reflectance infrared Fourier transform spectra in the mid-infrared region (mid-DRIFTS) as affected by soil grinding (sieved to <2 mm or ground to <0.5 mm), and one-to-four analytical replicates *Within* (same) or *Across* (different) subsamples. Spectra were averaged from one-to-four soil analytical replicates in both methods of soil replication (*Across* and *Within*). Spectra were scaled by mean centering prior to PCA. Instrument settings: wavenumber window 4000–700 cm^{-1}, resolution 4 cm^{-1}, and 24 co-added scans

with replication *Within* the same subsample had lower spectral variability than the coarser particle sizes (<2.0 mm) combined with replication *Across* different subsamples.

The regions at which more variability occurred were markedly different between the grinding sizes (Figure 4). For the <2.0 mm samples, there were several regions of high standard deviations at approximately 1750 to 1250 cm^{-1}, in which wavenumbers are attributed to carboxyl vibrations of carboxyl groups, aldehydes, ketones, and esters (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 or fulvic acid. While for the <0.5 mm samples, there was a pronounced increase in the variability at approximately 1250 to 800 cm^{-1}, where wavenumbers are attributed to organic components C-O bonds in both polyalcoholic and ether functional groups (Spaccini & Piccolo, 2007), C-O of polysaccharides or similar substances (Senesi, D'Orazio, & Ricca, 2003), and mineral components Si-O of silicate (Senesi et al., 2003) and carbonates calcite and dolomite (Mirzaeitalarposhti, Demyan, Rasche, Cadisch, & Müller, 2016).

Increasing the number of analytical replicates (from one-to-four replicates) increased the averaged spectral variability in all grinding sizes and replication methods (Figure 4). The changes in variability as affected by the number of analytical replicates occurred to a greater extent in the replication method *Across*. Additionally, the effect among different numbers of analytical replicates was greater for the <2.0 mm samples than <0.5 mm samples, and it was also greater for the replication method *Across* when compared to the replication *Within*.

Wavenumber importance in the SVM models was specific for each soil variable, and it was affected by both soil grinding and replication methods (Supplemental Figure S4). This analysis was specifically done using the spectra averaged from four analytical replicates. Changes were more apparent in the comparison between the two grinding sizes, but there were changes between replication methods mostly for sand and pH in the <0.5 mm samples.

The described spectral changes impacted the prediction outcomes from SVM models (Table 3; Figure 5). Changes in prediction outcomes followed a similar trend to those described for the spectral changes, as changes were more evident for the different grinding sizes than either replication methods or number of analytical replicates. For most evaluated soil properties (clay, sand, TOC, and POXC), except pH, soil grinding to <0.5 mm resulted in greater prediction accuracy than sieving solely to <2.0 mm based on $RMSE_V$ (Tables 3, S1). However, some results obtained with <2.0 mm were comparable to those obtained with <0.5 mm,

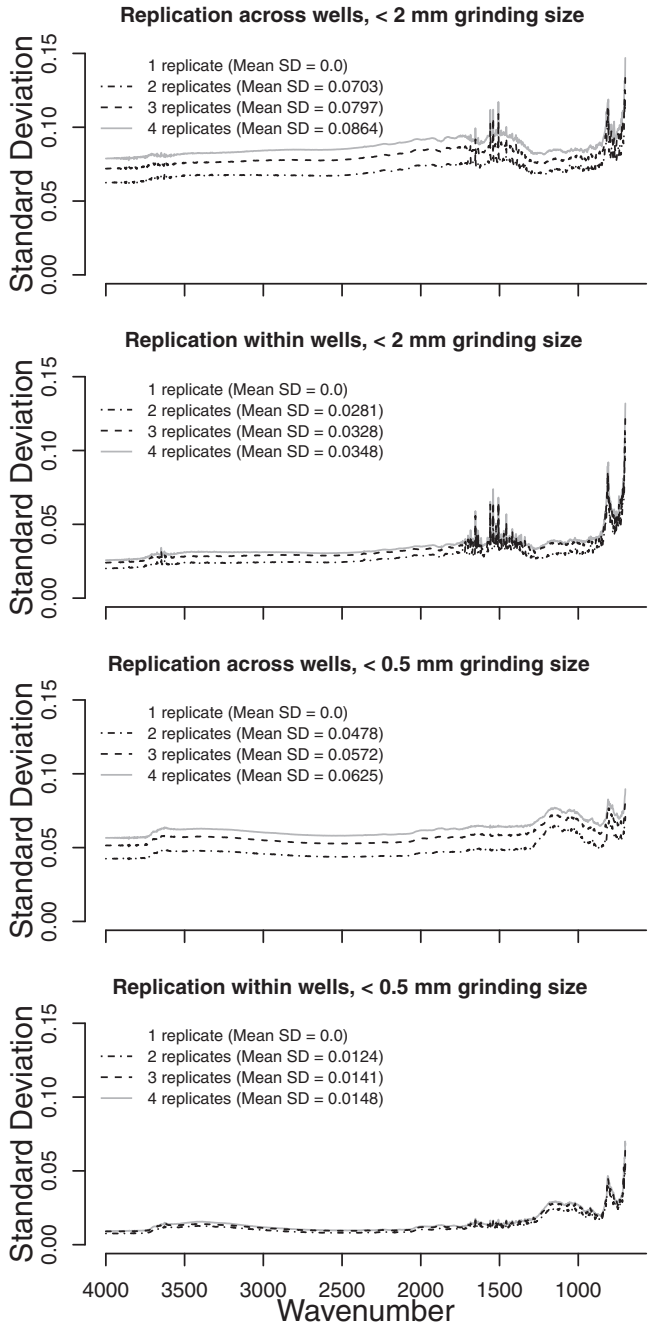

FIGURE 4 Diffuse reflectance infrared Fourier transform absorbance spectra variability in the mid-infrared region (mid-DRIFTS) as affected by soil grinding (sieved to <2 mm or ground to <0.5 mm), and one-to-four analytical replicates *Within* (same) or *Across* (different) subsamples. Soils were from four physiographic regions of the USA Midwest (total $n = 397$). The different replicate numbers mean that wavenumber-specific standard deviation was determined from four-to-two randomly selected analytical replicates for each sample and then standard deviations were averaged across all samples. Instruments settings: resolution 4 cm^{-1} and 24 co-added scans using a 24-well plate

e.g., pH (*Across* and *Within*) and POXC (<2.0 mm *Across* vs. <0.5 mm *Within*).

The number of analytical replicates affected the prediction outcomes, and the responses were specific for the grinding sizes and replication methods (Table 3; Figure 5; Supplemental Table S2). Most soil properties were predicted with greater accuracy when multiple replicates were used (three or four replicates) instead of fewer replicates (one or two replicates) in both sizes (<2.0 mm or <0.5 mm) and methods of replication (*Across* and *Within*).

4 | DISCUSSION

Soil particle size homogenization by grinding (<0.5 mm) affected the mid-DRIFTS predictions of soil properties and many factors could have contributed to these outcomes. Differences in particle size are sensed during mid-DRIFTS measurements, for example by soil-to-detector variability or eventual specular reflections (Le Guillou et al., 2015), and minimization of the resulting variability from these factors (Figures 2, 3, and 4) contributed to the best performance of SVM modeling of soil variables in the <0.5 mm spectra (Table 3; Figure 5). The size reduction of both larger primary particles and aggregates and homogenization in particle size distributions by finely grinding soils (Stenberg et al., 2010) help to increase representativeness of soil samples, once soil particles larger than the beam size could be overestimated in the spectra of soil samples solely sieved to <2 mm. The change in original particle size distributions (Stenberg et al., 2000) does not seem to have degraded prediction performance in finely ground soils. Destruction of soil aggregates by soil grinding to < 0.5 mm could have exposed previously entrapped organic matter and also individual silt and clay particles. Moreover, the presence of soil aggregates that are related to TOC concentrations (Blanco-Canqui & Lal, 2004; Mutuo, Shepherd, Albrecht, & Cadisch, 2006; Verchot, Dutaur, Shepherd, & Albrecht, 2011) could have contributed to heterogeneity of the exposed soil surface. Therefore, predictions could have been negatively affected due to this variability in <2.0 mm samples. This effect promoted by aggregates would be more expected in clay-rich soils, where aggregation of primary particles is stronger than coarser soils. Finally, surrogate prediction of other soil variables may also benefit from the particle size homogeneity when prediction accuracy is maintained or improved. The development of surrogate calibrations, in which some properties are totally or partially predicted because of their correlation or covariation with certain other soil properties, is a well-known topic in infrared spectroscopy (Chang, Laird, Mausbach, & Hurburgh, 2001; Reeves, 2010; Stenberg et al., 2010).

Besides the effect of soil particle size heterogeneity, the expected increase in the overall spectra reflectance (i.e.,

TABLE 3 Model prediction performance of diffuse reflectance infrared Fourier transform spectroscopy in the mid infrared region using sieved (<2.0 mm) and ground (<0.5 mm) soil, and spectra replicated from the same (*Within*) or different (*Across*) soil subsamples. For both replication methods, spectra were averaged from four analytical replicates. Soils were from four physiographic regions of the U.S. Midwest and each region comprised one quarter of the total dataset.[a]

Soil property	Grinding size (<mm)	Spectra replication	Calibration set (75% of dataset)			Validation set (25% of dataset)		
			RMSE	RPD	R^2	$RMSE_V$	RPD_V	$R^2{}_V$
Sand (g kg^{-1})	2	Within	83	2.60	0.89	81	1.78	0.79
(n = 397)		Across	80	2.74	0.89	81	1.90	0.80
	0.5	Within	54	4.18	0.95	66	2.36	0.87
		Across	77	2.89	0.90	58	1.96	0.77
Clay (g kg^{-1})	2	Within	50	2.87	0.91	33	3.35	0.91
(n = 397)		Across	49	3.01	0.91	42	2.47	0.85
	0.5	Within	35	4.32	0.95	31	3.49	0.91
		Across	32	4.75	0.96	38	2.54	0.84
pH	2	Within	0.43	2.16	0.85	0.40	2.47	0.85
(n = 397)		Across	0.40	2.43	0.87	0.41	2.27	0.83
	0.5	Within	0.31	3.14	0.92	0.43	2.00	0.81
		Across	0.50	1.84	0.80	0.43	2.08	0.81
TOC (%)	2	Within	0.57	1.58	0.78	0.22	3.29	0.90
(n = 397)		Across	0.56	1.56	0.78	0.21	3.39	0.93
	0.5	Within	0.53	1.79	0.81	0.15	3.78	0.94
		Across	0.52	1.83	0.82	0.13	4.75	0.95
POXC (mg kg^{-1})	2	Within	90	3.14	0.92	78	2.51	0.83
(n = 392)		Across	91	2.95	0.92	70	3.31	0.92
	0.5	Within	93	3.00	0.91	72	2.87	0.91
		Across	91	3.04	0.92	74	2.89	0.91

[a]RMSE, root mean squared error; RPD, residual prediction deviation; R^2, coefficient of determination; TOC, total organic carbon; POXC, permanganate oxidizable carbon.

reduction in absorbance) (Stenberg et al., 2010) resulting from grinding soils (<0.5 mm) was verified in our study (Figure 4; Supplemental Figure S4). However, this effect does not seem to have degraded prediction outcomes from mid-DRIFTS (Table 3; Figure 5). This increase in reflectiveness could be happening as an effect of the increase in surface area of highly reflective materials such as metals/metalloids and quartz. A preeminent increase in peak sizes attributed to quartz (~2080 to 1754 cm^{-1}, Nguyen et al., 1991) can be verified in the finely ground soils when compared to sieved soils (Supplemental Figure S4, fourth quartile of sand comparing <0.5 mm and <2.0 mm sample sizes). A similar response of increased quartz peak sizes in more ground soil samples (<2.0 mm versus <0.1 mm) was verified by Le Guillou et al. (2015) in mid-DRIFTS. Working in the near-infrared region and evaluating the effects of soil texture on DRIFTS predictions, Stenberg, Jonsson, and Börjesson (2002) verified that presence of quartz in sandy soils increased light scattering and that was enough to mask absorptions of organic compounds, and consequently reduce prediction performance of soil organic matter. However, masking of TOC spectral features by quartz may depend on TOC soil content (Waruru, Shepherd, Ndegwa, & Sila, 2016). Similarly, greater reflectiveness/specular reflection of other metals or metalloids as affected by soil grinding could be speculated based on the PCA loadings at the wavenumbers attributed to O-H and Si-O functional groups (3723 to 3565 cm^{-1}) (Figure 3) and wavenumber-specific standard deviations at the wavenumbers attributed to Si-O functional groups (1080 to 950 cm^{-1}). These functional groups respectively correspond to hydroxyl stretching (kaolinite and others) (3723 to 3686 cm^{-1}), 2:1 layer alumino-silicate (3686 to 3565 cm^{-1}), and Si-O of silicates (1080 to 950 cm^{-1}) (Ahlrichs, 1968; Nguyen et al., 1991; Russell, 1987; Wada & Greenland, 1970). The PCA loadings (Figure 3) and wavenumber-specific variability (Figure 4) of these functional groups were more intense in the <0.5 mm samples when compared to <2.0 mm samples.

The studies reporting the effect of soil grinding over the prediction outcomes are inconsistent (Table 1), and this study adds to the broader debate. Moreover, there are no studies to date that have evaluated the interactive effects of grinding and analytical replicates in the mid-infrared region. Making a direct comparison between our results and those obtained in previous studies in mid-DRIFTS

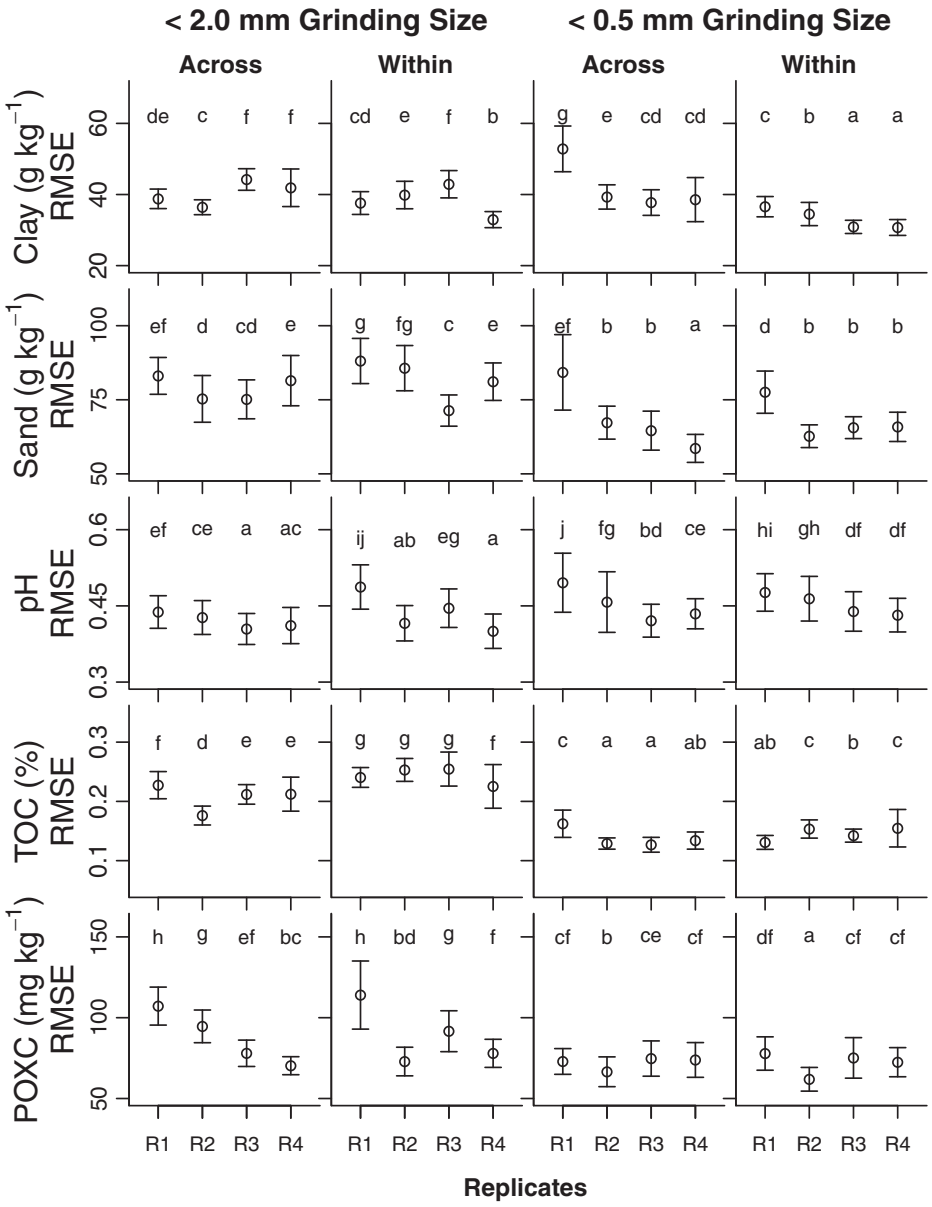

FIGURE 5 Bootstrapped root mean squared error ($RMSE_V$) of independent validation sets as affected by soil grinding (sieved to <2 mm or ground to <0.5 mm) and one-to-four analytical replicates *Within* (same) or *Across* (different) subsamples in diffuse reflectance infrared Fourier transform spectroscopy in the mid-infrared region (mid-DRIFTS). The different replicates number means that (absorbance) spectra were averaged from one to four randomly selected analytical replicates (R1 to R4). The means multiple comparison was across grinding sizes and replication methods. Different letters indicate significant differences by the Tukey test ($p < .05$). Bards denote standard deviation of means. TOC: Total organic carbon. POXC: Permanganate oxidizable carbon.

(Table 1) is complex because of differences in experimental procedures and soil characteristics. These differences include different grinding sizes and methods to grind soils; sample pretreatment (e.g., drying temperatures); types of soils, range of soil properties, and number of soil samples; and modeling approaches comprising spectral treatments and multivariate methods. As described earlier, the two grinding sizes used in our study (<2.0 and <0.5 mm) were primarily defined based on the results found by Le Guillou et al. (2015) that although their prediction accuracy improved with more grinding, they found no differences between <1.0 mm to <0.25 mm sizes (including <0.5 mm size), and those were more accurate than <2.0 mm. All the studies (including ours) compared <2.0 mm samples to finer grinding sizes, and most studies evaluated finer particle size distributions when compared to our study (<0.2 mm or <0.1 mm versus 0.5 mm in our study). Our study included more soil variables, and a larger (number of samples) and more diverse (in terms of soil properties range) set of samples than most studies, except for Janik et al. (2016) that used a larger number of soil samples

to test for soil texture variables. The modeling approaches were also different. We developed predictions using SVM regression models, while all other studies used partial least squares regression models. In our study, prediction performances of SVM models outperformed partial least squares models for almost all developed predictions (data not shown).

Recommendations on how much grinding is needed for mid-DRIFTS should also consider the time and effort required to prepare the soil samples as well as the objective of the analysis (Le Guillou et al., 2015). For predicting soil properties where increased accuracy is required, grinding generally reduces error. Grinding soil to <0.5 mm reduced $RMSE_V$ (*Within*/*Across*) for sand (23–40%), clay (6.5–10%), SOC (47–61%), and POXC (8.3%) [*Within*], but not for pH and POXC [*Across*] (Supplemental Table S1). However, other strategies can reduce errors even further than the ones obtained in this study. In another study, Deiss et al. (2020a) found using the same dataset that broadening spectra resolution up to 32 cm^{-1} in <2.0 mm samples significantly reduced the size of errors to a more accurate or similar level of accuracy as obtained in this study with <0.5 mm samples and 4 cm^{-1} resolution. Moreover, Reeves (2010) stated that multivariate models could at least partially help ameliorate the problem of poor spectral quality. Strategies such as tuning SVM regression models can help to improve prediction accuracy of soil properties in mid-DRIFTS (Deiss et al., 2020b).

In the present study, our experimental design allowed us to systematically compare the effect of both particle size and replication across a suite of soil variables. This contributed to some of the differences found in our results when compared to those from other studies (Table 1). For example, Janik et al. (2016) used a different set of samples for each grinding size, whereas using the same set of soils for each soil property would allow a more direct comparison between the prediction performances. Le Guillou et al. (2015) used the same spectral treatment for all grinding sizes for development of their prediction models. As we have shown in our study, grinding to <0.5 mm fundamentally changes the nature of the spectral data by decreasing both spectral wavenumber variability and PCA scores dispersion, changing loading vectors distribution across the wavenumbers, and leading to a distinct wavenumber importance allocation in support vector machine models. This spectral response to soil grinding was expected (Figures 2, 3, and 4) and consistent with other studies (e.g., Baldock et al., 2013; Le Guillou et al., 2015; Stumpe et al., 2011). The final goal of any modeling is to optimize the prediction power for model use on unknown samples and our results indicated that different spectral treatments were selected for the different grinding sizes to generate the most accurate prediction outcomes. Also, three of the four studies (Table 1) used cross-validation instead of working with independent validation sets. The cross-validation likely overestimates calibration stability (Stenberg et al., 2010; Stumpe et al., 2011), and creates uncertainty about the reproducibility of the developed models for future use in unknown samples.

Evaluating the interactive effects of grinding and analytical replicates has been done using other regions of the electromagnetic spectrum. Using near-infrared spectroscopy, Barthès, Brunet, Ferrer, Chotte, and Feller (2006) evaluated how the interaction between sample preparation methods (soil grinding and drying) and number of replicates affect predictions of total carbon and nitrogen in 123 soil samples originating from tropical regions of Africa and America. They obtained better prediction accuracies with more finely ground soils, but those predictive outcomes were comparable to results obtained with coarsely ground soils by increasing the number of analytical replicates (using different soil subsamples) (Barthès et al., 2006). However, how these interactive effects of soil grinding and sample replication would respond in the more sensitive mid-infrared region was previously unsettled. We verified that increasing the number of replicates in <2.0 mm samples improved prediction of clay, sand, pH, and POXC, but the improvement did not equal the prediction accuracy obtained with < 0.5 mm samples with comparable sample replication. Mid-infrared spectroscopy has greater potential to be affected by sample preparation methods because it has a greater intensity and specificity of the absorption features when compared to near-infrared frequencies (Gholizadeh et al., 2013). Moreover, Le Guillou et al. (2015) suggested that soil grinding has a lesser effect on vis–NIR spectra because the energy of light in this region is stronger than mid-IR, and the beam aperture of the instruments is often larger.

According to literature, increasing the number of analytical replicates could counteract spectral variability, increase sample representativeness, and account for differences in particle size and packing density (e.g., Mirzaeitalarposhti et al., 2017; Peng et al., 2014; Riedel et al., 2018; Terhoeven-Urselmans et al., 2010; Zhang et al., 2018). However, as mentioned earlier, most studies did not explicitly evaluate how varying numbers of spectral replicates affected spectral characteristics and prediction performance. Conversely, Peng et al. (2014) evaluated the effect of number of analytical replicates in mid-DRIFTS using different subsamples and found that the prediction performance of TOC and clay markedly improved from one to three replicates, and there was no difference then up to ten replicates. However, the effect of using the same or different soil subsamples to generate spectral replicates was still not clear. Replicating spectra in the same soil subsample produced less variable spectra then among different subsamples because soils are intrinsically heterogeneous (Figure 4). Inconsistent sample preparation (e.g., sample loading, packing and leveling) could add to that variability between different soil subsamples when compared to replicating spectra in the same soil subsample. However, this reduction in spectral variability

did not translate in accuracy improvements for most soil variables. Replicating spectra in *Within* wells reduced $RMSE_V$ for clay (18%) and POXC (2.7%) with <0.5 mm samples, and for Clay (21%) and pH (2.4%) with <2 mm samples (Supplemental Table S2). Even though counter intuitive, increasing the number of replicates increased the overall absorbance variability. This possibly occurred because of the absorbance variability within each wavenumber among spectral replicates of the same soil sample that resulted in greater averaged variability across all wavenumbers with more spectral replicates, especially when spectra was replicated *Across* subsamples. Our results indicate that four spectral replicates often improve prediction accuracy at either *Within* or *Across* methods of sample replication and that spectra can be replicated in the same subsample instead of different subsamples.

5 | CONCLUSION

High-throughput mid-DRIFTS is a valuable technique used to measure soil properties, but inappropriate methodological procedures can limit its predictive applications. A better understanding of how soil grinding and sample analytical replication can affect spectral properties enables mid-DRIFTS prediction optimization. Based on our results, finely grinding samples (<0.5 mm) and acquiring at least four spectral replicates are effective ways to build more accurate models for mid-DRIFTS predictions of soil variables. Recommendations on how much grinding is needed and how to replicate spectra for mid-DRIFTS should consider the tradeoffs between accuracy required for a specific application and the time and effort needed to prepare the soil samples as well as equipment availability. Grinding soils improved prediction accuracy for most soil variables, but required greater investment in equipment, time, and labor. The predictions with <2.0 mm samples increased spectral variability and generated greater prediction uncertainty, though did not jeopardize predictive applications and enabled greater laboratorial efficiency. Spectra replication within a single sample was equally as effective as multiple subsamples and this approach allows reducing the number of sub-samples to be scanned. Reading a single sample can be facilitated with autosamplers capable of reading multiple places within a single sample cup or well. Other factors such as acquisition parameters and chemometric approaches can be considered for prediction accuracy optimization. These recommendations will allow users to further optimize mid-DRIFTS prediction performance of soil physical, chemical, and biological variables in a high-throughput framework.

ACKNOWLEDGMENTS

We thank Bethany Herman, Nicole C. Hoekstra, Thomas Doohan, Mason Gingery, Madison Campbell, and Meredith Mann for their help in the laboratory. We recognize the National Cooperative Soil Survey in Ohio, which was the source of soil samples for this study. 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
Steve W. Culman https://orcid.org/0000-0003-3985-257X
M. Scott Demyan https://orcid.org/0000-0001-6198-3774

REFERENCES

Ahlrichs, J. L. (1968). Hydroxyl stretching frequencies of synthetic Ni-, AI-, AND Mg-hydroxy interlayers in expanding clays. *Clays and Clay Minerals*, *16*, 63–71. https://doi.org/10.1346/CCMN.1968.0160108

Baes, A. U., & Bloom, P. R. (1989). Diffuse reflectance Fourier transform infrared (DRIFT) spectroscopy 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

Barthès, B. G., Brunet, D., Ferrer, H., Chotte, J. L., & Feller, C. (2006). Determination of total carbon and nitrogen content in a range of tropical soils using near infrared spectroscopy: Influence of replication and sample grinding and drying. *Journal of Near Infrared Spectroscopy*, *14*, 341–348. https://doi.org/10.1255/jnirs.686

Barthès, B. G., Kouakoua, E., Moulin, P., Hmaidi, 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

Blanco-Canqui, H., & Lal, R. (2004). Mechanisms of carbon sequestration in soil aggregates. *Critical Reviews in Plant Sciences*, *23*, 481–504. https://doi.org/10.1080/07352680490886842

Burt, R. (2011). Soil survey laboratory information manual. Soil Survey Investigations Report No. 45 (Version 2.0). USDA NRCS, National Soil Survey Center, Lincoln, NE.

Canty, A. (2017). The Package 'boot': Reference manual. Comprehensive R Archive Network. Retrieved from https://cran.r-project.org/web/packages/boot/boot.pdf (accessed 24 Jan. 2019).

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

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: Kellogg Biological Station.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. UK: Cambridge Univ. Press. https://doi.org/10.1017/CBO9780511802843

Deiss, L., Margenot, A. J., Demyan, M. S., Culman, S. W. (2020a). Optimizing acquisition parameters in diffuse reflectance infrared Fourier transform spectroscopy of soils. *Soil Science Society of America Journal*. https://doi.org/10.1002/saj2.20028

Deiss, L., Margenot, A. J., Culman, S. W., Demyan, M. S. (2020b). Tuning support vector machines regression models improves prediction T accuracy of soil properties in MIR spectroscopy. *Geoderma*, *365*, 114227. https://doi.org/10.1016/j.geoderma.2020.114227

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

Hesse, M., Meier, H., & Zeeh, B. (2005). Spektroskopische Methoden in der Organischen Chemie. (In German.) Georg Thieme Verlag, Stuttgart. https://doi.org/10.1055/b-002-46985

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

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

Kuhn, M. (2018). The Package 'caret': Reference manual. Comprehensive R Archive Network. Retrieved from https://cran.r-project.org/web/packages/caret/caret.pdf (accessed 14 Jan. 2019).

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. Comprehensive R Archive Network. Retrieved from https://cran.r-project.org/web/packages/emmeans/emmeans.pdf (accessed 25 Jan. 2019).

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. San Diego, CA: Academic Press. https://doi.org/10.1002/bimj.4710240520

Martens, H., & Naes, T. (1989). *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. Retrieved from https://CRAN.R-project.org/package=e1071 (accessed 21 June 2017).

Mirzaeitalarposhti, R., Demyan, M. S., Rasche, F., Cadisch, G., & Müller, T. (2016). Overcoming carbonate interference on labile soil organic matter peaks for midDRIFTS analysis. *Soil Biology and Biochemistry*, *99*, 150–157. https://doi.org/10.1016/j.soilbio.2016.05.010

Mirzaeitalarposhti, R., Scott, M., Rasche, F., Cadisch, G., & Müller, T. (2017). Catena 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

Mutuo, P. K., Shepherd, K. D., Albrecht, A., & Cadisch, G. (2006). Prediction of carbon mineralization rates from different soil physical fractions using diffuse reflectance spectroscopy. *Soil Biology and Biochemistry*, *38*, 1658–1664. https://doi.org/10.1016/j.soilbio.2005.11.020

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

Ohio Division of Geological Survey (ODGS). (1998). Physiographic regions of Ohio. State of Ohio, Dep. of Natural Resources, Columbus.

Ohio Supercomputer Center (OSC). (1987). *Ohio Supercomputer Center*. Columbus, OH: Ohio Supercomputer Center.

Oksanen, J. (2018). The Package 'vegan': Reference manual. Comprehensive R Archive Network. Retrieved from https://cran.r-project.org/web/packages/vegan/vegan.pdf (accessed 15 Jan. 2019).

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

Core Team, R. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Retrieved from https://www.R-project.org/ (accessed 21 June 2017).

Reeves, J. B., & Smith, D. B. (2009). The potential of mid- and near-infrared diffuse reflectance spectroscopy for determining major- and trace-element concentrations in soils from a geochemical survey of North America. *Applied Geochemistry*, *24*, 1472–1481. https://doi.org/10.1016/j.apgeochem.2009.04.017

Reeves, J. B. (2010). Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? *Geoderma*, *158*, 3–14. https://doi.org/10.1016/j.geoderma.2009.04.005

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

Russell, J. D. (1987). Infrared spectroscopy of inorganic compounds. In H. Willis (Ed.), *Laboratory methods in infrared spectroscopy*. New York: John Wiley & Sons. https://doi.org/10.1016/B978-0-12-523450-4.50005-5

Senesi, N., D'Orazio, V., & Ricca, G. (2003). Humic acids in the first generation of EUROSOILS. *Geoderma*, *116*, 325–344. https://doi.org/10.1016/S0016-7061(03)00107-1

Soil Survey Staff. (2014). Kellogg soil survey laboratory methods manual. Soil Survey Investigations Report No. 42, Version 5.0. USDA-NRCS, 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. 1. Chemical fractionation and infrared spectroscopy. *Journal of Agricultural and Food Chemistry*, *55*, 2293–2302. https://doi.org/10.1021/jf0625398

Stenberg, B., Jonsson, A., & Börjesson, T. (2000). Snabbmetoder för bestämning av lerhalt och kalkbehov med hjälp av NIR-analysis. Rep. No. 2000–2. ODAL FoU, Lidköping, Sweden.

Stenberg, B., Jonsson, A., & Börjesson, T. (2002). Near infrared technology for soil analysis with implications for precision agriculture. In A. Davies & R. Cho (Eds.), *Near infrared spectroscopy: Proceedings of the 10th International Conference* (pp. 279–284). Chichester: NIR Publishing.

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

Stevens, A., & Ramirez-Lopez, L. (2015). An introduction to the prospect r package. GitHub, San Francisco, CA. Retrieved from 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

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

USDA-NRCS. (2006). Land resource regions and major land resource areas of the United States, the Caribbean, and the Pacific Basin. Handbook 296. USDA- NRCS, Washington, DC. Retrieved from https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_050898.pdf (accessed 26 Aug. 2019).

Varmuza, K., & Filzmoser, P. (2009). *Introduction to multivariate statistical analysis in chemometrics*. Boca Raton, FL: CRC Press. https://doi.org/10.1201/9781420059496

Verchot, L. V., Dutaur, L., Shepherd, K. D., & Albrecht, A. (2011). Organic matter stabilization in soil aggregates: Understanding the biogeochemical mechanisms that determine the fate of carbon inputs in soils. *Geoderma*, *161*, 182–193. https://doi.org/10.1016/j.geoderma.2010.12.017

Viscarra-Rossel, R. A. V., & Webster, R. (2012). Predicting soil properties from the Australian soil visible–near infrared spectroscopic database. *European Journal of Soil Science*, *63*, 848–860. https://doi.org/10.1111/j.1365-2389.2012.01495.x

Wada, K., & Greenland, D. (1970). Selective dissolution and differential infrared spectroscopy for characterization of 'amorphous' constituents in soil clays. *Clay Minerals*, *8*, 241–254. https://doi.org/10.1180/claymin.1970.008.3.02

Waruru, B. K., Shepherd, K. D., Ndegwa, G. M., & Sila, A. M. (2016). Estimation of wet aggregation indices using soil properties and diffuse reflectance near infrared spectroscopy: An application of classification and regression tree analysis. *Biosystems Engineering*, *152*, 148–164. https://doi.org/10.1016/j.biosystemseng.2016.08.003

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

Zhang, L., Yang, X., Drury, C., Chantigny, M., Miller, J., Reynolds, D., … District, S. (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, Culman SW, Demyan MS. Grinding and spectra replication often improves mid-DRIFTS predictions of soil properties. *Soil Sci Soc Am J*. 2020;1–16. https://doi.org/10.1002/saj2.20021