

INFRARED SPECTROSCOPY FOR SOIL CARBON ACCOUNTING AND SOIL
QUALITY ASSESSMENT

A Dissertation

Presented to the Faculty of the Graduate School
of Cornell University

In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

by

Sonam Rinchen Sherpa

August 2019

© 2019 Sonam Rinchen Sherpa

INFRARED SPECTROSCOPY FOR SOIL CARBON ACCOUNTING AND SOIL QUALITY

Sonam Rinchen Sherpa, Ph. D.

Cornell University 2019

The future of global food security and economic stability continue to raise increasing concern, as the human population and thus demand for agricultural resources are rising at an alarming rate, while the frequency of extreme weather events leading to drought, flooding, and heat stress are projected to increase. Furthermore, the impacts of climate change on crop productivity and water availability are already apparent, and if current trends of increasing atmospheric CO₂ levels and land degradation persist, the stability of whole food systems could be at risk. Building soil organic carbon (SOC) is essential for restoring degraded soils and enhancing soil health, as increasing SOC can reduce compaction, improve drainage and aeration, increase water holding and cation exchange capacity, and enhance nutrient cycling and fertility, thereby improving agronomic productivity and resource use efficiency in agroecosystems. Implementing soil management practices to build SOC can provide win-win opportunities for farmers and resource managers, resulting in improved soil health, increased fertilizer use efficiency, and increased resiliency to drought and heat stress, while serving to mitigate climate change by sequestering atmospheric carbon. However, the excessive sampling and analysis costs required to measure baseline levels and monitor changes in soil quality, currently limits our ability to establish

effective policies for incentivizing sustainable soil management practices.

Mid infrared (MIR) and visible and near infrared (VNIR) spectroscopy have been proposed as accurate and low cost options for predicting multiple soil quality indicators, and may be capable of meeting current needs of a low cost reliable method for quantifying soil quality and SOC. In this dissertation, I explore potential applications for MIR and VNIR spectroscopy by developing and testing low-cost, reliable measurement-based approaches for quantifying soil quality and SOC at farm-to-regional scales. Both VNIR and MIR spectroscopy successfully predicted SOC and other soil quality indicators with an acceptable level of accuracy often comparable to that of standard wet chemistry methods for soil assessment, yet measurement costs for both VNIR and MIR predictions was roughly an order of magnitude less than standard wet chemistry methods.

BIO GRAPHICAL SKETCH

Sonam Sherpa was born on August 16, 1983 in Pequannock, New Jersey to Judy and Topgyal Sherpa. He has two older brothers, Jayson Rome and Pema Sherpa. Sonam attended high school at the Windsor Academy in Morristown, New Jersey. Upon completion of high school, he spent one year travelling around India and Nepal, studying Tibetan culture and religion.

Having experienced and embraced his cultural heritage Sonam returned to the United States ready to begin his higher education. In August of 2003, Sonam enrolled in the Pre-Environmental Science and Forestry program at Alfred State College, State University of New York in Alfred, New York. He completed his Associate in Arts degree in the May of 2005, and by August 2005 he was continuing his education at Cornell University Ithaca, New York. In January 2007, Sonam completed his Bachelors of Science at Cornell with a dual degree in Natural Resources and Plant Sciences. Upon completion of his undergraduate degree, Sonam continued to explore his passion for ecology and the environment as a staff member of the Cornell Soil Health Team, until he began graduate studies at Cornell University in August of 2011.

Dedicated to the loving memory of Jayson Evert Rome, Khenpo Sonam Gyatso Rinpoche, and Tashi Chopal Sherpa. Their love, guidance, and vigilant support enabled this journey and continues to illuminate my path forward.

ACKNOWLEDGMENTS

First, I would like to thank my family and friends for all of their love and support, both during the good times and the bad times. I would not be here today without all of you.

I would like to give a special thanks to my mother Judy, father Topgyal, and brother Pema. Words cannot explain my gratitude for the unending love, compassion, and support that you have shown throughout this journey.

I am extremely thankful to have had the opportunity to peruse my Ph.D. under the guidance of Dr. David Wolfe, who fostered an environment of both intellectual and creative freedom, yet always pushed me to challenge myself. I would also like to thank my committee members, Drs. Harold van Es, Peter Woodbury, and Stephen Reiners. Their critical feedback and positive support has allowed me to develop into a better scientist. Many thanks to Dr. Richard Ferguson and the USDA-NRCS Kellogg Soil Survey Laboratory for their help with obtaining the MIR spectral library and soils database used in this study, and for their guidance on database management and soil spectroscopy approaches. Field and laboratory work could have been possible without the help of Jeff Beem-Miller, Ryan Devlin, Stefan Karkuff, Jonathan Comstock, Paul Thomas, and Jessica Leonard. Finally, thank you to the United States Department of Agriculture's National Institute for Food and Agriculture, for so graciously providing financial support for my research and studies.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH	iv
DEDICATION	v
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	vii
CHAPTER 1. EVALUATING VNIR SPECTROSCOPY FOR FARM-LEVEL SOIL CARBON ACCOUNTING IN MAIZE CROPPING SYSTEMS	1
CHAPTER 2. MOVING BEYOND THE BLACK BOX WITH MID INFRARED SPECTROSCOPY AND PREDICTION OF SOIL QUALITY INDICATORS	41
CHAPTER 3. GROUND-TRUTHING MID-IR SPECTROSCOPY FOR VERIFYING SOIL CARBON: IS THE ACCURACY OF STANDARD SOIL ANALYSES WORTH THE COST?	78

CHAPTER 1

EVALUATING VNIR SPECTROSCOPY FOR FARM-LEVEL SOIL CARBON ACCOUNTING IN MAIZE CROPPING SYSTEMS

Abstract

Our ability to quantify soil organic carbon (SOC) baseline levels and change at the farm-level is currently constrained by the highly variable nature of SOC, and excessive sample size requirements. This study evaluated visible and near-infrared diffuse reflectance spectroscopy (VNIR) as a rapid and inexpensive tool for quantifying SOC content at the farm-level for maize cropping systems in NY and IA, USA. Partial least squares regression models were constructed using 3919 samples collected from 15 farms in NY and 2 in IA, with farms reflecting the major soil types, management practices, and biophysical regions for maize production. Soils were sampled at 0-10, 10-20, 20-30, 30-50, and 50-75 cm depth increments. Stratification by depth improved SOC prediction accuracy, and errors were further reduced when carbonate containing soils were removed. VNIR successfully predicted SOC, allowing for accurate SOC accounting both within-field and among farms. Spiking the calibration sets with local samples reduced the root mean square error (RMSE) for farms between 1 to 62 %. The influence of soil texture, SOC content, and soil variability on model performance at the farm-level was explored. Spiking was most effective when spectral characteristic of a farm were poorly represented in the calibration set, but for most sites, the small improvement in accuracy would not justify additional costs. Benefits from VNIR were greatest at farm sites where SOC was

highly variable, as reduced per-sample costs would allow more thorough sampling and characterization of SOC spatial variability.

Introduction

Soil organic carbon (SOC) is a key property for maintaining soil productivity and environmental quality, yet historical land use practices associated with intensive agriculture have reduced SOC stocks throughout the globe by 78 ± 12 Pg C (Zomer et al., 2017). Additionally, there has been growing interest in sequestering SOC in agricultural soils as a strategy for mitigating greenhouse gas emissions and lowering atmospheric CO₂ levels (Paustian et al., 2016; Griscom et al., 2017). Much of the appeal of SOC-based mitigation strategies is due to the co-benefits for farmers resulting from increased soil organic matter (SOM), such as increases soil microbial diversity and biogeochemical cycling of nutrients, improved soil structure, nutrient and water retention, and resilience of soils (Wolfe, 2013).

Successful development and implementation of policies that incentivize farmer adoption of practices that build SOC will require monitoring and establishment of baseline levels. Furthermore, low-cost approaches for quantifying soil C change are crucial if a C economy for the agricultural sector is to develop (de Gruijter et al., 2016). The challenge is that SOC is highly variable in both space and time, and so accurately quantifying SOC stocks in agroecosystems can be cost prohibitive due to excessive soil sampling and laboratory analyses costs (Sherpa et al., 2016). Standard laboratory analyses for measuring SOC (i.e. dry or wet combustion) are time consuming, expensive, and can require hazardous reagents. and so there is need for

rapid, robust, and reliable techniques for assessing SOC. Visible and near-infrared diffuse reflectance spectroscopy (VNIR) has been used as a safe, non-destructive, fast, precise, and inexpensive approach to predict soil properties, including SOM, C, N, pH, and soil texture (Chang et al., 2001; Cozzolino and Morón, 2006; Vasques et al., 2009).

Successful SOC prediction by VNIR has been widely documented, demonstrating this technology's potential as a tool for low cost SOC assessment; however, small sample sizes with narrow geographic coverage are a common limitation among soil spectroscopy studies, raising questions regarding the extent to which such findings may be extrapolated to new unknown sites (Cohen et al., 2005; Viscarra Rossel et al., 2006; Igne et al., 2010; Terhoeven-Urselmans et al., 2010). Increasing the spectral diversity and therefore sample size of VNIR spectral libraries generally allows for better representation of local soil diversity, but excessive sample diversity will hinder model performance. Calibrations developed for broad geographic regions (i.e. global, continental, or national scale), tend to display poor performance. Soil-spectral relationships can vary spatially and by soil taxonomy. Additionally, relationships may be nonlinear, and the absorption spectra of one soil component can be masked, distorted, or shifted by variation in other soil components with overlapping absorption features (Price, 1994). Strategies for improving the accuracy of VNIR prediction models have largely focused on developing optimized calibrations that attempt to account for this soil-spectral variability.

Approaches to optimizing VNIR calibrations can be grouped into two broad classes: 1) stratification approaches, where a subset of the full spectral library is selected based

on soil chemical, spectral, or geographic characteristics, and individual calibrations are developed for the subsets (Gogé et al., 2014; Nocita et al., 2015; Wight et al., 2016); and 2) spiking approaches, where a subset of samples from the local-site are incorporated into calibration models developed from a broader spectral library (Brown, 2007; Gogé et al., 2012; Guerrero et al., 2014). There is a fundamental difference in how these approaches account for local soil-spectral relationships of new unknown sites. Stratification requires a spectral library large enough to encompass the spectral variation of all potential unknown soils, and samples are strategically removed until models are calibrated to the specific local conditions of unknown sites. Conversely, spiking is well suited to smaller regional-scale spectral libraries, but requires that for each new unknown site, a subset of samples are analyzed by wet chemistry and included in model calibration. From a practical standpoint for SOC assessment, there are trade-offs between low-cost measurement and accuracy. Stratification has the obvious benefit of reduced analytical costs and therefore increased sample sizes, but it can require a larger initial investment to develop the spectral library. On the contrary, substantial model improvements gained from spiking would likely justify the increased analytical costs.

Soil organic matter, carbonates, and clay minerals all have known absorption features in the VNIR region, and represent major components of most soils. Stratification based on soil textural classes or depth may improve the performance of VNIR prediction models, as SOM, carbonate, and clay content often vary as a function of soil depth and taxonomy (Viscarra Rossel and Chen, 2011). Wight et al. (2016), constructed artificial soils to investigate the effects of organic matter type and

concentration, soil texture, clay mineralogy, and their combined interactions on VNIR SOC predictability. While clay mineralogy had no effect on SOC predictability, the total clay content of soils was the primary characteristic influencing VNIR SOC model performance, and stratification by texture class was recommended whenever possible. Organic matter type also seemed to affect SOC predictability.

Even though quartz has no spectral features in the VNIR region, increasing sand content can increase SOC absorption depths, confounding soil-spectral relationships, and diluting the spectral signature of other soil components (Stenberg et al., 2010; Vasques et al., 2010). Additionally, SOC predictability has been shown to depend on the degree of association between SOC and sand particles, and the speciation of SOM (Vasques et al., 2010). Guerrero et al. (2014), compared different strategies for selecting spiking subsets when predicting SOC at the local scale. Calibrations performed best when spiking samples were evenly distributed across the space defined by the first three principal components of the spectra, and spiking with a small local-subset resulted in SOC predictions with more accuracy than geographically stratified models. While both the stratification and spiking approaches have been shown to improve SOC predictions, which strategy is best for predicting SOC at a farm is still unclear.

If VNIR is to be used as a tool for establishing SOC baseline levels and monitoring temporal change, or for verifying the effectiveness of policy incentives, then predictions must be accurate enough at the farm scale to detect SOC change with sufficient statistical confidence (de Gruijter et al., 2016). Validating SOC prediction models at local farm sites provides a more robust estimate of model performance,

better reflecting real world conditions and potential unknown soils, and yet few studies have evaluated the performance of VNIR SOC predictions at individual farms. To our knowledge, SOC prediction by VNIR has not been evaluated for a large set of farms representative of the typical biophysical characteristics and management practices for a single crop production system.

The aim of this study was to evaluate the potential for using VNIR spectroscopy as a low-cost reliable method for quantifying SOC at the farm-level in maize cropping systems, and to test VNIR's capacity for accurately predicting SOC at new farm sites located within the study regions. To accomplish this, we integrated spatially explicit databases, including biophysical, land use, land management, and crop production data. Selection criteria were developed to ensure that farm evaluation sites were representative of key climactic conditions, soil types, and field management practices for maize (*Zea mays* L.) production in New York, USA. In Iowa, USA similar stratification was performed, but sampling was limited to 2 sites for logistical reasons. Our specific objectives were to:

- 1) develop a geospatial biophysical landscape stratification approach for selecting farm-level evaluation sites representative of the predominant environmental and management conditions for maize production in the study regions;
- 2) evaluate depth, soil carbonates, texture, and SOM as soil properties with potential as stratification criteria for VNIR model calibration,
- 3) quantify the benefit of including local samples (spiking) in model calibration when predicting SOC at individual farms, and identify site conditions where inclusion of local samples is recommended; and

4) assess the accuracy of SOC predictions at individual farm sites, and evaluate VNIR's potential as a tool for farm-level SOC accounting in maize cropping systems of NY and IA.

Materials and Methods

Farm site selection and geospatial biophysical stratification

Farm sites were selected based on a geospatial stratification plan aimed at identifying key representative combinations of soil and climate characteristics along with key combinations of management practices for maize production systems in NY and to a lesser extent in IA. Geospatial analyses utilized a variety of spatially explicit data layers including land cover, land use, soils, terrain characteristics, maize yield data, and climatic conditions throughout the study region. The USDA National Agricultural Statistics Service Cropland Data Layer was used to determine the location of maize production, and of other crops grown in rotation with maize, notably soybean and alfalfa. The location of maize production throughout the study region encompassing NY and IA was determined, and the non-irrigated maize yields were calculated based on soil characteristics and historical yield data, and growing degree days based on 30-year records (1971-2000) from individual weather stations near potential field sampling sites.

The four predominant Major Land Resource Areas (MLRAs) for maize production acreage in NY were identified, and these regions were used as the coarsest level of geospatial stratification. The selection of sites was further refined based on the availability of reliable, detailed, and long-term management records. Additional

priority was given to sites with contrasting management practices either on the farm or within the nearby vicinity. The final stratification scheme incorporated the major soil texture classes, crop rotations, tillage, and nutrient management practices representative of maize acreage for the different climactic regions (Table 1). A total of 3919 soil samples were collected from 15 farm sites in NY and 2 farm sites in IA between the years of 2011 and 2014. Two distinct classes of farm sites were sampled: 1) long-term agronomic field experiments, and 2) commercial farms. Soil samples from commercial farms were obtained in collaboration with cooperating maize growers, where the management history of sampled fields reflected typical practices for that farm. Detailed information regarding the soils and management history for individual farms is provided in Table 1.

Sampling design

Prior to field sampling, geospatial data layers (management boundaries, soil series, and terrain indices) were obtained for each site, and uploaded to a geographic information system. These data were then used to identify a sampling area representative of the management unit in question for each farm. For most sites, soil cores were collected from within a 1 ha sampling area, where sampling locations formed 5 equally distanced *equilateral triangles* (20 m length to each side) (Conant and Paustian, 2002). For 3 sites (Intensive Sites), high density sampling was conducted, so that SOC estimates could be evaluated at finer spatial scales, such as at the within-field level. Intensive sampling ranged from 51 cores collected over a 1 ha sampling area (NY-C5), 188 soil cores over 20 ha (NY-W3), and 232 cores collected

over 25 ha (IA-N1). Long term tillage field experiments were conducted at three of the sites (IA-N2, NY-N1, NY-N2); samples were collected along transects for each experimental treatment. At one of the long term experimental sites (NY-N2), 23 soil cores (115 samples) were collected from uncultivated regions of the farm with a long history (>20 yrs) in sod. Uncultivated regions of the farm were sampled to estimate the potential for increasing SOC due to changes in soil management.

Table 1. Summary of the soil characteristics, cropping history, and biophysical regions for the 17 farms in this study.

Farm	Region	MLRA†	Site type	Cropping‡	Soil Order§	Texture Class§	– SOC % –	
							n	Range
IA-N1	Northern IA	103	commercial	CS/COA	Mollisol	Loam/clay loam/ silty clay loam	1149	0.03-5.42
IA-N2	Northern IA	103	experimental	CC/CS/COA	Mollisol	Clay loam	97	0.20-4.84
NY-C1	Central NY	101	experimental	CC/AA	Alfisol	Loam	59	0.00-2.78
NY-C2	Central NY	101	commercial	CSW	Alfisol	Gravelly silt loam	115	0.23-2.82
NY-C3	Central NY	101	commercial	CC	Alfisol	Silt loam	91	0.01-3.58
NY-C4	Central NY	101	commercial	CS	Alfisol	Silt loam/loam	144	0.10-5.28
NY-C5	Central NY	140	experimental	CAA	Alfisol	Gravelly loam	253	0.27-3.50
NY-E1	Eastern NY	144A	commercial	CC/WCS/AA	Inceptisol, Alfisol	Silt loam	105	0.11-4.10
NY-E2	Eastern NY	144A	experimental	CAA	Inceptisol	Gravelly sandy loam	96	0.08-1.92
NY-N1	Northern NY	142	experimental	CC	Inceptisol	Silt loam	152	0.01-3.51
NY-N2	Northern NY	142	experimental	CC/G	Alfisol	Silty clay	305	0.03-9.34
NY-W1	Western NY	101	commercial	CSW	Alfisol	Cobbly loam	94	0.02-3.09
NY-W2	Western NY	101	commercial	CS	Alfisol	Cobbly loam	100	0.06-2.56
NY-W3	Western NY	101	commercial	CS	Alfisol	Silt loam	908	0.09-3.44
NY-W4	Western NY	140	commercial	CA	Inceptisol, Alfisol	Fine sandy loam/silt loam	99	0.13-3.52
NY-W5	Western NY	140	commercial	CA	Inceptisol	Channery loam	64	0.20-4.96
NY-W6	Western NY	140	commercial	CA	Alfisol	Gravelly silt loam	200	0.16-3.24

† MLRA description: 101, Ontario-Erie Plain and Finger Lakes Region; 140, Glaciated Allegheny Plateau and Catskill; 144A, New England and Eastern New York Upland, Southern Part; 142, St. Lawrence-Champlain Plain; 103, Central Iowa and Minnesota Till Prairies

‡ Cropping rotations have been simplified to represent general sequence over 10 years of management prior to sampling; "/" indicates cropping on multiple fields: C = maize, O = oats, S = soybean, A = alfalfa, W = wheat, G = grass sod.

§ Soil taxonomic data were obtained from the national Soil Survey Geographic (SSURGO) database (Soil Survey Staff, 2014).

Soil laboratory analyses

Soil samples were oven dried at 35°C for a minimum of 48 hr, and passed through a 2 mm sieve. A 20-g subsample was dried at 105°C for 48 hr and weighed, so that the soil moisture content could be adjusted to a 105°C oven-dry basis. Soil organic C concentration was obtained by subtracting soil inorganic C (SIC) content from total C content (TC). Total C was measured by dry combustion (LECO CN 2000; LECO, St. Joseph, MI). Soil inorganic C (SIC) was predicted by MIR spectroscopy using a nationwide partial least squares regression (PLSR) model, which was developed by the Kellogg Soil Survey Laboratory (KSSL) of the United States Department of Agriculture-Natural Resources Conservation Service (USDA-NRCS). Measured SIC values used in model calibration were determined by pressure-calculator (Sherrod et al., 2002). The SIC and MIR analyses were performed at KSSL according to methods 4Ea1a1a1 and 7A7, respectively, of the KSSL Methods Manual, Soil Survey Investigations Report (SSIR) No. 42 (Soil Survey Staff, 2014). To test the performance of the KSSL nationwide carbonate prediction model, SIC was measured by pressure-calculator for a subset of samples from this study (n = 989). The accuracy of SIC predictions was excellent ($R^2 = 0.95$ and RMSE = 0.16 %), and so we felt justified using MIR predicted SIC content to determine SOC content. Soil particle size distribution (sand, silt, and clay content) was assessed using a rapid quantitative method developed by Kettler et al. (2001); where 3% hexametaphosphate ((NaPO₃)_n) is used to disperse the soil sample, and a combination of sieving and sedimentation steps is used to separate size fractions.

Spectral measurement

Soil VNIR (350-2500 nm) diffuse reflectance spectra were collected in the laboratory from air-dried and 2-mm sieved samples using a Fieldspec Pro hyperspectral sensor (Analytical Spectral Devices, Boulder, CO). The equipment had a spectral resolution of 3 nm from 350 nm to 1000 nm and 10 nm from 1000 to 2500 nm. During scanning, soils were placed in a 4cm diameter optical petri dishes (Electron Microscopy Sciences, Hatfield, PA), housed in an ASD high intensity muglight to minimize measurement errors associated with stray light and specular reflected components. Each sample was scanned 50 times, rotated 90°, and scanned again 50 times. Spectra were averaged to reduce noise associated with the detection equipment. Instrument performance was monitored through the use of a Spectralon (Labsphere, Inc., North Sutton, NH) white reference material with a known reflectance. A white reference spectrum was taken every 8 samples, and if irregular values were obtained, the previous 8 spectra were discarded, and the instrument was recalibrated.

Spectral preprocessing and modeling

Spectral preprocessing and chemometric analyses were performed using OPUS QUANT 2 software version 7.2 (BrukerOptik GmbH 2012, Ettlingen, Germany). Several spectral regions were removed (350-420, 961-1019, 1771-1829, and 2481-2500 nm) due to a low signal-to-noise ratio caused by the splicing of individual sensors within the spectrometer. Soil properties were predicted from VNIR spectra by partial least squares regression (PLSR). The OPUS QUANT 2 search routine was used to optimize spectral preprocessing and wavelength range selection by automatically and iteratively testing spectral preprocessing options in various wavelength ranges. Spectral preprocessing options included: 1) no data treatment, 2) subtraction of a constant offset,

3) subtraction of a linear function (SLF), 4) vector normalization (VN), 5) minimum-maximum scaling, 6) multiplicative scatter correction (MSC), 7) 1st derivative, 8) 2nd derivative, 9) 1st derivative and SLF, 10) 1st derivative and VN, and 11) 1st derivative and MSC (Conzen, 2006). All calibration results were ranked and the method with the lowest prediction error was used to construct the model. To minimize the potential for over-fitting, the maximum number of prediction factors was restricted to 20.

For both the stratified models and the model calibrated using the full dataset, spectra were divided into separate calibration (75%) and evaluation sets (25%). Sample sets were selected based on a principal component analysis performed on the spectra, where the Kennard-Stone algorithm was used to distribute calibration samples equally across the final principal component space. This approach helps ensure that the spectral characteristics of the entire database are well represented in both the calibration and internal evaluation sets (Gogé et al., 2014; Guerrero et al., 2015).

Model stratification by soil depth and carbonates

Soil depth and carbonate presence were evaluated as criteria for model stratification, where individual *depth stratified models* were constructed for each soil depth increment. Depth increments were further stratified based on the absence or presence of carbonates, and individual PLSR models developed. Stratification by carbonates had a negligible effect for depths D10, D20, and D30, and so only the results for D50 and D75 were reported. Additionally, a *full dataset model* including samples from all depths and sites was constructed. To compare the full dataset model with the depth stratified models, SOC predictions were separated by depth increment, and evaluation statistics were calculated for the individual depths. Similarly, SOC

predictions from the five depth stratified models were pooled, and evaluation statistics were calculated for all of the depths combined, for comparison with the results from the full dataset model.

Farm-level evaluation and spiking analysis

To better assess VNIR's potential as a tool for farm-level SOC accounting, a leave-one-farm-out evaluation approach was employed. Individual *site not included* VNIR models were constructed for each of the 17 farm sites, by withholding the respective farm from model calibration. The removed farm sites served as independent evaluation sets for the 17 site not included models, providing a more robust estimate of model performance when predicting SOC at new farm sites. An additional *Spiking 50%* VNIR model was constructed with the above described PCA approach used to split each farm site into two sample sets of equal size, where half of the samples were allocated to model calibration, and half to model evaluation. The accuracy of SOC predictions was evaluated at each farm site for models constructed with 0 % and 50 % of local samples allocated to model calibration.

VNIR-predicted SOC stocks to 30 cm

Soil organic C stocks to a depth of 30 cm were calculated using SOC content predicted by the full dataset model, combined with measured bulk density values. Bulk density for the fine-earth fraction was calculated as described by (Holmes et al., 2012) using the following formula:

$$\rho_b = \frac{M_{soil} - M_{rock}}{V_{soil} - \frac{M_{rock}}{\rho_{rock}}} \quad (1)$$

where M_{soil} is the mass of the <2 mm fraction, V_{soil} is the volume of the <2 mm fraction, M_{rock} is the sample mass of the >2 mm fraction, and ρ_{rock} is the rock density, assumed to be 2.6 g cm⁻³. Soil water content was used to adjust M_{soil} to a 105°C oven-dry basis.

Soil organic C stocks (per unit area) to a depth of 30 cm were calculated with the following equation:

$$SOC_{stocks} = \rho_b * SOC_{conc} * D * (1 - CF) \quad (2)$$

where ρ_b is the bulk density of the fine-earth fraction, SOC_{conc} is the SOC concentration, D is sampling depth, and CF is the fraction of coarse material >2 mm on a volume basis. The accuracy of VNIR-predicted SOC stocks to a depth of 30 cm was evaluated for each farm site.

Model evaluation

The performance of VNIR for predicting SOC content was evaluated based on several common chemometric indicators; the coefficient of determination (R^2), the root mean square error of prediction (RMSE), and model bias (Bellon-Maurel et al., 2010). To address some of the more applied questions related to VNIR's practical capacity and limitations for SOC accounting, model performance was evaluated in the context of quantifying SOC at varying spatial scales ranging from among farm to within-field.

The RMSE represents the mean error expected when new unknown samples are measured, and was estimated by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_{pred.} - Y_{meas.})^2}{n}} \quad (3)$$

where $Y_{pred.}$ is the predicted value; $Y_{meas.}$ is the measured value; and n is the number of observations.

Bias reflects the inherent structural error often originating from equipment and other

experimental sources, and was defined as:

$$\text{Bias} = \bar{Y}_{pred.} - \bar{Y}_{meas.} \quad (4)$$

where $\bar{Y}_{pred.}$ is the mean of the predicted values; and $\bar{Y}_{meas.}$ is the mean of the measured values.

The relative error (RE), or RMSE relative to the mean, was defined as:

$$RE = \frac{RMSE}{mean} * 100 \quad (5)$$

where $RMSE$ is the mean error of SOC predictions, and $mean$ is the mean measured SOC content of the evaluation set.

The relative improvement gained from spiking (RI), or percent reduction in RMSE at a given farm site, was defined as:

$$RI = \frac{RMSE_{no.} - RMSE_{Spk.}}{RMSE_{no.}} * 100 \quad (6)$$

where $RMSE_{no.}$ is the mean error for the site not included model, and $RMSE_{Spk.}$ is the mean error for the spiking 50% model at a given farm site.

Results and Discussion

Measured SOC descriptive statistics

The SOC concentration across all sites and all depths ranged from 0 to 9.34 % (Table 1), with a highly skewed SOC distribution where 75 % of samples showed SOC concentrations below 2.29 %. Averaging across all sites, SOC concentration decreased with soil depth from a mean of 2.57 % at D10 to 0.64 % at D75 (Table 2). There was little variation in SOC concentration among the Mollisols samples collected from IA-N1 and IA-N2, but high variation among the 15 farm sites in NY. Site NY-N2 had the highest SOC concentration (9.34 %), well above the upper limits of the other sites in NY (Table 1).

Depth

When compared across all sites and depths, the accuracy of SOC predictions by VNIR was satisfactory, with the full dataset model having an $R^2 = 0.84$, RMSE = 0.43, and RE = 34 (Table 3; Fig. 1a). Stratifying calibration sets by depth improved the accuracy of SOC predictions at all individual depth increments (Table 3). When pooled across all depths, stratified VNIR models resulted in excellent SOC predictability with $R^2 = 0.88$, RMSE = 0.38, and RE = 23.

Depth stratified VNIR models effectively predicted SOC concentration for depths down to 30 cm, but predicted SOC poorly for subsoils. Prediction accuracy increased with depth for D10, D20, and D30, while this trend was reversed for D50 and D75 (Table 3). The mean measured SOC concentration and RMSE were (0.99 and 0.34 %), and (0.72 and 0.44 %) for the D50 and D75 depth models, respectively. Errors for the subsoil models were within the range of other models, but were much greater relative to the evaluation set mean. Several authors have noted depth effects on the performance of SOC predictions by VNIR (Lee et al., 2009; Morgan et al., 2009; Bellon-Maurel and McBratney, 2011). Surface soils are exposed to more frequent and severe disturbance by biological and environmental processes such as organic matter inputs, erosion, and tillage. These disturbances can result in increased heterogeneity of SOM content and composition, which may explain why D30 outperformed depths D10 and D20.

Table 2. Summary statistics for soil organic carbon (SOC), carbonates (CaCO₃), clay, and sand contents. Results are presented for the whole dataset and individual soil depth increments.

Depth	n	Mean	SD	Median	Min	Max	Mean	SD	Median	Min	Max
		SOC %					CaCO ₃ %				
All	3919	1.64	1.06	1.66	0.00	9.34	2	7	0	0	55
0-10cm	806	2.57	0.94	2.37	0.54	9.34	0	1	0	0	14
10-20cm	805	2.22	0.73	2.11	0.37	5.17	0	1	0	0	24
20-30cm	806	1.69	0.85	1.66	0.09	4.84	0	3	0	0	30
30-50cm	781	0.94	0.71	0.77	0.00	5.28	3	8	0	0	44
50-75cm	721	0.64	0.56	0.46	0.01	3.35	6	12	0	0	55
		Clay %					Sand %				
All	3918	15	11	14	0	76	29	19	32	0	95
0-10cm	806	15	9	12	0	49	33	17	32	0	63
10-20cm	805	15	9	12	0	63	33	17	32	0	65
20-30cm	806	16	11	14	0	69	31	18	31	0	70
30-50cm	780	19	14	15	0	70	32	20	32	0	91
50-75cm	721	20	13	16	0	76	33	21	33	0	95

Table 3. Summary of evaluation statistics for VNIR predicted soil organic C (SOC) % concentration by the full dataset model and depth stratified models. Results are presented for each individual soil depth increment and all depths combined.

Depth	SOC %											
	Full dataset model						Depth stratified models					
	-measured- mean	n	—VNIR predictions— R ² RMSE Bias RE†				-measured- mean	n	—VNIR predictions— R ² RMSE Bias RE			
All Depths	1.27	980	0.84	0.43	0.03	34	1.63	982	0.88	0.38	-0.01	23
0-10cm	2.79	108	0.79	0.58	-0.08	21	2.60	202	0.81	0.45	0.00	17
10-20cm	2.11	90	0.78	0.40	0.01	19	2.19	202	0.84	0.33	0.02	15
20-30cm	1.52	201	0.84	0.37	-0.02	24	1.51	202	0.87	0.31	0.01	20
30-50cm	0.96	295	0.79	0.35	-0.04	37	0.99	195	0.78	0.34	0.02	35
50-75cm	0.62	292	0.21	0.48	-0.04	78	0.72	181	0.52	0.44	0.02	60

†Relative error (RE) = 100* RMSE * (measured mean SOC)⁻¹.

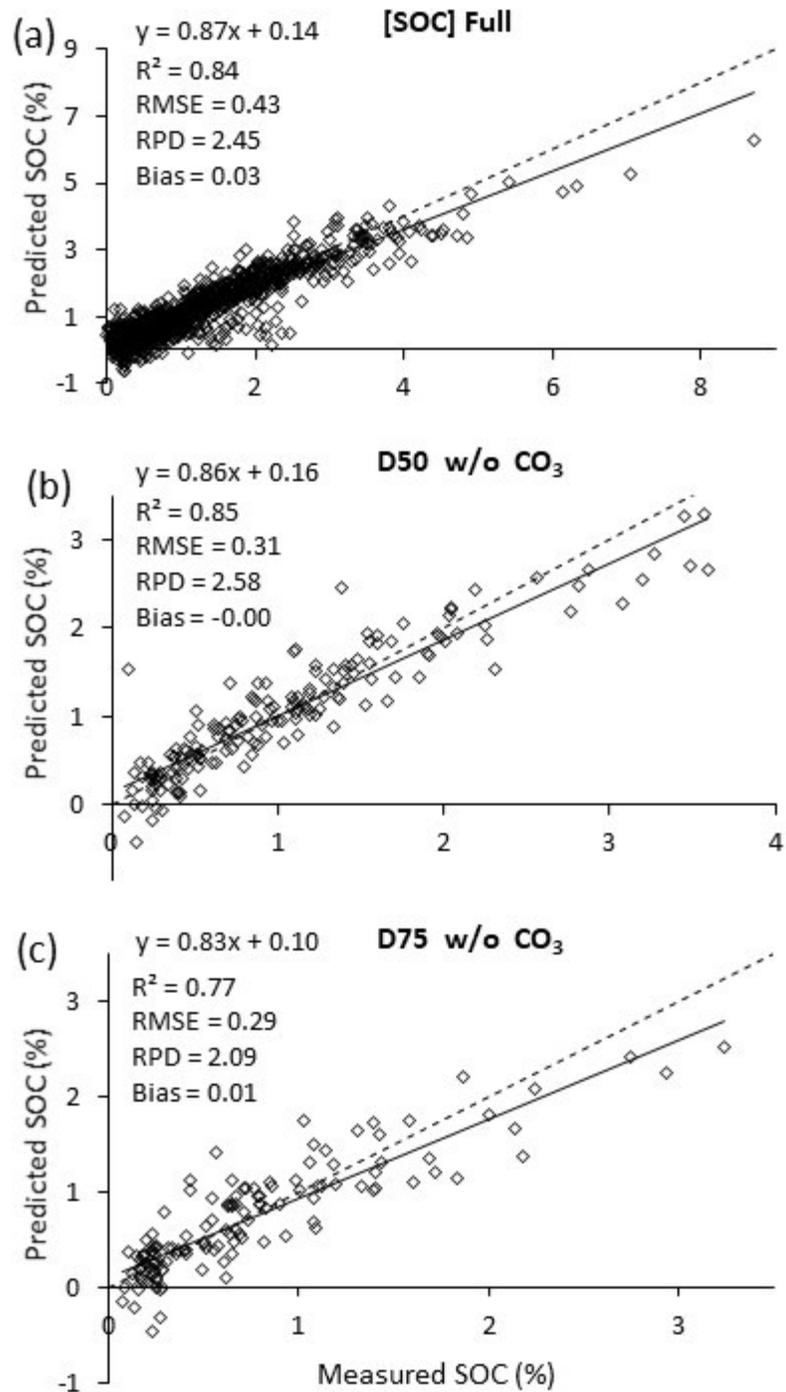


Fig. 1. (a) Predicted vs. measured SOC concentration for the full dataset model. Dash line is 1:1 line, solid line is regression line. Predicted vs. measured. (b) D50 predicted SOC with carbonate containing samples removed from the model. (c) D75 predicted SOC with carbonate containing samples removed from the model. The dashed line represents the 1:1 line, and the solid line is regression line.

Carbonates

The performance of VNIR predictions for depths D50 and D75 greatly improved when samples with carbonates ($n = 114$ and 198 for D50 and D75, respectively) were excluded from calibration and evaluation sets (Fig. 1b, 1c). Samples with carbonates had a slightly higher mean reflectance than samples without carbonates at all wavelengths from 400 to 2400 nm (Fig. 2a). Such shifts in reflectance may be due in part to soil depth, as carbonates were seldom observed in the top 30 cm, and SOC generally decreased with depth while carbonates tended to increase. Reeves (2010) found that near infrared (NIR) reflectance increased with carbonate concentration, and previous studies have indicated that the presence of carbonates can degrade SOC predictions (McCarty et al., 2002; Reeves and Smith, 2009; Bilgili et al., 2010). Carbonates may mask spectral features related to SOC, and soils with low SOC content relative to carbonates can pose methodical challenges for both SOC measurement and prediction (e.g. reduced signal to noise). The presence of carbonates appeared to be limiting the accuracy of SOC predictions at depths below 30 cm, contributing to the high prediction errors observed for subsoils.

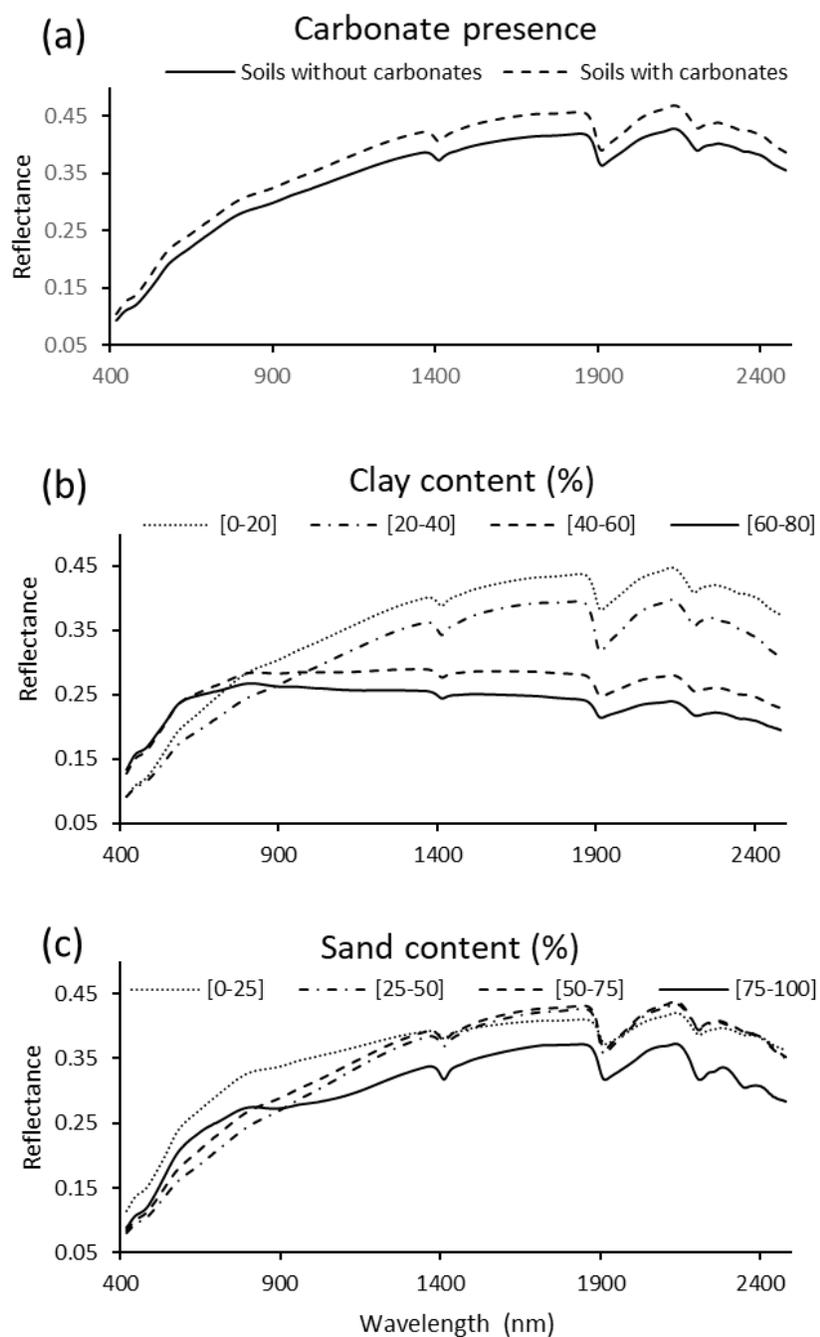


Fig. 2. (a) Mean spectra of samples with or without carbonate. Mean reflectance of all samples, computed for arbitrary (b) clay and (c) sand classes. The clay classes are 0-20 %, 20-40 %, 40-60 %, 60-80 % and the sand classes are 0-25 %, 25-50 %, 50-75 %, 75-100 %.

RI to prediction accuracy with spiking

The benefit gained from spiking varied greatly among farms, where SOC predictability showed almost no relative improvement (RI) for NY-C3 (RI = 1 %), whereas IA-N1 and NY-N2 displayed vast improvements from spiking, with RIs of 62 % and 49 %, respectively (Table 4). Additional sites that benefited substantially from spiking included NY-E1, NY-W3, NY-N1, and NY-E2, with RIs of 34 %, 30 %, 26 %, and 26 %, respectively. While spiking did improve the accuracy of SOC predictions for all farms (mean RI of 19 %), benefits were minimal for most, with 11 of the 17 sites having a RI of less than 15 %. The cost of measuring SOC by standard analytical methods (i.e. dry combustion) is far higher than prediction by VNIR, and so for most farm sites in this study, it is unlikely that the small level of accuracy gained by spiking would be justified, given the large increase in analytical costs associated with using 50 % of local samples for model calibration.

SOC and soil texture effects on RI

Of the soil properties measured, soil texture and SOC content appeared to be the main factors influencing RI among the farms, with sites characterized by textural extremes (i.e. high sand, silt, or clay content) benefiting most from spiking (Table 4). For example, NY-N2 had the highest mean clay content of any site at 41%, and IA-N1 was one of only two other sites with clay soils (i.e., clay loam and silty clay loam texture classes; Table 1). IA-N1 and NY-N2 also displayed the largest range in SOC concentration among farm sites, with maximum SOC values of 5.42 % and 9.34 %, respectively (Table 1). For the soils evaluated in this study, mean VNIR reflectance tended to decrease as clay content increased, with spectral differences most apparent when clay content was greater than 40 % (Fig 2b). Sand content appeared to have less effect on VNIR

spectra (Fig. 2c). Soil texture is a major factor controlling SOC formation and depletion in managed soils (Burke et al., 1989), and many clay minerals have distinct absorbance features in the VNIR region which can distort or mask the spectral signature of SOC (Brown et al., 2005, 2006).

Soil-spectral variability among the farms was evaluated by principal component analysis of the raw VNIR spectra for all samples, revealing unique spectral characteristics associated with IA-N1 and NY-N2 (Fig. 3). A biplot of the first two principal components shows VNIR spectra grouping into three distinct clusters clearly associated with IA-N1, NY-N2, and all other farm sites. Additionally, the IA-N1 and NY-N2 sites displayed the greatest RIs from spiking, suggesting that these soils were poorly represented in the calibration set, potentially due to their high clay and SOC content (Table 4). Silt loam soils account for the bulk of maize cropping area in the study region, and so VNIR models would likely perform well most farm sites, as high clay soils are atypical in these systems.

Farm sites characterized by high sand content and low SOC also displayed greater benefits from spiking, as was observed for NY-E2 and NY-E1 (Table 4). Farms in the NY-E region had the lowest mean SOC concentration, with soils tending to be sandier. Only the NY-N1 site had a very high mean silt content at 80 % silt, but RI from spiking was substantial. Sites characterized by textural and SOC extremes benefited most from spiking, suggesting a close linkage among a soil's texture, SOC content, and VNIR spectra. These results demonstrate the potential effect of soil texture on SOC prediction accuracy by VNIR, and highlight the importance of ensuring that soil texture conditions at a new farm site are represented by the soil samples used for VNIR model calibration.

Table 4. Summary of measured soil properties by individual farm site for soil organic C (SOC) %, Clay %, Sand %, and carbonates (CaCO₃) %. Farm-level validation statistics for predicted SOC content by VNIR models with and without local samples used in model calibration are also presented.

	Measured soil properties										predicted SOC %				
	SOC %		Clay %		Sand %		CaCO ₃ %		Site not included†	Spiking 50%‡	Spiking RI§	Conc	stock 30cm		
	n	Mean	SD	Mean	SD	Mean	SD	Mean	SD	RMSE	RMSE	Relative Error ¶	Relative Error ¶		
IA-N1	1149	1.98	1.00	17	5	42	9	1.1	4.1	0.95	0.36	62	18	12	
IA-N2	97	2.65	1.33	28	4	28	9	1.4	3.6	0.42	0.37	12	14	7	
NY-C1	59	1.17	0.88	13	1	37	3	13.8	16.2	0.46	0.39	14	34	18	
NY-C2	115	1.33	0.77	14	3	29	5	0.2	1.2	0.35	0.31	10	24	10	
NY-C3	91	1.67	0.75	11	2	36	4	7.2	14.2	0.35	0.35	1	21	11	
NY-C4	144	1.54	1.19	13	3	33	7	10.8	16.6	0.56	0.53	5	35	15	
NY-C5	253	1.92	0.88	18	3	11	7	0.1	0.3	0.33	0.28	13	15	10	
NY-E1	105	0.88	0.72	19	11	31	26	0.0	0.1	0.62	0.46	26	52	39	
NY-E2	96	0.74	0.52	5	2	65	13	0.0	0.0	0.56	0.37	34	50	20	
NY-N1	152	1.03	0.96	7	2	13	6	3.7	4.6	0.42	0.31	26	30	15	
NY-N2	305	1.72	1.73	41	17	18	18	1.0	2.9	1.03	0.52	49	30	14	
NY-W1	94	1.38	0.93	9	3	47	9	7.6	14.0	0.32	0.30	6	22	30	
NY-W2	100	1.34	0.74	8	4	48	9	8.4	14.4	0.42	0.37	10	28	15	
NY-W3	908	1.28	0.88	15	6	20	11	0.9	3.2	0.59	0.41	30	32	14	
NY-W4	99	1.1	0.73	7	4	49	10	0.0	0.0	0.30	0.26	13	23	14	
NY-W5	64	1.86	1.29	9	3	38	9	0.0	0.0	0.44	0.42	5	23	15	
NY-W6	200	1.57	0.70	11	5	42	7	1.0	3.7	0.57	0.52	9	33	16	

† 0 % of local samples from a given farm site were used for model calibration.

‡ 50 % of local samples from a given farm site were used for model calibration.

§ Relative improvement in RMSE gained by including 50 % of local samples in model calibration: $Spiking\ RI = 100 * [(Site\ not\ included) - (Spiking\ 50\ %)] * [Site\ not\ included]^{-1}$.

¶ Relative error is the farm-level SOC prediction error relative to farm-level mean measured SOC content: $Relative\ error = 100 * RMSE * (mean\ SOC)^{-1}$. Results are presented for both SOC concentration (Conc) and SOC stocks to 30 cm (Stock 30cm).

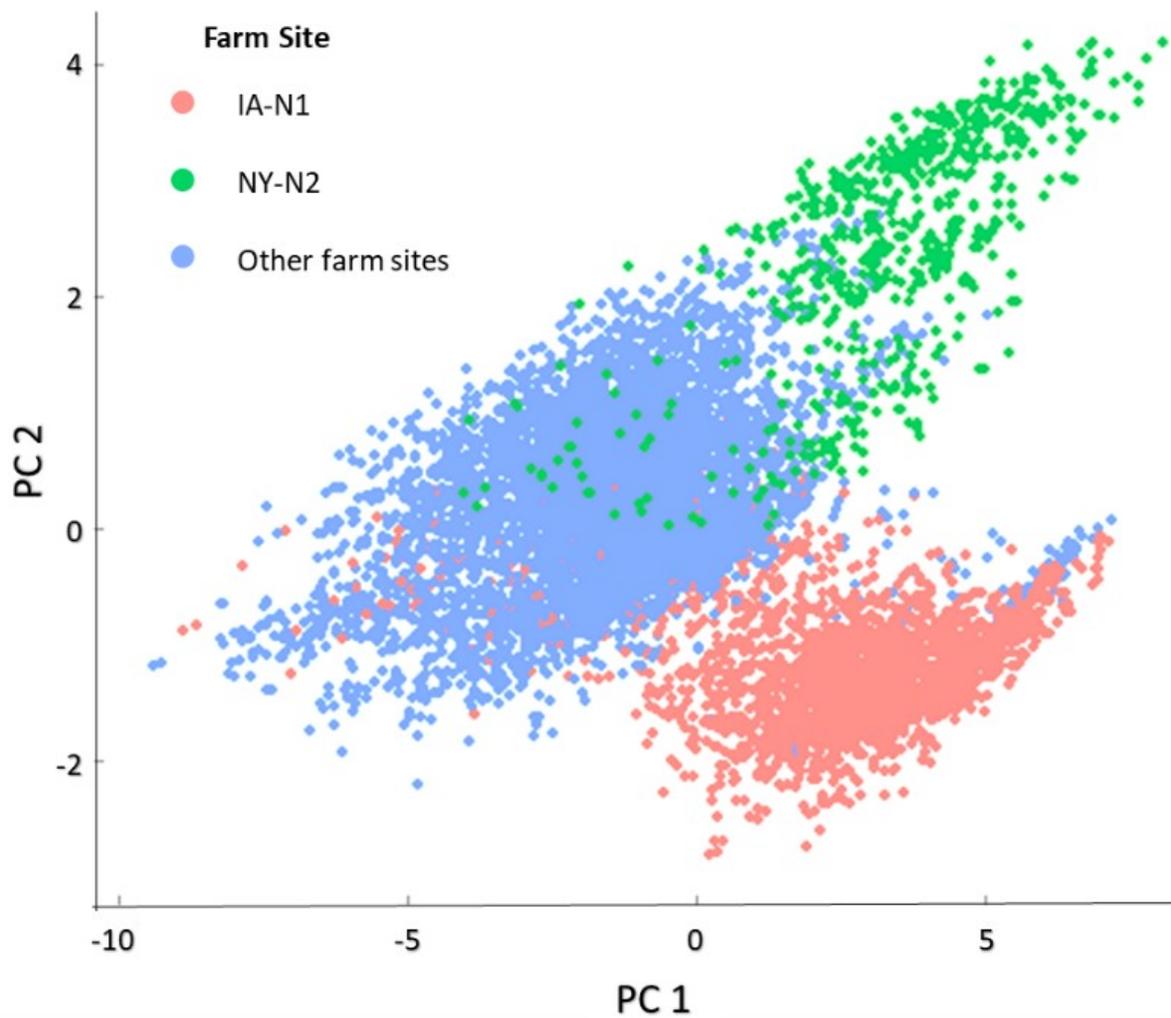


Fig 3. Biplot of principal component (PC) scores for the first two PCs, PCA based on raw spectra of all samples.

Sampling intensity effects on RI

The sampling intensity also appeared to be influencing RI from spiking. Sites with larger sample sizes and mean distance between sampling locations tended to benefit more from spiking. For instance, sample sizes at IA-N1, NY-N2, and NY-W3 were considerably larger than most other sites at 1149, 305, and 908, respectively (Table 1). Sample size alone does not explain improvements gained from spiking, as NY-C5 had a sample size of 253, yet spiking resulted in a much smaller RI of 13 %. NY-C5 had the highest sampling density of any site, with 253 soil samples from 51 cores collected over a 1 ha grid. The spatial extent of the sampling area and degree of soil spatial variability at a farm, will strongly influence the evaluation set sample statistics for that site. For instance, soils with higher clay content can have a greater potential for storing SOC and an increased “equilibrium” SOC content, which could in turn increase the potential magnitude of treatment effects (i.e. tillage, crop, organic amendments, etc.), or other environmental factors affecting SOC flux at a site (i.e. topography, aspect, soil moisture, etc.). A large sample size and sampling area increases the likelihood of capturing regions of a farm or field with extreme soil texture and SOC values. Uncommon soil conditions and especially multivariate outliers (e.g. soils with both a high SOC and sand content) can affect soil-spectral relationships characterized by an evaluation set, and thus a model’s predictive capacity at that site.

Applications for spiking

Few studies have evaluated how sample statistics of the calibration and evaluation sets affect SOC prediction accuracy, and even fewer have evaluated how the sample statistics of ancillary soil properties relate to model performance (Kuang and Mouazen, 2011). Based on our results, spiking is suggested for sites with clay soils, or for sites expected to have a large range in SOC

concentration. Long-term effects of land use can affect both SOM quality and content, and so additional priority should be placed on spiking when sites are characterized by stark differences in historic land use. This was the case for NY-N2, which had the highest mean clay content, increases in SOC content due to tillage effects were greater than for any other sites (data not shown), and SOC predicted without spiking resulted in the lowest accuracy observed for any model (RMSE = 1.03). NY-N2 displayed high spatial variability in inherent soil properties, and a long history of tillage effects has changed SOC content and likely SOM quality. Interactions among these factors alter soil-spectral relationships, reduce prediction accuracy, and decrease chances of detecting differences among tillage treatments (i.e. SOC change). NY-N2 is a university experimental research farm, where long-term tillage experiments have been in place for over 20 years, and provides an example of where benefits from spiking were maximized.

SOC concentration prediction accuracy

The accuracy of SOC predictions at individual farm sites varied considerably among the farms, with the RMSE ranging from 0.26 % for NY-W4 to 0.53 % for NY-C4 (Table 4). Some sites displayed a low RMSE and large RI from spiking, but when viewed in the context of the farm's mean SOC content, RMSE was quite high (Fig. 4a). This trend was most apparent for the NY-E1 and NY-E2 farms, where because of their sandy textured soils, mean SOC content was low and SOC predictions improved considerably from spiking, but errors were very large relative to the site mean. To account for this discrepancy, model performance was evaluated based on the relative error (RE), where RMSE was expressed as the percentage error relative to the mean (Table 4). Geospatial trends in the RMSE and SOC levels among farm sites were observed, with sites located within the same geographic region tending to be more similar. For example, the

lowest SOC levels were observed for sites within the eastern NY region, while sites in northern IA had higher SOC levels. For many of the farm sites, soil texture and management history appeared to be the primary drivers of SOC variation. The NY-N1 and NY-N2 sites are located only 56 km from each other, both are university research farms with long-term maize tillage experiments in place, and yet their mean and maximum SOC content were vastly different (NY-N1: mean = 1.03, max = 3.51; NY-N2: mean = 1.72, max = 9.34). Soils at the NY-N1 farm were predominantly silt, while NY-N2 had clay soils and samples were also collected from fallow regions of the farm to estimate SOC storage potential. These stark contrasts in soil texture and management history are likely driving farm-level differences in SOC content and RMSE observed for NY-N1 and NY-N2. Farms where substantial amounts of soil carbonates were present tended to have higher RMSEs than nearby sites without carbonates but with similar SOC levels (Table 4). Soil carbonates are likely influencing SOC prediction errors at the NY-C1 and NY-C4 farms, particularly for subsoils where carbonates were greatest.

Quantifying farm-level SOC stock differences by VNIR

Farm-level SOC stocks were highly variable, with mean SOC stocks to 30 cm for individual farms ranging from 37.7 Mg ha⁻¹ for NY-E-2, to 104.6 Mg ha⁻¹ for IA-N2. Prediction errors for SOC stocks ranged from an RMSE of 5.5 Mg ha⁻¹ (NY-C2) to 17.8 Mg ha⁻¹ (NY-E1), with an average RMSE of 9.2 Mg ha⁻¹ (Fig. 4b). The RMSE is the most common way to measure SOC prediction accuracy, but provides little information without a frame of reference for the relative magnitude of errors. If VNIR is to be used as a tool for farm-level SOC accounting, then SOC predictions must be accurate enough to detect SOC stock differences between farms. Differences

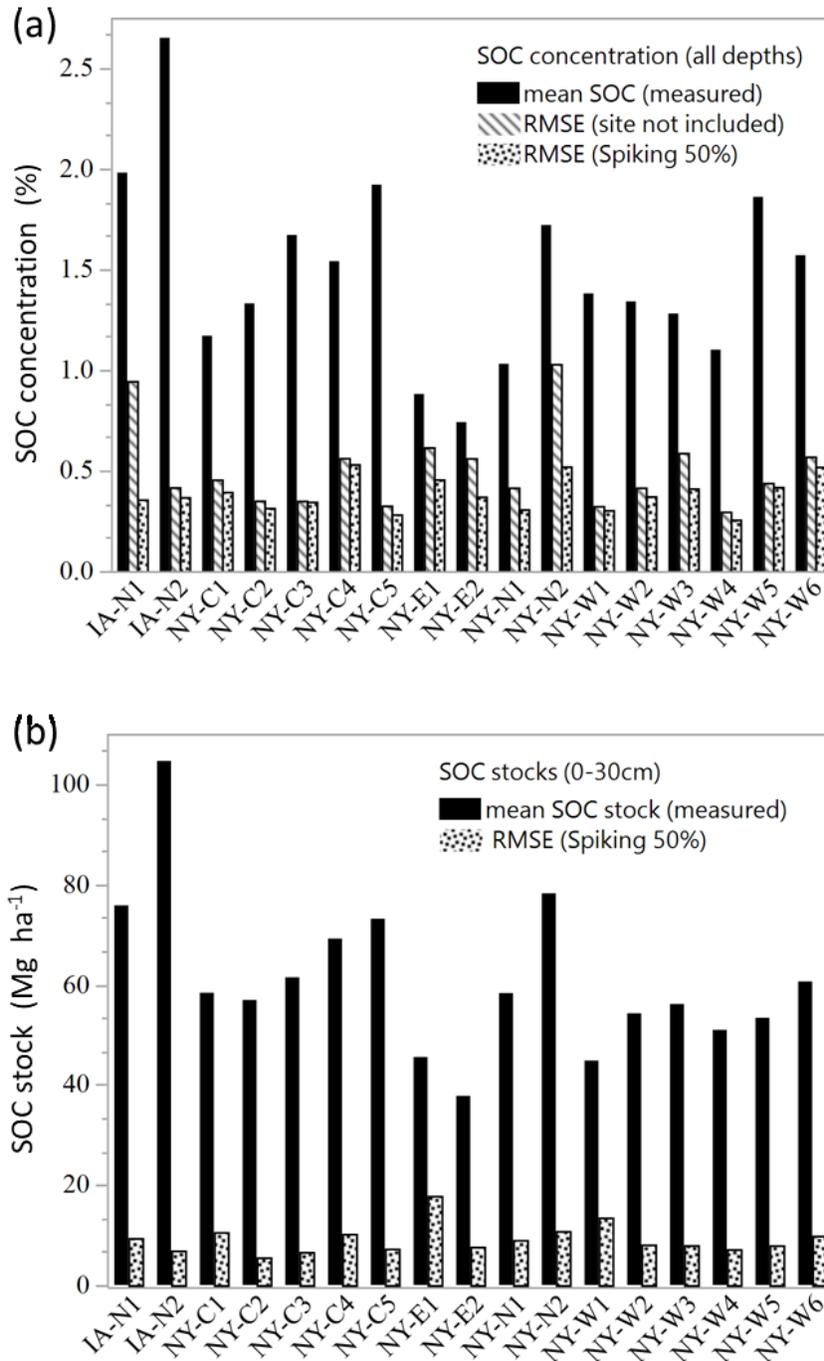


Fig. 4. (a) Mean SOC concentration for each farm site, and SOC root mean square error of prediction (RMSE) for models developed with (Spiking 50%) and without (site not included) local samples used in calibration. (b) Mean SOC stock to 30 cm for each farm site, and RMSE of SOC stocks predicted by the Spiking 50% model.

in mean SOC stocks among farm sites ranged from 0 Mg ha⁻¹ to 67 Mg ha⁻¹, with a mean difference of 17 Mg ha⁻¹ (data not shown). Differences in mean SOC stocks at the farm-level often far exceeded SOC prediction errors for the individual farms, and for most farm sites VNIR effectively predicted SOC (Fig. 4b).

Quantifying field-level SOC stock differences by VNIR

Expressing VNIR predictions as SOC stocks to a depth of 30 cm reduced RE at all farm sites except NY-W1, with RE ranging from 7 % for IA-N2 to 39 % for NY-E1 (Table 4). While prediction accuracy at sites with larger REs was adequate for detecting SOC differences at the farm level, such high errors are likely insufficient for quantifying management effects or field-level differences across the farm. As was previously noted, this was the case for the NY-E1 and NY-E2 farms, where prediction errors were high relative to the range of measured SOC values observed at the site. The potential magnitude for SOC change due to management would also be small for these sites, and therefore difficult to detect given the relatively large prediction errors. There is no straightforward approach for relating RMSE to the minimum detectable difference, but when compared to the expected differences or SOC variation at a site, RMSE provides essential information for evaluating VNIR's performance at an individual farm.

Long term field experiments with treatments comparing no-till (NT) and moldboard plowing (MP) were conducted at three of the sites, allowing further evaluation of VNIR's performance for farm-level SOC accounting. Treatment differences provide an estimate for potential SOC change over time in response to management, and were not evaluated for statistical significance. Experimental treatment differences in mean SOC stocks at the three sites

were NY-C1 (2.9 Mg ha⁻¹), NY-N1 (16.6 Mg ha⁻¹), and NY-N2 (22.0 Mg ha⁻¹), while the RMSE for these sites were 10.6 Mg ha⁻¹, 9.0 Mg ha⁻¹, and 10.8 Mg ha⁻¹, respectively. For the NY-N2 farm site, in addition to sampling experimental plots, 23 soil cores to a depth of 75 cm were collected from uncultivated regions (UR) of the farm with a long history (>20 yrs) in sod; there was a 28 Mg ha⁻¹ difference in mean SOC stocks between UR and CT. At the NY-C1 site, VNIR was unable to quantify field-level differences in SOC stocks, as tillage had a small effect on SOC, and RMSE was three times greater than differences in mean SOC stocks. At the NY-N1 and NY-N2 sites, field level differences in mean SOC stocks were much greater, and could be accurately quantified by VNIR-predicted SOC. In this study, VNIR's predictive capacity varied by site, where the ability to detect field-level treatment differences was primarily dependent on the extent of SOC variation at a site.

Quantifying within-field SOC stock variation by VNIR

The accuracy of VNIR predictions were evaluated at the within-field scale for the intensive sites NY-C5, NY-W3, and IA-N1, where sampling was conducted at an intensity of 51 cores 1 ha⁻¹, 188 cores 20 ha⁻¹, and 232 cores 25 ha⁻¹, respectively. The standard deviation of the measured SOC values was used to estimate within-field spatial variability for the intensive sites. For NY-W3 and IA-N1, prediction errors were much lower than the inherent SOC variability of the site, with VNIR resulting in an acceptable level of accuracy for quantifying within-field SOC spatial variability (IA-N1: RMSE=9.4, SD=18.6), (NY-W3: RMSE=8.0, SD=14.4). The NY-C5 farm site had a lower relative error than other intensive sites, and VNIR accurately quantified mean SOC-stocks, but because of the low degree of variability at this spatial scale i.e. (1 ha), VNIR was not able to quantify within-field SOC variability (NY-C5: RMSE=7.3, SD=8.1).

Conclusions

Stratification by depth substantially improved model performance for all depth increments, with further improvements gained by adding an additional level of stratification for presence of carbonates at depths below 30 cm. Depth stratification was only evaluated by individual depth increment and for all depths combined, representing the two extremes. Stratification regroups the data into smaller sample sets, but each added grouping demands additional time and resources for data management, preprocessing, and predictive modeling, all of which must be taken into account when choosing a calibration strategy.

The spiking approach of including local samples can greatly improve predictions by VNIR models at previously unsampled farm sites. Improvements to model performance gained from spiking will be greatest for sites with unique spectral characteristics that are poorly represented by calibrations. In this study, these include sites characterized by atypical SOC concentration, clay content, or texture class. If the low-cost advantage of VNIR spectroscopy is to be maintained, the benefits of including additional local samples must be weighed against the additional costs. For many farm sites, accurate SOC predictions were obtained for models without local samples included in calibration (Fig. 4a). For these sites spiking provided little benefit, and would not justify the additional cost.

Based on results obtained from 17 farm sites representing the major growing conditions for maize production systems in the study region; VNIR successfully predicted SOC, allowing for accurate SOC accounting at spatial scales ranging from within-field to among farms. Even at the within-field scale, surface soils displayed very high SOC spatial variation, with SOC concentration at D10 ranging by as much as 4.5 % for a single field. The high measurement

sensitivity of standard analytical methods can't compensate for a sampling intensity which is inadequate for capturing SOC spatial variability within a management unit. VNIR-predicted SOC is less accurate than analytical measurement, but its cost-savings can allow for more thorough sampling and characterization of SOC spatial variability. The benefits of improved method accuracy and sensitivity must be carefully weighed against that of increased sample size and spatial coverage especially when farm sites or fields display high SOC spatial variability (i.e. measured SOC SD >> RMSE).

A farm-level evaluation approach, where SOC predictions are evaluated for models developed both with and without local samples, should be employed whenever possible. Comparing VNIR models across multiple independent farms provides a better representation of real world conditions, facilitates more robust findings, and will allow new insights to be gained. For future studies, we recommend further investigation of the relationship between the evaluation set sample statistics, and the relative performance of VNIR models evaluated across spatial scales.

REFERENCES

- Bellon-Maurel, V., E. Fernandez-Ahumada, B. Palagos, J.M. Roger, and A. McBratney. 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC - Trends Anal. Chem.* 29(9): 1073–1081. doi: 10.1016/j.trac.2010.05.006.
- Bellon-Maurel, V., and A. McBratney. 2011. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. *Soil Biol. Biochem.* 43(7): 1398–1410. doi: 10.1016/j.soilbio.2011.02.019.

- Bilgili, V.A., H.M. van Es, F. Akbas, A. Durak, and W.D. Hively. 2010. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *J. Arid Environ.* 74(2): 229–238. doi: 10.1016/j.jaridenv.2009.08.011.
- Brown, D.J. 2007. Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. *Geoderma* 140: 444–453. doi: 10.1016/j.geoderma.2007.04.021.
- Brown, D.J., R.S. Bricklemyer, and P.R. Miller. 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. *Geoderma* 129(3–4): 251–267. doi: <http://dx.doi.org/10.1016/j.geoderma.2005.01.001>.
- Brown, D.J., K.D. Shepherd, M.G. Walsh, M. Dewayne Mays, and T.G. Reinsch. 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma* 132(3–4): 273–290. doi: 10.1016/j.geoderma.2005.04.025.
- Burke, I.C., C.M. Yonker, W.J. Parton, C. V Cole, D.S. Schimel, and K. Flach. 1989. Texture, climate, and cultivation effects on soil organic matter content in US grassland soils. *Soil Sci. Soc. Am. J.* 53(3): 800–805.
- Chang, C.W., D. A. Laird, M.J. Mausbach, and C.R. Hurburgh. 2001. Near-infrared reflectance spectroscopy–Principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* 65(2): 480. doi: 10.2136/sssaj2001.652480x.
- Cohen, M.J., J.P. Prenger, and W.F. DeBusk. 2005. Visible-near infrared reflectance spectroscopy for rapid, nondestructive assessment of wetland soil quality. *J. Environ. Qual.* 34(4): 1422–1434. doi: 10.2134/jeq2004.0353.

Conant, R.T., and K. Paustian. 2002. Spatial variability of soil organic carbon in grasslands: Implications for detecting change at different scales. *Environ. Pollut.* 116: 127–135. doi: 10.1016/S0269-7491(01)00265-2.

Cozzolino, D., and a. Morón. 2006. Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil organic carbon fractions. *Soil Tillage Res.* 85(1–2): 78–85. doi: 10.1016/j.still.2004.12.006.

Gogé, F., R. Joffre, C. Jolivet, I. Ross, and L. Ranjard. 2012. Optimization criteria in sample selection step of local regression for quantitative analysis of large soil NIRS database. *Chemom. Intell. Lab. Syst.* 110(1): 168–176. doi: 10.1016/j.chemolab.2011.11.003.

Gogé, F., C. Gomez, C. Jolivet, and R. Joffre. 2014. Which strategy is best to predict soil properties of a local site from a national Vis-NIR database? *Geoderma* 213: 1–9. doi: 10.1016/j.geoderma.2013.07.016.

Griscom, B.W., J. Adams, P.W. Ellis, R.A. Houghton, G. Lomax, D.A. Miteva, W.H. Schlesinger, D. Shoch, J. V. Siikamäki, P. Smith, P. Woodbury, C. Zganjar, A. Blackman, J. Campari, R.T. Conant, C. Delgado, P. Elias, T. Gopalakrishna, M.R. Hamsik, M. Herrero, J. Kiesecker, E. Landis, L. Laestadius, S.M. Leavitt, S. Minnemeyer, S. Polasky, P. Potapov, F.E. Putz, J. Sanderman, M. Silvius, E. Wollenberg, and J. Fargione. 2017. Natural climate solutions. *Proc. Natl. Acad. Sci.* 114(44): 11645–11650. doi: 10.1073/pnas.1710465114.

de Gruijter, J.J., A.B. McBratney, B. Minasny, I. Wheeler, B.P. Malone, and U. Stockmann. 2016. Farm-scale soil carbon auditing. *Geoderma* 265: 120–130. doi: 10.1016/j.geoderma.2015.11.010.

Guerrero, C., B. Stenberg, J. Wetterlind, R. a. Viscarra Rossel, F.T. Maestre, a. M. Mouazen, R. Zornoza, J.D. Ruiz-Sinoga, and B. Kuang. 2014. Assessment of soil organic carbon at local scale with spiked NIR calibrations: Effects of selection and extra-weighting on the spiking subset. *Eur. J. Soil Sci.* 65(2): 248–263. doi: 10.1111/ejss.12129.

Guerrero, C., J. Wetterlind, B. Stenberg, A.M. Mouazen, M.A. Gabarrón-Galeote, J.D. Ruiz-Sinoga, R. Zornoza, and R.A. Viscarra Rossel. 2015. Do we really need large spectral libraries for local scale SOC assessment with NIR spectroscopy? *Soil Tillage Res.* doi: 10.1016/j.still.2015.07.008.

Holmes, K.W., A. Wherrett, A. Keating, and D. V Murphy. 2012. Meeting bulk density sampling requirements efficiently to estimate soil carbon stocks. *Soil Res.* 49(8): 680–695. <http://dx.doi.org/10.1071/SR11161>.

Igné, B., J.B. Reeves, G. McCarty, W.D. Hively, E. Lundc, and C.R. Hurburgh. 2010. Evaluation of spectral pretreatments, partial least squares, least squares support vector machines and locally weighted regression for quantitative spectroscopic analysis of soils. *J. Near Infrared Spectrosc.* 18(3): 167–176. doi: 10.1255/jnirs.883.

Kettler, T.A., J.W. Doran, and T.L. Gilbert. 2001. Simplified method for soil particle-size determination to accompany soil-quality analyses. *Soil Sci. Soc. Am. J.* 65: 849–852. doi: 10.2136/sssaj2001.653849x.

Kuang, B., and a. M. Mouazen. 2011. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms. *Eur. J. Soil Sci.* 62(August): 629–636. doi: 10.1111/j.1365-2389.2011.01358.x.

- Lee, K.S., D.H. Lee, K. a Sudduth, S.O. Chung, N.R. Kitchen, and S.T. Drummond. 2009. Wavelength identification and diffuse reflectance estimation for surface and profile soil properties. *Trans. ASABE*. 52(3): 683–695.
- McCarty, G.W., J.B. Reeves, V.B. Reeves, R.F. Follett, and J.M. Kimble. 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Sci. Soc. Am. J.* 66(2): 640. doi: 10.2136/sssaj2002.0640.
- Morgan, C.L.S., T.H. Waiser, D.J. Brown, and C.T. Hallmark. 2009. Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy. *Geoderma* 151(3–4): 249–256. doi: 10.1016/j.geoderma.2009.04.010.
- Nocita, M., a Stevens, B. van Wesemael, M. Aitkenhead, M. Bachmann, B. Barthès, E. Ben Dor, D.J. Brown, M. Clairotte, a Csorba, P. Dardenne, J. a. M. Demattê, V. Genot, C. Guerrero, M. Knadel, L. Montanarella, C. Noon, L. Ramirez-Lopez, J. Robertson, H. Sakai, J.M. Soriano-Disla, K.D. Shepherd, B. Stenberg, E.K. Towett, R. Vargas, and J. Wetterlind. 2015. Soil spectroscopy: An alternative to wet chemistry for soil monitoring. : 139–159. doi: 10.1016/bs.agron.2015.02.002.
- Paustian, K., J. Lehmann, S. Ogle, D. Reay, G.P. Robertson, and P. Smith. 2016. Climate-smart soils. *Nature* 532(7597): 49–57. doi: 10.1038/nature17174.
- Price, J.C. 1994. How unique are spectral signatures? *Remote Sens. Environ.* 49(3): 181–186. doi: 10.1016/0034-4257(94)90013-2.
- 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(1–2): 3–14. doi: 10.1016/j.geoderma.2009.04.005.

Reeves, J.B., and D.B. Smith. 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. *Appl. Geochemistry* 24(8): 1472–1481. doi: 10.1016/j.apgeochem.2009.04.017.

Sherpa, S.R., D.W. Wolfe, and H.M. van Es. 2016. Sampling and data analysis optimization for estimating soil organic carbon stocks in agroecosystems. *Soil Sci. Soc. Am. J.* 80(5): 1377. doi: 10.2136/sssaj2016.04.0113.

Sherrod, L.A., G. Dunn, G.A. Peterson, and R.L. Kolberg. 2002. Inorganic carbon analysis by modified pressure-calimeter method. *Soil Sci. Soc. Am. J.* 66(1): 299–305.

Stenberg, B., R.A. Viscarra Rossel, A.M. Mouazen, and J. Wetterlind. 2010. Visible and near infrared spectroscopy in soil science. *Adv. Agron.* 107:163–215.

doi:10.1016/S0065-2113(10)07005-7

Terhoeven-Urselmans, T., T.-G. Vagen, O. Spaargaren, and K.D. Shepherd. 2010. Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library. *Soil Sci. Soc. Am. J.* 74(5): 1792. doi: 10.2136/sssaj2009.0218.

USDA-NASS. 2011. National agricultural statistics service. <http://www.nass.usda.gov/>

Vasques, G.M., S. Grunwald, and W.G. Harris. 2010. Spectroscopic models of soil organic carbon in Florida, USA. *J. Environ. Qual.* 39(3): 923–934. doi: 10.2134/jeq2009.0314.

Vasques, G.M., S. Grunwald, and J.O. Sickman. 2009. Modeling of soil organic carbon fractions using visible–near-infrared spectroscopy. *Soil Sci. Soc. Am. J.* 73(1): 176. doi: 10.2136/sssaj2008.0015.

Viscarra Rossel, R. A., D.J. Walvoort, A. B. McBratney, L.J. Janik, and J.O. Skjemstad. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131(1–2): 59–75. doi: 10.1016/j.geoderma.2005.03.007.

Wight, J.P., A.J. Ashworth, and F.L. Allen. 2016. Organic substrate, clay type, texture, and water influence on NIR carbon measurements. *Geoderma* 261: 36–43. doi: 10.1016/j.geoderma.2015.06.021.

Wolfe, D.W. 2013. Climate Change Solutions from the Agronomy Perspective. pp. 11–29. *In* Hillel D. and C. Rosenzweig (eds). *Handbook of Climate Change and Agroecosystems: Global and Regional Aspects and Implications*. Imperial College Press. London.

Zomer, R.J., D.A. Bossio, R. Sommer, and L. V Verchot. 2017. Global sequestration potential of increased organic carbon in cropland soils. *Sci. Rep.* 7(1): 15554. doi: 10.1038/s41598-017-15794-8.

CHAPTER 2

MOVING BEYOND THE BLACK BOX WITH MID INFRARED SPECTROSCOPY AND PREDICTION OF SOIL QUALITY INDICATORS

Abstract

Effective monitoring of soil management strategies to improve soil quality (SQ) will be essential for meeting expanding food demand. Mid infrared spectroscopy (MIR) shows promise as an accurate and low cost option for predicting multiple SQ indicators, yet there is still considerable uncertainty regarding the influence of geographic and taxonomic coverage on MIR model calibration performance, and a limited understanding of the mechanisms underlying these empirical prediction models. We evaluated MIR's potential for predicting 12 common SQ indicators: total nitrogen, total carbon, carbonates (CO_3), bulk density, sand, silt, and clay content, cation exchange capacity (CEC), pH, water retention -33 kPa (WR 0.33), water retention -200 kPa (WR 2), and water retention -1500 kPa (WR 15) by partial least squares regression (PLSR) for Mollisols soils of the US. Additionally, we developed a conceptual framework for evaluating the soil-spectral mechanisms underlying PLSR prediction models. Soil spectra and reference data for 5712 soil samples were obtained from a publicly available United States Department of Agriculture-Natural Resources Conservation Service (USDA-NRCS) database, with soils encompassing 418 different soil series from the Mollisol order, 26 US states, and 243 US counties. The best models in terms of root mean square error (RMSE) were as follows: CO_3 (20 g kg^{-1}), total C (4.0 g kg^{-1}), CEC ($2.5 \text{ cmol}_c \text{ kg}^{-1}$), clay (3.9 %), sand (7.3 %), and silt (8.0 %). The poorest models were bulk density (0.15 g cm^{-3}), WR 0.33 (5.1 %), pH (0.5), WR 15 (3.1 %), and WR 2 (3.9 %). Multivariate analysis of the measured soil properties, prediction

accuracies, and PLSR β -coefficients revealed considerable overlap in the spectral regions utilized by the different PLSR models. Only the models for CO₃, total C, clay, and sand appear to rely heavily on their unique spectral features. As matric potential became more negative, models for predicting water retention relied less on spectral features associated with total C and silt and relied more on features associated with clay. The observed trends were in agreement with fundamental soil-spectral theory, demonstrating the proposed conceptual framework's potential for evaluating the underlying soil-spectral mechanisms of MIR PLSR models. This new insight into how MIR prediction models utilize spectral information, and demonstration that multiple SQ indicators can be accurately predicted for Mollisol soils of the US, addresses some of the common concerns preventing more widespread use of MIR analysis for SQ indicators.

Introduction

Improving soil quality (SQ) will be a key step towards increasing the resiliency of food systems, thereby reducing risks associated with climate change and extreme weather events. Although switching to management practices that build SQ can involve some fixed and variable costs, over the long term input costs often go down, and high SQ soils have more resilience to drought and flooding, and less soil erosion and chemical runoff into waterways during heavy rainfall events (de Gruijter et al., 2016; Wolfe, 2019; Wolfe, 2013). It is little surprise that the scientific community, governments, and even the business community have all turned their attention towards improving SQ as a potential strategy for addressing global food security and economic stability in the face of climate change. Soil quality encompasses a more holistic view of the soil system, combining soil biological, physical and chemical components to assess a soil's functional capacity to provide ecosystem services. This requires a more comprehensive

suite of measurements, and therefore increases analytical costs. Excessive analytical costs currently limit our ability to establish SQ baseline levels and monitor changes in response to management (Grunwald et al., 2011; Sherpa et al., 2016). Reducing analytical costs will allow for larger sample sizes, increased spatial resolution of management, and a more widespread adoption of soil conservation practices.

Mid-infrared diffuse reflectance spectroscopy (MIR) is inexpensive, high throughput, safe, nondestructive, and allows for the simultaneous estimation of multiple soil properties (Gobrecht et al., 2014; Reeves III, 2012; Soriano-Disla et al., 2014). The bonds of organic and mineral components of soil absorb energy in the mid-infrared region, and MIR captures the spectral signature associated with these bonds. There is still a great deal of uncertainty regarding which soil properties can be effectively predicted and how best to develop calibration models (Stenberg and Viscarra Rossel, 2010). The effectiveness of this approach is dependent on the vibrational modes associated with a given soil property, and so method accuracy will vary for different soil properties and calibration sets. The taxonomic diversity and geographic coverage of calibration and validation sets have a strong influence on the performance of MIR prediction models, as taxonomic and geographic similarity often translates to the chemical composition and spectral signature of soils (Savvides et al., 2010). Calibration sets and therefore MIR analysis of soils encompassing broader taxonomic or geographic coverage tend to be less accurate. The goal is to develop the broadest models possible, while maintaining an acceptable level of error for a given application.

Future advancements to the field of soil spectroscopy will likely require the scanning of archived soils and utilization of preexisting soil attribute and spectral databases. Large soil-spectral databases are already available for public use in several regions of the world. The United

States Department of Agriculture-Natural Resources Conservation Service (USDA-NRCS) currently curates one of the World's largest MIR spectral libraries and soils databases, with the Kellogg Soil Survey Laboratory (KSSL) Soil Characterization database containing analytical data for over 300,000 soil samples and the associated MIR spectra for over 60,000 full profile samples. The KSSL carries out the widest range of soil analytical methods of any facility worldwide, and houses a full time permanent staff of research scientists and technicians to help ensure continuity in trained personnel. These data meet high standards in terms of quality assurance, contain soils representing all of the major US soil taxonomic groupings, reference data for a range of soil chemical and physical properties, and are available to the public at no charge.

The complexity of correlations among spectral features and soil components often necessitates spectral preprocessing and sophisticated multivariate methods to infer relationships between soil properties and spectra. Partial least squares regression (PLSR) has been the most widely used approach in previous spectroscopy studies. While PLSR is an invaluable tool for revealing hidden relationships existing within MIR spectra, the complexity of this approach makes model interpretation inherently challenging. One of the main criticisms of MIR applications for soil analyses has been the convention towards a "black box" approach to modeling soil properties, where calibration models are seldom checked against fundamental IR principles, and the mechanistic underpinnings of models often ignored (Brown et al., 2006; Wight et al., 2016). Each PLSR model (MIR-PLSR) utilizes a unique combination of spectral regions, where the weighting of individual regions is selected during model calibration. Partial least squares regression utilizes spectral information indiscriminately, that is MIR-PLSR models may also incorporate indirect spectral features, or spectral information associated with correlated properties. In this way, soil

properties with weak or even no direct absorption features in the mid-IR region may be successfully modeled and predicted (Chang et al., 2001). Soil organic matter, carbonates, clay minerals, silica, and water all have strong absorption bands in the mid-IR region (Reeves III, 2012), and so MIR-PLSR models for total C, total N, CO₃, sand, silt, and clay likely rely on direct spectral features. Conversely, those soil properties which are operationally defined (i.e. water retention, CEC, bulk density, pH) will invariably rely on the indirect spectral features of correlated soil properties with prominent absorption bands. Chang et al. (2001), explored this phenomenon for 33 SQ indicators predicted by near infrared spectroscopy (NIR) from 802 soils, comparing correlations among measured values, and predicting secondary properties by multiple linear regression (MLR) using the primary properties as independent variables. Although this work was limited in terms of soil types, in general, when secondary properties predicted by MLR displayed high R^2 values, high R^2 values were also obtained for corresponding NIR-PLSR models. For example, CEC was predicted by MLR using moisture, total C, particle-size distribution, and macro-aggregation as predictors, resulting in R^2 values of 0.89 and 0.81 for MLR and NIR-PLSR models, respectively. Some secondary properties could not be explained by their covariance with primary properties, yet could be accurately predicted by NIR-PLSR. The authors attributed this discrepancy to the potential shortcomings of MLR and/or the spectral influence of other unmeasured soil properties.

Since the influential work by Chang *et al.* (2001), the mechanistic interpretation of NIR/MIR predictive models has largely focused on evaluating the PLSR beta coefficients (β -coefficients) (Baldock et al., 2013; Janik et al., 2007b; Terhoeven-Urselmans et al., 2010). Partial least squares regression models assign a weighting to each wave number in the form of β -coefficients, where the magnitude of the β -coefficients for a given spectral region reflects its contribution to

the PLSR model (Baldock *et al.*, 2013). Thus the β -coefficient spectra provide a graphical representation of model function, a window into the black box, allowing for the identification of key MIR spectral features and the assessment of their relative contribution to MIR-PLSR models. Baldock *et al.* 2013, developed individual MIR-PLSR models for predicting the content of C and its component fractions in Australian soils, evaluating correlations in the β -coefficients to assess the level of independence of the MIR-PLSR models. Additionally, positive and negative peaks within the β -coefficient spectra were identified, and peaks were checked for consistency with known absorption bands. While mechanistic analyses are becoming more common in the soil spectroscopy literature, there is still much to be known regarding how MIR prediction models utilize spectral information for a given soil property and calibration set. Key questions for a given model include: what spectral regions are most influential; what soil properties/compounds are associated with those peaks; and whether the model reflects our conceptual understanding of soil science and IR fundamentals. The inherent challenge of deciphering soil-spectral relationships is the primary reason that MIR is not widely used in soil analysis (Stenberg and Viscarra Rossel, 2010). Improved understanding about how the chemical composition of soils influence their spectral features, and thus prediction errors, is vital for developing optimized calibration models which are reliable yet capable of handling a diverse range of unknown materials (Gobrecht *et al.*, 2014).

The overall objective of this study was to evaluate the feasibility of using a publicly available national scale soils database and spectral library to develop MIR prediction models for 12 common SQ indicators. The Mollisols soil Order was chosen because it represents a broad taxonomic and geographic range, is well represented in the KSSL database, is the predominant soil order of the U.S. covering 21 % of the country (200 million ha), accounts for about 916

million ha worldwide, and is one of the most productive agricultural soils in the world making it of global importance in terms of food security and economic stability (Liu et al., 2012) (Fig. 1). A second objective of this study was to develop a conceptual framework for interpreting the underlying soil-spectral mechanisms of MIR-PLSR models, so as to gain insight on the complex network of correlations among measured soil properties, PLSR β -coefficients, and spectral features, and to evaluate if prediction models conform to soil science and IR spectroscopy fundamental theory. The overarching goal of this approach was to move from correlation to causation, that is from a purely statistical to a more mechanistic understanding of how MIR-PLSR models function.

The specific objectives were to:

1. Evaluate the performance of MIR as a tool for predicting twelve common soil quality indicators for Mollisol soils of the US.
2. Determine if MIR-PLSR models of primary soil properties are independent of each other, and check for structure/hierarchy in model dependence among primary and secondary soil properties.
3. Evaluate how primary soil properties contribute to the spectral features associated with secondary soil properties, by comparing correlations among MIR-PLSR β -coefficients, correlations among measured values, and RPD values.

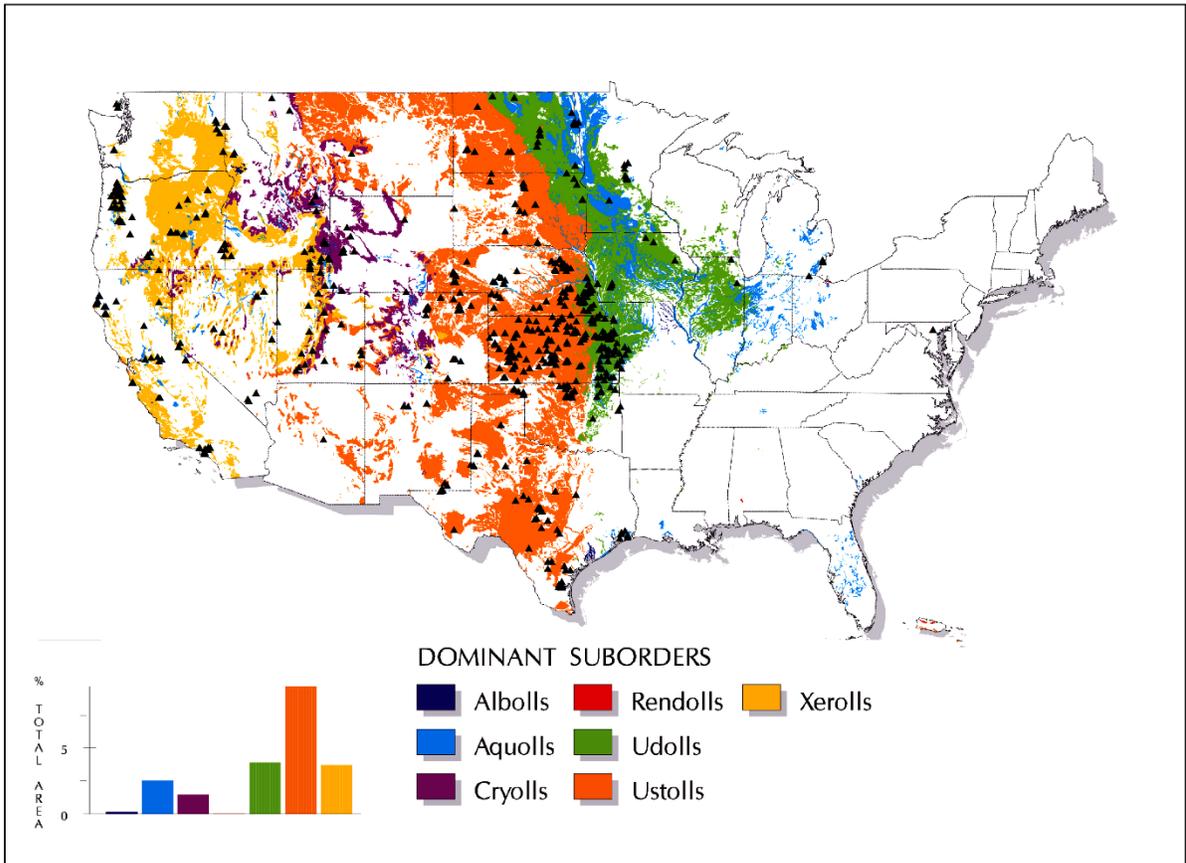


Fig. 1. Map of the dominant soil suborders of the Mollisol order in the United States (USDA-NRCS). ▲ Indicates the location of georeferenced soil samples in the full data set.

Methods

2.1 MIR spectral library and soil attribute database

Soil wet chemistry data and MIR spectra were obtained from the USDA-NRCS Kellogg Soil Survey Laboratory (KSSL) in Lincoln, NE. Soils were sampled by horizon, and generally all horizons within a pedon were collected. Soil wet chemistry and MIR analyses were performed at KSSL according to the KSSL Methods Manual, Soil Survey Investigations Report (SSIR) No. 42 (Soil Survey Staff. 2014). KSSL data were obtained for the following thirteen soil analyses, with the KSSL Method Codes provided in parentheses (Soil Survey Staff. 2014): total carbon and nitrogen by dry combustion (4H2a), carbonates (CO₃) by electronic manometer (4E1a1a1a1), cation exchange capacity (CEC) by ammonium acetate displacement (4B1a1a), pH by 1:2 calcium chloride (4C1a2a2), bulk density by saran-coated clods (3B1), water retention by pressure-plate extraction at -33 kPa (WR 0.33) (3C1c), -200 kPa (WR 2) (3C1e1a), and -1500 kPa (WR 15) (3C2a1a), total sand, silt, and clay content by pipette (3A1a1a), and MIR diffuse reflectance spectroscopy (7A7).

2.2. Soil samples and measured properties

The soils included in this study were collected from throughout the United States, and captured the typical range of geographic and taxonomic diversity observed in US Mollisols (Fig 1.), with horizons often sampled to a depth of 2 m or greater. Surface and subsurface soil processes are intimately linked, and so while including subsurface soils in MIR model calibration and validation adds additional challenges related to normality, it greatly increases the practical applications of this approach. The mineralogy and profile morphology of Mollisols are extremely diverse (Allen and Fanning, 1983).

While there were a total of 5712 soils in the full MIR spectral library used in this study,

individual soil properties displayed a range of sample sizes. The KSSL primarily supports the efforts of the NRCS and the National Cooperative Soil Survey, and so data availability reflects the needs associated with various soil conservation and survey projects. Most soil properties had a sample size over 5200, but CO₃, WR 0.33, WR 2, and bulk density had sample sizes of 2803, 2721, 2408, and 2721, respectively. However, CO₃ was only measured for those soils displaying a positive effervescence upon exposure to dilute HCl (fizz test), yet the fizz test was performed on all soils (data not shown). Soils without a CO₃ measurement are assumed to have a value of 0 %, and while these data were excluded from calibration and validation sets, it is worth noting that soil organic carbon content can be calculated by difference for all total C measurements.

2.3. Spectroscopic measurements

Soil spectra were collected in the laboratory from air-dried and ground samples using a Bruker Vertex 70 FT-IR MIR spectrometer (Bruker Optik GmbH, Germany). The instrument was equipped with a broadband KBr beam splitter, liquid N₂-cooled mercury–cadmium telluride (MCT) detector, and an automated high throughput microplate reader attachment (Bruker HTS-XT). Approximately 20 mg of soil was loaded into each well of a 96 well aluminum microplate. Each soil sample was added to four different wells resulting in four replicate spectra for each sample. Soil spectra were measured from 600 to 4000 cm⁻¹ at a resolution of 4 cm⁻¹, with each spectra representing the average of 32 repeated scans.

2.4. Model constraints and data exclusions

MIR prediction models were constrained by soil Order, therefore only U.S. Mollisol soils were used during model development and evaluation. The Rendoll suborder is somewhat atypical in that it is of humid tropical regions, often forms on calcareous materials such as shell bars, and accounts for less than 0.5 % of U.S. Mollisols by land area (NRCS-Soil Taxonomy).

Therefore Rendolls, which accounted for only two samples in the spectral library, were excluded. Additionally, 31 samples collected from the O horizon were excluded, as this layer very rarely occurs in Mollisols. There were 5746 samples originating from 1117 soil pedons, representing 26 states and 243 counties in the KSSL spectral library which met our criteria. The comprehensive nature of the Soil Characterization database allowed for a large sample size with extensive geographic and taxonomic representation for U.S. Mollisols (Fig. 1.).

2.5. Calibration and validation data sets

There is a lack of consensus regarding the optimal strategy for selecting calibration and validation samples, but it is generally agreed that the calibration sets should be large enough, and be representative of future unknowns in terms of variation in spectral properties and variation in soil properties (Soriano-Disla et al., 2014; Stenberg et al., 2010). Due to the large sample size and comprehensive coverage of the soils data base used in this study (Fig. 1.), the MIR spectral library was assumed to be representative of US Mollisols as a population. The full dataset was divided into three subsets: 1) *independent validation*, 2) *internal validation*, and 3) *calibration*. The independent validation set was developed based on the criteria that all validation samples must have measurement data for all soil properties of interest. This resulted in an independent validation set consisting of 286 randomly selected soil samples from 50 different pedons.

An additional internal validation set was selected by taking a principal component analysis (PCA) of the spectra for the remaining samples, and then removing 25 % of the samples using the Kennard-Stone algorithm to ensure equal distribution in the final PCA space. This approach ensures that the spectral variation of the entire database is well represented in both the calibration and internal validation sets (Gogé et al., 2014; Guerrero et al., 2015). To minimize spatial dependence between validation and calibration data sets, no calibration and validation samples

originated from the same soil pedon.

The *internal validation* set was selected to ensure adequate spectral representation in both the calibration and validation sets, while the *independent validation* set was selected as an additional random check where all samples have measured values for all SQ indicators. The *independent validation* provided a more conservative estimate of errors, and so all subsequent discussion of error statistics will refer to the this data set unless stated otherwise.

2.6. Spectral preprocessing and MIR prediction

Spectral preprocessing and chemometric analyses were performed using OPUS QUANT 2 software version 7.2 (BrukerOptik GmbH 2012, Ettlingen, Germany). The OPUS Quick Compare feature was used to evaluate the repeatability of spectral replicates, and detect potential errors which may have occurred during scanning. Thirty four samples displayed spectral characteristics which were highly suspect, and so these samples were removed. For the remaining samples, replicate spectra were averaged, and mean spectra were used for subsequent analyses. Soil properties were predicted from MIR spectra by partial least squares regression (PLSR), using the PLS 1 algorithm in QUANT 2 (Martens and Naes, 1989). The OPUS QUANT 2 search routine was used to optimize spectral preprocessing and wavelength range selection by automatically and iteratively testing spectral preprocessing options in various wavelength ranges (Bornemann et al., 2008). Spectral preprocessing options included: 1) no data treatment, 2) subtraction of a constant offset, 3) subtraction of a linear function (SLF), 4) vector normalization (VN), 5) minimum-maximum scaling, 6) multiplicative scatter correction (MSC), 7) 1st derivative, 8) 2nd derivative, 9) 1st derivative and SLF, 10) 1st derivative and VN, and 11) 1st derivative and MSC. All calibration results were ranked and the method with the lowest prediction error was used to construct the model. To minimize the potential for over-fitting, the

maximum number of prediction factors were restricted to 20. A summary of the optimized MIR-PLSR models is provided in Table 1.

2.7. MIR-PLSR model evaluation

The performance of MIR-PLSR models was evaluated using a range of common chemometric indicators (Bellon-Maurel et al., 2010). The measured values ($Y_{meas.}$) and MIR-PLSR predicted values ($Y_{pred.}$) for each soil property were used to obtain the coefficient of determination (R^2), and the root mean square error of prediction (RMSE) Eq. [1]. The RMSE provides an estimate of the average prediction error for the MIR PLSR models. Systematic errors in MIR-PLSR models were evaluated based on the error of means, where the mean of the predicted values ($\bar{Y}_{pred.}$) and the mean of the measured values ($\bar{Y}_{meas.}$) were used to estimate bias Eq. [2]. The relative performance of MIR-PLSR models was assessed based on the ratio of performance to deviation (RPD), where the standard deviation (SD) of the measured values used in the validation set was divided by the RMSE Eq. [3].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_{pred.} - Y_{meas.})^2}{n}} \quad \text{Eq. [1]}$$

$$\text{Bias} = \bar{Y}_{pred.} - \bar{Y}_{meas.} \quad \text{Eq. [2]}$$

$$RPD = \frac{SD}{RMSE} \quad \text{Eq. [3]}$$

2.8. Overview of conceptual framework

One of the main goals of this work was to develop a mechanistic framework for evaluating how MIR-PLSR models function in terms of soil-spectral relationships; specifically—how models for predicting different soil properties utilize overlapping spectral information. This framework is built on several key assumptions which are based on widely accepted principles from the soil science and spectroscopy literature. 1) Soils are composed of materials with known absorption features in the MIR region, and the content of these materials will be reflected in the absorbance intensity in those regions. 2) Some soil properties directly represent physical components of the soil (Primary properties), while other soil properties are operationally defined, that is they represent a function of the soil (e.g., WR, CEC) rather than a specific physical component (Secondary properties). 3) Many soil properties are highly correlated, displaying a positive or negative relationship in the measured values. 4) The spectral features of one soil component may mask, distort, or shift the features of another component. 5) MIR prediction models rely on two types of distinct soil-spectral relationships: those spectral features directly associated with a predicted property (Direct spectral features), and spectral features associated with correlated soil properties (Indirect spectral features). A flowchart describing the basic overview of the methods applied, and theory behind the study is provided in Fig. 2.

Table 1. Descriptive statistics for measured soil quality (SQ) indicators, and summary statistics for the PLSR models for each predicted SQ indicator; only data for the independent validation set is shown. The optimized spectral pretreatments applied for each PLSR model are also provided.

		total N	total C	CO ₃	sand	silt	clay	WR .33	WR 2	WR 15	CEC	bulk density	pH	
		g kg ⁻¹	g kg ⁻¹	g kg ⁻¹	%	%	%	%	%	%	cmol _c kg ⁻¹	g cm ⁻³		
Summary statistics for measured properties		<i>Independent Validation Set</i>												
	n	286	286	276	286	286	286	286	286	286	286	286	286	286
	Mean	1.0	18.8	80.1	28.4	41.2	30.4	30.4	19.12	14.08	20.15	1.59	7.43	
	SD	1.0	16.5	116.2	21.8	15.1	12.9	12.9	7.06	5.45	8.37	0.21	0.79	
	Min	0.0	0.4	0.1	1.1	0.7	2.9	2.9	2.37	1.60	2.65	1.09	5.1	
	Q25	0.4	8.2	2.9	9.4	32.0	22.2	22.2	14.56	10.55	14.76	1.43	7.12	
	Median	0.8	14.4	31.1	24.2	40.4	28.6	28.6	18.49	13.21	19.53	1.58	7.62	
	Q75	1.2	23.7	115.2	43.2	51.6	37.3	37.3	23.81	17.12	24.76	1.73	7.85	
Max	7.6	97.1	624.3	96.4	72.8	72.1	72.1	49.52	36.23	54.93	2.08	9.94		
Summary statistics for PLSR models		<i>Independent Validation Set</i>												
	RMSE	0.45	4.0	15.8	7.3	8.0	3.9	5.13	3.93	3.07	2.45	0.15	0.51	
	RPD	2.2	4.7	7.43	3.07	1.9	3.3	1.53	1.79	1.78	3.42	1.38	1.67	
	R ²	0.79	0.96	0.98	0.90	0.76	0.90	0.58	0.70	0.70	0.91	0.49	0.65	
	Bias	-0.02	1.93	2.2	1.7	-0.5	-0.3	-0.68	-0.15	-0.28	-0.15	0.01	0.20	
	Slope	0.87	0.92	0.93	0.95	0.94	0.90	0.66	0.76	0.80	0.92	0.58	0.69	
Intercept	0.2	-0.4	3.2	-0.2	3.1	3.4	8.90	4.80	3.10	1.80	0.66	2.10		
Summary of PLSR models		<i>PLSR Models for SQ Indicators</i>												
	PLSR components [†]	14	17	16	17	19	20	17	19	20	19	16	17	
	Preprocessing [‡]	1st Der + SNV	1st Der + MSC	1st Der + SLF	1st Der + SNV	1st Der + SNV	1st Der + SNV	1st Der + SNV	1st Der + SNV	1st Der	1st Der + SNV	1st Der + SNV	1st Der + SNV	1st Der + SNV
Spectral regions (cm ⁻¹) [§]	600 - 3662	939 - 4003	939 - 2985	600 - 3662	600 - 4002	600 - 1618 - 4004	600 - 1618 - 4004	600 - 1281; 2303; 2982 - 3323	602 - 1281; 2303; 1618 - 4004	600 - 1281; 2301; 1618 - 4004	600 - 1281; 2301; 1618 - 4004	600 - 1281; 2301; 1618 - 4004	600 - 1281; 2301; 1618 - 4004	1279 - 1963; 2299 - 4004

[†] Number of PLSR components included in optimized PLSR models.

[‡] Spectral preprocessing techniques: 1st Der, first derivative; MSC, multiplicative scatter correction; SNV, standard normal variate normalization; SLF, subtraction of a linear function.

[§] Spectral regions included in optimized PLSR models.

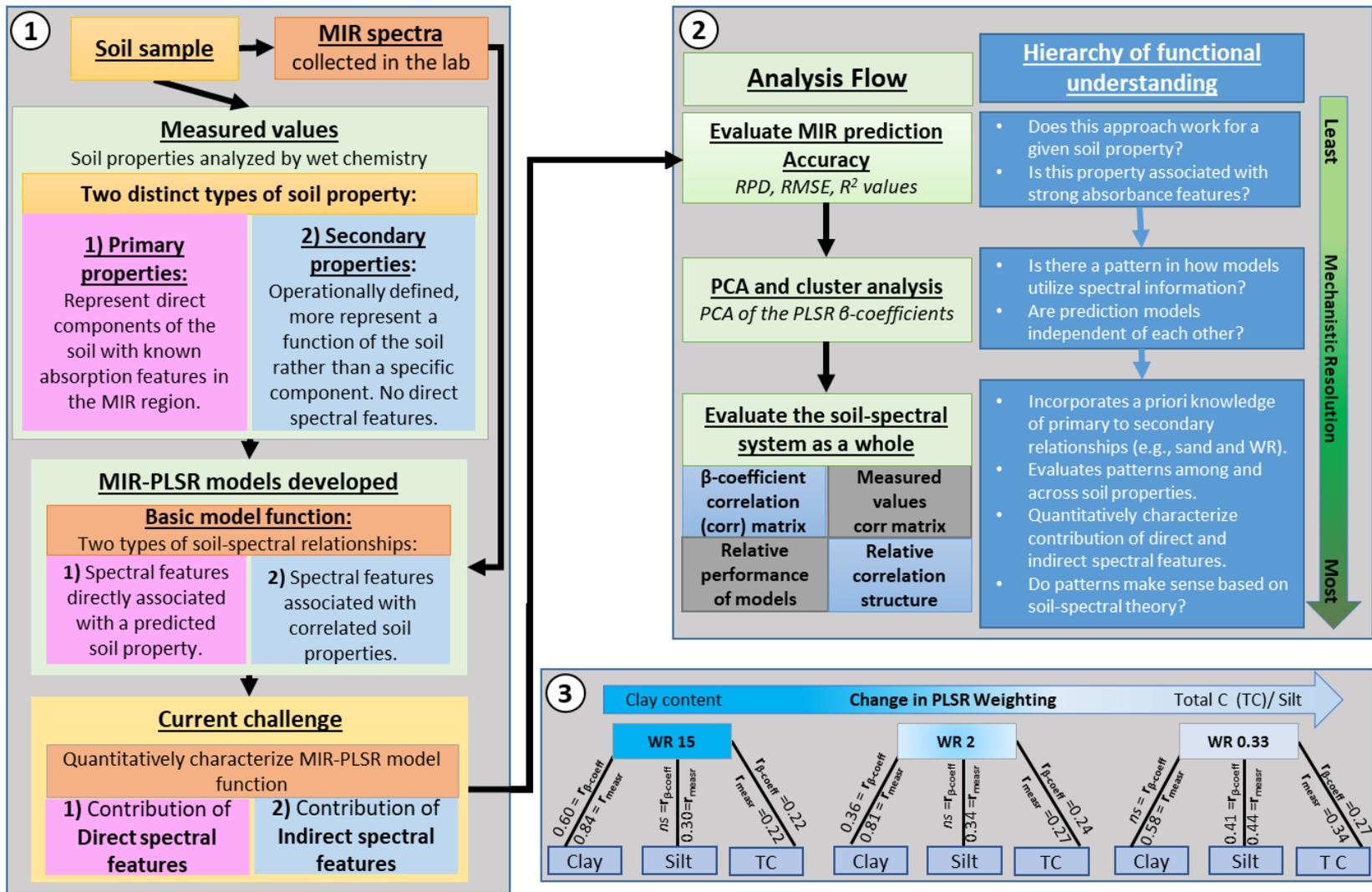


Fig. 2. Flowchart outlining the methodology applied, and basic theory behind the study. Box 1, Development of “black box” models; Box 2, Conceptual framework for evaluating MIR-PLSR model function; Box 3, Example evaluating patterns in the correlations among secondary (water retention (WR)) and primary (clay, silt, and TC content) properties, across secondary soil properties (WR at different matrix potentials). Where $r_{\beta\text{-measr}}$ and $r_{\beta\text{-coeff}}$ are the correlations for the measured values and PLSR β -coefficients, respectively.

Identifying structure/hierarchy in MIR-PLSR model dependence

To gain insight into the similarities among MIR-PLSR models, specifically, to identify which models utilize overlapping spectral regions, the MIR-PLSR β -coefficients for the different models were explored by principal component analysis (PCA) and variable clustering (Fig. 4; Table 2). The variable clustering algorithm developed by Cortez *et al.*, (2009) was applied, as it allows for additional variable interpretation by assigning MIR-PLSR models to groups based on the similarity of their β -coefficients. Models which grouped together rely on similar spectral features. Variable clustering and PCA were performed with JMP Pro Version 10.0.2 (SAS Institute).

Evaluating the soil-spectral system as a whole

Many of the different mineral and organic components of soil absorb energy in the MIR region, and the PLSR algorithm will use any spectral information with a strong enough relationship to the soil property being predicted. A correlation matrix of the measured values was used to identify correlated soil properties, and a correlation matrix of the MIR-PLSR β -coefficients was used to identify correlations among soil properties in terms of how their MIR-PLSR models utilize spectral information. The relative correlation structure was evaluated by comparing correlations for the measured values with correlations for the MIR-PLSR β -coefficients. We then evaluated how the relative correlation structure changed across soil properties in relation to our current understanding of soil science theory (Fig. 2; Box 3). In this way, model dependence among the primary and secondary soil properties was assessed, and the accuracy of predictions used to estimate the contribution of direct and indirect spectral features to each MIR-PLSR model. Additionally, primary soil properties were evaluated for their contribution to spectral features utilized by MIR-PLSR models of secondary soil properties.

Results and Discussion

Summary of measured soils data

Descriptive statistics of the SQ indicators for the *independent validation* data set are provided in Table 1. Descriptive statistics for the *calibration* and *internal validation* data sets are provided in Table S-1 of the supplementary materials. Soils were representative of all the major soil textural classes, and displayed a wide range in particle size distribution, with particle size classes ranging as follows: sand (0.1-99.8 %), silt (0.2-85.7 %), and clay content (0.0-87.8 %). The CO₃ content of soils ranged from none, to a composition of mostly CO₃, at 690 g kg⁻¹. A series of common transformations were applied to the measured soils data, but a normal distribution could not be achieved. The *independent validation* set tended to have a reduced range of values and standard deviation, potentially due to its smaller sample size, but in general properties displayed similar parameters across the different data sets, suggesting that calibration and validation sets are representative of variation in measured SQ indicators.

Evaluating MIR performance for predicting soil quality indicators

Summary statistics describing MIR-PLSR model performance for the *independent validation* data set are provided in Table 1. Results for the *Internal Validation* data set are provided in Table S-2 of the supplementary materials. As was observed in the measured data, MIR-PLSR models tended to display similar error statistics when predicting SQ indicators for the calibration and validation sets (Table S-2). The performance of MIR at predicting SQ indicators for Mollisols of the US, ranged from fair to excellent ($R^2 = 0.49 - 0.98$; RPD = 1.4 - 7.4). The RMSE associated with a given soil property will depend in part on the population parameters of the validation set, (i.e. range, mean, and standard deviation), and so this value

should be interpreted in conjunction with other measures of error such as RPD. The acceptable level of error will ultimately depend on the specific objectives or management goals associated with a given project. The best MIR predictions were obtained for CO₃, total C, CEC, clay, and sand (Table 1; Fig. 3), with error statistics of ($R^2 = 0.98$; RPD = 7.4), ($R^2 = 0.96$; RPD = 4.7), ($R^2 = 0.91$; RPD = 3.4), ($R^2 = 0.91$; RPD = 3.3), and ($R^2 = 0.90$; RPD = 3.1), respectively. The RPD values for these predicted SQ indicators are all higher than 3, and may be considered adequate for analytical purposes (Janik et al., 2009).

Prediction accuracies for the remaining SQ indicators were comparable with or better than previously reported results (Janik et al., 2007a; Sila et al., 2016; Terhoeven-Urselmans et al., 2010). However, for some SQ indicators, such as water retention, there have been very few studies reporting MIR predictions (Soriano-Disla et al., 2014). Janik *et al.*, (2007a) demonstrated successful prediction of soil water retention from MIR spectra for a small sample set ($n = 96$) of widely variable soil types from southern Australia, obtaining prediction accuracies of ($R^2 = 0.54$; RMSE = 6.7) and ($R^2 = 0.72$; RMSE = 4.3), based on a validation set of measurement data for soil water retention at field capacity (-33 kPa) and permanent wilting point (-1500 kPa), respectively. The MIR prediction accuracies for soil water retention observed in this study were comparable to that of Janik *et al.*, (2007a), with predictions for WR 0.33 and WR 15 displaying accuracies of ($R^2 = 0.60$; RMSE = 5.1) and ($R^2 = 0.70$; RMSE = 3.1), respectively. Similarly, the accuracy of MIR predictions were comparable to that of pedotransfer functions for estimating water retention, which require costly analytical measurements such as particle size distribution, bulk density, and soil organic matter (Cornelis et al., 2001).

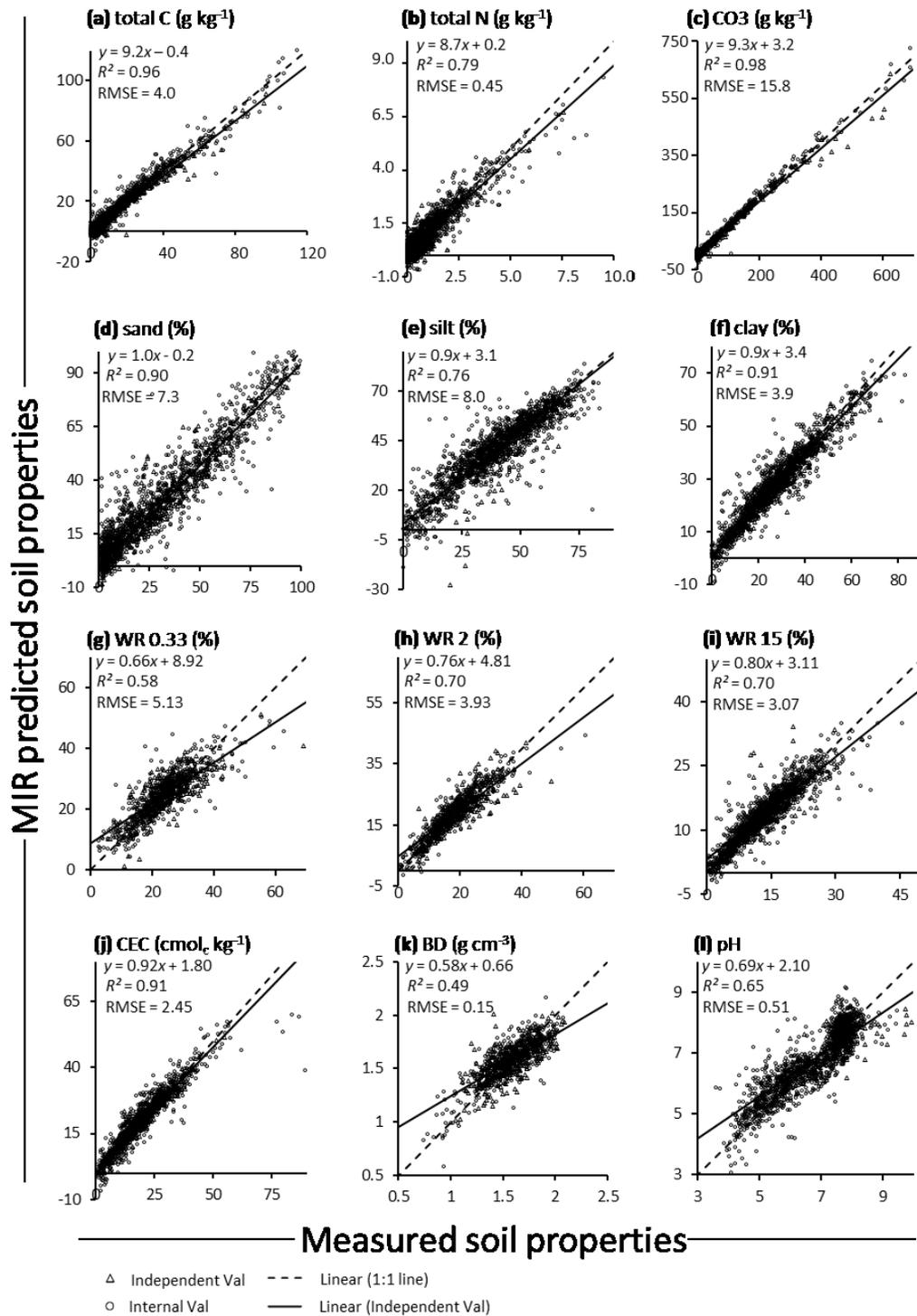


Fig. 3. Scatter plots comparing lab measured and MIR predicted values for (a) total N, (b) total C, (c) carbonates (CO₃), (d) sand, (e) silt, and (f) clay content, water retention at (g) 33 kPa (WR 0.33), (h) -200 kPa (WR 2), and (i) -1500 kPa (WR 15), (j) CEC, (k) bulk density (BD), and (l) pH. Plots include both internal (Internal Val) and independent validation (Independent Val) sets; however, regressions only reflect independent validation sets.

literature, summarizing the median R^2 values and number of published studies for MIR predictions of various soil physical, chemical, and biological properties. The MIR predictions presented in this study for CO_3 , total C, CEC, clay, sand, and silt are in general more accurate than those of previous studies reviewed by Soriano-Disla et al., (2014), where the median R^2 values and number of studies reviewed (n) for CO_3 , total C, CEC, clay, sand, and silt were ($R^2 = 0.95$; n = 3), ($R^2 = 0.93$; n = 16), ($R^2 = 0.85$; n = 13), ($R^2 = 0.80$; n = 14), ($R^2 = 0.83$; n = 11), and ($R^2 = 0.63$; n = 8), respectively. The high level of MIR accuracy observed in this study is particularly noteworthy, as few previous works have included such a wide spatial extent of sampling locations, large sample sizes, and diversity of soil taxa represented in calibration and validation sets.

There has been a well-documented trend in the soil spectroscopy literature where prediction errors generally increase as the spatial scale of calibration sets increase, with accurate prediction of soil properties for large and diverse areas proving especially difficult (Nocita et al., 2015; Stenberg and Viscarra Rossel, 2010). The poor performance of MIR for large-scale spectral libraries can often be traced to three main factors: the wider range and higher variability of soil properties, variation in the relationships between soil properties and spectral features, and inconsistencies in sampling protocols, instrumentation, and analytical methods (Nocita et al., 2015). The range and variability of soil properties appears to be the dominant factor influencing prediction accuracy (Stenberg et al., 2010), yet in our study soil properties displayed large ranges, high variability in measured properties, and MIR predictions with low errors (Table 1, Fig. 3). It is not clear how the chemical composition of soils and the spatial scale of calibration sets interact to influence variation in soil-spectral relationships, and so the soils included in this study were limited to the Mollisols order. This restricted the taxonomic distance among samples,

while allowing for large geographic distances between samples, and an adequate sampling density to provide thorough taxonomic and geographic representation (Fig. 1).

The relatively high accuracy of MIR predictions observed in this study is likely due in part to the high quality of the KSSL MIR spectral library and soils database. For example, field sampling was conducted by trained personnel, and soils were sampled by horizon for all samples in the database. Additionally, KSSL analytical techniques follow a rigorous quality assurance and quality control (QA/QC) protocol, with detailed metadata recorded for all analyses. The robust nature of the KSSL MIR spectral library and soils database, combined with the availability of detailed metadata, allowed for large sample sizes, consistency in sampling and analytical methods, low analytical errors, and thorough taxonomic and geographic representation of the Mollisols soil order, thereby allowing us to overcome many of the obstacles plaguing previous studies where large scale soil-spectral databases were evaluated (Nocita et al., 2015).

3.2. Identifying structure/hierarchy in MIR-PLSR model dependence

The level of independence, or the degree to which MIR-PLSR models rely on the spectral features of correlated properties, was evaluated by PCA and variable clustering of the MIR-PLSR β -coefficients for the different SQ indicators (Fig. 4.; Table 2; Table S-3). As can be seen from the variable clustering summary provided in Table 2, MIR-PLSR models grouped into five clusters, with clustering explaining 69 percent of the overall variation in MIR-PLSR models. The proportion of variation explained within each cluster ranged from 0.59-1.00, suggesting that a strong dependence structure exists among the MIR-PLSR models. The only SQ indicator with its own cluster was CO₃, and so it was the only independent MIR-PLSR model, i.e. relying primarily on direct spectral features.

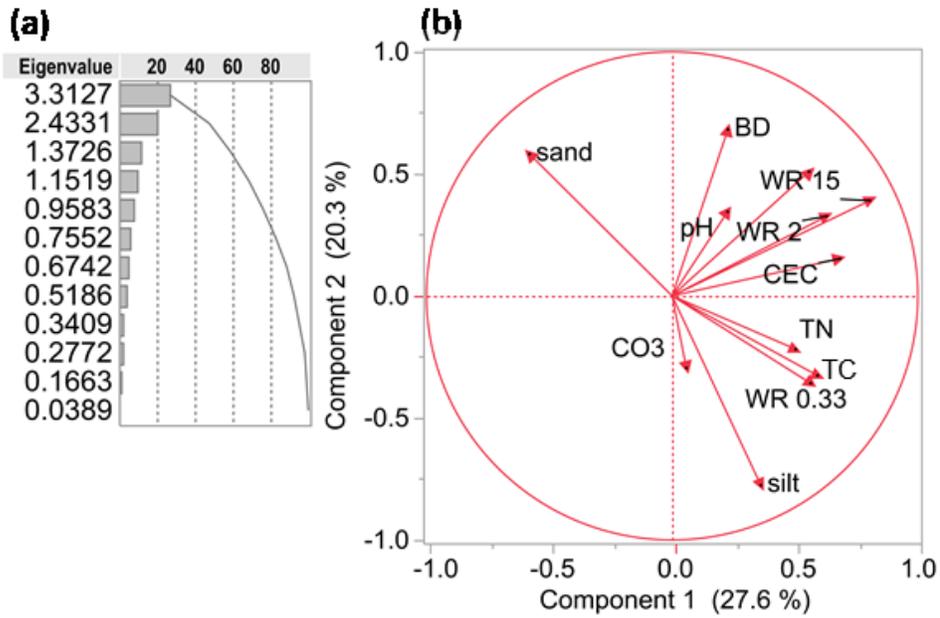


Fig. 4. Principal component analysis of the β -coefficients from partial least squares regression (PLSR) models for: total N (TN), total C (TC), carbonates (CO_3), sand, silt, and clay content, cation exchange capacity (CEC), bulk density, pH, and water retention at -33 kPa (WR 0.33), -200 kPa (WR 2), and -1500 kPa (WR 15).

Table 2. Summary of the principal component analysis-cluster analysis. Principal component analysis performed on the β -coefficients of the MIR-PLSR models.

PLSR- β -coefficient cluster analysis					
cluster id	explained variation	cluster members [†]	R^2 own-cluster	R^2 next-cluster	R^2 ratio
1	1.00	CO ₃	1.00	0.03	0.00
2	0.60	clay	0.68	0.07	0.34
		WR 15	0.72	0.29	0.39
		bulk density	0.47	0.05	0.56
		WR 2	0.50	0.14	0.58
3	0.59	total C	0.81	0.08	0.21
		total N	0.75	0.03	0.26
		WR 0.33	0.23	0.17	0.94
4	0.95	silt	0.95	0.09	0.06
		sand	0.95	0.13	0.06
5	0.59	pH	0.59	0.05	0.44
		CEC	0.59	0.22	0.53

Proportion of variation explained by clustering: 0.69

[†] MIR-PLSR models associated with each cluster: total N, total C, carbonates (CO₃), sand, silt, and clay content, water retention -33 kPa (WR 0.33), water retention -200 kPa (WR 2), and water retention -1500 kPa (WR 15), cation exchange capacity (CEC), bulk density, and pH.

The first variable listed within each cluster is the most representative member. For example, the MIR-PLSR models for clay, WR 15, bulk density, and WR 2 were all in cluster 2, that is, all models rely on similar spectral information, with clay being the most representative variable (Table 2.). The R^2 ratio measures the relative closeness between a variable's current cluster and the next most similar cluster. The members of cluster 3 included total C, total N, and

WR 0.33, with total C having the most representative MIR-PLSR model. Water retention at -33 kPa had an R^2 ratio of 0.94, implying that there is considerable overlap with the spectral information utilized by cluster 4, i.e. the sand/silt cluster. The proportion of variation explained within cluster 4 was 0.95, suggesting that MIR-PLSR models for silt and sand were highly dependent, but in general shared little similarity with models for other primary soil properties. Cluster 5 contained the models for pH and CEC. The MIR-PLSR models for WR 0.33, WR 2, bulk density, CEC, and pH all appear to display considerable overlap in their use of spectral information, with R^2 ratios of 0.94, 0.58, 0.56, 0.53, and 0.44, respectively. Based on the PCA and variable clustering analysis of the β -coefficients, it appears that soil properties in clusters 1, 2, and 3 were predominantly associated with the direct spectral features of a single primary soil property, i.e. associated with CO_3 , clay, and total C content, respectively. The relative contribution of sand and silt to cluster 4 cannot be determined from this analysis, while cluster 5 appears to be associated with a combination of spectral features from both clay and total C content.

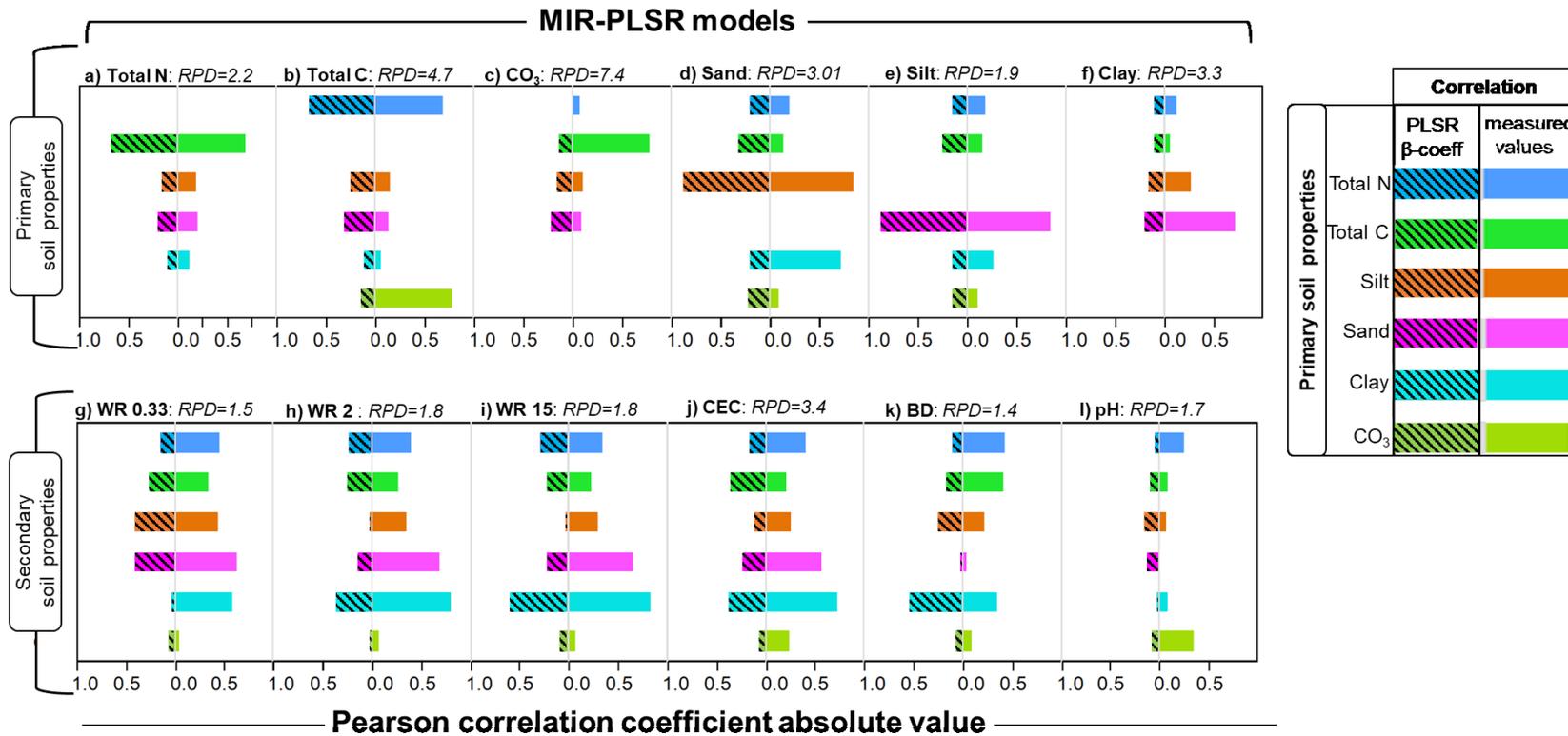


Fig. 5. Mechanistic evaluation of the soil-spectral system as reflected by model performance (*RPD*), and the relative correlation structure of the measured properties (measured values) and model β -coefficients (PLSR β -coeff). Correlations among primary soil properties (5a-f; total N, total C, carbonates (CO₃), sand, silt, and clay content,) and correlations between primary and secondary soil properties (5g-l; water retention at -33 kPa (WR 0.33), -200 kPa (WR 2) , and -1500 kPa (WR 15), cation exchange capacity (CEC), bulk density (BD), and pH) are presented.

Evaluating the soil-spectral system as a whole

A conceptual framework for model function was developed using the correlation structure of the β -coefficients, the R^2 /RPD of MIR-PLSR models, and the correlation structure of the measured SQ indicators (Fig. 5.). Soil properties associated with strong absorption features in the MIR region tend to display better predictability with higher R^2 values and lower errors, and MIR-PLSR models for these properties tend to rely more heavily on direct spectral features (Reeves III et al., 2006). Most MIR-PLSR models displayed strong dependence with regard to model weighting by spectral region (Fig. 5.; Table 3).

Soil properties displaying correlations among their measured values tended to exhibit increased dependency among their respective MIR-PLSR models. However, primary soil properties with prominent direct spectral features may have highly correlated measured values, but little or no correlation for the PLSR β -coefficients. For example CO_3 and total C measurements were highly correlated ($r = 0.79$), while there was very little overlap in the spectral regions utilized by their respective MIR-PLSR models ($r = 0.14$) (Fig. 5.b, 5.c; Table 3.). This is likely due to the fact that total C is composed of both the organic and inorganic fractions of soil C, while CO_3 measurements solely reflect the inorganic C content. Therefore, the MIR-PLSR model for total C must rely on the spectral features of both the inorganic and organic C fractions, while models for CO_3 may rely entirely on the absorbance features of inorganic C bonds. A similar trend was observed between the sand and clay models, where the measured values were correlated ($r = -0.72$), but MIR-PLSR models displayed a much lower correlation ($r = -0.20$) (Fig. 5.d, 5.f, Table 3.). In this instance, the sand and clay fractions each have a distinct chemical composition with prominent absorption features (Fe-oxides, layer alumina silicates, and

Table 3a. Correlation matrix between measured soil properties.

	total N	total C	CO ₃	sand	silt	clay	WR 0.33	WR 2	WR 15	CEC	bulk density	pH
total N	1.00											
total C	0.69***	1.00										
CO ₃	-0.07***	0.79***	1.00									
sand	-0.19***	-0.13***	-0.08***	1.00								
silt	0.19***	0.14***	0.10***	-0.86***	1.00							
clay	0.11***	0.06***	<i>ns</i>	-0.72***	0.27***	1.00						
WR 0.33	0.45***	0.34***	<i>ns</i>	-0.63***	0.44***	0.58***	1.00					
WR 2	0.39***	0.27***	<i>ns</i>	-0.69***	0.34***	0.81***	0.77***	1.00				
WR 15	0.34***	0.22***	-0.06**	-0.66***	0.30***	0.84***	0.77***	0.95***	1.00			
CEC	0.39***	0.20***	-0.23***	-0.57***	0.25***	0.73***	0.72***	0.83***	0.86***	1.00		
bulk density	-0.42***	-0.41***	-0.08**	<i>ns</i>	-0.21***	0.35***	-0.30***	<i>ns</i>	0.08***	<i>ns</i>	1.00	
pH	-0.24***	0.09***	0.34***	<i>ns</i>	-0.06***	0.09***	-0.05*	-0.05**	0.04**	-0.04**	0.10***	1.00

Table 3b. Correlation matrix between β -coefficients of the MIR-PLSR prediction models for the different soil properties.

	total N	total C	CO ₃	sand	silt	clay	WR 0.33	WR 2	WR 15	CEC	bulk density	pH
total N	1.00											
total C	0.67***	1.00										
CO ₃	<i>ns</i>	0.14***	1.00									
sand	-0.20***	-0.31***	-0.21***	1.00								
silt	0.14***	0.24***	0.15***	-0.89***	1.00							
clay	0.10***	0.09***	<i>ns</i>	-0.20***	-0.15***	1.00						
WR 0.33	0.15***	0.27***	0.06**	-0.40***	0.41***	<i>ns</i>	1.00					
WR 2	0.24***	0.24***	<i>ns</i>	-0.13***	<i>ns</i>	0.36***	0.20***	1.00				
WR 15	0.28***	0.22***	-0.08**	-0.22***	<i>ns</i>	0.60***	0.43***	0.60***	1.00			
CEC	0.17***	0.35***	-0.06**	-0.23***	0.11***	0.38***	0.32***	0.33***	0.55***	1.00		
bulk density	-0.10***	-0.17***	-0.06**	<i>ns</i>	-0.24***	0.54***	-0.25***	0.25***	0.37***	0.17***	1.00	
pH	<i>ns</i>	0.09***	-0.07**	0.12***	-0.15***	<i>ns</i>	0.13***	0.24***	0.28***	0.17***	0.18***	1.00

* Significant at $P < 0.05$ ** Significant at $P < 0.01$ *** Significant at $P < 0.001$ *ns*: not significant at $P < 0.05$

Al-oxides for clays and quartz minerals for sand), allowing for independent MIR-PLSR models with correlated measured properties. While MIR-PLSR models of primary soil properties all displayed some level of dependence, only the models for CO₃, total C, clay, and sand appear to be relying heavily on direct spectral features (Fig. 5.b, 5.c, 5.d, 5.f). The CO₃, total C, clay, and sand models all displayed excellent prediction accuracy with RPD > 3, and all but sand were the most representative variable or their respective clusters (Table 2). Sand and silt were equally representative of cluster 4, with nearly identical R^2 own-cluster, next-cluster, and ratios.

The MIR-PLSR models for total N and silt appear to be relying more on the spectral features of total C and sand, respectively. This would make sense, as mean total N concentrations were an order of magnitude less than that of total C, and would likely result in weaker spectral features. Sand and silt models were highly dependent, with correlations of $r = -0.86$ and $r = -0.89$ for the measured properties and β -coefficients, respectively. Both sand and silt displayed similar trends in their relative correlation structure, but silt tended to have weaker correlations for the β -coefficients. Further interpretation of correlations among primary soil properties would require additional analyses of the absorption features associated with MIR-PLSR β -coefficient spectra (Fig. 6.), such as band assignments for known soil compounds.

Of the secondary soil properties, only CEC displayed excellent predictions with an RPD = 3.4 (Fig.4.j). As observed with the sand and silt models, there was an increase in the correlation of the β -coefficients relative to the measured properties for CEC and total C, with correlations of 0.35 and 0.20, respectively. Cation exchange capacity appeared to be relying most heavily on the spectral features of clay followed by total C, then sand, with correlations of 0.38, 0.35, and -0.23, respectively (Table 3b). Water retention at -33 kPa was highly correlated with clay content, 0.58 correlation, yet their MIR-PLSR models were independent. The model for WR 0.33 appeared to

be utilizing spectral information associated with sand, silt, and total C. This is in accordance with basic soil-water principles, as water retention at field capacity (-33 kPa) is primarily dependent on the large pore spaces associated with sand sized particles and soil organic matter. Similarly, the models for WR 2 and WR 15 showed an increasing reliance on the spectral features associated with clay, and a decreasing reliance on features associated with total C (Fig. 5.; Fig. 2, Box 3). This makes sense, as water retention at the permanent wilting point (-1500 kPa) is largely dependent on small pore spaces in clay sized material (Soriano-Disla et al., 2014).

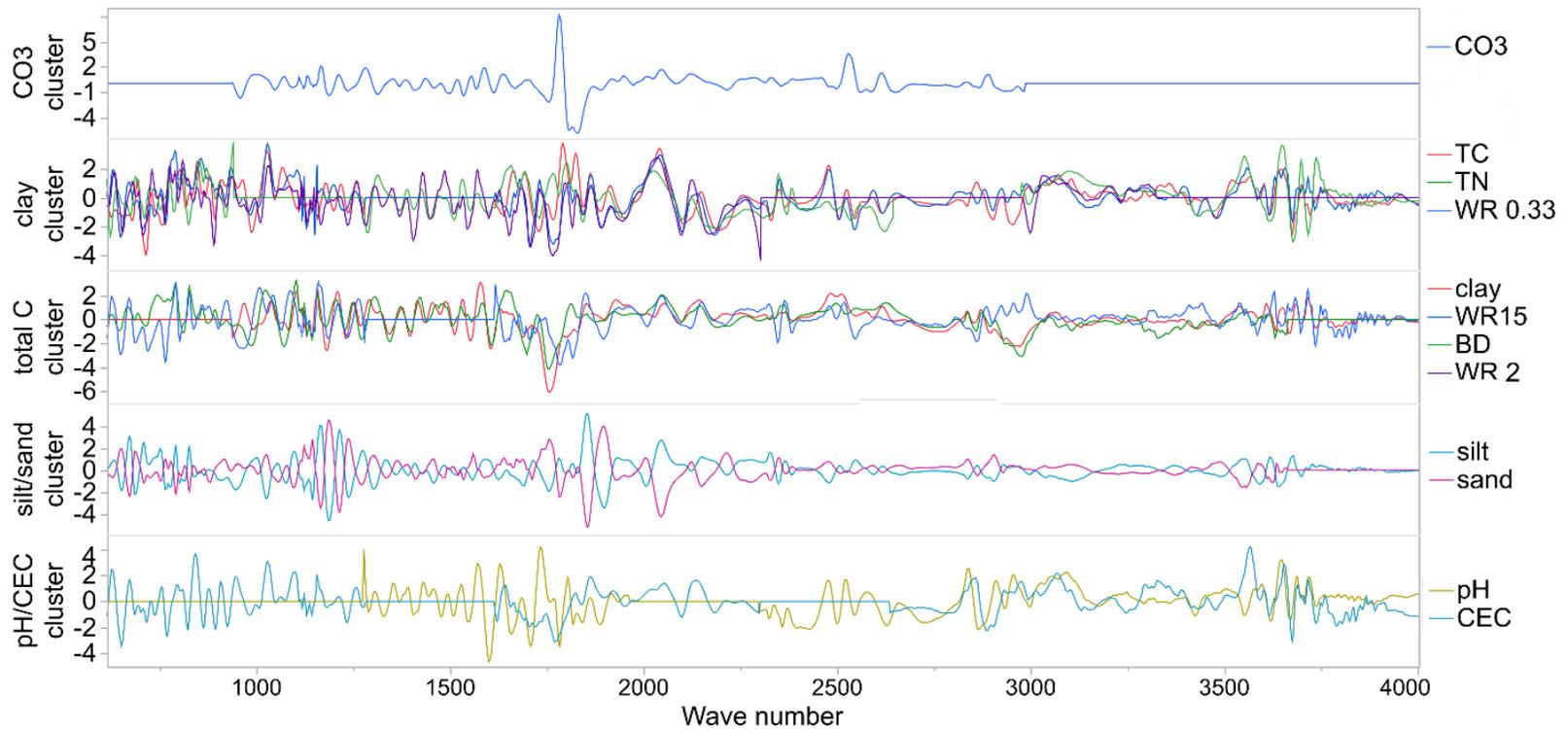


Fig. 6. Spectra of the β -coefficients from the partial least squares regression (PLSR) models for: total C (TC), total N (TN), carbonates (CO₃), sand, silt, and clay content, cation exchange capacity (CEC), bulk density (BD), pH, and water retention at -33 kPa (WR 0.33), -200 kPa (WR 2), and -1500 kPa (WR 15). The grouping of spectra within plots is based on a PCA-cluster analysis of the β -coefficients.

The standardized MIR-PLSR β -coefficient spectra for the different SQ indicators were grouped together based on their PCA cluster assignment, and plotted so that model function could be visualized and qualitatively assessed (Fig. 6). There was clear overlap in the spectral features utilized by models grouped in the same cluster, further supporting the effectiveness of the variable clustering algorithm, and the validity of MIR-PLSR model interpretations. Spectral preprocessing methods may influence the interpretation of β -coefficient spectra, as these approaches may accentuate and mask features in the original MIR spectra. The first derivative transformation was applied for all models to remove additive spectral effects, as well as various corrective functions to minimize multiplicative effects associated with sample grinding and optical set-up (Shepherd and Walsh, 2002) (Table 1). The sand and silt models displayed a strong negative correlation, and β -coefficient spectra were essentially inverses of each other (Fig. 6.d). Models assigned to the clay cluster also showed considerable overlap in spectral regions of importance, most notably the regions around 3500-3800, 1800-2100, 1600-1790, 1000-1200, and the region below 1000. Similar trends were observed for models assigned to the total C cluster (Fig. 6.b, 6.c). The CEC and pH models appear to rely on a combination of spectral features associated with nearly of all the primary soil properties.

Conclusions

A large sample size, high quality analytical data, and detailed metadata are essential for minimizing many of the sources of error, which commonly plague studies evaluating large-scale spectral libraries and soils databases. Developing MIR-PLSR models specific to Mollisols was an effective strategy for calibration development. This approach allowed for the simultaneous

prediction of multiple SQ indicators over large spatial scales, with soils representing high sample diversity in terms of soil type, land use, and climate. Additionally, high data quality and dimensionality allows for complex and robust multivariate analyses of the mechanistic function of MIR-PLSR models, that takes into account *a priori* knowledge of relationships between Primary (e.g., soil texture) and Secondary (e.g. WR) soil properties. This conceptual framework allows interpretation of MIR-PLSR models by multivariate analysis of the β -coefficients and evaluation of the relative performance and correlation structure for each SQ indicator. This provides a window into the “black box”, and allows for the development of new insight into how MIR-PLSR models utilize spectral information for primary and secondary soil properties.

REFERENCES

- Allen, B.L., Fanning, D.S., 1983. Composition and soil genesis, in: Pedogenesis and Soil Taxonomy. Elsevier, New York, pp. 141–192. doi:10.1016/S0166-2481(08)70601-9
- 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. doi:10.1071/SR13077
- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC - Trends in Analytical Chemistry* 29, 1073–1081. doi:10.1016/j.trac.2010.05.006
- Bornemann, L., Welp, G., Brodowski, S., Rodionov, a., Amelung, W., 2008. Rapid assessment of black carbon in soil organic matter using mid-infrared spectroscopy. *Organic Geochemistry*

39, 1537–1544. doi:10.1016/j.orggeochem.2008.07.012

Brown, D.J., Shepherd, K.D., Walsh, M.G., Dewayne Mays, M., Reinsch, T.G., 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma* 132, 273–290.

doi:10.1016/j.geoderma.2005.04.025

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. doi:10.2136/sssaj2001.652480x

Cornelis, W.M., Ronsyn, J., Meirvenne, M. Van, Hartmann, R., 2001. Evaluation of pedotransfer functions for predicting the soil moisture retention curve. *Soil Sci. Soc. Am. J.* 65, 638–648.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., Reis, J., 2009. Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems* 47, 547–553.

doi:10.1016/j.dss.2009.05.016

de Gruijter, J.J., McBratney, A.B., Minasny, B., Wheeler, I., Malone, B.P., Stockmann, U., 2016. Farm-scale soil carbon auditing. *Geoderma* 265, 120–130. doi:10.1016/j.geoderma.2015.11.010

Gobrecht, A., Roger, J.M., Bellon-Maurel, V., 2014. Major issues of diffuse reflectance NIR spectroscopy in the specific context of soil carbon content estimation. A review., *Advances in Agronomy*. doi:10.1016/B978-0-12-420225-2.00004-2

Gogé, F., Gomez, C., Jolivet, C., Joffre, R., 2014. Which strategy is best to predict soil properties of a local site from a national Vis-NIR database? *Geoderma* 213, 1–9.

doi:10.1016/j.geoderma.2013.07.016

Grunwald, S., Thompson, J.A., Boettinger, J.L., 2011. Digital soil mapping and modeling at continental scales: Finding solutions for global issues. *Soil Science Society of America Journal* 75, 1201. doi:10.2136/sssaj2011.0025

Guerrero, C., Wetterlind, J., Stenberg, B., Mouazen, A.M., Gabarrón-Galeote, M.A., Ruiz-Sinoga, J.D., Zornoza, R., Viscarra Rossel, R.A., 2015. Do we really need large spectral libraries for local scale SOC assessment with NIR spectroscopy? *Soil and Tillage Research*. 155, 501–509. doi:10.1016/j.still.2015.07.008

Janik, L.J., Forrester, S.T., Rawson, A., 2009. The prediction of soil chemical and physical properties from mid-infrared spectroscopy and combined partial least-squares regression and neural networks (PLS-NN) analysis. *Chemometrics and Intelligent Laboratory Systems* 97, 179–188. doi:10.1016/j.chemolab.2009.04.005

Janik, L.J., Merry, R.H., Forrester, S.T., Lanyon, D.M., Rawson, a., 2007a. Rapid prediction of soil water retention using mid infrared spectroscopy. *Soil Science Society of America Journal* 71, 507. doi:10.2136/sssaj2005.0391

Janik, L.J., Skjemstad, J.O., Shepherd, K.D., Spouncer, L.R., 2007b. The prediction of soil carbon fractions using mid-infrared-partial least square analysis. *Australian Journal of Soil Research* 45, 73–81. doi:10.1071/SR06083

Liu, X., Lee Burras, C., Kravchenko, Y.S., Duran, A., Huffman, T., Morras, H., Studdert, G., Zhang, X., Cruse, R.M., Yuan, X., 2012. Overview of Mollisols in the world: Distribution, land use and management. *Canadian Journal of Soil Science* 92, 383–402. doi:10.4141/cjss2010-058

Martens, H., Naes, T., 1989. *Multivariate calibration*. Wiley, New York.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., BenDor, E., Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J.A.M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J.M., Shepherd, K.D., Stenberg, B., Towett, E.K., Vargas, R., Wetterlind, J., 2015. *Soil spectroscopy: an alternative to wet chemistry for soil monitoring*.

Adv. Agron. 132, 139–159. doi:10.1016/bs.agron.2015.02.002

Reeves III, J.B., 2012. Mid-infrared spectral interpretation of soils: Is it practical or accurate? *Geoderma* 189–190, 508–513. doi:10.1016/J.geoderma.2012.06.008

Reeves III, J.B., Follett, R.F., McCarty, G.W., Kimble, J.M., 2006. Can near or mid-infrared diffuse reflectance spectroscopy be used to determine soil carbon pools? *Communications in Soil Science and Plant Analysis* 37, 2307–2325.

Savvides, A., Corstanje, R., Baxter, S.J., Rawlins, B.G., Lark, R.M., 2010. The relationship between diffuse spectral reflectance of the soil and its cation exchange capacity is scale-dependent. *Geoderma* 154, 353–358. doi:10.1016/j.geoderma.2009.11.007

Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal* 66, 988. doi:10.2136/sssaj2002.9880

Sherpa, S.R., Wolfe, D.W., van Es, H.M., 2016. Sampling and data analysis optimization for estimating soil organic carbon stocks in agroecosystems. *Soil Science Society of America Journal* 80, 1377. doi:10.2136/sssaj2016.04.0113

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. doi:10.1016/j.chemolab.2016.02.013

Soil Survey Staff, 2014. *Kellogg Soil Survey Laboratory Methods Manual. Version 5.0.* USDA-NRCS, Washington D.C

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. doi:10.1080/05704928.2013.811081

Stenberg, B., Viscarra Rossel, R.A., 2010a. Diffuse reflectance spectroscopy for high-resolution soil sensing, in: Viscarra Rossel, R.A., McBratney, A.B., Minasny, B. (Eds.), *Proximal Soil Sensing*. pp. 29–47. doi:10.1007/978-90-481-8859-8_3

Stenberg, B., Viscarra Rossel, R. a., Mouazen, A.M., Wetterlind, J., 2010b. Visible and near infrared spectroscopy in soil science, *Advances in Agronomy*. 107, pp. 163–215.

doi:10.1016/S0065-2113(10)07005-7

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. doi:10.2136/sssaj2009.0218

Wight, J.P., Ashworth, A.J., Allen, F.L., 2016. Organic substrate, clay type, texture, and water influence on NIR carbon measurements. *Geoderma* 261, 36–43.

doi:10.1016/j.geoderma.2015.06.021

Wolfe DW. 2019. *The New York Soil Health Roadmap*. Cornell University. 13 pp.

(www.newyorksoilhealth.org).

Wolfe, D.W., 2013. Climate change solutions from the agronomy perspective, in: Daniel, H., Cynthia, R. (Eds.), *Handbook of Climate Change and Agroecosystems: Global and Regional Aspects and Implications*. Imperial College Press, pp. 11–29. doi:10.1142/9781848169845_0002

CHAPTER 3

GROUND-TRUTHING MID-IR SPECTROSCOPY FOR VERIFYING SOIL CARBON: IS THE ACCURACY OF STANDARD SOIL ANALYSES WORTH THE COST?

Abstract

The cost of accurately quantifying soil organic carbon (SOC) change in response to land management is a primary constraint to more widespread implementation of programs to incentivize SOC sequestration. Mid infrared diffuse reflectance spectroscopy (MIR) has been proposed as a low cost accurate method for establishing SOC baseline levels and verifying change, but there is presently little information regarding the cost and error associated with MIR compared to standard soil testing methods. To evaluate the relative cost and measurement error of MIR compared to standard soil testing laboratories, 25 reference soil samples from the Mollisols Order, were obtained from the North American Proficiency Testing Program (NAPT), and total C, total N, soil carbonates, SOC, cation exchange capacity (CEC), and particle size distribution were predicted by partial least squares regression from MIR spectra. This paper presents trends in the cost and accuracy of MIR predicted soil properties compared to typical results obtained from soil testing laboratories participating in the NAPT Proficiency Assessment Program for U.S. Mollisols. Total C, carbonates, and SOC predictions by MIR displayed an analytical level of accuracy comparable to that of North American soil testing laboratories, while the cost of quantifying SOC by standard laboratory measurement was three and half times greater than MIR prediction. Our analyses further revealed that analytical method cost and measurement error were positively correlated, where standard methods for CEC and particle size

distribution displayed both the highest costs and measurement errors. Current approaches to model evaluation cannot separate analytical errors associated with evaluation samples from prediction errors, presenting a “chicken or the egg” dilemma for soil spectroscopy; measurement errors and costs associated with reference methods will need to be considered in future studies evaluating the performance of spectroscopic approaches such as MIR.

Introduction

The human population and thus global demand for food and clean water are rising at an alarming rate, while historic land use practices have left one-third of all soils and more than half of agricultural soils moderately or severely degraded with a reduced functional capacity to provide ecosystem services (Godfray et al., 2010; FAO, 2015). Furthermore, the frequency of extreme weather events leading to drought, flooding, and heat stress are projected to increase, with the impacts of climate change on crop productivity and water availability already apparent in many parts of the world (Wheeler and von Braun, 2013; Wolfe, 2013; Dhankher and Foyer, 2018). Serving to regulate the flow and storage of water and nutrients for the terrestrial biosphere, soil is the cornerstone of the water, food, energy nexus, and the future of global food security and economic stability will depend on the successful implementation of management strategies to restore degraded soils across the globe (Kahiluoto et al., 2014; Hatfield et al., 2017). Soil-centric management responses to climate change have largely focused on building soil organic carbon (SOC), as increasing SOC can reduce compaction, improve drainage and aeration, increase water holding and cation exchange capacity, and enhance nutrient cycling and fertility, thereby improving agronomic productivity and resource use efficiency in agroecosystems (Lal, 2004; Glenk and Colombo, 2011). In addition to providing an adaptation

strategy through increased resilience of agricultural systems, soil-based approaches can serve as climate change mitigation strategies, and can result in substantial reductions in greenhouse gas (GHG) emissions (i.e. methane, nitrous oxide), and sequester some atmospheric carbon dioxide (CO₂) in the form of SOC (Paustian et al., 2016). While the concept of building SOC to mitigate agricultural GHG emissions and enhance food security continues to gain interest among the scientific community, the cost of accurately quantifying and verifying soil management effects on SOC change remains to be the primary factor preventing widespread implementation of programs incentivizing SOC sequestration (Conant et al., 2011; Kahiluoto et al., 2014; Paustian et al., 2016; Griscom et al., 2017).

Mid infrared diffuse reflectance spectroscopy (MIR) has shown great promise as a rapid, low cost, and accurate option for measuring soil properties, and could potentially provide a way forward for overcoming the barriers currently impeding SOC-based GHG mitigation programs (Stenberg et al., 2010; Nocita et al., 2015). The mineral and organic components of soils contain chemical bonds that absorb energy in the mid-IR region of the electromagnetic spectrum. MIR relies on multivariate statistical approaches such as partial least squares regression (PLSR) to utilize this spectral information, allowing for the quantitative determination of a range of soil properties including total C and N, soil carbonates, SOC, soil texture, and cation exchange capacity (CEC). One of the main challenges faced by soil spectroscopy has been the task of developing calibrations capable of predicting soil properties for large and diverse areas with an acceptable level of accuracy, as soil-spectral relationships may differ vastly among taxonomically distant soils (Nocita et al., 2015). To account for this soil-spectral variability calibrations may be developed from large and taxonomically comprehensive soil-spectral libraries paired with high quality soil databases, such as the USDA-NRCS Kellogg Soil Survey

Laboratory (KSSL) Soil Characterization database, which is publicly available at no charge. For instance, Sherpa et al. (2019) used the KSSL database to predict 12 soil quality indicators by MIR for approximately 5200 Mollisol soils from across the US, achieving comparable or lower errors than typically observed for local scale models. While the cost-effectiveness of soil spectroscopic approaches to soil analysis is mentioned in nearly every paper published on the subject, few previous studies have attempted even a simple cost-error comparison with traditional wet chemistry methods (O'Rourke and Holden, 2011; Schwartz et al., 2012; Tóth et al., 2013).

The dearth of information regarding the relative cost and error of MIR analysis compared to standard soil testing methods, is due in part from the lack of a straightforward way to compare MIR evaluation statistics (root mean square error of prediction (RMSE), coefficient of determination (R^2), and bias), with the typical error statistics used for evaluating standard analytical methods. Furthermore, the insufficient assessment of standard analytical method errors presents a “chicken or the egg” dilemma for soil spectroscopy, where model evaluation doesn't allow for errors in the initial analytical measurements to be separated from MIR prediction errors, and so all error is assumed due to MIR. Another possible reason for the lack research on MIR's cost-effectiveness is the abstract nature of error interpretation, for example a statistical power analysis or an evaluation of sample size requirements for detecting change would be a simple but valuable addition to most studies, yet have been neglected in both soil spectroscopy and soil C research (Cohen, 1992; Kravchenko and Robertson, 2011). Robust quantification of the relative cost and method error for MIR predictions is essential if this technology is to be widely adopted for SOC verification. Organizations such as the North American Proficiency Testing (NAPT) Program maintain large soil archives encompassing a wide diversity of control

samples, oversee inter-laboratory sample exchange and statistical evaluation of the analytical data, and provide proficiency assessment and certification programs for commercial soil testing laboratories. This pooling of resources allows for more robust method evaluation, which would otherwise be cost prohibitive for any single laboratory or research group, yet to our knowledge such a dataset has never been used to evaluate the performance and cost of MIR predictions compared to standard soil testing laboratories.

The objective of this study was to use NAPT reference soil materials and published soil testing results to compare the accuracy and cost of MIR predicted total C and N, soil carbonates, SOC, soil texture, and CEC with the accuracy and cost of standard analytical measurements performed by soil testing laboratories participating in the NAPT program. To assess MIR's cost-effectiveness for SOC verification in agricultural systems, the statistical power of each method was adjusted to account for measurement error, and the analytical budgets required for quantifying differences in SOC were compared by power analysis.

Materials and Methods

Reference soil materials & Standard analytical methods for soil analysis

Reference soil materials used in this study were obtained from the North American Proficiency Testing Program, as part of their inter-laboratory sample exchange. The NAPT published median and median absolute deviation (MAD) values, reflect the results submitted by soil testing laboratories participating in the program. NAPT quarterly reports for soil testing results from 2009-2017 were downloaded from (<https://www.naptprogram.org/content/laboratory-results>), and individual reference soil materials were selected based on the criteria that soils be of the Mollisols Order, sample was available, and

that there were no potential conflicts with APHIS-regulation. Twenty-five NAPT reference soil samples met our requirements, and were used for all three evaluation datasets compared throughout this study (Fig. 1). Values used for dataset 1) NAPT published results, were taken directly from the published quarterly reports. A comparison of methods used by NAPT participating soil testing laboratories is in Table 1, where the methodology identified by the local Land Grant University were followed.

Additionally, NAPT reference soil materials were sent to the USDA-NRCS Kellogg Soil Survey Laboratory in Lincoln, NE, for MIR analysis and measurement of soil properties by standard methods to be used in evaluation datasets 2) MIR predictions, and 3) KSSL Analytical measurement, respectively (Fig. 1). Methods for these data were conducted according to the KSSL Methods Manual following protocols for total carbon and total nitrogen (4H2a), carbonates (4E1a1a1a1), cation exchange capacity (4B1a1a), total sand, silt, and clay content (3A1a1a), and MIR diffuse reflectance spectroscopy (7A7).

Table 1. Comparison of soil analytical methods used by NAPT participating laboratories (NAPT) and Kellogg Soil Survey Laboratory (KSSL). Descriptive statistics for the sample sizes of NAPT participating laboratories are also presented.

	Laboratory Method comparison		Sample size for ‡ NAPT published results			
	NAPT †	KSSL	min	max	mean	Std Dev
Total C	Soil Total C (combustion)	Carbon, Total NCS	23	36	29	4
Total N	Soil TN (combustion)	Nitrogen, Total NCS	31	44	39	4
CO3	CaCO3 Content	Calcium Carbonate Equivalent, HCl, <2mm	8	19	14	4
SOC	(Total C) - 0.12*(CO3)	(Total C) - 0.12*(CO3)	7	12	9	2
CEC	CEC - Cation Displacement	CEC and Cations, NH4OAc, 2M KCl displacement	15	24	20	3
Sand	Particle Size Analysis-Hydrometer	PSDA, Routine, Pipet	34	47	38	4
Silt	Particle Size Analysis-Hydrometer	PSDA, Routine, Pipet	34	47	38	4
Clay	Particle Size Analysis-Hydrometer	PSDA, Routine, Pipet	34	47	38	4

† Values derived using the methodology identified by the local Land Grant University.

‡ Sample size descriptive statistics reflect the number of NAPT participating laboratories submitting results for a given reference soil material.

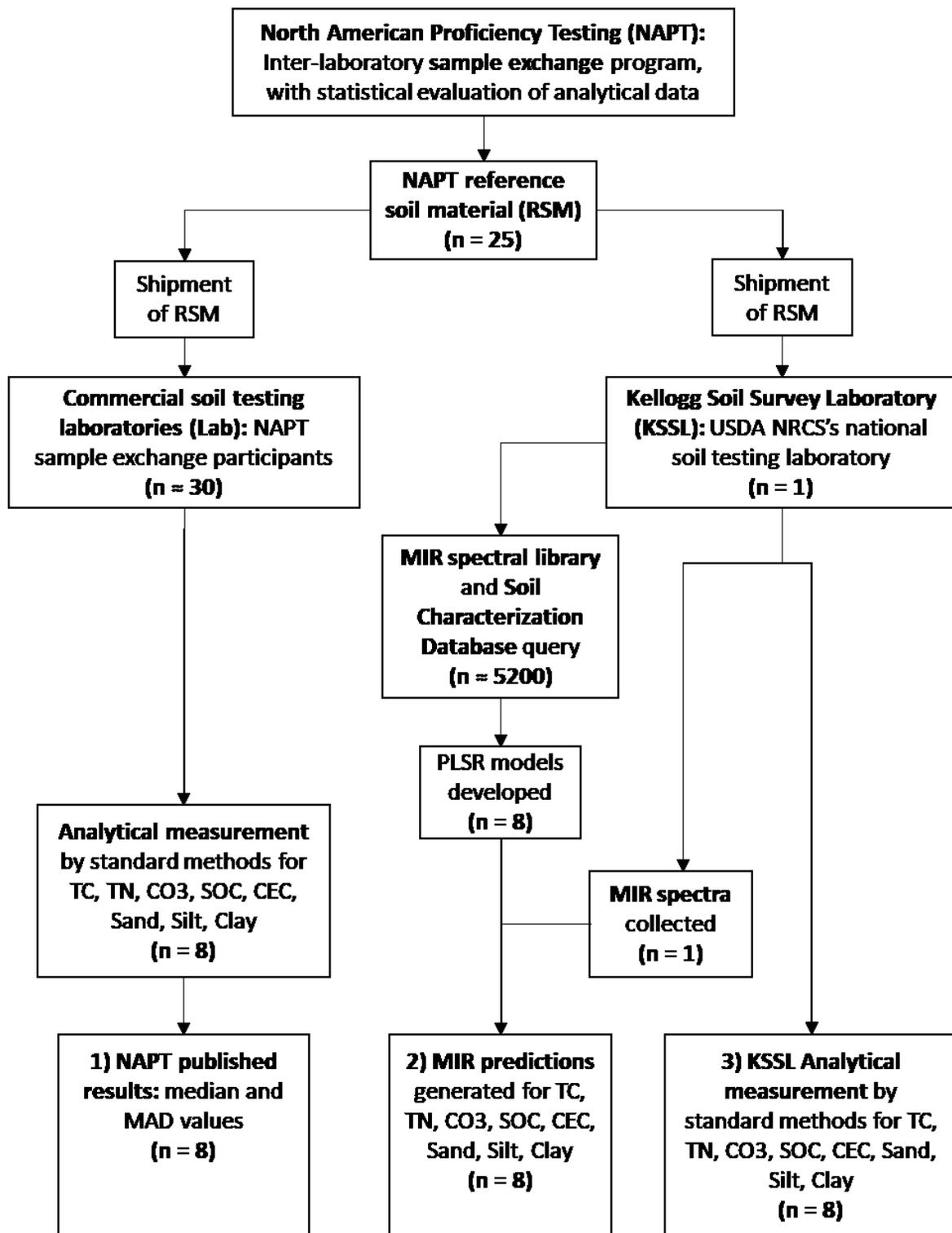


Fig. 1. Flow chart outlining the steps for generating the three evaluation datasets used to estimate method errors: 1) NAPT published results, 2) MIR predicted values, and 3) KSSL measured values. Measured and predicted soil properties included total C (TC), total N (TN), carbonates (CO₃), soil organic carbon (SOC), cation exchange capacity (CEC), and Sand, Silt, and Clay content.

MIR predictions

Soil properties were predicted by partial least squares regression (PLSR) from MIR spectra. The MIR spectral and soil attribute data used in model development were obtained from the KSSL Soil Characterization database. The approximately 5200 soil samples included in this study were collected from throughout the United States, and captured the typical range of geographic and taxonomic diversity observed in U.S. Mollisols. These data were initially collected for various NRCS and National Cooperative Soil Survey projects. For a detailed description of spectral preprocessing and calibration methods used in the development of individual models, see Sherpa et al. 2019.

NAPT simulated values

NAPT simulated values were used to estimate the R^2 , RMSE, and bias associated with measurements performed by North American soil testing laboratories, so that the accuracy of MIR predictions, KSSL measurements, and standard soil testing laboratories could be directly compared. Random observations were generated for each soil property and reference soil material, where simulated values were constrained to a normal distribution with mean and standard deviation equal to the NAPT published median and MAD. However, because NAPT removes outliers prior to publishing laboratory performance results and setting control limits, simulated estimates for laboratory measurement error were optimistic, with actual errors likely to be higher.

Accuracy of MIR compared to standard analytical methods

The accuracy of MIR predicted soil properties were evaluated using a range of typical chemometric indicators (Bellon-Maurel et al., 2010). The measured values ($Y_{meas.}$) and predicted soil property values ($Y_{pred.}$) for each property were used to obtain the coefficient of determination (R^2), and the root mean square error of prediction (RMSE) Eq. [1]. The RMSE provides an estimate of the average prediction error for MIR models. The systematic error in predicted soil properties was evaluated based on the error of means, where the mean of the predicted values ($\bar{Y}_{pred.}$) and the mean of the measured values ($\bar{Y}_{meas.}$) were used to estimate bias Eq. [2]. In order to explore different sources of error contributing to MIR predictions, measured values obtained by KSSL were also evaluated using chemometric indicators, where ($\bar{Y}_{pred.}$) was equal to soil properties by KSSL analytical measurement. The median absolute deviation (MAD) was used to estimate measurement errors associated with soil testing laboratories participating in the NAPT program. This value reflects measurement variability among laboratories for a given reference soil material and measured soil property.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_{pred.} - Y_{meas.})^2}{n}} \quad \text{Eq. [1]}$$

$$\text{Bias} = \bar{Y}_{pred.} - \bar{Y}_{meas.} \quad \text{Eq. [2]}$$

Cost of MIR compared to standard analytical methods

Estimates for the cost incurred by soil testing laboratories was obtained from KSSL, personal communication. Estimates for the cost charged by soil testing laboratories represent the average cost charged by the UC Davis Analytical Laboratory and Cornell Nutrient Analysis Laboratory (<https://anlab.ucdavis.edu/>; <https://cnal.cals.cornell.edu/>).

Economic optimization and power analysis for verifying SOC

Power analysis allows researchers to determine the sample size required for detecting population differences at a given statistical power, and is essential for estimating the cost associated with monitoring SOC levels and verifying change in response to management (Kravchenko and Robertson, 2011). Phillips and Jiang (2016) demonstrated that measurement error affected the power of statistical tests, and when present in the data but not accounted for, that sample size requirements determined by power analysis would be underestimated. The approaches presented by Phillips and Jiang (2016) have large ramifications for assessing MIR's cost-effectiveness and determining the economic optimum strategy for SOC verification, allowing sampling plans to be optimized for detecting SOC change based on measurement cost, measurement error, and site variability. The statistical power of SOC estimates based on MIR predictions and NAPT published results were adjusted to account for measurement error according to Phillips and Jiang (2016), and the analytical budgets required for quantifying differences in SOC were compared by power analysis. An R^2 value was calculated for NAPT published results, where median and MAD values were used in sum of squares calculations. R^2 values associated with MIR predictions and NAPT participating labs were then used to adjust the effect size of power calculations. Estimates for SOC site variability used in power analysis were taken from Sherpa et al. (2016), where strategies for SOC sampling optimization at the agroecosystem scale were compared for different land uses. An SOC standard deviation (SD) of 0.52 and 1.10 (%) SOC were used to compare power analysis results under scenarios of low spatial variability and high spatial variability, respectively.

Results and Discussion

Evaluation datasets

The NAPT reference soil materials (n=25 independent soil samples) displayed a wide range of values and large standard deviation for all soil properties measured, providing strong evidence for the robustness of the evaluation datasets used in this study (Table 2). SOC based on the NAPT published median ranged from 0.38-5.85%, sufficiently encompassing the range of values likely required for agricultural C verification programs (Kern, 1994). The slight difference in range observed for total C (TC; 0.42-5.91%) and SOC is due to measurement error associated with carbonate (CO₃) determination, as SOC was determined by difference from TC and CO₃.

Descriptive statistics for the three evaluation datasets: 1) NAPT published median values, 2) KSSL measured values, and 3) MIR predicted values, were all in close agreement (Table 1). Calibration models for MIR were developed using spectral and soil attribute data obtained from the NRCS KSSL Soil Characterization Database, and so any systematic errors present in the KSSL analytical data will likely be reflected in MIR predictions. The wide range of SOC concentration and soil texture included in the evaluation datasets provided a thorough representation of U.S. agricultural soils from the Mollisols Order (Liu et al., 2012), with the goal that findings from this study would be directly applicable to researchers and policy makers working within the large geographic region of immense agronomic importance, encompassed by these soils (21 % of U.S. land area; 200 million ha).

Table 2. Descriptive statistics of the evaluation dataset (n= 25), based on three separate measurement estimates for each soil property.

	NAPT†				KSSL‡				MIR§			
	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
Total C (%)	2.38	1.58	0.42	5.91	2.47	1.67	0.38	6.34	2.51	1.64	0.50	6.65
Total N (%)	0.20	0.11	0.04	0.54	0.23	0.13	0.03	0.60	0.20	0.10	0.06	0.52
SOC (%)	2.04	1.29	0.38	5.85	2.09	1.38	0.38	6.34	2.13	1.37	0.50	6.65
CO3 (%)	3.5	6.2	0.3	22.0	3.2	6.9	0.0	23.8	3.2	6.8	0.0	22.9
Sand (%)	34	24	10	84	29	28	3	87	29	28	-1	98
Silt (%)	41	17	9	64	45	21	6	75	48	17	5	75
Clay (%)	25	16	4	54	26	18	2	57	22	14	0	46
CEC (cmol kg ⁻¹)	21	7	3	35	19	7	3	29	21	11	-5	47

† Results are based on the published NAPT quarterly reports.

‡ Results are based on laboratory measurements performed at KSSL.

§ Results are based on MIR predictions.

Accuracy of MIR compared to standard analytical methods

This section first documents the typical accuracy of North American soil testing laboratories, to provide a frame of reference for evaluating the performance of MIR predictions next, and finally explores various sources of error and their contribution to the accuracy of MIR predictions.

Accuracy of North American soil testing laboratories

All of the standard soil analytical methods evaluated displayed some degree of measurement error, with measurements for CEC, texture, and CO₃ tending to be more prone to measurement error, as was evident by the larger MAD values reported by laboratories participating in the NAPT program (Fig. 2). TC and TN displayed the lowest measurement error, though errors for TN were considerably greater than TC (Fig. 2a, 2b). As noted previously, SOC was determined by difference from TC and CO₃ values, therefore SOC measurement errors were primarily attributed to soil carbonate measurements and the challenge of accurately quantifying inorganic C in soils. Measurement errors associated with individual reference soil materials were highly variable, that is, some samples were more prone to error than others were, though trends in the magnitude of error were not consistent among soil properties. For example, sand content displayed a negative relationship with MAD, where the measurement errors tended to decrease with increasing sand content, while the opposite relationship was observed for CEC and clay content (Fig. 2e, 2g, 2h). The NAPT reference soil materials and published results provide a unique perspective regarding the relative accuracy of “gold-standard” methods for soil analysis, and allow for the first time, the performance of MIR predicted soil properties to be framed within the context of typical soil measurement errors.

Accuracy of MIR predicted soil properties

MIR produced excellent predictions for TC, CO₃, and SOC, explaining 98-99% of the variability in NAPT published median values, and for all intents and purposes displayed analytical accuracies equal to that of NAPT participating soil testing laboratories (Fig. 2). When CO₃ measurements were used in the determination of SOC, MIR achieved a higher accuracy than the standard method, with predicted values clustered tight to the regression line relative to measurement error bars representing the control limits set by NAPT for individual reference soils (Fig. 2a, 2c, 2d). The NAPT serves to “provide the USDA NRCS with a means to assess laboratories for NRCS related work” (NAPT 2018). Laboratory performance scores reflect the percent of measurements falling within NAPT defined control limits for individual reference materials and soil properties.

MIR predicted values passed NAPT certification requirements of 60% method performance for all properties but CO₃, sand, and silt (Table 3). Performance improved substantially when predictions were evaluated relative to laboratory measurements performed by KSSL, increasing total performance score from 65% to 71%. A performance score of 80% is required for NAPT certification, and so MIR predictions did not meet NAPT requirements. Due to the small number of soil properties evaluated (n = 8) total performance score was strongly influenced by a single low score, and even KSSL passed by only a small margin with an 82% total performance score.

As was observed in the measured values reported by NAPT, predictions by MIR for CEC and texture tended to have higher errors than predictions for TC and TN, suggesting that individual reference soil materials possess an inherent variability and propensity for measurement errors, which is reflected in both MIR predictions and standard analytical

measurements. The observed trends in method performance have important implications for interpreting the relative accuracy of MIR predictions, as statistical methods for model assessment assume that evaluation samples represent the “true” value, with a measurement error equal to zero. The large measurement error (i.e. MAD) observed for some NAPT published results would typically remain undetected, as analytical error associated with evaluation samples cannot be separated from prediction error, giving rise to the “chicken or the egg” dilemma currently faced by the soil spectroscopy community.

Exploring sources of error in MIR predictions

The contribution from individual bias components to total error had a strong influence on the potential interpretation of results from MIR predictions, where evaluation datasets characterized by high measurement error produced misleading results in some cases. Approaches for assessing the quality of MIR predictions have largely relied on R^2 , RMSE, and to a lesser extent bias, where accuracy is inferred from RMSE and R^2 , and bias measures the individual contributions from systematic and random errors. Given sufficient replication, bias can be further decomposed into individual components for method bias, laboratory bias, and sample bias. Conventional indicators for assessing the quality of MIR predictions such as RMSE and R^2 , do not account for measurement errors associated with the evaluation samples, and may therefore be inadequate for assessing MIR’s relative performance given the large analytical errors typical of many soil analyses.

Performance scores presented in Table 3, represent the percentage of observations landing within control limits unique to each sample, which when combined with R^2 , allowed for the accuracy of MIR predictions to be evaluated within the context of typical measurement errors

for soil analysis. For example, evaluating MIR's performance based on R^2 , would suggest that CEC and clay were predicted somewhat poorly, with low R^2 values of 0.83 and 0.85, respectively. However, when viewed in the context of performance scores, CEC and clay were predicted with greater accuracy than any other soil property, achieving scores of 88% and 76%, respectively (Fig. 2g, 2h; Table 3). Similarly, at an R^2 of 0.91, typical interpretation would suggest that MIR predictions for sand were more accurate than for other size fractions, yet with a performance score of 52%, values would not pass NAPT certification requirements.

The high level of data-dimensionality achieved by integrating the NAPT published dataset with typical approaches to model evaluation allowed for further decomposition of systematic errors into separate components of bias for 1) measurement error due to MIR predictions and 2) measurement error due to analytical values used in model evaluation (Table 3). In general, bias in KSSL measurements relative to the NAPT published median ($KSSL \sim NAPT$) could be combined with MIR bias relative to KSSL measurements ($MIR \sim KSSL$) to obtain bias for MIR relative to NAPT ($MIR \sim NAPT$), such that $(KSSL \sim NAPT) + (MIR \sim KSSL) = (MIR \sim NAPT)$. When evaluated relative to NAPT published results, systematic errors in MIR predictions were largely associated with errors in the evaluation dataset, such as differences in the methodology applied by KSSL and NAPT, rather than due to the predictions themselves (Table 3).

Table 3. Summary of evaluation statistics for soil properties where values were obtained by MIR prediction, standard analytical methods performed by KSSL, and simulated measurements based on the NAPT published median and MAD values. NAPT performance scores (score) and measurement errors are presented relative to NAPT published median values, and relative to KSSL measured values.

	relative to NAPT published median									relative to KSSL		
	MIR			KSSL			NAPT‡			MIR		
	RMSE	bias	score†	RMSE	bias	score	RMSE	bias	score	RMSE	bias	score
TC	0.24	0.13	68%	0.14	0.09	100%	0.08	0.00	100%	0.18	0.04	68%
TN	0.03	0.01	76%	0.04	0.04	44%	0.01	0.00	100%	0.06	-0.03	60%
CO3	0.86	-0.30	44%	0.76	-0.31	40%	0.68	0.00	100%	0.45	0.01	96%
SOC	0.25	0.10	68%	0.16	0.06	96%	0.28	0.02	100%	0.17	0.04	76%
CEC	4.75	0.35	88%	5.30	-1.59	96%	4.10	0.60	100%	7.48	1.94	88%
Sand	9.81	-4.96	52%	6.78	-4.47	72%	4.02	0.72	100%	9.28	-0.49	64%
Silt	14.25	7.16	48%	5.77	3.70	88%	3.09	-0.06	100%	14.65	3.46	48%
Clay	6.58	-2.65	76%	2.62	1.27	100%	3.50	0.11	100%	8.03	-3.93	72%

† Scores reflect the percentage of measurements/predictions falling with the control limits set by NAPT for individual reference materials and soil properties. Control limits set to +/- 2.5 MADs or 10% of the NAPT published median.

‡ Evaluation statistics for simulated measurements, where simulated values were assumed a normal distribution with mean and standard deviation equal to the NAPT published median and MAD.

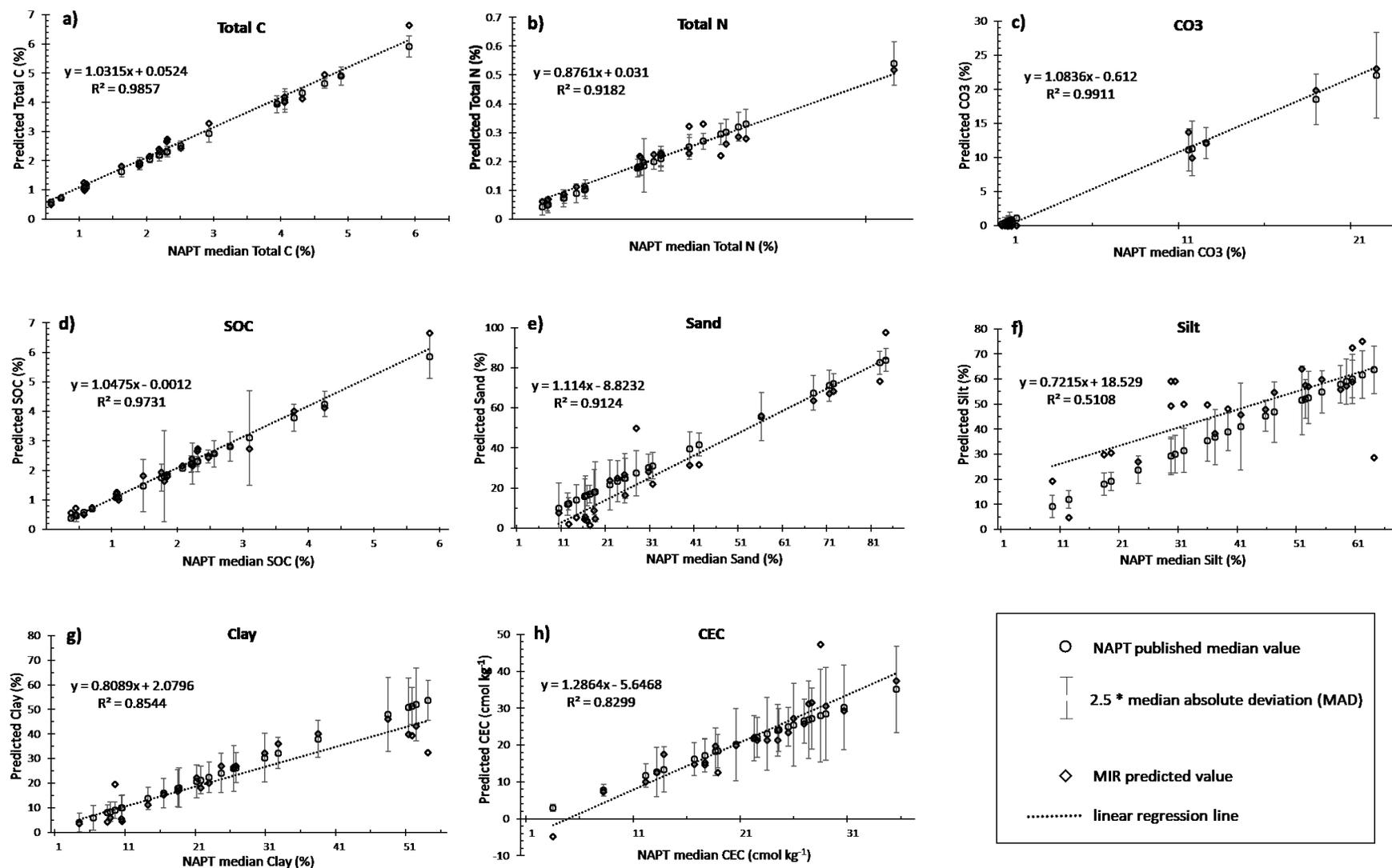


Fig. 2. MIR predicted value versus published NAPT median value for **a)** Total C, **b)** Total N, **c)** carbonates (CO₃), **d)** soil organic C (SOC), **e)** Sand, **f)** Silt, **g)** Clay, and **h)** cation exchange capacity (CEC). The published NAPT median value with error bars of + or - 2.5 median absolute deviations (MAD) is also shown.

Evaluating MIR predictions relative to KSSL measured values tended to improve model performance compared to error statistics computed using NAPT published median values. KSSL measured values reflected the highest level of laboratory quality assurance (QA)/quality control (QC) among the evaluation data sets, and were used in both model calibration and model evaluation (Fig. 1; Table 3). Comparing MIR predicted values relative to KSSL measured values reduced sources of bias such as method discrepancies and inter-laboratory variability. The influence of method error on bias was most apparent for sand, silt, and clay; in this case, KSSL and NAPT use different methods for texture, which is likely the main source of bias (Table 1).

Evaluating the cost-effectiveness of MIR

Predicting soil properties by MIR provided substantial cost savings, with some predicted soil properties displaying accuracies comparable to standard soil testing laboratories. Error statistics derived from NAPT simulated values were similar in magnitude to that of MIR predictions for SOC, CEC, clay, and CO₃ (Table 3). NAPT removes outliers prior to publishing results and so actual errors for standard soil testing laboratories are likely to be higher. At \$1.80 per sample, MIR analysis had the lowest cost of any measurements performed by KSSL, and was capable of simultaneously predicting all 8 soil properties (Table 4). Measurement costs varied greatly among soil properties, both in terms of the average cost charged by commercial labs, and the cost incurred by KSSL for performing soil analyses. The economic benefit from MIR increased hugely as additional soil properties were included, such that the cost charged and cost incurred for all 8 soil analyses were roughly 82 and 30 times greater than MIR (Table 4). To the authors' surprise, method cost seemed to increase with method error, this is a noteworthy finding, as the (high-error, high-cost) and (low-error, low-cost) scenario offers an opportunity to

Table 4. Comparison of method costs for individual soil properties, and practical combinations. Estimates are provided for the average cost charged by commercial soil testing laboratories (Lab_c), analytical measurement cost incurred by Kellogg Soil Survey Laboratory (KSSL_c), and the cost of MIR analysis incurred by KSSL (MIR_c).

	Soil Property	Lab_c^\dagger	KSSL_c	MIR_c
Individual soil property cost	TC	17.00	2.46	1.80
	TN	17.00	2.46	1.80
	TC, TN	23.00	2.46	1.80
	CO ₃	24.50	3.60	1.80
	SOC	32.50	6.06	1.80
	CEC	36.00	25.45	1.80
	Sand	62.00	21.98	1.80
	Silt	62.00	21.98	1.80
	Clay	62.00	21.98	1.80
	Texture	62.00	21.98	1.80
	Practical cost	TC, TN, CO ₃ , SOC	50.50	6.06
CEC		36.00	25.45	NA
Sand, Silt, Clay		62.00	21.98	NA
Total Bundle cost		148.5	53.49	1.80

† Average cost of soil analysis charged by Cornell University and UC Davis, soil testing laboratories.

maximize benefits from economic optimization, particularly when estimates of MIR performance account for the measurement errors associated with standard methods.

The cost incurred for CEC and soil texture were 14 and 12 times greater than MIR, yet both CEC and texture displayed some of the highest analytical measurement errors. MIR predictions for CEC and Clay on the other hand, achieved the highest NAPT performance scores of any predicted soil property. Although not included in the subsequent section on economic optimization for SOC verification, incorporating low-cost soil texture and CEC measurements into SOC-based GHG mitigation programs would be valuable. Soil texture and clay mineralogy play an important role in SOC dynamics, and clay content is believed to be the main factor influencing SOC accumulation in U.S. Mollisols (Kern, 1994). MIR predicted texture could serve as a covariate for SOC prediction, informing models, optimizing sampling designs, and setting expected rates for SOC change. In addition, CEC predictions would allow for quantification of co-benefits such as improved ecosystem function through enhanced nutrient cycling and storage.

Economic optimization and power analysis for verifying SOC

Based on the results from 25 NAPT reference soil samples, MIR served as a low-cost accurate tool for quantifying SOC in U.S. Mollisols, as the cost incurred for determining SOC by wet chemistry was nearly three and a half times greater than prediction by MIR, while the measurement accuracy gained was minimal to none (Table 3; Table 4). MIR successfully predicted TC, CO₃, and SOC with an analytical level of accuracy comparable to typical North American soil testing laboratories.

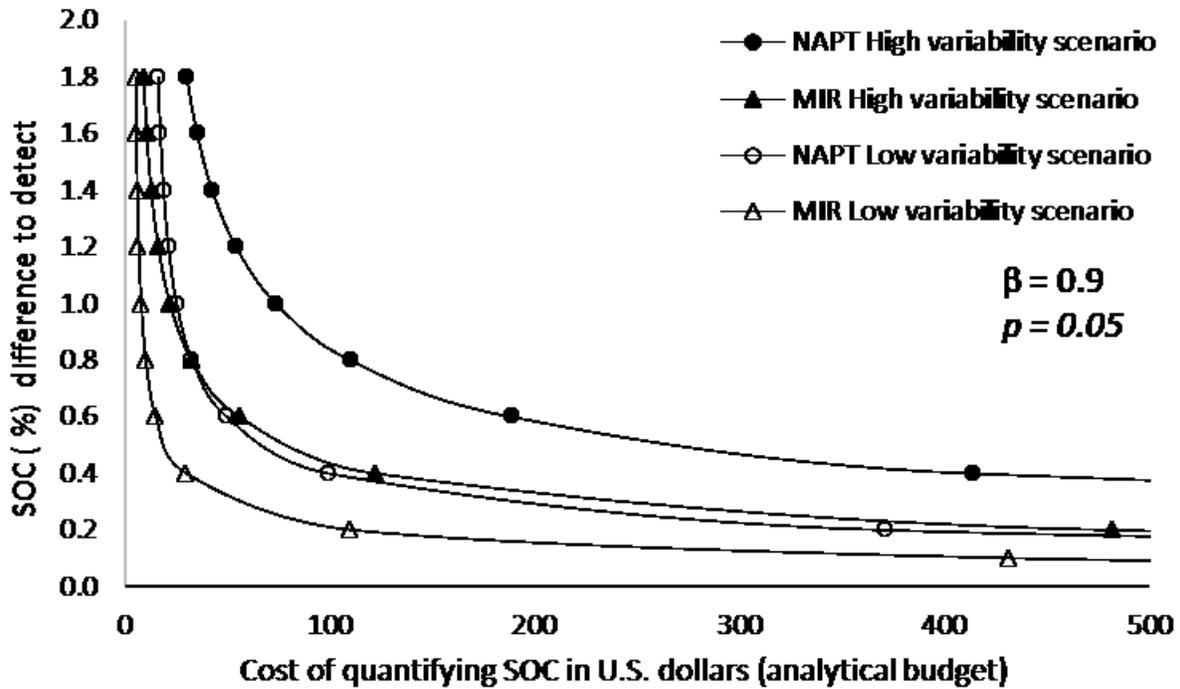


Fig. 3 Power analysis for estimating the minimum detectable difference in SOC based on analytical budget, where effect size was adjusted for measurement errors associated with MIR predictions or NAPT soil testing laboratories. Results are presented for two scenarios, sites with high SOC variability ($SD=1.10$) and sites with low SOC variability ($SD=0.52$). High and low estimates of variability, based on observations by Sherpa et al. (2016) for SOC to a depth of 30cm among different land uses of a complex agroecosystem. Cost of analytical measurements reflect the costs incurred by KSSL.

The cost-effectiveness of MIR for verifying SOC was dependent on analytical budget and SOC variability, with cost savings decreasing with increasing analytical budget. Additionally, the cost savings achieved by MIR increased with SOC variability of a site, where the economic benefit from MIR effectively doubled as SOC standard deviation increased from 0.52 to 1.10, (Fig. 3). For example, given an analytical budget of \$100 USD, under the low variability scenario, replacing standard analytical methods with MIR predictions improved the minimum detectable difference (MDD) by 0.2 % SOC, while MDD improved by 0.4 % SOC under the high variability scenario (Fig. 3). Power calculations used in Fig. 3, were adjusted so that effect of size accounted for measurement errors, but adjustments were negligible because of MIR's low error.

Conclusions

The results presented in this study provide a frame of reference for interpreting the relative magnitude of cost and error associated with both MIR predictions and standard soil testing laboratories. This overview of options for soil analysis reflect a gradient of cost-error relationships, and can be used to optimize future planning and design of sampling schemes for quantifying SOC based on analytical budget, SOC difference to detect, and site variability.

MIR predicted TC, CO₃, and SOC displayed an analytical level of accuracy comparable to that of North American soil testing laboratories, while the cost of quantifying SOC by standard laboratory measurement was three and half times greater than MIR. Based on these findings, MIR shows great promise as low-cost accurate tool for verifying SOC in U.S. Mollisol

soils.

Performance and accuracy are relative terms, defined based on the distance of an observation from some reference point. Current approaches for evaluating model performance do not account for uncertainty in the reference point's location; yet large measurement errors in the reference method can have a strong influence on estimates of model performance. Incorporating reference soil materials into the regular evaluation of spectroscopic approaches, as well as the development of performance metrics capable of accounting for measurement errors in the initial reference values, will hasten the widespread adoption of this technology.

REFERENCES

- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC - Trends in Analytical Chemistry* 29: 1073–1081. doi:10.1016/j.trac.2010.05.006
- Cohen, J. 1992. A Power Primer. *Quant. methods Psychol.* 112: 155–159.
- Conant, R.T., S.M. Ogle, E.A. Paul, and K. Paustian. 2011. Measuring and monitoring soil organic carbon stocks in agricultural lands for climate mitigation. *Front. Ecol. Environ.* 9(3): 169–173.
- Dhankher, O.P., and C.H. Foyer. 2018. Climate resilient crops for improving global food security and safety. *Plant Cell Environ.* 41(5): 877–884. doi: 10.1111/pce.13207.
- FAO. 2015. Intergovernmental Technical Panel on Soils. *Status of the World's Soil Resources.*

- Glenk, K., and S. Colombo. 2011. Designing policies to mitigate the agricultural contribution to climate change: an assessment of soil based carbon sequestration and its ancillary effects. *Clim. Change*. 105: 43–66. doi: 10.1007/s10584-010-9885-7.
- Godfray, H.C.J., J.R. Beddington, I.R. Crute, L. Haddad, D. Lawrence, J.F. Muir, J. Pretty, S. Robinson, S.M. Thomas, and C. Toulmin. 2010. Food security : The challenge of feeding 9 billion people. *Science*. 327: 812–819. doi: 10.1126/science.1185383.
- Griscom, B.W., J. Adams, P.W. Ellis, R.A. Houghton, G. Lomax, D.A. Miteva, W.H. Schlesinger, D. Shoch, J. V. Siikamäki, P. Smith, P. Woodbury, C. Zganjar, A. Blackman, J. Campari, R.T. Conant, C. Delgado, P. Elias, T. Gopalakrishna, M.R. Hamsik, M. Herrero, J. Kiesecker, E. Landis, L. Laestadius, S.M. Leavitt, S. Minnemeyer, S. Polasky, P. Potapov, F.E. Putz, J. Sanderman, M. Silvius, E. Wollenberg, and J. Fargione. 2017. Natural climate solutions. *Proc. Natl. Acad. Sci.* 114(44): 11645–11650. doi: 10.1073/pnas.1710465114.
- L. Hatfield, Jerry & Sauer, Thomas & M. Cruse, Richard. (2017). Soil: The Forgotten Piece of the Water, Food, Energy Nexus. *Advances in Agronomy*. 10.1016/bs.agron.2017.02.001.
- Kahiluoto, H., P. Smith, D. Moran, and J.E. Olesen. 2014. Enabling food security by verifying agricultural carbon. *Nat. Publ. Gr.* 4(5): 309–311. doi: 10.1038/nclimate2209.
- Kern, J.S. 1994. Spatial patterns of soil organic carbon in the contiguous United States. *Soil Sci. Soc. Am. J.* 58(2): 439–455.
- Kravchenko, A.N., and G.P. Robertson. 2011. Whole-profile soil carbon stocks : the danger of assuming too much from analyses of too little. *Soil Sci. Soc. Am. J.* 75(1): 235–240. doi: 10.2136/sssaj2010.0076.
- Lal, R. 2004. Soil carbon sequestration to mitigate climate change. *Geoderma* 123: 1–22. doi: 10.1016/j.geoderma.2004.01.032.

Liu, X., C. Lee Burras, Y.S. Kravchenko, A. Duran, T. Huffman, H. Morras, G. Studdert, X.

Zhang, R.M. Cruse, and X. Yuan. 2012. Overview of Mollisols in the world: Distribution, land use and management. *Can. J. Soil Sci.* 92(3): 383–402. doi: 10.4141/cjss2010-058.

Nocita, Marco & Stevens, Antoine & Wesemael, Bas & Aitkenhead, Matthew & Bachmann, Martin & Barthès, B & Ben-Dor, Eyal & Brown, David & Clairotte, Michael & Csorba, Adam & Dardenne, Pierre & Demattê, José Alexandre & Genot, Valérie & Guerrero, Cesar & Knadel, Maria & Montanarella, Luca & Noon, Carole & Ramirez-Lopez, Leonardo & Robertson, Jean & Wetterlind, Johanna. (2015). Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. *Advances in Agronomy*. 10.1016/bs.agron.2015.02.002.

O'Rourke, S.M., and N.M. Holden. 2011. Optical sensing and chemometric analysis of soil organic carbon - a cost effective alternative to conventional laboratory methods? *Soil Use Manag.* 27(2): 143–155. doi: 10.1111/j.1475-2743.2011.00337.x.