


8-2018

# A Comparison of VNIR and MIR Spectroscopy for Predicting Various Soil Properties

Joshua R. Gates

*University of Nebraska-Lincoln*, [jgates@huskers.unl.edu](mailto:jgates@huskers.unl.edu)

Follow this and additional works at: <http://digitalcommons.unl.edu/natresdiss>

 Part of the [Hydrology Commons](#), [Natural Resources and Conservation Commons](#), [Natural Resources Management and Policy Commons](#), [Other Environmental Sciences Commons](#), and the [Water Resource Management Commons](#)

---

Gates, Joshua R., "A Comparison of VNIR and MIR Spectroscopy for Predicting Various Soil Properties" (2018). *Dissertations & Theses in Natural Resources*. 264.

<http://digitalcommons.unl.edu/natresdiss/264>

This Article is brought to you for free and open access by the Natural Resources, School of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Dissertations & Theses in Natural Resources by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

A COMPARISON OF VNIR AND MIR SPECTROSCOPY FOR PREDICTING  
VARIOUS SOIL PROPERTIES

by

Joshua R. Gates

A THESIS

Presented to the Faculty of  
The Graduate College at the University of Nebraska  
In Partial Fulfillment of Requirements  
For the Degree of Master of Science

Major: Natural Resource Sciences

Under the Supervision of Professor Paul Hanson

Lincoln, Nebraska

August, 2018

# A COMPARISON OF VNIR AND MIR SPECTROSCOPY FOR PREDICTING VARIOUS SOIL PROPERTIES

Joshua R. Gates, M.S.

University of Nebraska, 2018

Advisor: Paul Hanson

Soil plays an important role in our daily lives, namely producing food, cleaning water and storing carbon. The ability to rapidly and cost-effectively quantify the various components of soils can help us understand and better manage this important resource. This study aims to compare the ability of visible near-infrared (VNIR) spectroscopy and mid-infrared (MIR) spectroscopy to quickly and accurately predict various important soil properties (electrical conductivity, soil pH, cation exchange capacity, exchangeable cations, phosphorus, carbon, beta-glucosidase enzyme activity and nitrogen). Prediction models were developed using partial least squares regression (PLSR) techniques. Three different calibration sampling methods were tested along with various spectral preprocessing techniques to find the best predictive ability of VNIR and MIR. Soil components related to carbon, nitrogen and cation exchange capacity had good predictive ability ( $R^2 > 0.8$ ) by both VNIR and MIR, but MIR was more accurate. Electrical conductivity, sodium cations and phosphorus were poorly predicted by both ( $<0.71$ ). VNIR models were not as robust as MIR models but could be potentially useful for qualitative analyses when rapid analyses are preferred over methods that are more accurate. MIR predictions overall yielded more accurate predictions than VNIR and could potentially be used as a surrogate method for timely laboratory techniques for spectrally active soil components.

## Table of Contents

List of Figures .....	vi
List of Tables .....	vi
List of Equations .....	vii
<b>CHAPTER 1. INTRODUCTION .....</b>	<b>1</b>
1.1 Purpose of Study .....	1
1.2 Objectives .....	2
<b>CHAPTER 2. BACKGROUND &amp; LITERATURE REVIEW .....</b>	<b>3</b>
2.1 Introduction .....	3
2.2 Electromagnetic Radiation and Matter .....	4
2.3 Spectral Preprocessing .....	8
2.4 Chemometrics .....	10
2.5 Predicting Soil Properties .....	13
2.5.1 Electrical Conductivity .....	14
2.5.2 Soil pH .....	15
2.5.4 Cation Exchange Capacity and Exchangeable Cations .....	18
2.5.5 Carbon .....	19
2.5.6 $\beta$ -Glucosidase .....	22
2.5.7 Nitrogen .....	22
2.6 Summary .....	23
<b>CHAPTER 3. MATERIALS &amp; METHODS .....</b>	<b>25</b>
3.1 Study Area .....	25
3.2 Reference Data Collection .....	26
3.2.1 Air-dry Oven-dry Ratio .....	26
3.2.2 Electrical Conductivity .....	26
3.2.3 Soil pH .....	27
3.2.4 Water Soluble Phosphorus .....	28
3.2.5 Cation Exchange Capacity .....	30
3.2.6 Extractable Bases .....	32
3.2.7 Calcium Carbonate .....	36
3.2.9 Total Phosphorus .....	39
3.2.10 Soil Organic Carbon .....	41
3.2.11 Particle Size Distribution Analysis .....	42

	iv
3.3 Collection of Spectra.....	43
3.3.1 Vis-Near Infrared Spectroscopy .....	43
3.3.2 Mid-Infrared Spectroscopy .....	46
3.4 Statistical Analyses .....	47
3.4.1 Spectral Preprocessing.....	47
3.4.2 Model Creation and Validation.....	48
3.4.3 Model Performance and Selection.....	49
CHAPTER 4. RESULTS .....	51
4.1. Basic Soil Properties .....	51
4.1.1 Electrical Conductivity .....	57
4.1.2 Soil pH .....	57
4.1.3 Cation Exchange Capacity and Exchangeable Cations .....	59
4.1.4 Phosphorus.....	64
4.1.5 Carbon.....	66
4.1.6 $\beta$ -Glucosidase.....	69
4.1.7 Nitrogen .....	70
4.2 Calibration Sampling .....	71
4.2.1 Random Sampling.....	71
4.2.2 Stratified Sampling .....	74
4.2.3 Kennard Stone Sampling .....	77
CHAPTER 5. DISCUSSION & CONCLUSION.....	79
5.1 General Soil Properties .....	79
5.1.1 Electrical Conductivity .....	79
5.1.2 Soil pH .....	80
5.1.3 Cation Exchange Capacity and Extractable Cations.....	81
5.1.4 Phosphorus.....	82
5.1.5 Carbon.....	83
5.1.6 $\beta$ -Glucosidase.....	84
5.1.7 Total Nitrogen.....	84
5.2 Calibration Sampling .....	85
5.2.1 Random Sampling.....	85
5.2.2 Stratified Sampling .....	86
5.2.3 Kennard-Stone Sampling .....	86
5.3 Conclusions and Further Work .....	87

	v
REFERENCES .....	90
APPENDIX A: R CODE .....	96
APPENDIX B: SOIL REFERENCE DATA .....	112
APPENDIX C: SOIL PROPERTY CORRELATIONS .....	124
APPENDIX D: MODEL SCATTER PLOTS.....	128

## List of Figures

- Figure 2.1 The electromagnetic spectrum (page 5)  
Figure 2.2 A soil absorbance spectrum (page 7)  
Figure 3.1 Study area and location of sites (page 44)  
Figure 3.2 VNIR raw reflectance spectra (page 44)  
Figure 3.3 VNIR detector correction (page 45)  
Figure 3.4 VNIR raw absorbance spectra (46)  
Figure 3.5 MIR raw absorbance spectra (47)  
Figure 4.1 Distribution of measured soil properties (page 50)  
Figure 4.2 Relationship between total C and measured soil properties (page 54)  
Figure 4.3 Relationship between clay percent and measured soil properties (page 55)  
Figure 4.4 Goodness of fit plots for EC (page 57)  
Figure 4.5 Goodness of fit plots for 1:1 H<sub>2</sub>O pH (page 58)  
Figure 4.6 Goodness of fit plots for 1:2 CaCl<sub>2</sub> pH (page 59)  
Figure 4.7 Goodness of fit plots for CEC (page 60)  
Figure 4.8 Goodness of fit plots for Ca (page 61)  
Figure 4.9 Goodness of fit plots for Mg (page 62)  
Figure 4.10 Goodness of fit plots for Na (page 63)  
Figure 4.11 Goodness of fit plots for K (page 64)  
Figure 4.12 Goodness of fit plots for water P (page 65)  
Figure 4.13 Goodness of fit plots for total P (page 66)  
Figure 4.14 Goodness of fit plots for total C (page 67)  
Figure 4.15 Goodness of fit plots for SOC (page 68)  
Figure 4.16 Goodness of fit plots for CaCO<sub>3</sub> (page 69)  
Figure 4.17 Goodness of fit plots for  $\beta$ G (page 70)  
Figure 4.18 Goodness of fit plots for total N (page 71)

## List of Tables

- Table 3.1 Qualitative model performance based on RPD and R<sup>2</sup> (page 50)  
Table 4.1 Descriptive statistics of measured soil properties (page 52)  
Table 4.2 Correlation with spectrally active properties (page 52)  
Table 4.3 Goodness of fit statistics for overall best VNIR models (page 56)  
Table 4.4 Goodness of fit statistics for overall best MIR models (page 56)  
Table 4.5 Goodness of fit statistics for VNIR with random sampling (page 73)  
Table 4.6 Goodness of fit statistics for MIR with random sampling (page 74)  
Table 4.7 Goodness of fit statistics for VNIR with stratified sampling (page 76)  
Table 4.8 Goodness of fit statistics for MIR with stratified sampling (page 76)  
Table 4.9 Goodness of fit statistics for VNIR with Kennard Stone sampling (page 78)  
Table 4.10 Goodness of fit statistics for MIR with Kennard Stone sampling (page 78)

## List of Equations

- Equation 1: Beer's Law (page 6)
- Equation 2: Reflectance (page 6)
- Equation 3: Reflectance to absorbance (page 6)
- Equation 4: MSC correction coefficient estimation (page 8)
- Equation 5: MSC corrected spectrum (page 8)
- Equation 6: SNV corrected spectrum (page 9)
- Equation 7: Detrending 2<sup>nd</sup> order polynomial regression (page 9)
- Equation 8: Detrended spectra (page 9)
- Equation 9: RMSEP (page 12)
- Equation 10: RPD (page 12)
- Equation 11: AD:OD ratio (page 26)
- Equation 12: Absorption to concentration (page 29)
- Equation 13: Extractable P conversion to soil P (page 29)
- Equation 14: CEC (page 32)
- Equation 15: Cation concentration (page 33)
- Equation 16: BG activity (page 36)
- Equation 17: Manometer correction (38)
- Equation 18: Calcium carbonate equivalent (page 38)
- Equation 19: Total carbon % (page 39)
- Equation 20: Total nitrogen % page (39)
- Equation 21: Major element concentration (page 41)
- Equation 22: SOC % (page 42)
- Equation 23: Reflectance to Absorbance (page 44)
- Equation 24: RMSECV (page 49)
- Equation 25: RPD (page 49)
- Equation 26: Bias (page 49)



## CHAPTER 1. INTRODUCTION

### 1.1 Purpose of Study

Soil is a complex and heterogeneous system with many physical, chemical, and biological processes that govern our food and fiber production, water infiltration, contaminant remediation and carbon sequestration. Understanding these processes can be difficult and time-consuming. Conventional analytical techniques often ignore the complex behaviors of soils by trying to draw relationships between physical, chemical, and biological properties that have been analyzed and quantified independently from subsamples and may further complicate our understanding of soils as a system by disrupting the soil equilibrium with chemical extractions (Viscarra Rossel et al., 2006). These techniques are often slow and expensive when cost-effective, and efficient analytical techniques are needed to better quantify our greatest natural resource.

Spectroscopic techniques such as: mass spectroscopy (MS), nuclear magnetic resonance (NMR), ultraviolet (UV), visible (VIS), near-infrared (NIR) and mid-infrared (MIR), may in some cases be an alternative to traditional soil analyses (Janik et al., 1998; Nocita et al., 2015). Great advances have been made in the last few decades to increase the cost-effectiveness, speed, and accuracy of spectroscopic analytical methods (Bellon-Maurel & McBratney, 2011). These methods are able to quickly analyze soil as a whole without the need for hazardous chemicals and have the added benefit of being able to measure multiple soil properties from a single soil sample (Viscarra Rossel et al., 2006). Development of these techniques is crucial for our understanding of soils as a system and for rapidly assessing and quantifying soil properties so that we may best utilize them.

## 1.2 Objectives

The main objective of this thesis is to compare the predictive ability between VNIR and MIR spectroscopy for various soil properties at the regional scale. The soil properties of interest are electrical conductivity (EC), pH, cation exchange capacity (CEC) and extractable cations ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ), water-soluble phosphorus ( $\text{H}_2\text{O P}$ ), total phosphorous (P), total carbon (C), soil organic carbon (SOC), calcium carbonate ( $\text{CaCO}_3$ ),  $\beta$ -glucosidase enzyme activity ( $\beta\text{G}$ ), and total nitrogen (N). A variety of spectral preprocessing techniques were tested to improve spectral signals in order to increase model performance. Additionally, three different model calibration sampling schemes were tested to find the best overall partial least squares regression (PLSR) model for each property.

## CHAPTER 2. BACKGROUND & LITERATURE REVIEW

### 2.1 Introduction

Soil is one of the most important natural resources on Earth. It provides the structure that allows us to produce food, fiber, and fuel that we depend on every day. Soil provides a habitat for microbes that decompose, process and recycle nutrients; filter water and remediate contaminants; soils can sequester carbon from carbon dioxide in the atmosphere (Lal, 2015). Therefore, understanding these complex soil processes and quantifying soil properties is important for ensuring our growing population has access to fresh water and continues to produce adequate food supplies.

Assessments of soil quality require measurements of many soil parameters over large areas to be of use in mitigating threats to soil health and in implementing management policies. Unfortunately, analytical laboratory techniques for measuring soil properties can vary widely between laboratories, hindering the exchange of comparable quantitative information (Nocita et al., 2015). Traditional laboratory techniques for analyzing soil chemical, physical and biological properties can be time-consuming, may utilize hazardous chemicals and are destructive to the sample. Most laboratory methods only measure one soil property at a time and therefore, it may be difficult to make correlations on a system as heterogeneously complex as soil (Stenberg et al., 2010). Spectroscopic techniques allow for the simultaneous characterization of various soil constituents in a cost-effective and timely manner (Viscarra Rossel et al., 2006).

The benefits of spectroscopy are rapid, timely, cost-effective and nondestructive analysis of soil. Soil samples are processed minimally by oven drying, grinding and sieving and no hazardous chemicals are involved. Many samples can be scanned in a

short period of time and the instruments are easy to use and require very little training (Viscarra Rossel et al., 2006), allowing a tremendous amount of information to be quickly gathered about the soil. However, spectroscopic predictions may lack the accuracy of traditional analytical laboratory methods, especially for those properties that are not spectrally active.

Choosing between VNIR or MIR spectrometers usually comes down to cost. VNIR spectrometers utilize less expensive technology, are often smaller, and can be easily accommodated in a backpack for field use (Knadel et al., 2013). However, they tend to be less accurate and have larger prediction errors. Depending on the soil property, this may not be acceptable for precision agricultural practices, but if the goal is to collect large amounts of coarse data, such as environmental monitoring or assessing changes in soil carbon content, the slight increase in accuracy gained from MIR may not be worth the cost (Viscarra Rossel et al., 2006).

## 2.2 Electromagnetic Radiation and Matter

Spectroscopy is the study of how electromagnetic (EM) radiation and matter interact. The electromagnetic (EM) spectrum is composed of the various frequencies of electromagnetic radiation that have energy (photons) as is illustrated in figure 1. EM radiation can be measured by its wavelength (nm or  $\mu\text{m}$ ) or by its wavenumber ( $\text{cm}^{-1}$ ). It is more common to use wavelength when referring to relatively short wavelengths such as visible (400-700 nm) light or near-infrared (700-2500nm) radiation. Spectroscopists commonly switch to wavenumbers when referring to mid-infrared (4000-400  $\text{cm}^{-1}$ ) radiation because it directly relates to the amount of energy of the radiation (energy decreases as wavelength increases) (Viscarra Rossel et al., 2006).

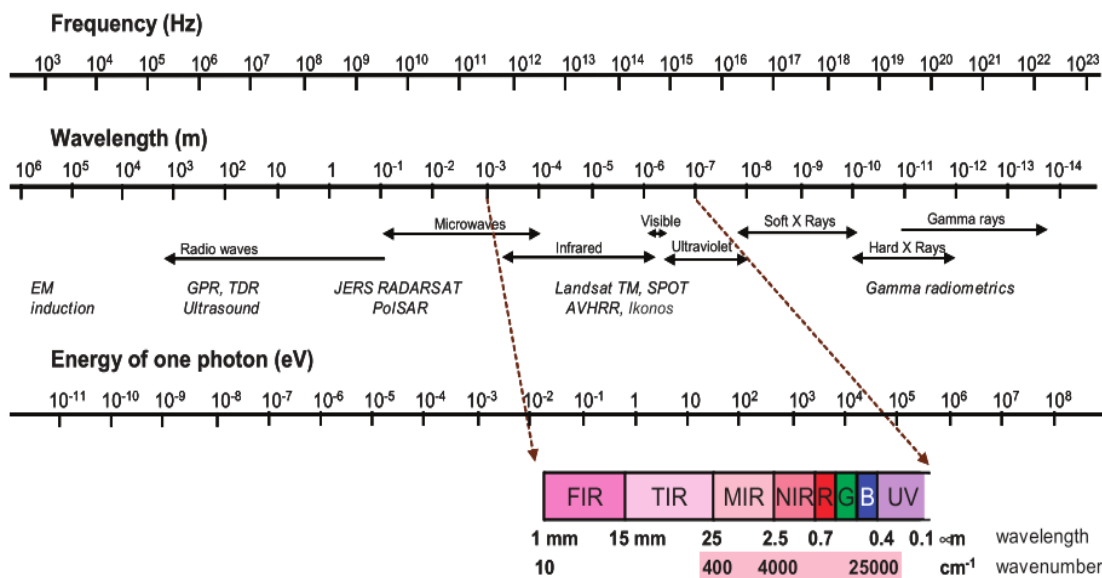


Figure 2.1 The electromagnetic (EM) spectrum highlighting the far infrared (FIR), thermal infrared (TIR), mid-infrared (MIR), near-infrared (NIR), red (R), green (G), blue (B), and ultraviolet (UV) regions (McBratney et al., 2003)

EM radiation interacts with matter in three ways: reflection, transmission, and absorption. In the visible region of the spectrum, the colors you see are examples of reflected radiation, and the colors you do not see are examples of absorbed radiation. Molecules are composed of elements bonded together. These bonds are continuously moving and have specific vibrational modes (stretching and bending) that occur when exposed to visible and infrared radiation. When molecular bonds are exposed to levels of radiation (energy) that match a bond's vibrational mode, the bond will absorb the photon, inducing an excited vibrational state that has energy equal to the absorbed photon (Clark, 2017). The frequency (wavelength or wavenumber) at which this absorption occurs can be used to identify and quantify specific molecules.

Absorbance cannot be directly measured but it can be calculated according to Beer's Law, which states that the intensity of light transmitted through a substance

decreases exponentially as the distance the light travels (path length) increases, concentration of the absorbing substance increases, and how strongly the substance can absorb light (extinction coefficient). More simply stated, absorbance is linearly related to the concentration of the absorbing substance (equation 1):

$$(1) \quad A = -\log\left(\frac{I_o}{I}\right) = \epsilon lc$$

where,  $A$  is absorbance (Au),  $I_o$  is the intensity of transmitted radiation,  $I$  is the intensity of incident radiation,  $\epsilon$  is the extinction coefficient,  $l$  is the path length and  $c$  is the concentration of the substance (Osibanjo et al., (2017). In soil spectroscopy, reflectance is easier to measure than transmittance and is used as a replacement for transmittance, where reflectance is calculated as the ratio of reflected radiation to incident radiation (equation 2) (Bellon-Maurel and McBratney, 2011):

$$(2) \quad R = \frac{R_o}{R_I}$$

where,  $R$  is reflectance,  $R_o$  is the amount of radiation reflected and  $R_I$  is the incident radiation. Using Beer's Law (equation 3), we can calculate absorbance from reflectance (Stenberg et al., 2010):

$$(3) \quad A = \log \frac{1}{R}$$

Vibrational spectroscopy works according to these assumptions, that absorbance is linearly related to the concentration of the substance, and that absorption occurs at specific wavelengths for specific molecular bonds (Clark, 2017). Many organic compounds have distinct absorption features in the MIR region, making MIR spectroscopy a valuable tool. These absorption features occur at the frequency required to excite a molecular bond or functional group to its first excited vibrational state or fundamental vibration. At higher frequencies there is more energy available and the

molecule can be excited to additional vibrational states or overtones. Since energy and frequency are proportional, the first overtone occurs at twice the wavenumber of the fundamental vibration and absorbs approximately twice the energy (Shay and Holmes, 2015). Organic compounds in the VNIR region contain a multitude of absorption bands related to first and second overtones, and combination bands (more than one fundamental vibration occurring simultaneously). These absorption features tend to be broader, less defined and more complex than the fundamental vibrational features in the MIR, making it more difficult to accurately quantify concentrations of the absorbing substance (BenDor et al., 1997; Zornoza et al., 2008; Bellon-Maurel and McBratney, 2011). Figure 2.2 shows an absorbance spectrum of soil.

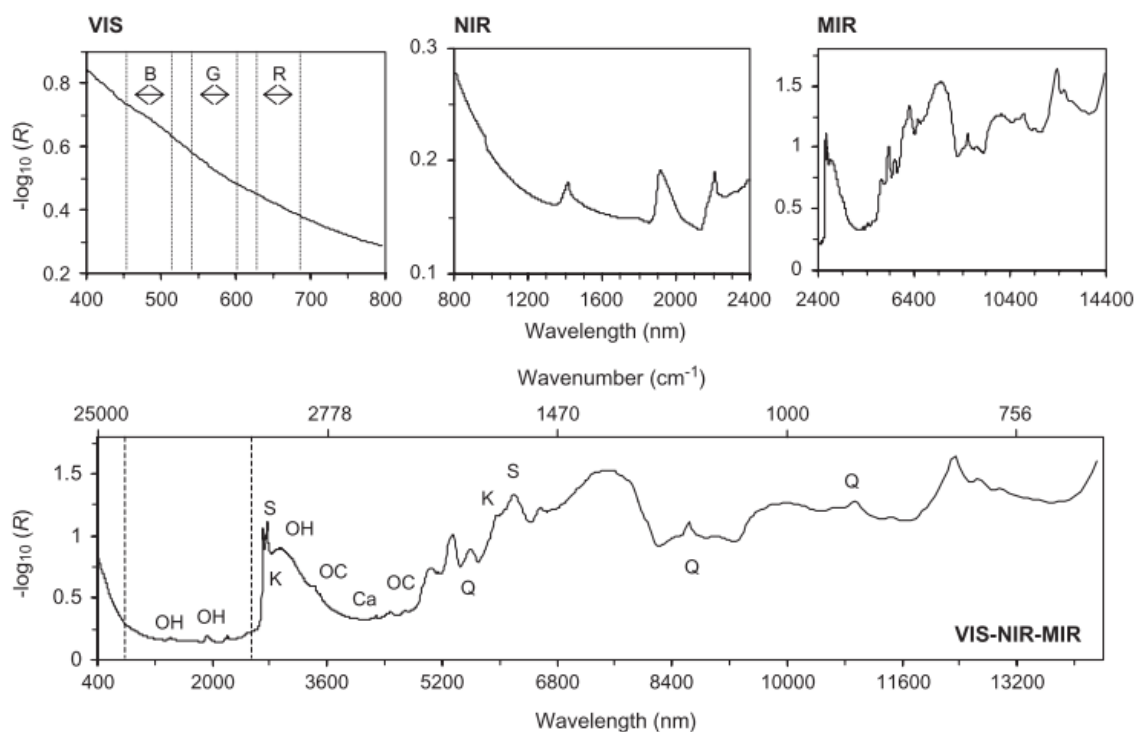


Figure 2.2 Soil spectrum shown in the visible (VIS), near infrared (NIR) and mid-infrared (MIR). Absorption bands of kaolinite (K), quartz (Q), smectite (S), calcite (Ca), organic compounds (OC) and soil free water (OH). Dashed lines in the bottom figure indicate the boundaries between the visible and near infrared and the near infrared and mid-infrared, respectively. (Viscarra Rossel et al., 2006).

### 2.3 Spectral Preprocessing

Variations in the spectra, often caused by the light scattering effects of quartz sand or by instrument drift, sometimes require spectral preprocessing to improve adsorption features by reducing the noise (Rinnan et al., 2009). Spectral preprocessing techniques can be generally divided into three categories: scatter correction, spectral smoothing, and spectral derivatives.

The three commonly used scatter correction techniques are; multiplicative scatter correction (MSC), standard normal variate (SNV), and spectral detrending. MSC is one of the most widely used preprocessing technique to increase signals by reducing the effects of light scattering or path length (Rinnan et al., 2009). The MSC algorithm generates new spectra in two steps, first by estimating the correction coefficients by performing ordinary least squares regression on a spectrum (equation 4):

$$(4) \quad x_i = b_0 + b_1 \cdot x_{ref} + \varepsilon$$

where,  $x_i$  is the  $i$ th uncorrected spectrum,  $x_{ref}$  is the reference spectrum obtained by taking the average of all the spectra by wavelength,  $b_0$  and  $b_1$  are scalar parameters, and  $\varepsilon$  are the residuals. The next step corrects the spectrum with the correction coefficients (equation 5):

$$(5) \quad x_{msc} = \frac{x_i - b_0}{b_1}$$

where,  $x_{msc}$  is the  $i$ th corrected spectrum (Rinnan et al., 2009). SNV, another popular preprocessing technique for scatter correction, has a similar form to MSC (equation 6) but differs from MSC by operating on spectra independently. SNV is performed by centering the spectrum and then dividing by the standard deviation of the spectrum:



$$(6) \quad x_{snv} = \frac{x_i - a_0}{a_1}$$

where,  $x_{snv}$  is the spectrum of corrected values,  $x_i$  is the original spectrum,  $a_0$  is the mean absorbance value of the spectrum and  $a_1$  is the standard deviation of the absorbance values (Rinnan et al., 2009). Detrending removes the mean function by first fitting a 2<sup>nd</sup> order polynomial (equation 7):

$$(7) \quad x_i = a\lambda^2 + b\lambda + c + e_i$$

where,  $x_i$  is the  $i$ th spectrum;  $\lambda$  is the vector of wavelengths;  $a$ ,  $b$ , and  $c$  are coefficients estimated by ordinary least squares and  $e_i$  are the residuals, then the detrended spectrum is found by subtracting the mean function from the spectral response (equation 8):

$$(8) \quad x_i^* = x_i - (a\lambda^2 + b\lambda + c) = e_i$$

where,  $x_i^*$  is the detrended spectrum (Barnes et al., 1989).

Spectral smoothing techniques are also common and involve averaging sections of spectra by a given gap size (Zornoza et al., 2008). Window smoothing averages the values of each section sequentially by overlapping with the previously averaged gap. A second type of smoothing is Savitsky-Golay (SG) smoothing which operates in the same way as window smoothing but averages all the gaps at once and can enhance absorption peaks (Burns and Ciurczak, 2001).

The third category of signal processing, spectral derivatives, have long been used with other spectroscopic techniques to improve signals (Rinnan et al., 2009). A first-order spectral derivative is obtained by finding the rate of change between points determined by the gap size from a smoothed spectrum. A second-order derivative is then obtained by finding the derivatives in a first-order transformed spectrum.

## 2.4 Chemometrics

Quantifying soil properties from MIR and VNIR spectral data requires the use of chemometric techniques. Chemometric techniques aim to extract relevant information from complex systems using mathematical and statistical techniques. For prediction purposes, techniques involve multivariate data analysis and model calibration and validation (Geladi, 2003).

Following preprocessing of spectra, multivariate techniques such as, multiple linear regression (MLR), principal component analysis (PCA), and partial least squares regression (PLSR) are used to analyze spectra. MLR cannot be applied to datasets that are collinear, such as spectra, and require the variables to be reduced, likely by band selection. This technique loses much of the information that may have been present when using the whole spectra (Bertrand et al., 2001).

PCA and PLSR are much better suited to handle spectra because they serve two purposes. They both are used to convert a set of highly correlated variables to a set of independent variables and to reduce data dimensionality by extracting the most important variables (Geladi, 2003; Tinti et al., 2015). PCA decomposes the matrix of predictor variables (spectra) into principal components (latent variables) that maximize the variance among variables. These principle components are independent of each other and can be used as new predictor variables in a regression model. PCA works well when there are more observations than original variables ( $n > k$ ) but is limited when there are more variables than observations, as is usually the case with spectral datasets.

Much like PCA, PLSR decomposes the collinear variables into latent variables but has the benefit of having the response variable annotated to the predictor variables

(the same as in multiple linear regression). PLSR attempts to maximize both the variance between response variables and latent variables (Wold et al., 2001). Because PLSR takes the response variable into account, it is capable of analyzing data with fewer observations than variables. For this reason, PLSR is more useful than PCA for building regression models from smaller datasets (Tinti et al., 2015).

Other multivariate techniques often used for predicting soil properties from spectral data include artificial neural networks (ANN) and regression trees (RF) and have modeled soil properties successfully (Minasny and Mcbratney, 2008; Wijewardane et al., 2016). While these methods are powerful if given enough input data, they can be very complicated and often cannot explain how soil spectra and soil properties are related. Additionally, these methods are prone to overfitting if not properly validated (Martens, 2001).

Any model created from multivariate methods must be validated in order to determine the model's performance, a process commonly done by splitting the data into calibration and validation sets. The calibration set is used to train the model with and the validation set is used to test the model. If the dataset is large, this can typically be accomplished by randomly assigning some independent subset of the data as the training set and the rest to the test set. It is necessary to have enough observations to best describe the relationship between the spectra and response in order to train the model, usually done by randomly assigning two-thirds of the data to the calibration set and the remaining third to the validation set (Estienne et al., 2001). When the number of observations is small, a leave-one-out or jack-knifing approach may be used to increase model "robustness" and has the benefit of easily detecting outliers (Martens et al., 2001). This is

done by training the calibration set on the whole dataset minus a subset of the data and validating the model with the subset. This process is repeated with a new subset that is independent of the first and is continued until every observation has been used to test the calibration set.

The many differences in modeling techniques can make it hard to compare and assess model performance across different studies. The coefficient of determination ( $R^2$ ) can be used to compare the degree of relationship between the observed and predicted values and is often used to test goodness of fit for calibration models. However, this value is dependent on the range of the dataset and should not be the only criteria used for characterizing model performance (Davies & Fearn, 2006). Reeves and Smith (2009) considered threshold values for model performance as: very good ( $R^2 > 0.9$ ), useful ( $R^2$  from 0.7-0.9), and not useful ( $R^2 < 0.7$ ). Janik et al., (1998) suggested that models with high correlation ( $R^2 > 0.9$ ) are a suitable replacement for traditional laboratory techniques and those from 0.7-0.9 may be a useful surrogate for analysis that are time-consuming and costly. The standard error of prediction (SEP) or root mean square error of prediction (RMSEP) are statistical parameters commonly used to describe model prediction performance (Estienne et al., 2001; Bellon-Maurel et al., 2010). RMSEP is computed as the sum of squares of the differences between the predicted and the actual values for the validation/test set (equation 9):

$$(9) \quad RMSEP = \sqrt{\frac{\sum_{i=1}^{n_t} (\hat{y}_i - y_i)^2}{n_t}}$$

where  $n_t$  is the number of samples in the test subset,  $\hat{y}$  is the predicted value and  $y$  is the true value. The RMSEP, like  $R^2$ , is affected by the range of the measured quantity and must be standardized first to allow for comparisons across studies and populations

(Bellon-Maurel et al., 2010). A common way this is done is to divide the standard deviation of the response by the RMSEP (equation 10) (Davies and Fearn, 2006; Bellon-Maurel et al., 2010). This result is the Ratio of Performance to Deviation (RPD), with higher values indicating a better fitted model. Chang et al. (2001) defined threshold values for model reliability as excellent ( $RPD > 2.0$ ), fair ( $1.4 < RPD < 2$ ), and poor ( $RPD < 1.4$ ).

$$(10) \quad RPD = \frac{\sigma}{\sqrt{\frac{\sum_{i=1}^m (y_t - y_i)^2}{m}}}$$

The ratio of performance to inter-quartile distance (RPIQ), is another metric for model comparison proposed by Bellon-Maurel et al. (2010). RPIQ is similar to RPD but is found by dividing the interquartile range (Q3-Q1) by the RMSEP, making it better metric when data are non-normal.

## 2.5 Predicting Soil Properties

Many authors have found that spectrally active soil properties (such as soil organic matter, carbonates, clay minerals and texture) and soil properties that are highly correlated with the spectrally active properties show promise at being predicted with the visible and infrared spectrum (Farmer, 1968; Pirie et al., 2005; Stenberg et al., 2010). VNIR and MIR soil spectra contain useful information that arises from molecular vibrations in soil properties and can be used with chemometric analyses to make predictions about the concentrations of these properties. The VNIR region has spectral features that relate to overtones and combinations of fundamental bond vibrations of organic compounds, soil water, and iron content. The MIR region shows distinct absorption peaks from fundamental vibrations of organic compounds, carbonates quartz,

feldspars, silicates and clay minerals (Chabrilat et al., 2013; Viscarra Rossel et al., 2006; Stenberg et al., 2010).

Accurate and rapid predictions are valuable for producers who use variable-rate fertilization, for high-resolution soil mapping and for those without access to lab facilities (Shepherd and Walsh, 2007; Stenberg et al., 2010). Soil electrical conductivity, pH, plant nutrients, CEC, carbon content and enzyme activity all play a role in plant growth and soil quality. Rapid and accurate predictions of soil properties can allow producers to apply the appropriate amount of fertilizer to maximize yield and reduce cost and create high-resolution maps for environmental monitoring.

### 2.5.1 Electrical Conductivity

Soil electrical conductivity (EC) or salinity, is a measure of salt content in a soil. Soil salinity tends to be elevated in irrigated croplands, resulting in limited plant water availability, decreased soil permeability and alteration of exchangeable cation composition, all which negatively impact crop productivity (Corwin and Lesch, 2003). The use of MIR and VNIR to predict soil electrical conductivity has been widely studied (Minasny et al., 2009; Viscarra Rossel et al., 2006; Pirie et al., 2005; Janik et al., 2009) with varied and often poor results since EC is not spectrally active. Todorova et al. (2011) found a moderate correlation with EC and the NIR spectra ( $R^2 = 0.74$ ) and had moderate prediction performance (RPD = 1.5). Zornoza et al. (2008) had poor model predictions for EC using NIR ( $R^2 = 0.57$ ; RPD = 1.73). Viscarra Rossel et al. (2006) found that MIR predicted EC better than NIR ( $R^2 = 0.38$  and, 0.04 respectively), but Pirie et al. observed the opposite ( $R^2 < 0.10$ ; RPD = 1.0 and  $R^2 = 0.20$ ; RPD = 1.1 for MIR and VNIR). Minasny et al. (2009) predicted EC with MIR ( $R^2 = 0.79$ ; RPD = 2.2) and

believes this success was due to having a large range of values in the dataset. EC has been found to be modeled poorly when the sample range is small and poorly distributed (Minasny et al., 2009). Other authors report good predictions when electrolyte concentrations are in equilibrium with exchange sites, allowing EC to be predicted indirectly by correlations with CEC and clay mineralogy (Minasny et al., 2008; Janik et al., 2009).

### 2.5.2 Soil pH

Soil pH is the measure of the hydrogen ion activity and is used to measure soil acidity. It is an extremely important soil parameter that governs nutrient availability, microbial activity, soil organic matter transformations and many other soil properties (Todorova et al., 2011). Soil pH is not a spectrally active soil property but was predicted when correlated with soil organic acids, carbonates and soil minerals (Reeves, 2001; Sarathjith et al., 2014; Todorova et al., 2011). Todorova et al. (2011) found good model predictions for pH using NIR ( $R^2 = 0.91$ ; RPD = 2.3) and, Sarathjith et al. (2014) also found strong correlations with pH and organic carbon and had similarly good predictions ( $R^2 > 0.76$ ; RPD > 2.07). However, when there was not a strong correlation between pH, organic matter or clay mineralogy, predictions were poor. Chang et al. (2001) and Islam et al. (2003) observed little correlation between pH and spectrally active components, therefore, their pH models were poorer ( $R^2 < 0.71$ ; RPD < 2). Reeves et al., (2001) found that MIR predicted pH better than NIR and attributed this to MIR's ability to distinguish soil mineral composition. Viscarra Rossel et al. (2006) also found that MIR spectroscopy performed better than NIR and had smaller errors ( $R^2 = 0.86$ ; RMSEP = 0.10 for MIR and  $R^2 = 0.57$ ; RMSEP = 0.57 for NIR). This improvement was likely due to increased sensitivity to clay minerals in the

MIR region. In summary, soil pH can be predicted with VNIR and MIR spectroscopy, but only when secondary correlations with spectrally active soil components (SOM, carbonates and clay mineralogy) occur.

### 2.5.3 Phosphorous

Phosphorous has been relatively well studied in spectroscopy as P is an essential plant nutrient and excessive P can contribute to eutrophication and algal blooms in catchment basins (Bogrekci and Lee, 2005; Abdi et al., 2012). Current P extraction methods available such as Mehlich 3 (Mehlich, 1984), Olsen (Olsen et al., 1954), Bray 1, Bray 2 (Bray and Kurtz, 1945), and water (Morel et al., 2000) are expensive, destructive and time-consuming (Abdi et al., 2012). There is no one standard extraction method for all soils, as extraction methods are determined by the soils pH. In a recent review of VNIR literature by Stenberg et al. (2010), varying correlations for P predictions ( $R^2$  from 0.23-0.92) were found and the authors suggested that some of the variation in P predictability may be explained by the multitude of methods that are used to determine reference P content, which can be a part of different P fractions (total, extractable, available).

Total P has been predicted successfully with VNIR by Bogrekci & Lee (2005) ( $R^2 = 0.92$ , RMSE = 273.3 mg/kg P) and Todorova et al. (2011) ( $R^2 = 0.89$ , RPD = 2.0). One study by Reeves and Smith (2009) observed that MIR performed much better at predicting total P concentrations than NIR (MIR:  $R^2 = 0.85$ , RPD = 2.6 ;NIR:  $R^2 = 0.09$ , RPD = 1.1). Although it predicted well, total P may not be as meaningful to managers and producers who are interested in plant available P or extractable P. Plant available P is the form of P that producers are most interested in because it is the amount of P available



to the crop. However, there are no methods that directly measure plant available P, so extractable P is used to relate to plant available P (Tiessen and Moir, 1993). Bogrekci and Lee (2005) found that Mehlich III extracted-P correlated slightly better with VNIR than water extracted-P ( $R^2 = 0.86$  for Mehlich III P and  $R^2 = 0.77$  for water-soluble P). Chang et al. (2001) found that Mehlich III extractable cations were better predicted with VNIR than cations extracted by ammonium acetate, however, Mehlich III P was poorly predicted ( $R^2 = 0.40$ , RPD = 1.18).

MIR spectroscopy also portrays large variations in extractable P predictions. Janik et al., (2009) observed good predictions with water extractable P ( $R^2 = 0.85$ , RPD 2.5) and very poor correlation for Bray 1 extractable P ( $R^2 = 0.28$ ). A study by Minasny et al., (2009) found similar results, with water extractable P ( $R^2 = 0.79$ , RPD = 2.1) performing better than Bray 1 extractable P ( $R^2 = 0.04$ , RPD = 0.9).

Phosphorus is not a spectrally active component in the soil, therefore it cannot be directly measured, although it may be correlated with SOC, clay content and CEC (Minasny et al., 2009). This may explain why total P is well predicted by MIR and VNIR but not NIR. Total P would be highly correlated with clay mineralogy and clay content, both of which are well predicted by MIR, while VNIR and MIR are very good at predicting SOM (Bellon-Maurel and McBratney, 2011). Water extractable P, which is the P sorbed to the soil, is related to the surface area and charge of the soil, which may explain why it has been predicted relatively well. Because MIR and VNIR spectroscopy are better at predicting SOC and clay mineralogy, they may be better at predicting P fractions than NIR.

#### 2.5.4 Cation Exchange Capacity and Exchangeable Cations

Cation exchange capacity is important because it is a measure of the soils' buffering capacity and can inform the producer of a soils' ability to hold exchangeable cations (Sparks, 2003). CEC ( $\text{cmol kg}^{-1}$ ) is determined by measuring the amounts of Ca, Mg, and Na cations and summing their quantities. While cation exchange capacity is not a spectrally active component in the soil, it is related to clay mineralogy and SOM content, both of which are spectrally active properties (Stenberg et al., 2010). Chang et al. (2001) using VNIR and Zornoza et al. (2008) with NIR were successful in predicting CEC at regional scales ( $R^2 = 0.81$ ; RPD = 2.28 and  $R^2 = 0.92$ ; RPD = 3.46 respectively). A study by Groeningen et al. (2003) found effective cation exchange capacity better predicted by NIR ( $R^2 = 0.83$ ; RPD = 2.36) than MIR ( $R^2 = 0.56$ ; RPD = 1.54) at the field scale. However, a comparison of MIR and VNIR predictions by Viscarra Rossel et al. (2006) indicated that CEC had a better response in the MIR than VNIR, with average  $R^2$  values of 0.88 and 0.73 respectively. This may be due to the MIR's higher sensitivity to clay mineralogy and organic matter.

Exchangeable cations such as  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{K}^+$ , and  $\text{Na}^+$  play an important role in plant nutrition. The exchangeable cations calcium and magnesium have been predicted successfully with both NIR and MIR spectroscopy while exchangeable potassium and sodium generally show poorer correlations (Groeningen et al., 2003; Minasny et al., 2009; Pirie et al., 2005; Janik et al., 2009). In a review of VNIR studies, Stenberg et al. 2010 found exchangeable Ca, Mg, K, and Na to have highly variable correlations ( $R^2$  in parenthesis): Ca (0.75-0.89), Mg (0.53-0.82), K (0.11-0.55, and Na (0.09-0.44), suggesting that locally present co-variation of spectrally active components determines

the success of the predictions. Groeningen et al. (2003) had much success in predicting exchangeable magnesium ( $R^2 = 0.82$ ; RPD of 2.27) and exchangeable calcium ( $R^2 = 0.80$ ; RPD = 2.17) with VNIR but had very poor results for potassium ( $R^2 = 0.11$ ; RPD = 1.10). Chang et al. (2001) had similar results with VNIR spectroscopy, strong predictions for Ca ( $R^2 = 0.75$ ; RPD = 1.94) and Mg ( $R^2 = 0.68$ ; RPD = 1.75) and poor predictions for K ( $R^2 = 0.55$ ; RPD = 1.44) and Na ( $R^2 = 0.09$ ; RPD = 0.92). Similar predictions for Ca, Mg, K, and Na are observed with MIR models (Janik et al., 2009; Pirie et al., 2005). Janik et al. (2009) observed Ca ( $R^2 = 0.87$ ; RPD = 3.0), Mg ( $R^2 = 0.90$ ; RPD = 3.1), K ( $R^2 = 0.34$ ;) and Na ( $R^2 = 0.31$ ), indicating very good model performance for Ca and Mg. Pirie et al. (2005) observed Ca ( $R^2 = 0.69$ ; RPD = 1.8), Mg ( $R^2 = 0.76$ ; RPD = 2.1), K ( $R^2 = 0.46$ ; RPD = 1.3) and Na ( $R^2 = 0.20$ ; RPD = 1.3) and suggests that the poor performance for K and Na are due to their small range of values in the dataset. A study by Minasny et al. (2009) suggests this may be the case, as they observed good correlations and model performance for exchangeable Na ( $R^2 = 0.76$ ; RPD = 2.0) by having a range 1.5 times greater than that of Pirie et al. (2005). Ca and Mg are predicted well by both MIR and VNIR when there is a wide range of values in the dataset.

#### 2.5.5 Carbon

Soil carbon plays a very important role in soil systems and is fundamental in the carbon cycle. Soil organic carbon (SOC) is the amount of carbon stored in the soil and a significant component of soil organic matter (SOM). SOC is often used as an indicator for soil health (Weil and Magdoff., 2004). It affects bulk density and promotes soil aggregation, both of which have effects on soil water capacity and nutrient transport. Sequestering carbon in soils can increase food productivity while reducing greenhouse

gas emissions (Lal, 2015). It is important that we have rapid and cost-effective methods for quantifying soil carbon stocks.

There has been a good deal of research focusing on predicting soil carbon with VNIR (Stenberg et al., 2010; Chang et al., 2001, Viscarra Rossel et al., 2006; Bellon-Maurel and McBratney, 2011) and MIR techniques (Minasny et al., 2008; Wijewardane et al., 2018). In a comprehensive study of recent literature by Bellon-Maurel and McBratney (2011), MIR spectroscopy performed slightly better at predicting soil carbon than NIR, with MIR prediction errors about 10-40% lower than those by NIR and these findings have since been supported by other authors (Henaka Arachchi et al., 2016; Vohland et al., 2014). Ge et al. (2014) found that MIR models performed slightly better than VNIR models for organic carbon ( $R^2 = 0.8$  and  $RPD = 2.26$ ). Researchers suggest that MIR outperforms Vis and NIR as a result of the fundamental molecular vibrations that occur in the MIR spectrum. The fundamental bond vibrations from H-C, H-N, and H-O showed up as distinct absorption peaks that were linearly related to carbon content (Chang et al., 2001; Bellon-Maurel and McBratney, 2011). Minasny et al. (2008) had excellent predictions for total C using MIR spectroscopy with an RPD value of 6.25 and RMSEP of 0.24% and Wijewardane et al. (2018) also observed excellent predictions for total C ( $R^2 = 0.95$  and  $RPD = 4.44$ ).

While the NIR spectrum does not contain the distinct absorption features of the bond vibrations seen in the MIR region, it does contain weak overtones and combinations of these fundamental vibrations. This can make qualitative assessments of C content difficult by just looking at the spectrum. However, bands in the visible spectrum (410, 570 and 660 nm) and in the NIR (1875-1905 nm and 2200-2210 nm) have strong

correlations with soil organics and are found useful for predicting SOC (Viscarra Rossel et al., 2006; Vohland et al., 2014). Due to the relationship between soil color and organic matter, some studies have achieved better results with a combination of the visible and NIR spectrum than the NIR region itself. (Islam et al., 2003; Viscarra Rossel et al., 2006). Islam et al. (2003) improved  $R^2$  and RMSEP values for SOC predictions (NIR:  $R^2 = 0.68$ , RMSEP = 0.45%; VNIR:  $R^2 = 0.81$ , RMSEP = 0.35%). Chang et al., (2001) reported excellent results for total carbon ( $R^2 = 0.87$ ; RPD = 2.79; RMSEP = 7.86%) using the VNIR spectrum and Wijewardane et al., (2016) had similar results for total C ( $R^2 = 0.83$ ; RPD = 2.41; RMSEP = 7.38%).

Predictions for SOM and SOC are highly variable, where predictions tend to be poorer over larger, coarser scales with wide variations in C concentrations and better at field scales where C pools do not vary as much (Stenberg et al., 2010). Stenberg et al., (2010) also found that calibrations for SOM could be improved by removing the sandiest of soils from the calibration dataset due to light scatter masking soil organic content in higher quartz content in some sandy soils.

MIR spectroscopy shows excellent predictions for total carbon content and organic carbon content due to the fundamental molecular vibrations which can be linearly related to the amount of carbon present (Bellon-Maurel and McBratney, 2011). VNIR spectroscopy has good performance but tends to not perform as well as MIR because it relates broader, less distinct absorption peaks to carbon content. These absorption features can be easily masked by the scattering effects of quartz, as is the case for sandy soils (Stenberg et al., 2002).

### 2.5.6 $\beta$ -Glucosidase

Soil microbial activity can be used as an indicator of soil quality (Cohen et al., 2005). Measuring microbial activity has advantages over measuring chemical properties due to its interactive and dynamic nature.  $\beta$ -glucosidase producing microbes are of particular importance when it comes to SOC. The  $\beta$ -glucosidase ( $\beta$ G) enzyme is involved in converting cellulose into glucose, which in turn provides energy for microbes (Cohen et al., 2005; Dick et al., 2013). Because of this direct relationship with soil organic carbon,  $\beta$ G activity can be used to monitor rapid changes in SOC that occur from changes in soil management (Bandick and Dick, 1999).  $\beta$ -glucosidase activity, due to its relationship with soil organic carbon, has been modeled successfully with infrared spectroscopy (Cohen et al., 2005; Zornoza et al., 2008; Dick et al., 2013) Cohen et al. (2005) had excellent predictions for  $\beta$ -glucosidase activity for wetland soils using VNIR ( $R^2 = 0.96$ ; RPD = 2.64) as did Zornoza et al. (2008), who predicted  $\beta$ -glucosidase successfully ( $R^2 = 0.93$ ; RPD = 3.66) using NIR. Dick et al. (2013) found that excluding the visible region of the spectrum reduced noise and improved correlations. Infrared spectroscopy has been observed to be capable of predicting  $\beta$ -glucosidase enzyme activity by correlation with SOC and shows the potential for rapid assessments of soil quality.

### 2.5.7 Nitrogen

Nitrogen, like phosphorus, is considered one of the most important macronutrients in agriculture and is a common fertilizer component (Todorova et al., 2011). Plant available N, such as ammonium and nitrate are dynamic and thus estimates of these nitrogen forms are quite variable and hard to predict (Stenberg et al., 2010; Viscarra

Rossel et al., 2006; Minasny et al., 2009). Viscarra Rossel et al. (2006) found no correlation across the VNIR and MIR spectra with nitrate-N. Minasny et al. (2009) observed similar results for nitrate-N using MIR ( $R^2 = 0.08$ , RPD = 0.9). However, total N and total Kjeldahl N (TKN), which contain larger fractions of N and tend to be less dynamic, have been predicted with very good accuracy using NIR (Chang et al., 2001; Todorova et al., 2011; Zornoza et al., 2008).

Total N is the total amount of nitrogen in the soil including all organic and inorganic forms while TKN includes the organic and ammonium form of nitrogen. Chang et al. (2001) observed  $r^2$  values of 0.85 and RPD values of 2.52. Todorova et al. (2011) observed similar results as Chang et al. (2001) ( $R^2 = 0.91$  and RPD = 2.3). Zornoza et al. (2008) had very good predictions for TKN ( $R^2 = 0.96$  and RPD = 4.88). MIR predictions were similar to VNIR for total N. Minasny et al. (2009) observed  $R^2$  value of 0.76 and an RPD of 2 indicating moderate model performance. Reeves et al. (2001) observed a good fit for total N ( $R^2 = 0.95$ ) but model performance was not reported. VNIR and MIR spectroscopy were equally effective in predicting total N, but both failed to predict plant available N. Quantifying soil nitrate and ammonium was likely difficult due to their low concentrations and their dynamic nature (Janik et al., 1998).

## 2.6 Summary

Infrared spectroscopy can be used to predict a wide range of spectrally active soil properties. In particular, soil organic content and total nitrogen are usually predicted very well with both VNIR and MIR. For spectrally inactive properties like pH and CEC, prediction accuracy is determined by the amount of co-variation these properties have with soil organic matter and the clay mineralogy of the soil. Additionally, datasets that

have large response ranges tend to have better prediction accuracy than those that have smaller ranges, as is the case for exchangeable Na. In general, studies have shown that MIR predictions are more accurate than VNIR for predicting soil carbon content and soil properties closely related to SOM due to the fundamental molecular vibrations of organic compounds. However, this slight increase in prediction accuracy may not offset the costs associated with the more expensive MIR spectrometers. Visible and infrared spectroscopic methods can collect a significant quantity of information on various soil properties rapidly and nondestructively, making these methods ideal for sampling of coarse data over large spatial regions or fine resolution sampling at the field scale.



## CHAPTER 3. MATERIALS & METHODS

### 3.1 Study Area

Soil samples used in this study were from a subset of samples collected for the National Rapid assessment Carbon analysis (RaCA) by the United States Department of Agriculture Natural Resources Conservation Service (USDA-NRCS) (Soil Survey Staff and Loecke, 2016). 27 sites were randomly chosen across Nebraska and Kansas within major land resource area (MLRA) 5, figure 3.1. The sites included a variety of land use/land cover (LULC) classes: pasture (n = 6), rangeland (n = 11), cropland (n = 7) and conservation reserve program (CRP) land (n = 3). At each site, samples were collected from a central pedon or profile (0-100 cm) and from two satellite pedons (0 – 100 cm), 30 meters away, on either side of the central pedon. A total of 156 samples used in this study came from A horizons, 0-5 cm and 5 cm to the bottom of the A horizon. Sub-horizons not classified as A horizons were omitted from the study.

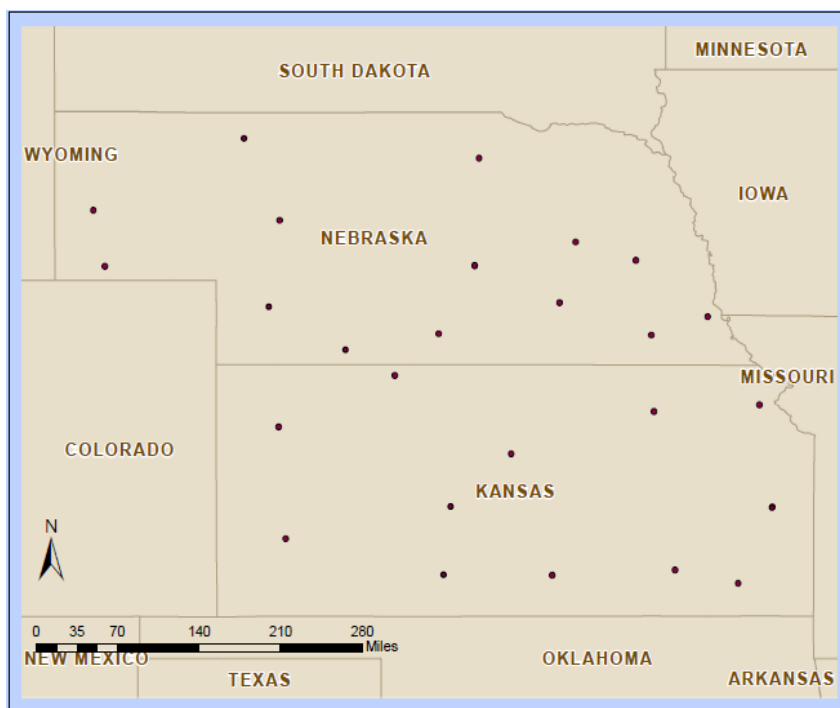


Figure 3.1. Study area and locations of 27 sampled sites.

### 3.2 Reference Data Collection

All analyses were performed at the NRCS Kellogg Soil Survey Laboratory. Soils were air-dried and ground to < 2-mm. The following analyses were performed: electrical conductivity (EC), pH, cation exchange capacity (CEC), extractable bases ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{K}^+$ ,  $\text{Na}^+$ ), water-soluble phosphorus ( $\text{H}_2\text{O P}$ ), total phosphorous (TP), total carbon (TC), soil organic carbon (SOC), calcium carbonate ( $\text{CaCO}_3$ ),  $\beta$ -Glucosidase enzyme activity ( $\beta\text{G}$ ), and total nitrogen (TN). All samples were analyzed according to the USDA-NRCS Soil Survey Laboratory Methods Manual (Burt et al., 2014). Soil standards with known properties were run for all analyses for quality control.

#### 3.2.1 Air-dry Oven-dry Ratio

Prior to analyses, the air-dry/oven-dry (AD/OD) ratio was determined to allow conversion of soils to an oven-dry basis and was calculated by procedure 3D1 in the USDA-NRCS Soil Survey Laboratory Methods Manual (Burt et al., 2014). About a 5-gram sample of soil was placed into a metal tin and weighed. The tin was then placed into a  $110^\circ\text{C}$  oven for 12-16 hours. Sample tins were removed and allowed to cool for a few minutes then weighed to the nearest milligram (mg). The AD/OD was calculated according to the following equation:

$$(11) \quad AD/OD \text{ ratio} = \frac{(air-dry\ mass\ of\ soil + measuring\ tin)}{(oven-dry\ mass\ of\ soil + measuring\ tin)}$$

#### 3.2.2 Electrical Conductivity

Electrical conductivity is used to predict soluble salt concentrations and was performed according to procedure 4F1a1a1. EC was determined as the conductivity of a soil and water solution after 24 hours. Equipment used included: an electronic balance with  $\pm 1.0$  mg sensitivity, conductivity bridge and conductivity cell, with automatic

temperature adjustment,  $25 \pm 0.1^\circ \text{C}$  (Markson Model 1056, Amber Science, Eugene, OR), 30-mL plastic cups with lids (Sweetheart Cup Co. Inc., Owings Mills, MD) and disposable 10-mL plastic pipets. Reagents included reverse osmosis (RO) water (ASTM III grade of reagent water) and 0.01 N Potassium chloride (KCl) (conductivity of  $1.412 \text{ mmhos cm}^{-1}$  at  $25^\circ \text{C}$ ).

5 grams of soil sample were weighed and placed in plastic cups. 10-ml of RO water were added and mixed to create a soil solution, then were allowed to equilibrate overnight. The conductivity bridge was standardized with RO water (blank) and 0.01 N KCl. Samples were placed under the conductivity cell tube and the solution is drawn into the cell from suction created by a pipet. The conductivity readings were measured from the bridge and recorded to the nearest  $0.01 \text{ mmhos cm}^{-1}$ .

### 3.2.3 Soil pH

Soil pH was determined by two commonly used methods, 1:1 water pH (procedure 4C1a2a1) and 1:2 Calcium Chloride ( $\text{CaCl}_2$ ) (procedure 4C1a2a2). Soil pH was determined from an equilibrated soil solution of water and  $\text{CaCl}_2$ . Equipment used included: calibrated measuring scoop ( $\approx 20 \text{ g}$  capacity); 120-mL disposable paper cups (Solo Cup Co., No. 404); reagent dispenser (0-30mL); wooden beverage stirring sticks; 250-mL polyethylene titration beakers; automatic titrator, Metrohm Titroprocessors (Metrohm Ltd., Riverview, FL; Brinkman Instruments, Inc., Westbury, NY); combination pH-reference electrode, Metrohm part no. 6.01210.100 (Brinkman Instruments, Inc., Westbury, NY). Reagents included: RO water; Borax pH buffers, pH 4.00, pH 7.00 and pH 9.18 (Beckman, Fullerton, CA); 0.02 M  $\text{CaCl}_2$ .

With the measuring scoop, about 20-g of soil were placed into paper cups. 20 mL of RO water was dispensed into samples and stirred. The sample cups were allowed to stand for one hour and stirred occasionally. After one hour, cups were loaded into the sample changer. The pH meter was calibrated with pH buffer solutions. The computer then automatically stirred the sample for 30 seconds, waited 1 minute, positioned electrode into solution and collected 1:1 water pH reading. Then 20-mL of CaCl<sub>2</sub> solution was added to the sample, stirred for 30 seconds and a 1:2 CaCl<sub>2</sub> pH reading was collected after 1 minute. The electrode and stirrer were automatically rinsed before continuing to the next sample. The pH readings were recorded to the nearest 0.01 unit.

#### 3.2.4 Water Soluble Phosphorus

Water soluble P is the amount of P available to a water solution and attempts to approximate soil solution P. This was determined by procedure 4D2a1a1. Samples were shaken in a water solution for 30 minutes and centrifuged. The supernatant was then reacted with a coloring solution and soluble P was determined colorimetrically.

Equipment used included: an electronic balance,  $\pm 1.0$ -mg sensitivity; Eberbach 6000 mechanical shaker (Everbach Corp., Ann Arbor, MI); 50-mL polyethylene centrifuge tubes; 0.45- $\mu$ m syringe filters, (Whatman); Leur-lok 10-mL syringes; Centra GP-8 centrifuge (Thermo IEC, Needham Heights, MA); volumetric flasks, 2-L, 1-L, 100-mL, and 25-mL; dark plastic bottles, 2-L; plastic cuvettes, 4.5-mL, 1-cm light path (Daigger Scientific Inc., Vernon Hills, IL), Cary 60 UV-Visible spectrophotometer, (Varian Inc., Palo Alto, CA) and a computer with Cary WinUV software (Varian Inc., Palo Alto, CA). Reagents used were: reverse osmosis deionized (RODI) water (ASTM I grade of reagent water); sulfuric acid (H<sub>2</sub>SO<sub>4</sub>), concentrated, 36 N, trace pure grade; ascorbic acid;

molybdate solution (reagent A); reagent B, made by dissolving 2.112-g of ascorbic acid in 400 mL of reagent A; stock standard P solution (SSPS), 100.0 mg P L<sup>-1</sup>; working stock standard P solution (WSSPS) 10.0 mg P L<sup>-1</sup>; standard P calibration solutions (SPCS) 0.0, 0.2, 0.4, 0.6 and 0.8 mg P L<sup>-1</sup>; a 0.1 mg P L<sup>-1</sup> quality control solution and blanks.

2.5 grams of soil and 25-mL of RODI water were added to 50-mL centrifuge tubes. The samples were placed in a mechanical shaker for 30 minutes at 200 oscillations min<sup>-1</sup> at room temperature. After shaking, the samples were centrifuged for 20 minutes at 3000 rpm. The supernatant was decanted into a 10-mL syringe in which a syringe filter was attached. The plunger was inserted, and the solution was filtered into a cup. 2-mL of sample solution was pipetted into a plastic cup and 4-mL of reagent B and 19-mL of RODI water was added. The samples were allowed to sit for 20 minutes to let the color develop and then transferred to cuvettes. The spectrophotometer was set to 882-nm and auto-zeroed with the calibration blank (0.0 mg P L<sup>-1</sup>). The calibration curve was made with the SPCS standards. The minimum fit for calibration was 0.9900. The cuvettes were placed in the spectrophotometer and run with QC samples in between batches. Any samples that had readings outside of the calibration were diluted with extracting solution. The sample concentration was calculated from equation 12:

$$(12) \quad \textit{Absorption} = \textit{Slope} \times \textit{Concentration}$$

Conversion of extracted P (mg L<sup>-1</sup>) to soil P (mg kg<sup>-1</sup>) was according to equation 13:

$$(13) \quad \textit{SoilP}(\textit{mgkg}^{-1}) = [(A \times B \times C_1 \times C_2 \times R \times 1000)/E]$$

Where A is the sample extract reading (mg kg<sup>-1</sup>); B is extract volume (L); C<sub>1</sub> is the dilution factor (125); C<sub>2</sub> is the dilution factor (if required); R is the AD/OD ratio and E is the sample weight (g).

### 3.2.5 Cation Exchange Capacity

Cation exchange capacity is the sum of negative charges per unit mass and is found by the sum of extractable bases, expressed as centimoles per kg ( $\text{cmol}(+) \text{kg}^{-1}$ ) and is performed by procedure 4B1a1a1a1. CEC was determined by saturating soil exchange sites with an index cation, ammonium ( $\text{NH}_4^+$ ), then an ethanol wash was used to remove excess cations from the soil solution. Finally, potassium chloride (KCl) was used to displace  $\text{NH}_4^+$  cations absorbed to exchange sites. CEC was measured by the amount of  $\text{NH}_4^+$  cations in solution by steam distillation and titration. Equipment used included: an electronic balance  $\pm 1.0$ -mg sensitivity; mechanical vacuum extractor, 24-place (SAMPLETEX, MAVCO Industries, Lincoln, NE); 60-mL polypropylene tubes (for extracting, reservoirs and tared extraction tubes), rubber tubing, 3.2 ID x 1.6 OD x 6.4 mm; Kjeltex Auto 2300 Sampler System (Perstorp Analytical, Malö, Sweden); 250-mL straight neck digestion tubes; 0.45- $\mu\text{m}$  syringe filters (Whatman); wash bottles; plastic vials and a Centra, GP-8 centrifuge (Thermo IEC, Needham Heights, MA). Reagents used included reverse osmosis deionized (RODI) water; 1 N, ammonium acetate solution ( $\text{NH}_4\text{OAc}$ ), pH 7.0; ethanol ( $\text{CH}_3\text{CH}_2\text{OH}$ ), 95%, U.S.P.; Nessler's reagent; 2 M potassium chloride solution; 4% (w/v) boric acid with bromocresol green-methyl red indicator (0.075 % bromocresol green and 0.05 % methyl red) (Chempure, Plymouth, MI); 0.1 N, standardized hydrochloric acid (HCl); and 1 M sodium hydroxide (NaOH).

2.5 g of soil were weighed to the nearest mg and placed into polypropylene extraction tubes. Tubes were placed on the extractor and connected to a tared extraction tube with rubber tubing.  $\text{NH}_4\text{OAc}$  was used to remove cations from exchange sites. First, about 20-mL of  $\text{NH}_4\text{OAc}$  were added to the sample in the extraction tube. Empty

reservoir tubes were then placed on top of the extraction tubes and allowed to stand for 30 minutes. Then the  $\text{NH}_4\text{OAc}$  solution was extracted until about 2 mL of solution was above the soil level. 40-mL of  $\text{NH}_4\text{OAc}$  were added to the reservoir tubes and then allowed to extract overnight. The reservoir tubes and the tared extraction tubes were then removed from the extractor and the tared extraction tubes weighed. The extractant was shaken manually and dispensed into small plastic vials for analysis of extracted cations:  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{K}^+$ ,  $\text{Na}^+$  (procedure 4B1a1b1-4). The tared extraction tubes were reconnected to their respective extraction tubes. Ethanol was then used to remove  $\text{NH}_4\text{OAc}$  from the soil solution. The extraction tubes were filled to the 20-mL mark with ethanol. Reservoir tubes were placed onto extraction tubes and allowed to stand for 30 minutes. The ethanol solution was extracted through the soil until about the solution was 2-mL above the soil. 45-mL of ethanol were added to the reservoir tubes and extracted until 2-mL of solution remained above the soil. The tared extraction tubes were removed, and the ethanol discarded. The empty tared extraction tubes were replaced, and the process was repeated with another 55 mL of ethanol. After this rinsing, a sample of ethanol was taken and collected on a spot plate. Nessler's reagent was used to test for any remaining  $\text{NH}_4^+$  in solution. If a yellow, red to reddish-brown precipitate forms, another rinse was performed. A new tared extracting tube was then attached to the extraction tube on the extractor. 2 M  $\text{KCl}$  was then used to replace the  $\text{NH}_4^+$  on exchange sites. The extraction tube was filled to the 20-mL mark with  $\text{KCl}$  solution and allowed to stand for 30 minutes. The  $\text{KCl}$  solution was extracted through the soil until 2 mL of solution remained above the soil. Then 40 mL of  $\text{KCl}$  solution were added to the reservoir tube and extracted for another 45 minutes. The extracted  $\text{KCl}$  solution was transferred to a 250-mL digestion

tube for steam distillation using the Kjeltex autosampler. The distillation and titration were performed automatically. CEC was calculated by equation 14:

$$(14) \quad CEC = [Titer \times N \times 100 \times R]/[E]$$

where: CEC is cation exchange capacity (meq 100 g<sup>-1</sup>); Titer is titer of sample (mL); N is normality of HCl titrant; 100 is the conversion factor to 100 g basis; R is the AD/OD ratio, and E is the soil sample weight (g).

### 3.2.6 Extractable Bases

Extractable bases (Ca<sup>+</sup>, Mg<sup>+</sup>, K<sup>+</sup>, Na<sup>+</sup>) extracted from an ammonium acetate (NH<sub>4</sub>Ac) extraction (procedure 4B1a1) were considered to be exchangeable bases from cation exchange sites. Determining extractable bases consisted of diluting the NH<sub>4</sub>Ac extract from the CEC procedure with an ionization suppressant, and then measured by atomic absorption spectrophotometry (AAS). Equipment used included: an electronic scale, ± 1.0-mg sensitivity; Analyst 300 atomic absorption spectrophotometer (AAS) with double-beam optical system (Perkin-Elmer Corp., Norwalk, CT); AS-90 autosampler (Perkin-Elmer Corp., Norwalk, CT); computer with AA WinLab software (Perkin-Elmer Corp., Norwalk, CT); single-stage regulator, acetylene; digital diluter/dispenser, with syringes 10000 and 1000 µL, gas tight, MicroLab 500 (Hamilton Co., Reno, NV); plastic test tubes, 15-mL, 16mm x 100; polyethylene containers; and a peristaltic pump.

Reagents used included: 1:1 HCl:RODI; 1 N ammonium acetate solution (NH<sub>4</sub>OAc), pH 7.0 (solution used to extract cations in previous procedure 4B1a1); 2 N ammonium acetate solution (NH<sub>4</sub>OAc), pH 7.0; 65,000 mg L<sup>-1</sup> stock lanthanum ionization suppressant solution (SLISS); 2,000 mg L<sup>-1</sup> working lanthanum ionization suppressant solution (WLISS); 1000 mg L<sup>-1</sup>: Ca, Mg, K and Na primary stock standards solutions



(PSSS), high purity; working stock mixed standards solution (WSMSS), High, Medium, Low, Very Low, and Blank; mixed calibration standard solutions (MCSS), High, Medium, Low, Very Low, and Blank; compressed air with water and oil traps; and acetylene gas, purity 99.6%.

The NH<sub>4</sub>OAc sample extract solutions were diluted with an ionization suppressant, WLISS, 1-part extract to 20 parts WLISS. 1 mL of diluted sample solutions were transferred to test tubes in the sample changer to be analyzed by the AAS. The AAS was calibrated using the MCSS solutions where fit rejection criteria is <0.99. If samples exceed calibration standard, then the samples were diluted with 1 N NH<sub>4</sub>Ac followed by a 1:20 dilution with WLISS. The MCSS was ran as a quality control standard every 12 samples to ensure the instrument retained calibration. Analyte readings were recorded to the nearest 0.01 mg L<sup>-1</sup> (extract concentration) and converted to meq 100 g<sup>-1</sup> by equation 15:

$$(15) \quad \text{AnalyteConcentration}(\text{meq}100\text{g}^{-1}) = \frac{[A \times \frac{(B_1 - B_2)}{B_3}] \times C \times R \times 100}{[1000 \times E \times F]}$$

where A is analyte (Ca, Mg, K, Na) concentration in extract (mg L<sup>-1</sup>); B<sub>1</sub> is weight of extraction syringe and extract (g); B<sub>2</sub> is weight of tared extraction syringe (g); B<sub>3</sub> is the density of 1 N NH<sub>4</sub>OAc at 20° C (1.0124 g cm<sup>-3</sup>); C is the dilution, if required; R is the AD/OD ratio; 100 is the conversion factor to 100-g basis; 1000 is the mL to L conversion; E is the soil sample weight (g); and F is the equivalent weight (mg meq<sup>-1</sup>) where: Ca<sup>2+</sup> = 20.04 mg meq<sup>-1</sup>; Mg<sup>2+</sup> = 12.15 mg meq<sup>-1</sup>; K<sup>+</sup> = 39.10 mg meq<sup>-1</sup>; and Na<sup>+</sup> = 22.99 mg meq<sup>-1</sup>. Extractable cations were reported to the nearest 0.1 meq 100 g<sup>-1</sup>.

### 3.2.6 β-Glucosidase

The  $\beta$ -glucosidase enzyme was measured according to procedure 6A5a1a1.

Samples were divided into control and treatment sets. A buffer solution was added to both and p-nitrophenyl- $\beta$ -glucoside (PNG) was then added to the treatment set. Both sets were incubated for an hour at 37° C. After incubation, the reactions were stopped, and the samples were analyzed on a spectrophotometer. Concentrations were determined by the difference in concentrations between the treatment and control samples. Equipment used included: an electronic balance,  $\pm 1.0$  mg sensitivity; volumetric flasks (50-mL, 100-mL and 1-L); plastic amber bottles, 1-L; disposable plastic cups, 1 oz.; 0.45- $\mu$ m syringe filters (Whatman); 50-mL centrifuge tubes, (Falcon); water bath; Cary 60 UV-Visible spectrophotometer (Varian Inc., Palo Alto, CA), vortexer; 10-mL syringes; cuvettes; 500-mL beakers; pH meter; magnetic stir bars; and magnetic stir plate. Reagents used included: reverse osmosis deionized (RODI) water; 1 M sodium hydroxide (NaOH); 0.5 M NaOH; 0.1 M hydrochloric acid; modified universal buffer (MUB) stock solution (prepared by dissolving 12.10 g tris(hydroxymethyl)aminomethane (THAM), 11.60 g maleic acid (C<sub>4</sub>H<sub>4</sub>O<sub>4</sub>), 14.00 g citric acid (C<sub>6</sub>H<sub>8</sub>O<sub>7</sub>), and 6.30 g boric acid (H<sub>3</sub>BO<sub>3</sub>) in 488 mL 1 M NaOH in a 1-L volumetric flask and bringing to volume with RODI water); MUB working solution (pH 6.0) (prepared by placing 200 mL of MUB stock solution in a 500-mL beaker with a magnetic stir bar on a magnetic stirrer, while stirring, 0.1 M HCl was added until pH reached 6.0, then solution was transferred to a 1-L volumetric flask and brought to volume with RODI water); 0.05 M p-nitrophenyl- $\beta$ -D-glucoside (PNG) (prepared by dissolving 1.308 g of PNG in 80 mL MUB working solution (pH 6.0) in a 100 mL volumetric flask, and brought to volume with MUB working solution); 0.5 M calcium chloride dihydrate (CaCl<sub>2</sub>•H<sub>2</sub>O); 0.1 M 2-amino-2-(hydroxymethyl)-1-3-

propanediol (THAM), pH 10; 0.1 M THAM, pH 12; p-nitrophenol stock standard solution (NPSSS), (prepared in a fume hood by dissolving 1.00 g p-nitrophenol in 700 mL of RODI water in a 1-L volumetric flask and brought to volume with RODI water); p-nitrophenol working stock standard solution (NPWSS), (prepared in a fume hood by adding 1 mL NPSSS in a 100 mL volumetric flask and brought to volume with RODI water); and standard p-nitrophenol calibration solutions (NPCS), (prepared in 6 centrifuge tubes by adding: 5 mL, 4 mL, 3 mL, 2 mL, 1 mL, and 0 mL NPWSS respectively, then 0 mL, 1 mL, 2 mL, 3 mL, 4 mL, and 5 mL of RODI water were added, respectively. Then 0.5 M CaCl<sub>2</sub> and 4.0 mL of THAM (pH 12) were added to each tube, capped and mixed).

PNG solution, THAM (pH 10), and MUB working solution were placed in a water bath and warmed to 37° C. Centrifuge tubes were divided into control and treatment sets. 1 gram of soil sample was placed into a control tube and a treatment tube. 4 mL of MUB working solution were added to both control and treatment tubes. 1 mL of PNG solution was then added to the treatment tubes. All the tubes were capped and mixed by vortex for about 1 second. All samples were then placed in the water bath for 1 hour. 1 mL of 0.5 M CaCl<sub>2</sub> and 4 mL of THAM (pH 12) were added to all samples and mixed to stop the reaction. To make the solution matrix the same between control and treatment tubes, 1 mL of PNG was added to the control tubes and vortexed. The caps were removed, and the samples allowed to rest for 5 minutes before filtering. Solutions were then aspirated out of the test tubes with a syringe. Syringe filters were attached to the syringes and the solutions were filtered into plastic cups. Samples were decanted into cuvettes and read on the spectrophotometer at 410 nm. The NPWSS solutions were used

to create the calibration curve. If the treatment sample exceeded the high standard, both the treatment and control samples were diluted with THAM (pH 10). A blank sample and soil standard were ran as a quality control.  $\beta$ -glucosidase activity was determined by equation 16:

$$(16) \quad \beta G = R \times [(D_t \times A_t / E_t) - (D_c \times A_c / E_c)] / T$$

where:  $\beta G$  is the p-nitrophenol concentration in the soil per hour ( $\mu\text{g} \cdot \text{g}^{-1} \cdot \text{h}^{-1}$ );  $A_t$  is the amount of p-nitrophenol in treatment sample ( $\mu\text{g}$ );  $A_c$  is the amount of p-nitrophenol in the control sample ( $\mu\text{g}$ );  $D_t$  is the dilution factor of the treatment sample;  $D_c$  is the dilution factor of the control sample;  $E_t$  is the sample weight of the treatment sample (g);  $E_c$  is the sample weight of the control sample (g);  $R$  is the AD/OD ratio; and  $T$  is the incubation time (h).

### 3.2.7 Calcium Carbonate

Calcium Carbonate is a part of the inorganic fraction of carbon and was measured according to procedure 4E1a1a. A soil sample was treated with HCl to test for the presence of carbonates (procedure 1B152b5) and if effervescence occurred, the sample was measured for carbonates by manometrically measuring the amount of carbonate converted to  $\text{CO}_2$  when reacted with HCl. Equipment used included: an electronic balance,  $\pm 0.10$ -mg sensitivity; porcelain spot plate; 120-mL wide-mouth clear glass threaded weighing bottles; machined PVC caps for threaded weighing bottles, with hole drilled in center; O-rings, 3.2 x 50.8 x 57.2 mm; Flanged stopper, (No. 03-255-5, Fisher Scientific,); PLC-200 Series manometer (Omega Engineering, Stanford, CT); 23-gauge hypodermic needle; 10-mL size 11 gelatin capsules (Torpac Inc., Fairfield, NJ); and Eberbach 6140 mechanical rotating shaker (Everbach Corp., Ann Arbor, MI). Reagents

used included: reverse osmosis deionized (RODI) water; methyl red indicator; 3 N hydrochloric acid (HCl) (made by diluting 500 mL of concentrated HCl with 1500 mL RODI water with a few crystals of methyl red indicator); 1 N HCl; glycerin, USP; and CaCO<sub>3</sub> assay dried basis 100.01% (Ultrex).

Samples to be tested for carbonates were first tested for an effervescent reaction when subject to HCl (procedure 1B1b2b5). A gram of soil was placed in a porcelain spot plate and mixed with RODI water. A few drops of 1 N HCl were added to the solution to check effervescence. When a sample was classified as having any degree of effervescence or had a 1:2 CaCl<sub>2</sub> pH greater than 6.95, it was tested for carbonates. The manometer was calibrated using three replicates of CaCO<sub>3</sub> standards (0, 0.025, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, and 0.75 g) that had been dried in an oven at 110° C for 2 hours. Soil sample weight was determined by the degree of effervescence: 2-g for samples with Very slight to Slight effervescence; 1-g for Strong effervescence; and 0.5 g for Violent effervescence. The samples were placed in a 120-mL bottle. A gelatin capsule containing 10 mL of 3 N HCl was placed in the bottle and capped. The initial pressure in the bottle was released by piercing the stopper with a hypodermic needle for 5 to 10 seconds. Once the HCl dissolved the gelatin capsule, the bottle was shaken at 140 rpm for 10 minutes. Samples stand for 40 minutes, then shaken for another 10 minutes. The HCl converts the carbonates in the soil into carbon dioxide (CO<sub>2</sub>) which was measured with the manometer. The manometer was auto-zeroed before readings were taken. Readings were recorded to the nearest mm Hg. Manometer readings were corrected according to equation 17:

$$(17) \quad CR = (MR - BR)$$

where: CR is the corrected reading in (mm Hg); MR is the manometer reading (mm Hg); and BR is the blank reading (mm Hg) obtained by the average of three blanks. The corrected reading of the standards was used to create the regression equation for estimating CaCO<sub>3</sub> % from equation 16:

$$(18) \quad CCE = [(CR \times Slope + Intercept)/E] \times R$$

where: CCE is the calcium carbonate equivalent (%); CR is the corrected reading (mm Hg); E is the sample weight; and R is the AD/OD ratio. The calculated CCE was the percentage of carbonate per oven-dry gram of soil.

### 3.2.8 Total Carbon and Nitrogen

Total C and N were measured to get the total concentrations of carbon and nitrogen (both organic and inorganic forms) in the soil by dry combustion (procedure 4H2a1-3) by measuring the amounts of carbon dioxide (CO<sub>2</sub>) and nitrogen gas (N<sub>2</sub>) from combustion. Equipment used included: Elementar varioEL/Elementar varioEL III (Elementar Analysensysteme GmbH, Hanau, Germany); combustibles: quartz ash finger (quartz), quartz bridge, combustion tube, reduction tube, gas purification (U-tube, GL 18), support tube (65 mm) protective tube, O<sub>2</sub> lance (150 mm rapid N), tin boats (4 x 4 x 11 mm), and tin foil cups (Elementar Americas Inc., Mt. Laurel, NJ; Alpha Resources Inc., Stevensville, MI); and computer with varioEL software (Elementar Analysensysteme GmbH, Hanau, Germany). Reagents used included: sulfanilic acid calibration standard (41.6% C, 4.1% H, 8.1% N, 27.7% O, and 18.5% S); copper sticks; corundum balls, high purity, alumina spheres (3-5 mm); cerium dioxide (1-2 mm); tungsten oxide powder; tungsten trioxide granulate; quartz wool; silver wool; phosphorus

pentoxide, Sicapent (Elementar America, Inc., Mt. Laurel, NJ); helium carrier gas (99.996% purity); and oxygen combustion gas (99.995% purity).

Soil samples were first finely ground to  $< 180 \mu\text{m}$ . A 0.100 g of tungsten oxide powder was placed in tin foil and then the soil sample (0.100-0.05 g) was packed into the foil and placed into the carousel of the automatic sample feeder of the elemental analyzer. The sample weight was recorded by the computer and the combustion process begins. The carbon and nitrogen in the sample were measured by the amount of  $\text{CO}_2$  and  $\text{N}_2$  after combustion. A quality control sample was performed every 35 to 40 samples. Total C was calculated by equation 19:

$$(19) \quad C = C_i \times R$$

where, C is the total carbon percent on an oven-dry basis,  $C_i$  is the carbon percent read by the machine, and R is the AD/OD ratio. Total nitrogen was calculated from equation 20:

$$(20) \quad N = N_i \times R$$

where, N is the total nitrogen percent on an oven-dry basis,  $N_i$  is the nitrogen percent read by the machine, and R is the AD/OD ratio. Total C was recorded to the nearest 0.01% and total N was recorded to the nearest 0.001%.

### 3.2.9 Total Phosphorus

Total phosphorus was determined by the Major Element analysis (procedure 4H1b1a1a1-11) where samples undergo microwave acid digestion, and elements were analyzed by inductively coupled plasma atomic emission spectroscopy (ICP-AES). Equipment used included: an electronic balance,  $\pm 0.1$ -mg sensitivity; 10-mL pipettes (Omnifit Corp.); 100-mL Nalgene volumetric flasks; 60-mL polypropylene bottles with cap; CEM Mars 5 microwave, with 14 position-HP500 PLUS vessels and rotor (CEM

Corp., Mathews, NC); desiccator; disposable aluminum-weighing dishes; 500-mL polypropylene volumetric flask; 500-mL polypropylene containers, with screw caps; electronic digital pipettes, 2500  $\mu\text{L}$  and 10 mL, (Rainin Instrument Co., Woburn, MA); Perkin-Elmer Optima 3300 Dual View ICP-AES (Perkin-Elmer Corp., Norwalk, CT); 45-MHz free running RF generator (Perkin-Elmer Corp., Norwalk, CT); computer with WinLab software (Perkin-Elmer Corp., Norwalk, CT ); compressed gases of argon (99.996% purity) and nitrogen (99.999% purity); AS-90 autosampler (Perkin-Elmer Corp., Norwalk, CT); quartz torch, part no. N069-1662 (Perkin-Elmer Corp., Norwalk, CT); and an alumina injector (2.0 mm I.D.), part no. N069-5362 (Perkin-Elmer Corp., Norwalk, CT). Reagents used included reverse osmosis deionized (RODI) water; calcium sulfate desiccant; 48 % hydrofluoric acid (HF), low trace metal content; 12 N concentrated hydrochloric acid (HCl), trace pure grade; 16 N concentrated nitric acid ( $\text{HNO}_3$ ), trace pure grade; 4.5% boric acid solution; 1.9% boric acid solution; and primary standards, 1000 mg P L<sup>-1</sup> (High Purity Standards, Charleston, SC).

A 250 mg sample of finely ground soil (<75  $\mu\text{m}$ ) was weighed into a 100-mL Teflon vessel in which 9.0 mL of  $\text{HNO}_3$  and 3.0 mL of HCl were then added. 4.0 mL of HF was added, and the vessels were placed in sleeves, covered and placed in a rotor. The digestion rotor was placed in a microwave oven and digested at 1200 Watts at 100% power for 10 minutes at 180° C, then 1200 Watts at 70% power for 9.5 minutes at 180° C and then cooled and vented for 15 minutes. After venting, the vessels were removed from oven and 20 mL of  $\text{H}_3\text{BO}_3$ , 4.5%, was added. The vessels were then re-digested in microwave oven at: 1200 Watts, 100% power at 160° C, then maintained 160° C for 10 minutes, and allowed to vent and cool for 15 minutes. Then digestate were transferred to



100-mL nalgene volumetric flasks in which 1.9 %  $H_3BO_3$  was added to achieve a concentration of 2.1%  $H_3BO_3$ . The flasks were mixed and allowed to stand overnight to dissolve any metal fluorides. Then flasks were filled to volume with 1.9 %  $H_3BO_3$ , mixed and decanted into polypropylene containers. Working standards included blanks, reference soil sample and a National Institute of Standards and Technology (NIST) standard reference. Instrument alignment and gas pressures were checked prior to use to obtain optimum readings and to maximize signal to noise ratio. Detection limits were calculated by using 3 times the standard deviation of 10 readings of the blank standard solution and was considered the lower detection limits for each element. Any readings below detection were set to zero. Blank standard solution was used to dilute samples that exceeded the high concentration standard and the diluted sample was rerun for all elements. The calculation of element concentration ( $mg\ kg^{-1}$ ) from solution concentration ( $mg\ L^{-1}$ ) was determined from equation 21:

$$(21) \quad \text{AnalyteConcentration} = [A \times B \times C \times R \times 1000]/E$$

where, analyte concentration is the element concentration in the soil ( $mg\ kg^{-1}$ ); A is the sample extract reading ( $mg\ L^{-1}$ ); C is the dilution, if performed; R is the AD/OD ratio; 1000 is the conversion factor to kg-basis; and E is the sample weight.

### 3.2.10 Soil Organic Carbon

Soil organic carbon was not measured directly but was calculated by subtracting the inorganic carbon fraction (procedure 4E1a1a) from the total carbon content of the soil (procedure 4H2a1). SOC was calculated from equation 22:

$$(22) \quad SOC = \%TC - 0.12 \times \%CaCO_3$$

where, SOC is the soil organic carbon percentage; %TC is the total carbon content percentage (4H2a1); 0.12 is the percent of carbon in carbonate; and %CaCO<sub>3</sub> is the percent of carbonates in the sample.

### 3.2.11 Particle Size Distribution Analysis

Particle size analysis was performed for each site on samples from the central pedon or from a satellite if the central pedon did not have enough sample. This was done to see if clay content correlated with other soil properties. Clay content and sand content were determined according to procedure 3A1a1a.

### 3.3 Collection of Spectra

#### 3.3.1 Vis-Near Infrared Spectroscopy

VNIR spectra were collected on dry ground (<2 mm) samples. Soil samples were pressed into a sample holder and scanned 100 times and averaged to obtain the reflectance spectra. Equipment used were: LabSpec® 2500 spectrometer with a spectral range of 350-2500 nm range with 4 nm and 10 nm spectral resolution in the visible and near-infrared, respectively.(Analytical Spectral Devices (ASD) Inc., Boulder, CO); spectrometer light source, Muglight, Tungsten Quartz Halogen bulb, with sample tray (ASD Inc., Boulder, CO); fiber optic cables; fiber optic magnifying glass, Fiber Checker (ASD Inc., Boulder CO); sample pucks; Spectralon white reference panel (Labshpere, North Sutton, NH); wavelength standard (); sample press; number 7 rubber stopper; wavelength verification software; Indico Pro software (ASD Inc., Boulder, CO), spatula; oil-less air compressor; microfiber cloth; and isopropyl alcohol, 70 %.

Prior to spectral scans, the fiber optic cable was checked for any breaks by attaching one end of the fiber optic cable to the magnifying glass and holding the other end up to a light source. The light source was warmed for 3 hours before any scans were taken. Sample pucks and puck with white reference/wavelength standard were cleaned with a microfiber cloth and isopropyl alcohol before being used. The spectrometer was checked for accuracy with the wavelength verification software by collecting the baseline with the white reference and then scanning the wavelength standard. Soil samples were loaded loosely into sample pucks. Excess soil was scraped off the sample puck with the spatula. The sample puck face was placed down onto the number 7 rubber stopper and then the soil was pressed into the puck at about 46 psi. The white reference panel was run before scans to create a baseline and after every 10 samples, creating a new baseline to

reduce any effects of instrument drift. Spectra produced were the average of 100 reflectance scans (figure 3.2).

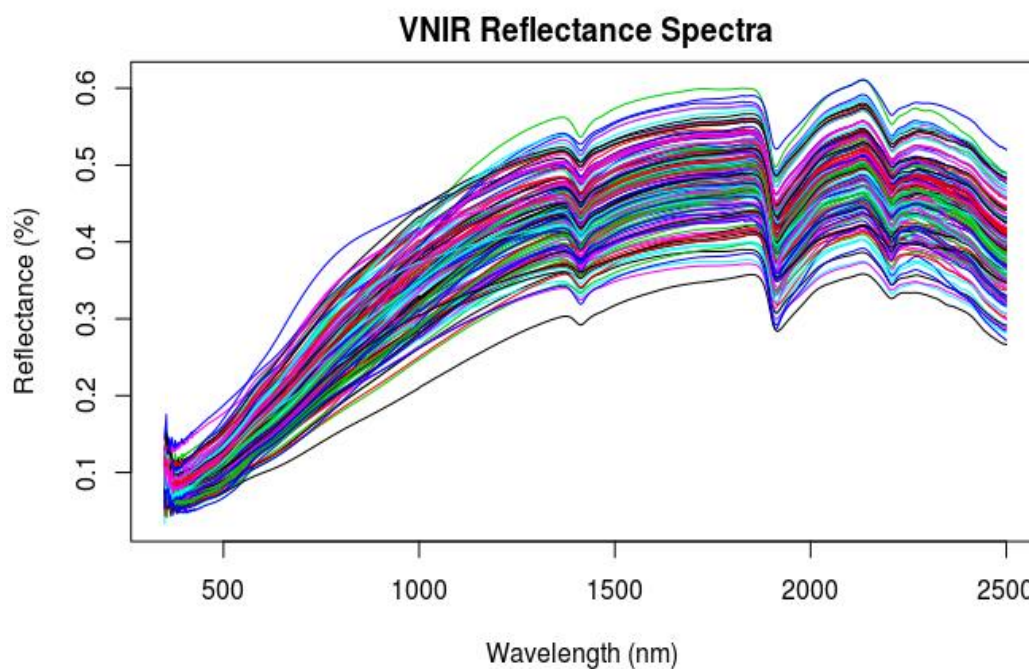


Figure 3.2 VNIR raw reflectance spectra (350-2500 nm) of 156 samples collected with LabSpec 2500 spectrometer.

The Labspec 2500 contains 3 detectors, with each operating in its own spectral range, 350 to 1000 nm, 1000-1830 nm and 1830-2500 nm. There was a shift in values at 1000 nm and 1830 nm that was corrected by linear interpolation. The “spliceCorrection” function in the “prospectr” package (Stevens & Ramirez-Lopez., 2013) in R (R Core Team., 2017) was used to correct the spectra (figure 3.3).

The VNIR reflectance spectra were converted to absorbance by equation 23:

$$(23) \quad A = \log(1/R)$$

where, A is absorbance (Au) and R is reflectance. The wavelengths below 400 nm were cut due to noise (figure 3.4).

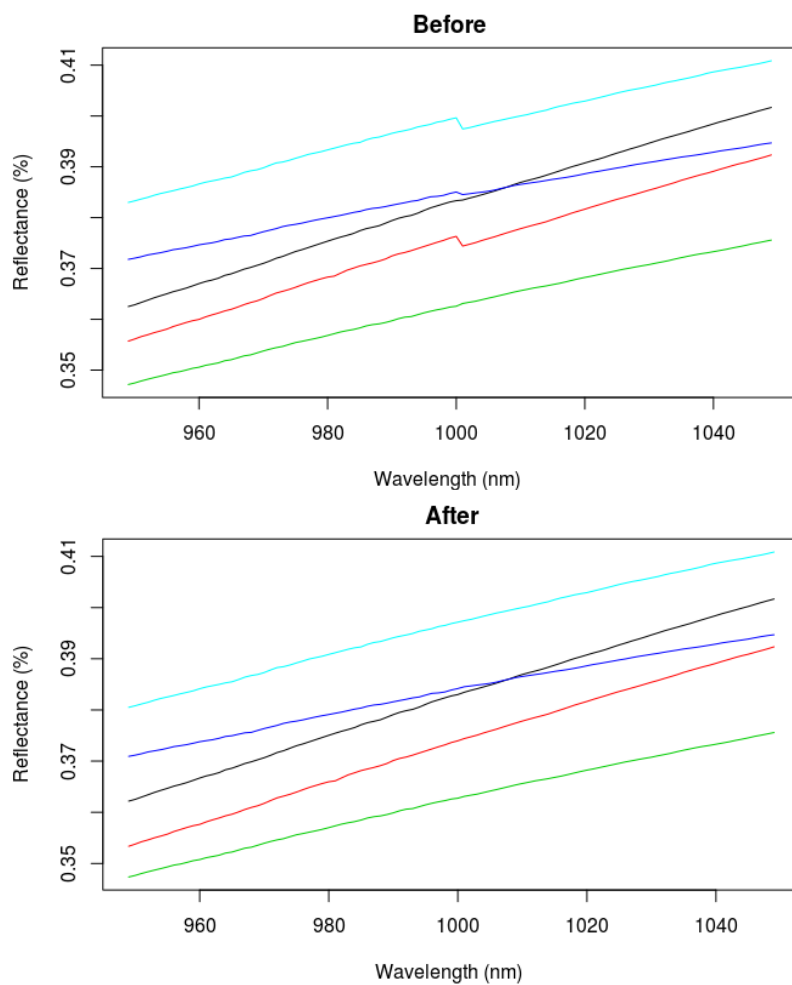


Figure 3.3 Example of VNIR spectra at 1000 nm before and after splice correction.

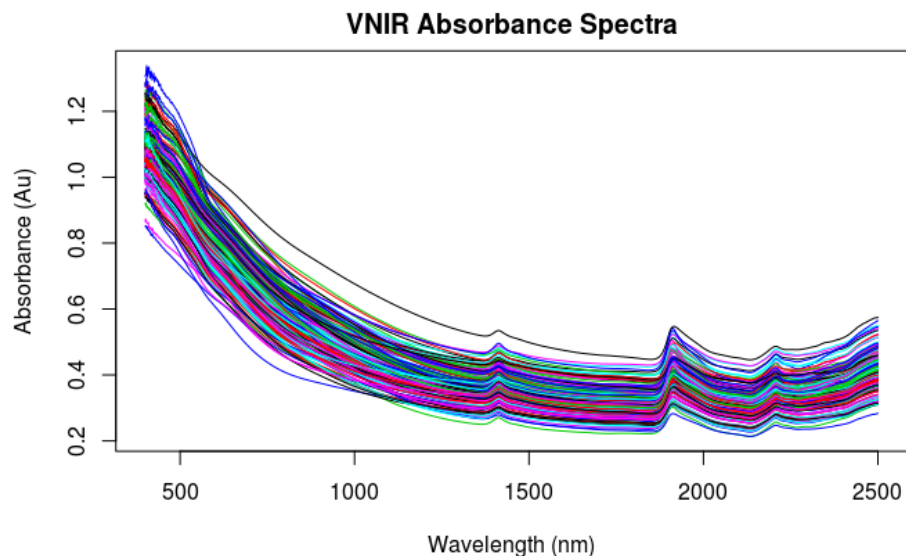


Figure 3.4 VNIR raw absorbance transformed spectra (400-2500 nm) of 156 samples collected.

### 3.3.2 Mid-Infrared Spectroscopy

MIR spectra were obtained on dry ground (<180  $\mu\text{m}$ ) samples. Four sub-samples were pressed into small pucks in a sample tray, scanned 100 times each and averaged. The resulting MIR spectra came from the average of the 4 sub-samples, however, one sample had a significant outlier between the spectral scans and was removed prior to averaging. Equipment used were: Vertex 70 spectrometer, with XTS sample port, spectral range of 7498-600  $\text{cm}^{-1}$  and 2  $\text{cm}^{-1}$  resolution (Bruker Corp., Billerica, MA); sample tray; press; Opus software (Bruker Corp., Billerica, MA); and inert cooling gas, nitrogen.

The spectrometer scans the built-in white reference in the sample tray prior to scanning a sample. Each MIR spectrum was the average of each puck which in turn was the average of 100 scans (figure 3.5). The NIR region between 7498 and 4000  $\text{cm}^{-1}$  were

removed from the spectra and the MIR region from 4000-600  $\text{cm}^{-1}$  was retained.

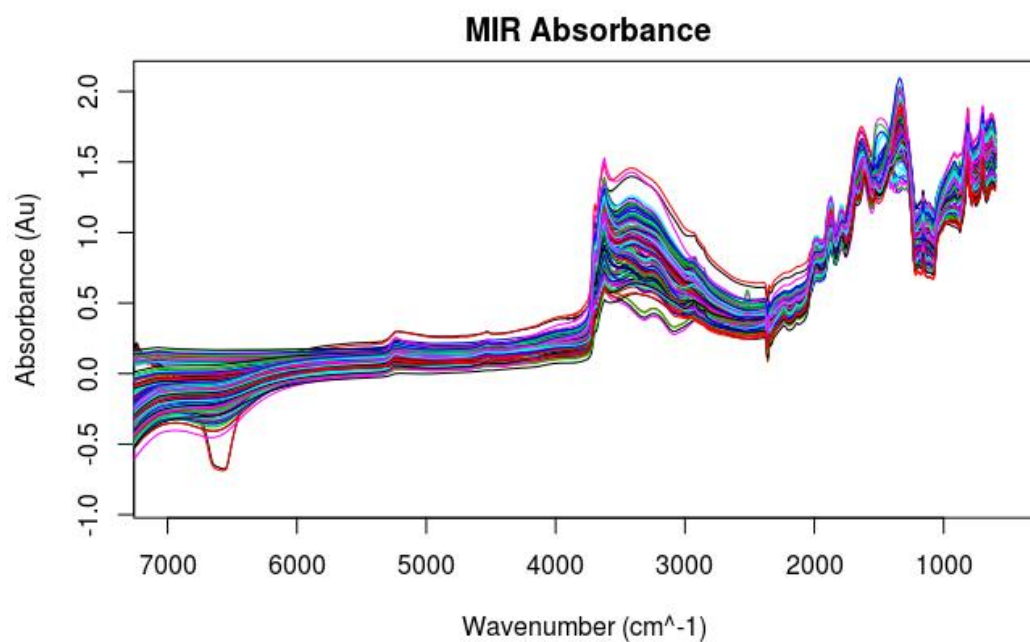


Figure 3.5 MIR raw absorbance spectra (7498 to 600  $\text{cm}^{-1}$ ) of 156 samples collected with Vertex 70v FTIR spectrometer.

### 3.4 Statistical Analyses

Basic descriptive statistics were performed for all soil properties to determine mean, range, standard deviation, and skew. Pearson's correlation coefficient was determined to see if there were relationships between the spectrally active soil properties (total carbon, nitrogen, soil organic content, and clay content) and the remaining soil properties.

#### 3.4.1 Spectral Preprocessing

Different spectral preprocessing techniques were applied to the spectra before modeling. Preprocessing techniques used to improve spectral signal and scatter correction were: multiplicative scatter correction (MSC), standard normal variate (SNV), spectral

detrending (Detrend), Savitsky-Golay smoothing (Smooth), 1<sup>st</sup> and 2<sup>nd</sup> derivatives performed after Savitsky-Golay smoothing (1-D and 2-D), MSC performed after Savitsky-Golay smoothing (SG + MSC), and SNV performed after Savitsky-Golay smoothing (SG + SNV). These were done using the statistical software R (version 3.4.3) (R Core Team., 2017) and the code can be found in Appendix A. Packages used for preprocessing were, “pls” (Mevik et al., 2016) and “prospectr” (Stevens & Ramirez-Lopez., 2013).

### 3.4.2 Model Creation and Validation

Three different calibration sampling strategies were tested, random sampling (RS), stratified sampling (SS), and Kennard Stone sampling (KSS). For each sampling method, the samples were first grouped by soil profile to avoid pseudo-replication. For the RS, a random seed generator in R was used to randomly select 75% of the samples into the training set while the other 25% were allocated to the test set. SS allocated all of the samples from the satellite pedons into the training set and the central pedons into the test set. The KSS algorithm projected the spectra into principal component space and sampled spectra based on furthest distance or dissimilarity from neighbor. The furthest neighbor was allocated into the test set and removed from the sample set, then the next furthest neighbor was removed and placed into the sampling set. This was done iteratively until the desired number of samples were selected. The spectra were projected into 20 dimensions via principal component analysis and the dissimilarity metric was determined by the Mahalanobis distance using the “KenStone” function from the “prospectr” package (Stevens & Ramirez-Lopez., 2013), allocating 81 samples to the training set and 75 samples to the test set. Within each calibration sampling scheme, the 9



different preprocessing techniques (raw absorbance, MSC, SNV, detrended, smoothed, smoothed + MSC, smoothed + SNV, SG 1<sup>st</sup> D, and SG 2<sup>nd</sup> D) were tested. Partial least squares regression models were built on the training set of each calibration sampling scheme and each of the above preprocessing techniques with the “pls” package in R (Mevik et al., 2016). The test set was then used to validate the model and to choose the number of latent variables retained. The number of latent variables chosen for the model was based on the number of components that had the lowest RMSEP when the test set was fitted. To reduce over complicating the model and overfitting, the number of components was restricted to a maximum of 20.

### 3.4.3 Model Performance and Selection

The criterion used to evaluate models were based on the coefficient of determination ( $R^2$ ), the ratio of performance deviation (RPD) and bias. RPD values were calculated by dividing the root mean square error of cross-validation (RMSECV), also known as RMSEP, from the standard deviation of the test set (equations 24 and 25),

$$(24) \quad RMSECV = \sqrt{\sum_{i=1}^m \frac{(\hat{y}_i - y_i)^2}{m}}$$

$$(25) \quad RPD = \frac{\sigma}{RMSECV}$$

where;  $m$  is the number of samples in the test set;  $\hat{y}_i$  is the predicted value,  $y_i$  the observed value, and  $\sigma$  is the standard deviation of the soil property in the test set (Bellon-Maurel et al., 2010). Bias of the models were also calculated to determine the trueness or accuracy of the models and calculated from equation 26:

$$(26) \quad Bias = \bar{\hat{y}} - \bar{y}$$

where,  $\bar{y}$  is the mean of the predicted values and  $\bar{y}$  is the mean of the actual values. Models with a bias close to 0 were preferred, indicating more accuracy (Bellon-Maurel et al., 2010).

The threshold RPD values used to test model performance in this study were ones developed by Chang et al., (2001), where excellent models have  $RPD > 2$ , fair models have RPD from 1.4-2, and non-reliable models were considered  $RPD < 1.4$ . Assessment of model performances were qualitatively defined by combining threshold values of RPD described by Chang et al. (2001) and  $R^2$  values adapted from Reeves and Smith (2009) (seen in table 3.1). RPD was given more weight than  $R^2$  values because it is a standardized metric that allows for comparing models across different study areas. Selection of the best models for each soil property within each calibration sampling method were first determined by RPD then by  $R^2$  value if RPD values were the same.

$R^2$	$RPD < 1.4$	$1.4 < RPD < 2$	$RPD > 2$
$< 0.7$	Very poor	Poor	-
0.7-0.8	Poor	Fair	Good
0.8-0.9	-	Good	Very good
$> 0.9$	-	-	Excellent

## CHAPTER 4. RESULTS

### 4.1. Basic Soil Properties

The basic statistics of the measured soil properties for the 156 samples are shown in table 4.1. Measured soil reference data can be found in Appendix B. Most soil properties had slightly skewed distributions and can be seen in figure 4.1. EC, extractable  $\text{Na}^+$ , and  $\text{CaCO}_3$  displayed highly skewed distributions, were most samples had undetectable concentrations and there were relatively few samples with high concentrations. Table 4.2 and figures 4.2-4.3 show the correlations between spectrally active soil properties and the other measured soil properties.  $\beta\text{G}$  had the highest correlations with spectrally active properties related to carbon and nitrogen ( $r > 0.7$ ). CEC had strong correlation with both clay content ( $r = 0.94$ ) and total N ( $r = 0.74$ ). Clay content had moderate correlations with  $\text{Ca}^{2+}$  ( $r = 0.52$ ) and total P ( $r = 0.62$ ). All soil properties had very poor relationships associated with  $\text{CaCO}_3$ , except for EC,  $\text{Ca}^{2+}$  and  $\text{Na}^+$  which were moderately correlated. The full correlation table for all measured soil properties and scatter plots showing the relationship between spectrally active and the remaining soil properties can be found in Appendix C. Table 4.3 and Table 4.4 show the model performances of the best calibration models for each soil property and indicate which calibration sampling strategy and spectral preprocessing technique was used for VNIR and MIR respectively.

Table 4.1 Descriptive statistics of measured soil properties for 156 soil samples in this study

Property	Mean	Median	Std. Dev	Min	Max	Range	Skew
EC (dS m <sup>-1</sup> )	0.40	0.28	0.53	0.05	5.62	5.57	6.73
1:1 H <sub>2</sub> O pH	6.16	6.00	0.90	4.44	8.58	4.14	0.75
1:2 CaCl <sub>2</sub> pH	5.70	5.49	0.96	3.94	8.29	4.35	0.69
CEC (cmol(+) kg <sup>-1</sup> )	18.52	19.58	9.30	2.21	40.46	38.25	-0.04
Ca <sup>2+</sup> (cmol(+) kg <sup>-1</sup> )	16.77	14.93	13.03	1.55	65.19	63.64	1.52
Mg <sup>2+</sup> (cmol(+) kg <sup>-1</sup> )	3.00	2.50	2.43	0.30	15.76	15.46	2.44
Na <sup>+</sup> (cmol(+) kg <sup>-1</sup> )	0.18	0.00	1.19	0.00	11.69	11.69	7.97
K <sup>+</sup> (cmol(+) kg <sup>-1</sup> )	1.04	0.98	0.82	0.00	4.88	4.88	1.47
H <sub>2</sub> O P (mg kg <sup>-1</sup> )	1.01	0.90	0.34	0.73	3.23	2.51	3.10
Total P (mg kg <sup>-1</sup> )	476.20	481.47	214.56	44.01	1041.29	997.28	0.07
Total C (%)	2.03	1.71	1.25	0.24	6.99	6.75	1.07
SOC (%)	1.97	1.63	1.82	0.22	5.13	4.91	0.81
CaCO <sub>3</sub> (%)	0.50	0.04	1.83	0.00	16.83	16.83	6.29
β-glucosidase (μg g <sup>-1</sup> h <sup>-1</sup> )	103.93	80.00	84.85	0.00	434.00	434.00	1.04
Total N (%)	0.20	0.17	0.11	0.03	0.51	0.48	0.81
Clay (%)*	24.29	25.31	12.85	2.86	50.10	47.24	-0.15
Sand (%)*	31.99	19.12	32.34	1.04	95.82	94.78	0.85

\* Analyses were performed on samples from central pedons, n = 53.

Table 4.2 Correlation with spectrally active properties.

	Total C	SOC	CaCO <sub>3</sub>	Total N	Clay %*
EC	0.52	0.45	0.56	0.48	0.14
1:1 H <sub>2</sub> O pH	0.20	0.14	0.38	0.15	0.21
1:2 H <sub>2</sub> O pH	0.27	0.20	0.47	0.22	0.24
CEC	0.66	0.68	0.07	0.72	0.94
Ca <sup>2+</sup>	0.59	0.50	0.64	0.54	0.52
Mg <sup>2+</sup>	0.41	0.35	0.45	0.41	0.47
Na <sup>+</sup>	0.28	0.19	0.56	0.20	0.02
K <sup>+</sup>	0.39	0.37	0.25	0.46	0.44
H <sub>2</sub> O P	0.27	0.31	-0.08	0.35	0.07
Total P	0.49	0.48	0.16	0.54	0.62
βG	0.70	0.72	0.11	0.74	0.31

\*Correlations with central pedon samples.

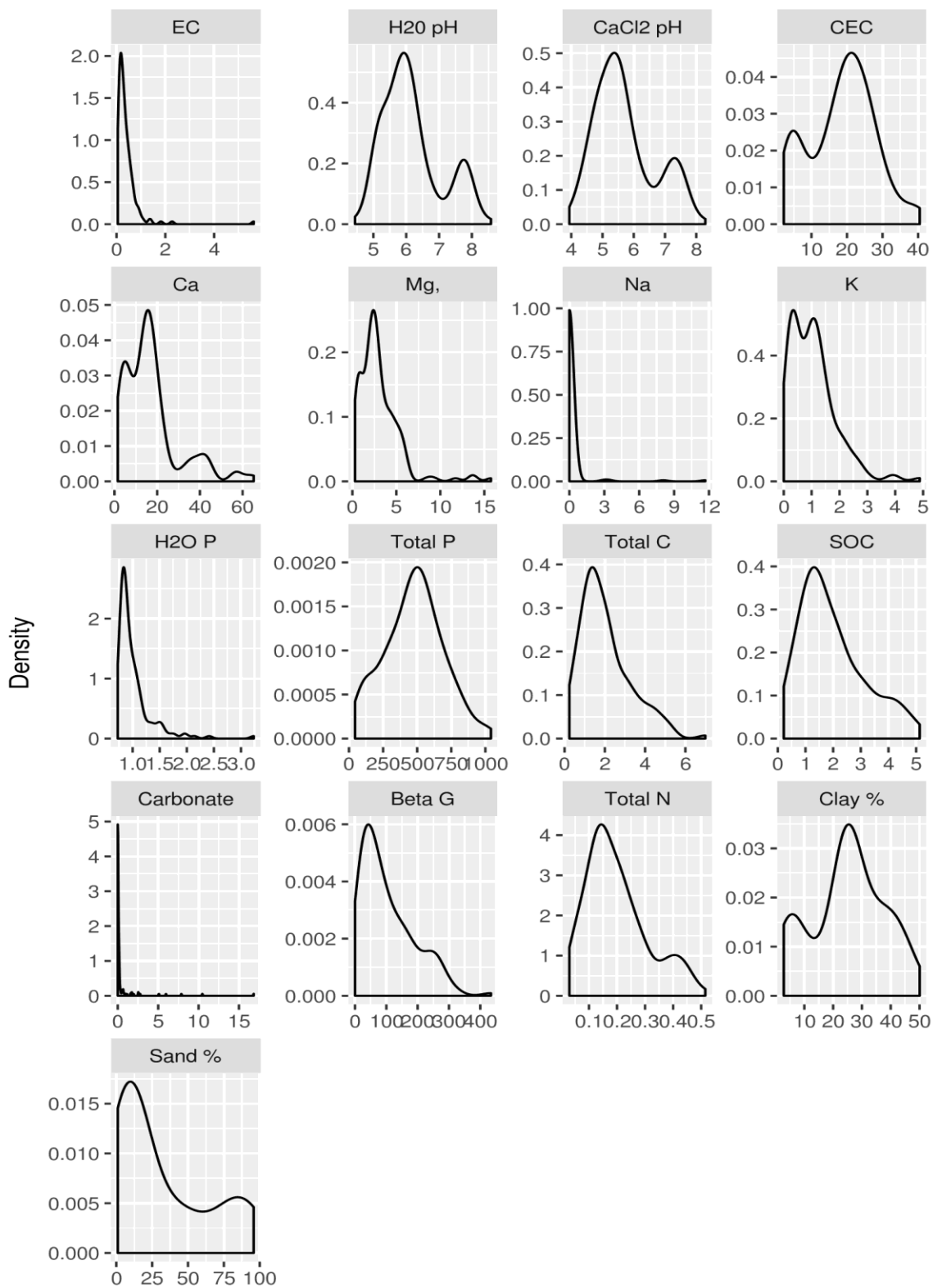


Figure 4.1 Distribution and range of measured soil properties for 156 samples (n = 53 for Clay% and Sand%).

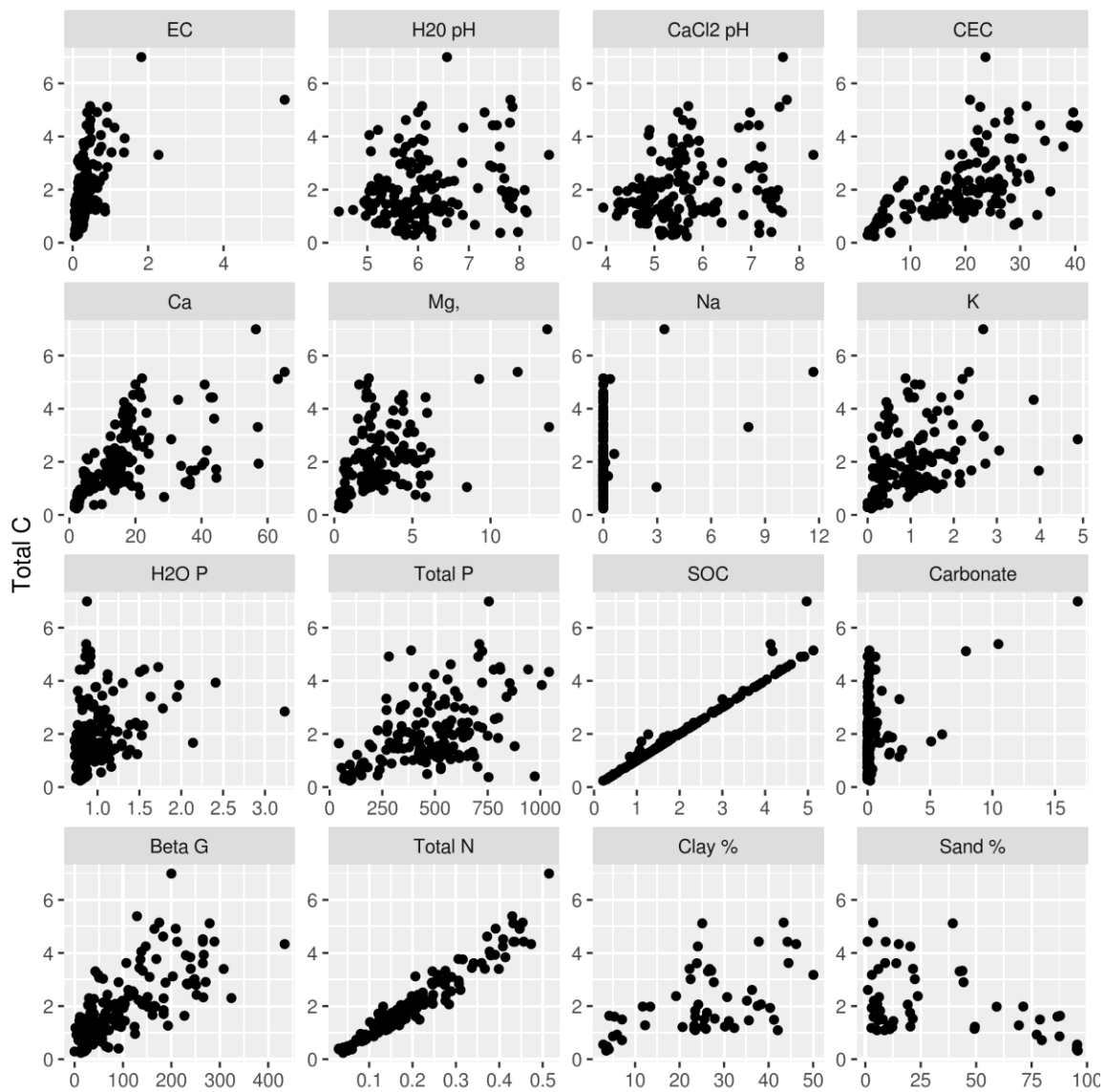


Figure 4.2 Scatterplots showing the relationships between total C and measured soil properties for 156 samples (n = 53 for Clay% and Sand%).

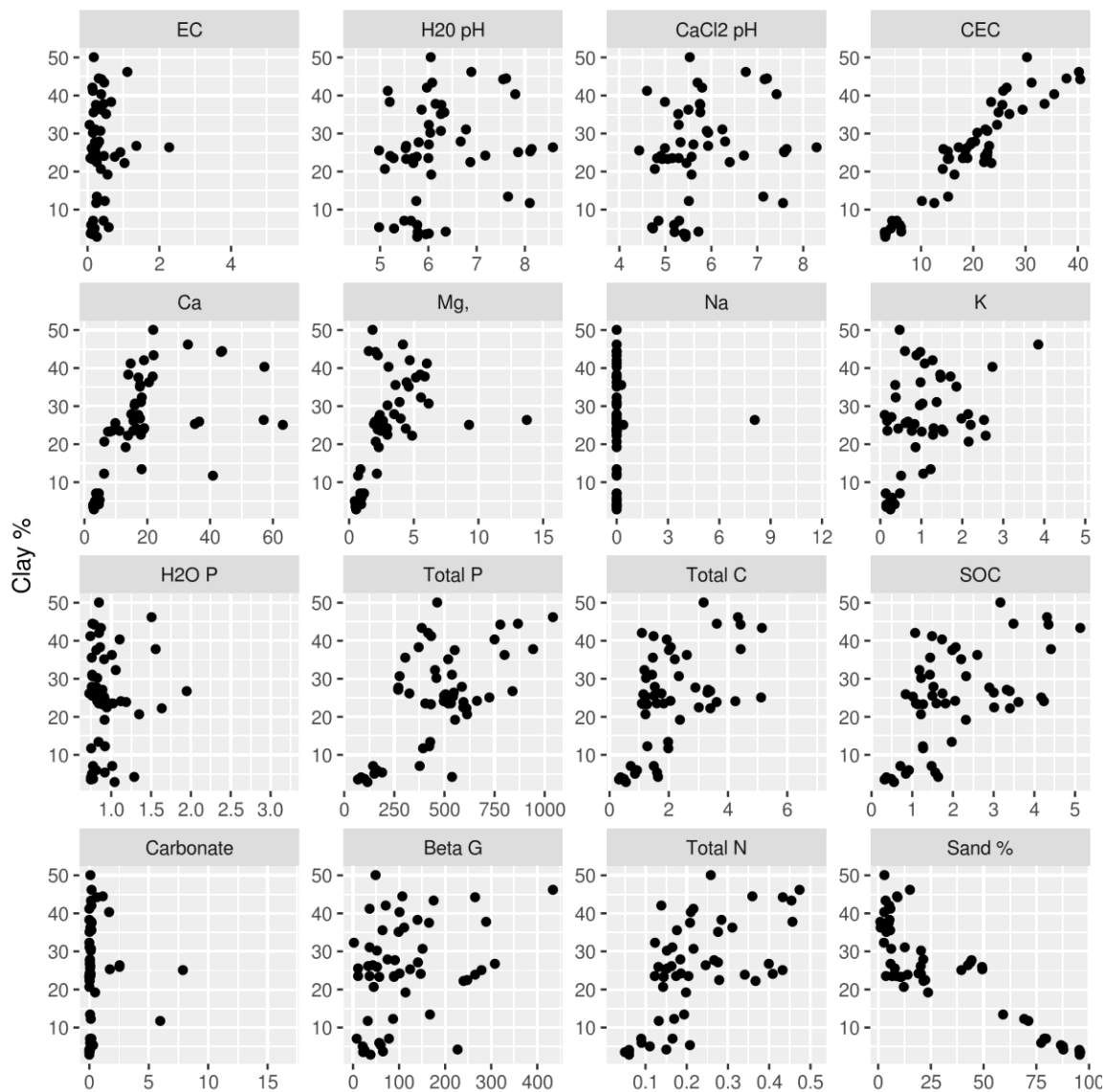


Figure 4.3 Scatterplots showing the relationships between total C and measured soil properties for 156 samples ( $n = 53$  for Clay% and Sand%).

Table 4.3 Goodness of fit summary for overall best VNIR models.

Property	Train R <sup>2</sup>	RMSE	Test R <sup>2</sup>	RMSECV	RPD	Bias	Method
EC	0.83	0.24	0.60	0.24	1.60	-0.03	SS
1:1 H <sub>2</sub> O pH	0.69	0.51	0.70	0.46	1.83	0.01	KSS
1:2 CaCl <sub>2</sub> pH	0.79	0.45	0.73	0.47	1.93	0.00	KSS
CEC	0.94	2.23	0.82	4.11	2.39	-0.28	RS
Ca <sup>2+</sup>	0.86	4.58	0.81	6.28	2.29	-0.47	SS
Mg <sup>2+</sup>	0.93	0.54	0.75	1.15	2.06	-0.03	SS
Na <sup>+</sup>	0.83	0.50	0.63	0.67	1.65	-0.02	SS
K <sup>+</sup>	0.74	0.42	0.69	0.45	1.81	0.02	SS
H <sub>2</sub> O P	0.70	0.18	0.54	0.23	1.49	0.01	RS
Total P	0.41	154.90	0.58	157.20	1.57	-5.47	RS
Total C	0.87	0.43	0.84	0.55	2.49	-0.16	RS
SOC	0.84	0.44	0.85	0.53	2.62	-0.16	RS
CaCO <sub>3</sub>	0.77	0.77	0.48	1.46	1.40	-0.15	KSS
βG	0.72	42.85	0.67	52.51	1.75	6.45	SS
Total N	0.79	0.05	0.88	0.04	2.92	-0.01	RS

Table 4.4 Goodness of fit summary for overall best MIR models.

Property	Train R <sup>2</sup>	RMSE	Test R <sup>2</sup>	RMSECV	RPD	Bias	Method
EC	0.98	0.09	0.63	0.19	1.66	-0.02	RS
1:1 H <sub>2</sub> O pH	0.85	0.39	0.81	0.33	2.29	-0.02	KSS
1:2 CaCl <sub>2</sub> pH	0.98	0.15	0.91	0.25	3.35	0.02	KSS
CEC	1.00	0.57	0.99	1.01	8.82	0.15	KSS
Ca <sup>2+</sup>	0.99	1.43	0.96	1.99	5.10	-0.09	KSS
Mg <sup>2+</sup>	0.95	0.45	0.89	0.77	3.07	-0.01	SS
Na <sup>+</sup>	0.62	0.76	0.50	0.78	1.42	0.21	SS
K <sup>+</sup>	0.67	0.49	0.83	0.25	2.48	-0.01	RS
H <sub>2</sub> O P	0.43	0.25	0.50	0.24	1.43	-0.04	RS
Total P	0.52	145.70	0.72	116.90	1.90	-16.06	SS
Total C	0.99	0.11	0.98	0.21	6.70	-0.04	RS
SOC	0.99	0.10	0.98	0.18	7.72	-0.04	RS
CaCO <sub>3</sub>	1.00	0.12	0.88	0.44	2.86	-0.14	KSS
βG	0.86	29.91	0.86	39.44	2.33	-5.89	SS
Total N	0.98	0.01	0.96	0.02	5.15	-0.00	RS



#### 4.1.1 Electrical Conductivity

EC ranged from 0.05 to 5.62 dS m<sup>-1</sup> and had a mean of 0.40 dS m<sup>-1</sup>. The best observed VNIR model, sampled by SS and preprocessed by detrending the spectra was considered poor (RPD = 1.60 and R<sup>2</sup> = 0.60). The best MIR model, sampled by RS and preprocessed by the 2<sup>nd</sup> derivative, was considered poor (RPD = 1.66 and R<sup>2</sup> = 0.63). Both models performed similarly, while the MIR model had a better training fit and smaller errors overall. Calibration models are shown in figure 4.4.

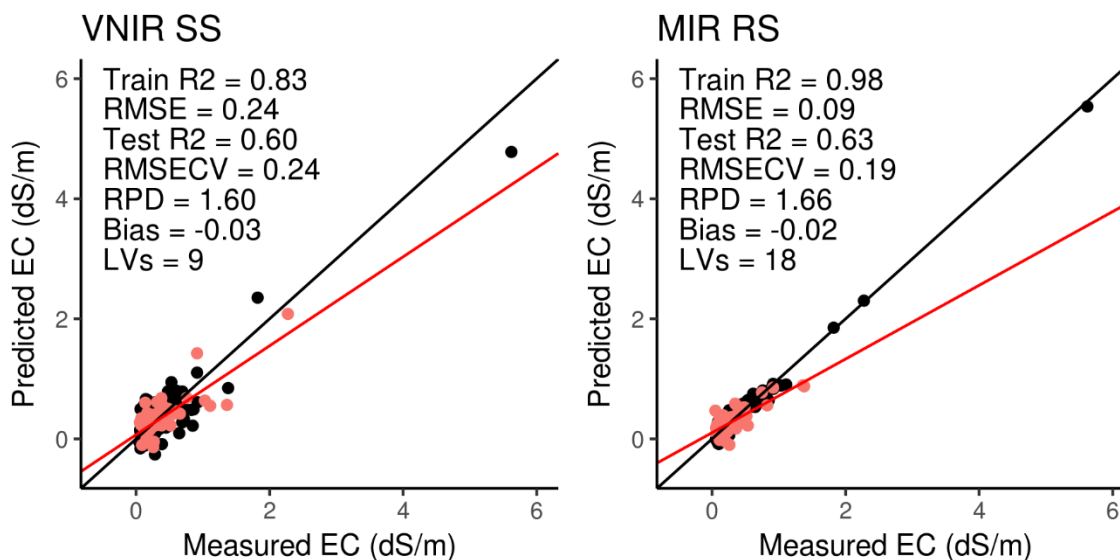


Figure 4.4. Goodness of fit plots showing relationship of PLS predicted EC values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

#### 4.1.2 Soil pH

1:1 H<sub>2</sub>O pH ranged from 4.44 to 8.58 and had a mean of 6.16. The best VNIR model, sampled by KSS and preprocessed by SNV, was considered fair (RPD = 1.83 and an R<sup>2</sup> = 0.70). The best MIR model, sampled by KSS and preprocessed by SG-smoothing + MSC, was considered good (RPD = 2.29 and R<sup>2</sup> = 0.81). MIR performed better than

VNIR, however, both had significant outliers in the training set that failed to be fitted and resulted in higher RMSE than RMSECV values. Both models had a tendency to overpredict at low pH and underpredict at high pH. Calibration models are shown in figure 4.5.

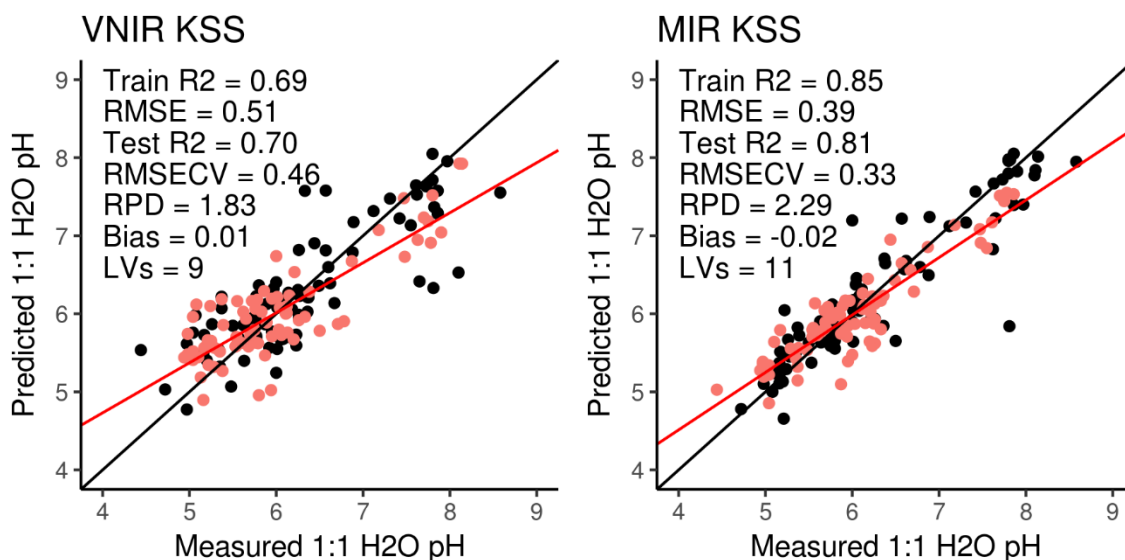


Figure 4.5. Goodness of fit plots showing relationship of PLS predicted 1:1 H<sub>2</sub>O pH values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R<sup>2</sup> (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

1:2 CaCl<sub>2</sub> pH ranged from 3.94 to 8.29 and had a mean of 5.7. The best VNIR model, sampled by KSS and preprocessed by SNV, was considered fair (RPD = 1.93 and R<sup>2</sup> = 0.73). The best MIR model, sampled by KSS and preprocessed by detrending, was considered excellent (RPD = 3.35 and R<sup>2</sup> = 0.91). MIR had very good training and test fits, while VNIR had a few notable outliers in the training set. The MIR model had prediction errors of a quarter pH unit while VNIR errors were less than half a pH unit. Calibration models are shown in figure 4.6.

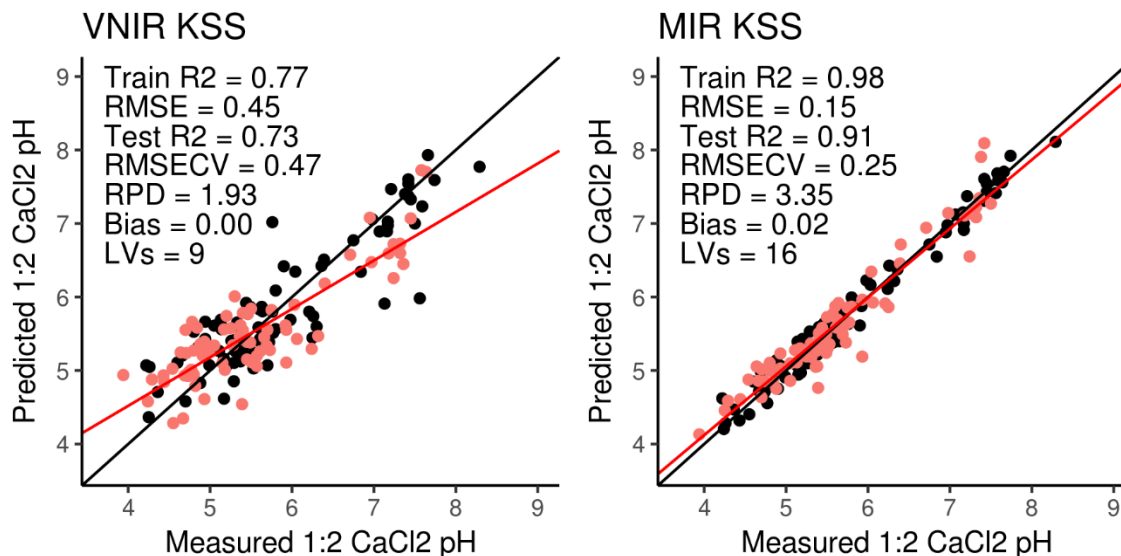


Figure 4.6. Goodness of fit plots showing relationship of PLS predicted 1:2 CaCl<sub>2</sub> pH values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R<sup>2</sup> (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

#### 4.1.3 Cation Exchange Capacity and Exchangeable Cations

CEC ranged from 2.21 to 40.46 cmol(+) kg<sup>-1</sup> and had a mean of 18.52 cmol(+) kg<sup>-1</sup>

<sup>1</sup>. The best VNIR model, sampled by RS and no preprocessing of spectra, was considered very good (RPD = 2.39 and R<sup>2</sup> = 0.82). Calibration models are shown in figure 4.7. The best MIR model, sampled by KSS and the first derivative, was considered excellent (RPD = 8.82 and R<sup>2</sup> = 0.99). Both models performed well, with the MIR model having a much better fit and lower errors. VNIR had outliers that were generally underpredicted.

Calibration models are shown in figure 4.7.

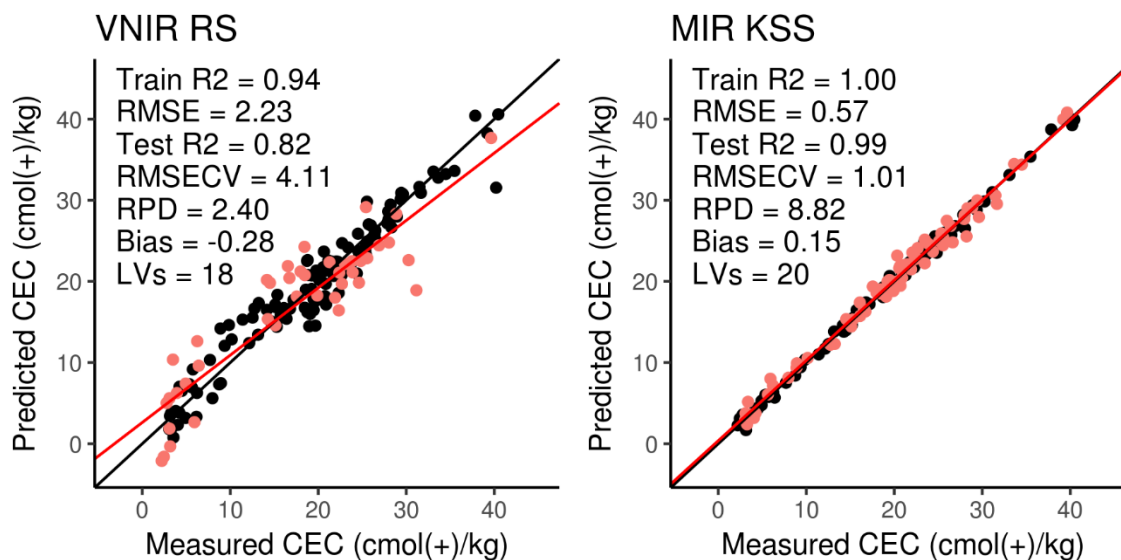


Figure 4.7. Goodness of fit plots showing relationship of PLS predicted CEC values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

$\text{Ca}^{2+}$  ranged from 1.55 to 65.19  $\text{cmol}(+)/\text{kg}^{-1}$  and had a mean of 16.77  $\text{cmol}(+)/\text{kg}^{-1}$

<sup>1</sup>. The best VNIR model, sampled by SS and preprocessed by SG-smoothing, was considered very good (RPD = 2.29 and  $R^2 = 0.81$ ). The best MIR model, sampled by KSS and preprocessed by detrending, was considered excellent (RPD = 5.10 and  $R^2 = 0.96$ ). While both models performed well, the MIR model had a much better fit. The VNIR model had error values three times as high as the MIR errors and had notable outliers in the training and test set. Samples tended to be underpredicted at high concentrations and a few samples were predicted as having negative concentrations in both the training and test sets with the VNIR model. Calibration models are shown in figure 4.8.

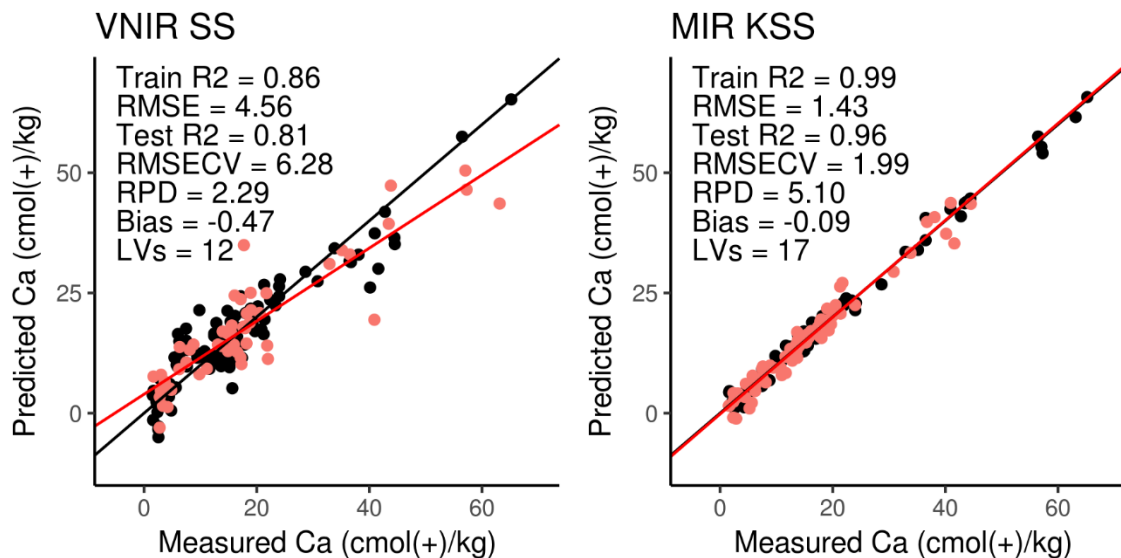


Figure 4.8. Goodness of fit plots showing relationship of PLS predicted  $\text{Ca}^{2+}$  values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

$\text{Mg}^{2+}$  ranged from 0.30 to 15.76  $\text{cmol}(+) \text{kg}^{-1}$  and had a mean of 3.00  $\text{cmol}(+) \text{kg}^{-1}$

<sup>1</sup>. The best VNIR model, sampled by SS and preprocessed by smoothing + SNV, was considered good (RPD = 2.06 and  $R^2 = 0.75$ ). The best MIR model, sampled by SS and preprocessed by detrending, was considered very good (RPD = 3.01 and  $R^2 = 0.89$ ). Both VNIR and MIR models performed well and had similar fit and error values in the training set. The MIR model had better predictive ability, with better observed test set fit and error values. However, the MIR model had a few outliers that were underpredicted. The VNIR model had outliers that were both over and underpredicted, and like MIR, tended to underpredict samples at higher concentrations. Calibration models are shown in figure 4.9.

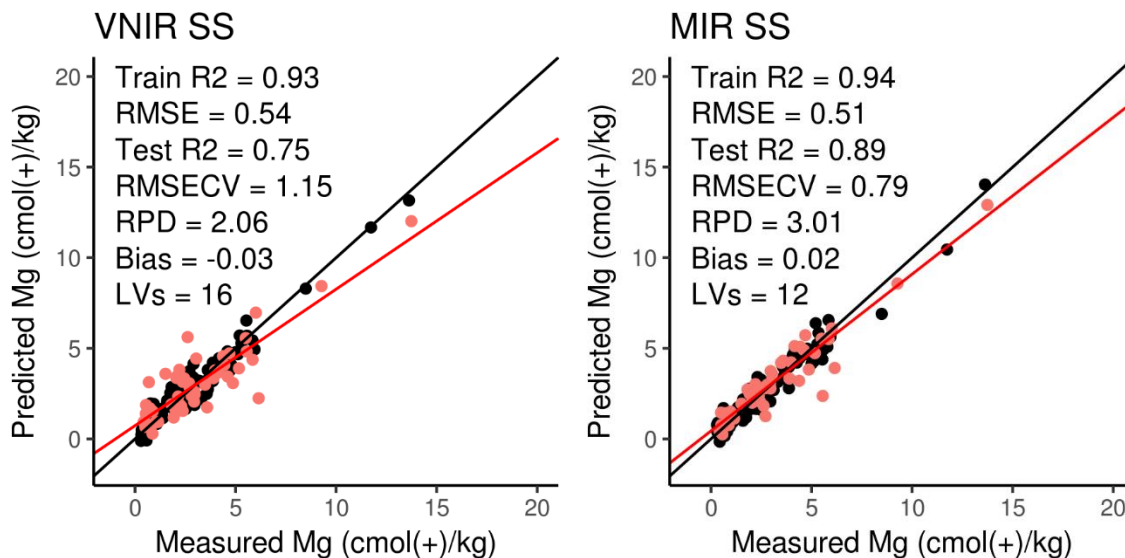


Figure 4.9. Goodness of fit plots showing relationship of PLS predicted  $\text{Mg}^{2+}$  values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

$\text{Na}^+$  ranged from 0.00 to 11.69  $\text{cmol}(+) \text{kg}^{-1}$  and had a mean of 0.18  $\text{cmol}(+) \text{kg}^{-1}$ .

The best VNIR model, sampled by SS and preprocessed by SNV, was considered poor (RPD = 1.65 and  $R^2 = 0.63$ ). The best MIR model, sampled by SS and preprocessed by the 2<sup>nd</sup> Derivative, was considered poor (RPD = 1.42 and  $R^2 = 0.50$ ). VNIR had better training fit than MIR and subsequently had slightly better predictions than MIR. Both VNIR and MIR had significant outliers in the training set, where both VNIR and MIR failed to fit the same samples by either over predicting or under predicting. Likewise, predictions of the test sets were poor, with samples having no measurable concentrations of  $\text{Na}^+$  both being under and overpredicted. Calibration models are shown in figure 4.10.

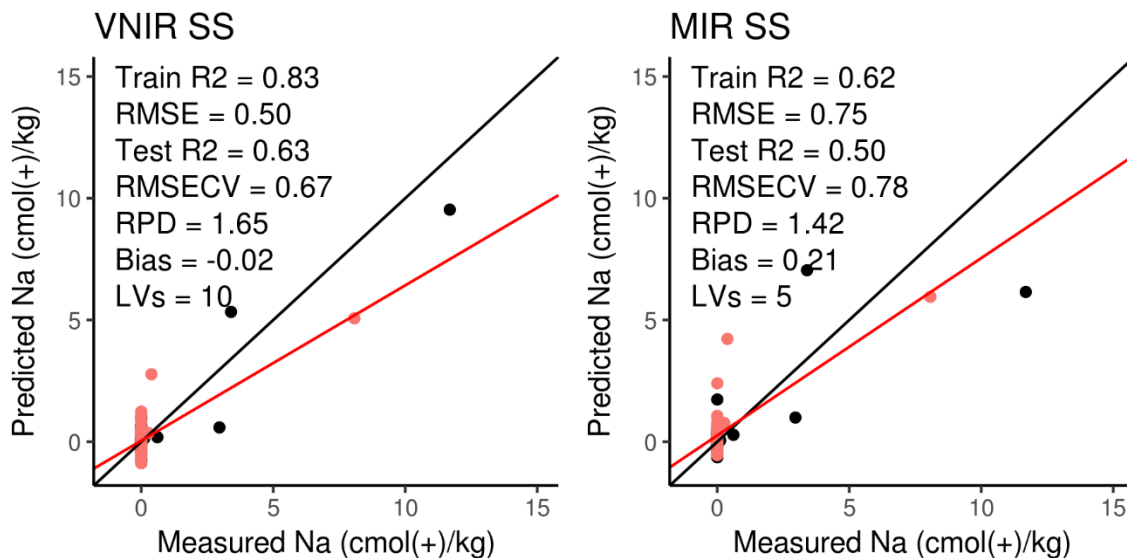


Figure 4.10. Goodness of fit plots showing relationship of PLS predicted  $\text{Na}^+$  values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

$\text{K}^+$  ranged from 0.00 to 4.88  $\text{cmol}(+) \text{kg}^{-1}$  and had a mean of 1.04  $\text{cmol}(+) \text{kg}^{-1}$ .

The best VNIR model, sampled by SS and preprocessed by SG-smoothing + SNV, was considered poor to fair (RPD = 1.81 and  $R^2 = 0.69$ ). The best MIR model, sampled by RS and preprocessed by detrending, was considered very good (RPD = 2.47 and  $R^2 = 0.83$ ). The VNIR model had a better training fit than the MIR model ( $R^2 = 0.74$  and  $R^2 = 0.68$  for VNIR and MIR respectively) but had poorer predictions when the test set was fitted, while the MIR test set was better fit than the training set. Both VNIR and MIR training sets tended to underpredict  $\text{K}^+$  at higher concentrations with. Calibration models are shown in figure 4.11.

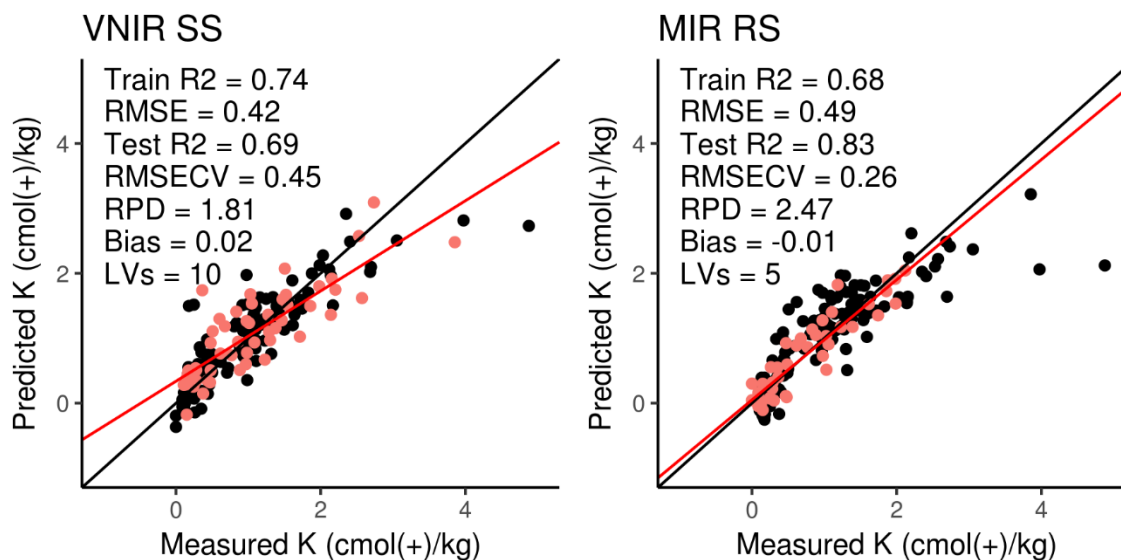


Figure 4.11. Goodness of fit plots showing relationship of PLS predicted K<sup>+</sup> values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R<sup>2</sup> (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

#### 4.1.4 Phosphorus

Water soluble P ranged from 0.73 to 3.23 mg kg<sup>-1</sup> and had a mean of 1.01 mg kg<sup>-1</sup>.

<sup>1</sup>. The best VNIR model, sampled by RS and preprocessed by SG-1<sup>st</sup> derivative, was considered poor (RPD = 1.49 and R<sup>2</sup> = 0.54). The best MIR model, sampled by RS and preprocessed by SG-2<sup>nd</sup> derivative, was considered poor (RPD = 1.43 and R<sup>2</sup> = 0.50). Both models performed similarly but VNIR had better training fit and slightly better predictions than MIR. Both VNIR and MIR models tended to underpredict H<sub>2</sub>O P at higher concentrations. Calibration models are shown in table 4.12.



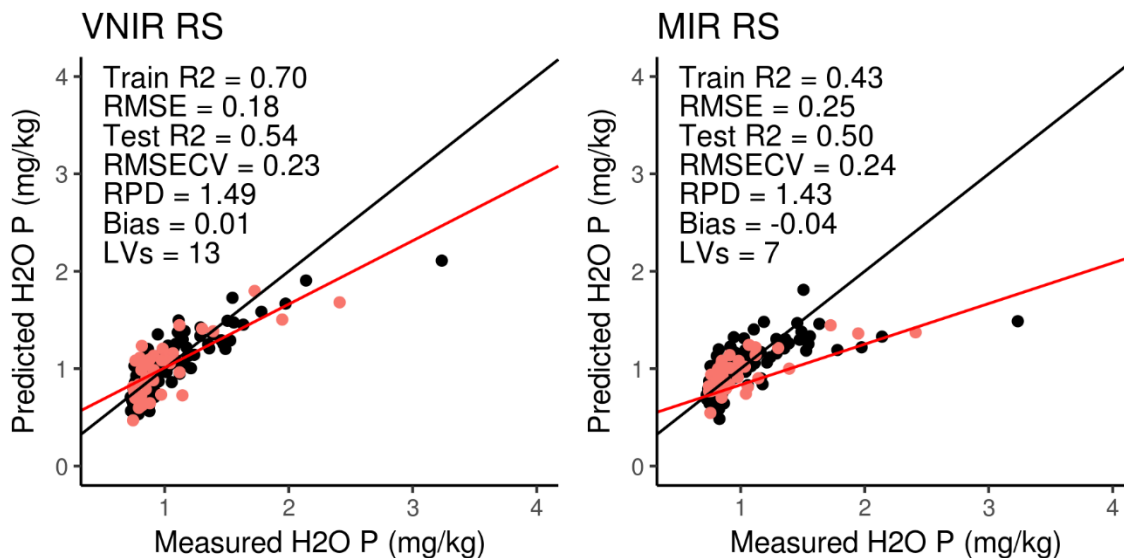


Figure 4.12. Goodness of fit plots showing relationship of PLS predicted H<sub>2</sub>O P values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R<sup>2</sup> (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

Total P ranged from 44 to 1041.3 mg kg<sup>-1</sup> and had a mean of 476.2 mg kg<sup>-1</sup>. The best VNIR model, sampled by RS and preprocessed by SNV, was considered poor (RPD = 1.57 and R<sup>2</sup> = 0.58). The best MIR model, sampled by SS and preprocessed by SG-2nd derivative, was considered fair (RPD = 1.90 and R<sup>2</sup> = 0.72). The MIR model performed better than the VNIR model, but both had poor fit of the test set with many outliers. The VNIR model predictions tended to be overpredicted at low to mid concentrations and underpredicted at higher concentrations. The MIR model showed a similar trend but was not as extreme. Calibration models are shown in figure 4.13.

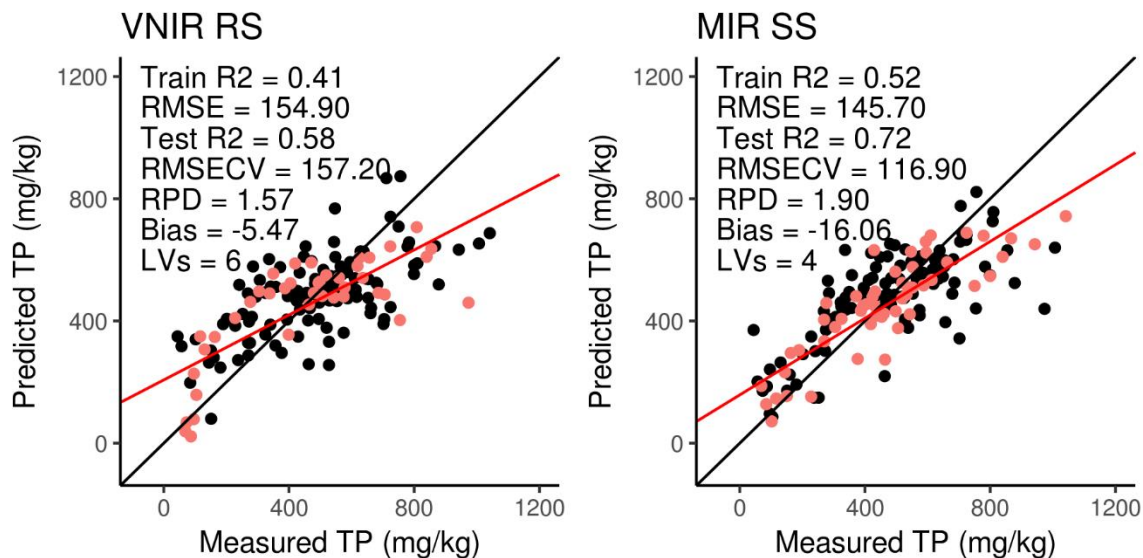


Figure 4.13. Goodness of fit plots showing relationship of PLS predicted Total P values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

#### 4.1.5 Carbon

Total C ranged from 0.24 to 6.99 % and had a mean of 2.03 %. The best VNIR model, sampled by RS and no preprocessing, was considered very good (RPD = 2.49 and  $R^2 = 0.84$ ). The best MIR model, sampled by RS and preprocessed by MSC, was considered excellent (RPD = 6.70 and  $R^2 = 0.98$ ). Total carbon was well predicted by both VNIR and MIR, with MIR predictions significantly better, with prediction error values of 0.20%, half that of those observed with VNIR. Calibration models are shown in figure 4.14.

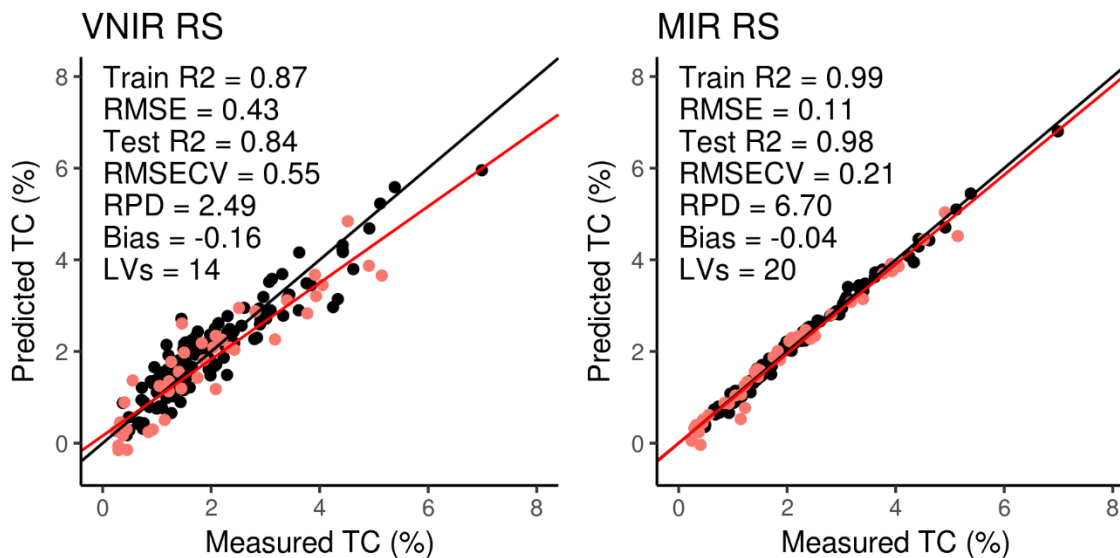


Figure 4.14. Goodness of fit plots showing relationship of PLS predicted Total C values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R<sup>2</sup> (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

SOC ranged from 0.22 to 5.13 % and had a mean of 1.97%. The best VNIR model, sampled by RS and no preprocessing, was considered very good (RPD = 2.62 and R<sup>2</sup> = 0.85). The best MIR model, sampled by RS and preprocessed by detrending, was considered excellent (RPD = 7.80 and R<sup>2</sup> = 0.98). MIR predictions for SOC were better than those for VNIR. The VNIR model performed well and had SOC predictions within half a percent. VNIR predictions tended to be underpredicted at higher concentrations. At low SOC concentrations, both VNIR and MIR tended to underpredict. Calibration models are shown in figure 4.15.

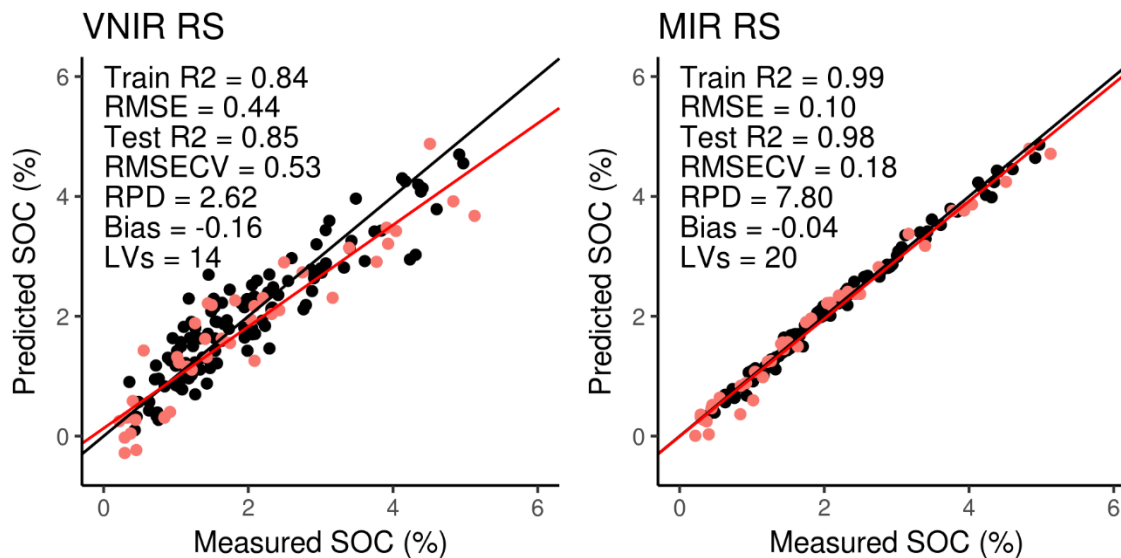


Figure 4.15. Goodness of fit plots showing relationship of PLS predicted SOC values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

CaCO<sub>3</sub> ranged from 0.00 to 16.83 % and had a mean of 0.50 %. The best VNIR model, sampled by KSS and preprocessed by SG-1st derivative, was considered very poor (RPD = 1.40 and R<sup>2</sup> = 0.48). The best MIR model, sampled by KSS and preprocessed by the 2nd derivative, was considered very good (RPD = 2.86 and R<sup>2</sup> = 0.88). MIR significantly had better predictions than VNIR, were observed predictions for CaCO<sub>3</sub> were within a half of a percent. Most of the error occurred at low concentrations or when there was no measurable CaCO<sub>3</sub>. VNIR had poor fit and many outliers in the training set and tended to underpredict CaCO<sub>3</sub> concentrations. Calibration models are shown in figure 4.16.

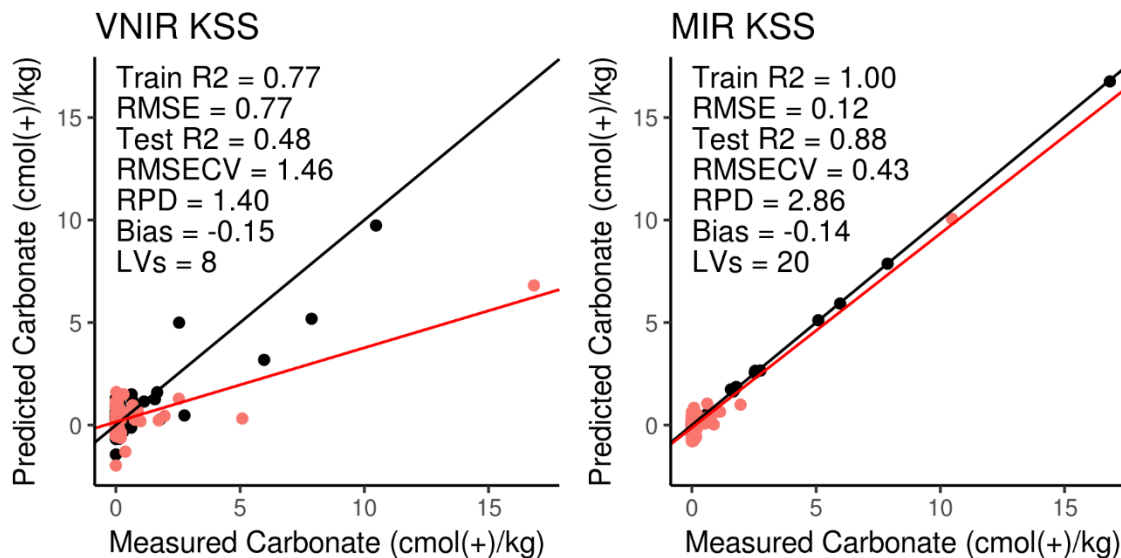


Figure 4.16. Goodness of fit plots showing relationship of PLS predicted  $\text{CaCO}_3$  values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

#### 4.1.6 $\beta$ -Glucosidase

$\beta$ G activity ranged from 0.00 to 434  $\mu\text{g g}^{-1} \text{h}^{-1}$  and had a mean of 103.93  $\mu\text{g g}^{-1} \text{h}^{-1}$

<sup>1</sup>. The best VNIR model, sampled by SS and no preprocessing, was considered poor to fair (RPD = 1.75 and  $R^2 = 0.67$ ). The best MIR model, sampled by SS and preprocessed by detrending, was considered very good (RPD = 2.47 and  $R^2 = 0.83$ ). MIR performed better than VNIR and had better training set fit. However, the training set had a few notable outliers that were underpredicted. The VNIR model also underpredicted the same samples as MIR. In the test set, VNIR tended to overpredict low  $\beta$ G activity and MIR underpredicted. Both VNIR and MIR underpredicted  $\beta$ G activity at high activity.

Calibration models are shown in figure 4.17.

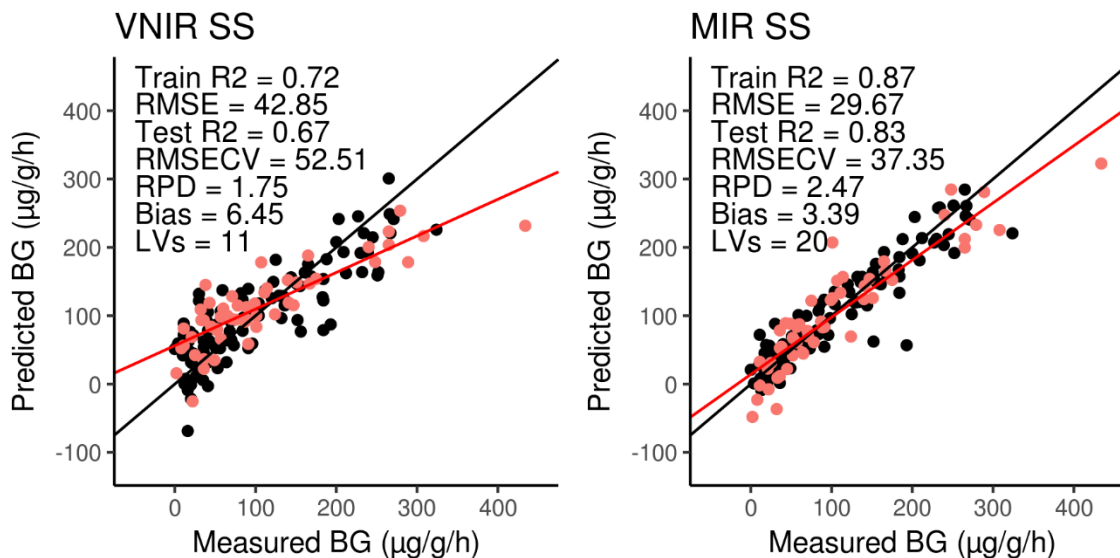


Figure 4.17. Goodness of fit plots showing relationship of PLS predicted  $\beta$ G activity from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

#### 4.1.7 Nitrogen

Total N ranged from 0.03 to 0.51 % and had a mean of 0.20 %. The best VNIR model, sampled by RS and preprocessed by SG-smoothing, was considered very good (RPD = 2.92 and  $R^2 = 0.88$ ). The best MIR model, sampled by RS and no preprocessing, was considered excellent (RPD = 5.15 and  $R^2 = 0.96$ ). Predictions for total N were very good with both VNIR and MIR, with MIR having better predictions. The training set for VNIR was not fit as well as MIR and resulted in more outliers. However, the test set for VNIR was fit better and had lower errors. Calibration models are shown in figure 4.18.

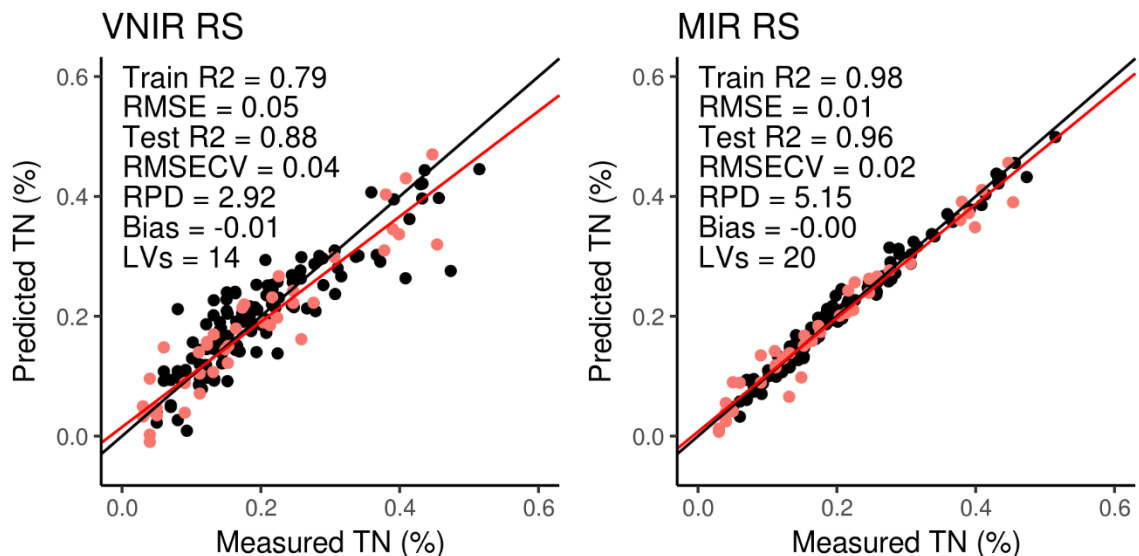


Figure 4.18. Goodness of fit plots showing relationship of PLS predicted Total N values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross-validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

## 4.2 Calibration Sampling

The three model calibration schemes had varying degrees of success in creating useful prediction models, where the performance of the calibration model was often dependent on the soil property of interest. For many of the spectrally active soil properties, the RS calibration scheme provided the best predictions due to many training samples. When the soil property had highly skewed distributions or was only present in a few locations, the SS calibration scheme worked better. KSS sampling worked best on spectrally active soil properties that tended to be location dependent.

### 4.2.1 Random Sampling

The random sample calibration performance, created by a randomly assigning 75% of the samples to the training set and 25% of samples to the test set, can be seen in table 4.5-6 and scatter plots of calibration models can be found in Appendix D. The RS

calibration method had the best prediction performance for both VNIR and MIR for predicting CEC, H<sub>2</sub>O P, SOC and total N, compared to the other calibration sampling methods. The VNIR models had a wide range of fit ( $R^2$ ) values for the test set, ranging from -84.09 to 0.88 and likewise RPD values (0.11-3.46). MIR models were improved but still showed equally varied performance ( $R^2$  -31.84 to 0.98 and RPD 0.18 to 8.14, respectively).

Both VNIR and MIR models had very good to excellent performance for predicting CEC, total N and total C ( $R^2 > 0.82$  and RPD  $> 2.39$ ). VNIR predictions for CEC were very good ( $R^2 = 0.82$  and RPD = 2.39). MIR performed better and had excellent model performance ( $R^2 = 0.99$  and RPD = 8.54) for CEC but had a slightly higher bias than VNIR (0.34 vs -0.28, for MIR and VNIR respectively). Predictions for total N with VNIR were very good ( $R^2 = 0.88$  and RPD = 2.92) but MIR predictions were better ( $R^2 = 0.96$  and RPD = 5.15). Total C predictions are considered very good for VNIR ( $R^2 = 0.84$  and RPD = 2.49) and excellent for MIR ( $R^2 = 0.98$  and RPD = 6.70). Organic carbon was also well predicted by both VNIR and MIR ( $R^2 = 0.85$ , RPD = 2.62 and  $R^2 = 0.98$ , RPD = 7.72, for VNIR and MIR respectively). Despite total C and SOC showing good predictions, carbonate failed to be modeled with VNIR ( $R^2 = -1.12$ ) and had poor fit with MIR ( $R^2 = 0.66$ ).  $\beta$ G was poorly predicted with VNIR ( $R^2 = 0.66$  and RPD = 1.74) and was improved with MIR and considered good ( $R^2 = 0.74$  and RPD = 2.00). The extractable cations had the most varied performance for both VNIR and MIR. Na<sup>+</sup> had the poorest predictive ability and could not be modeled by VNIR ( $R^2 = -84.09$ ) or MIR ( $R^2 = -32.49$ ). Ca<sup>2+</sup> predictions were considered excellent ( $R^2 = 0.94$  and RPD = 4.20) with MIR while VNIR had poor predictive ability ( $R^2 = 0.54$  and RPD =



1.49). VNIR had poor predictive ability for  $Mg^{2+}$  ( $R^2 = 0.60$  and  $RPD = 1.60$ ) and  $K^+$  ( $R^2 = 0.60$  and  $RPD = 1.60$ ) while MIR had good predictive ability ( $R^2 = 0.77$ ,  $RPD = 1.86$ ;  $R^2 = 0.83$ ,  $RPD = 2.48$ , for  $Mg^{2+}$  and  $K^+$ , respectively). Electrical conductivity was also poorly predicted for both VNIR and MIR ( $R^2 = 0.57$  for VNIR and  $R^2 = 0.63$  for MIR). Of the two pH methods, 1:2  $CaCl_2$  pH was better predicted than 1:1  $H_2O$  pH. VNIR pH models were poor overall, with  $CaCl_2$  pH slightly higher ( $R^2 = 0.68$ ,  $RPD = 1.78$ ;  $R^2$  of 0.56,  $RPD = 1.53$ , for  $CaCl_2$  pH and  $H_2O$  pH respectively). MIR predicted  $CaCl_2$  pH with very good performance ( $R^2 = 0.86$  and  $RPD = 2.71$ ) but had similarly poor predictions for  $H_2O$  pH ( $R^2 = 0.66$  and  $RPD = 1.74$ ) as VNIR. Both VNIR and MIR poorly predicted  $H_2O$  P ( $R^2 < 0.54$  and  $RPD < 1.49$ ). MIR had the best prediction for total P but was still considered poor ( $R^2 = 0.68$  and  $RPD = 1.80$ ).

Table 4.5 Goodness of fit statistics for VNIR with random sampling.								
Property	Train $R^2$	RMSE	Test $R^2$	RMSECV	RPD	Bias	LVs	Preprocessing
EC	0.56	0.39	0.57	0.20	1.54	-0.02	6	SG + MSC
1:1 $H_2O$ pH	0.61	0.56	0.56	0.58	1.53	-0.23	3	1-D
1:2 $CaCl_2$ pH	0.48	0.71	0.68	0.49	1.78	-0.03	1	Detrend
CEC	0.94	2.23	0.82	4.11	2.39	-0.28	18	Raw
Ca <sup>2+</sup>	0.82	5.83	0.54	6.89	1.49	0.61	7	SG + MSC
Mg <sup>2+</sup>	0.84	0.94	0.60	0.93	1.60	0.06	9	SG + SNV
Na <sup>+</sup>	0.07	1.32	-84.09	0.38	0.11	0.15	1	Raw
K <sup>+</sup>	0.70	0.47	0.60	0.39	1.60	0.09	9	MSC
$H_2O$ P	0.70	0.18	0.54	0.23	1.49	0.01	13	1-D
Total P	0.41	154.90	0.58	157.20	1.57	-5.47	6	SNV
Total C	0.87	0.43	0.84	0.55	2.49	-0.16	14	Raw
SOC	0.84	0.44	0.85	0.53	2.62	-0.16	14	Raw
$CaCO_3$	0.70	1.14	-1.12	0.79	0.70	-0.05	8	MSC
$\beta G$	0.61	54.01	0.66	45.78	1.74	-8.14	8	SG + SNV
Total N	0.79	0.05	0.88	0.04	2.92	-0.01	14	Smooth

Property	Train R <sup>2</sup>	RMSE	Test R <sup>2</sup>	RMSECV	RPD	Bias	LVs	Preprocessing
EC	0.98	0.09	0.63	0.19	1.66	-0.02	18	2-D
1:1 H <sub>2</sub> O pH	0.97	0.16	0.66	0.51	1.74	-0.23	19	2-D
1:2 CaCl <sub>2</sub> pH	0.94	0.25	0.86	0.32	2.71	-0.15	11	2-D
CEC	0.99	0.93	0.99	1.15	8.54	0.34	14	Raw
Ca <sup>2+</sup>	0.98	1.88	0.94	2.45	4.20	0.84	10	1-D
Mg <sup>2+</sup>	0.97	0.38	0.77	0.80	1.86	0.29	16	1-D
Na <sup>+</sup>	0.70	0.76	-32.49	0.24	0.18	-0.03	6	1-D
K <sup>+</sup>	0.67	0.49	0.83	0.25	2.48	-0.01	5	Detrend
H <sub>2</sub> O P	0.43	0.25	0.50	0.24	1.43	-0.04	7	2-D
Total P	0.80	89.31	0.68	137.04	1.80	-18.38	20	SG + SNV
Total C	0.99	0.11	0.98	0.21	6.70	-0.04	20	Detrend
SOC	0.99	0.10	0.98	0.18	7.72	-0.04	20	Detrend
CaCO <sub>3</sub>	0.85	0.80	0.66	0.31	1.78	-0.01	5	2-D
βG	0.94	22.05	0.74	39.71	2.00	-1.75	19	2-D
Total N	0.98	0.01	0.96	0.02	5.15	-0.00	20	Raw

#### 4.2.2 Stratified Sampling

The stratified sampling scheme, using the satellite soil profiles to train the models and the central soil profiles as the test set had very good performance with MIR and moderate performance with VNIR, as seen in tables 4.7-8. Additional scatter plots of the calibration models can be found in Appendix D. SS performed better than the other sampling methods for Mg<sup>2+</sup>, Na<sup>+</sup>, and βG with both VNIR and MIR. MIR models for predicted CEC, Ca<sup>2+</sup>, total C, and SOC are considered excellent (R<sup>2</sup> > 0.96 and RPD > 5.01), while the best prediction for VNIR was very good (R<sup>2</sup> = 0.81 and RPD = 2.29 for Ca<sup>2+</sup>). VNIR outperformed MIR in predicting EC and Na<sup>+</sup>, however the models were still poor (R<sup>2</sup> = 0.60, RPD = 1.60; R<sup>2</sup> = 0.63, RPD = 1.65, for EC and Na<sup>+</sup> respectively). Models for pH were poor for VNIR (R<sup>2</sup> from 0.59-0.62 and RPD from 1.58-1.64 for H<sub>2</sub>O

pH and CaCl<sub>2</sub> pH, respectively) and good to very good for MIR ( $R^2$  from 0.77-0.86 and RPD from 2.11-2.69 for H<sub>2</sub>O pH and CaCl<sub>2</sub> pH, respectively). VNIR predicted CEC and Ca<sup>2+</sup> about the same as MIR and are considered good to very good ( $R^2$  from 0.78-0.81 and RPD from 2.14-2.29 for CEC and Ca<sup>2+</sup> respectively). MIR had excellent models for CEC ( $R^2 = 0.97$  and RPD = 6.10) and Ca<sup>2+</sup> ( $R^2 = 0.96$  and RPD = 5.01). Predictions for Mg<sup>2+</sup> and K<sup>+</sup> with VNIR were fair to good ( $R^2 = 0.75$ , RPD = 2.06;  $R^2 = 0.69$ , RPD = 1.81 for Mg<sup>2+</sup> and K<sup>+</sup>, respectively) while MIR predictions are very good ( $R^2 = 0.89$ , RPD = 3.07;  $R^2 = 0.82$ , RPD = 2.36 for Mg<sup>2+</sup> and K<sup>+</sup>, respectively). Predictions for P were varied, with  $R^2$  from 0.03-0.72. H<sub>2</sub>O P was not predicted by either VNIR ( $R^2 = 0.03$ ) or MIR ( $R^2 = 0.12$ ). Total P had fair model performance with MIR ( $R^2 = 0.72$  and RPD = 1.90) and poor performance with VNIR ( $R^2 = 0.57$  and RPD = 1.55). Total C and SOC models performed poor to fair with VNIR ( $R^2 = 0.70$ , RPD = 1.84 for total C and  $R^2 = 0.66$ , RPD = 1.72 for SOC). MIR had excellent predictions for both total C ( $R^2 = 0.96$  and RPD = 5.05) and SOC ( $R^2 = 0.97$  and RPD = 5.76). Carbonate content was modeled poorly with MIR ( $R^2 = 0.57$ ) and very poorly with VNIR ( $R^2 = 0.17$ ).  $\beta$ G had very good predictions with MIR ( $R^2 = 0.81$  and RPD = 2.33) and had poor predictions with VNIR ( $R^2 = 0.67$  and RPD = 1.75). Total N was poorly predicted by VNIR ( $R^2 = 0.65$  and RPD = 1.71) and had very good predictions with MIR ( $R^2 = 0.89$  and RPD = 2.98).

Table 4.7 Goodness of fit statistics for VNIR with stratified sampling.

Property	Train R <sup>2</sup>	RMSE	Test R <sup>2</sup>	RMSECV	RPD	Bias	LVs	Preprocessing
EC	0.83	0.24	0.60	0.24	1.60	-0.03	9	Detrend
1:1 H <sub>2</sub> O pH	0.68	0.50	0.59	0.58	1.58	-0.07	4	1-D
1:2 CaCl <sub>2</sub> pH	0.74	0.50	0.62	0.57	1.64	-0.12	4	1-D
CEC	0.85	3.49	0.78	4.58	2.14	-0.21	12	Smooth
Ca <sup>2+</sup>	0.86	4.58	0.81	6.28	2.29	-0.47	12	Smooth
Mg <sup>2+</sup>	0.93	0.54	0.75	1.15	2.06	-0.03	16	SG + SNV
Na <sup>+</sup>	0.83	0.50	0.63	0.67	1.65	-0.02	10	SNV
K <sup>+</sup>	0.74	0.42	0.69	0.45	1.81	0.02	10	SG + SNV
H <sub>2</sub> O P	0.25	0.32	0.03	0.24	1.03	0.14	5	SNV
Total P	0.63	127.70	0.57	143.90	1.55	26.28	11	SG + MSC
Total C	0.88	0.44	0.70	0.69	1.84	0.01	14	SG + SNV
SOC	0.93	0.30	0.66	0.71	1.72	-0.00	19	Smooth
CaCO <sub>3</sub>	0.57	1.32	0.17	1.27	1.11	-0.21	5	Detrend
βG	0.72	42.85	0.67	52.51	1.75	6.45	11	Raw
Total N	0.86	0.04	0.65	0.07	1.71	-0.05	14	SG + SNV

Table 4.8 Goodness of fit statistics for MIR with stratified sampling.

Property	Train R <sup>2</sup>	RMSE	Test R <sup>2</sup>	RMSECV	RPD	Bias	LVs	Preprocessing
EC	0.40	0.46	0.50	0.27	1.42	0.01	2	Detrend
1:1 H <sub>2</sub> O pH	0.55	0.59	0.77	0.43	2.11	0.01	5	MSC
1:2 CaCl <sub>2</sub> pH	0.87	0.36	0.86	0.35	2.69	-0.13	7	MSC
CEC	0.99	0.91	0.97	0.97	6.10	0.49	11	2-D
Ca <sup>2+</sup>	0.99	0.89	0.96	2.87	5.01	-0.11	20	1-D
Mg <sup>2+</sup>	0.95	0.45	0.89	0.77	3.07	-0.01	13	Detrend
Na <sup>+</sup>	0.62	0.76	0.50	0.78	1.42	0.21	5	2-D
K <sup>+</sup>	0.70	0.45	0.82	0.34	2.36	0.00	6	Raw
H <sub>2</sub> O P	0.25	0.32	0.13	0.23	1.08	0.12	3	SNV
Total P	0.52	145.70	0.72	116.90	1.90	-16.06	4	2-D
Total C	0.99	0.11	0.96	0.25	5.05	-0.10	15	MSC
SOC	0.99	0.14	0.97	0.21	5.76	-0.05	10	1-D
CaCO <sub>3</sub>	0.61	1.25	0.57	0.91	1.55	-0.25	2	Detrend
βG	0.86	29.91	0.81	39.44	2.33	-5.89	20	Detrend
Total N	0.98	0.02	0.89	0.04	2.98	-0.01	10	2-D

### 4.2.3 Kennard Stone Sampling

Predictions using the KSS algorithm are shown in tables 4.9-10 and Appendix D. KSS calibrations had the best performance over the other sampling techniques for pH. MIR predictions overall, were better than those of VNIR. MIR had excellent predictions for 1:2 CaCl<sub>2</sub> pH, CEC, Ca<sup>2+</sup>, total C, SOC, and total N ( $R^2 > 0.91$  and RPD  $> 3.35$ ). VNIR's predictive ability was relatively poor, where the best model had fair performance ( $R^2 = 0.74$  and RPD = 1.98 for Ca<sup>2+</sup>). Total C had fair prediction performance ( $R^2 = 0.72$  and RPD = 1.91) with VNIR, while MIR had excellent performance ( $R^2 = 0.98$  and RPD = 6.44). VNIR had poor performance predicting SOC ( $R^2 = 0.68$  and RPD = 1.79) while MIR had excellent predictions ( $R^2 = 0.98$  and RPD = 6.95). Carbonate was very poorly predicted by VNIR ( $R^2 = 0.48$  and RPD = 1.40), while MIR had very good performance ( $R^2 = 0.88$  and RPD = 2.86).  $\beta$ G was modeled poorly by both VNIR ( $R^2 = 0.57$  and RPD = 1.54) and MIR ( $R^2 = 0.68$  and RPD = 1.78). Total N model for VNIR had poor fit ( $R^2 = 0.69$ ) but good prediction performance (RPD = 1.80) while MIR had excellent fit and prediction performance ( $R^2 = 0.95$  and RPD = 4.47). VNIR had fair model performance with CEC, Ca<sup>2+</sup>, 1:1 H<sub>2</sub>O pH, and 1:2 CaCl<sub>2</sub> pH ( $R^2 > 0.70$  and RPD  $> 2.83$ ). MIR had better predictive ability than VNIR for these properties, with models considered very good to excellent ( $R^2 > 0.81$  and RPD  $> 2.29$ ). VNIR and MIR had fair and good predictions for Mg<sup>2+</sup> ( $R^2 = 0.74$ , RPD = 1.97;  $R^2 = 0.89$ , RPD = 3.07 for VNIR and MIR respectively). K<sup>+</sup> was poorly predicted with VNIR ( $R^2 = 0.63$  and RPD = 1.65) and had good predictions with MIR ( $R^2 = 0.77$  and RPD = 2.09). VNIR could not predict Na<sup>+</sup> ( $R^2 = -1.03$ ) and MIR likewise had very poor predictions for Na<sup>+</sup> ( $R^2 = 0.47$  and RPD =

1.39). EC was also very poorly predicted by both VNIR ( $R^2 = 0.24$ ) and MIR ( $R^2 = 0.16$ ).

Total P and H<sub>2</sub>O P were not modeled well by MIR ( $R^2 < 0.63$ ) or VNIR ( $R^2 < 0.49$ ).

Table 4.9 Goodness of fit statistics for VNIR with Kennard Stone sampling.

Property	Train $R^2$	RMSE	Test $R^2$	RMSECV	RPD	Bias	LVs	Preprocessing
EC	0.91	0.21	0.24	0.24	1.15	-0.06	11	MSC
1:1 H <sub>2</sub> O pH	0.69	0.51	0.70	0.46	1.83	0.01	9	SNV
1:2 CaCl <sub>2</sub> pH	0.79	0.45	0.73	0.47	1.93	0.00	9	SNV
CEC	0.89	3.35	0.71	4.44	1.88	0.66	12	Raw
Ca <sup>2+</sup>	0.86	5.52	0.74	5.30	1.98	-0.36	10	Raw
Mg <sup>2+</sup>	0.83	0.97	0.74	1.01	1.97	0.08	7	1-D
Na <sup>+</sup>	0.91	0.48	-1.03	0.56	0.71	0.05	10	2 <sup>nd</sup> Derivative
K <sup>+</sup>	0.85	0.30	0.63	0.51	1.65	-0.18	15	SG + SNV
H <sub>2</sub> O P	0.79	0.11	0.40	0.31	1.30	-0.07	19	SG + MSC
Total P	0.79	106.50	0.49	138.60	1.41	25.02	13	Raw
Total C	0.83	0.59	0.72	0.50	1.91	-0.04	10	Raw
SOC	0.79	0.61	0.68	0.54	1.79	-0.02	9	Raw
CaCO <sub>3</sub>	0.77	0.77	0.48	1.46	1.40	-0.15	8	1-D
βG	0.75	45.27	0.57	50.83	1.54	-6.25	9	MSC
Total N	0.80	0.05	0.69	0.05	1.80	-0.01	10	Raw

Table 4.10 Goodness of fit statistics for MIR with Kennard Stone sampling.

Property	Train $R^2$	RMSE	Test $R^2$	RMSECV	RPD	Bias	LVs	Preprocessing
EC	0.79	0.31	0.16	0.21	1.10	0.07	8	1-D
1:1 H <sub>2</sub> O pH	0.85	0.39	0.81	0.33	2.29	-0.02	11	SG + MSC
1:2 CaCl <sub>2</sub> pH	0.98	0.15	0.91	0.25	3.35	0.02	16	Detrend
CEC	1.00	0.57	0.99	1.01	8.82	0.15	20	1-D
Ca <sup>2+</sup>	0.99	1.43	0.96	1.99	5.10	-0.09	17	Detrend
Mg <sup>2+</sup>	0.98	0.32	0.89	0.59	3.07	-0.09	20	2-D
Na <sup>+</sup>	0.71	0.54	0.47	0.97	1.39	-0.02	6	2-D
K <sup>+</sup>	0.78	0.44	0.77	0.33	2.09	0.03	8	Smooth
H <sub>2</sub> O P	0.24	0.34	0.42	0.20	1.31	0.02	5	SG + SNV
Total P	0.87	79.18	0.63	123.95	1.65	8.55	20	SG + MSC
Total C	0.99	0.12	0.98	0.16	6.44	0.02	16	MSC
SOC	0.99	0.11	0.98	0.15	6.95	0.03	16	MSC
CaCO <sub>3</sub>	1.00	0.12	0.88	0.44	2.86	-0.14	20	2-D
βG	0.84	35.72	0.68	44.28	1.78	11.68	12	MSC
Total N	0.99	0.01	0.95	0.02	4.47	0.00	17	MSC

## CHAPTER 5. DISCUSSION & CONCLUSION

### 5.1 General Soil Properties

Due to the large study area, many of the soils were dissimilar, portraying wide ranges of analyte values. Generally, it is considered good to have a wide range of analyte values because it can increase a model's predictive performance and make the model more robust but in cases of greater skewness in the analyte's distribution, the model's performance can be weakened (Minasny et al., 2009). A few properties had elevated skewness, most notably  $\text{CaCO}_3$ , extractable  $\text{Na}^+$ , and EC. The high skew in the  $\text{CaCO}_3$  was expected considering samples were collected from surface horizons where carbonates are generally not present in these soils. Additionally, the soils within the sampling area are under different moisture regimes where conditions are wetter in the east and much drier in the west. While  $\text{CaCO}_3$  has been effectively modeled from other studies (Reeves and Smith, 2009) due to its specific absorption bands in both the VNIR and MIR, the skewed distribution likely made carbonates difficult to model in this study. Extractable  $\text{Na}^+$  and EC also exhibited skewness, and the two properties unsurprisingly were correlated (0.83). This skewness may have led to the inability to predict these properties.

#### 5.1.1 Electrical Conductivity

Soil electrical conductivity models had slight predictive ability. MIR was found to have slightly more predictive ability than VNIR for predicting EC, similar to findings by Viscarra Rossel et al. (2006). The distribution of EC values in the sample set was very skewed and supports Minasny et al., (2009) that suggested poor EC predictions are due to a poorly distributed samples and that good predictions for EC are dependent on a large

range of values. Additionally, there was no correlation with CEC ( $r = 0.17$ ) or any of the spectrally active soil components, making predictions of EC indirectly from CEC unlikely. The highly skewed and poor distribution of EC values in the samples were likely the reason predictions were not better.

### 5.1.2 Soil pH

MIR models for pH were altogether better than models from VNIR and 1:2 CaCl<sub>2</sub> pH was much better than 1:1 H<sub>2</sub>O pH. In the training sets of both VNIR and MIR models for 1:1 H<sub>2</sub>O pH, two samples failed to fit. One surface sample, in particular, was underpredicted by both, having a measured pH of 7.81 and predicted pH of 6.33 and 5.84 by VNIR and MIR respectively. This sample had measurable carbonates but also had a high SOC of 4.51%, which likely accounts for the underpredicted value as carbonates were generally modeled poorly. Predictive ability of pH depends on correlation with soil organic carbon (Sarathjith et al., 2014), soil organic acids and carbonates (Reeves, 2010), however water pH in these samples was poorly correlated ( $r < 0.30$ ), leading to poor predictions.

1:2 CaCl<sub>2</sub> pH in these samples had slight correlation with carbonates ( $r = 0.47$ ) and poor correlation with SOC ( $r = 0.20$ ). The lack of correlation between soil organics and pH in these samples is likely the reason that VNIR could not predict pH as well as MIR. Despite having a weak correlation with the soil organic fraction and carbonates, MIR had good predictions for pH. Minasny et al., (2009) hypothesized that the ability of MIR to predict pH is due to its relationship with exchangeable cations. The observed correlation between pH and extractable Ca<sup>2+</sup> was strong ( $r > 0.74$ ), supporting this hypothesis. Given the high correlation with pH and Ca<sup>2+</sup>, it is possible that MIR is



indirectly correlating pH with the clay mineralogy. However, pH and CEC had lower correlation ( $r = 0.30$ ) than pH and  $\text{Ca}^{2+}$ , suggesting that MIR is correlating pH with carbonates and CEC combined. There is not enough information on the samples' clay mineralogy to conclude this, these findings and those by Minasny et al., (2009) suggest that MIR may be capable of predicting pH without a strong correlation with soil organics. While both VNIR and MIR models for 1:2  $\text{CaCl}_2$  pH were improved over 1:1  $\text{H}_2\text{O}$  pH, only the MIR model is considered useful for predictions.

### 5.1.3 Cation Exchange Capacity and Extractable Cations

MIR was much better at predicting CEC than VNIR for every calibration scheme. These findings are supported by those of Viscarra Rossel et al., (2006) and indicate that MIR has a higher sensitivity to clay mineralogy and organic matter. VNIR models for CEC were still good and are comparable to findings by Chang et al., (2001) using VNIR. The VNIR model predicted a few samples as having negative values. However, these samples were extremely sandy (> 90 %) and may have increased noise due to light scatter.

Of the extractable cations,  $\text{Ca}^{2+}$  was most easily predicted followed by  $\text{Mg}^{2+}$  and  $\text{K}^+$ .  $\text{Na}^+$  was not predicted well in most cases.  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ , and  $\text{K}^+$  were all better predicted with MIR. The predictability of the cations increased when cations were correlated with CEC ( $r = 0.62, 0.50, 0.47,$  and  $0.5$  for  $\text{Ca}^{2+}, \text{Mg}^{2+}, \text{K}^+,$  and  $\text{Na}^+$ , respectively), indicating that the cations are predicted by their relationship with clay mineralogy. As the range of values increased, the predictive ability for both VNIR and MIR increased, supporting the idea that model performance and robustness are improved by having a greater range in values (Minasny et al., 2009).  $\text{Na}^+$  had a very skewed

distribution and was not correlated with any spectrally active soil properties or secondarily active properties making it difficult to achieve good predictions.

#### 5.1.4 Phosphorus

H<sub>2</sub>O P was not predicted well by either VNIR or MIR. This may be due to not having any correlation with spectrally active soil properties and having a small range of values (2.51 mg kg<sup>-1</sup>) and a skewed distribution. H<sub>2</sub>O P predictions were slightly better than those observed by Cohen et al., (2005), who had very poor results ( $R^2 = 0.37$  and RPD = 1.20) with wetland samples using VNIR. Cohen et al., (2005) also observed a similarly small range of values (0.94 mg kg<sup>-1</sup>). Janik et al., (2009) showed good results for H<sub>2</sub>O P with MIR ( $R^2 = 0.84$ ). Janik et al., (2009) had a 16 h extraction time for H<sub>2</sub>O P, allowing for more soluble P to come into equilibrium and therefore had a much greater range (898 mg kg<sup>-1</sup>) of soluble P values. The results cannot be compared directly but indicate longer extraction times for soluble P should be explored further to see if range plays a major role in its prediction of P. Neither VNIR nor MIR models were robust enough to predict H<sub>2</sub>O P.

Total P had better predictions when modeled with MIR than with VNIR, which supports findings by Reeves and Smith (2009). Total P predictions observed with VNIR were not as good as those previously seen in the literature with  $R^2 > 0.89$  (Bogrekci & Lee, 2005; Todorova et al., 2011). This may be due to differences in sample size and sample range. Todorova et al., (2011) had a total sample range 5 times greater than the sample range in this study.

Total P models performed better than H<sub>2</sub>O P models. This increase in performance is likely due to the forms of P measured. Water soluble P relates to the P in

the soil solution whereas total P also includes forms of P within clay structures. Total P showed some correlation with CEC and clay ( $r = 0.62$  for both CEC and clay) whereas  $H_2O$  P did not ( $r = 0.20$  and  $r = 0.07$  for CEC and clay, respectively). Abdi et al., (2012) also observed better performance with total P models than  $H_2O$  P and suggested that this was due to the fraction of organic P in the soil correlated with organic matter. Observed P correlations with total N may support Abdi et al., (2012), where correlations with total P were better than those of  $H_2O$  P ( $r = 0.54$  vs  $0.35$ ).

#### 5.1.5 Carbon

Total C, SOC, and  $CaCO_3$ , were better predicted by MIR than VNIR. The fundamental absorption bands in the MIR are better predictors than the overtones of these bands in the VNIR region (Bellon-Maurel and McBratney, 2011). Knox et al. (2015) reported similar findings, showing that MIR outperformed VNIR for both total C and SOC. VNIR predictions for total C in this study were comparable to VNIR predictions observed by Chang et al. (2001) over the coterminous U.S. ( $R^2 = 0.87$  and  $RPD = 2.79$ ). Predictions for total C performed better and had lower errors than those for SOC with VNIR. The opposite was observed with MIR, where SOC models outperformed Total C models but both models for total C and organic C were excellent with MIR and very good with VNIR.  $CaCO_3$  prediction was not as good as expected, as all VNIR models performed very poorly ( $R^2 < 0.48$ ). MIR models exhibited a wide range of fit ( $R^2$  from 0.58-0.87) and had poor to very good predictive ability. The overall poor predictive performance can be attributed to having only surface soil samples where carbonates are unlikely to form, resulting in a heavily skewed distribution of carbonates in the samples.

### 5.1.6 $\beta$ -Glucosidase

$\beta$ G had better predictions with MIR than with VNIR. The training set for MIR had a better fit and fewer outliers than VNIR. One sample was consistently predicted at having lower  $\beta$ G activity with both VNIR and MIR, however this may be due to measurement error.  $\beta$ G had strong correlations with SOC ( $r = 0.72$ ), Total C ( $r = 0.70$ ), and Total N ( $r = 0.74$ ), all of which were better predicted with MIR.  $\beta$ G's relationship with soil organic matter allows it to be predicted better with MIR. The VNIR models performed as well as those in other studies. Cohen et al., (2005) observed good results ( $R^2 = 0.79$  and  $RPD = 2.64$ ) with VNIR. Zoronoza et al., (2008) had very good results ( $R^2 = 0.96$  and  $RPD = 2.64$ ) when just using the NIR region between 1372-2272 nm. Likewise, Dick et al., (2013) had improvements in model performance when using the NIR region (1100-2498 nm) over the VNIR (400-2498 nm) and attributed this to more background noise present in the visible region.

### 5.1.7 Total Nitrogen

MIR had excellent predictions for Total N in all sampling schemes while VNIR had only one model that could be considered very good. The best MIR model for total N performed as well as MIR models by Reeves et al., (2001) ( $R^2 = 0.95$ ) but better than MIR models observed by Minasny et al., (2009) ( $R^2 = 0.76$  and  $RPD = 2.0$ ). VNIR models performed slightly better than NIR (1000-2500 nm) models ( $R^2 = 0.73$  and  $RPD = 2.20$ ) from Abdi et al., (2012) but were in agreement with Todorova et al., (2011) who observed similar results using the NIR spectra between 1001-2500 nm ( $R^2 = 0.91$  and  $RPD = 2.3$ ). Nitrogen is a big component of soil organics and had a strong correlation

with SOC ( $r=0.97$ ). Total N was also found to have a good correlation with CEC ( $r = 0.72$ ) possibly making it easier to predict with MIR.

## 5.2 Calibration Sampling

The calibration sampling used for creating prediction models can play a large role in model performance. A significant number of samples are needed to train the model and still have enough independent samples left over to test the model performance. Although the number of samples can be considered large ( $n=156$ ), they are not all independent from each other considering 2 samples were taken from the same soil profile. This effectively cut the number of independent samples in half, making an appropriate calibration sampling scheme vital to avoid pseudoreplication and over-fitting the models.

### 5.2.1 Random Sampling

The random sampling scheme generally provided the best predictions except in cases when the soil properties had a skewed distribution. The increase in performance is likely due to a larger number of samples used to train the model compared to the rest of the calibration schemes, and thus cannot be directly compared with the other calibration schemes. When the soil component had a skewed distribution, samples may not have been representative between training and test sets, resulting in poorer models. In the case of  $\text{Na}^+$ , the test set contained samples with a  $\text{Na}^+$  concentration of zero, except for one sample. This sampling scheme was able to predict the one sample with  $\text{Na}^+$  with an error of  $0.08 \text{ cmol}(+) \text{ kg}^{-1}$  but failed to predict a majority of samples as having zero concentrations of  $\text{Na}^+$ . Random sampling for calibration models is often used because of its simplicity and many authors report good calibration results for spectrally active soil properties (Estienne et al., 2001; Van Groenigen et al., 2003; Dick et al., 2013).

### 5.2.2 Stratified Sampling

The stratified sampling scheme did not perform as well as RS for many properties, despite the advantage of having an even distribution of sites in the training and test sets. The SS scheme predicted  $Mg^{2+}$ ,  $Na^+$ , and  $\beta G$  concentrations the best because there was less variation of these properties within site location. Two of the 27 sites had measurable amounts of  $Na^+$ , which may explain why  $Na^+$  predictions were most improved with this scheme. Wijewardane et al., (2016) found that SOC models stratified by soil textural class and land use class outperformed global models. While the SS scheme in this study is not the same as Wijewardane et al., (2016), it attempts to create an even distribution of samples with the same soil textural classes and land use class in both the training and test sets. This increased the model performance for soil properties that were spatially dependent by having comparable ranges of values in the training and test sets, which is important for good calibrations according to Davies and Fearn, (2006).

### 5.2.3 Kennard-Stone Sampling

Of all properties evaluated, only  $CaCO_3$  was best modeled with the Kennard-Stone sampling scheme. This makes sense due to  $CaCO_3$  being spectrally active, where KSS selects training and testing samples based on dissimilarity between spectra, allowing for a uniform distribution in training and test sets. KSS did not significantly outperform the other calibration sampling techniques for the rest of the soil properties, despite other authors finding it an effective method for calibration sampling of spectral data (Liu et al., 2014; Zhang et al., 2017). Liu et al. (2014) found KSS sampling to be the best calibration sampling scheme for their study area, which consisted of mixed land use classes (cropland, forest, and meadows) and had a smaller study area and finer sampling

resolution than this study. Wetterlind et al., (2013) found that KSS worked well in capturing the variation in datasets when the sampling size is large. The average performance of KSS in this study may be attributed to the large spatial distribution of the study area, coarse sampling resolution and a relatively small number of independent samples. This suggests that KSS may not be an optimal calibration sampling scheme for small datasets with coarse sampling resolutions.

### 5.3 Conclusions and Further Work

Infrared spectroscopic methods paired with PLSR can be a useful tool for rapidly predicting certain soil properties. MIR and VNIR proved useful in predicting spectrally active soil properties such as total C, SOC and total N, and secondary properties like CEC and pH. MIR models were more robust than VNIR models for these properties and could potentially be used as a surrogate for traditional laboratory analyses or ground truthing/validation for airborne and satellite remote sensing platforms. VNIR models for these properties were good but had more error and may not be a suitable replacement for laboratory analyses when accuracy is important. However, they could prove useful for qualitative analysis when traditional techniques are timely, such as CEC. Although MIR had better predictions than VNIR in most instances, VNIR outperformed MIR under certain calibration sampling schemes, although predictions for both these properties were considered poor.

Soil properties with little correlation to spectrally active properties, such as P, Na<sup>+</sup>, and EC, or those that had highly skewed distributions in the sample set (CaCO<sub>3</sub>) tend to be modeled poorly by both VNIR and MIR. A greater sample size/density could improve these models, especially for underrepresented soil properties that had low range or heavily skewed distributions. Likewise, further research should be done to validate

both MIR and VNIR models with independent sites to test the robustness of the models in predicting these soil properties at the regional scale.

Standardization or normalization of non-normally distributed properties should be considered if sample size cannot be easily increased, as many of these techniques assume normally distributed data. Additional calibration sampling techniques such as leave-one-out cross-validation or selecting calibration samples for the training set based on the highest and lowest values for each property should be tested. Due to the differences in training set and test set sizes between the tested calibration sampling methods, direct comparisons cannot be made as to which method is best. While finding the optimal calibration sampling schemes is important and plays a vital role in model validation, that is beyond the scope of this thesis, where the objective was to compare the predictive ability of VNIR and MIR spectroscopic methods. Likewise, this thesis only looked at PLS regression, a classical multivariate statistical technique, for modeling complex soil properties, when more advanced machine learning algorithms such as artificial neural networks and regression trees have been used with much success. This study indicates that that the MIR spectrum contains more attainable and useable information than VNIR with the use of classical statistical techniques for predicting soil properties over a regional area with coarse sampling resolution.

MIR spectrometers are costlier than VNIR spectrometers, involve more sample preparation (finer grinding of sample), but predictions tend to be much more accurate. The MIR region of a soil spectra contains distinctive adsorption peaks related to fundamental bond vibrations that can be used to detect clays, minerals and organic compounds, which can be visually observed in the spectra whereas the VNIR region is



complex and requires chemometric analysis to access useful information. Due to the need for quantitative information, this minimizes the differences between the spectral regions, where chemometric methods are required to create predictive models. The VNIR spectrum combined with multivariate analysis techniques can predict useful predictions of these soil properties with minimal processing, whereas MIR is capable of more accurate predictions, but at the cost of more sample processing and time. Ultimately, the choice between these two techniques becomes a tradeoff between instrumentation cost and accuracy.

## REFERENCES

- Abdi, D., G.F. Tremblay, N. Ziadi, G. Bélanger, and L.É. Parent. 2012. Predicting soil phosphorus-related properties using near-infrared reflectance spectroscopy. doi:10.2136/sssaj2012.0155.
- Bandick, A.K., and R.P. Dick. 1999. Field management effects on soil enzyme activities. *Soil Biol. Biochem.* 31(11): 1471–1479. doi: 10.1016/S0038-0717(99)00051-6.
- Bellon-Maurel, V., E. Fernandez-ahumada, B. Palagos, J. 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. *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.
- BenDor, E., Y. Inbar, Y. Chen, and E. Ben-Dor. 1997. The reflectance spectra of organic matter in the visible near infrared and short wave infrared region(400-2500nm) during a controlled decomposition process. *Remote sens. Environ* 61(May 1996): 1–15.
- Bertrand, D., E.M. Qannari, and E. Vigneau. 2001. Latent root regression analysis: An alternative method to PLS. *Chemom. Intell. Lab. Syst.* 58(2): 227–234. doi: 10.1016/S0169-7439(01)00161-7.
- Bogrekci, I., and W.S. Lee. 2005. Spectral soil signatures and sensing phosphorus. *Biosyst. Eng.* 92(4): 527–533. doi: 10.1016/j.biosystemseng.2005.09.001.
- Bray, R.H., and L.T. Kurtz. 1945. Determination of total, organic, and available forms of phosphorus in soils. *Soil Sci.* 59: 39-45.
- Burns, DA., and E.W. Ciurczak. *Handbook of near-infrared analysis*. Second edition. Marcel Dekker; New York: 2001. revised and expanded
- Burt, R. 2014. Soil survey laboratory methods manual. Soil Survey Investigations Report No. 42, Version 5. U.S. Department of Agriculture, Natural Resources Conservation Service. [https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/ref/?cid=nrcs142p2\\_054247](https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/ref/?cid=nrcs142p2_054247) (accessed May. 2017).
- Chabrillat, S., E. Ben-Dor, R.A.V. Rossel, and J.A.M. Demattê. 2013. Quantitative soil spectroscopy. *Appl. Environ. Soil Sci.* 2013(i). doi: 10.1155/2013/616578.
- 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): 480–490. doi: 10.2136/sssaj2001.652480x.

- Clark, J. (2017) The Beer-Lambert Law. LibreTexts.  
[https://chem.libretexts.org/Core/Physical\\_and\\_Theoretical\\_Chemistry/Spectroscopy/Electronic\\_Spectroscopy/Electronic\\_Spectroscopy\\_Basics/The\\_Beer-Lambert\\_Law](https://chem.libretexts.org/Core/Physical_and_Theoretical_Chemistry/Spectroscopy/Electronic_Spectroscopy/Electronic_Spectroscopy_Basics/The_Beer-Lambert_Law)  
(accessed 5 Feb 2018).
- 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.
- Corwin, D.L., and S.M. Lesch. 2003. Application of Soil Electrical Conductivity to Precision Agriculture. *Agron. J.* 95(3): 455. doi: 10.2134/agronj2003.0455.
- Davies, A., and T. Fearn. 2006. Back to basics: calibration statistics. *Spectrosc. Eur.* 18(2): 31–32.  
[http://www.eurolabdanmark.dk/documents/Netvaerk/Kemi/SpecEurope\\_BackToBasics\\_CalibrationStats\\_TD.pdf](http://www.eurolabdanmark.dk/documents/Netvaerk/Kemi/SpecEurope_BackToBasics_CalibrationStats_TD.pdf).
- Dick, W.A., B. Thavamani, S. Conley, R. Blaisdell, and A. Sengupta. 2013. Prediction of  $\beta$ -glucosidase and  $\beta$ -glucosaminidase activities, soil organic C, and amino sugar N in a diverse population of soils using near infrared reflectance spectroscopy. *Soil Biol. Biochem.* 56: 99–104. doi: 10.1016/j.soilbio.2012.04.003.
- Estienne, F., L. Pasti, V. Centner, B. Walczak, F. Despagne, D.J. Rimbaut, O.E. De Noord, and D.L. Massart. 2001. A comparison of multivariate calibration techniques applied to experimental NIR data sets: Part II. Predictive ability under extrapolation conditions. *Chemom. Intell. Lab. Syst.* 58(2): 195–211. doi: 10.1016/S0169-7439(01)00159-9.
- Farmer, V.C., and J.D. Russell. 1967. Infrared Absorption Spectrometry in Clay Studies. *Clays Clay Miner.* 15(1): 121–142. doi: 10.1346/CCMN.1967.0150112.
- Ge, Y., J.A. Thomasson, and C.L.S. Morgan. 2014. Mid-infrared attenuated total reflectance spectroscopy for soil carbon and particle size determination. *Geoderma* 213: 57–63. doi: 10.1016/j.geoderma.2013.07.017.
- Geladi, P. 2003. Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochim. Acta - Part B At. Spectrosc.* 58(5): 767–782. doi: 10.1016/S0584-8547(03)00037-5.
- Groenigen, J.W. Van, C.S. Mutters, W.R. Horwath, and C. Van Kessel. 2003. NIR and DRIFT-MIR spectrometry of soils for predicting soil and crop parameters in a flooded field. *Plant Soil* 250(1): 155–165. doi: 10.1023/A:1022893520315.
- Henaka Arachchi, M.P.N.K., D.J. Field, and A.B. McBratney. 2016. Quantification of soil carbon from bulk soil samples to predict the aggregate-carbon fractions within using near- and mid-infrared spectroscopic techniques. *Geoderma* 267(December 2017): 207–214. doi: 10.1016/j.geoderma.2015.12.030.
- Islam, K., B. Singh, and a McBratney. 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Aust. J. Soil Res.* 41: 1101–1114. doi: Doi 10.1071/Sr02137.

- Janik, L.J., S.T. Forrester, and A. Rawson. 2009. Chemometrics and Intelligent Laboratory Systems The prediction of soil chemical and physical properties from mid-infrared spectroscopy and combined partial least-squares regression and neural networks ( PLS-NN ) analysis. *Chemom. Intell. Lab. Syst.* 97(2): 179–188. doi: 10.1016/j.chemolab.2009.04.005.
- Janik, L.J., R.H. Merry, and J.O. Skemstad. 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? *Aust. J. Exp. Agric.* 38: 681–696. doi: 10.1071/EA98132.
- Knadel, M., B. Stenberg, F. Deng, A. Thomsen, and M.H. Greve. 2013. Comparing predictive abilities of three visible-near infrared spectrophotometers for soil organic carbon and clay determination. *J. Near Infrared Spectrosc.* 21(1): 67–80. doi: 10.1255/jnirs.1035.
- Knox, N.M., S. Grunwald, M.L. McDowell, G.L. Bruland, D.B. Myers, and W.G. Harris. 2015. Geoderma Modelling soil carbon fractions with visible near-infrared ( VNIR ) and mid-infrared ( MIR ) spectroscopy. *Geoderma* 239–240: 229–239. doi: 10.1016/j.geoderma.2014.10.019.
- Lal, R. 2015. Sequestering carbon and increasing productivity by conservation agriculture. *J. Soil Water Conserv.* 70(3): 55A–62A. doi: 10.2489/jswc.70.3.55A.
- Liu, Y., Q. Jiang, T. Fei, J. Wang, T. Shi, K. Guo, X. Li, and Y. Chen. 2014. Transferability of a visible and near-infrared model for soil organic matter estimation in riparian landscapes. *Remote Sens.* 6(5): 4305–4322. doi: 10.3390/rs6054305.
- Martens, H. 2001. Reliable and relevant modelling of real world data: A personal account of the development of PLS Regression. *Chemom. Intell. Lab. Syst.* 58(2): 85–95. doi: 10.1016/S0169-7439(01)00153-8.
- Mehlich, A. 1984. Mehlich-3 soil test extractant--A modification of mehlich-2 extractant. *Commun. Soil. Sci. Plant Anal.* 15, 1409-1416.
- Mevik, B.H., M., R. Wehrens and K.H. Liland (2016). pls: Partial least squares and principal component regression. R package version 2.6-0. <https://CRAN.R-project.org/package=pls>
- Minasny, B., and A.B. Mcbratney. 2008. Chemometrics and Intelligent Laboratory Systems Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. 94: 72–79. doi: 10.1016/j.chemolab.2008.06.003.
- Minasny, B., A.B. McBratney, G. Tranter, and B.W. Murphy. 2008. Using soil knowledge for the evaluation of mid-infrared diffuse reflectance spectroscopy for predicting soil physical and mechanical properties. *Eur. J. Soil Sci.* 59(5): 960–971. doi: 10.1111/j.1365-2389.2008.01058.x.
- Minasny, B., G. Tranter, A.B. Mcbratney, D.M. Brough, and B.W. Murphy. 2009. *Geoderma* Regional transferability of mid-infrared diffuse reflectance

- spectroscopic prediction for soil chemical properties. *Geoderma* 153(1–2): 155–162. doi: 10.1016/j.geoderma.2009.07.021.
- Morel, C., H. Tunney, D. Plénet, and S. Pellerin. 2000. Transfer of Phosphate Ions between Soil and Solution: Perspectives in Soil Testing. *J. Environ. Qual.* 29:50-59. doi:10.2134/jeq2000.00472425002900010007x
- Nocita, M., A. Stevens, B. van Wesemael, M. Aitkenhead, M. Bachmann, B. Barth??s, E. Ben Dor, D.J. Brown, M. Clairotte, A. Csorba, P. Dardenne, J.A.M. Dematt??, 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. *Adv. Agron.* 132: 139–159. doi: 10.1016/bs.agron.2015.02.002.
- Olsen, S.R., C.V. Cole, F.S. Watanabe, and L.A. Dean. 1954. Estimation of available phosphorus in soils by extraction with sodium bicarbonate. U.S. Dep. of Agric. Circ. 939.
- Osibanjo, R., Rachael Curtis, and Lai Zijuan. 2017. Infrared: Theory. LibreTexts. [https://chem.libretexts.org/Core/Physical\\_and\\_Theoretical\\_Chemistry/Spectroscopy/Vibrational\\_Spectroscopy/Infrared\\_Spectroscopy/Infrared%3A\\_Theory](https://chem.libretexts.org/Core/Physical_and_Theoretical_Chemistry/Spectroscopy/Vibrational_Spectroscopy/Infrared_Spectroscopy/Infrared%3A_Theory) (accessed 5 Feb. 2018).
- Pirie, A., B. Singh, and K. Islam. 2005. Ultra-violet, visible, near-infrared, and mid-infrared diffuse reflectance spectroscopic techniques to predict several soil properties. *Aust. J. Soil Res.* 43: 713–721.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Reeves, J.B. 2010. *Geoderma* 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., G.W. McCarty, and V.B. Reeves. 2001. Mid-infrared diffuse reflectance spectroscopy for the quantitative analysis of agricultural soils. *J. Agric. Food Chem.* 49(2): 766–772. doi: 10.1021/jf0011283.
- Reeves, J.B., and D.B. Smith. 2009. *Applied Geochemistry* 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.
- Rinnan, Å., F. van den Berg, and S.B. Engelsen. 2009. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends Anal. Chem.* 28(10):

- 1201–1222. doi: 10.1016/j.trac.2009.07.007.
- Sarathjith, M.C., B.S. Das, S.P. Wani, and K.L. Sahrawat. 2014. Dependency Measures for Assessing the Covariation of Spectrally Active and Inactive Soil Properties in Diffuse Reflectance Spectroscopy. *Soil Sci. Soc. Am. J.* 78(5): 1522. doi: 10.2136/sssaj2014.04.0173.
- Shepherd, K.D., and M.G. Walsh. 2007. Infrared Spectroscopy---Enabling an Evidence-Based Diagnostic Surveillance Approach to Agricultural and Environmental Management in Developing Countries. *J. Near Infrared Spectrosc.* 15(1): 1–19. <http://www.osapublishing.org/jnirs/abstract.cfm?URI=jnirs-15-1-1>.
- Sparks, D.L. 2003. *Environmental soil chemistry*. Academic Press, San Diego.
- Stenberg, B., R. a V Rossel, a M. Mouazen, and J. Wetterlind. 2010. Visible and Near Infrared Spectroscopy in Soil Science. *Adv. Agron. Vol 107* 107(10): 163–215. doi: 10.1016/s0065-2113(10)07005-7.
- Tinti, A., V. Tugnoli, S. Bonora, and O. Francioso. 2015. Recent applications of vibrational mid-infrared (IR) spectroscopy for studying soil components: A review. *J. Cent. Eur. Agric.* 16(1): 1–22. doi: 10.5513/JCEA01/16.1.1535.
- Todorova, M., S. Atanassova, H. Lange, and D. Pavlov. 2011. Agriculture and Environment Estimation of total N , total P , pH and electrical conductivity in soil by near-infrared reflectance spectroscopy. 3(1): 50–54.
- Viscarra Rossel, R.A., D.J.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.
- Vohland, M., M. Ludwig, S. Thiele-Bruhn, and B. Ludwig. 2014. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. *Geoderma* 223–225(1): 88–96. doi: 10.1016/j.geoderma.2014.01.013.
- Weil, R.R., and F. Magdoff. 2004. Significance of soil organic matter to soil quality and health. In: F. Magdoff and R.R. Weil, editors, *Soil organic matter in sustainable agriculture*. CRC Press, Boca Raton, FL. p. 1–44.
- Wetterlind, J., B. Stenberg and R.A. Viscarra Rossel. 2013. Soil analysis using visible and near infrared spectroscopy. In *Plant Mineral Nutrients: Methods and Protocols*. (ed.) by Frans J. M. Maathuis, New York: Humana Press, Springer, pp 95-107. Published in serie: *Methods in molecular biology*, nr. 953.
- Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.
- Wijewardane, N.K., Y. Ge, S. Wills, and Z. Libohova. 2018. Predicting physical and chemical properties of us soils with a mid-infrared reflectance spectral library. *Soil Sci. Soc. Am. J.* 82(3). doi: 10.2136/sssaj2017.10.0361.

- Wijewardane, N.K., Y. Ge, S. Wills, and T. Loecke. 2016. Prediction of Soil Carbon in the Conterminous United States: Visible and Near Infrared Reflectance Spectroscopy Analysis of the Rapid Carbon Assessment Project. *Soil Sci. Soc. Am. J.* 80(4): 973. doi: 10.2136/sssaj2016.02.0052.
- Wold, S., M. Sjöström, and L. Eriksson. 2001. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 58(2): 109–130. doi: 10.1016/S0169-7439(01)00155-1.
- Zhang, L., G. Li, M. Sun, H. Li, Z. Wang, Y. Li, and L. Lin. 2017. Kennard-Stone combined with least square support vector machine method for noncontact discriminating human blood species. *Infrared Phys. Technol.* 86: 116–119. doi: 10.1016/j.infrared.2017.08.020.
- Zornoza, R., C. Guerrero, J. Mataix-Solera, K.M. Scow, V. Arcenegui, and J. Mataix-Beneyto. 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biol. Biochem.* 40(7): 1923–1930. doi: 10.1016/j.soilbio.2008.04.003.

## APPENDIX A: R CODE

## Sample R code and output

## Required libraries

```
library(readxl)      #to read excel files
library(ggplot2)     #to make plots
library(pls)         #to perform partial least squares regression
library(prospectr)   #to perform spectral preprocessing
```

## VNIR

## Reading in VNIR data and initial processing

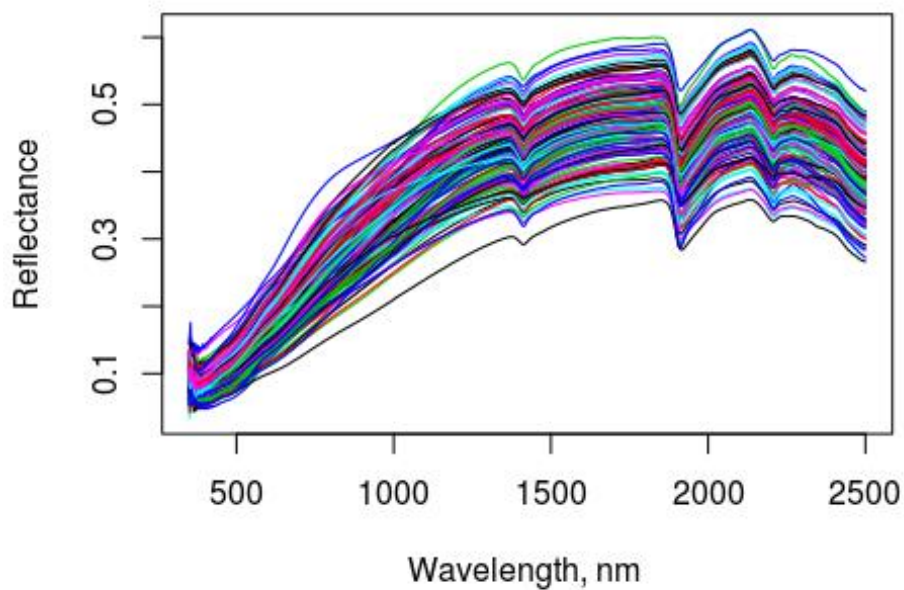
```
#import reference data
ref <- read_xlsx(path = "/home/jg8s/Dropbox/ref_data.xlsx", sheet = 1,
col_names = TRUE)

#import vnir files
vnir_spec <- read_xlsx(path =
"/home/jg8s/Dropbox/Spectra/VNIR/All_spectra.xlsx", sheet = 2,
col_names = TRUE)

# transpose spectral files-> samples as rows, wavelengths as columns
Wavelength <- vnir_spec$Wavelength
vnir_spec <- data.frame(t(vnir_spec))
colnames(vnir_spec) <- Wavelength ## Set wavelength as colnames
vnir_spec <- vnir_spec[-1,] ## Remove first redundant first row

# Plot reflectance
xax <- "Wavelength, nm " ## Set x-axis title
yax <- "Reflectance"      ## Set y-axis title
matplot(x= Wavelength, y = t(vnir_spec), xlab = xax, ylab = yax, type =
"l", lty = 1)
```



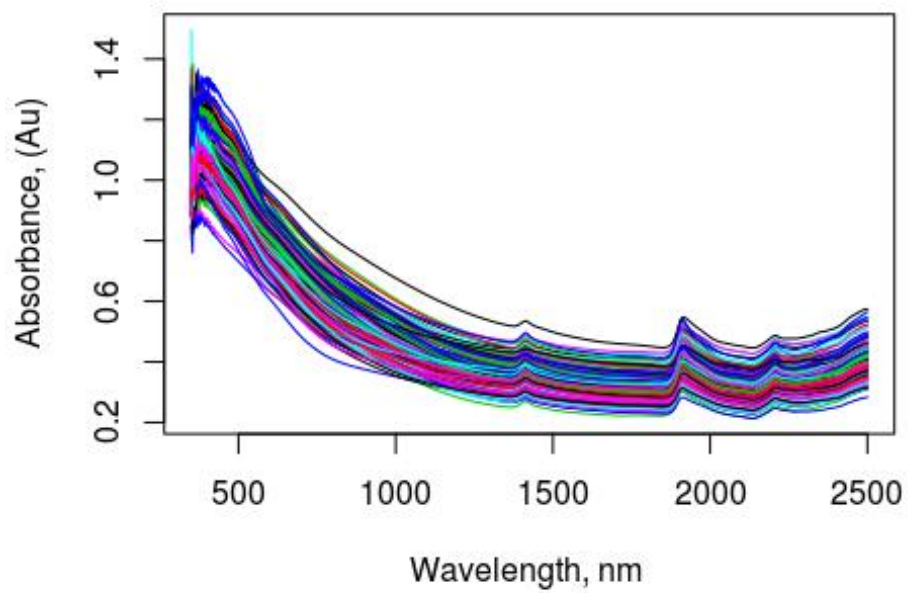


Splice correction between detectors

```
# Splice correction at 1000 nm and 1830 nm  
corrected_spc <- spliceCorrection(X = vnir_spec, wav = Wavelength,  
splice = c(1000,1830))
```

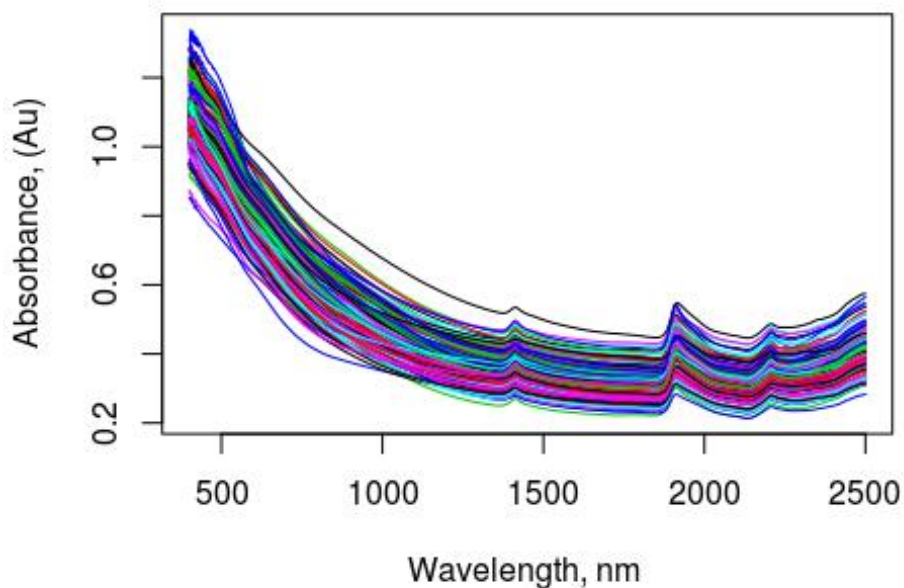
Reflectance to Absorbance Conversion

```
# Transform to absorbance ( log(1/R)  
abs_spec <- log10(1/corrected_spc)  
matplot(x= Wavelength, y = t(abs_spec), xlab = xax, ylab = "Absorbance,  
(Au)", type = "l", lty = 1)
```



Remove spectra below 400nm

```
raw_abs <- abs_spec[,51:2151]  
matplot(x= colnames(raw_abs), y = t(raw_abs), xlab = xax, ylab =  
"Absorbance, (Au)", type = "l", lty = 1)
```



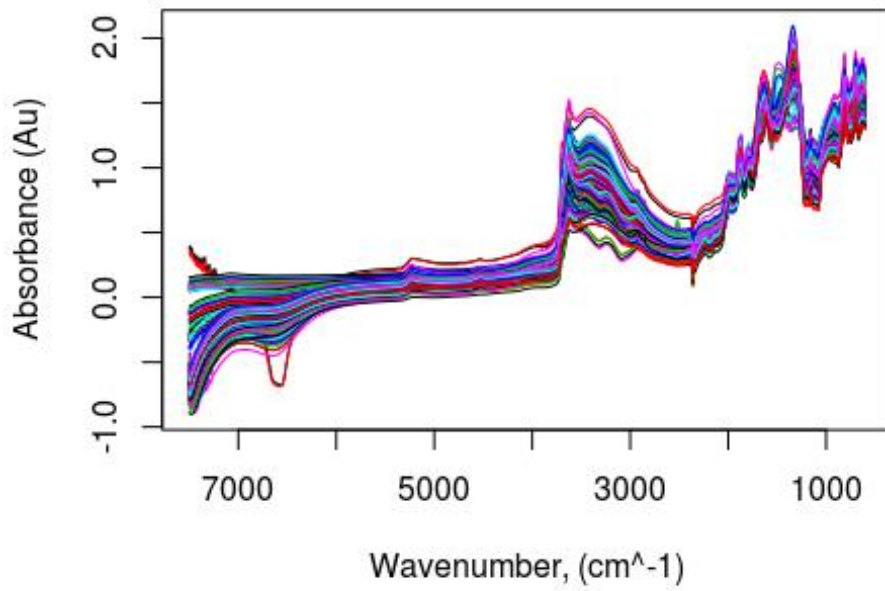
## MIR

Reading in MIR data

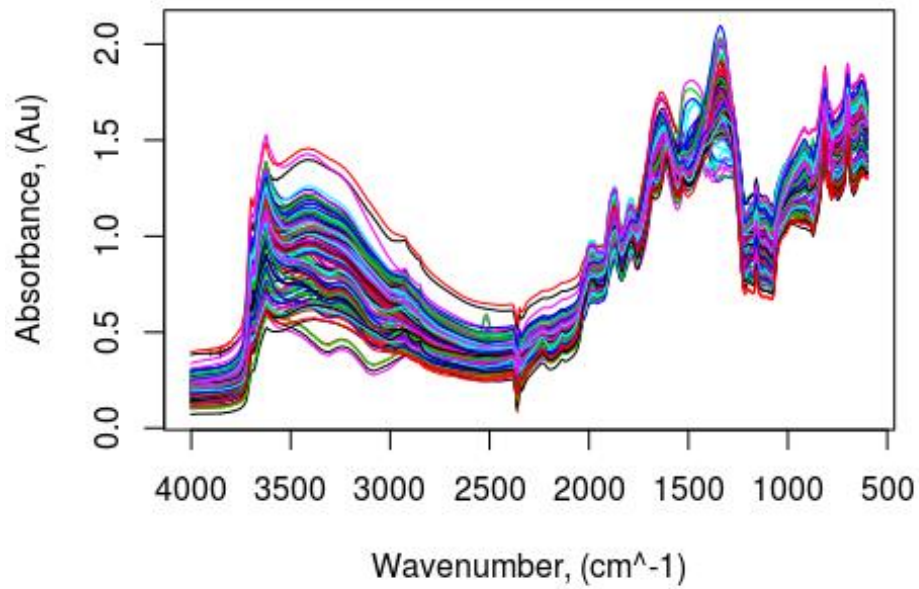
```
#import reference data
ref <- read_xlsx(path = "/home/jg8s/Dropbox/ref_data.xlsx", sheet = 1,
col_names = TRUE)
#import MIR files
mir_spec <- read_xlsx(path = "/home/jg8s/Dropbox/Pedometrics
code/working_dir/mir_avg.xlsx", col_names = TRUE)

# transpose -> samples as rows, wavelengths as columns
Wavelength <- mir_spec$Wavelength
mir_spec <- data.frame(t(mir_spec))
colnames(mir_spec) <- Wavelength
mir_spec <- mir_spec[-1,]

# Plot Absorbance
matplot(x= (Wavelength), y = t(mir_spec), xlab = "Wavenumber, (cm^-1)",
ylab = "Absorbance (Au)", type = "l", lty = 1, xlim = rev(c(600,7500)))
```



```
# remove Wavelengths below 4000 1/cm  
mir_trim <- mir_spec[,1814:3578]  
matplot(x= colnames(mir_trim), y = t(mir_trim), xlab = "Wavenumber,  
(cm^-1)", xlim = rev(c(600,4000)), ylab = "Absorbance, (Au)", type =  
"l", lty = 1)
```

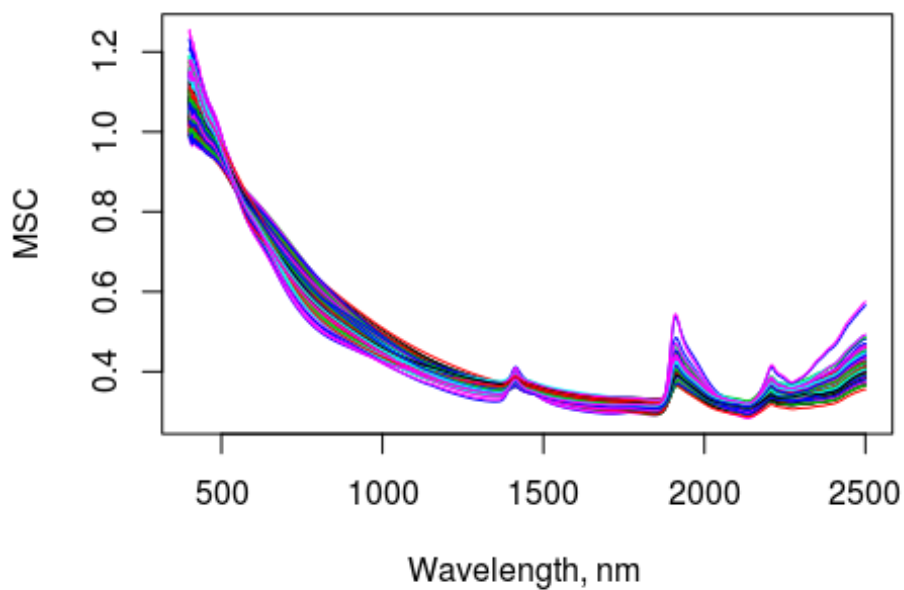


### Spectral Preprocessing

Examples given from VNIR spectra but are the same for MIR.

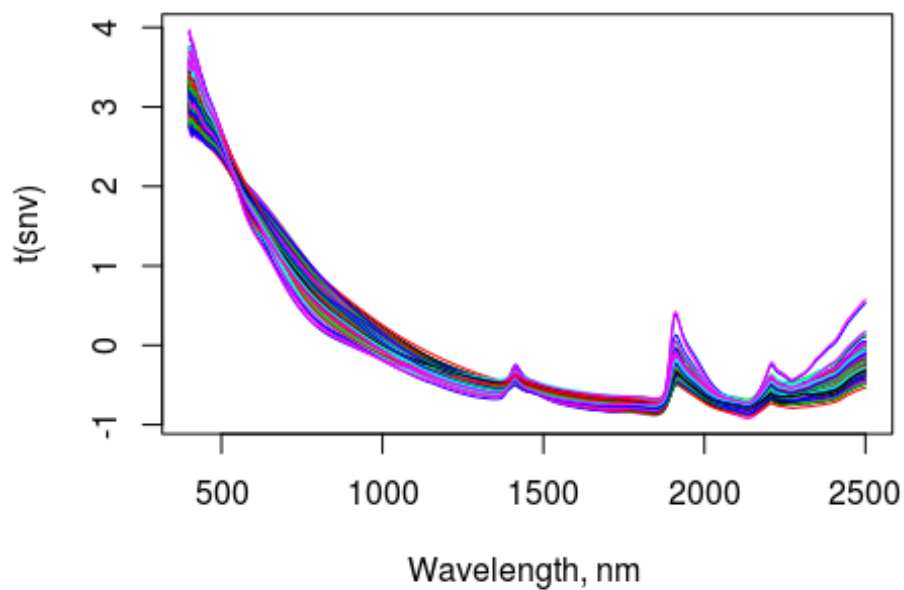
#### Multiplicative Scatter Correction

```
msc <- msc(raw_abs, reference = NULL)
matplot(x= colnames(msc), y = t(msc), xlab = xax, ylab = "MSC", type =
"l", lty = 1)
```



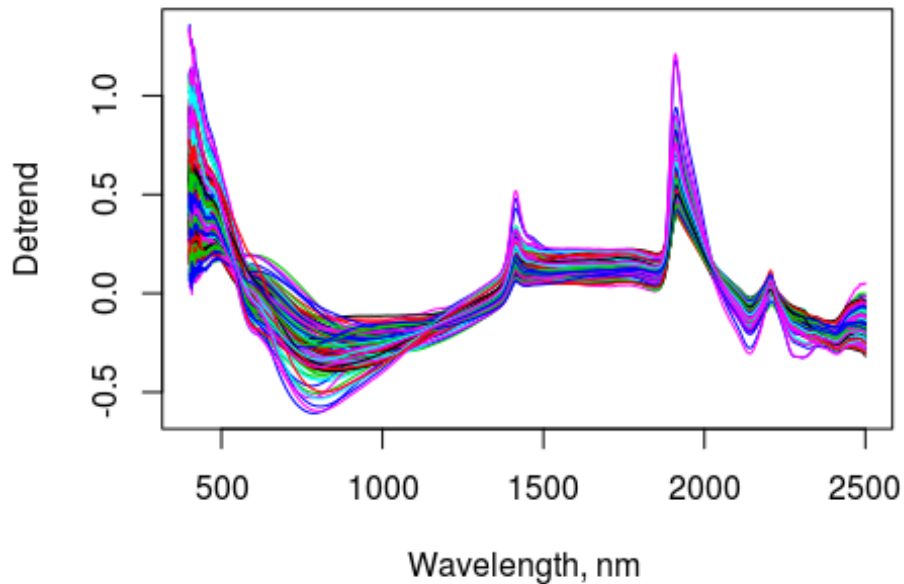
Standard Normal Variate

```
snv <- standardNormalVariate(X = raw_abs)  
matplot(x= colnames(snv), y = t(snv), xlab = xax, type = "l", lty = 1)
```



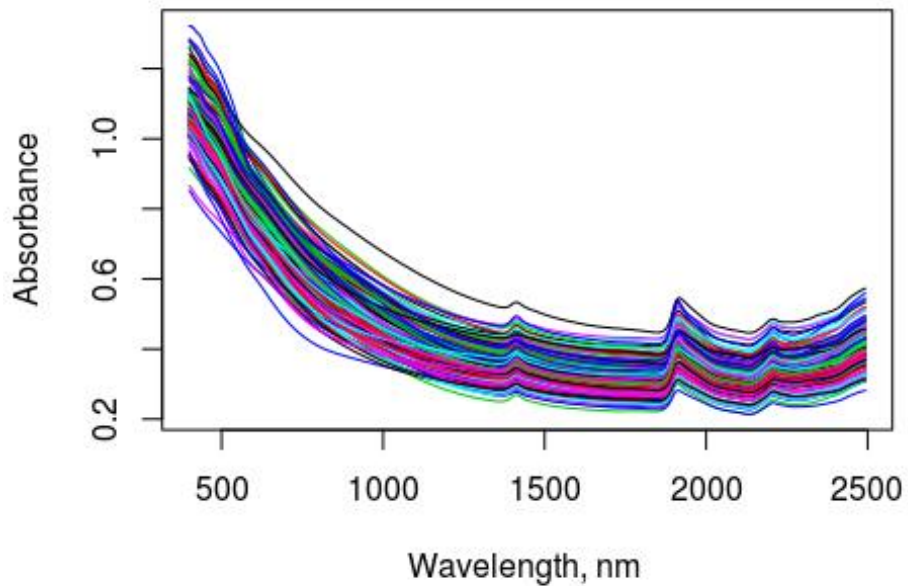
### Detrending

```
wav <- as.numeric(colnames(raw_abs)) ## set wavelengths
det <- detrend(raw_abs, wav)
matplot(x= colnames(det), y = t(det), xlab = xax, ylab= "Detrend", type
= "l", lty = 1)
```



### Savitzky-Golay smoothing

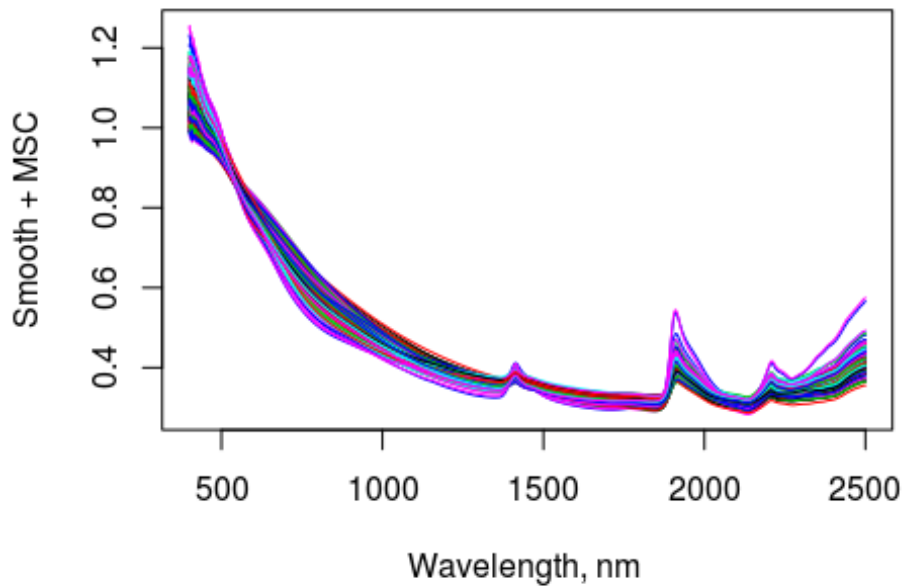
```
# Note:preprocessing performed on full spectra before trimming. This
process trims the ends of the spectra by half the window size.
# Window size = 15, derivative = 0, polynomial = 0
smooth_spec <- savitzkyGolay(X = abs_spec, m = 0, p = 0, w = 15)
smooth_abs <- smooth_spec[,44:2137] ## remove wavelengths below 400nm
matplot(x= colnames(smooth_abs), y = t(smooth_abs), xlab = xax, ylab =
"Absorbance", type = "l", lty = 1)
```



#### SG smoothing + MSC

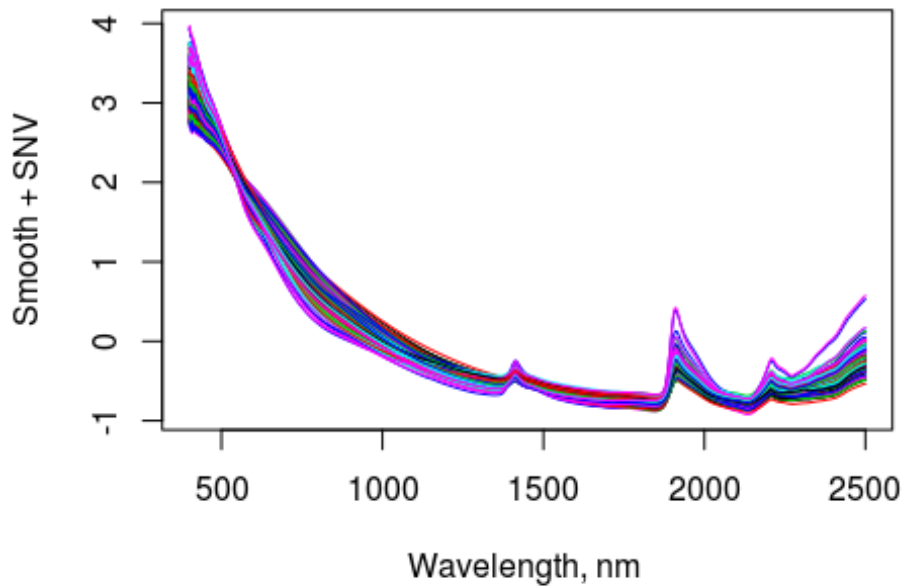
```
# Note:preprocessing performed on full spectra before trimming. This
process trims the ends of the spectra by half the window size. Spectra
is smoothed, then transformed by MSC and then trimmed.
# Window size = 15, derivative = 0, polynomial = 0
smooth_spec <- savitzkyGolay(X = abs_spec, m = 0, p = 0, w = 15)
msc2 <- msc(smooth_spec, reference = NULL)
msc2 <- msc2[,44:2137] ## Remove spectra below 400nm
matplot(x= colnames(msc), y = t(msc), xlab = xax, ylab = "Smooth +
MSC", type = "l", lty = 1)
```





#### SG smoothing + SNV

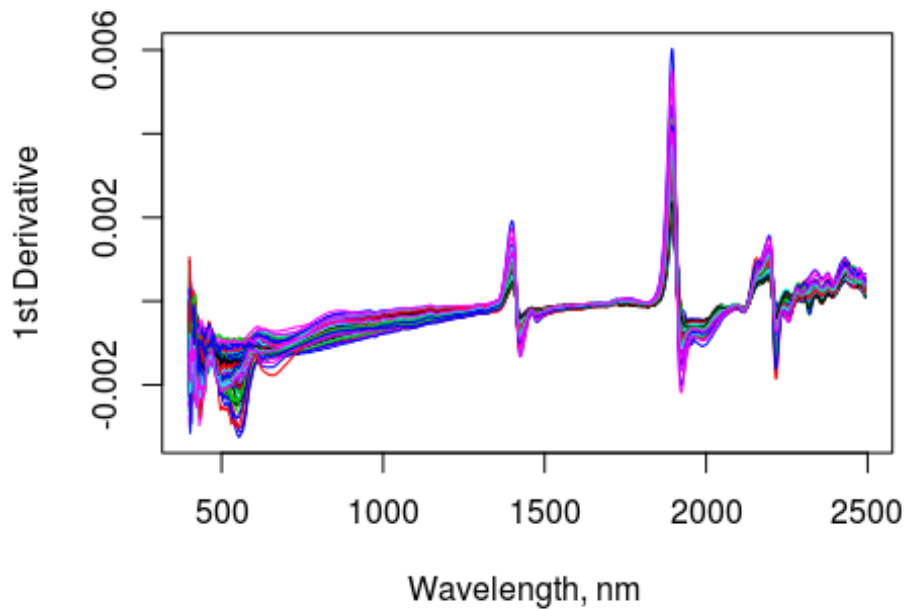
```
# Note:preprocessing performed on full spectra before trimming. This
process trims the ends of the spectra by half the window size. Spectra
is smoothed, then transformed by SNV and then trimmed.
# Window size = 15, derivative = 0, polynomial = 0
smooth_spec <- savitzkyGolay(X = abs_spec, m = 0, p = 0, w = 15)
snv2 <- standardNormalVariate(smooth_spec)
snv2 <- snv2[,44:2137] ## Remove spectra below 400nm
matplot(x= colnames(snv), y = t(snv), xlab = xax, ylab = "Smooth +
SNV", type = "l", lty = 1)
```



#### Savitsky-Golay 1st Derivative

# Note: preprocessing performed on full spectra before trimming. This process trims the ends of the spectra by half the window size. Spectra is smoothed and transformed by taking the first derivative and then trimmed.

```
# Window size = 15, derivative = 1, polynomial = 1
smooth_spec <- savitzkyGolay(X = abs_spec, m = 1, p = 1, w = 15)
smooth_1d <- smooth_spec[,44:2137]
matplot(x= colnames(smooth_1d), y = t(smooth_1d), xlab = xax, ylab =
"1st Derivative", type = "l", lty = 1)
```



#### Savitsky- Golay 2nd Derivative

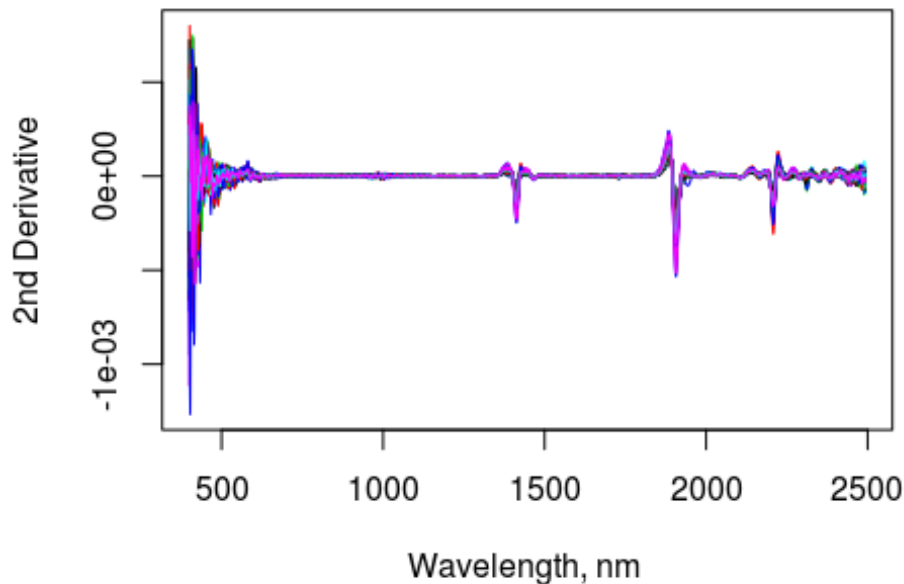
# Note: preprocessing performed on full spectra before trimming. This process trims the ends of the spectra by half the window size. Spectra is smoothed and transformed by taking the first derivative and then trimmed.

# Window size = 15, derivative = 2, polynomial = 2

```
smooth_spec <- savitzkyGolay(X = abs_spec, m = 2, p = 2, w = 15)
```

```
smooth_2d <- smooth_spec[,44:2137]
```

```
matplot(x= colnames(smooth_2d), y = t(smooth_2d), xlab = xax, ylab = "2nd Derivative", type = "l", lty = 1)
```



### Calibration Sampling Schemes

Combine Reference and Spectral data

```
soil_dat <- cbind(ref[,!(colnames(ref) %in%
as.character(colnames(raw_abs)))] ## create new data frame; insert
preprocessed spectra spectra here.
soil_dat$spec <- raw_abs ## adds raw spectra as a vector for each
sample; insert preprocessed spectra here.
```

### Random Sampling

```
# plot_id contains unique soil profiles and will be used to create
independent sets

n_profiles <- nlevels(as.factor(soil_dat$plot_id))
# randomly sample 75% of profiles for calibration
set.seed(4556)
cal_profile_id <- sample(x= levels(as.factor(soil_dat$plot_id)), size =
n_profiles *0.75)

cal1 <- soil_dat[soil_dat$plot_id %in% cal_profile_id,] ## training set
val1 <- soil_dat[!soil_dat$plot_id %in% cal_profile_id,] ## test set
```

### Stratified sampling

```
# Central pedons -> test; Satellites -> train
cal2 <- soil_dat[54:156,] ## last 2/3 of samples are from satellite
pedons
val2 <- soil_dat[1:53,] ## First 53 samples are from central pedons
```

### Kennard-Stone Sampling

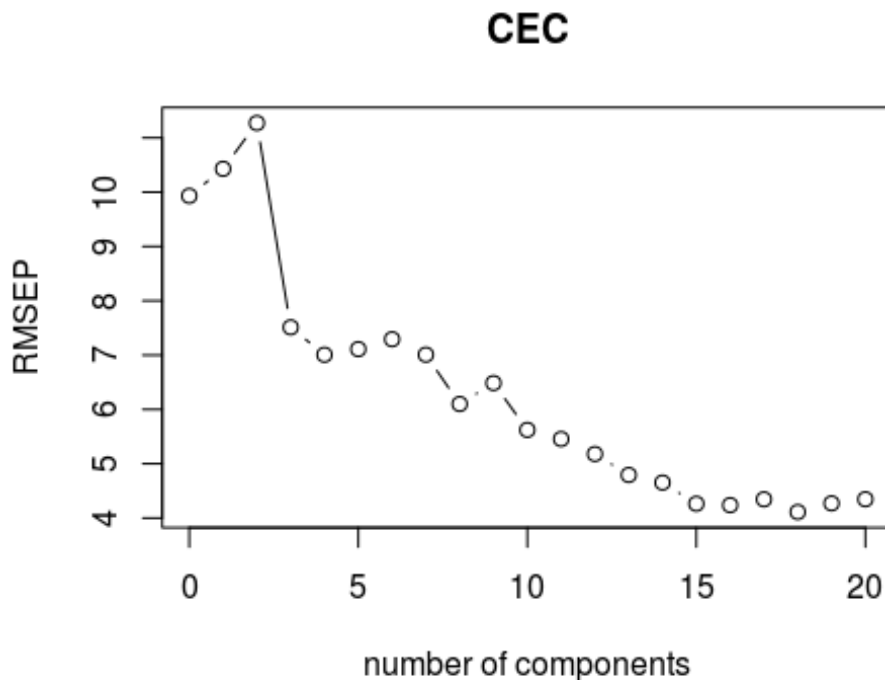
```
# KSS will sample spectra based on PC scores
set.seed(7823)
# choose sample size for calibration set = 81, based on Mahalanobis
distance in 20d principle component space, and grouped by soil profile.
kss <- kenStone(X= soil_dat$spec, k = 81, metric= "mahal", pc = 20,
group = as.factor(soil_dat$plot_id))

cal3 <- soil_dat[kss$model,] ## training
val3 <- soil_dat[kss$test,] ## test
```

### Partial least squares regression and prediction models

PLS model example for CEC using raw VNIR spectra and Random sampling.

```
## pls model built using the training set
pls_CEC <- plsr(CEC ~ spec, data = cal1 )
## visualize component with smallest error
validationplot(pls_CEC, val.type = "RMSEP", newdata = val1, ncomp =
1:20, type = "b")
```



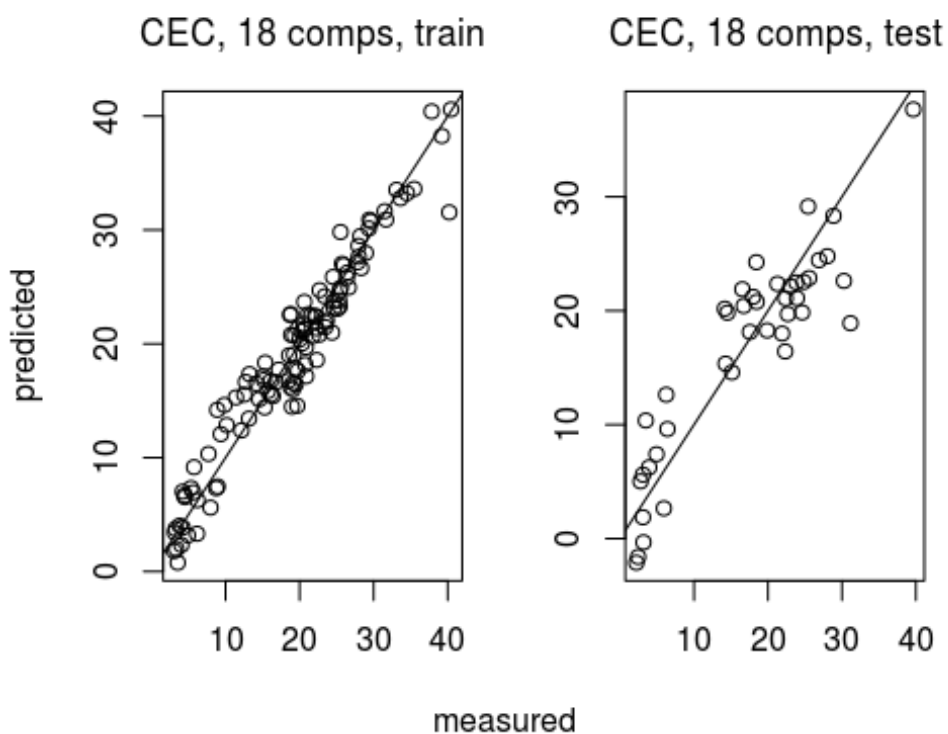
```
## choose number of latent variables to use from above
n_comp <- 18
## validate model by fitting test set
CEC_predicted <- predict(pls_CEC, ncomp = n_comp, newdata = val1)
## root mean square error
RMSEP(pls_CEC, estimate = "all", newdata = val1, ncomp= n_comp)
```

```
##      (Intercept) 18 comps
## train      9.042   2.229
## test       9.932   4.110

## fit
R2(pls_CEC, estimate = "all", newdata = val1, ncomp = n_comp)

##      (Intercept) 18 comps
## train      0.0000   0.9392
## test      -0.0435   0.8213

## prediction plots
predplot(pls_CEC, ncomp = n_comp, which = c("train", "test"), newdata =
val1, line = TRUE)
```



```
#Bias
mean(CEC_predicted) - mean(val1$CEC)

## [1] -0.2759712

#RPD
cec.sd <- sd(val1$CEC)
cec.rmsep <- sqrt((sum((CEC_predicted - val1$CEC)^2))/nrow(val1))
RPD <- cec.sd/cec.rmsep
RPD

## [1] 2.395048
```

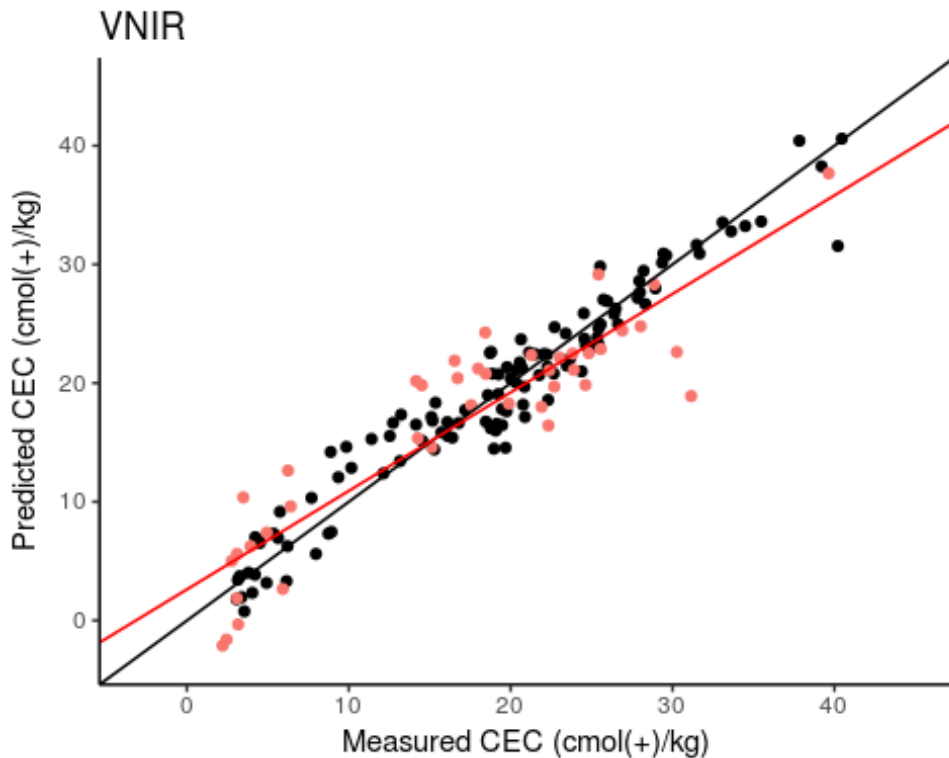
ggplot graphs

```
CEC_train <- predict(pls_CEC, ncomp= n_comp)

x <- val1$CEC
y <- CEC_predicted
cec<- cbind(x,y)
write.csv(cec, file = "VNIR_cec.csv") ##write to .csv allows one to get
slope and intercept after export to excel, used below.

VNIR_cec <- ggplot(data = NULL)+
  geom_point(aes(x= cal1$CEC, y= CEC_train))+
  geom_point(aes(x= val1$CEC, y= CEC_predicted, color= "red"),
show.legend = FALSE ) +
  geom_abline()+
  geom_abline(slope = 0.8299, intercept = 2.6197, color= 'red')+
  labs(x= "Measured CEC (cmol+)/kg)", y= "Predicted CEC (cmol+)/kg)",
title= "VNIR")+
  scale_x_continuous(limits= c(-3,45))+
  scale_y_continuous(limits = c(-3,45))+
  theme_classic()
```

VNIR\_cec



```
ggsave(filename = "vnir_cec.png", plot=VNIR_cec, path =
"/home/jg8s/Dropbox/Thesis/Plots/", width = 3, height = 3, units =
"in") ## exports and saves as .png in directory.
```

## APPENDIX B: SOIL REFERENCE DATA

Collected soil reference data and spectra can be downloaded from:

<https://github.com/josh-g8s/MS-thesis>

rcasiteid	plot id	sno id	LU	HZ	Top	Bottom	Lat	Lon	OSD	Clay	Sand	AD:OD
C0505C01	C0505C01-1	150984	C	Ap1	0	5	41.23	-96.83	Aksarben	38.30	5.39	1.026
C0505C01	C0505C01-1	150985	C	Ap2	5	19	41.23	-96.83	Aksarben	41.21	5.82	1.029
C0504R04	C0504R04-1	151006	R	A	0	5	42.68	-101.70	Valentine	2.86	95.56	1.004
C0504R04	C0504R04-1	151007	R	AC1	5	15	42.68	-101.70	Valentine	3.53	95.82	1.005
C0504R13	C0504R13-1	151031	R	A1	0	5	41.71	-101.26	Valentine	3.70	95.49	1.005
C0504R13	C0504R13-1	151032	R	A2	5	13	41.71	-101.26	Valentine	4.10	95.41	1.005
C0508C18	C0508C18-1	151127	C	Ap1	0	5	40.36	-99.28	Holdrege	23.88	13.96	1.033
C0508C18	C0508C18-1	151128	C	Ap2	5	35	40.36	-99.28	Holdrege	31.07	12.63	1.033
C0503C01	C0503C01-1	151146	C	Ap1	0	5	39.51	-95.29	Kennebec	32.29	2.75	1.027
C0503C01	C0503C01-1	151147	C	Ap2	5	21	39.51	-95.29	Kennebec	23.54	3.70	1.019
C0501P01	C0501P01-1	151839	P	Ap	0	5	38.30	-95.13	Collinsville	27.12	43.78	1.021
C0501P01	C0501P01-1	151840	P	A	5	12	38.30	-95.13	Collinsville	27.70	44.29	1.066
C0505R03	C0505R03-1	157520	R	Ap	0	5	38.29	-95.13	Wagstaff	43.37	3.54	1.060
C0505R03	C0505R03-1	157521	R	A	5	22	38.29	-95.13	Wagstaff	50.06	2.97	1.063
C0501R05	C0501R05-1	169739	R	A	0	5	41.16	-103.43	Busher	13.42	59.23	1.019
C0501R05	C0501R05-1	169740	R	Ack	5	35	41.16	-103.43	Busher	11.73	71.24	1.017
C0508R07	C0508R07-1	169780	R	A1	0	5	39.43	-96.60	Benfield	30.69	6.19	1.031
C0508R07	C0508R07-1	169781	R	A2	5	16	39.43	-96.60	Benfield	35.56	5.68	1.037
C0510P06	C0510P06-1	170365	P	Ap	0	5	37.39	-95.56	Lanton-like	24.10	20.29	1.021
C0510P06	C0510P06-1	170366	P	A1	5	19	37.39	-95.56	Lanton-like	26.15	20.19	1.019
C0503X05	C0503X05-1	170473	X	Ap1	0	5	39.25	-101.27	Bridgeport	46.20	15.08	1.052
C0503X05	C0503X05-1	170474	X	Ap3	9	28	39.25	-101.27	Bridgeport	40.34	2.89	1.050
C0511X02	C0511X02-1	184119	X	Ap1	0	5	37.92	-101.18	Bridgeport	25.30	49.41	1.019
C0511X02	C0511X02-1	184120	X	Ap2	5	12	37.92	-101.18	Bridgeport	25.93	49.26	1.018
C0508C12	C0508C12-1	184133	C	Ap1	0	5	38.30	-99.13	Harney	23.31	10.46	1.029
C0508C12	C0508C12-1	184134	C	Ap2	5	13	38.30	-99.13	Harney	23.31	11.05	1.030
C0508C35	C0508C35-1	184274	C	Ap1	0	5	39.86	-99.83	Holdrege	20.67	12.18	1.019



rcasiteid	plot id	smp id	LU	HZ	Top	Bottom	Lat	Lon	OSD	Clay	Sand	AD:OD
C0508C35	C0508C35-1	184275	C	Ap2	5	14	39.86	-99.83	Holdrege	23.52	8.69	1.022
C0505R02	C0505R02-1	184447	R	A1	0	5	37.55	-96.34	Sogn	44.480	8.943	1.056
C0505R02	C0505R02-1	184448	R	A2	5	11	37.55	-96.34	Sogn	44.250	9.205	1.056
C0506R14	C0506R14-1	184449	R	A1	0	5	40.17	-100.44	Coly	22.479	22.336	1.032
C0506R14	C0506R14-1	185000	R	A2	5	14	40.17	-100.44	Coly	24.203	19.122	1.034
C0511P10	C0511P10-1	188006	P	Ap1	0	5	40.56	-95.93	Nodaway	37.793	1.035	1.037
C0511P10	C0511P10-1	188007	P	Ap2	5	24	40.56	-95.93	Nodaway	36.268	1.170	1.035
C0502R05	C0502R05-1	188010	R	A	0	5	41.83	-103.57	Fluvaquentic Endoaquoll	25.090	39.467	1.029
C0502R05	C0502R05-1	188011	R	Ak	5	23	41.83	-103.57	Fluvaquentic Endoaquoll	26.369	42.562	1.024
C0508R03	C0508R03-1	188015	R	A1	0	5	41.17	-98.83	Uly	22.236	21.512	1.019
C0508R03	C0508R03-1	188016	R	A2	5	29	41.17	-98.83	Uly	27.900	21.399	1.025
C0511C08	C0511C08-1	188122	C	Ap1	0	5	40.34	-96.64	Kennebec	23.518	6.667	1.021
C0511C08	C0511C08-1	188123	C	Ap2	5	19	40.34	-96.64	Kennebec	25.542	7.891	1.021
C0511P02	C0511P02-1	188130	P	A1	0	5	41.45	-97.58	Inavale	#N/A	#N/A	1.002
C0511P02	C0511P02-1	188131	P	A2	5	19	41.45	-97.58	Inavale	5.963	77.295	1.006
C0508C29	C0508C29-1	188440	C	A	0	5	40.68	-101.39	Kuma	12.258	69.255	1.015
C0508X01	C0508X01-1	188448	X	Ap1	0	5	38.93	-98.38	Harney	37.535	5.017	1.040
C0508X01	C0508X01-1	188449	X	Ap2	5	15	38.93	-98.38	Harney	42.056	5.270	1.041
C0511P04	C0511P04-1	206317	P	A1	0	5	40.73	-97.78	Hobbs	26.748	5.876	1.023
C0511P04	C0511P04-1	206318	P	A2	5	24	40.73	-97.78	Hobbs	35.130	3.868	1.030
C0508P09	C0508P09-1	210070	P	Ap1	0	5	37.49	-99.22	Harney	19.217	23.669	1.022
C0508P09	C0508P09-1	210071	P	Ap2	5	24	37.49	-99.22	Harney	30.206	20.330	1.033
C0504R05	C0504R05-1	210110	R	A	0	5	42.45	-98.78	Ipaga	5.362	86.662	1.009
C0504R05	C0504R05-1	210111	R	A	5	20	42.45	-98.78	Ipaga	5.038	87.591	1.006
C0508R20	C0508R20-1	210369	R	A1	0	5	37.48	-97.87	Albion	7.076	78.762	1.010
C0508R20	C0508R20-1	210370	R	A2	5	32	37.48	-97.87	Albion	7.072	79.613	1.007

<b>rcasiteid</b>	<b>plot id</b>	<b>smp id</b>	<b>LU</b>	<b>HZ</b>	<b>Top</b>	<b>Bottom</b>	<b>Lat</b>	<b>Lon</b>	<b>OSD</b>	<b>Clav</b>	<b>Sand</b>	<b>AD:OD</b>
C0504R05	C0504R05-2	213001	R	A1	0	5	42.45	-98.78	lpage	#N/A	#N/A	1.006
C0504R05	C0504R05-2	213002	R	A2	5	18	42.45	-98.78	lpage	#N/A	#N/A	1.005
C0504R05	C0504R05-4	213009	R	A1	0	5	42.45	-98.78	lpage	#N/A	#N/A	1.006
C0504R05	C0504R05-4	213010	R	A2	5	17	42.45	-98.78	lpage	#N/A	#N/A	1.005
C0511P04	C0511P04-2	213691	P	A1	0	5	40.73	-97.78	Hobbs	#N/A	#N/A	1.027
C0511P04	C0511P04-2	213692	P	A2	5	20	40.73	-97.78	Hobbs	#N/A	#N/A	1.027
C0511P04	C0511P04-4	213700	P	A1	0	5	40.73	-97.78	Hobbs	#N/A	#N/A	1.021
C0511P04	C0511P04-4	213701	P	A2	5	38	40.73	-97.78	Hobbs	#N/A	#N/A	1.027
C0508P09	C0508P09-2	213830	P	Ap1	0	5	37.49	-99.22	Harnev	#N/A	#N/A	1.019
C0508P09	C0508P09-2	213831	P	Ap2	5	19	37.49	-99.22	Harnev	#N/A	#N/A	1.020
C0508P09	C0508P09-4	213840	P	Ap1	0	5	37.49	-99.22	Harnev	#N/A	#N/A	1.017
C0508P09	C0508P09-4	213841	P	Ap2	5	16	37.49	-99.22	Harnev	#N/A	#N/A	1.019
C0508R20	C0508R20-2	214429	R	A1	0	5	37.48	-97.87	Albion	#N/A	#N/A	1.007
C0508R20	C0508R20-2	214430	R	A2	5	34	37.48	-97.87	Albion	#N/A	#N/A	1.003
C0508R20	C0508R20-4	214438	R	A1	0	5	37.48	-97.87	Albion	#N/A	#N/A	1.006
C0508R20	C0508R20-4	214439	R	A2	5	27	37.48	-97.87	Albion	#N/A	#N/A	1.004
C0501R05	C0501R05-2	214754	R	A	0	5	41.16	-103.43	Busher	#N/A	#N/A	1.017
C0501R05	C0501R05-2	214755	R	ACK	5	38	41.16	-103.43	Busher	#N/A	#N/A	1.010
C0501R05	C0501R05-4	214763	R	A	0	5	41.16	-103.43	Busher	#N/A	#N/A	1.014
C0504R13	C0504R13-2	215596	R	A1	0	5	41.71	-101.26	Valentine	#N/A	#N/A	1.003
C0504R13	C0504R13-2	215597	R	A2	5	10	41.71	-101.26	Valentine	#N/A	#N/A	1.002
C0504R13	C0504R13-4	215606	R	A1	0	5	41.71	-101.26	Valentine	#N/A	#N/A	1.002
C0504R13	C0504R13-4	215607	R	A2	5	16	41.71	-101.26	Valentine	#N/A	#N/A	1.003
C0511P10	C0511P10-2	215915	P	Ap1	0	5	40.56	-95.93	Nodaway	#N/A	#N/A	1.035
C0511P10	C0511P10-2	215916	P	Ap2	5	20	40.56	-95.93	Nodaway	#N/A	#N/A	1.032
C0511P10	C0511P10-4	215923	P	Ap1	0	5	40.56	-95.93	Nodaway	#N/A	#N/A	1.027
C0511P10	C0511P10-4	215924	P	Ap2	5	23	40.56	-95.93	Nodaway	#N/A	#N/A	1.026

<b>rcasiteid</b>	<b>plot id</b>	<b>smp id</b>	<b>LU</b>	<b>HZ</b>	<b>Top</b>	<b>Bottom</b>	<b>Lat</b>	<b>Lon</b>	<b>OSD</b>	<b>Clav</b>	<b>Sand</b>	<b>AD:OD</b>
C0502R05	C0502R05-2	215931	R	Ak	0	5	41.83	-103.57	Fluvaquentic Endoaquoll	#N/A	#N/A	1.029
C0502R05	C0502R05-4	215942	R	Ak	0	5	41.83	-103.57	Fluvaquentic Endoaquoll	#N/A	#N/A	1.023
C0508R03	C0508R03-2	215955	R	A1	0	5	41.17	-98.83	Ulv	#N/A	#N/A	1.022
C0508R03	C0508R03-2	215956	R	A2	5	30	41.17	-98.83	Ulv	#N/A	#N/A	1.020
C0508R03	C0508R03-4	215966	R	A1	0	5	41.17	-98.83	Ulv	#N/A	#N/A	1.019
C0508R03	C0508R03-4	215967	R	A2	5	26	41.17	-98.83	Ulv	#N/A	#N/A	1.024
C0508X01	C0508X01-2	216168	X	Ap1	0	5	38.93	-98.38	Harnev	#N/A	#N/A	1.033
C0508X01	C0508X01-2	216169	X	Ap2	5	23	38.93	-98.38	Harnev	#N/A	#N/A	1.038
C0508X01	C0508X01-4	216179	X	Ap1	0	5	38.93	-98.38	Harnev	#N/A	#N/A	1.032
C0508X01	C0508X01-4	216180	X	Ap2	5	20	38.93	-98.38	Harnev	#N/A	#N/A	1.039
C0505R02	C0505R02-2	216395	R	A	0	7	37.55	-96.34	Sogn	#N/A	#N/A	1.040
C0505R02	C0505R02-4	216398	R	A	0	8	37.55	-96.34	Sogn	#N/A	#N/A	1.038
C0503X05	C0503X05-2	216822	X	Ap1	0	5	39.25	-101.27	Bridgeport	#N/A	#N/A	1.032
C0503X05	C0503X05-2	216823	X	Ap2	5	22	39.25	-101.27	Bridgeport	#N/A	#N/A	1.027
C0503X05	C0503X05-4	216832	X	Ap1	0	5	39.25	-101.27	Bridgeport	#N/A	#N/A	1.023
C0503X05	C0503X05-4	216833	X	Ap1	5	16	39.25	-101.27	Bridgeport	#N/A	#N/A	1.022
C0511P02	C0511P02-2	218005	P	A1	0	5	41.45	-97.58	Inavale	4.21	87.81	1.004
C0511P02	C0511P02-2	218006	P	A2	5	30	41.45	-97.58	Inavale	#N/A	#N/A	1.002
C0511P02	C0511P02-4	218014	P	A1	0	5	41.45	-97.58	Inavale	#N/A	#N/A	1.004
C0511P02	C0511P02-4	218015	P	A2	5	32	41.45	-97.58	Inavale	#N/A	#N/A	1.003
C0511C08	C0511C08-2	218059	C	Ap1	0	5	40.34	-96.64	Kennebec	#N/A	#N/A	1.019
C0511C08	C0511C08-2	218060	C	Ap2	5	20	40.34	-96.64	Kennebec	#N/A	#N/A	1.019
C0511C08	C0511C08-4	218067	C	Ap1	0	5	40.34	-96.64	Kennebec	#N/A	#N/A	1.018
C0511C08	C0511C08-4	218068	C	Ap2	5	20	40.34	-96.64	Kennebec	#N/A	#N/A	1.018
C0511X02	C0511X02-2	218166	X	Ap1	0	5	37.92	-101.18	Bridgeport	#N/A	#N/A	1.008
C0511X02	C0511X02-2	218167	X	Ap2	5	12	37.92	-101.18	Bridgeport	#N/A	#N/A	1.007
C0511X02	C0511X02-4	218176	X	Ap1	0	5	37.92	-101.18	Bridgeport	#N/A	#N/A	1.021

<b>rcasiteid</b>	<b>plot id</b>	<b>smp id</b>	<b>LU</b>	<b>HZ</b>	<b>Top</b>	<b>Bottom</b>	<b>Lat</b>	<b>Lon</b>	<b>OSD</b>	<b>Clav</b>	<b>Sand</b>	<b>AD:OD</b>
C0511X02	C0511X02-4	218177	X	Ap2	5	19	37.92	-101.18	Bridgeport	#N/A	#N/A	1.020
C0505R03	C0505R03-2	218185	R	Ap	0	5	38.29	-95.13	Wagstaff	#N/A	#N/A	1.031
C0505R03	C0505R03-2	218186	R	A	5	23	38.29	-95.13	Wagstaff	#N/A	#N/A	1.031
C0505R03	C0505R03-4	218195	R	Ap	0	5	38.29	-95.13	Wagstaff	#N/A	#N/A	1.034
C0505R03	C0505R03-4	218196	R	A	5	20	38.29	-95.13	Wagstaff	#N/A	#N/A	1.033
C0508C35	C0508C35-2	218426	C	Ap1	0	5	39.86	-99.83	Holdrege	#N/A	#N/A	1.021
C0508C35	C0508C35-2	218427	C	Ap2	5	11	39.86	-99.83	Holdrege	#N/A	#N/A	1.021
C0508C35	C0508C35-4	218439	C	Ap1	0	5	39.86	-99.83	Holdrege	#N/A	#N/A	1.013
C0508C35	C0508C35-4	218440	C	Ap2	5	11	39.86	-99.83	Holdrege	#N/A	#N/A	1.018
C0506R14	C0506R14-2	218592	R	A1	0	5	40.17	-100.44	Colv	#N/A	#N/A	1.028
C0506R14	C0506R14-2	218593	R	A2	5	21	40.17	-100.44	Colv	#N/A	#N/A	1.024
C0506R14	C0506R14-4	218603	R	A1	0	5	40.17	-100.44	Colv	#N/A	#N/A	1.027
C0506R14	C0506R14-4	218604	R	A2	5	15	40.17	-100.44	Colv	#N/A	#N/A	1.026
C0508R07	C0508R07-2	220218	R	Ap1	0	5	39.43	-96.60	Benfield	#N/A	#N/A	1.036
C0508R07	C0508R07-2	220219	R	Ap2	5	17	39.43	-96.60	Benfield	#N/A	#N/A	1.025
C0508R07	C0508R07-4	220228	R	Ap1	0	5	39.43	-96.60	Benfield	#N/A	#N/A	1.030
C0508R07	C0508R07-4	220229	R	Ap2	5	31	39.43	-96.60	Benfield	#N/A	#N/A	1.046
C0508C12	C0508C12-2	220275	C	Ap1	0	5	38.30	-99.13	Harney	#N/A	#N/A	1.017
C0508C12	C0508C12-2	220276	C	Ap2	5	16	38.30	-99.13	Harney	#N/A	#N/A	1.017
C0508C12	C0508C12-4	220285	C	Ap	0	5	38.30	-99.13	Harney	#N/A	#N/A	1.014
C0508C12	C0508C12-4	220286	C	A	5	19	38.30	-99.13	Harney	#N/A	#N/A	1.014
C0508C18	C0508C18-2	220385	C	A	0	5	40.36	-99.28	Holdrege	#N/A	#N/A	1.019
C0508C18	C0508C18-2	220386	C	AC1	5	30	40.36	-99.28	Holdrege	#N/A	#N/A	1.024
C0508C18	C0508C18-4	220395	C	A	0	5	40.36	-99.28	Holdrege	#N/A	#N/A	1.019
C0508C18	C0508C18-4	220396	C	AC	5	36	40.36	-99.28	Holdrege	#N/A	#N/A	1.020
C0508C29	C0508C29-2	220798	C	Ap	0	5	40.68	-101.39	Kuma	#N/A	#N/A	1.008
C0508C29	C0508C29-2	220799	C	A	5	32	40.68	-101.39	Kuma	#N/A	#N/A	1.009

<b>rcasiteid</b>	<b>plot_id</b>	<b>smpl_id</b>	<b>LU</b>	<b>HZ</b>	<b>Top</b>	<b>Bottom</b>	<b>Lat</b>	<b>Lon</b>	<b>OSD</b>	<b>Clay</b>	<b>Sand</b>	<b>AD:OD</b>
C0508C29	C0508C29-4	220806	C	Ap	0	5	40.68	-101.39	Kuma	#N/A	#N/A	1.007
C0508C29	C0508C29-4	220807	C	A	5	26	40.68	-101.39	Kuma	#N/A	#N/A	1.008
C0504R04	C0504R04-2	220902	R	Ap	0	5	42.68	-101.70	Valentine	#N/A	#N/A	1.002
C0504R04	C0504R04-2	220903	R	A	5	15	42.68	-101.70	Valentine	#N/A	#N/A	1.002
C0504R04	C0504R04-4	220911	R	Ap1	0	5	42.68	-101.70	Valentine	#N/A	#N/A	1.004
C0504R04	C0504R04-4	220912	R	Ap2	5	30	42.68	-101.70	Valentine	#N/A	#N/A	1.002
C0505C01	C0505C01-2	220940	C	Ap1	0	5	41.23	-96.83	Aksarben	#N/A	#N/A	1.026
C0505C01	C0505C01-2	220941	C	Ap2	5	25	41.23	-96.83	Aksarben	#N/A	#N/A	1.029
C0505C01	C0505C01-4	220951	C	Ap1	0	5	41.23	-96.83	Aksarben	#N/A	#N/A	1.027
C0505C01	C0505C01-4	220952	C	Ap2	5	21	41.23	-96.83	Aksarben	#N/A	#N/A	1.029
C0501P01	C0501P01-2	219519	P	A1	0	5	38.30	-95.13	Collinsville	#N/A	#N/A	1.019
C0501P01	C0501P01-2	219520	P	A2	5	17	38.30	-95.13	Collinsville	#N/A	#N/A	1.019
C0510P06	C0510P06-2	220761	P	Ap1	0	5	37.39	-95.56	Lanton-like	#N/A	#N/A	1.021
C0510P06	C0510P06-2	220762	P	Ap2	5	22	37.39	-95.56	Lanton-like	#N/A	#N/A	1.020
C0501P01	C0501P01-4	219525	P	A1	0	5	38.30	-95.13	Collinsville	#N/A	#N/A	1.020
C0501P01	C0501P01-4	219526	P	A2	5	13	38.30	-95.13	Collinsville	#N/A	#N/A	1.016
C0503C01	C0503C01-2	219876	C	Ap1	0	5	39.51	-95.29	Kennebec	#N/A	#N/A	1.016
C0503C01	C0503C01-2	219877	C	Ap2	5	31	39.51	-95.29	Kennebec	#N/A	#N/A	1.020
C0503C01	C0503C01-4	219884	C	Ap1	0	5	39.51	-95.29	Kennebec	#N/A	#N/A	1.016
C0503C01	C0503C01-4	219885	C	Ap2	5	37	39.51	-95.29	Kennebec	#N/A	#N/A	1.019
C0510P06	C0510P06-4	220771	P	Ap1	0	5	37.39	-95.56	Lanton-like	#N/A	#N/A	1.021
C0510P06	C0510P06-4	220772	P	Ap2	5	26	37.39	-95.56	Lanton-like	#N/A	#N/A	1.018

smpl id	EC	pH H2O	pH CaCl2	CEC	Ca	Mg	Na	K	H2O P	TP	TC	SOC	CaCO3	BG	TN
150984	0.66	5.2	4.99	23.40	13.95	5.50	0	1.46	0.86	371.66	2.07	2.07	0.00	140	0.28
150985	0.15	5.16	4.6	25.73	14.70	6.01	0	1.08	0.74	433.89	1.49	1.49	0.00	36	0.22
151006	0.26	5.77	5.44	3.17	3.02	0.54	0	0.25	1.04	116.37	0.55	0.55	0.00	38	0.06
151007	0.14	5.95	5.44	3.09	2.77	0.50	0	0.15	0.74	68.81	0.32	0.32	0.00	22	0.05
151031	0.09	6.01	5.4	3.27	2.96	0.57	0	0.17	0.77	101.94	0.48	0.48	0.00	65	0.06
151032	0.11	5.78	5.2	3.07	2.79	0.51	0	0.15	0.75	84.36	0.37	0.36	0.06	25	0.06
151127	0.76	5.76	5.57	22.27	18.75	2.21	0	1.50	1.19	595.15	3.61	3.61	0.04	265	0.34
151128	0.21	6.78	6.24	22.32	17.99	3.90	0	1.37	0.76	537.51	1.44	1.43	0.02	36	0.17
151146	0.06	6.01	5.29	24.55	18.24	5.55	0	0.38	1.06	453.24	1.18	1.18	0.00	2	0.12
151147	0.08	6	5.29	18.86	15.49	2.71	0	0.17	0.97	532.42	1.60	1.60	0.00	11	0.17
151839	0.27	6.01	5.61	19.47	17.30	2.32	0	0.27	0.89	269.42	3.33	3.32	0.05	141	0.28
151840	0.21	5.8	5.33	19.68	17.14	2.37	0	0.11	0.82	270.16	2.90	2.89	0.04	91	0.27
157520	0.46	6.08	5.7	31.15	22.00	2.22	0	0.88	0.87	386.79	5.14	5.13	0.14	175	0.45
157521	0.17	6.05	5.53	30.27	21.87	1.81	0	0.47	0.85	464.00	3.18	3.17	0.09	49	0.26
169739	0.26	7.65	7.13	15.17	18.19	0.88	0	1.23	0.84	429.04	1.98	1.97	0.06	167	0.19
169740	0.24	8.1	7.56	12.53	40.90	0.69	0	0.51	0.75	393.70	1.98	1.27	5.96	32	0.13
169780	0.36	6.26	5.9	22.69	16.08	6.15	0	1.03	0.76	276.43	2.34	2.32	0.13	151	0.22
169781	0.17	6.33	5.76	24.82	17.73	3.58	0.27	0.36	0.76	304.58	1.46	1.44	0.17	64	0.18
170365	0.46	5.21	4.9	22.25	16.56	4.38	0	0.44	1.12	496.88	4.25	4.23	0.11	147	0.41
170366	0.12	5.53	4.99	18.73	15.59	2.62	0	0.17	0.73	325.10	1.75	1.74	0.08	33	0.16
170473	1.11	6.89	6.75	40.21	32.93	4.16	0	3.85	1.51	1041.29	4.33	4.31	0.18	434	0.47
170474	0.38	7.8	7.42	35.48	57.29	3.04	0	2.74	1.11	750.03	1.93	1.73	1.66	101	0.21
184119	0.21	8.11	7.58	15.12	35.15	1.92	0	0.83	0.91	542.85	1.22	1.02	1.72	124	0.15
184120	0.21	8.14	7.64	14.27	36.51	2.00	0	0.68	0.80	505.54	1.14	0.84	2.54	52	0.13
184133	0.13	5.7	5.05	15.29	7.51	2.81	0	1.54	0.96	432.12	1.27	1.27	0.00	89	0.14
184134	0.13	5.54	4.93	15.10	8.11	2.79	0	1.01	0.88	432.35	1.13	1.13	0.00	57	0.14

smo id	EC	pH H2O	pH CaCl2	CEC	Ca	Mg	Na	K	H2O P	TP	TC	SOC	CaCO3	BG	TN
184274	0.37	5.1	4.77	14.15	6.32	2.06	0	2.15	1.35	613.01	1.21	1.21	0.00	45	0.14
184275	0.14	5.68	5.15	15.36	8.78	2.44	0	1.52	1.02	518.72	1.09	1.09	0.01	36	0.12
184447	0.31	7.61	7.21	37.83	43.82	1.52	0	0.60	0.76	866.88	3.62	3.49	1.13	107	0.36
184448	0.39	7.55	7.16	40.46	43.44	2.05	0	0.98	0.79	778.67	4.42	4.35	0.63	265	0.43
184449	0.26	6.87	6.4	22.06	17.95	2.96	0	1.29	0.94	594.21	3.01	3.01	0.02	248	0.28
185000	0.27	7.18	6.71	22.66	19.01	2.94	0	1.30	0.82	662.30	2.06	2.05	0.02	101	0.19
188006	0.45	6.15	5.75	33.62	21.72	5.85	0	1.71	1.56	942.47	4.43	4.41	0.17	289	0.46
188007	0.38	5.86	5.5	29.43	20.52	4.44	0	0.98	1.01	799.94	2.61	2.60	0.08	111	0.31
188010	0.91	7.86	7.59	22.71	63.15	9.27	0.38	2.20	0.91	724.22	5.12	4.17	7.88	279	0.43
188011	2.27	8.58	8.29	17.19	57.05	13.74	8.07	2.53	0.79	546.77	3.31	3.00	2.54	43	0.25
188015	1.03	5.69	5.46	23.46	13.86	4.87	0	2.57	1.63	609.88	3.40	3.40	0.03	240	0.37
188016	0.32	6.67	6.3	20.33	14.91	3.48	0	2.14	0.76	587.12	1.53	1.52	0.05	75	0.18
188122	0.22	5.28	4.82	17.99	11.16	2.69	0	0.77	0.84	404.73	1.83	1.82	0.08	89	0.20
188123	0.15	4.98	4.43	18.47	9.82	2.37	0	0.62	0.76	498.14	1.50	1.49	0.06	11	0.15
188130	0.08	6.26	5.62	3.10	1.67	0.55	0	0.48	0.88	227.81	0.44	0.44	0.00	70	0.04
188131	0.10	5.77	5.19	5.92	4.18	0.98	0	0.28	0.81	162.91	0.93	0.92	0.07	57	0.09
188440	0.48	5.75	5.51	10.16	6.24	2.14	0	1.05	0.92	424.26	1.28	1.26	0.12	87	0.17
188448	0.23	6.27	5.76	25.56	17.16	5.16	0	1.47	0.81	551.27	2.01	1.98	0.20	165	0.21
188449	0.14	5.97	5.8	26.39	18.93	4.69	0	1.27	0.85	418.39	1.09	1.07	0.17	71	0.14
206317	1.36	5.54	5.93	23.03	17.69	3.98	0	1.98	1.95	839.70	3.40	3.39	0.04	308	0.40
206318	0.52	6.26	5.28	26.90	17.65	4.61	0	1.86	0.91	518.81	2.20	2.20	0.02	99	0.28
210070	0.56	6.06	5.57	16.39	13.11	2.29	0	0.86	0.92	553.01	2.38	2.32	0.49	114	0.20
210071	0.15	6.04	5.93	20.77	15.96	2.97	0	0.97	0.82	460.07	1.23	1.22	0.12	52	0.15
210110	0.59	4.98	4.71	6.15	4.83	0.74	0	0.28	0.92	189.34	1.60	1.57	0.31	60	0.21
210111	0.20	5.29	4.73	4.20	3.34	0.43	0	0.24	0.76	150.37	0.86	0.84	0.15	21	0.11
210369	0.44	5.5	5.3	5.35	4.57	1.15	0	0.48	1.01	376.72	1.50	1.47	0.17	78	0.17
210370	0.16	5.64	4.85	4.46	3.68	0.87	0	0.13	0.77	144.28	0.72	0.71	0.06	8	0.09

smo	id	EC	pH H2O	pH CaCl2	CEC	Ca	Mg	Na	K	H2O P	TP	TC	SOC	CaCO3	BG	TN
213001	0.35	5.79	5.18	4.93	4.82	0.68	0	0.35	0.97	130.41	1.22	1.22	0.00	96	0.11	
213002	0.26	5.25	4.74	3.93	2.55	0.39	0	0.26	0.79	96.12	0.84	0.84	0.00	26	0.09	
213009	0.63	5.04	4.88	5.63	4.44	0.81	0	0.50	1.04	201.58	1.50	1.50	0.00	82	0.15	
213010	0.21	5.17	4.73	4.04	2.52	0.44	0	0.35	0.75	180.76	0.74	0.74	0.00	20	0.08	
213691	1.38	5.62	5.38	28.03	16.51	3.78	0	1.88	2.41	723.52	3.93	3.93	0.00	267	0.39	
213692	0.54	6.08	5.65	24.62	16.98	4.60	0	1.99	1.03	624.79	2.09	2.09	0.00	104	0.25	
213700	0.54	5.65	5.49	20.64	12.47	3.62	0	2.70	1.78	638.55	2.96	2.94	0.17	245	0.31	
213701	0.48	6.21	5.3	25.97	15.71	4.17	0	2.03	1.11	628.61	2.41	2.41	0.00	112	0.28	
213830	0.66	6.05	5.98	19.26	15.67	2.52	0	1.27	1.02	681.99	2.88	2.87	0.14	188	0.24	
213831	0.13	6.37	6.26	18.59	14.58	2.48	0	1.32	0.73	466.74	1.19	1.19	0.00	11	0.11	
213840	0.83	6.36	5.63	16.54	13.70	2.66	0	0.97	1.12	577.74	2.51	2.49	0.14	122	0.22	
213841	0.19	6.18	5.8	16.72	11.61	2.46	0	0.98	0.74	479.35	1.26	1.26	0.00	193	0.11	
214429	0.61	5.08	4.78	7.97	5.23	1.38	0	0.45	1.16	228.12	2.13	2.12	0.14	59	0.19	
214430	0.14	5.24	4.7	3.81	2.34	0.57	0	0.24	0.83	159.30	0.73	0.73	0.00	12	0.07	
214438	0.53	5.37	5.13	7.70	5.64	1.33	0	0.44	1.24	272.29	2.08	2.08	0.00	84	0.20	
214439	0.10	5.37	4.8	4.51	2.86	0.73	0	0.11	0.82	56.67	0.72	0.72	0.00	11	0.07	
214754	0.35	6	6.97	18.77	20.20	1.27	0	2.17	1.01	479.90	2.80	2.80	0.00	251	0.27	
214755	0.25	7.8	7.45	11.41	44.46	0.62	0	0.98	0.81	397.15	1.72	1.11	5.09	40	0.12	
214763	0.23	7.86	7.5	14.54	21.32	0.72	0	0.92	0.95	424.01	1.70	1.70	0.00	184	0.15	
215596	0.10	6.12	5.4	4.22	3.16	0.68	0	0.24	0.97	251.86	0.95	0.95	0.00	125	0.08	
215597	0.09	5.95	5.37	3.55	2.37	0.70	0	0.20	0.87	237.40	0.65	0.63	0.20	56	0.07	
215606	0.05	5.75	5.14	2.77	1.60	0.46	0	0.09	0.82	104.00	0.29	0.29	0.00	0	0.04	
215607	0.06	6.26	5.67	3.48	2.20	0.64	0	0.09	0.79	95.68	0.24	0.22	0.20	13	0.04	
215915	0.41	6.11	5.92	34.49	23.35	5.94	0	1.38	1.98	1006.92	3.84	3.84	0.03	239	0.41	
215916	0.32	6.38	6.03	31.49	22.40	5.57	0	0.99	1.15	783.81	2.56	2.55	0.08	120	0.29	
215923	0.49	6.03	5.69	28.86	18.92	4.42	0	1.61	1.30	854.61	3.91	3.91	0.02	230	0.38	
215924	0.28	6.33	5.66	25.56	16.18	4.07	0	1.06	0.98	618.64	2.08	2.08	0.00	69	0.23	



smo id	EC	pH H2O	pH CaCl2	CEC	Ca	Mg	Na	K	H2O P	TP	TC	SOC	CaCO3	BG	TN
215931	1.82	6.57	7.66	23.67	56.48	13.62	3.40	2.68	0.87	755.82	6.99	4.97	16.83	200	0.51
215942	5.62	7.82	7.74	20.85	65.19	11.73	11.69	2.35	0.86	710.78	5.38	4.13	10.47	129	0.43
215955	0.91	7.81	5.77	25.42	19.30	4.40	0	2.12	1.73	807.98	4.52	4.51	0.07	265	0.41
215956	0.23	6.62	6.21	19.89	12.83	4.26	0	1.74	0.88	686.00	1.15	1.15	0.02	51	0.12
215966	0.53	6.49	5.6	19.46	12.47	3.57	0	1.42	1.04	724.28	2.23	2.23	0.00	132	0.24
215967	0.22	6.44	6.04	21.53	14.80	4.37	0	1.38	0.83	652.93	1.17	1.16	0.06	65	0.14
216168	0.28	6.14	5.64	27.98	18.02	5.11	0	1.55	0.95	547.96	3.12	3.12	0.00	203	0.21
216169	0.13	6.57	6.39	29.61	21.40	5.21	0	0.97	0.83	527.68	0.76	0.76	0.00	23	0.09
216179	0.50	6.6	6.36	29.36	24.00	4.79	0	1.72	0.94	509.61	2.30	2.29	0.10	324	0.22
216180	0.28	7.12	7.17	28.94	28.67	5.84	0	0.93	0.81	527.63	0.68	0.63	0.38	19	0.08
216395	0.39	7.31	6.98	39.65	40.96	1.60	0	1.23	0.85	705.66	4.91	4.83	0.62	165	0.45
216398	0.37	7.47	6.95	39.21	42.78	2.37	0	1.08	0.85	810.07	4.42	4.39	0.30	212	0.44
216822	0.53	7.7	7.24	31.67	41.60	3.29	0	3.05	1.46	447.42	2.42	2.33	0.76	252	0.25
216823	0.33	7.78	7.36	28.32	38.10	2.53	0	2.41	1.03	285.21	1.67	1.57	0.88	85	0.18
216832	0.92	7.48	7.24	28.21	30.83	3.83	0	4.88	3.23	472.05	2.84	2.76	0.68	232	0.31
216833	0.46	7.75	7.32	25.53	36.74	2.70	0	3.97	2.14	543.84	1.67	1.55	0.99	161	0.18
218005	0.16	6.36	5.72	6.21	4.58	0.93	0	0.35	1.29	538.01	1.64	1.64	0.00	227	0.15
218006	0.07	6.2	5.53	3.38	2.38	0.49	0	0.16	0.86	463.00	0.49	0.46	0.29	23	0.05
218014	0.31	6.71	6.32	8.74	7.60	1.59	0	0.81	1.55	411.69	2.33	2.32	0.10	266	0.22
218015	0.15	6.12	5.7	4.92	4.02	1.08	0	0.33	1.16	701.71	0.76	0.76	0.00	57	0.08
218059	0.62	5.38	4.93	20.56	13.41	3.18	0	1.30	1.39	568.11	2.34	2.34	0.00	151	0.22
218060	0.85	5.04	3.94	20.00	10.95	2.80	0	0.88	0.91	678.79	1.32	1.32	0.00	49	0.14
218067	0.28	5.04	4.44	20.33	10.75	2.96	0	1.11	1.53	653.95	1.94	1.94	0.00	70	0.17
218068	0.10	4.94	4.29	20.70	11.68	3.11	0	0.52	0.84	878.30	1.54	1.54	0.00	18	0.14
218166	0.13	7.62	7.17	6.41	7.46	0.88	0	0.28	0.86	754.34	0.37	0.37	0.00	23	0.03
218167	0.11	7.97	7.42	6.24	9.82	0.73	0	0.25	0.80	973.54	0.40	0.40	0.00	91	0.03
218176	0.27	7.73	7.38	16.78	33.83	2.49	0	1.25	0.89	799.27	1.85	1.66	1.57	184	0.16

smo	id	EC	pH	H2O	pH	CaCl2	CEC	Ca	Mg	Na	K	H2O	P	TP	TC	SOC	CaCO3	BG	TN
218177	0.22	7.8	7.42	15.72	44.53	2.52	0	0.84	0.82	268.74	1.40	1.07	2.76	61	0.10				
218185	0.65	6	5.49	27.95	20.01	2.08	0	1.10	0.92	281.20	4.92	4.92	0.00	209	0.39				
218186	0.16	5.87	5.31	25.11	19.70	1.79	0	0.43	0.83	340.45	3.03	3.02	0.06	59	0.26				
218195	0.47	5.76	5.59	27.84	21.24	2.08	0	0.93	0.91	574.61	4.62	4.60	0.14	183	0.37				
218196	0.14	5.75	5.13	25.42	17.17	1.82	0	0.43	0.82	520.08	3.07	3.07	0.00	51	0.26				
218426	0.66	5.87	4.54	16.09	8.88	2.90	0	1.83	1.29	513.05	1.54	1.54	0.00	73	0.16				
218427	0.17	5.35	4.8	16.11	9.46	2.88	0	1.34	0.99	616.27	1.09	1.09	0.06	35	0.12				
218439	0.61	4.72	4.36	12.14	5.40	1.76	0	2.14	1.48	660.45	1.24	1.24	0.00	36	0.13				
218440	0.17	5.35	4.88	13.16	7.49	2.11	0	1.61	1.05	434.90	1.00	1.00	0.00	29	0.11				
218592	0.40	7.42	7.07	24.64	24.15	2.90	0	1.22	1.09	336.31	2.91	2.91	0.00	271	0.26				
218593	0.28	7.86	7.45	19.78	36.49	2.10	0	1.14	0.84	552.54	1.30	1.09	1.78	39	0.11				
218603	0.38	7.63	7.18	23.83	23.99	2.91	0	1.18	1.07	419.43	2.81	2.75	0.53	234	0.25				
218604	0.33	7.9	7.32	22.33	40.14	2.33	0	1.10	0.91	635.33	1.87	1.63	1.96	91	0.16				
220218	0.32	6.2	5.45	25.35	16.93	4.69	0.00	0.71	0.88	358.06	3.10	3.07	0.21	155	0.29				
220219	0.28	5.95	4.91	21.14	13.72	3.87	0.11	0.34	0.81	322.67	1.97	1.95	0.20	152	0.19				
220228	0.72	5.34	5.17	24.51	14.77	5.00	0.61	0.72	0.84	341.62	2.30	2.27	0.21	30	0.22				
220229	0.14	6.88	6.84	33.09	20.83	8.49	2.96	0.46	0.83	241.06	1.05	1.05	0.00	21	0.12				
220275	0.11	6.5	4.8	14.17	7.48	2.50	0	1.39	1.39	494.46	1.40	1.40	0.00	84	0.13				
220276	0.08	5.39	4.64	14.51	7.61	2.37	0	1.02	1.14	574.60	1.05	1.04	0.05	41	0.11				
220285	0.86	4.44	4.26	13.23	6.26	2.21	0	1.59	1.08	525.53	1.18	1.18	0.00	30	0.13				
220286	0.24	4.97	4.22	12.74	6.02	2.11	0	1.23	0.96	395.89	1.01	1.01	0.00	20	0.12				
220385	0.49	5.6	5.61	18.47	15.29	1.92	0	1.50	1.36	529.00	1.99	1.99	0.00	156	0.19				
220386	0.36	6.13	6.06	20.88	13.69	3.33	0	1.00	0.98	465.00	1.33	1.33	0.00	19	0.13				
220395	0.69	6.23	5.49	18.97	15.17	1.85	0	1.64	1.49	704.12	2.29	2.29	0.00	184	0.22				
220396	0.26	6.23	6.25	19.08	15.67	2.78	0	1.10	0.81	494.90	1.32	1.32	0.00	17	0.13				
220798	0.59	5.62	5.25	9.85	6.23	1.97	0	1.16	1.07	427.58	1.44	1.44	0.00	104	0.14				
220799	0.30	5.72	5.46	9.36	5.97	1.86	0	1.07	0.86	293.93	0.89	0.89	0.00	43	0.10				

smpl_id	EC	pH H2O	pH CaCl2	CEC	Ca	Mg	Na	K	H2O P	TP	TC	SOC	CaCO3	BG	TN
220806	0.48	5.37	5.17	8.88	5.69	1.75	0	1.11	1.21	324.53	1.45	1.45	0.00	98	0.15
220807	0.32	5.76	5.45	8.93	5.78	1.87	0	0.98	0.88	356.54	1.00	1.00	0.00	53	0.12
220902	0.16	5.57	5.11	2.45	1.68	0.32	0	0.00	0.85	73.02	0.45	0.45	0.00	16	0.05
220903	0.09	5.92	5.37	2.21	1.55	0.30	0	0.00	0.81	86.62	0.29	0.29	0.00	20	0.04
220911	0.37	5.26	4.94	5.75	4.28	0.77	0	0.44	1.12	44.01	1.65	1.64	0.09	91	0.14
220912	0.11	5.63	5.17	3.15	2.58	0.34	0	0.12	0.82	151.16	0.44	0.43	0.09	16	0.05
220940	0.34	5.16	4.55	24.38	14.94	5.78	0	1.01	0.84	648.52	2.09	2.09	0.00	91	0.22
220941	0.19	5.13	4.24	26.63	12.40	5.29	0	0.62	0.81	585.59	1.96	1.95	0.05	36	0.21
220951	0.46	5.48	4.7	25.43	13.49	5.35	0	1.17	0.88	564.76	2.05	2.05	0.03	144	0.23
220952	0.15	4.97	4.25	26.49	13.84	5.54	0	0.87	0.80	483.04	1.47	1.47	0.00	41	0.16
219519	0.29	5.89	5.38	21.76	16.17	2.31	0	0.40	0.97	461.23	3.76	3.74	0.13	136	0.32
219520	0.15	6.25	5.05	18.79	14.04	2.14	0	0.17	0.84	275.19	2.27	2.19	0.66	40	0.20
220761	0.75	5.04	4.88	23.90	17.80	2.63	0	0.49	1.12	559.68	4.05	4.04	0.10	138	0.38
220762	0.14	5.17	4.66	21.28	15.37	2.25	0	0.30	0.81	338.06	2.04	2.04	0.00	28	0.17
219525	0.45	5.89	5.73	22.39	17.39	2.41	0	0.47	0.98	398.83	3.77	3.77	0.00	169	0.31
219526	0.17	5.94	5.39	17.56	12.84	2.14	0	0.12	0.85	471.22	1.75	1.75	0.00	64	0.15
219876	0.18	5.22	4.69	18.43	11.70	2.91	0	0.15	1.06	656.59	2.43	2.43	0.00	68	0.21
219877	0.06	5.8	4.67	21.91	15.12	3.87	0	0.12	0.82	349.62	1.45	1.43	0.20	10	0.12
219884	0.12	5.55	4.84	19.13	13.52	2.96	0	0.25	1.18	573.77	1.38	1.37	0.08	41	0.13
219885	0.07	5.83	5.36	19.75	13.67	3.15	0	0.18	0.86	456.95	0.93	0.92	0.02	4	0.09
220771	0.69	5.07	4.94	23.47	16.34	3.62	0	0.49	1.06	483.90	3.44	3.43	0.11	134	0.34
220772	0.13	5.77	5.26	19.24	14.96	2.95	0	0.38	0.82	306.82	1.54	1.53	0.08	36	0.15

## APPENDIX C: SOIL PROPERTY CORRELATIONS

Correlation matrix of the collected soil reference data.																	
	EC	1:1 pH	1:2 pH	CEC	Ca	Mg	Na	K	H2O P	TP	TC	SOC	CaCO3	BG	TN	Clay%	Sand%
EC	1.00	0.18	0.26	0.17	0.46	0.56	0.83	0.44	0.24	0.28	0.52	0.45	0.56	0.31	0.48	0.14	-0.10
1:1 pH	0.18	1.00	0.94	0.28	0.73	0.19	0.26	0.33	0.01	0.29	0.20	0.14	0.38	0.30	0.15	0.21	0.01
1:2 pH	0.26	0.94	1.00	0.30	0.77	0.25	0.31	0.39	0.03	0.30	0.27	0.20	0.47	0.36	0.22	0.24	-0.02
CEC	0.17	0.28	0.30	1.00	0.62	0.55	0.05	0.47	0.20	0.62	0.66	0.68	0.06	0.48	0.72	0.94	-0.83
Ca	0.46	0.73	0.77	0.62	1.00	0.54	0.44	0.49	0.05	0.45	0.59	0.50	0.64	0.37	0.54	0.52	-0.31
Mg	0.56	0.19	0.25	0.55	0.54	1.00	0.61	0.46	0.08	0.42	0.49	0.42	0.50	0.24	0.48	0.47	-0.47
Na	0.83	0.26	0.31	0.05	0.44	0.61	1.00	0.21	-0.08	0.09	0.28	0.19	0.56	-0.01	0.20	0.02	0.04
K	0.44	0.33	0.39	0.47	0.49	0.46	0.21	1.00	0.61	0.50	0.39	0.37	0.25	0.54	0.46	0.44	-0.45
H2OP	0.24	0.01	0.03	0.20	0.05	0.08	-0.08	0.61	1.00	0.34	0.27	0.31	-0.08	0.51	0.35	0.07	-0.20
TP	0.28	0.29	0.30	0.62	0.45	0.42	0.09	0.50	0.34	1.00	0.49	0.48	0.16	0.46	0.54	0.62	-0.63
TC	0.52	0.20	0.27	0.66	0.59	0.49	0.28	0.39	0.27	0.49	1.00	0.99	0.40	0.70	0.96	0.54	-0.44
SOC	0.45	0.14	0.20	0.68	0.50	0.42	0.19	0.37	0.31	0.48	0.99	1.00	0.24	0.72	0.97	0.56	-0.47
CaCO3	0.56	0.38	0.47	0.06	0.64	0.50	0.56	0.25	-0.08	0.16	0.40	0.24	1.00	0.11	0.29	-0.02	0.14
BG	0.31	0.30	0.36	0.48	0.37	0.24	-0.01	0.54	0.51	0.46	0.70	0.72	0.11	1.00	0.74	0.31	-0.24
TN	0.48	0.15	0.22	0.72	0.54	0.48	0.20	0.46	0.35	0.54	0.96	0.97	0.29	0.74	1.00	0.60	-0.51
*Clay %	0.14	0.21	0.24	0.94	0.52	0.47	0.02	0.44	0.07	0.62	0.54	0.56	-0.02	0.31	0.60	1.00	-0.86
*Sand%	-0.10	0.01	-0.02	-0.83	-0.31	-0.47	0.04	-0.45	-0.20	-0.63	-0.44	-0.47	0.14	-0.24	-0.51	-0.86	1.00

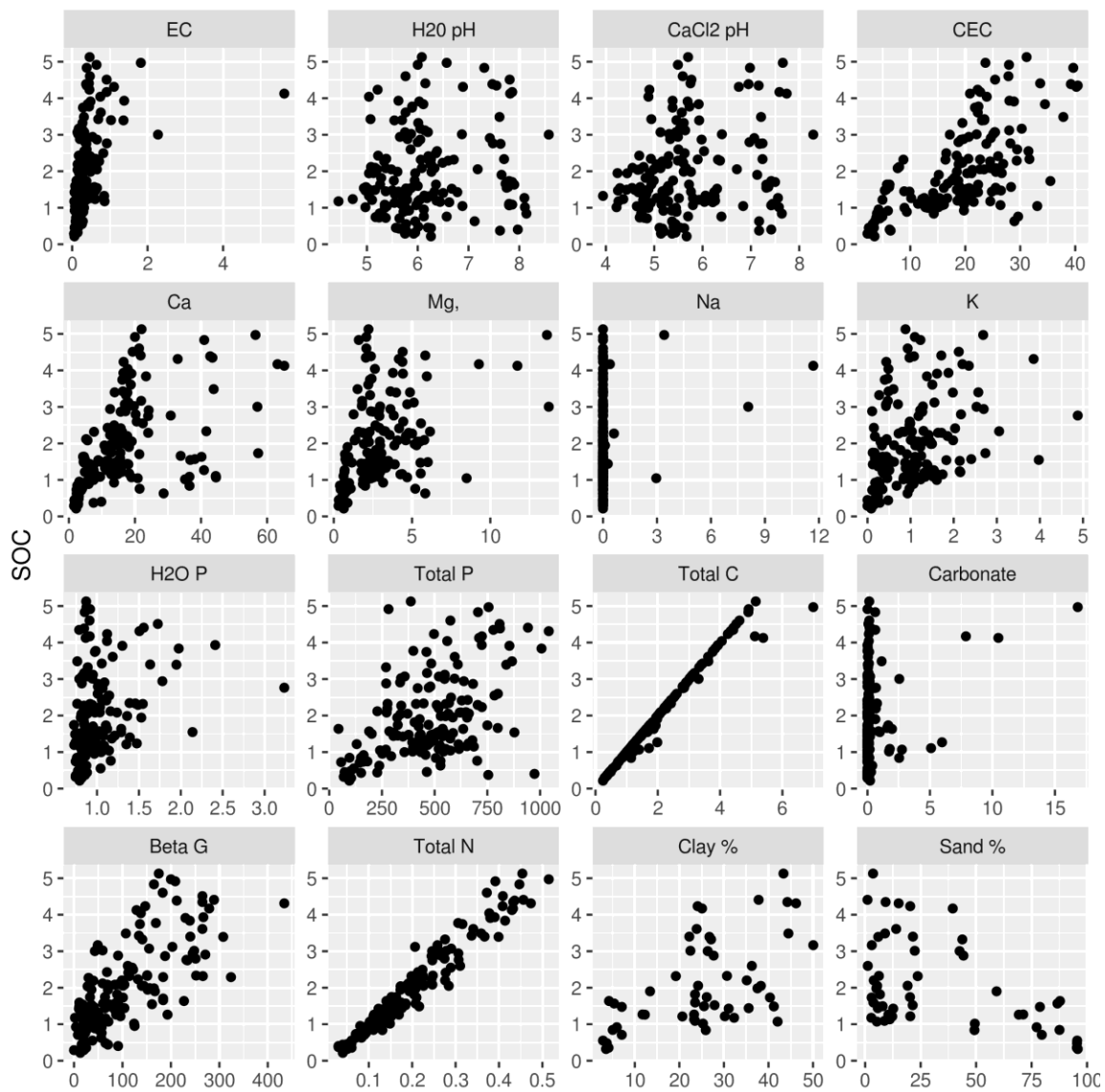


Figure C1 Scatter plot matrix of SOC relationship with the remaining measured soil properties.

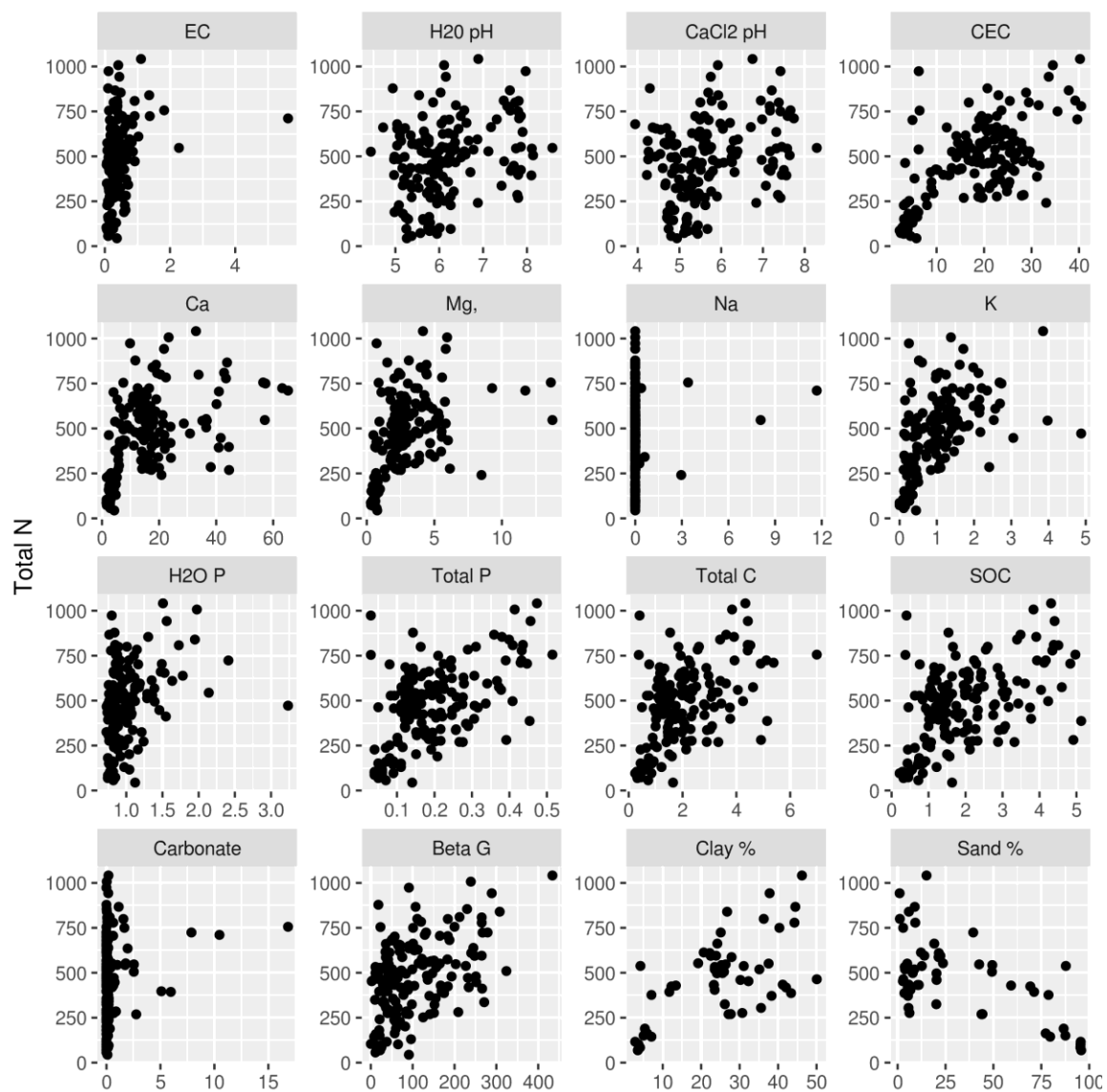


Figure C.2 Scatter plot matrix of Total N relationship with the remaining measured soil properties.

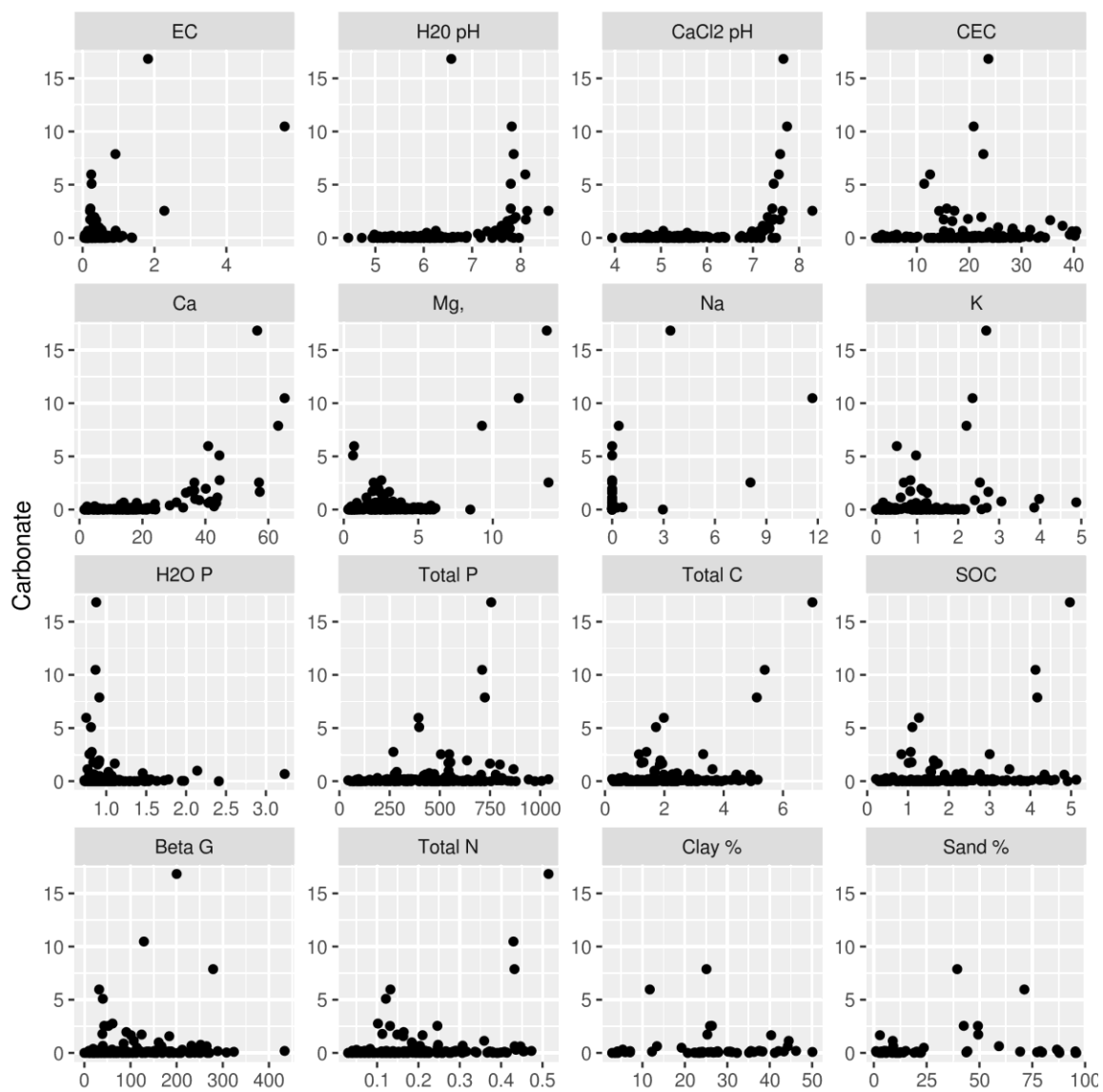


Figure C.3 Scatter plot matrix of  $\text{CaCO}_3$  relationship with the remaining measured soil properties.

## APPENDIX D: MODEL SCATTER PLOTS

Scatterplots are shown below for VNIR and MIR models with the best predictive ability within each calibration sampling scheme by measured soil property.



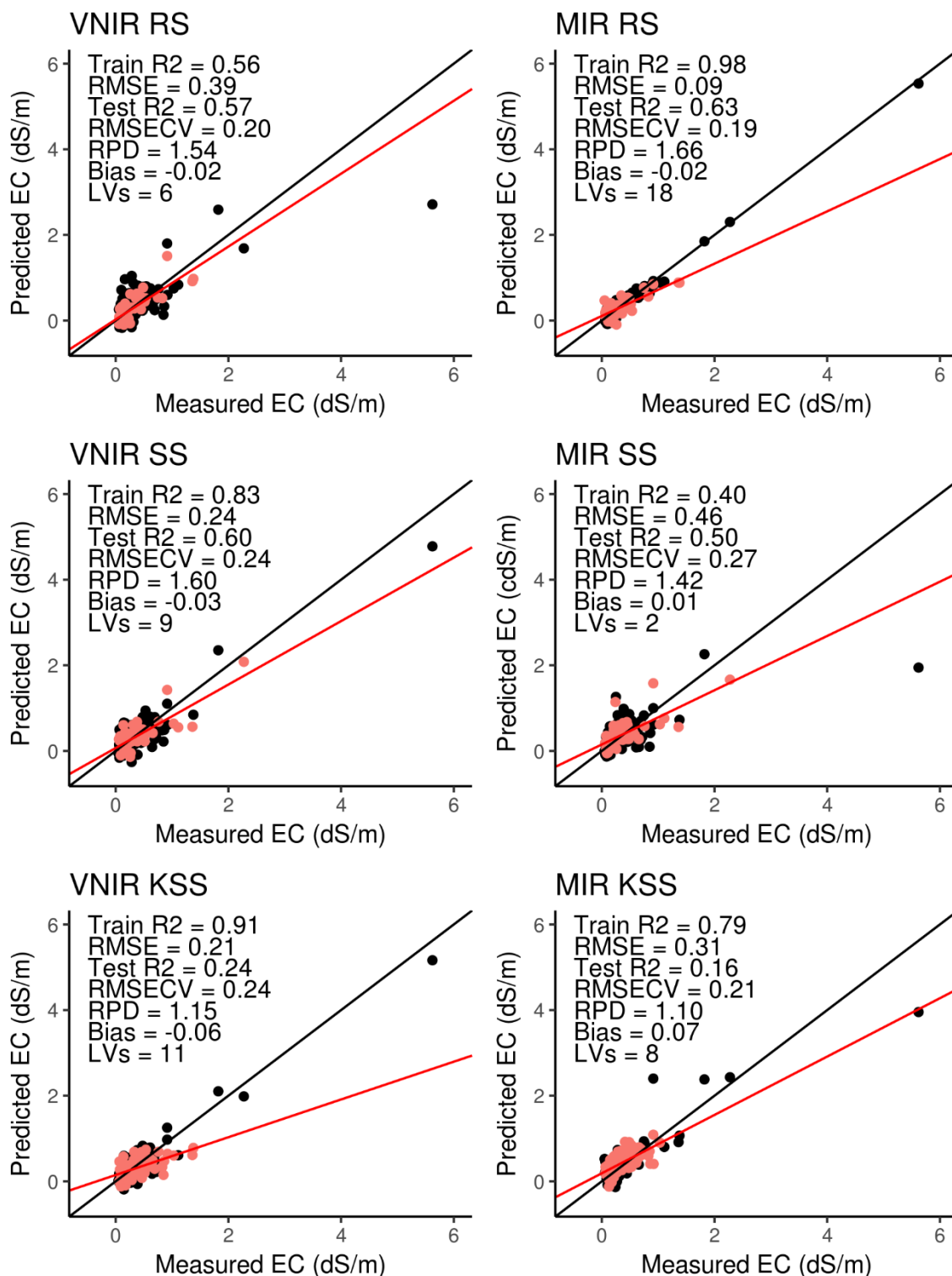


Figure 4.4. Goodness of fit plots showing relationship of PLS predicted EC values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

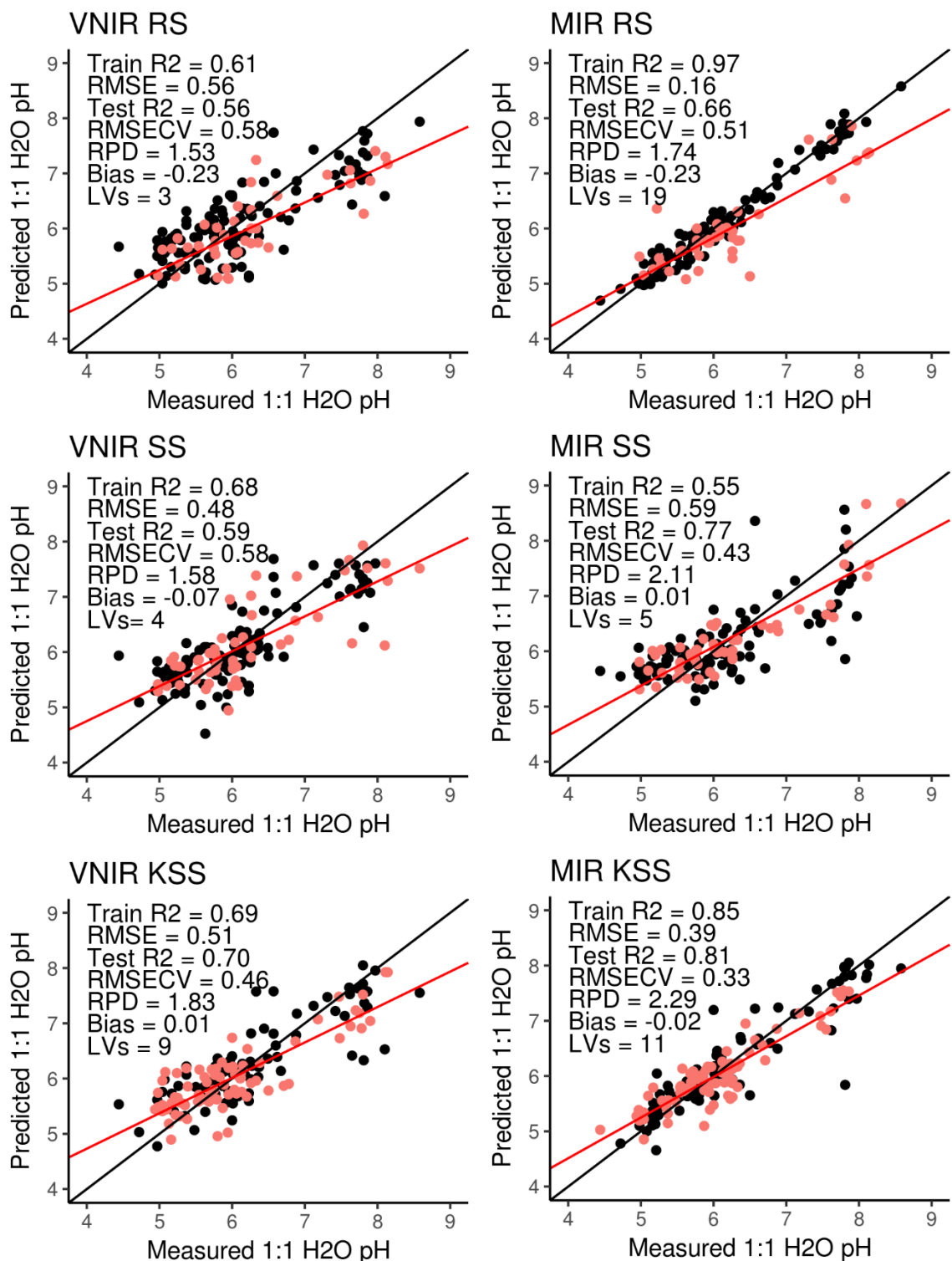


Figure 4.5. Goodness of fit plots showing relationship of PLS predicted 1:1 H<sub>2</sub>O pH values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R<sup>2</sup> (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

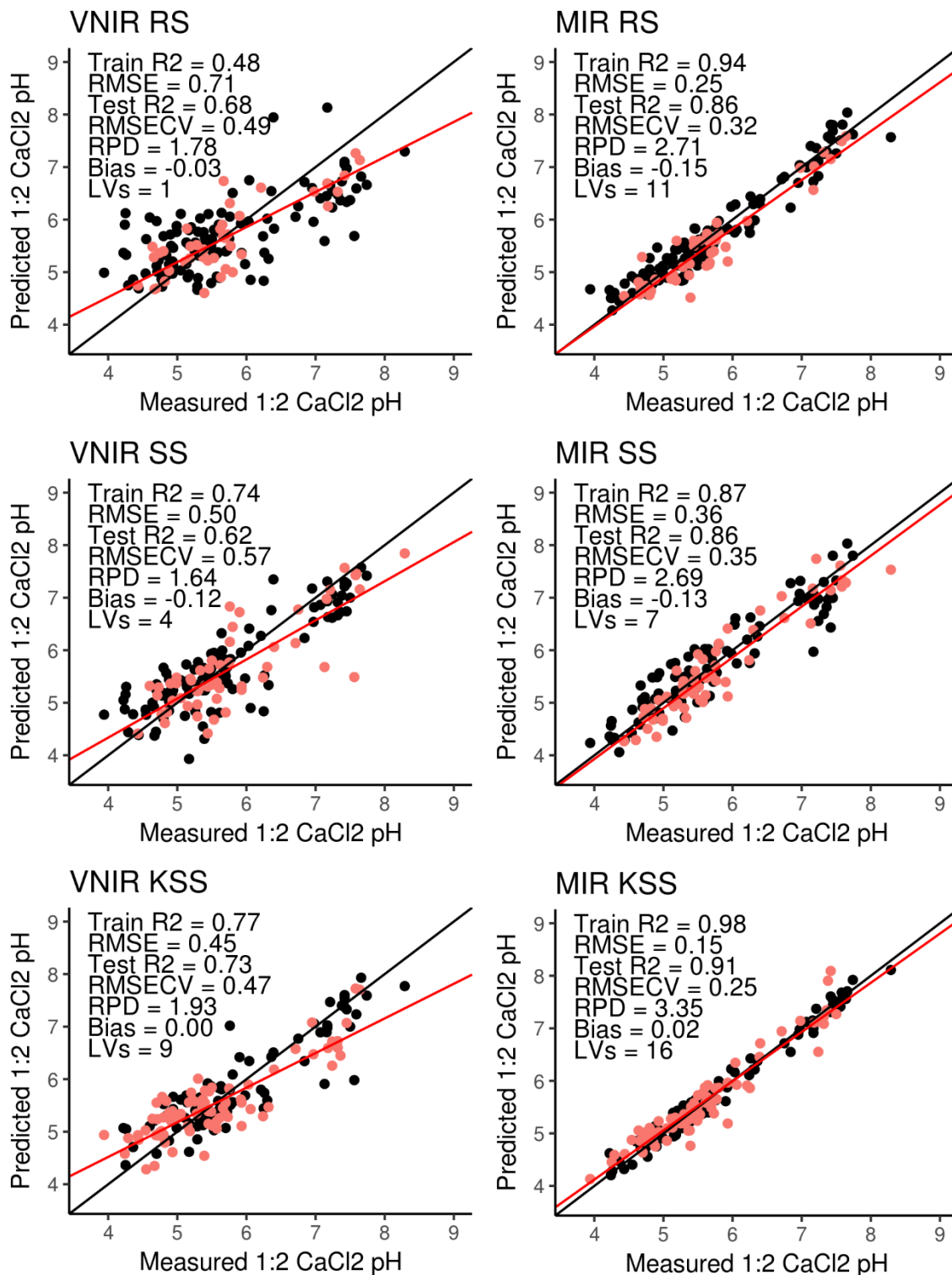


Figure 4.6. Goodness of fit plots showing relationship of PLS predicted 1:2 CaCl<sub>2</sub> pH values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R<sup>2</sup> (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

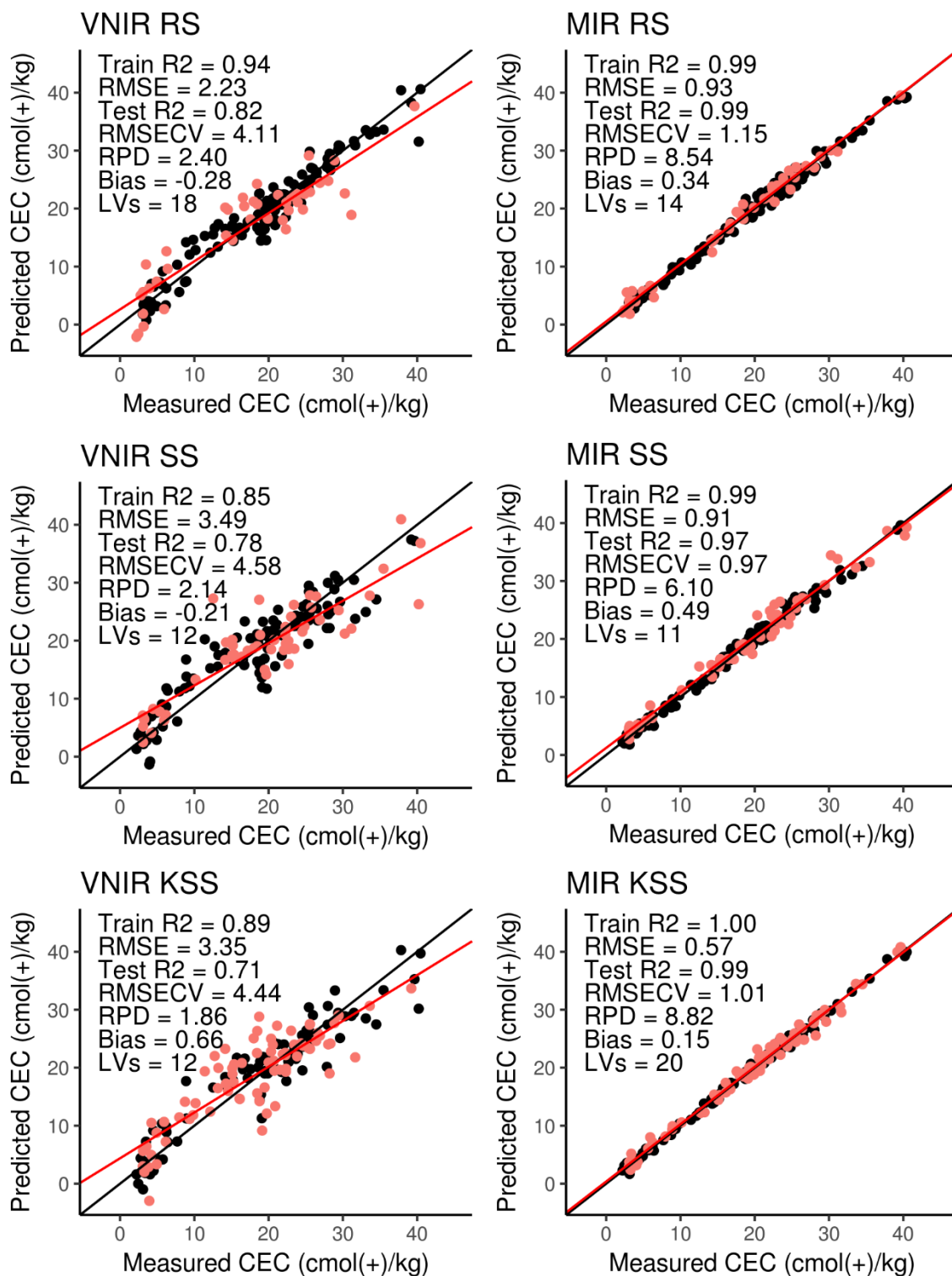


Figure 4.7. Goodness of fit plots showing relationship of PLS predicted CEC values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

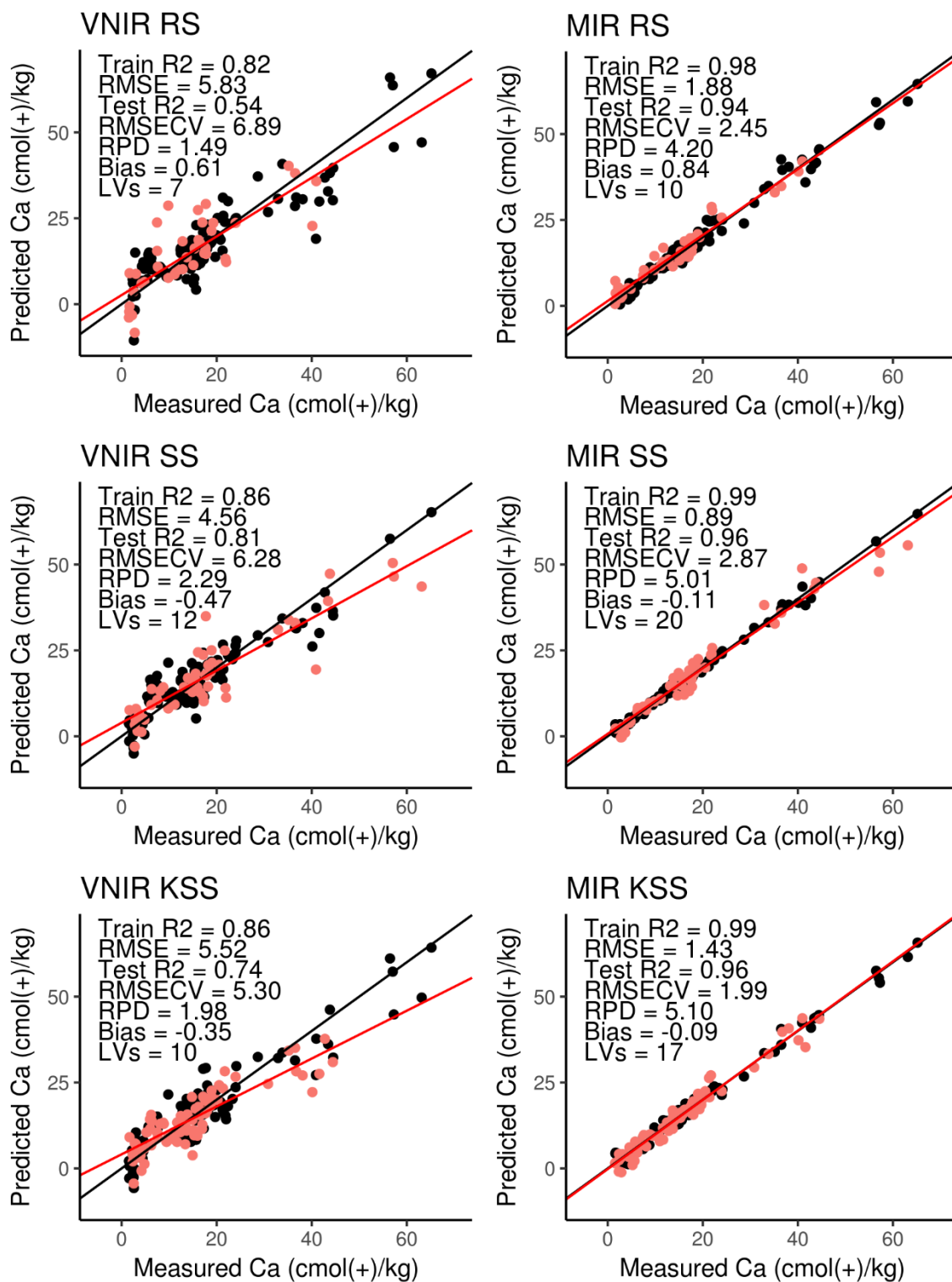


Figure 4.8. Goodness of fit plots showing relationship of PLS predicted  $\text{Ca}^{2+}$  values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

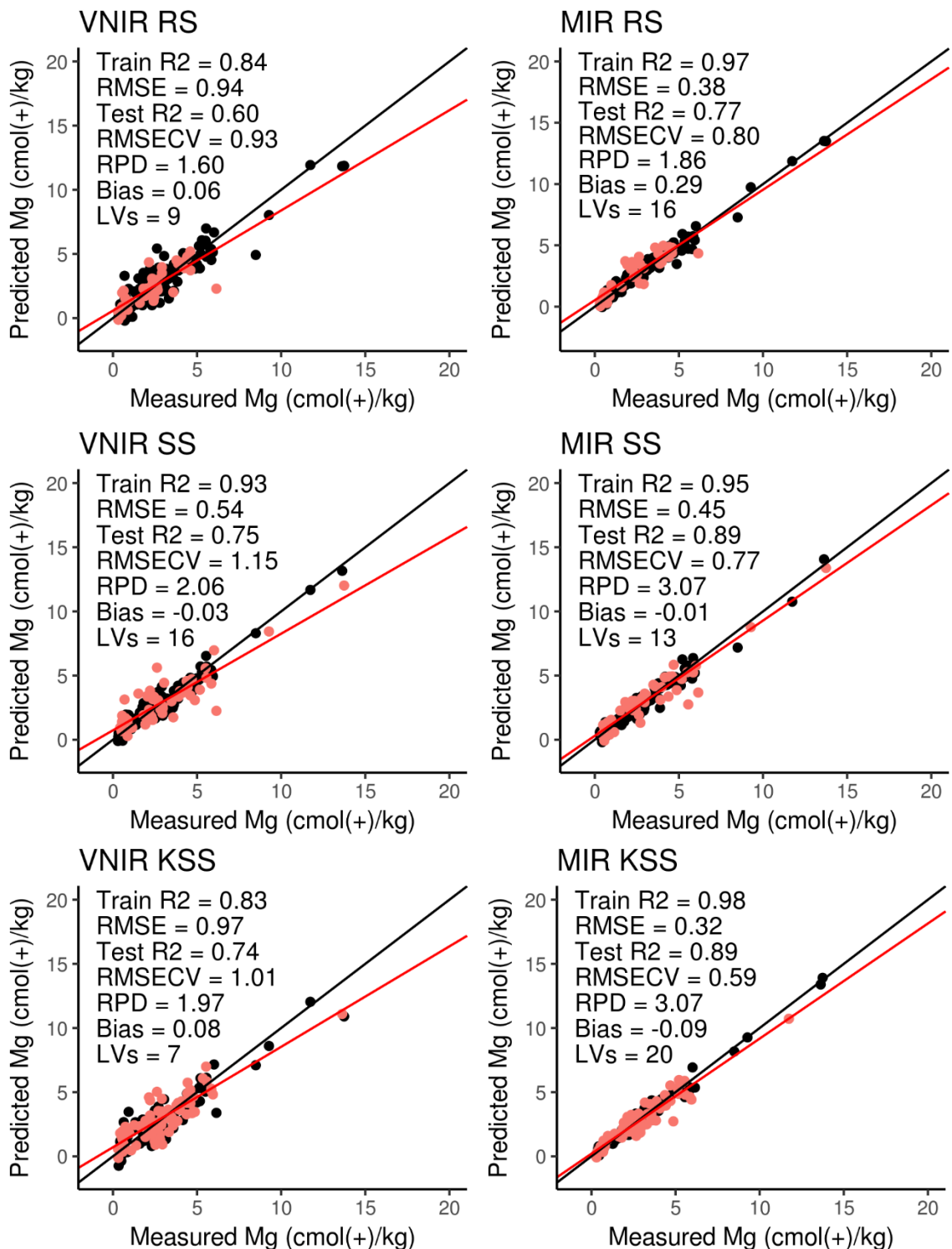


Figure 4.9. Goodness of fit plots showing relationship of PLS predicted  $Mg^{2+}$  values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

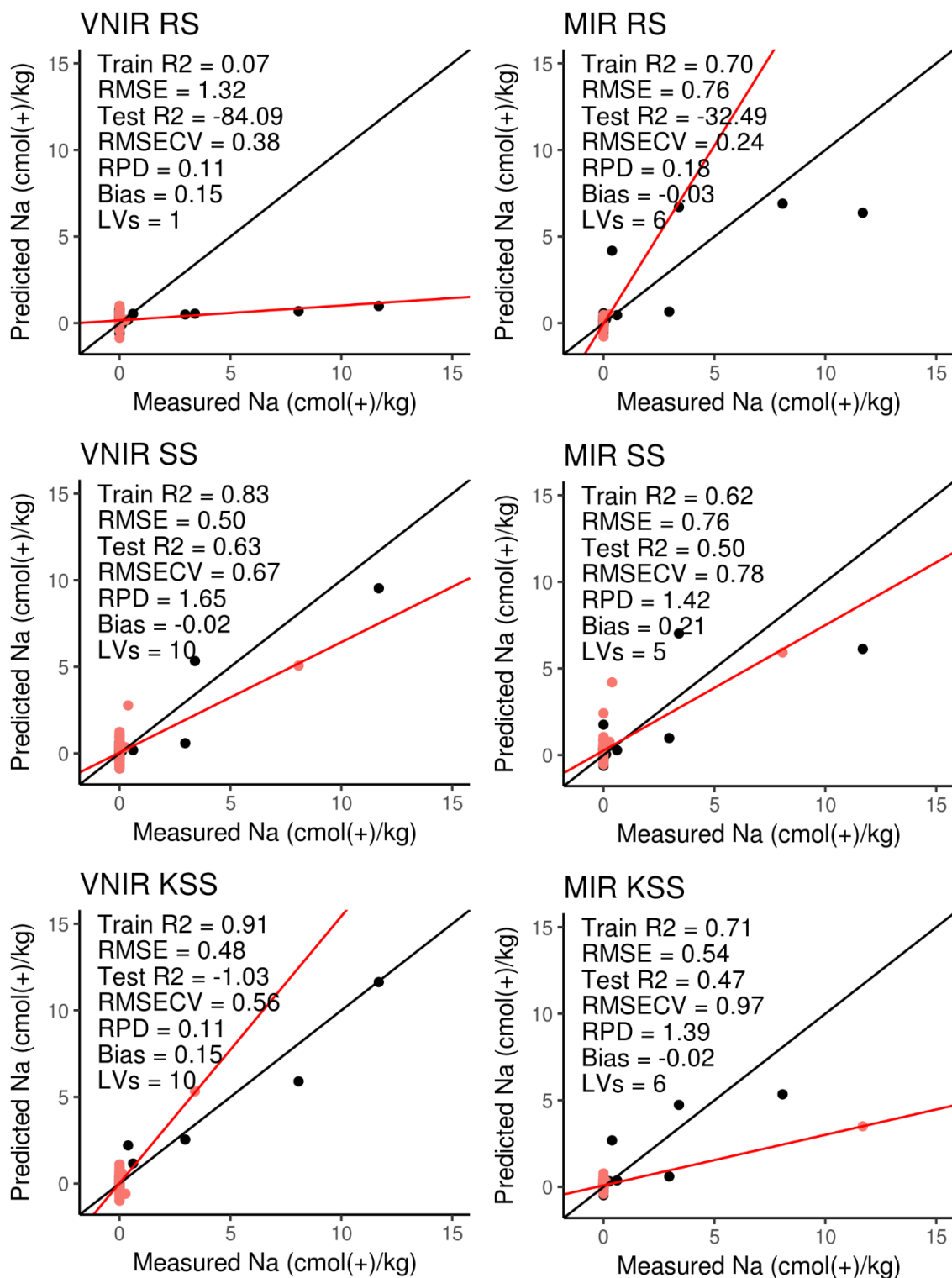


Figure 4.10. Goodness of fit plots showing relationship of PLS predicted  $\text{Na}^+$  values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

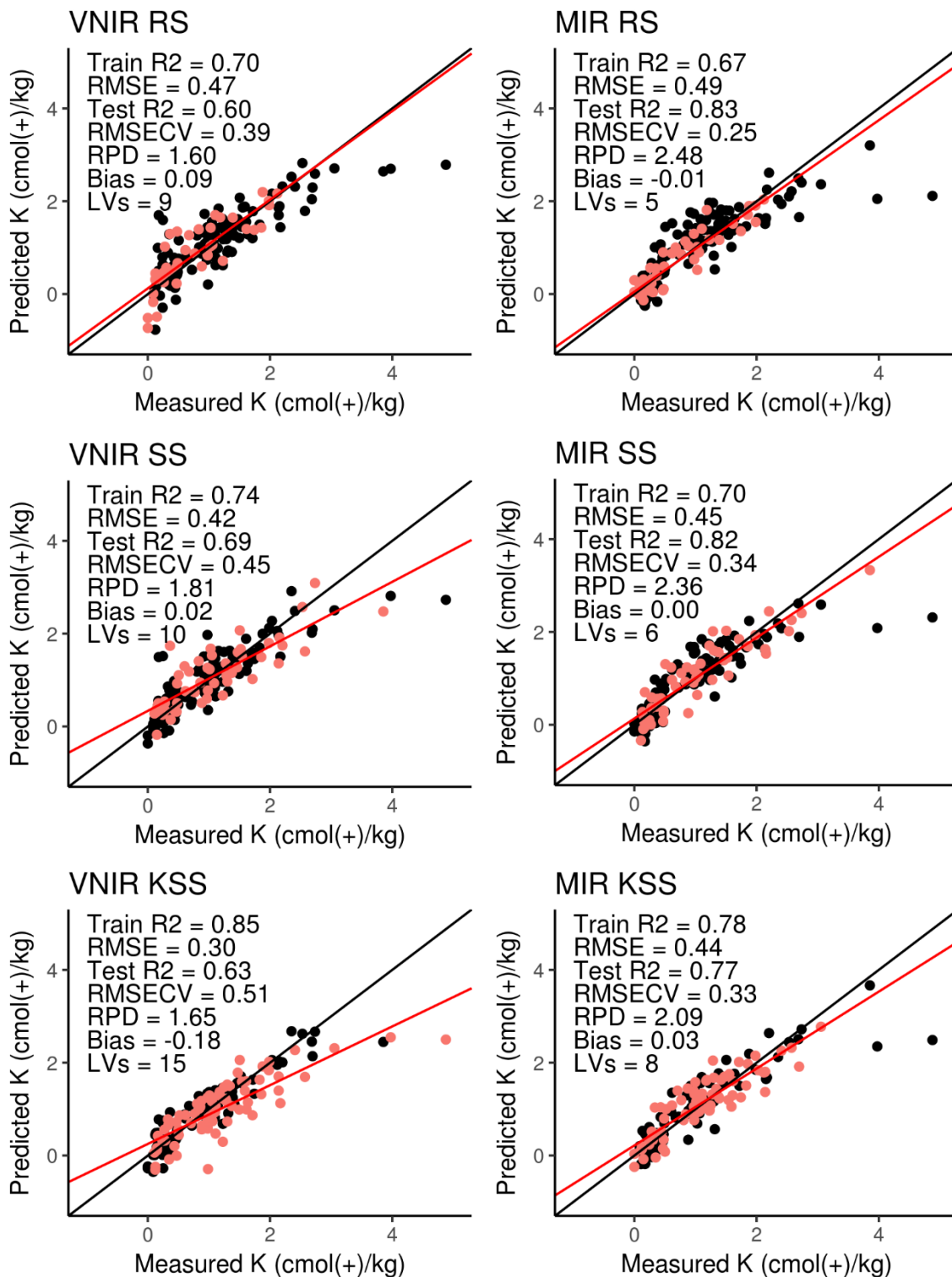


Figure 4.11. Goodness of fit plots showing relationship of PLS predicted  $K^+$  values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).



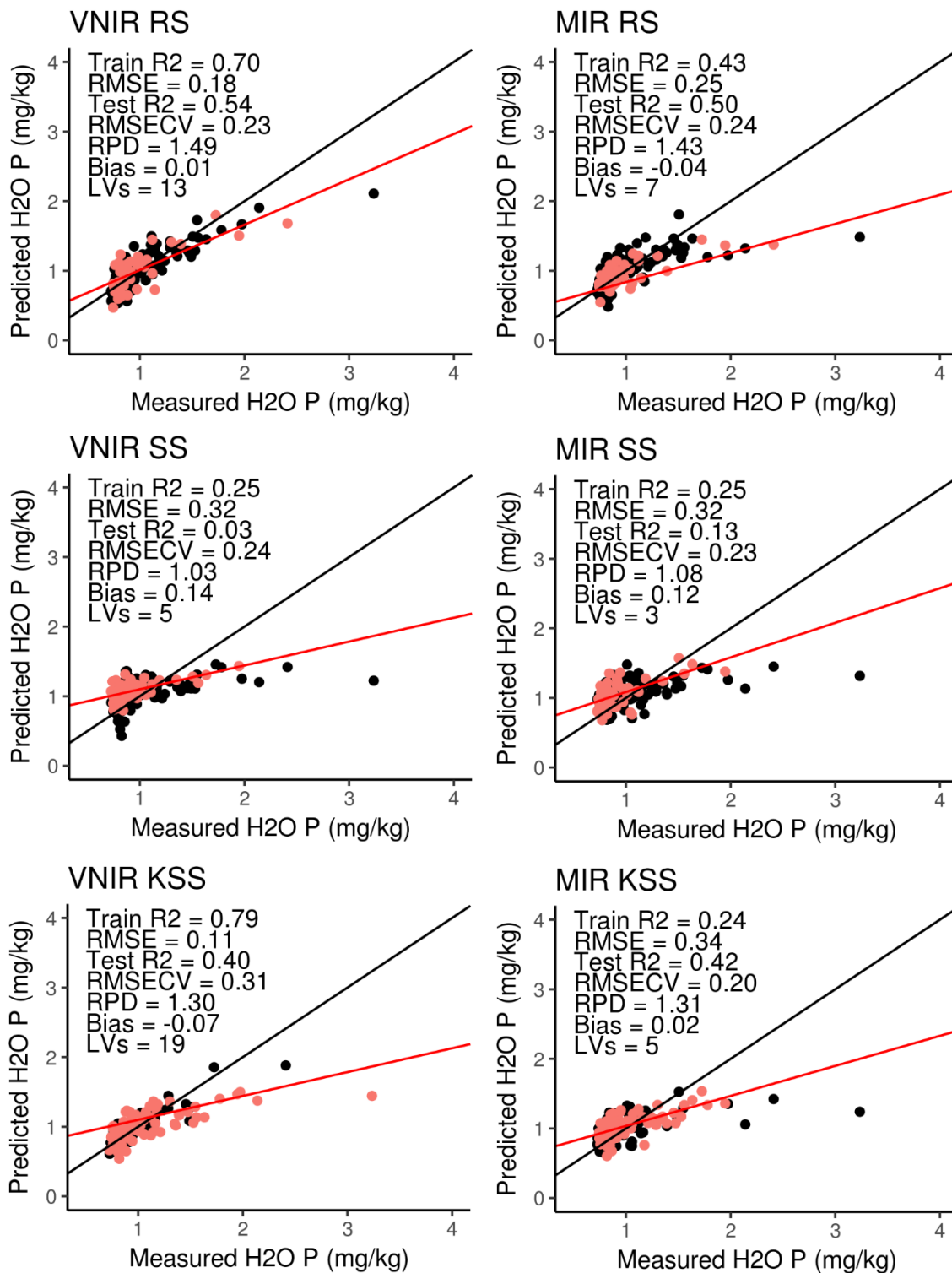


Figure 4.12. Goodness of fit plots showing relationship of PLS predicted H<sub>2</sub>O P values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R<sup>2</sup> (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

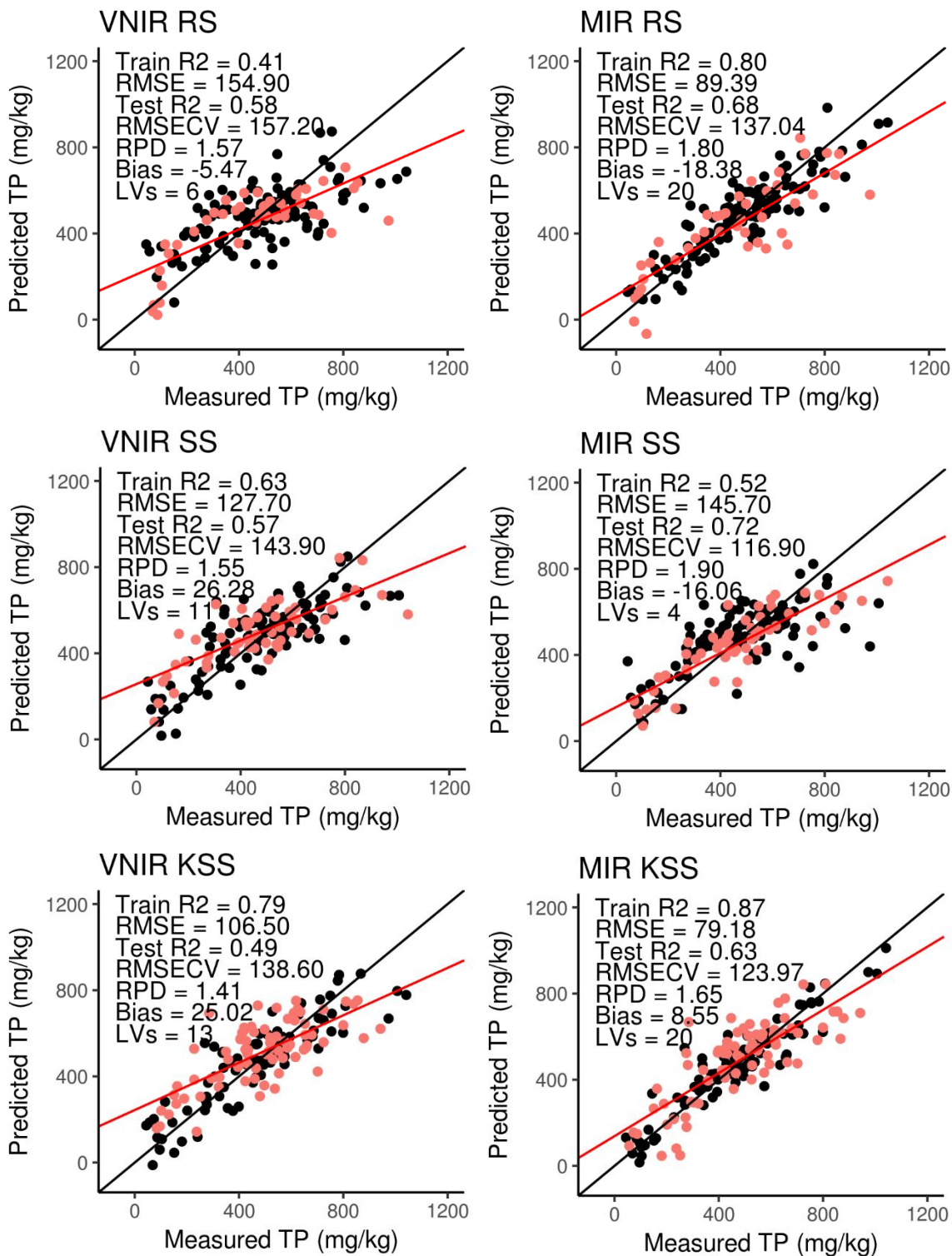


Figure 4.13. Goodness of fit plots showing relationship of PLS predicted Total P values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R<sup>2</sup> (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

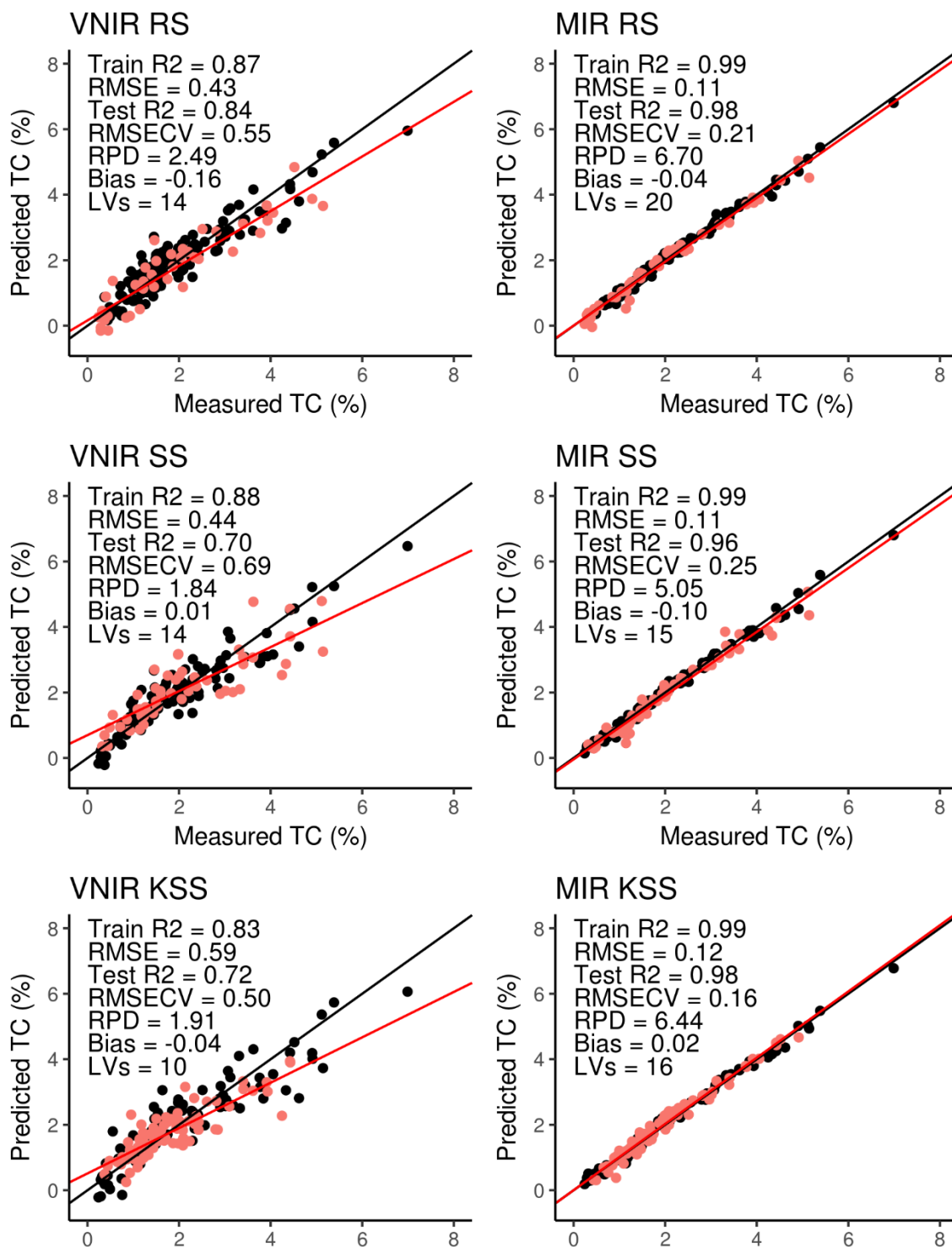


Figure 4.14. Goodness of fit plots showing relationship of PLS predicted Total C values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

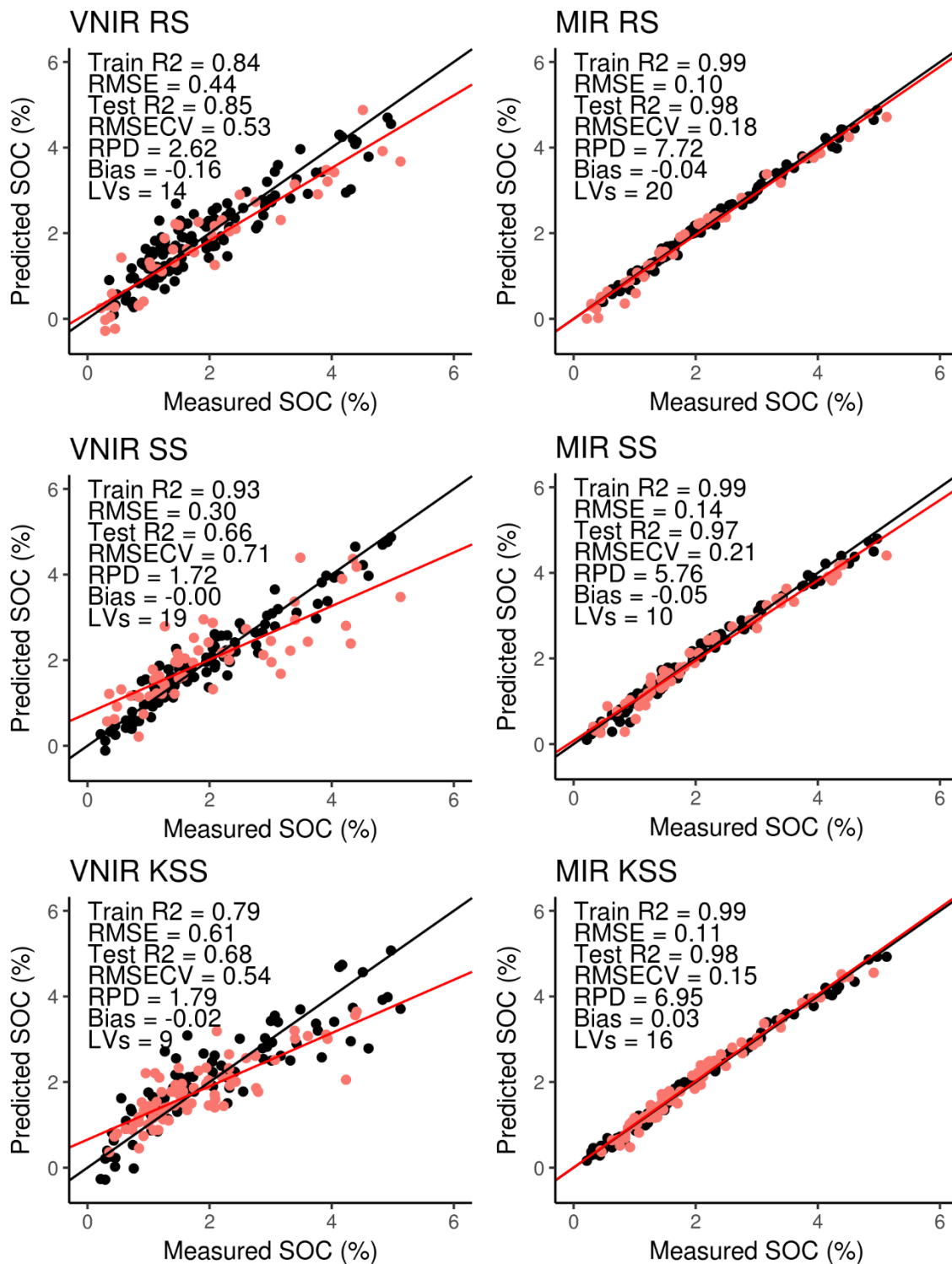


Figure 4.15. Goodness of fit plots showing relationship of PLS predicted SOC values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R<sup>2</sup> (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

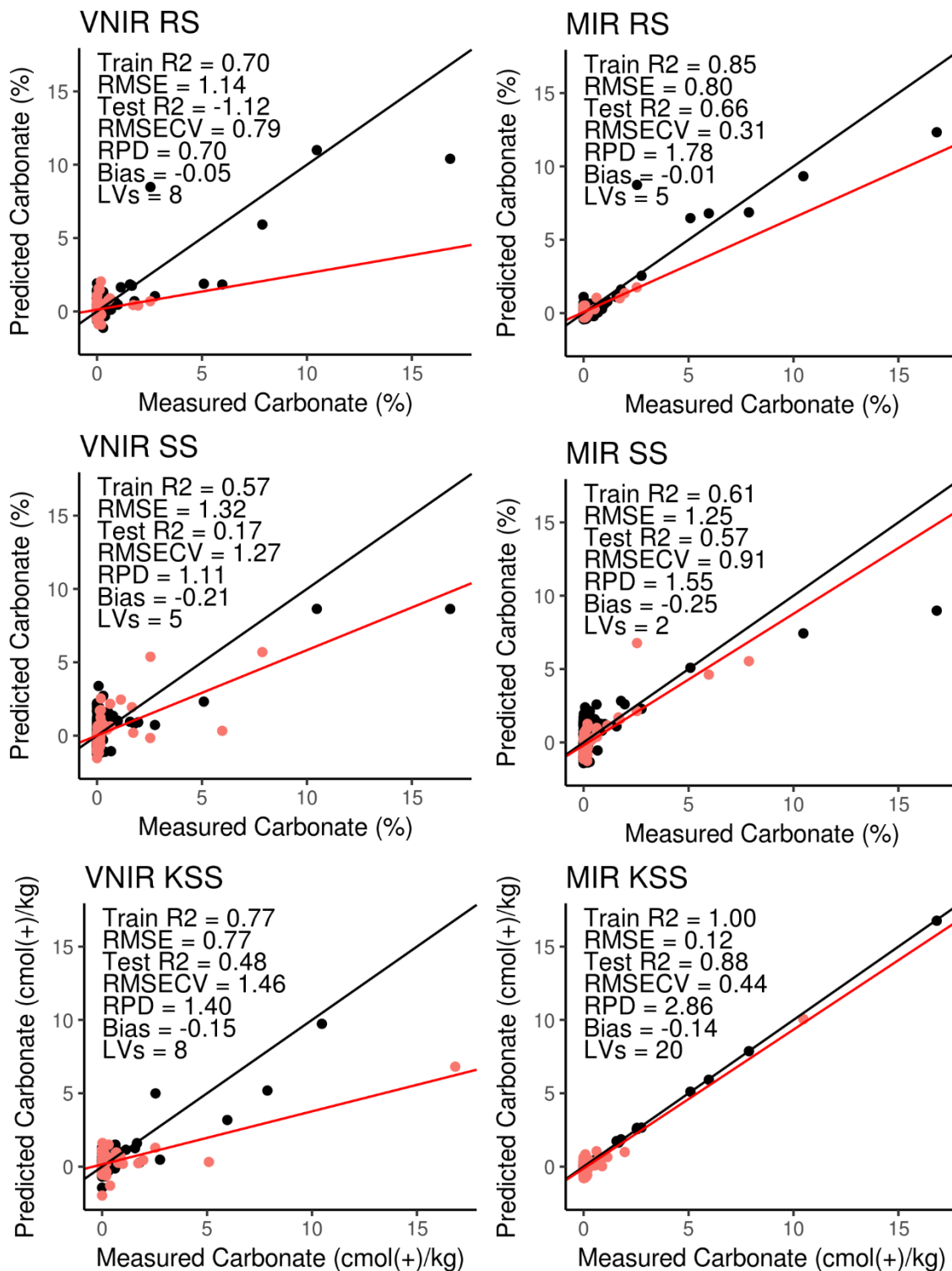


Figure 4.16. Goodness of fit plots showing relationship of PLS predicted  $\text{CaCO}_3$  values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

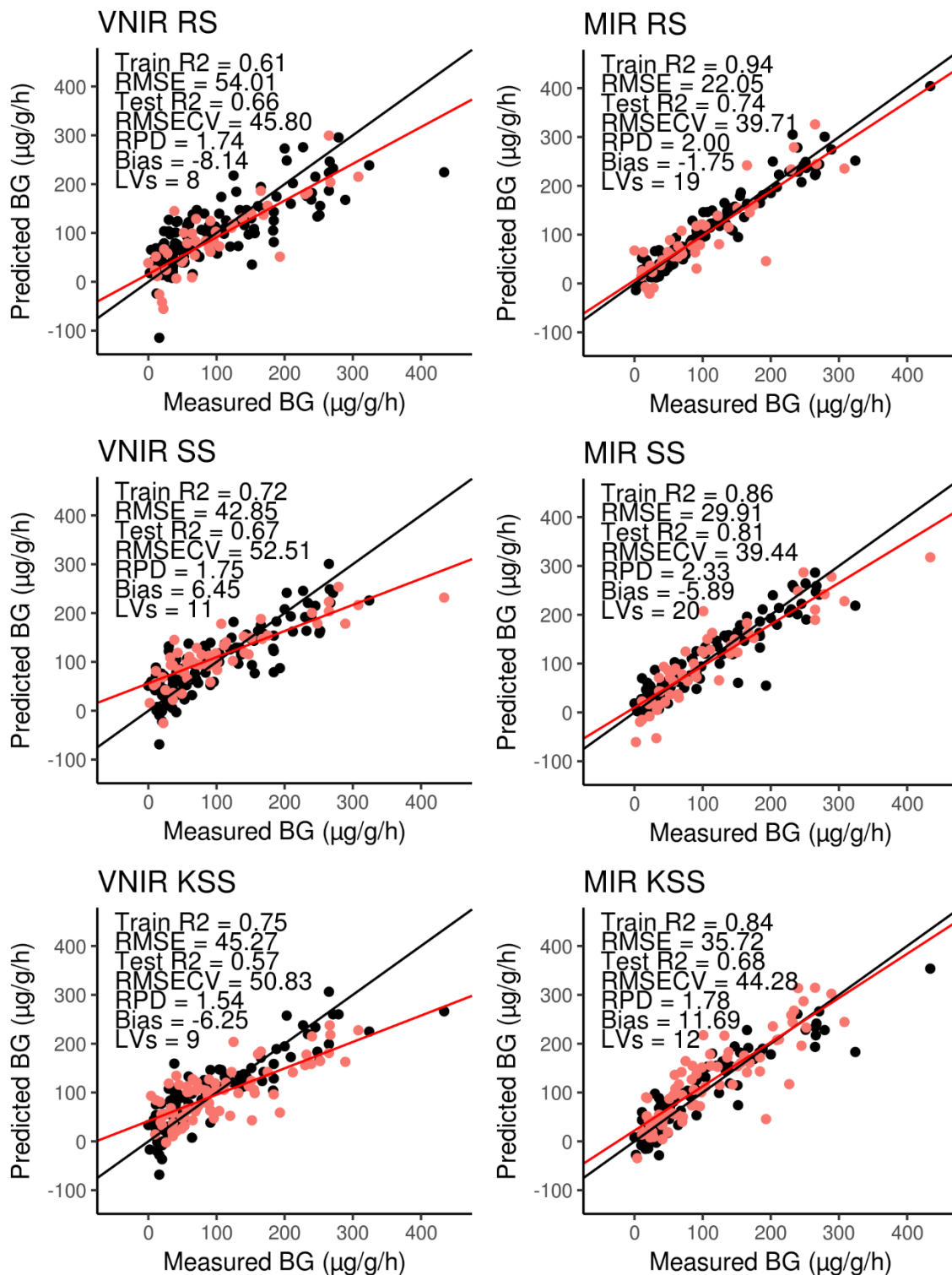


Figure 4.17. Goodness of fit plots showing relationship of PLS predicted  $\beta\text{G}$  activity from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red),  $R^2$  (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).

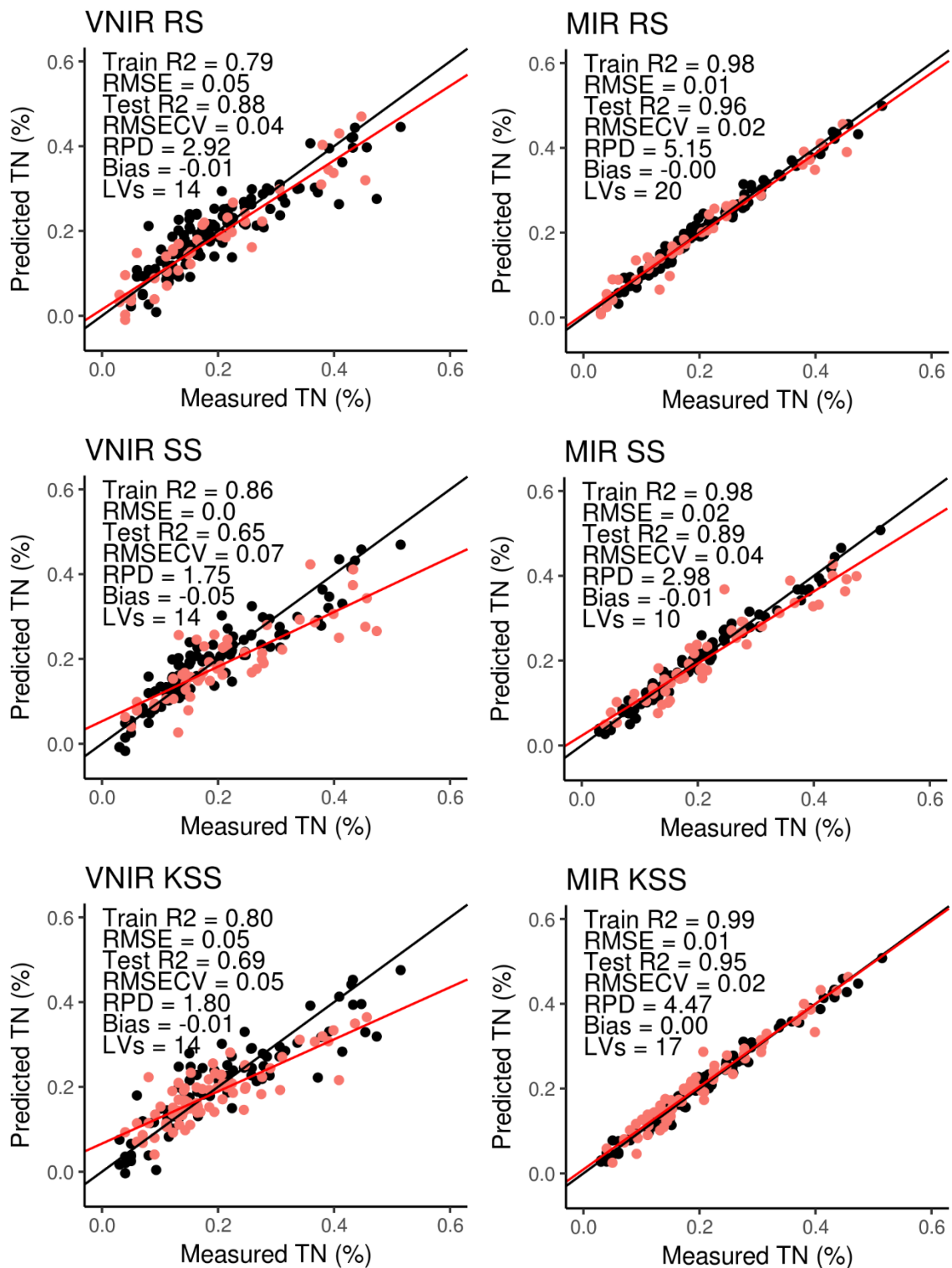


Figure 4.18. Goodness of fit plots showing relationship of PLS predicted Total N values from VNIR and MIR versus traditional laboratory methods. Showing 1:1 line (black), regression slope (red), training set (black), test set (red), R2 (coefficient of determination), RMSE (root mean square error of training set), RMSECV (root mean square error of cross validation), RPD (ratio of performance to deviation), and LVs (number of PLS latent variables).