DIFFUSE REFLECTANCE SPECTROSCOPY FOR THE CHARACTERIZATION OF CALCAREOUS GLACIAL TILL SOILS FROM NORTH CENTRAL MONTANA

by

Genevieve Christine Steward

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Land Resources and Environmental Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

August 2006

APPROVAL

of a thesis submitted by

Genevieve Christine Steward

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Division of Graduate Education.

Dr. David Brown

Approved for the Department of Land Resources and Environmental Sciences

Dr. Jon Wraith

Approved for the Division of Graduate Education

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Only the copyright holder may grant requests for permission for extended quotation from or reproduction of this thesis in whole or in parts.

Genevieve Christine Steward

August 25, 2006

ACKNOWLEDGEMENTS

I would like to thank my advisor, David Brown, for making this project possible and for pushing me to run that extra mile (or twenty . . .). I learned the most within these last few miles and for that I am appreciative. Thanks to my committee members, Jon Wraith and Rick Engel, for their feedback and constructive criticism. I am grateful to Ross Bricklemeyer for sharing his expertise of the study area as well as the soil samples for this project. I express my gratitude to the pit crew of the Landscape Pedology lab including Melisa Borino, Katie Tillerson, Kevin Dalton, and Austin Allen for their technical support. I would also like to acknowledge the Big Sky Carbon Sequestration Partnership as well as the National Research Initiative Competitive Grant for funding this project.

Most importantly, I know that I couldn't have finished the marathon without the constant support of my family; I appreciate the support of Mom, Dad, Bonnie, Serena, Julia and Becca Gee more than they will ever know. The last three years have been the most challenging of my life, and without these people I wouldn't have come close to the finish line. Thank you all.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT.)

TABLE OF CONTENTS (CONT.)

TABLE OF CONTENTS (CONT.)

LIST OF TABLES

LIST OF TABLES (CONT.)

Table Page

LIST OF FIGURES

LIST OF FIGURES (cont.)

Figure Page

LIST OF FIGURES (cont.)

Figure Page

ABSTRACT

Diffuse reflectance spectroscopy (DRS) is a method of soil carbon (C) quantification. In this study, the Vis-NIR (350 – 2500 nm) and MIR (2500-25000 nm) regions were evaluated to determine respective predictive accuracies of soil organic and inorganic carbon (SOC and SIC, respectively). The dataset included 315 soil samples of glacial till origin, obtained from six independent farm sites within the Golden Triangle region of Montana, with depths ranging from 0-100 cm.

For Vis-NIR analysis, Local vs. Regional vs. Global calibration sets were compared by six-fold cross validation by site of C predictions developed by Partial Least Squares (PLS) regression and Boosted Regression Trees (BRT). First derivative spectral data was used along with four preparation methods: (i) field moist and (ii) dry cores, (iii) 2-mm sieved ("Sieved") and (iv) milled samples (<200-µm, "Milled") were used to evaluate the potential application to *in-situ* analysis. The most accurate SOC predictions were from Milled samples using a Local calibration set. SOC predictions were a result of SOM electronic absorptions within the visible region. The most accurate SIC predictions were from Sieved samples with a Local calibration set.

Within the MIR analysis, spectral transformations, KBr dilution and Regional vs. Local calibrations were evaluated using the same independent validation procedure of PLS calibrations. Spectral transformations included: absorbance (AB), first derivative (D) and Kubelka-Munk (KM) from samples prepared both with and without KBr dilution. The most accurate SOC predictions were obtained from models developed with the D transformation of Local spectral data. SOC predictions were a result of SOM, clay and carbonate absorptions. The most accurate SIC prediction was from the KM transformation of Local spectral data. KBr diluted samples gave comparable regional predictive accuracy to neat samples. Specular reflection was found in neat sample spectral signatures; with Local core addition, neat samples built more accurate prediction models despite these distortions.

In the end, it was determined that the MIR region provides more accurate predictions of soil C for calcareous glacial till soils of north central Montana. But, the Vis-NIR region presents an accurate method while existing as the less expensive and more efficient route.

CHAPTER 1

INTRODUCTION

Diffuse reflectance spectroscopy (DRS) is used as a reliable, efficient and inexpensive means of characterizing soils (Dunn et al., 2002; McCarty et al., 2002; Brown et al., 2005). Traditional laboratory methods are often more time consuming, can require extensive training of personnel, produce waste, and leave the soil sample distorted and unusable for future analyses (Chang et al., 2002; Dunn et al., 2002). With the rise of precision (site-specific) agriculture, characterization for more soil samples from a limited area is necessary and conventional laboratory analysis is too costly (Dunn et al., 2002). As well, soil carbon (C) data is particularly important for future global C research as soils are looked to as a possible atmospheric C sink or source (Brown et al., 2005).

Diffuse reflectance spectroscopy is used for material analysis using the visible to near infrared (Vis-NIR, 400-2500 nm) and mid infrared (MIR, 2500-25000 nm) regions. Recent studies focus on the prediction of a suite of soil characteristics that are normally determined individually with traditional laboratory methods. There is extensive research in the area of Vis-NIR as well as MIR DRS predictive soil analyses (Ben-Dor & Banin, 1995; Dunn et al., 2002; Shepherd & Walsh, 2002; Chang & Laird, 2002; McCarty et al., 2002; Islam et al., 2003; Brown et al., 2005; Viscarra Rossel et al., 2006). What is lacking in the research is adequate validation of the prediction models developed; adequate validation involves applying the prediction model to a dataset independent of the dataset used to calibrate the prediction model. Many of these studies have been

validated using a random selection procedure that is not always guaranteed independent (Brown et al., 2005). Even more studies have been published solely reporting calibration results that are inapplicable to outside data (unless independently validated).

The predictive accuracies of the MIR and Vis-NIR regions have been evaluated; the question as to which region provides more accurate predictions is still left unanswered. Several studies have investigated this problem (McCarty et al., 2002; Reeves III et al., 2002; Viscarra Rossel et al., 2006) but again, "independent" validation procedures have left much to be desired. For future soil analysis needs, it is important to know which region can most efficiently provide accurate information.

Soil sample preparation is an essential topic of concern that could pave the way toward more accurate DRS predictions and/or more efficient analysis procedures. The Vis-NIR spectrometer has been developed as a piece of equipment safe and convenient to use *in-situ* yet there have only been two studies reporting the accuracy of this data (Sudduth and Hummel, 1992; Waiser et al., in review) – Sudduth and Hummel reporting on SOM analysis and Waiser et al. reporting on clay quantification. In MIR research, the issue of neat sample (as is) analysis vs. KBr-diluted sample analysis is still up for debate. Specular reflection in the MIR results in the distortion of prominent spectral absorptions unless samples are diluted in a non-absorbing matrix (Olinger and Griffiths, 1993). But, neat sample analysis was determined qualitatively comparable to diluted sample analysis (Nguyen et al., 1991), therefore MIR quantitative analyses moved directly to this method because of its simplicity. With the possibility of spectral distortion in the MIR region,

predictive accuracy as well as qualitative spectral results of neat vs. diluted samples should be examined further.

Organic matter (SOM) is a complex soil component that has been difficult to consistently characterize with spectroscopy. Organic constituents have been analyzed individually using transmission spectroscopy in the MIR (Gerasimowicz, Byler and Susi, 1986; Baes and Bloom, 1989; Celi et al, 1996), but because of the complexity surrounding these constituents, they have been difficult to analyze within soil in its entirety.

There are questions lingering relating to the predictive accuracy of organic C. First and foremost, are these predictions truly relating to SOM absorptions or are they derived from more distinct absorptions such as those from clay or iron oxides (Clark et al., 1999; Brown et al., 2005)? Organic matter quantification and characterization is important for many areas of soil research; the details pertaining to infrared analysis of SOM should be better understood for future applications.

As a result of these fundamental elements missing in the current research, the following project was organized around the focal question: which region, NIR or MIR, is the most efficient method of analysis for rapid and convenient predictions of soil C? To answer this question, several objectives were approached:

1) construct a robust and independently validated soil C prediction model using the most efficient and accurate calibration dataset available.
2) qualitatively examine soil spectra within different preparation methods.

3) investigate the effects of soil organic matter on soil spectral signatures as well as its relation to SOC predictions.

The subsequent thesis is organized into four additional chapters. Chapter two is a review of literature surrounding DRS, chapter three meets the preceding objectives utilizing the NIR region, chapter four meets the objectives utilizing the MIR region, and chapter five is a summary and conclusion of the project.

References

Baes, A.U. and Bloom P.R., 1989. Diffuse reflectance and transmission fourier transform infrared (DRIFT) spectroscopy of humic and fulvic acids. Soil Sci. Soc. Am. J. 53: 695-700.

Ben-Dor, E. and Banin, A., 1995. Near infrared analysis (NIRA) as a method to simultaneously evaluate spectral featureless constituents in soils. Soil Science, 159(4): 259-270.

Brown, D.J., Brickelmeyer, R.S., and Miller, P.R., 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C predictions in Montana. Geoderma, 129: 251-267.

Celi, L., Schnitzer, M., and Negre, M., 1997. Analysis of carboxyl groups in soil humic acids by a wet chemical method, FTIR spectrophotometry, and solution-state C-13 NMR: a comparitive study. Soil Science, 162 (3): 189-200.

Chang, C.W., Laird D.A., Mausbach M.J., and Hurburgh C.R., 2001. Near-infrared reflectance spectroscopy – principal components regression analyses of soil properties. Soil Sci. Soc. Am. J., 65: 480-490.

Chang, C.W. and Laird, D.A., 2002. Near-infrared reflectance spectroscopic analysis of soil C and N. Soil Science, 167(2): 110-116.

Clark, R.N., 1999. Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy. In: N. Rencz (Editor), Remote Sensing for the Earth Sciences: Manual of Remote Sensing, 3 ed. John Wiley & Sons, New York, pp. 3-52.

Dunn, B.W., Beecher, H.G., Batten, G.D., Ciavarella, S., 2002. The potential of near-infrared reflectance spectroscopy for soil analysis – a case study from the Riverine Plain of south-eastern Australian. Australian Journal of Experimental Agriculture, 42: 607-614.

Gerasimowicz, WV, Susi, and Byler, 1986. Resolution enhanced FTIR Spectra of soil constituents – humic acid. Applied Spectroscopy, 40(4): 504.

Islam, K., Singh, B., and McBratney, A., 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. Australian Journal of Soil Research, 41: 1101-1114.

McCarty, G.W., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Sci. Soc. Am. J, 66: 640-646.

Nguyen, T., Janik, L.J., Raupach, M., 1991. Diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy in soil studies. Aust. J. Soil Res, 29: 49-67.

Olinger , J.M. and Griffiths, P.R., 1993. Effects of sample dilution and particle-size morphology on diffuse reflectance spectra of carbohydrate systems in the NIR and MIR. Applied Spectroscopy, 47(6): 687-694.

Reeves III, J.B., McCarty, G.W. and Mimmo, T., 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. Environmental Pollution 116: S277-S284.

Shepherd, K.D. and Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Sci. Soc. Am. J., 66: 988-998.

Sudduth, K.A. and Hummel, J.W., 1993. Portable, near-infrared spectrophotometer for rapid soil analysis. Transactions of the ASAE, 36(1): 185-193.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., and Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma, 131: 59-75.

CHAPTER 2

LITERATURE REVIEW

Fundamentals of Diffuse Reflectance Spectroscopy (DRS)

Theory

Spectroscopy is the study of light (including the non-visible wavelengths) emitted, reflected, or scattered from a sample (Springsteen, 1998). Light can be measured in units of frequency or wavelength according to the equation

$$\lambda\nu = c \qquad (1)$$

where λ is wavelength (m), ν is frequency (Hz, s^{-1}) and c is the speed of light, $2.998 \cdot 10^{8}$ m s^{-1}. With c as a constant, frequency can also be described as the reciprocal of wavelength or wavenumber (ύ)

$$\lambda^{-1} = \acute{\upsilon} \qquad (2)$$

This unit of measurement is analogous to wavelength and is used particularly in the mid infrared region.

Light travels in "packets" of energy, or photons, and each photon carries a specific amount of energy related to the frequency or wavelength of the light according to

$$E = h\nu \qquad (3)$$

where E is energy (J), *h* is Planck's constant ($6.626 \cdot 10^{-34}$ J · s). This equation summarizes the qualitative impact of light – higher frequency results in higher energy and thus a lower wavelength gives higher energy (Umland and Bellama, 1996). This

relationship can be viewed in Fig. 2.1, a diagram of the electromagnetic spectrum. The regions focused on here will be the visible (Vis, 400-700 nm), the near infrared (NIR, 700-2500 nm) and the mid infrared (MIR, 2500-25000 nm).

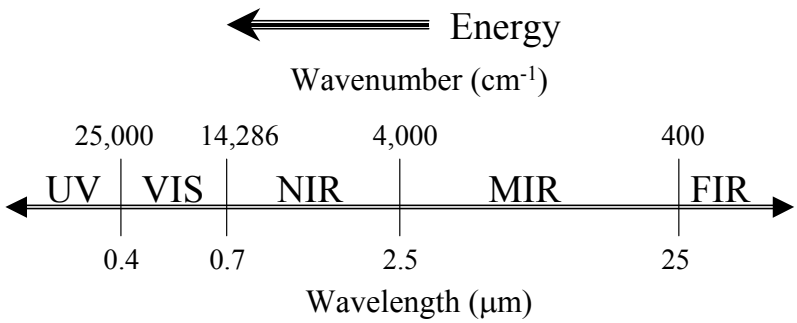

Figure 2.1. A diagram of the electromagnetic spectrum .

Solid materials have five reactions to light which can be qualified and quantified by spectroscopy: scattering, transmission, reflectance, diffraction and absorption (Workman, 1998). Transmission spectroscopy is based upon the relationship known as Beer's Law, a quantitative interpretation as to how photons are attenuated in relation to an intervening medium (Clark, 1999). When used in transmission spectroscopy, Beer's Law is stated as

$$I = I_o e^{-acl} \tag{4}$$

where I is the output intensity at a specific wavelength, I_o is the original intensity at the same wavelength, a is the wavelength and material specific absorptivity, c is the concentration of the analyte and l is the optical path length. This law only applies when

the analyte is diluted in a non-absorbing matrix such as KBr. The equation leads to a derivation of the linear relationship between absorbance and concentration (Duckworth, 1998)

$$A = \log(1/T) = acl \tag{5}$$

where A is absorbance and T is transmittance or I/I_o.

Similarly, the Kubelka-Munk function is a linear relation between analyte concentration and reflectance for diffuse reflectance spectroscopy -

$$f(R_\infty) = \frac{(1-R_\infty)^2}{2R_\infty} = \frac{k}{s} \tag{6}$$

where R_∞ is the reflectance after penetration and scattering throughout a non-absorbing matrix of infinite depth, and $k = \varepsilon C$ with ε, molar absorptivity, and C, concentration (Coates, 1998). This relationship only applies to samples diluted in a non-absorbing matrix.

Reflectance can be divided into three forms: specular, diffuse and diffuse specular. Diffuse reflectance is light that has been redirected over a range of angles after entering and passing through the sample (Stewart, 1998) and is a focus of spectral measurements. Specular reflectance is light that has been reflected as from a mirror (Stewart, 1998) and can distort spectral measurements. Stronger absorption features (fundamentals) are typically more affected by specular reflection than weaker absorptions (overtones and combinations) are. For this reason, the NIR region is generally not affected by specular reflectance interference while the MIR region is (Ollinger and Griffiths, 1993). Intense absorption features can even experience a complete band

inversion as a result of specular reflectance interference known as Restrahlen bands (Nguyen et al., 1991). Grinding materials into a fine powder and diluting in a non-absorbing matrix can reduce or eliminate specular reflectance (Coates, 1998). Ollinger and Griffiths (1993) compared KBr-diluted and undiluted samples of cellulose, sucrose and wheat starch in the NIR and MIR and found that specular reflectance in the MIR region does occur with neat sample preparation but does not with KBr-diluted samples.

Absorption

According to Clark (1999), there are two processes of absorption: vibrational and electronic. Vibrational absorption is a result of molecular bonds. Bonds between atoms of a molecule with a dipole moment vibrate at a specific frequency due to symmetric stretching, asymmetric stretching and bending (Umland and Bellama, 1986); for a molecule with N atoms, there are 3N-6 *fundamental* vibrations (referred to as $\acute{\upsilon}_1$, $\acute{\upsilon}_2$, $\acute{\upsilon}_3$, etc.). *Overtone* vibrations occur as a multiple of these fundamental vibrations (e.g. $2\acute{\upsilon}_1$, $3\acute{\upsilon}_1$, etc.) and *combination* vibrations occur as a result of the fundamental vibrations complementing each other (e.g. $\acute{\upsilon}_1 + \acute{\upsilon}_2$, $\acute{\upsilon}_2 + \acute{\upsilon}_3$, etc.) (Clark, 1999).

Fundamental vibrational absorptions occur in the MIR region, with overtones and combinations occurring at lower MIR wavelengths as well as in the NIR. The NIR region holds first overtones of C-H, O-H and N-H bond vibrations (Weyer and Lo, 2002) because of the close proximity of the relative fundamentals within the MIR region. For example, water fundamentals are found at 6269, 2738, and 2662 nm with overtones of the OH stretch at about 1400 nm and the combination of the H-O-H bend with the OH stretch

found near 1900 nm (Clark, 1999). Although fundamental absorptions are stronger than overtones and combinations, overtones are specific to each constituent and are more sensitive to changes in the chemical composition of the absorbing molecule leaving both regions (Wetzel, 1983).

Electronic absorption is a result of the individual energy states within the electron shell of an ion; in DRS analysis of soils, this type of absorption is related to metal ions within a mineral, particularly iron. As electrons change energy levels, the atoms absorb different wavelengths of light energy. Electronic absorptions are often the cause of color in minerals, particularly iron bearing minerals, and occur at lower wavelengths (higher energy states) within the NIR and Vis regions(Clark, 1999).

As individual molecules absorb and reflect light at different wavelengths, they will contribute to different peaks and troughs within a spectral signature. Hence, each mineral and soil constituent (e.g. kaolinite, hematite, carbonate, water, etc.) has its own particular spectral reflectance signature. Soils are a heterogeneous mixture of these constituents and therefore these signatures combine to form a complex and unique spectral signature of each individual soil.

Pertinent Mineralogical and Molecular Absorptions

The ultimate objective of spectroscopic analysis of soils has been to relate the properties analyzed by traditional laboratory procedures to the spectral features of the soil determined through spectroscopy. These properties include clay, sand, organic matter and water content (among many other characteristics) which all influence the amount of

energy that is reflected or absorbed. Exact locations of soil constituent absorptions have been determined for both the NIR and MIR regions. The following are key features relative to this study published in the literature thus far.

MIR

In the MIR region, fundamental and overtone absorptions of soil constituents have been established (Clark et al., 1990; Nguyen et al., 1991). The major influences to soil spectra in the MIR are clay and quartz absorptions which can seriously interfere with organic absorptions (Skjemstad et al., 1998) as well as carbonate absorptions (Nguyen et al., 1991). Hydroxyl stretching occurs at 3755 and 3651 cm^{-1}, both appearing with the presence of clay, and water bending occurs at 1595 cm^{-1} (Clark et al., 1990). Quartz absorption features occur at: 2000, 1870, 1790, 1080, 800-780, and 700 cm^{-1} (Nguyen et al., 1991). Carbonates show several peaks; the fundamental absorptions are at 1063, 879, 680 and cm^{-1} with overtones at 4000, 2800 and 2500 cm^{-1} (Clark, 1999).

Organics absorptions have also been identified in the MIR region. Gerasimowicz et al. (1986) used transmission analysis of humic acid to assign qualitative values to absorption peaks at 3750-3570 cm^{-1} (OH stretch), 3000-2800 cm^{-1} (aliphatic CH stretch), 1720-1500 cm^{-1} (carbonyl and amide stretching), 1480-1350 cm^{-1} (CH bend), 1300-1000 cm^{-1} (CO stretching and COH bending). Baes and Bloom (1989) used transmission and diffuse reflectance spectroscopy of humic and fulvic acids to assign the following bands: OH absorbance features at 3400 cm^{-1} for fulvic acid and 3300 cm^{-1} for humic acid (it was suggested that the shift in fulvic acid spectra is due to H-bonding), CH stretching

vibration absorbance at 3030 cm^{-1} for both, OH stretching of COOH at 2600 cm^{-1}, CO stretching vibration of COOH as a strong band at 1720 cm^{-1}, aromatic CC and amide NH stretching at 1610-1620 cm^{-1}, and C-OH stretch at 1270 cm^{-1}. Most of these absorption features are located within or close to absorptions of the aforementioned constituents making it difficult to definitively determine SOM absorptions in the MIR region.

Vis-NIR

Soil spectra in the Vis-NIR region are most influenced by absorption bands of iron oxides (870 and 1000 nm), overtones of water (1200, 1450, 1770 and 1950 nm) and hydroxyl stretching (1400 nm) and a metal-OH stretch combination (2200-2300 nm) (Baumgardner et al., 1985). Clark et al. (1990) established the prominent Vis-NIR soil spectra as that which shows water absorptions at approximately 1400 and 1900 nm as well as the metal-OH combination at 2200-2300. Carbonaceous soils show overtone features at 2500 and 2300-2350 nm and weaker combinations near 2120-2160, 1970-2000 and 1850-1870 nm (Clark, 1999).

Organics have been more difficult to characterize in the Vis-NIR region as it only holds overtones of CH, OH and NH fundamental absorptions found at low wavelengths of the MIR region. Workman (1998) assigned absorptions in the NIR to organics including 1700-1900 and 1100-1300 nm to CH first and second overtones, respectively, and 1450-1550 nm as the NH first overtone. Chang et al. (2002) and Workman (1998) both assigned a distinct stretch from 2300-2350 nm to humic acid. Most of the organic

absorption features would likely be overlapped by stronger absorptions of water and carbonates, explaining the difficulty in interpretation of Vis-NIR soil spectra.

Quantitative Spectroscopy

Chemometrics is the process through which we obtain chemical information about an "unknown" material using chemical data from a reference material. In spectroscopy, the process refers to utilizing spectral data to predict related chemical data (Duckworth, 1998). Regression is used as a prediction method, though one of the key problems in spectroscopy is the set of highly correlated predictor variables (spectral data). Principal component regression (PCR) is used with the objective of reducing highly correlated data into a smaller set of uncorrelated orthogonal vectors to use as a reduced set of predictor variables (Wold et al., 2001) and was one of the first methods to include all spectral data within the prediction (Chang et al., 2001).

Partial least squares (PLS) regression is one of the more effective and commonly used prediction methods in spectroscopy. The basis of this technique is decomposition of the spectral data into a set of eigenvectors and scores while using the concentration data from a set of "known values" (Duckworth, 1998). The method is similar to PCR, except that both X and Y variables are used to build vectors with the greatest predictive power, correlating new X variables with provided Y data (Wold et al., 2001). A calibration model is built using new uncorrelated and reduced X variables and known Y variables, and is then validated through application to a new set of data with properties falling within the range of the calibration set (McCarty et al., 2002).

Regression tree analysis, an alternative to traditional regression methods, was initially introduced into the social sciences in the 1960's (Sain and Carmack, 2002) but has only recently been applied to the field of spectroscopy (Goodacre et al., 2001; Kaihara et al., 2001; Johnson et al., 2004). Regression trees are alternatives to traditional regression methodologies that build on complex interactions between high-order data (Sain and Carmack, 2002). Boosting algorithms have been introduced to regression trees to weight predictors; throughout the model building process, weaker predictors are weighted more heavily than stronger predictors. In the end, heavier weights are applied to the most important variables. Friedman (2001) has published an outline of boosting algorithms and theory. Boosting trees are an additive approach, taking many small simple models to build a more complex model. It requires a large dataset but is a quick process. Random subsets are taken of the model to determine the best possible fit (termed "bagging"), thereby making overfitting unlikely (Breiman, 1996).

The Treenet ® package (Salford Systems, San Diego, CA) includes a stochastic gradient boosting program as does the R freeware statistical package (R Development Core Team, 2004). Gradient boosting is a statistical approach that optimizes the standard boosting reweighting procedure (Friedman, 2001). This predictive approach is insensitive to input errors because of the weighting procedure, requires no variable transformations, handles large amounts of data in a relatively short period of time. It is a new application to the world of spectroscopy that has thus far only been applied within global spectroscopic analysis of soils (Brown et al., 2006). Additional applications might lead to a successful means of prediction pending future research.

Empirical Results

Near infrared prediction models began with multiple regression, using only three absorption wavelengths as predictor variables and the properties obtained in the laboratory as the response variable (Dalal and Henry, 1986). Since then, different statistical analysis methods have been used including PLS regression (Rumpel et al, 2001; Dunn et al, 2002; Reeves and McCarty, 2002; Chang and Laird, 2002), PCR (Chang et al, 2001; Islam et al., 2003) and multivariate adaptive regression splines (Shepherd and Walsh, 2002) leading to the ability to use up to hundreds of spectral characteristics. Predictive applications in the MIR used multivariate analyses as well, including PCR and PLS (Janik et al, 1995 and 1998; Rumpel et al, 2001; Reeves et al, 2002; Bertrand et al, 2002; McCarty et al, 2002; Siebielec et al, 2004; Madari et al, 2005; Viscarra Rossel et al., 2006). These studies showed success in calibration models built around carbon, nitrogen, clay, moisture and metal contents, and other soil attributes.

Many C prediction models using near and mid infrared spectral data have been built, but most lack independent validation. The validation stage in infrared quantitative analysis is extremely critical (Ben-Dor and Banin, 1995), as is validation with an independent data set (Brown et al., 2005) and this is where current research falls short. In development of a prediction model, it is necessary to calibrate the model using the data collected followed by validation with samples excluded from the calibration procedure. It is then possible to determine how applicable the model would be outside of the immediate dataset, which is a goal behind exploration of spectral prediction models. Random selection or leave-one-out cross validation techniques do not guarantee sample

independence in most cases (Brown et al., 2005), and are therefore inappropriate methods of model validation. There have been an abundance of calibration models developed (Dalal and Henry, 1986; Morra et al, 1991; Ben-Dor and Banin, 1995; Chang et al, 2001; Chang and Laird, 2002; Dunn et al., 2002; Shepherd and Walsh, 2002; McCarty et al., 2002; Islam et al., 2003; Brown et al., 2005), but only a few studies have conducted independent validation of reported calibrations.

Some studies that include independent validation results (Ben-Dor and Banin, 1995; Dunn et al., 2002) do not divulge exactly how the validation sets were chosen. Other studies claim independent validation by "randomly choosing" the validation set (Chang and Laird, 2002; Shepherd and Walsh, 2002; Islam et al., 2003). Brown et al. (2005) showed that random validation sets overestimate predictive accuracy by comparing these results to truly independent validation (in this case, validation by site). With spatial correlation present in soils due to landscape positioning, randomly choosing a validation dataset does not guarantee independence (Brown et al., 2005). As well, geographic scale is often overlooked in the validation process. A larger physiographic scale or wider range of soil properties implies more robustness by the model developed and a wider range of potential applicability.

Studies of MIR predictive capability are equally lacking in application of independent validation techniques. Though previous investigations have conveyed positive MIR prediction capabilities, most studies have yet to validate models with independent and unique data sets. Rumpel et al. (2001) and Bertrand et al. (2002) used

MIR spectral data to predict soil characteristics. Predictive accuracy reported was that only of the calibration data.

Investigations comparing NIR-PLS prediction capabilities with those derived from MIR-PLS data are lacking independently validated predictions from a large geographic range as well (Reeves et al, 2002; McCarty et al, 2002; Siebielec et al, 2004; Madari et al., 2005; Viscarra Rossel et al., 2006). The studies that have developed carbon calibrations have displayed MIR calibration models with higher correlation ($r^2 > 0.70$) between predicted property values and measured values than those derived from NIR calibrations. McCarty et al. (2002) validated their regional model with an independent data set including 16 samples from a different state. Though small, their study represents the most independent validation set in the literature to date, demonstrating the MIR region as the more accurate region of prediction than NIR. Table 2.1 shows all independently validated soil C prediction models.

Table 2.1 NIR and MIR independently validated C prediction studies and statistics.

Author, Year	Dataset	Calib. (n)/ Val. (n)	SOC Range (g kg^{-1})	SIC Range (g kg^{-1})	Validation	Region	SOC RMSD	SOC r^2	SIC RMSD	SIC r^2
Ben-Dor & Banin, 1995	12 groups of Israeli soils n=91	30/ 49	SOM= 0.9-132	0-743	unk.*	VNIR	1.34	0.55		0.69
Dunn et al., 2002	550 sites, 1 region	270/ 90	6.4-30		unk.*	VNIR	0.25	0.66		
Shepherd & Walsh, 2002	SE Africa n=1011	674/ 337	3.3 - 34		30% random val.	VNIR	3.10	0.80		
Chang & Laird, 2002	108 mixtures from 5 samples	76/ 32	15.4-145	0-35.7	30% random val.	VNIR	6.20	0.89	1.54	0.96
McCarty et al., 2002	Midwest n=273	257/ 16	0.23-98	0-65.4	1 site val.	NIR	7.90	0.98	4.40	0.42
						MIR	6.00	0.98	0.70	0.82
Islam et al., 2003	11orders, 2 states, Australia n=161	121/ 40	0.6-49.5		30% random val.	VNIR	0.44	0.76		
Brown et al., 2005	6 sites, 1 region, 5 soil series n=284	236/ 47	1.93-15.8	0-26.1	6-fold holdout cross val. by site	VNIR	2.53	0.63	2.76	0.93
Brown et al., 2006	Global n=4184	3988/ 665	0-537	0-129	1/6 holdout cross val.	VNIR	9.00	0.82	6.20	0.83
Viscarra Rossel et al., 2006	1 field n=118	117/1	0.81-1.98		leave-one-out cross val.	NIR	0.18	0.60		
						MIR	0.15	0.73		

* Validation set selection procedure was left undisclosed.

References

Amundson, R., 2001. The carbon budget in soils. Annu. Rev. Earth Planet Sci., 29:535-562.

Baes, A.U. and Bloom P.R., 1989. Diffuse reflectance and transmission fourier transform infrared (DRIFT) spectroscopy of humic and fulvic acids. Soil Sci. Soc. Am. J. 53: 695-700.

Baldock, J.A. and Skjemstad, J.O., 2000. Role of the soil matrix and minerals in protecting natural organic materials against biological attack. Organic Geochemistry, 31: 697-710.

Baumgardner, M.F., Silva, L.F., Stoner, E.R., 1985. Reflectance properties of soils. Advances in Agronomy, 38: 1-45.

Ben-Dor, E. and Banin, A., 1995. Near infrared analysis (NIRA) as a method to simultaneously evaluate spectral featureless constituents in soils. Soil Science, 159(4): 259-270.

Ben-Dor, E. and Banin, A., 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Sci. Soc. Am. J, 59: 364-372.

Ben-Dor, E., Inbar, Y., and Chen, Y., 1997. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400-2500 nm) during a controlled decomposition process. Remote Sens. Environ., 61: 1-15.

Bertrand, I., Janik, L.J., Holloway, R.E., Armstrong, R.D., and McLaughlin, M.J., 2002. The rapid assessment of concentrations and solid phase associations of macro- and micronutrients in alkaline soils by mid-infrared diffuse reflectance spectroscopy. Aust. J. Soil Res., 40: 1339-1356.

Bouma, J., Stoorvogel, J., van Alphen, B.J., Booltink, H.W.G., 1999. Pedology, precision agriculture and the changing paradigm of agricultural research. Soil Sci. Soc. Am. J., 63: 1763-1768.

Brown, D.J., Brickelmeyer, R.S., and Miller, P.R., 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C predictions in Montana. Geoderma, 129: 251-267.

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

Breiman, L., 1996. Bagging predictors. Machine Learning, 24(2): 123-140.
Burrough, P.A., 1993. Soil variability: a late 20th century view. Soils and Fertilizers, 56: 529-562.

Cambardella, C.A. and Elliott, E.T., 1993. Methods for physical separation and characterization of soil organic matter fractions. Geoderma, 56: 449-457.

Celi, L., Schnitzer, M., and Negre, M., 1997. Analysis of carboxyl groups in soil humic acids by a wet chemical method, FTIR spectrophotometry, and solution-state C-13 NMR: a comparitive study. Soil Science, 162 (3): 189-200.

Chang, C.W., Laird D.A., Mausbach M.J., and Hurburgh C.R., 2001. Near-infrared reflectance spectroscopy – principal components regression analyses of soil properties. Soil Sci. Soc. Am. J., 65: 480-490.

Chang, C.W. and Laird, D.A., 2002. Near-infrared reflectance spectroscopic analysis of soil C and N. Soil Science, 167(2): 110-116.

Clark, R.N, King, T.V., Klejwa, M., Swayze, G., and Vergo, N., 1990. High spectral resolution reflectance spectroscopy of minerals, J. Geophys Res, 95 (12): 653-12,680.

Clark, R. N., Swayze, G.A., and Gallagher, A., King, T.V., and Calvin, W.M., 1993. The U.S. Geological Survey Digital Spectral Library: Version 1: 0.2 to 3.0 μm, U.S. Geol. Surv. Open File Rep., pp. 93-592. (available at http://speclab.cr.usgs.gov)

Clark, R.N., 1999. Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy. In: N. Rencz (Editor), Remote Sensing for the Earth Sciences: Manual of Remote Sensing, 3 ed. John Wiley & Sons, New York, pp. 3-52.

Coates, J, 1998.A review of sampling methods for infrared spectroscopy, in *Applied Spectroscopy: A Compact Reference for Practitioners.* Workman, J. and Springsteen, A. (ed.), Academic Press, Massachusetts, pp.49-91.

Dalal, R.C. and Henry, R.J., 1986. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. Soil Sci. Soc. Am. J., 50: 120-123.

Davies, A.M.C., Dennis, C., Grant, A., Hall, M.N., Robertson, A., 1987. Screening of tomato puree for excessive mould content by near infrared spectroscopy: a preliminary evaluation. J. Sci. Food Agric., 39: 349-355.

Duckworth, J.H., 1998. Spectroscopic Quantitative Analysis. In: Workman, J. and Springsteen, A. (Eds.), Applied Spectroscopy – A Compact Reference for Practitioners, Academic Press, Chestnut Hill, MA, pp. 93-107.

Dunn, B.W., Beecher, H.G., Batten, G.D., Ciavarella, S., 2002. The potential of near-infrared reflectance spectroscopy for soil analysis – a case study from the Riverine Plain of south-eastern Australian. Australian Journal of Experimental Agriculture, 42: 607-614.

Elliott, E.T. and Cambardella, C.A., 1991. Physical separation of soil organic matter. Agriculture, Ecosystems and Environment, 34: 407-419.

Foley, W.J., McIlwee, A., Lawler, I., Aragones, L., Woolnough, A.P., Berding, N., 1998. Ecological applications of near infrared reflectance spectroscopy- a tool for rapid, cost-effective prediction of the composition of plant and animal tissues and aspects of animal performance. Oecologia, 116: 293-305.

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Stat., 29(5): 1189-1232.

Gerasimowicz, WV, Susi, and Byler, 1986. Resolution enhanced FTIR Spectra of soil constituents – humic acid. Applied Spectroscopy, 40(4): 504.

Goodacre, R., Wade, W.G. and Alsberg, B.K., 2001. Chemometric analysis of DRIFT spectra using rule induction methods: application to the classification of eubacterium species. Applied Spectroscopy, 52(6): 823.

Griffith, S.M. and Schnitzer, M., 1975. Analytical characteristics of humic and fulvic acids extracted from tropical volcanic soils. Soil Sci. Soc. Amer. Proc., 39: 861-866.

Hassink, J., 1995. Decomposition rate constants of size and density fractions of soil organic matter. Soil Sci. Soc. Am. J., 59: 1631-1635.

Hassink, J., Whitmore, A., and Kubat, J., 1997. Size and density fractionation of soil organic matter and the physical capacity of soils to protect organic matter. European J. of Agronomy, 7: 189-199.

Islam, K., Singh, B., and McBratney, A., 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. Australian Journal of Soil Research, 41: 1101-1114.

Janik, L.J, and Skjemstad, J. O., 1995. Characterization and analysis of soils using mid-infrared partial least-squares. II. Correlations with some laboratory data. Aust. J. Soil Res., 33: 637-650.

Janik, L.J., 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? Australian Journal of Experimental Agriculture, 38: 681-96.

Johnson, T.J., Amonette, J.E., Valentine, N.B., Thompson, S.E. and Foster, N.S., 2004. Identification of sporulated and vegetative bacteria using statistical analysis of FTIR data. Applied Spectroscopy, 58 (2): 203-211.

Kaihara, M., Sato, Y., Sato, T., Takahashi, T. and Takahashi, N., 2001. Chemometrics for FTIR reflectance spectroscopy – the distinction between the specular and diffuse reflectance. Analytical Sciences, 17: Supp.

Lal, R., 1991. Soil structure and sustainability. Journal of Sustainable Agriculture, 1(4): 67-91.

MacCarthy, P. and Mark H.B., 1975. Direct measurement of infrared-spectra of humic substances in water by Fourier-transform infrared spectroscopy. J. of Ag. And Food Chem., 23 (3): 600-602.

Madari, B.E., Reeves, J.B., Coelho, M.R., Machado, P., De-Polli, H., 2005. Mid- and near-infrared spectroscopic determination of carbon in a diverse set of soils from the Brazilian National Soil Collection. Spectroscopy Letters, 38: 721-740.

McBratney, A.B., Santos, M.L.M., Minasny, B., 2003. On digital soil mapping. Geoderma, 117: 3-52.

McCarty, G.W., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Sci. Soc. Am. J, 66: 640-646.

Morra, M.J., Hall, M.H., and Freeborn, L.L., 1991. Carbon and nitrogen analysis of soil fractions using near-infrared reflectance spectroscopy. Soil Sci. Soc. Am. J., 55: 288-291.

Neyroud, J. and Schnitzer, M., 1974. The chemistry of high molecular weight fulvic acid fractions. Can. J. Chem, 52: 4123-4132.

Nguyen, T., Janik, L.J., Raupach, M., 1991. Diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy in soil studies. Aust. J. Soil Res, 29: 49-67.

Olinger , J.M. and Griffiths, P.R., 1993. Effects of sample dilution and particle-size morphology on diffuse reflectance spectra of carbohydrate systems in the NIR and MIR. Applied Spectroscopy, 47(6): 687-694.

Post, W.M. et al., 2001. Monitoring and verifying changes of organic carbon in soils. Climatic Change, 51: 73-99.

Reeves III, J.B., McCarty, G.W., and Mimmo, T., 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. Environmental Pollution 116: S277-S284.

Roberts, D.A., Gardner, M., Church, R., Ustin, S., Scheer, G. and Green, R.O., 1998. Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. Rem. Sens. of Env., 65 (3): 267-279.

Rumpel, C., Janik, L.J., Skjemstad, J.O., and Kogel-Knabner, I., 2001. Quantification of carbon derived from lignite in soils using mid-infrared spectroscopy and partial least squares. Organic Geochemistry, 32: 831-839.

Sain, S. and Carmack, P., 2002. Boosting Multivariate Regression Trees. *Computing Science and Statistics*, 34, 2002 Proceedings.

Schnitzer, M., 1969. Characterization of humic constituents by spectroscopy. In: McLaren, A.D. and Skujins, J. (eds.), Soil Biochemistry II, Marcel Dekker, New York, pp. 60-95.

Schuh, W.M., Cline, R.L, Sweeney, M.D., 1988. Comparison of a laboratory procedure and a textural model for predicting in situ soil water retention. Soil Sci. Soc. Am. J., 52: 1218-1227.

Schulten, H.R. and Leinweber, P., 2000. New insights into organic-mineral particles: composition, properties and models of molecular structure. Biology and Fertility of Soils, 30 (5-6): 399-432.

Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Progress in Physical Geography, 27(2): 171-197.

Shepherd, K.D. and Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Sci. Soc. Am. J., 66: 988-998.

Siebelec, G., McCarty, G.W., Stuczynski, T.I., and Reeves, J.B., 2004. Near- and mid-infrared diffuse reflectance spectroscopy for measuring soil metal content. J. Environmental Quality, 33: 2056-2069.

Six, J., Paustian, K., Elliott, E.T., and Combrink, C., 2000. Soil structure and organic matter: I. Distribution of aggregate-size classes and aggregate-associated carbon. Soil Sci. Soc. Am. J., 64: 681-689.

Skjemstad, J.O., Frost, R.L., and Barron, P.F., 1983. Structural units in humic acids from south-eastern Queensland soils as determined by C-13 NMR spectroscopy. Aust. J. Soil Res., 21: 539-547.

Springsteen, A., 1998. Reflectance Spectroscopy: An Overview of Classifications and Techniques. In: Workman, J. and Springsteen, A. (Eds.), Applied Spectroscopy – A Compact Reference for Practitioners, Academic Press, Chestnut Hill, MA, pp. 194-223.

Stevenson, F.J. and Goh, K.M., 1971. Infrared spectra of humic acids and related substances. Geochimica et Cosmochimica Acta 35 (5): 471.

Stevenson, F.J. and Goh, K.M., 1974. Infrared spectra of humic acids – elimination of interference due to hygroscopic moisture and structural changes accompanying heating with KBr. Soil Science, 117 (1): 34-41.

Stevenson, F.J., 1982. Nitrogen in Agricultural Soils. American Society of Agronomy, New York, pp. 23-30.

Stewart, 1998. Color and Solar Transmittance in *Applied Spectroscopy: A Compact Reference for Practitioners.* Workman, J. and Springsteen, A. (ed.), Academic Press, Massachusetts, pp. 344-382.

Sudduth, K.A. and Hummel, J.W., 1993. Portable, near-infrared spectrophotometer for rapid soil analysis. Transactions of the ASAE, 36(1): 185-193.

Tan, K.H., 1977. Infrared-absorption similarities between hymatomelanic acid and methylated humic acid. Soil Sci. Soc. Am. J., 39 (1): 70-73.

Tkachuk, R., 1981. Oil and protein analysis of whole rapeseed kernels by near infrared reflectance spectroscopy. JAOCS, pp. 819-821.

Umland, J.B. and Bellama, J.M, 1996. General Chemistry, 2nd edition. Brooks/Cole Publishing Company, Pacific Grove, CA, pp. 224-234.

Van Alphen, B.J. and Stoorvogel, J.J., 2000. A functional approach to soil characterization in support of precision agriculture. Soil Sci. Soc. Am. J., 64: 1706-1713.

Vinkler, P., Lakatos, B. and Meisel, J., 1976. Infrared spectroscopic investigations of humic substances and their metal complexes. Geoderma, 15 (3): 231-242.

Wetzel, D.L, 1983. Near-infrared reflectance analysis: Sleeper among spectroscopic techniques. Analytical Chemistry, 55(12): 1165A-1176A.

Weyer, L.G. and Lo, S.C., 2002. Spectra-Structure Correlations in the Near-infrared. In: Handbook of Vibrational Spectroscopy, John Wiley and Sons Ltd., pp. 1817-1837.

Whalley, W.R. and Leeds-Harrison, P.B., 1991. Estimation of soil moisture status using near infrared reflectance. Hydrological Processes, 5: 321-327.

Williams, P.C. and Norris, 1985. Determination of protein and moisture in wheat and barley by near-infrared transmission. J. Agric. Food Chem., 33: 239-244.

Wold, S., Sjostrom, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics Intelligence Laboratory Systems, 58: 109-130.

Workman, J., 1998. Ultraviolet, Visible and Near-Infrared Spectroscopy. In: Workman, J. and Springsteen, A. (Eds.), Applied Spectroscopy – A Compact Reference for Practitioners, Academic Press, Chestnut Hill, MA, pp. 29-47.

CHAPTER 3

VISIBLE TO NEAR INFRARED DIFFUSE REFLECTANCE SPECTROSCOPY TO CHARACTERIZE CARBON CONTENT OF GLACIAL TILL CALCAREOUS SOILS OF NORTH CENTRAL MONTANA

Abstract

Diffuse reflectance spectroscopy (DRS) has been considered a potential surrogate for traditional laboratory analyses used to characterize soils. Regression techniques, in particular Partial Least Squares (PLS) analysis, are employed to develop calibration models for predicting soil properties, including organic and inorganic soil C. It is difficult to judge from current research whether DRS has the potential to replace traditional laboratory protocols because studies lack independent model validation. There is a lack of information regarding sample preparation as well as independent validation of global vs. regional calibrations developed from visible-near infrared (Vis-NIR, 400-2500 nm) spectra. In this study, I used PLS regression to predict soil organic (SOC) and inorganic carbon (SIC) values with spectral data from the Vis-NIR region. A dataset including 315 soil samples from six different sites in north central Montana was used with 6-fold site-holdout to determine predictive accuracy of four different preparation methods: *in-situ* field-moist core, air-dried core, oven-dried crushed and sieved (< 2-mm) and oven-dried milled (< 200-µm). The milled preparation gave the most accurate predictions for soil organic C (SOC), with root mean squared deviation

(RMSD) = 1.64 g kg^{-1}, correlation coefficient (r^2 = 0.78), and residual prediction deviation (RPD) = 2.07. The sieved preparation gave the most accurate soil inorganic C (SIC) predictions, with RMSD = 2.81 g kg^{-1}, r^2 = 0.93, and RPD = 3.65. Water within *In-situ* moist cores visually influenced spectral signatures as well as negatively affected predictive accuracy; SOC predictive results provided RMSD = 2.89 g kg^{-1}, r^2 = 0.37, and RPD = 1.20. Dried cores were more accurate providing a RMSD = 2.76 g kg^{-1}, r^2 = 0.42, and RPD = 1.26. *In-situ* analysis did not prove analytical sufficiency for C predictions but could meet semi-quantitative needs for certain applications. We inspected local (individual validation site) vs. regional (6-farm dataset) vs. global (~4000 sample dataset from 6 different continents) calibration models and found regional with local addition to be the most successful means of predicting both SOC and SIC for our sample set. Organic matter (SOM) spectral influences were qualitatively examined and found to be relatively small in comparison to carbonate and water influences.

Introduction

Vis-NIR can provide rapid, inexpensive and non-destructive soil analysis (Dalal and Henry, 1986; Sudduth and Hummel, 1993; Ben-Dor and Banin, 1995; McCarty et al., 2002; Shepherd and Walsh, 2002; Dunn et al, 2002; Brown et al., 2005). Standard laboratory procedures for measuring soil properties can be complex, time-consuming, and expensive (Brown, 2005). Near infrared spectroscopy is known for its rapidity, convenience, simplicity and for its ability to analyze many constituents simultaneously

while producing no waste and leaving samples intact and chemically unaltered (Stark, 1986; Davies, 1987).

In-situ (in field) analysis is possible with a portable Vis-NIR spectrometer increasing its convenience, but *in-situ* predictive abilities have not been established. Sudduth and Hummel (1992) developed a field portable NIR spectrophotometer and validated its accuracy to quantify SOM. Dalal and Henry (1986) examined multiple linear regression predictions of SOC, N, and moisture from the spectra of coarsely ground (<2 mm) vs. finely ground (< .25 mm) samples and found standard error of prediction (SEP) higher for coarsely ground samples. Their results led them to claim that the NIR prediction technique would be less reliable for *in-situ* samples. Twenty years later, Waiser et al. (in review) found acceptable predictions for clay content using *in-situ* measurement. More research in the area of sample preparation, including sample homogeneity and moisture state, would be helpful to bring conclusion to the matter.

Vis-NIR advances are closely tied to new modeling techniques. Most of the recent studies have focused on calibrating and validating models using PLS regression with the intent of using spectral data to predict individual soil characteristics (Dunn et al., 2002; McCarty et al., 2002; Islam et al., 2003; Brown et al., 2005). Partial Least Squares regression is a multivariate calibration method that breaks down correlated spectral data (X) and soil property data (Y) into a new smaller set of uncorrelated variables that best describe all the variance in the data (Haaland and Thomas, 1988; Wold et al., 2001). In development of a prediction model, it is necessary to calibrate the model with laboratory data collected and then independently validate the model using soils that were not

included in the calibration procedure. There has been an abundance of prediction models developed with validation methods that are claimed to be independent (Dalal and Henry, 1986; Morra et al., 1991; Ben-Dor and Banin, 1995; Chang et al., 2001; Chang and Laird, 2002; Dunn et al., 2002; Islam et al., 2003), but selection process of the validation set has been omitted. The selection process of validation sets that is included in most studies has been a "random" selection procedure, which does not directly imply independence unless spatial relationships are accounted for (Brown et al., 2005). As NIR analysis is based on observations and laboratory processes, the validation stage is extremely critical (Ben-Dor and Banin, 1995). Validation with an independent data set is a crucial part of this stage (Brown et al., 2005) and so far, where much of the current research fails.

Calibration set characteristics and their influence over predictive accuracy have been examined. These include calibration set size (Shepherd and Walsh, 2002), range of characteristics (Morra et al., 1991), and geographical origin/applicability (i.e. regional vs. global, Brown et al., 2006). It was determined that a larger calibration set gives more accurate predictions, but a large range of values decreases predictive accuracy. Brown et al. (2006) display semi-quantitative predictive accuracy of different soil characteristics obtained with a large calibration set (~4000 samples) of international origin. Global, regional and local predictions have not yet been compared in the same study.

Several research questions will be addressed in this study including

1.) Is it possible to develop a robust soil C prediction model for calcareous glacial till soils of north central Montana using local, regional and

global calibrations? If so, which calibration set develops the more accurate prediction model?

2.) Can in-situ calibrations produce accurate predictions?

3.) Are SOC predictions based on absorptions of SOM in the Vis-NIR region or on other soil constituents such as carbonates, clay or Fe-oxides?

These questions will be answered with spectral data from the Vis-NIR region with the anticipation that overall predictive accuracy of this region can be evaluated for soil C.

Materials and Methods

The methods implemented in this study include collection and processing of soil samples, collection and processing of spectral data, construction of calibration models using local, regional and global spectral data, and independent validation of these models using a reserved dataset (in this case, one field site). A flow chart displays an overview of the sample presentation and modeling techniques used (Fig. 3.1).

Sampling Design

Study area and Core Selection. The study area consists of six farm sites located in north central Montana, bounded by the towns of Great Falls, Shelby and Havre (Fig. 3.2). The soils originated from glacial deposits that contain similar calcareous parent material throughout the six sites (Montagne et al., 1982). Vegetation used to include short prairie grasses but have been converted to wheat cropland. The sites are found within ustic

moisture and frigid temperature regimes. Topography is gently rolling slopes, < 5% in most cases.

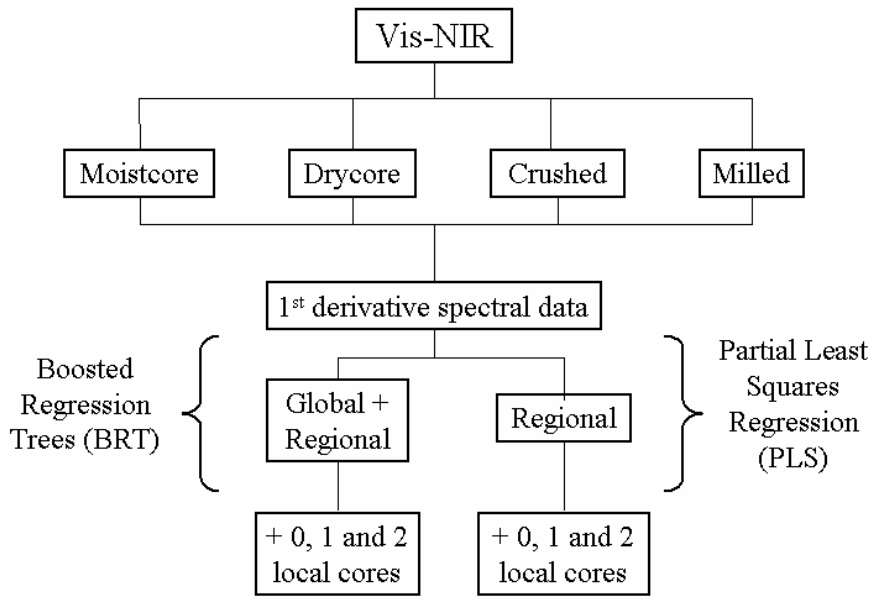

Figure 3.1. An overview of sample presentation, calibration sets and modeling techniques used in the study.

At each farm, a field of 32 ha was divided into four strips (8 ha) representing both till and no-till as well as different crop rotation management schemes. Within each strip four points were identified for sampling/monitoring of soil carbon changes over time (part of the CASMGS, Consortium for Agricultural Soils Mitigation of Greenhouse Gases, project). In total, there are 16 sampling points per farm, and at each point there are five cores of 50-100 cm depth taken in a star shaped pattern. The star locations were determined by landscape position, followed by digging soil pits to determine texture and

depth to carbonates. Pits that met depth to carbonate and texture requirements were chosen as core sampling locations. This sampling method was intended to build an accurate representation of the field. Each core was drawn in September 2002 and was divided into three depths: 0-10, 10-20, and 20-50 cm resulting in three samples per core. Two cores per star location were dug to 100 cm depth (where possible) resulting in four samples per core (one additional 50-100 cm sample).

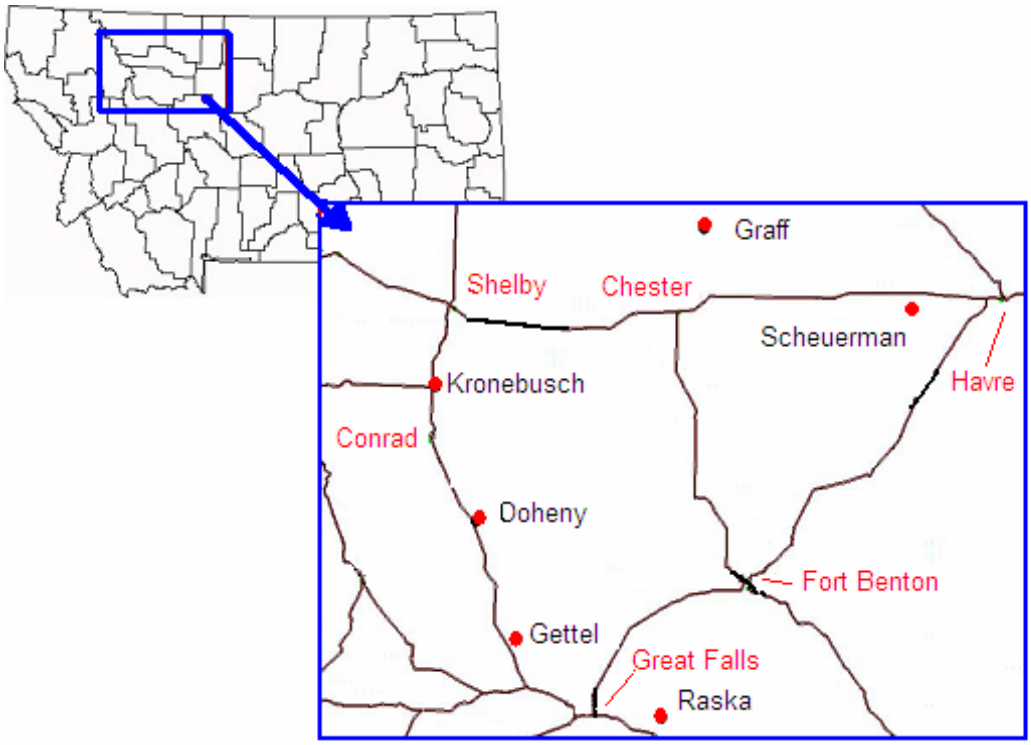

Figure 3.2. A map of Montana including the six sampling sites (identified by symbols labeled with names of farm owners) in the north central "Golden Triangle" region.

In total, 480 cores were collected from the six farm sites and a 94-core subset was selected for this study. The subset was built using the deepest of the five cores from each designated star site. If there were equivalent core depths at each site, a core was chosen randomly. The final subset therefore consisted of 315 individual soil samples. This dataset is referred to as "Regional" throughout the study.

Additional Datasets. An additional regional dataset (known as "Regional2") consisting of 404 samples was included with the original dataset to assess effects of calibration set size on predictive accuracy (Table 3.1). These samples also originated from the Golden Triangle region of north central Montana, but from different farm sites than the principal 315-sample dataset (see Brown et al., 2005 for complete sampling description). The cores were collected and divided in September 2001, in the same manner described above. Vis-NIR spectral reflectance data was collected using the same equipment and procedure described below, after samples had been dried at 40 ºC and passed through a 2-mm sieve.

A "Global" dataset, consisting of 3794 and 4184 samples with SOC and SIC measurements, respectively, was provided by the Natural Resources Conservation Service (NRCS). The Global dataset includes Vis-NIR spectral reflectance data obtained from air-dried, 2-mm sieved soils from six continents. This set was used along with the Regional dataset to build a global calibration model.

Local Core Addition. Spectrally dissimilar cores from the site being validated were added to the calibration set during the modeling procedures. These cores are

referred to as "Local" cores or validation site cores. The spectral difference was determined by a standardized Mahalanobis Distance (Eq. 1), computed with standardized scores derived from principal components analysis (PCA). PCA was performed (S-Plus® 6.0, Insightful Corp., Seattle, WA, USA) for each Crushed, Milled, Moistcore and Drycore first derivative reflectance datum. Scores for the components explaining at least 90% of the variance (or for the first ten components, whichever provides scores of at least 10 components), were standardized to provide $\hat{\chi}$ for each sample. Standardized component scores for samples from the five sites not related to the sample being analyzed were averaged to give $\hat{\chi}_a$ for each component. The Mahalanobis Distance was then computed for each sample as:

$$\textit{Mahalanobis Distance} = \sqrt{(\hat{\chi} - \hat{\chi}_a)} \qquad (1)$$

$\hat{\chi}$ is the individual sample standardized PCA score vector
$\hat{\chi}_a$ is the average standardized PCA score vector of remaining 5 sites

The sample with the greatest distance was considered the most spectrally dissimilar sample. This sample along with the corresponding samples from its core, were all removed from the validation set and added to the calibration set.

Sample Preparation.

Four methods of sample preparation including moist core, dry core, milled and 2-mm sieved states were compared in this study (Fig. 3.3). The predictive accuracy as well as the qualitative spectral influence of each preparation method was the basis of comparison.

Table 3.1 A descriptive table of datasets used in the study.

Dataset	# of Samples	Origin	Preparation
Global	3794 SOC and 4184 SIC	6 continents	Sieved*
Regional2	405	6 independent sites within NC Montana	Sieved
Regional	315	6 different independent sites within NC Montana	Moistcore, Drycore, Sieved and Milled
Local	0, 1 or 2 cores (0-8 samples)	validation site	Moistcore, Drycore, Sieved and Milled

* Carbon content determined by the Walkley Black method.

In-situ. Field-moist ("Moistcore") and dried intact cores ("Drycore") were scanned in the laboratory to simulate *in-situ* analysis. Considering collection location (a semi-arid region) as well as the late summer/early fall collection time, Moistcores had moisture contents near dry. The moist cores were cut lengthwise and scanned at four different points within each sample interval down the center of the core.

Sieved. After drying at 40 ºC, each sample was passed through a 2-mm sieve to remove rock fragments, surface plant litter, and coarse root material. This preparation method is referred to as "Sieved".

Milled. Approximately six grams of each crushed soil sample were milled to a powdered homogeneous particle size (<200-µm) for three minutes per sample using a

ball mill (Pica Blender Mill Model 2601, Cianflone Scientific Instruments Corp., Pittsburg, PA). This preparation method is referred to as "Milled".

Organic Matter Removal. Organic matter (SOM) was removed from Sieved samples prior to spectral analysis by treating a 13-g sub-sample with 20-40 ml of Na-hypochlorite. The treated sample was then placed in a 69 ˚C sonic water bath for 1 hr followed by 2 hrs in a 70 ˚C oven. Water was added to samples as necessary to prevent desiccation. Samples that appeared black after the initial Na-hypochlorite treatment were allowed to settle and clear supernatant was removed, before repeating the treatment again. After SOM had been removed, samples were rinsed four times with deionized water using a centrifuge at 10,000 rpm (25 minutes) and dried overnight in a 70 ˚C oven. After the rinsing and drying process, the soils were milled for three minutes to <200-µm using the ball mill described above.

Spectroscopy

Instrument. Samples with and without OM, and with different preparation methods, were analyzed using diffuse reflectance spectral analysis. Vis-NIR spectral reflectance scans were obtained with an ASD (Analytical Spectral Devices, Boulder, CO) "Fieldspec Pro FR" spectroradiometer. This device measures visible reflectance from 350-700 nm as well as NIR reflectance from 700-2500 nm with a 2 nm sampling resolution and spectral resolution of 3 nm at 700 nm and 10 nm at 1400 and 2100 nm. One spectral signature resulting from ten internally averaged scans between 350 and 2500

nm is recorded per "scan". The laboratory instrument is equipped with a "mug lamp" white light source. The Milled and Crushed samples were set on top of this mug lamp in an optical borosilicate glass Petri dish (Duran®). The scans were calibrated using a Spectralon ® white reference panel. Two scans were taken per sample, with a 90-degree rotation in between each, and were then averaged to obtain a composite scan for each sample.

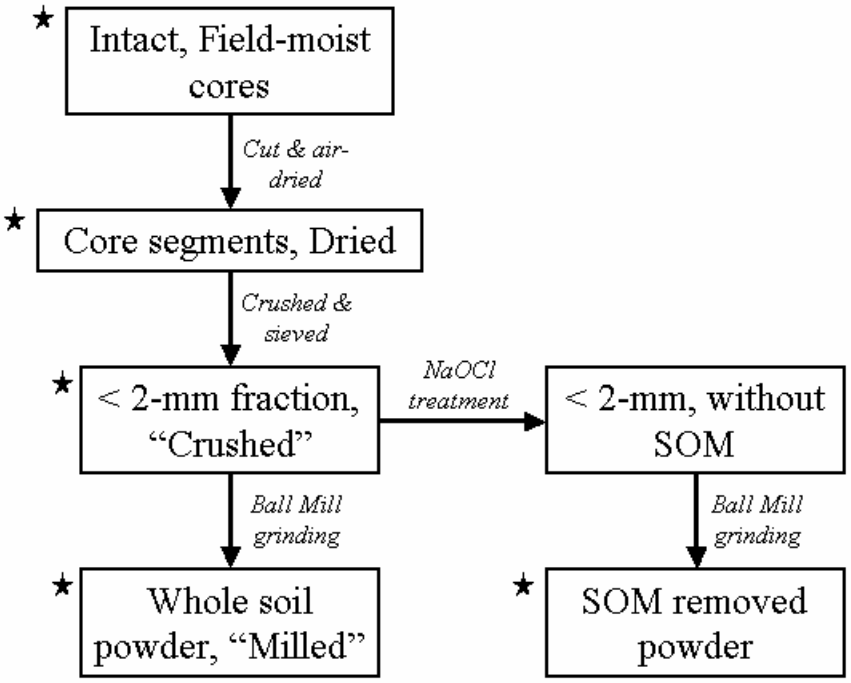

Figure 3.3. Flow chart of different sample preparation methods used with stars representing scanned presentations.

Core scanning. The ASD "contact probe" was used to obtain spectral data of intact soil cores. The probe has a white light source and quartz-sapphire window that can

be pressed against soil material. The probe connects to the Fieldspec Pro FR and spectral signatures are recorded and averaged in the manner described above. Moistcores were cut lengthwise and scanned at two different points within each sampling depth interval. Each point was scanned twice with the contact probe, with a 90º rotation of the probe between scans. Each set of four scans was averaged to compile a spectral representation of that particular depth. Moistcores were then cut into depth intervals, oven-dried at 40 ºC and scanned again in the same fashion to obtain Drycore spectral data.

Spectral Pretreatment. Spectral pretreatment methods were applied according to Brown et al. (2005). Cubic smoothing splines were fit to raw reflectance spectra using "R" software (R Development Core Team, 2004). First and second derivatives were taken of reflectance spectra to account for baseline effects (Duckworth, 1998). Second derivative data results were less successful than first derivative results and were not reported. Smoothed and transformed data were extracted at 10 nm intervals to ease data handling.

Conventional Laboratory Analyses

Carbon. Total carbon (TC) content was measured in the laboratory through dry combustion using a LECO C/N/S 2000 analyzer (LECO Corporation, St Joseph, MI, USA). Milled samples were oven dried overnight at 70˚ C, and 0.2 g samples of each soil were weighed out in preparation for this analysis. Additional pre-weighed 0.2 g samples were treated with 1 ml of 1 M H_3PO_4, and then dried in a 70˚ C oven overnight prior to

analysis. TC of the whole soil less TC of the acidified soil gives SOC content (assuming acid removes all IC). Fifteen samples were chosen to obtain duplicate C measurements for estimation of both SOC and SIC standard error of laboratory (SEL) measurements.

Particle Size Analysis. Particle size analysis was performed using the pipet method (Gee and Bauder, 1986), with sodium hypochlorite (NaOCl) added to 10 g of sieved samples to remove soil organic matter. After settling, the samples were wet sieved at 53 μm for sand content determination.

Mineralogy. Clay mineralogy was determined for a subset of samples using oriented and powder mount x-ray diffraction (XRD) techniques (Whittig and Allardice, 1986). The subset included four samples, two surface and two subsurface samples, chosen randomly from each site. The 24 clay samples were analyzed using Cu radiation with a Siemens D5000 X-ray Diffractometer (XRD), located in the Image and Chemical Analysis Laboratory on the Montana State University, Bozeman campus. X-ray diffraction was performed for four treatments: (i) Mg-saturation, (ii) Mg-saturation and solvation with glycerol, (iii) K-saturation, and (iv) K-saturation with 550 °C heat treatment. The clay solutions were mounted as oriented crystals on porous ceramic plates. Oriented samples were step scanned from 3 to 14 degrees using 0.02° 2-theta step increments at a rate of 2.5-3 deg/min with 0.48 second count times. Samples with a high degree of noise were rescanned at 2 deg/min.

Peaks were recognized with Diffraction Management System Software (DMSNT), version 1.37 (Scintag, Inc., now Thermo Electron Corporation, Waltham,

MA). Background noise was removed and boxcar smoothing applied. Peaks were identified manually and by automated selection through digital filtering. Ordinal classes were assigned to each sample based on peak intensity using the National Soil Survey Center (Lincoln, NE) six-class procedure where 0 indicates none present, 1 indicates a peak intensity range of 1-110 Å, 2 (110-360 Å), 3 (360-1120 Å), 4 (1120-1800 Å), and 5 represents an intensity of greater than 1800 Å.

Modeling: Local, Regional and Global

Partial Least Squares Regression

Model building was done using PLS-1 within Unscrambler 8.05 software (CAMO PROCESS AS, Oslo, Norway). To build calibration models, partial least squares (PLS) regression was used to quantify SOC and SIC content using Vis-NIR first derivative reflectance data. The number of components chosen for each calibration model follows the methodology of Brown et al. (2005). The strategy is based on visual examination of the decrease in residual variance with each component determined through cross validation of the calibration set. To avoid overfitting, the number of components chosen is based on the point where the variance stops decreasing.

Calibration models were built around the six site subsets, using a 6-fold site-holdout cross-validation method. For example, sites 2-6 were used to calibrate the model that was then validated exclusively on site one. As each site is geographically independent of the next, this calibration method provided for independent regional validation, simulating predictions at a new site within the region. In a separate step, cores

from the site being validated (referred to as "Local cores") were added to calibration sets to determine whether or not this improved the model's prediction capabilities.

Regional + Local calibration models were constructed using: (i) the five sites mentioned above, (ii) the five sites plus one Local core, and (iii) the five sites plus two Local cores.

Regional2 calibration models were built with the same (i)-(iii) PLS model building procedure used for local calibrations, except with the Regional2 dataset added to each step.

Loadings are a measure of how much each predictor (X-variable) contributes to explaining the response variation along each model component (Haaland and Thomas, 1988). The components that explain the most variability will be observed as more influential to the model. Loadings from the most accurate SOC PLS calibration models were observed to determine the influence of each wavelength within the SOC prediction

Boosted Regression Trees (BRT)

Global calibration models were built using stochastic gradient BRT modeling with Treenet® (Salford Systems Inc., San Diego, CA, USA) after Brown et al. (2006). A maximum of 800 trees was specified, with 6 nodes per tree. Two blocks of models were run, both included the focal 315-sample set + 1 and + 2 Local cores. The first block included the Global dataset in addition to the Regional dataset with Local additions and the second block included the Global with the Regional2 dataset with Local additions.

Predictive Statistics.

Predicted and measured values from the modeling procedures listed above were compared using the following statistics taken after Brown et al. (2005):

Mean Square Deviation: $$MSD = \frac{\sum_n \left(Y_{pred} - Y_{meas}\right)^2}{N} \quad (2)$$

Root Mean Square Deviation: $$RMSD = \sqrt{MSD} \quad (3)$$

Standard Error of Prediction: $$SEP = RMSD \times \sqrt{\frac{N}{(N-1)}} \quad (4)$$

Residual Prediction Deviation: $$RPD = (\textit{Standard Deviation of } Y_{val})/SEP \quad (5)$$

$$r^2 = \textit{square correlation} \quad (6)$$

$$Bias = \frac{\sum_n \left(Y_{pred} - Y_{meas}\right)}{N} \quad (7)$$

Principal Components Analysis.

First derivative reflectance data for all sample preparation methods were plotted in principal component space (S-Plus® 6.0, Insightful Corp., Seattle, WA, USA) to identify spectral patterns and the most spectrally dissimilar samples (Shepherd and Walsh, 2002; Brown et al., 2005a). The components explaining 90% of the variance were plotted in scatterplot matrix format. Crushed preparation gave the most distinct pattern so it was focused on.

Continuum Removal.

Continuum removal was performed according to Clark (1999) on reflectance data of crushed samples in the region from 2300-2350 nm to obtain carbonate band depths for each sample. Continuum is background absorption consistent throughout the soil spectral signature and was removed to determine the depth of the feature due to the carbonate content rather than background influences. A linear fit between the two endpoints of the absorption feature was removed from the band depth of the feature at 10 nm intervals to determine band depth using the following equation

$$D = 1 - (R_b/R_c) \quad (8)$$

where R_b is the reflectance at the band bottom and R_c is the reflectance at the continuum (Clark and Roush, 1984).

According to Clark (1999), the region from 2300-2350 nm is the $CaCO_3$ first overtone absorption region as well as the most distinct carbonate feature in the Vis-NIR region. Multiple linear regression was applied to the resulting band depths for all 315 local samples, using band depth as a predictor variable and SIC content as the response. Band depth was plotted in this region by ordinal SIC content groups to determine if there was a visual correlation between SIC content and band depth in this region.

Results

Soil Carbon and Texture Data

The distribution of SOC, SIC and texture among sites can be viewed in Table 3.2. Site 2 exhibits the highest average SOC and SIC contents, with average SOC at 12.28 and

average SIC at 24.15 g kg^{-1}. Overall, C contents are relatively low and evenly distributed throughout the six sites. SOC contents are evenly spread except that sites 2, 4 and 5 include a few samples with higher SOC than the other 3 sites. Site 2 is more calcareous than the other sites and includes five extreme samples that increase the spread of the dataset. This range of C content should add to the robustness of the model, including a wider range of C values that can be predicted. Overall, sand and clay contents are relatively evenly distributed with site 5 containing the most clay and site 6 containing the most sand on average. From the 15 replicated samples, SEL measurements for the combustion method were found to be 0.27 g kg^{-1} for SOC and 0.34 g kg^{-1} for SIC.

Table 3.2: Descriptive statistics for the six sites

Property (g kg^{-1})	Statistic	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
	Min:	4.1	7.5	3.5	6.4	4.9	3.8
SOC	Max:	12.5	18.9	13.5	17.2	19.1	10.8
	Mean:	7.6	12.3	7.8	11.2	10.6	7.0
	Min:	0.0	1.8	0.0	0.0	0.2	0.0
SIC	Max:	15.5	72.5	22.9	18.0	14.2	19.3
	Mean:	7.1	24.1	9.6	3.5	6.2	5.3
	Min:	311.8	212.8	168.0	174.4	296.8	180.6
Clay	Max:	574.5	519.2	374.5	497.5	710.4	344.5
	Mean:	443.9	343.6	266.0	338.6	558.1	251.0
	Min:	135.2	102.0	167.9	127.2	100.0	214.7
Sand	Max:	444.9	369.5	634.6	515.9	232.9	510.6
	Mean:	279.6	193.0	385.9	344.0	144.4	397.0
	N:	51	48	57	48	59	52

Mineralogy

Oriented clay mounts displayed mixed clay mineralogy including moderate amounts of vermiculite, smectites, micas and kaolinite within each site. Powder mount analysis displayed major peaks at 3.4 (quartz) and 3.1 (feldspar) d-spacing. There was

also a large unidentifiable peak shown at approximately 1.83 d-spacing for a sample from site 5.

Prediction Results

Regional Calibration Model Validation. The most accurate SOC calibration validations were derived from Milled preparation, plus two Local cores, providing a RMSD of 1.64 g kg^{-1}, r^2 of 0.78, and RPD of 2.07 (Table 3.4; Fig. 3.4). Standard error of prediction was 1.4 g kg^{-1} (Table 3.3). The most accurate SIC results came from Sieved preparation, with two Local cores added, providing a RMSD of 2.81 g kg^{-1}, r^2 of 0.93, and RPD of 3.65 (Table 3.6; Fig. 3.5). Standard error of prediction was 2.8 g kg^{-1} (Table 3.5). Milled sample SIC results gave a fair r^2 value at 0.87 but RMSD was higher at 4.00 g kg^{-1}. SEP comparisons display Local Sieved and Milled predictions as the most accurate (Tables 3.5 and 3.6). Calibration statistics for the most accurate models are presented in Table 3.7.

Table 3.3: SEP results from Local and Regional validations of calibrations developed with four different preparation methods.

	Soil Organic Carbon SEP values (g kg-1)			
# of Local Cores	Sample Preparation Methods			
	Moistcore	Drycore	Sieved	Milled
0	3.60	3.40	3.13	2.03
1	3.28	3.00	2.45	1.91
2	2.90	2.77	2.16	1.65

Table 3.4: PLS validation statistics for Regional and Local organic C calibration models developed using soils in four different preparation states.

	Soil Organic Carbon			
Sample Preparation	local cores added	RMSD (g kg^{-1})	r^2	RPD
Moistcore	0	3.59	0.15	0.96
	1	3.27	0.24	1.06
	2	2.89	0.37	1.20
Drycore	0	3.39	0.27	1.01
	1	2.99	0.36	1.16
	2	2.76	0.42	1.26
Crushed	0	3.12	0.28	1.10
	1	2.45	0.53	1.41
	2	2.15	0.61	1.59
Milled	0	2.02	0.66	1.70
	1	1.90	0.70	1.80
	2	1.64	0.78	2.07

Table 3.5: SEP results from Local and Regional validations of calibrations developed with four different preparation methods.

	Soil Inorganic Carbon SEP values (g kg-1)			
# of Local Cores	Sample Preparation Methods			
	Moistcore	Drycore	Sieved	Milled
0	9.41	7.39	6.34	7.72
1	6.96	5.42	4.02	6.86
2	6.17	5.13	2.82	4.01

Table 3.6: PLS validation statistics for Regional and Local inorganic C calibration models developed using soils in four different preparation states.

	Soil Inorganic Carbon			
Sample Preparation	local cores added	RMSD (g kg^{-1})	r^2	RPD
Moistcore	0	7.37	0.58	1.47
	1	5.40	0.77	1.98
	2	5.11	0.78	2.05
Drycore	0	6.18	0.75	1.75
	1	6.21	0.75	1.72
	2	5.96	0.75	1.76
Crushed	0	6.32	0.74	1.71
	1	4.01	0.89	2.67
	2	2.81	0.93	3.65
Milled	0	7.70	0.57	1.41
	1	6.83	0.63	1.56
	2	4.00	0.87	2.55

Table 3.7: Calibration statistics for Local SOC and SIC models.

Variable	Most Accurate Calibration Model	RMSD (g kg^{-1})	r^2	RPD
SOC	Milled + 2 cores	1.01	0.90	3.41
SIC	Crushed + 2 cores	2.67	0.94	4.11

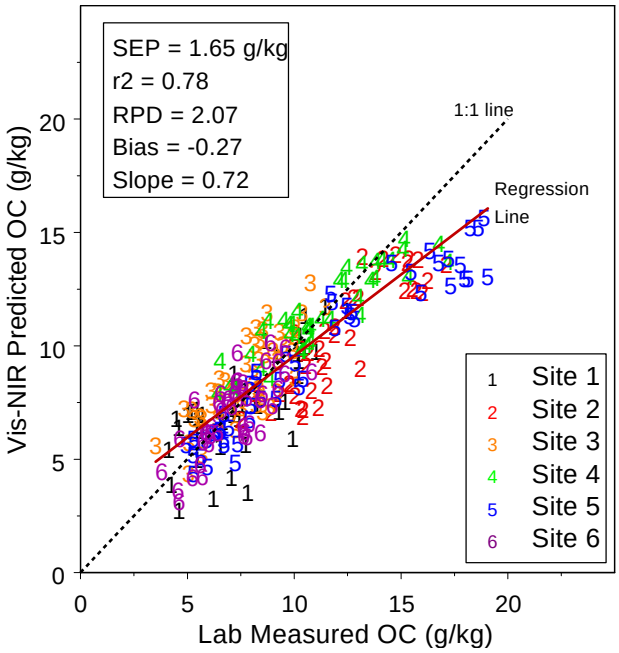

Figure 3.4. PLS validation results of the most accurate organic C Regional + 2 Local core calibration developed using Milled sample preparation.

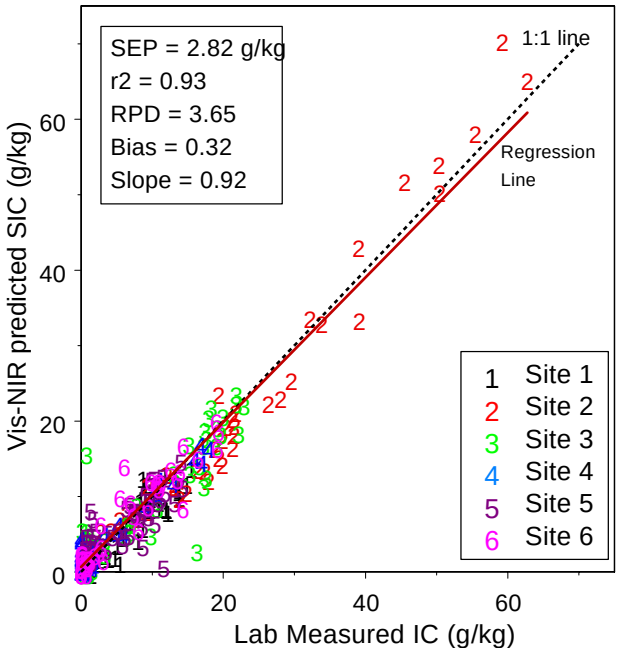

Figure 3.5. PLS validation results of the most accurate inorganic C Regional + 2 Local core calibration developed using Sieved sample preparation.

Local Core Addition. Prediction results were improved for all sample preparation methods with the addition of one or two validation site cores to the calibration set. For SOC predictions with Sieved preparation, RMSD = 3.12 g kg^{-1}, r^2 = 0.28, and RPD = 1.10 without the addition of Local cores. After the addition of two cores, results improved to RMSD = 2.15 g kg^{-1}, r^2 = 0.61, and RPD = 1.59. After the addition of two Local cores to SIC Moistcore Regional calibrations, validation RMSD decreased from 9.38 to 6.14 g kg^{-1}, r^2 increased from 0.27 to 0.67 and RPD increased from 1.15 to 1.71.

In-Situ Local Calibration Model Validation. Moistcore and Drycore preparation methods gave less accurate PLS validation results than the other preparation methods, with Moistcore being the least accurate. Moist- and Drycore SOC prediction results, with the addition of two Local cores, included RMSD values of 2.89 and 2.76 g kg^{-1} and r^2 values ranging from 0.37 and 0.42, respectively. For SIC predictions, RMSD values were higher, with 6.14 g kg-1 for Moist- and 5.11 for Drycore, r^2 of 0.67 for Moist- and 0.78 for Drycore, and RPD of 1.71 for Moist- and 2.05 for Drycore.

Regional2 Calibration Model Validation. Increasing the calibration data set with the Regional2 dataset did not appear to improve Regional SOC or SIC PLS validation results as much as the addition of two Local cores did. Organic C predictions, using the Sieved Regional + 2 Local cores + Regional2, gave RMSD = 2.26 g kg^{-1}, r^2 = 0.57 and RPD = 1.51. These results are less accurate than the Sieved validation results using Regional + Local calibrations. SIC prediction accuracy decreased with an additional

local dataset giving RMSD of 3.43 g kg^{-1}, r^2 of 0.90, and RPD of 2.99, which was not as accurate as the immediate Regional + two Local core calibration model results.

Global Calibration Model Validation. Global BRT calibration validation results do not show improvement over PLS Regional + Local calibration validation results (Tables 3.8 and 3.9). Global calibration validations, without Local core addition, are generally more accurate than their PLS counterparts. For all preparation methods, adding cores to a global calibration set from the site to be predicted does not significantly increase accuracy whereas in the PLS Regional modeling procedure, adding Local cores makes a noteworthy improvement resulting in more accurate final predictions (Figs. 3.6 and 3.7). The most accurate Global SOC validation results were from Milled spectra calibration without the addition of Local cores, providing a RMSD of 2.20 g kg^{-1}, r^2 of 0.60, and RPD of 1.56. The most accurate Global SIC validation results were from Sieved spectra calibration without the addition of Local cores, giving a RMSD of 4.0 g kg^{-1}, r^2 of 0.88, and RPD of 2.71.

Table 3.8: Global calibration validation statistics for organic C from models developed using four different sample preparations.

	Soil Organic Carbon			
	Local cores added	RMSD (g kg^{-1})	r^2	RPD
Global + Moistcore	0	3.63	0.04	0.91
	1	3.20	0.18	1.05
	2	3.13	0.19	1.08
Global + Drycore	0	4.06	0.14	0.85
	1	3.41	0.26	1.01
	2	3.48	0.24	1.00
Global + Crushed	0	2.74	0.43	1.25
	1	2.68	0.42	1.29
	2	2.51	0.48	1.37
Global + Milled	0	2.20	0.60	1.56
	1	2.24	0.59	1.53
	2	2.21	0.59	1.53

Table 3.9: Global calibration validation statistics for inorganic C from models developed using four different sample preparations.

	Soil Inorganic Carbon			
	Local cores added	RMSD (g kg^{-1})	r^2	RPD
Global + Moistcore	0	10.56	0.18	1.07
	1	8.30	0.45	1.34
	2	7.60	0.51	1.43
Global + Drycore	0	6.18	0.75	1.75
	1	6.21	0.75	1.72
	2	5.96	0.75	1.76
Global + Crushed	0	4.00	0.88	2.71
	1	4.01	0.87	2.67
	2	3.96	0.86	2.59
Global + Milled	0	4.73	0.86	2.29
	1	4.33	0.89	2.46
	2	4.29	0.86	2.38

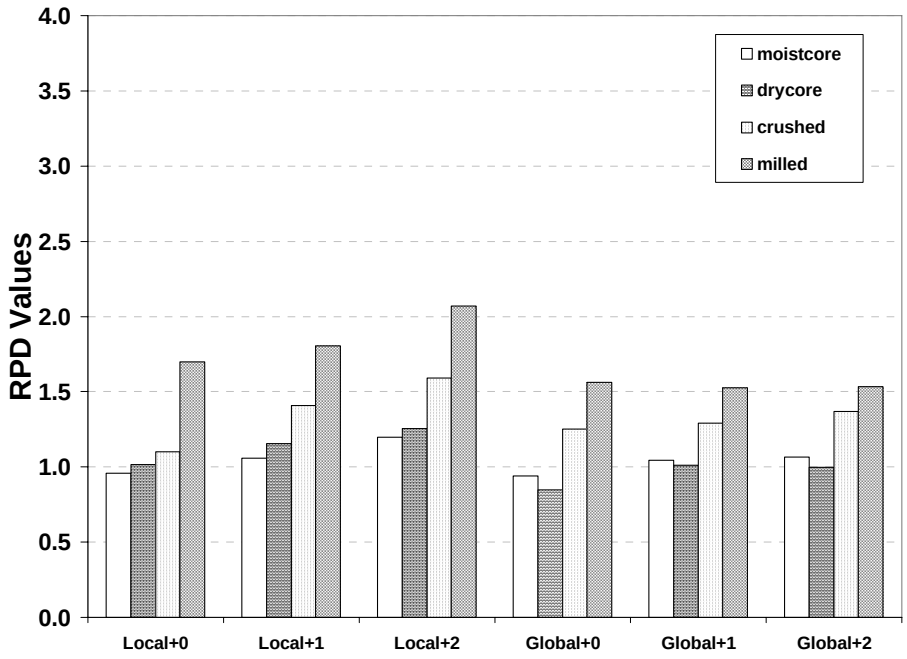

Figure 3.6. RPD values for organic C validations of Regional and Local calibrations developed from 4 different sample preparations in comparison to Global and Global + Local calibration validations in 4 different preparations.

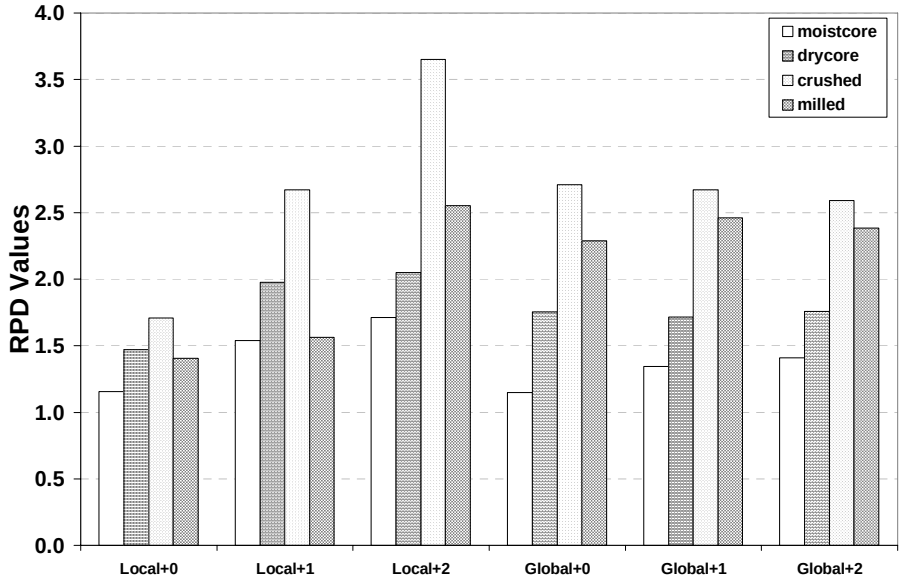

Figure 3.7. RPD values for inorganic C validations of Regional and Local calibrations developed from 4 different sample preparations in comparison to Global and Global + Local calibration validations in 4 different preparations.

When the Regional2 dataset was added to the Global + Regional calibration model, results were not improved from the Global + Regional validations or from PLS-Regional validations (Tables 3.10 and 3.11). The most accurate Global + Regional + Regional2 SOC validations resulted from Milled soils with 2 Local cores added, with RMSD of 2.37 g kg^{-1}, r^2 of 0.53 and RPD of 1.43. The most accurate SIC predictions resulted from Sieved soils without Local core addition, giving RMSD of 4.02 g kg^{-1}, r^2 of 0.87, and RPD of 2.69.

Table 3.10: Validation statistics from Global + Regional + Regional2 calibrations for organic C.

	Soil Organic Carbon			
	Local cores added	RMSD (g kg^{-1})	r^2	RPD
Global + Regional2 + Moistcore	0	3.53	0.09	0.94
	1	3.22	0.16	1.05
	2	3.16	0.21	1.07
Global + Regional2 + Drycore	0	3.59	0.20	0.96
	1	3.62	0.19	0.95
	2	3.43	0.26	1.01
Global + Regional2 + Crushed	0	2.48	0.49	1.38
	1	2.43	0.52	1.42
	2	2.35	0.53	1.45
Global + Regional2 + Milled	0	2.42	0.52	1.42
	1	2.39	0.53	1.43
	2	2.37	0.53	1.43

Table 3.11: Validation statistics from Global + Regional + Local calibrations for inorganic C.

	Soil Inorganic Carbon			
	Local cores added	RMSD ($g\ kg^{-1}$)	r^2	RPD
Global + Regional2 + Moistcore	0	9.83	0.26	1.15
	1	8.25	0.45	1.34
	2	7.70	0.50	1.41
Global + Regional2 + Drycore	0	5.90	0.77	1.83
	1	6.03	0.75	1.77
	2	5.80	0.75	1.81
Global + Regional2 + Crushed	0	4.02	0.87	2.69
	1	4.11	0.87	2.60
	2	3.97	0.86	2.58
Global + Regional2 + Milled	0	4.87	0.85	2.22
	1	4.64	0.85	2.30
	2	4.38	0.86	2.33

PLS SOC Loadings

Loadings were determined for the Milled + 2 Local cores SOC calibration model and the weighted average of these loadings is displayed in Fig. 3.8. The first component explained 55% of the variability and its highest loadings are within the visible region as well as at 1900 nm. The weighted average of the first ten components shows the highest peaks at 2360, 1900, 2020, 2180, 1400-1430, and at 530 nm.

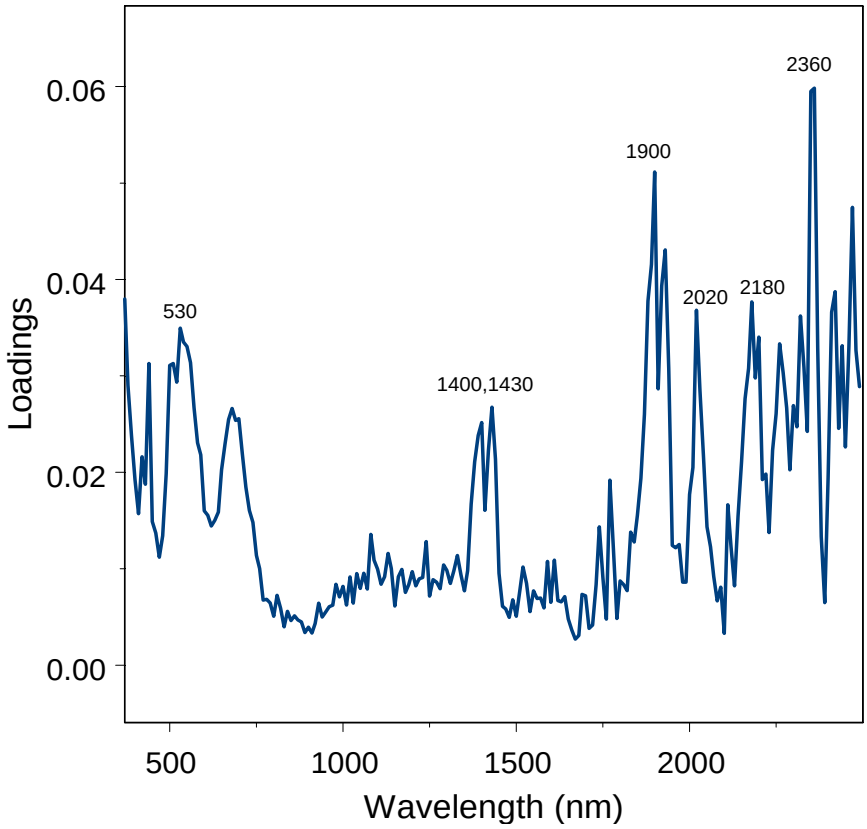

Figure 3.8. Loadings of the SOC calibration model developed with Milled sample presentation with each line representing components that explain the presented percentage of variability.

Principal Component Analysis

All plots were similar in nature, showing the same six outlying samples from Site 2 as well as the same three outlying samples from Site 5 and one Site 1 outlier (Fig. 3.9; for clarity, only the first 6 components are displayed). The spectra, carbon content and texture were checked for each sample; in all Site 2 cases, these samples contained higher SIC content than other samples. For Sites 5 and 1, these samples were from a greater depth (close to 1 m) than the other samples; this influences soil composition due to a lack of terrestrial ecosystem influences (weather, leaching, organic matter, etc.). Aside from

outliers, another point of interest within the PCA diagram was the spectral influence by location. Samples were generally grouped together according to site more than for any other soil characteristic.

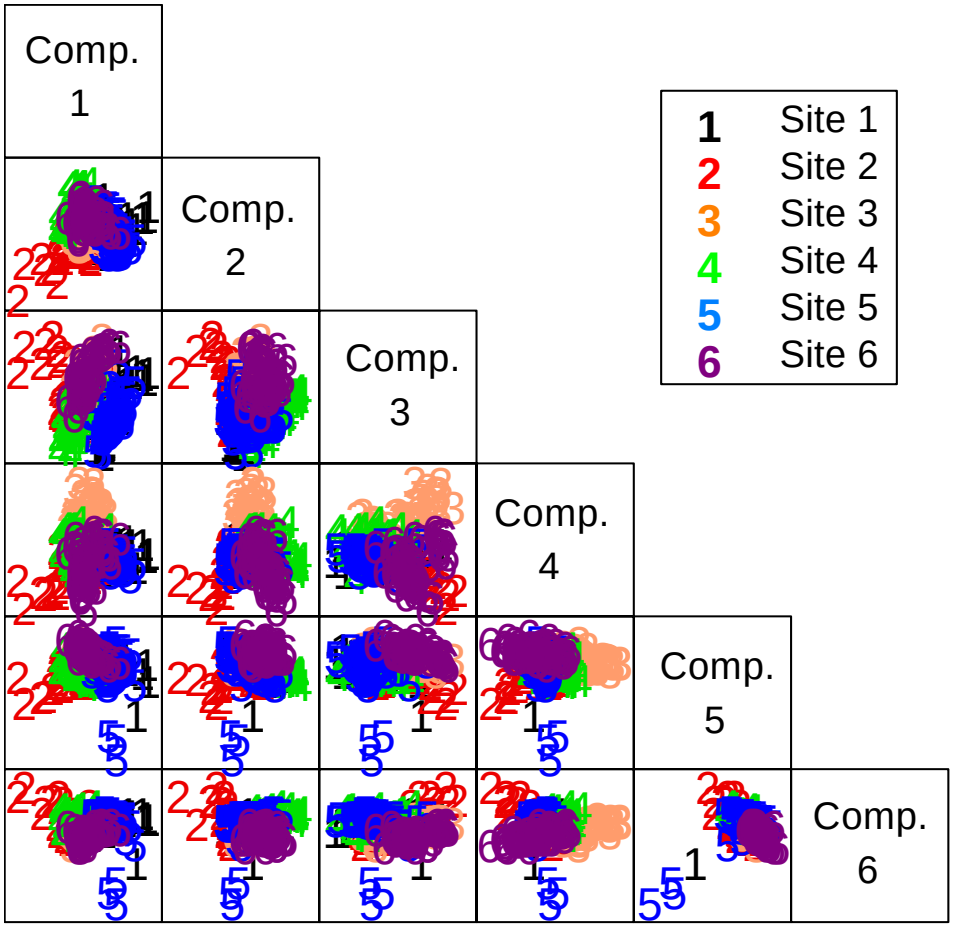

Figure 3.9. Scatterplot matrix diagram showing the first six principal components explaining 75% of the variability of the spectral response of the "Crushed" sample preparation method.

Preparation Methods

Three samples were compared spectrally with four different preparation methods (Figs. 3.10-3.12). Samples included were: sample 2 - holding the highest SIC content

(SIC = 72.5 g kg^{-1}, SOC = 9 g kg^{-1}), sample 4 - no SIC content but relatively high SOC content (SIC = 0, SOC = 17.2 g kg^{-1}), and sample 5 – containing the highest SOC content (SIC = 1 g kg^{-1}, SOC = 19.1 g kg^{-1}). These samples were qualitatively examined together to decipher spectral influences due to $CaCO_3$ and SOM content as well as to determine the effect sample preparation has on different sample types.

Sample 2 stood out as the most spectrally unique as a result of $CaCO_3$ influences (Fig. 3.10). The three samples chosen for comparison all held structural water and OH absorptions at 1400 and 1900 nm, as is stated in the literature (Hale and Quarry, 1973; Baumgardner, 1985; Ben-Dor and Banin, 1995; Workman, 1998; Clark, 1999). Another absorption shared by all three samples was at 2290 nm, which is interpreted to be due to clay/Al-OH stretching (Clark, 1999). In the visible region, sample 2 held the same reflectance features but higher reflectance values overall due to carbonate content. At 2360 nm, there is a large peak only for sample 2 while the other two samples only show relatively small absorptions.

The Moistcore preparation technique appeared the most spectrally unique as a result of water spectral influences. The most pronounced OH and H_2O absorbance features can be found at approximately 1400 and 1900 nm. Reflectance peaks at approximately 1500 and 2100 nm along with an absorbance trough at 1750 nm correspond to features in the spectra that have shifted to higher wavelengths due to water influences. The Milled preparation method appears to be the most visibly distinguished in the visible range of the spectrum, showing much higher reflectance than other methods.

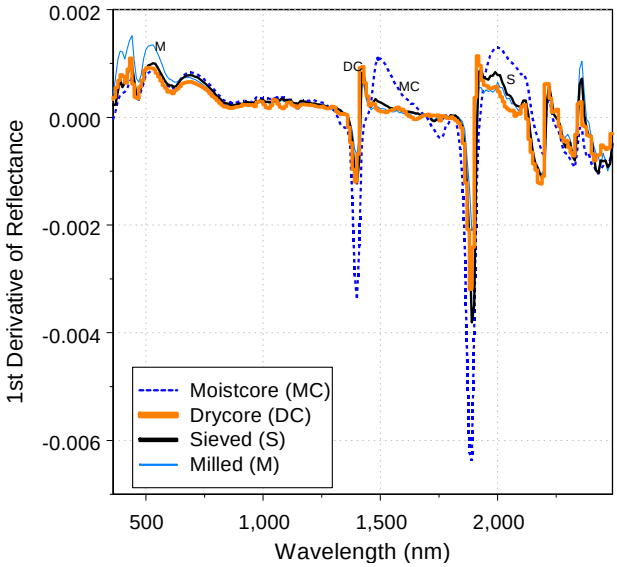

Figure 3.10. First derivative spectra of sample 2 (highest SIC content) in four different preparations.

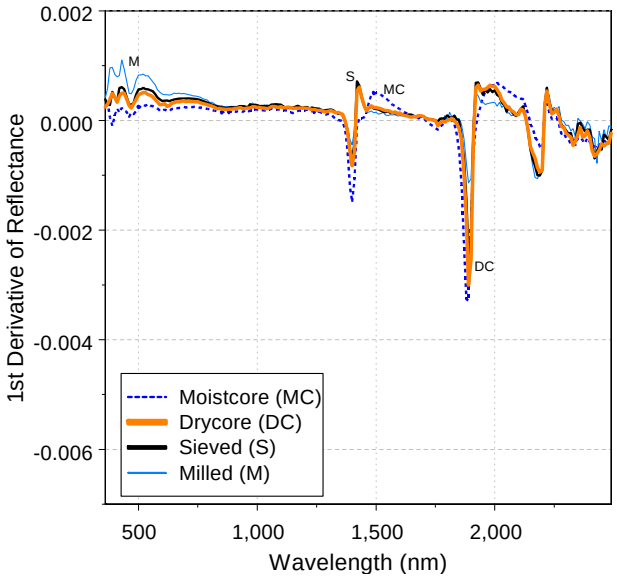

Figure 3.11. First derivative reflectance spectra of sample 4 (high SOC, no SIC) in four different preparations.

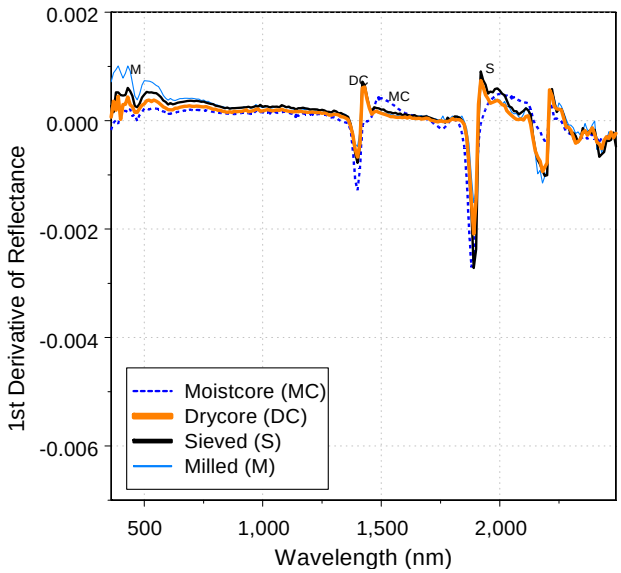

Figure 3.12. First derivative reflectance spectra of sample 5 (highest SOC content) in four different preparations.

Organic Matter Removal

Spectra of the same three samples with and without organic matter are plotted together and for visual comparison (Fig. 3.13). First derivative spectral signatures beyond 1000 nm are overlapping for samples with and without SOM; therefore, only reflectance spectra are displayed. Sample 2 has the highest SIC content, sample 4 has no SIC content and relatively high SOC content, and sample 5 has the highest SOC content. The most apparent difference displayed is higher reflectance by all samples after SOM has been removed.

Reflectance difference between samples with and without organic matter showed the largest difference in the visible region. Through grouping difference values by SOC content (Fig. 3.14), I was expecting to find the largest difference in spectral reflectance corresponding with greatest SOC content within areas of the spectrum pertinent to SOM

constituent signatures. The most influential place where this occurs is in the visible portion. More difference is observed here with higher SOC content groups. In the remainder of the spectrum, there isn't as much change. The correlation analysis shows the same results with highest correlation at 630 nm (Fig. 3.15).

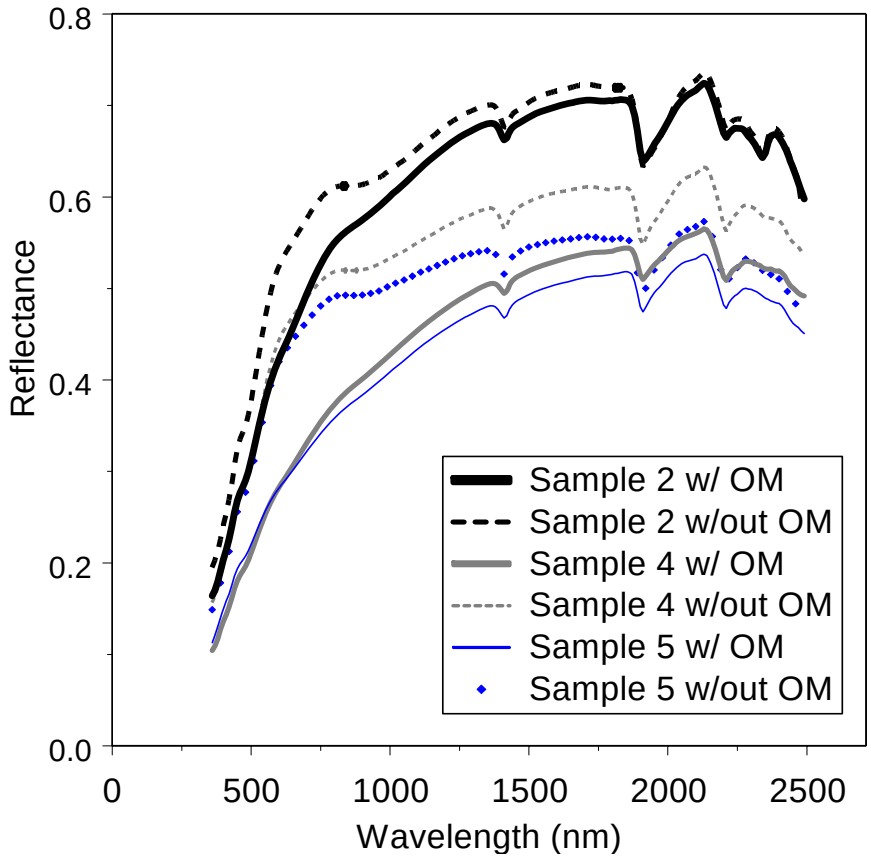

Figure 3.13. Reflectance spectra of three samples with and without organic matter.

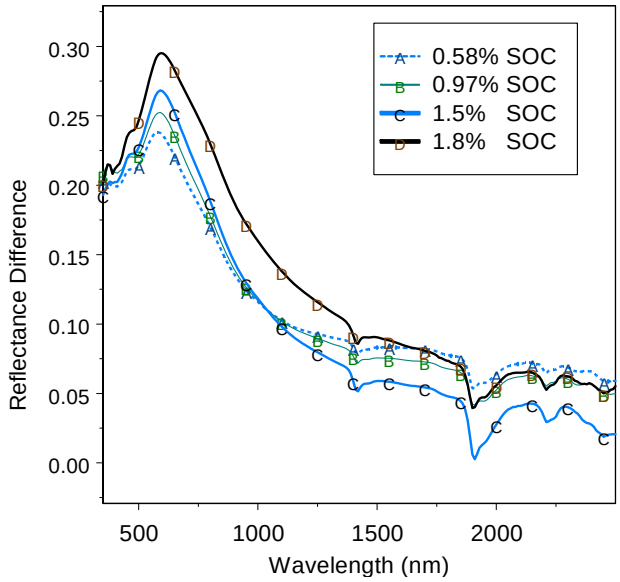

Figure 3.14. Reflectance difference between samples with and without organic matter grouped by average organic C content.

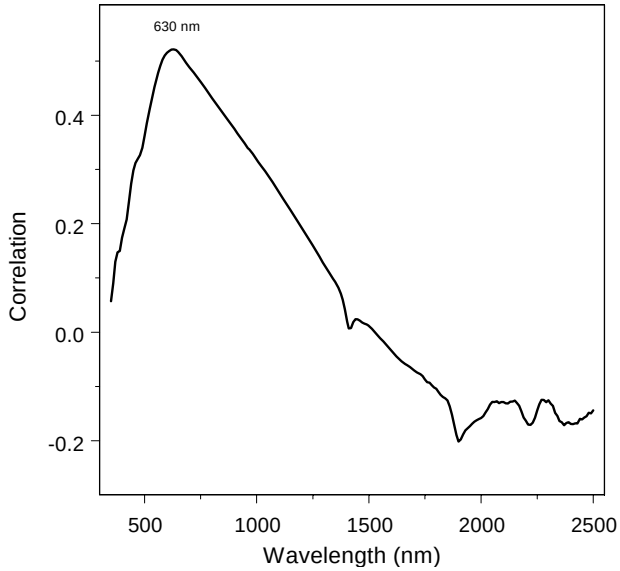

Figure 3.15. Correlation between spectral reflectance of samples with and without organic matter and SOC content.

Continuum Removal

Continuum removal analysis showed correlation between SIC and continuum removed spectra from 2300-2350 nm. Multiple linear regression results gave an r^2 value of 0.83. Continuum removed spectra averaged by ordinal SIC content groups displays consistency between SIC content and band depth (Fig. 3.16).

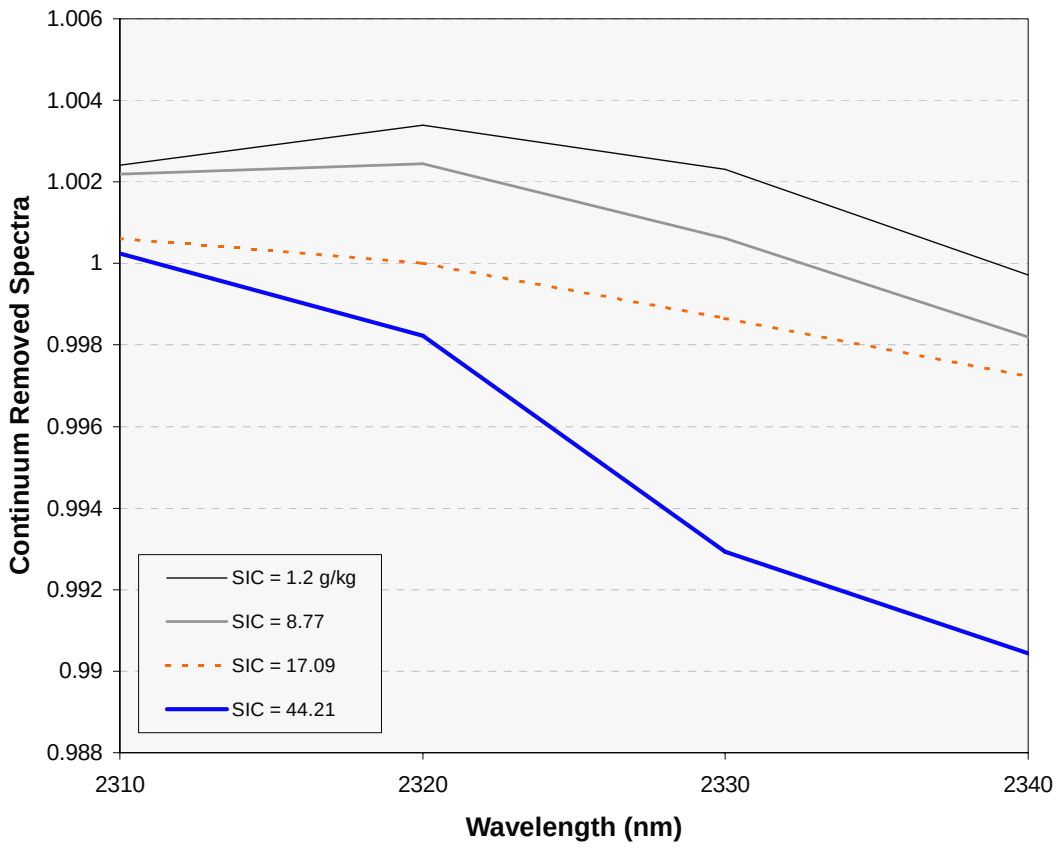

Figure 3.16. Continuum removed spectra from 2300-2350 nm by average SIC content groups.

Discussion

Prediction Results

Based on r^2 and RPD values, SOC predictions are less accurate than SIC predictions. This is a result of the small range of SOC content in our sample set. This smaller range also explains the lower RMSD values resulting from organic C predictions. Inorganic C has a greater range of values within this sample set, thereby increasing r^2, RPD and RMSD values. But, SOC prediction results do show lower RPD and SEP values thereby displaying more accurate predictions overall as a result of this small SOC content spread.

PLS prediction results exhibit Milled and Sieved as the most successful preparation methods for SOC and SIC predictions, respectively. The plots of different preparation methods (Figs. 3.10- 3.12) show the major difference between Sieved and Milled spectra in the visible region, which influences SOC predictive success. Particle coatings can dominate reflectance spectra (Clark, 1999); milling soils will disperse particle coatings thereby negatively influencing carbonate spectral signatures as well as SIC predictions.

In-situ simulated predictions, both dry and field-moist, proved to be less successful than *in-vitro* preparation methods. Organic C predictions gave low r^2 and RPD values, without much of a difference between moisture contents after the addition of two Local cores. RMSD values were all under 3.75 but with low r^2 and RPD values, SOC *in-situ* predictions can be classified as semi-quantitative. Inorganic C results

showed a stronger correlation between predicted and measured values but exhibited higher RMSD values (> 5.0), indicating some imprecision. After the addition of two Local cores, moisture content does not appear to influence predictive accuracy very much. These SIC *in-situ* prediction results are quantitative and indicate potential for effective application to on-site measurements.

Increasing the calibration set size with additional regional and global samples did improve predictive accuracy, though not as much as the addition of Local cores did. It can be interpreted that the most influential factor for prediction of this particular dataset is the inclusion of one or two extreme Local samples, rather than increasing the calibration set size with similar soil samples (Regional2) or adding a large amount of diverse samples. "Quality not quantity" applies here. Shepherd and Walsh (2002) conveyed more accurate predictions from larger calibration sets while examining calibrations made up from spectral data of highly weathered African soils. The soils analyzed in my study are relatively young, explaining more accuracy from site-specific calibrations rather than from increasing the calibration set size. If Local core additions are not available, the Global calibration appears to lead to most accurate predictions

The most accurate SOC and SIC Local PLS predictions gave SEP values conveying predictive potential comparable to laboratory accuracy when compared to carbon SEL values. Though SEP values for the PLS prediction method were more than triple the SEL values in this analysis, they still were less than 3 g kg^{-1} for both SOC and SIC, thereby displaying predictive potential.

Principal Component Analysis

PCA results show that the dominating influence on spectral similarity is immediate location or field. Though the samples are taken from six different sites within one physiographic region, samples were grouped to an extent by site. This conveys the importance of sample location when constructing calibrations for relatively young calcareous soils as the locality influence can affect model robustness.

Spectral Interpretation

After visually inspecting three samples, the four different preparation techniques appeared comparable showing the same major features within the Vis-NIR portion of the spectrum. Moisture and particle size seemed to be the most influential factors. In the scanning of field-moist cores, water influences the spectral signature significantly, which could explain the less successful prediction results. Milled sample reflectance features appear more distinct due to smaller sample particle size, which has been found to increase reflectance (Clark, 1999). Aggregate size difference between Sieved and Drycore does not appear to visually influence spectra that much, though predictive accuracy conveys a different message. Although spectral signatures follow each other consistently, predictions were less accurate all around for Drycore spectra calibrations.

Organic Matter Qualitative Interpretation

Difference between spectra with and without organic matter showed the greatest reflectance difference in the visible region. The highest correlation between the spectral difference of soils with and without SOM and SOC was also found in the visible region.

And finally, some of the higher PLS loadings were given within the visible region as well. According to Workman (1998), the region from 800-1000 nm is sensitive to organic third overtones. We might expect soils with organic matter to be more absorbing here except for the fact that third overtones are very difficult to distinguish, especially in spectra of soil in its entirety. As organic constituents are diluted within the soil matrix, these third overtone features are nearly impossible to decipher. There are iron oxide absorptions in this region at 870 and 1000 nm (Baumgardner et al., 1985), though very broad and non-distinct. A key finding here is that SOM correlation to the visible region can be explained by electronic absorptions of organic constituents rather than vibrational.

PLS Loadings

The strongest loadings within the visible region are explained by the color darkening effect of SOM on soils (Baumgardner, 1985) as well as the electronic absorptions explained above. The weight given to the 1900 nm band is related to structural water content (Clark et al., 1990), which would be relevant to clay content. The features at 2020 and 2180 nm are both organic stretching regions. At 2350 there is a distinct humic acid absorption (Chang et al., 2002) although this is most likely undetected behind more distinct carbonate absorptions at 2350 nm. This conveys indirect correlation between SOC predictions and carbonate content. At 2460 there is an aromatic CH absorption (Weyer and Lo, 2002). This loading weight analysis provides relevant information proving that there is indeed a relation between the spectral information and SOM.

Continuum Removal

Continuum removal analysis displayed a relation between band depth, or absorption, and particular soil constituents. Both the quantitative (regression) as well as the qualitative showed correlation between SIC content and band depth of this particular feature. The order of ordinal SIC content groups (Fig. 3.16) provides a visual of increasing band depth with increasing SIC content. Although the band depths appear consistent, a lower multiple regression r^2 value (< 0.90) could be a result of other influences in this region, including several organic absorptions between 2300-2350 nm (Elvidge, 1990; Workman, 1998; Chang et al., 2002).

Conclusions

The results of this study suggest that previous publications have presented over-optimistic interpretations of C predictive accuracy by Vis-NIR PLS methods. Studies claim "independent" validation while using a random selection procedure, forgetting to take spatial correlation into account. With accurate results for glacial till soil calibrations and a completely independent site-by-site validation process, Vis-NIR analysis shows potential for quantitative analyses of soil C. The results display that future application to young calcareous soils would obtain most accurate predictions using a regional calibration set in conjunction with spectrally diverse samples from the site to be predicted. *In-situ* analysis of soil C showed potential for semi-quantitative analysis of these complex soils; it is possible that investigation of other soil characteristics or perhaps another soil dataset could lead to more successful predictions. For soil C

predictions with this particular dataset, *in-vitro* analysis produced more predictive accuracy than *in-situ* methods.

Mixed mineralogy as well as a calcareous parent material gave this particular dataset a localized spectral fingerprint. Regional + Local calibrations proved to be the most successful method of prediction for these soils over Regional and Global. The addition of a limited number of samples from the validation site enhanced predictions using a regional calibration and could lead to accurate predictions of similar soils from new sites.

Vibrational absorptions within Vis-NIR soil spectra have been thought to be the major influence due to SOM. Within this study, electronic absorptions were found to be the key absorptions due to SOM content. The spectral difference between soils with and without SOM was most correlated to SOC content at 630 nm. Loadings from PLS SOC calibration models were also high within the visible region of the spectra. Future investigations focused on spectral features due to organic constituents should focus within this region of the spectra as opposed to further along, where vibrational absorptions occur.

References

Al-Abbas, A.H., Swain, P.H. and Baumgardner, M.F., 1972. Relating organic matter and clay content to multispectral radiance of soils. Soil Science, 114 (6): 477-485.

Baumgardner, M.F., Silva, L.F., Biehl, L.L. and Stoner, E., 1985. Reflectance properties of soils. Advances in Agronomy, 38: 1-45.

Ben-Dor, E. and Banin, A., 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Sci. Soc. Am. J, 59: 364-372.

Ben-Dor, E., Inbar, Y., and Chen, Y., 1997. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400-2500 nm) during a controlled decomposition process. Remote Sens. Environ., 61: 1-15.

Brown, D.J., Brickelmeyer, R.S., and Miller, P.R., 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C predictions in Montana. Geoderma, 129: 251-267.

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

Burrough, P.A., Beckett, P.T., and Jarvis, M.G. Relation between cost and utility in soil survey. J. of Soil Science, 22 (3): 359-&.

Chang, C.W., Laird D.A., Mausbach M.J., and Hurburgh C.R., 2001. Near-infrared reflectance spectroscopy – principal components regression analyses of soil properties. Soil Sci. Soc. Am. J., 65: 480-490.

Chang, C.W. and Laird, D.A., 2002. Near-infrared reflectance spectroscopic analysis of soil C and N. Soil Science, 167(2): 110-116.

Clark, RN and Roush, TL, 1984. Reflectance Spectroscopy – quantitative analysis techniques for remote sensing applications. J. of Geophysical Research, 89 (NB7): 6329-6340.

Clark, R.N., King, T.V., Klejwa, M., and Swayze, G., 1990. High Spectral Resolution Reflectance Spectroscopy of Minerals. J. of Geophysical Research, 95(B8): 12653-12680.

Clark, R.N., 1999. Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy. In: N. Rencz (Editor), Remote Sensing for the Earth Sciences: Manual of Remote Sensing, 3 ed. John Wiley & Sons, New York, pp. 3-52.

Couteaux, M.M., Berg, B. and Rovira, P., 2003. Near infrared reflectance spectroscopy for determination of organic matter fractions including microbial biomass in coniferous forest soils. Soil Biology and Biochemistry, 35 (12): 1587-1600.

Dalal, R.C. and Henry, R.J., 1986. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. Soil Sci. Soc. Am. J., 50: 120-123.

Daniel, KW, Tripathi, NK, and Honda, K, 2004. Analysis of VNIR spectral signatures for estimation of soil organic matter in tropical soils of Thailand. International J. of Remote Sensing, 25 (3): 643-652

Davies, A.M.C., Dennis, C., Grant, A., Hall, M.N., Robertson, A., 1987. Screening of tomato puree for excessive mould content by near infrared spectroscopy: a preliminary evaluation. J. Sci. Food Agric., 39: 349-355.

Dick, DP, Santos, JHZ, and Ferranti, EM, 2003. Chemical characterization and infrared spectroscopy of soil organic matter from two southern Brazilian soils. Revista Brasileira de Ciencia do Solo 27 (1): 29-39.

Duckworth, J. Spectroscopic Quantitative Analysis in *Applied Spectroscopy: A Compact Reference for Practitioners.* Workman, J. and Springsteen, A. (ed.), Academic Press, Massachusetts, 1998.

Dunn, B.W., Beecher, H.G., Batten, G.D., and Ciaverella, S., 2002. The potential of near-infrared reflectance spectroscopy for soil analysis – a case study from the Riverine Plain of south-eastern Australian. Australian Journal of Experimental Agriculture, 42: 607-614.

Ellerbrock, R.H. and Gerke, H.H., 2004. Characterizing organic matter of soil aggregate coatings and biopores by Fourier transform infrared spectroscopy. European J. of Soil Sci., 55 (2): 219-228.

Elvidge, C.D., 1990. Visible and near-infrared reflectance characteristics of dry plant materials. International J. of Remote Sensing, 11 (10): 1775-1795.

Fidencio, P.H., Poppi, R.J., de Andrade, J.C., and Cantarella, H., 2002. Determination of organic matter in soils using near-infrared spectroscopy and partial least squares regression. Communications in Soil Science and Plant Analysis, 33 (9-10): 1607-1615.

Haaland, D.M. and Thomas, E.V., 1988. Partial least-squares methods for spectral analyses .1. relation to other quantitative calibration methods and the extraction of qualitative information. Analytical Chemistry, 60 (11): 1193-1202.

Hale, GM and Querry, MR, 1973. Optical constants of water in 200 nm to 200 mum wavelength region. Applied Optics, 12 (3): 555-563.

Islam, K., Singh, B., and McBratney, A., 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. Australian Journal of Soil Research, 41: 1101-1114.

Martens H and Naes T, 1987. Multivariate calibration by data compression. John Wiley and Sons: Chichester, UK.

McCarty, G.W., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Sci. Soc. Am. J, 66: 640-646

McKissock, I., Gilkes, R.J. and van Bronswijk, W., 2003. The relationship of soil water repellency to aliphatic C and kaolin measured using DRIFT. Aust. J. of Soil Res., 41 (2): 251-265.

Morra, M.J., Hall, M.H., and Freeborn, L.L., 1991. Carbon and nitrogen analysis of soil fractions using near-infrared reflectance spectroscopy. Soil Sci. Soc. Am. J., 55: 288-291.

Osborne, B.G., Fearn, T., Miller, A.R., and Douglas, S., 1984. Application of near-infrared reflectance spectroscopy to the compositional analysis of biscuits and biscuit doughs. J. of Sci. of Food and Ag., 35 (1): 99-105.

Reeves III, J.B., McCarty, G.W., and Mimmo, T., 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. Environmental Pollution 116: S277-S284.

Reeves III, J.B., Francis, B.A. and Hamilton, S.K., 2005. Specular reflection and diffuse reflectance spectroscopy of soils. Applied Spectroscopy, 59 (1): 39-46.

Rumpel, C., Janik, L.J., Skjemstad, J.O., and Kogel-Knabner, I., 2001. Quantification of carbon derived from lignite in soils using mid-infrared spectroscopy and partial least squares. Organic Geochemistry, 32: 831-839.

Schmoldt, D.L., 2001. Precision agriculture and information technology. Computers and Electronics in Agriculture, 30 (1-3): 5-7.

Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Progress in Physical Geography, 27(2): 171-197.

Shepherd, K.D. and Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Sci. Soc. Am. J., 66: 988-998.

Sherrod, LA, Dunn, G, Peterson, GA, and Kolberg, RL, 2002. Inorganic carbon analysis by modified pressure-calcimeter method, Soil Sci. Soc. Am. J., 66 (1): 299-305.

Stark, E. and Luchter, K., 1986. Near-infrared analysis (NIRA) – a technology for quantitative and qualitative analysis. Applied Spectroscopy Reviews, 22 (4): 335-399.

Sudduth, K.A. and Hummel, J.W., 1993. Portable, near-infrared spectrophotometer for rapid soil analysis. Transactions of the ASAE, 36(1): 185-193.

Sylvester-Bradley, R., Lord, E., Sparkes, D.L., Scott, R.K., Wiltshire, J.J.J., and Orson, J., 1999. An analysis of the potential of precision farming in Northern Europe. Soil Use and Management, 15 (1): 1-8.

Tkachuk, R., 1981. Oil and protein analysis of whole rapeseed kernels by near infrared reflectance spectroscopy. JAOCS, pp. 819-821.

Whalley, W.R. and Leeds-Harrison, P.B., 1991. Estimation of soil moisture status using near infrared reflectance. Hydrological Processes, 5: 321-327.

Whittig, L. D., and Allardice, W. R., 1986. X-ray diffraction techniques. Methods of soil analysis, part 1. Physical and mineralogical methods-agronomy monograph no. 9 (2 ed). Soil Sci. Soc. Amer. Madison, WI., pp. 331-362.

Williams, P.C. and Norris, 1985. Determination of protein and moisture in wheat and barley by near-infrared transmission. J. Agric. Food Chem., 33: 239-244.

Wold, S., Sjostrom, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics Intelligence Laboratory Systems, 58: 109-130.

Workman, J., 1998. Ultraviolet, Visible and Near-Infrared Spectroscopy. In: Workman, J. and Springsteen, A. (Eds.), Applied Spectroscopy – A Compact Reference for Practitioners, Academic Press, Chestnut Hill, MA, pp. 29-47.

CHAPTER 4

MID-INFRARED DIFFUSE REFLECTANCE SPECTROSCOPY FOR CHARACTERIZATION OF GLACIAL TILL CALCAREOUS SOILS FROM NORTH CENTRAL MONTANA

Abstract

Diffuse Reflectance Spectroscopy (DRS) is a rapid and relatively inexpensive means of quantifying soil carbon (C). In this study, the mid-infrared region (MIR, 4000-600 cm^{-1}) and Partial least squares (PLS) regression were used together to determine soil organic C (SOC) and soil inorganic C (SIC) content of calcareous glacial till soils from north central Montana. A set of 315 samples from depths of 0-10, 10-20, 20-50 and 50-100 cm, was collected from six independent farm sites. SOC contents ranged from 3.5 to 19.29 g kg^{-1} and SIC contents ranged from 0 to 72.29 g kg^{-1}. PLS regression with with 6-fold site-holdout was used to determine predictive accuracy of KBr-diluted and neat samples, as well as different spectral transformations, including Kubelka Munk (KM), first derivative (D) and absorbance (AB). Regional vs. local calibrations were compared by addition of spectrally dissimilar cores from the validation site to the regional calibration model. The most accurate SOC prediction results were found using first derivative transformation with the addition of two local cores providing a root mean squared deviation (RMSD) of 1.3 g kg^{-1}, a coefficient of determination (r^2) of 0.86, and a residual prediction deviation (RPD) of 2.61. The most accurate SIC prediction results

were found with KM transformation with the addition of two local cores providing a RMSD of 1.95 g kg^{-1}, r^2 of 0.96, and RPD of 5.07. The addition of local cores improved predictive accuracy for neat samples with neat + local cores consistently emerging as more accurate predictions than KBr-diluted sample predictions. Qualitative analysis of KBr-diluted and undiluted spectra showed different absorption features between the two, particularly involving soil organic matter (SOM), carbonates and clay absorptions.

Introduction

Knowledge and quantification of the global carbon cycle within the pedosphere has become important as it acts as an atmospheric carbon sink and/or source within the global carbon budget. Evidence points to soils as a net source of atmospheric CO_2 as a result of land use change and agricultural activities over the past 200 years (Post, 2001). Slight increases or decreases in the net flux of CO_2 from soils may have a measurable bearing on atmospheric CO_2 levels (Amundson, 2001) and different land management techniques can easily alter this flux (Post, 2001). For these reasons, a robust and efficient protocol for monitoring soil C is needed (Post, 2001).

Diffuse reflectance spectroscopy (DRS) provides a rapid and inexpensive method of soil characterization and has been applied using both the visible to near infrared (Vis-NIR, 25000-4000 cm^{-1}) and MIR (4000-400 cm^{-1}) portions of the electromagnetic spectrum. Chemometrics are applied to MIR spectroscopy, usually by multivariate analysis techniques such as PLS regression, to develop a means of quantifying soil characteristics (McCarty et al., 2002). Though there have been many MIR-PLS results

published (Janik and Skjemstad, 1995; Janik et al., 1998; Rumpel et al., 2001; Bertrand et al., 2002; Reeves et al., 2002; Siebelec et al., 2004; Madari et al., 2005), none report independent validation. This leads us to question whether these models could be applied to soils outside of the calibration dataset, which is an important goal of model development (Brown et al., 2005). Spatially independent validation is required (Brown et al., 2005) and is lacking in the reported studies.

In previous studies, the MIR region has yielded more accurate PLS calibrations than the Vis-NIR region (McCarty et al., 2002; Reeves et al., 2002; Siebelec et al., 2004; Madari et al., 2005; Viscarra Rossel et al., 2006); in addition, the McCarty et al. (2002) study used independent validation by site to obtain more accurate C predictions from the MIR region. It has been hypothesized that predictive abilities result from the spectral features found in each region (Janik et al., 1998; McCarty et al., 2002); the MIR portion includes fundamental molecular absorptions as opposed to the less distinct second and third overtone absorptions found in the Vis-NIR portion (McCarty et al., 2002), which could explain more accurate predictions. Yet, with only one study utilizing independent validation for comparison, it is difficult to interpret applicability of these models as well as the efficiency and accuracy of the DRS method.

Specular reflection can distort or even invert (Reststrahlen effect) strong fundamental absorptions found in the MIR region (Nguyen et al., 1991; Perkins, 1993; Janik, 1998) unless samples are diluted with KBr. KBr is extremely hygroscopic and moisture contamination can alter spectral features, as can the physical pellet forming process (Baes and Bloom, 1989). Because of the inconvenience and problems associated

with the KBr sample dilution, neat sample spectral analysis has been investigated and determined qualitatively comparable (Nguyen, 1991; Reeves, 2003). Since these findings, neat sample presentation has been used regularly (Janik and Skjemstad, 1995; McCarty et al., 2002; Viscarra Rossel et al., 2006), but predictive capabilities of neat soil samples vs. diluted soil samples have not been thoroughly examined as of yet. Bertrand et al. (2002) looked at calibration results and found undiluted calibrations more accurate for certain constituents, and less accurate for others. In particular, SOC calibrations were more accurate for KBr-diluted samples but SIC calibrations were equivalent for both preparation methods. Further investigation is needed on the effects of sample dilution on predictive capabilities, particularly within independent predictions.

Spectral pretreatments most commonly used include: baseline correction, first and second derivatives, mean centering and Kubelka-Munk (KM) transformations. Oftentimes, pretreatment methods are mentioned but not discussed in detail (McCarty et al., 2002; Reeves et al., 2002; Siebelec et al, 2004; Madari et al., 2005). The KM transformation is applied to derive a linear relation between reflectance and analyte concentration. It is also discussed in the literature as a method to reduce diffuse scattering of soil (Coates, 1998; Workman, 2004) but has been claimed to reduce predictive accuracy of forages (Reeves and Reeves, 2002). Predictive accuracy of various soil constituents should be affected differently by assorted spectral transformations as they encompass a range of objectives. To my knowledge, the effects of spectral pretreatments on MIR-PLS modeling, including the KM transformation, have not been systematically studied or reported in the published soils literature thus far.

Several research questions will be addressed in this study including

1.) Is it possible to develop a robust soil C prediction model for calcareous glacial till soils of north central Montana using local and regional MIR calibrations? If so, which calibration set develops the more accurate prediction model?
2.) Can KBr-diluted calibrations produce accurate predictions? Does KBr dilution reduce specular reflection?
3.) Are SOC predictions based on absorptions of SOM in the MIR region or on other soil constituents such as carbonates, clay or Fe-oxides?

These questions will be answered with spectral data from the MIR region with the anticipation that overall predictive accuracy of this region can be evaluated for soil C.

Materials and Methods

The methods implemented in this study begin with the collection and processing of soil samples, with and without KBr dilution, as well as the collection and processing of spectral data, including three different spectral pretreatments. Construction of calibration models involved local and regional spectral data, from both diluted and undiluted samples, with different spectral pretreatments. The models were all independently validated using a reserved dataset (in this case, one field site). A flow chart displays an overview of the sample preparations and spectral pretreatments used (Fig. 4.1).

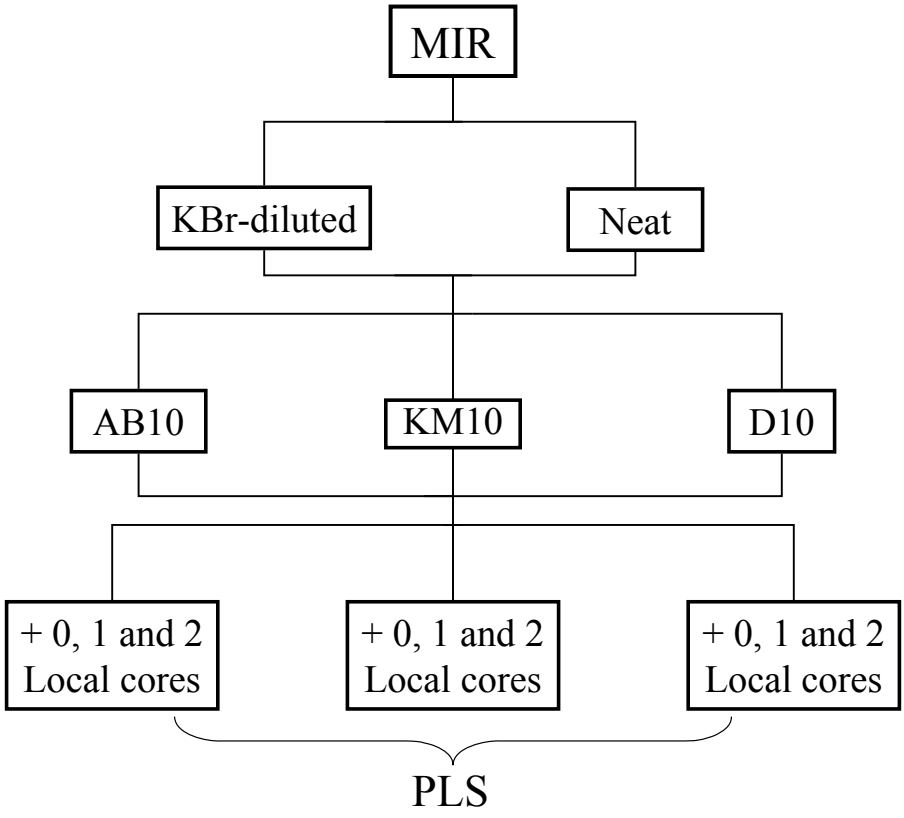

Figure 4.1. An overview of sample preparation, spectral pretreatments, calibration sets and the modeling technique used in the study.

Sampling Design

Study area and Core Selection. Six farm sites located in north central Montana, bounded by the towns of Great Falls, Shelby and Havre (Fig. 4.2), were used as the study area. The soils of these six sites all have originated from glacial deposits that contain similar calcareous parent material (Montagne et al., 1982). The sites are found within ustic moisture and frigid temperature regimes and formerly grew prairie grass but now

have all been converted to wheat cropland. Topography is gently rolling slopes, < 5% in most cases.

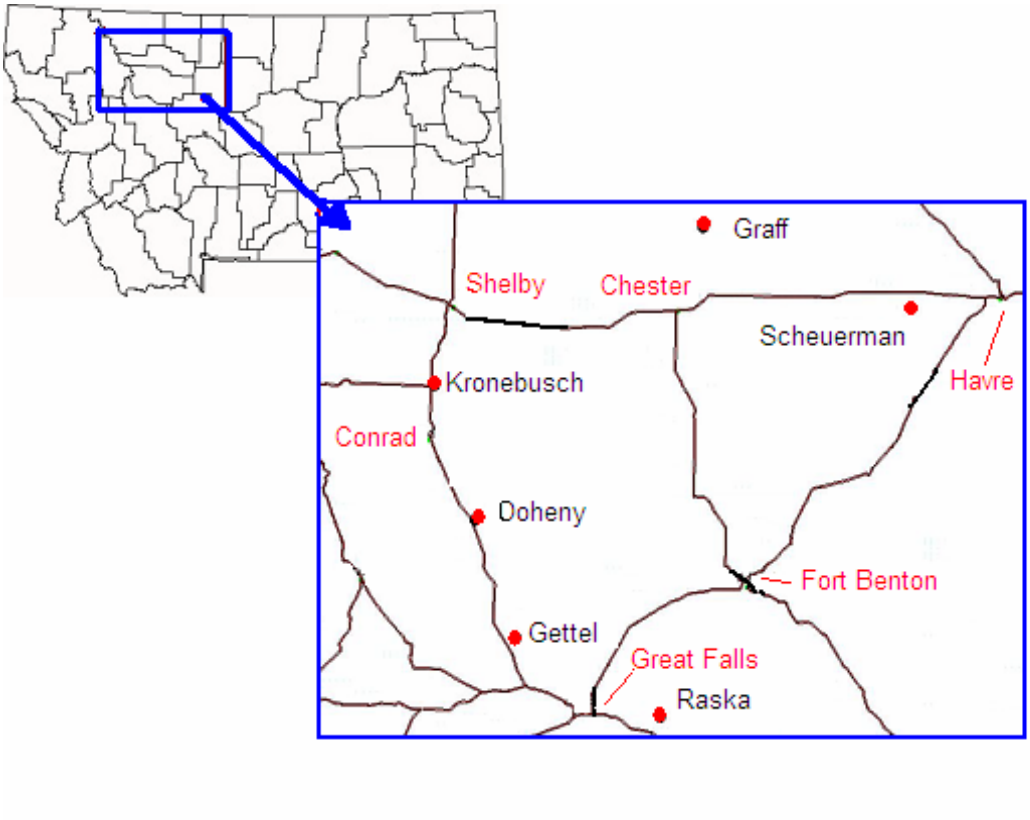

Figure 4.2. A map of Montana including the six sampling sites (identified by symbols labeled with names of farm owners) in the north central "Golden Triangle" region.

At each farm, a field of 32 ha was divided into four segments (8 ha) representing both till and no-till as well as different crop rotation management schemes. Within each segments four points were identified for sampling/monitoring of soil carbon changes over time (part of the CASMGS, Consortium for Agricultural Soils Mitigation of Greenhouse Gases, project). In total, there are 16 sampling points per farm, and at each point there are five cores of 50-100 cm depth taken in a star shaped pattern. The star locations were

determined by landscape position, followed by digging soil pits to determine texture and depth to carbonates. Pits that met depth to carbonate and texture requirements were chosen as core sampling locations. This sampling method was intended to build an accurate representation of the field. Each core was drawn in September 2002 and was divided into three depths: 0-10, 10-20, and 20-50 cm resulting in three samples per core. Two cores per star location were dug to 100 cm depth (where possible) resulting in four samples per core (one additional 50-100 cm sample).

In the end, 480 cores were amassed and a 94-core subset was chosen from them. This subset was constructed with the deepest of the five cores from each designated star site. If there were equivalent core depths at each site, a core was chosen randomly. The final subset consisted of 315 individual soil samples and is referred to as the "Regional" set throughout the study.

Sample Preparation

Four methods of sample preparation were compared in this study: KBr-diluted with SOM, KBr-diluted without SOM, neat (as is) with SOM and neat without SOM. The predictive accuracy as well as the qualitative spectral influence of each preparation method was the basis of comparison. An overview of sample preparation methods is shown in Fig. 4.3.

Neat Samples. After drying core segments at 40 ºC, each sample was passed through a 2-mm sieve to remove rock fragments, surface plant litter, and coarse root material. Approximately six grams of each sieved soil sample were milled to a powdered

homogeneous particle size (<200-µm) for three minutes per sample using a ball mill (Pica Blender Mill Model 2601, Cianflone Scientific Instruments Corp., Pittsburg, PA).

Organic Removal. Organic matter was removed from the sieved samples by treating a 13-g with 20-40 ml of Na-hypochlorite. The treated sample was then placed in a 69 ˚C sonic water bath for 1 hr followed by 2 hrs in a 70 ˚C oven. Water was added to samples as necessary to prevent desiccation. Samples that appeared black after the initial Na-hypochlorite treatment were allowed to settle and clear supernatant was removed, before repeating the treatment again. After SOM had been removed, samples were rinsed four times with deionized water using a centrifuge at 10,000 rpm (25 minutes) and dried overnight in a 70 ˚C oven. After the rinsing and drying process, the soils were milled for three minutes to <200-µm using the ball mill described above.

KBr Dilution. Milled sample portions, both with and without SOM, were homogenized with KBr in a 2% by weight ratio.

Spectroscopy

Instrument and Scanning. Samples with and without SOM were prepared for spectral analysis both with and without KBr dilution. Spectral reflectance scans were obtained with a Bruker Bruker Tensor 27 (2005) infrared spectrometer. The fine-ground samples were scanned in diffuse reflectance mode using aluminum micro-titer plates on a Bruker High-Throughput Screening Device (HTS-XT) using a MCT liquid-N-cooled detector. Light was recorded in absorbance units in the spectral range from 7990-400

cm^{-1} at a spectral resolution of 4 cm^{-1}, zero-filled to a resolution of 2 cm^{-1}. As this study pertains to the MIR region, we selected spectral data from 4000-400 cm^{-1}.

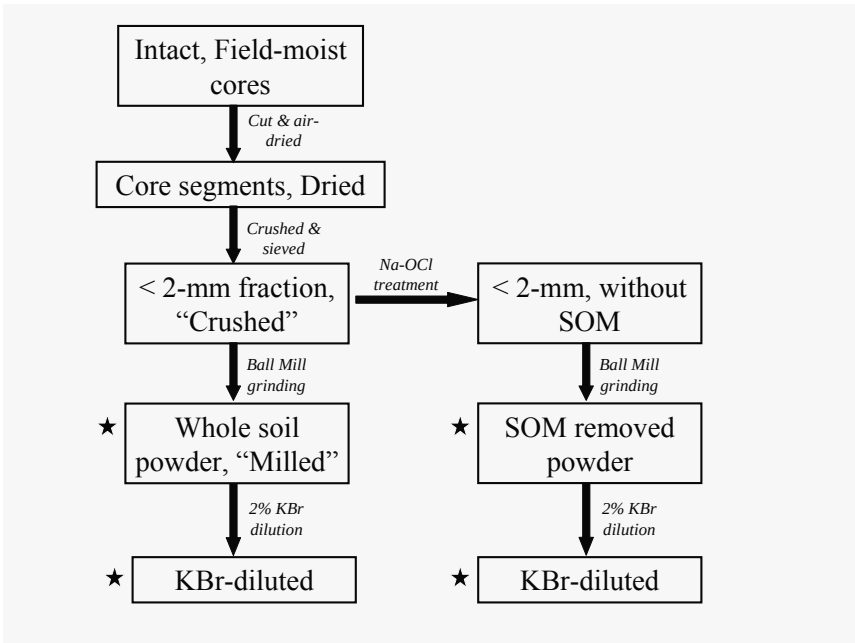

Figure 4.3. Flow chart of different sample preparation methods used with stars representing scanned presentations.

<u>Spectral Pretreatment</u>. Raw absorbance spectra was converted to reflectance data with

$$A = \log\left(\frac{1}{R}\right) \qquad (1)$$

where A is Absorbance and R is reflectance with a natural log transformation. Cubic smoothing splines were fit to reflectance data using "R" (R Development Core Team, 2004) and then transformed back to AB data. A KM transformation was performed on

reflectance data, to partly correct nonlinear intensities resulting from diffuse reflectance (Coates, 1998), using

$$\frac{K}{S} = \frac{\left(1 - R^2\right)}{2R} \qquad (2)$$

where R is reflectance, K is the absorption coefficient and S is the scattering coefficient. First derivative transformations of reflectance spectral data were also computed to account for noise and baseline effects (Duckworth, 1998). Smoothed AB, KM and D transformed data were extracted every 10 cm^{-1} to ease data handling. Multiplicative scatter correction was applied to absorbance data to adjust for particle-size differences but as all samples were milled to uniform size, no predictive improvements were observed and therefore, these results are not reported.

Conventional Laboratory Analyses

Carbon. Total carbon (TC) content was measured in the laboratory through dry combustion using a LECO C/N/S 2000 analyzer (LECO Corporation, St Joseph, MI, USA). Milled samples were oven dried overnight at 70˚ C, and 0.2 g samples of each soil were weighed out in preparation for this analysis. Additional 0.2 g samples were treated with 1 ml of 1 M H_3PO_4, then dried in a 70˚ C oven overnight prior to analysis. Total C of the whole soil less TC of the acidified soil gives the SOC content (assuming the acid removes all IC). Fifteen samples were chosen to obtain duplicate C measurements for standard error of laboratory (SEL) measurements for both SOC and SIC.

Texture. Sand, clay and silt contents were quantified for each sample using the pipette method (Gee and Bauder, 1986) referred to in the previous chapter.

Mineralogy. Clay mineralogy was determined for a subset of 24 samples using powder mounts as well as oriented clay mounts. The subset included 4 samples from each site chosen randomly, two surface and two subsurface samples. Methods and instrumentation used are described in the previous chapter.

Modeling

Model building was done using PLS-1 within Unscrambler 8.05 software (CAMO PROCESS AS, Oslo, Norway). MIR spectral data was utilized to build SOC and SIC calibration models using partial least squares (PLS) regression. The number of components employed for each calibration model was determined by a visual interpretation of the decrease in residual variance determined by cross validation (Brown et al., 2005). To avoid overfitting, the number of components chosen is based on the point where the variance stops decreasing.

Independent Validation

Calibration models were built around the six site subsets, using a 6-fold site-holdout cross-validation method. For example, sites 2-6 were used to calibrate the model that was then validated exclusively on site one. With geographic independence between sites, this calibration method provided independent regional validation, simulating predictions at an unknown site within the region. In a separate step, cores from the site

being validated (referred to as "Local cores") were added to calibration sets to determine whether or not this improved the model's prediction capabilities.

Regional + Local calibration models were constructed using: (i) the five sites mentioned above, (ii) the five sites plus one Local core, and (iii) the five sites plus two Local cores.

Local Core Addition. Samples from spectrally dissimilar cores from the validation site were added to the calibration set during the modeling procedures to compare Local vs. Regional calibrations. These cores are referred to as "Local" cores or validation site cores. The spectral difference was determined by a standardized Mahalanobis Distance (Eq. 1), computed with standardized scores derived from principal components analysis (PCA). PCA was performed (S-Plus® 6.0, Insightful Corp., Seattle, WA, USA) for each Crushed, Milled, Moistcore and Drycore first derivative reflectance datum. Scores for the components explaining at least 90% of the variance (or for the first ten components, whichever provided scores of at least 10 components), were standardized to provide $\hat{\chi}$ for each sample. Standardized component scores for samples from the five sites not related to the sample being analyzed were averaged to give $\hat{\chi}_a$ for each component. The Mahalanobis Distance was then computed for each sample as:

$$\textit{Mahalanobis Distance} = \sqrt{(\hat{\chi} - \hat{\chi}_a)} \qquad (1)$$

$\hat{\chi}$ is the individual sample standardized PCA score vector
$\hat{\chi}_a$ is the average standardized PCA score vector of remaining 5 sites

The sample with the greatest distance was considered the most spectrally dissimilar sample. This sample along with the corresponding samples from its core, were all removed from the validation set and added to the calibration set.

Loadings. Loadings are a measure of how much each predictor (X-variable) contributes to explaining the response variation along each model component (Haaland and Thomas, 1988). The components that explain the most variability will be observed as more influential to the model. Loadings from SOC PLS calibration models developed with KM transformed spectral data of KBr-diluted samples were observed to determine the influence of each wavelength within the SOC prediction

Predictive Statistics

Predicted and measured values from the modeling procedures listed above were compared using the following statistics taking after Brown et al., 2005:

Mean Square Deviation: $$MSD = \frac{\sum_n \left(\mathrm{Y}_{pred} - \mathrm{Y}_{meas}\right)^2}{N} \quad (4)$$

Root Mean Square Deviation: $$RMSD = \sqrt{MSD} \quad (5)$$

Standard Error of Prediction: $$SEP = RMSD \times \sqrt{\frac{N}{(N-1)}} \quad (6)$$

Residual Prediction Deviation: $RPD = (Standard\ Deviation\ of\ Y_{val})/SEP$ (7)

r^2 = *square correlation* (8)

$$Bias = \frac{\sum_n \left(\mathrm{Y}_{pred} - \mathrm{Y}_{meas}\right)}{N} \quad (9)$$

Principal Components Analysis.

KM, AB and D spectral data were each plotted in principal component space (S-Plus® 6.0, Insightful Corp., Seattle, WA, USA) to identify spectral outliers and/or patterns (Shepherd and Walsh, 2002; Brown et al., 2005). The components explaining 90% of the variance were plotted in scatterplot matrix format. D10 spectral treatment gave the most distinct pattern so it is focused on here.

Results

Soil Carbon and Texture Data

Overall, C contents are relatively low and evenly distributed throughout the sites (Table 4.1). SOC contents are evenly spread except that sites 2, 4 and 5 include a few samples with higher SOC than the other 3 sites. Site 2 is more calcareous than the other sites and includes five extreme samples that increase the range of the dataset. Sand and clay measurements are evenly distributed with site 5 containing the most clay and site 6 containing the most sand. From the 15 replicated samples, SEL measurements for the combustion method were found to be 0.27 g kg^{-1} for SOC and 0.34 g kg^{-1} for SIC

Mineralogy

Powder mount analysis revealed the dominating non-clay minerals to be quartz and feldspar at all six sites. A sample from site 5 displayed a dominating unidentifiable peak at approximately 1.83 d-spacing. Clay mineralogy uncovered through oriented clay

mounts conveyed mixed mineralogy throughout the six sites. All samples contain moderate amounts of vermiculite, kaolinite, micas and smectites.

Table 4.1: Descriptive statistics for the six sites.

Property (g kg^{-1})	Statistic	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
	Min:	4.1	7.5	3.5	6.4	4.9	3.8
SOC	Max:	12.5	18.9	13.5	17.2	19.1	10.8
	Mean:	7.6	12.3	7.8	11.2	10.6	7.0
	Min:	0.0	1.8	0.0	0.0	0.2	0.0
SIC	Max:	15.5	72.5	22.9	18.0	14.2	19.3
	Mean:	7.1	24.1	9.6	3.5	6.2	5.3
	Min:	311.8	212.8	168.0	174.4	296.8	180.6
Clay	Max:	574.5	519.2	374.5	497.5	710.4	344.5
	Mean:	443.9	343.6	266.0	338.6	558.1	251.0
	Min:	135.2	102.0	167.9	127.2	100.0	214.7
Sand	Max:	444.9	369.5	634.6	515.9	232.9	510.6
	Mean:	279.6	193.0	385.9	344.0	144.4	397.0
	N:	51	48	57	48	59	52

Prediction Results

KBr Dilution. SOC and SIC prediction results for KBr-diluted and undiluted samples are displayed in Tables 4.2-4.5 and calibration results for the most accurate models are shown in Table 4.6. One consistency is the accuracy of undiluted sample preparation over KBr-diluted samples with the addition of Local cores. Undiluted Local predictions are more accurate for SOC and SIC with all data transformations, except for the SIC KM10 prediction.

Spectral Pretreatment. Organic C and SIC show different predictive accuracy with different spectral transformations. Organic C shows most predictive accuracy in the first derivative transformation data with the addition of two validation site cores (Fig.

4.4) with RMSD = 1.3 g kg^{-1}, r^2 = 0.86 and RPD = 2.61. The SEP value here is 1.3 gkg^{-1} (Table 4.6). For SIC, the Kubelka Munk transformation with the addition of 2 cores gives the most accurate prediction results with RMSD < 2 g kg^{-1} and r^2 >0.95 (Fig 4.5). The SEP value here is 1.4 g kg^{-1} (Table 4.7).

Local Core Addition. Improvement in predictive accuracy occurs with each Local core addition, the most accurate predictions coming from the addition of two cores. After two cores from the validation site were added to the AB calibration set, validation results display r^2 values increasing from 0.59 to 0.81 and RMSD decreasing from 2.54 to 1.46 g kg^{-1}. Undiluted samples show more increase in accuracy than KBr-diluted samples for all transformations with the addition of Local cores.

Table 4.2: PLS organic C calibration validation statistics from models developed with neat samples using three different spectral pretreatments.

Transformation of undiluted sample spectra	# Local Cores	RMSD (g kg^{-1})	r^2	RPD
AB10	0	2.54	0.59	1.36
	1	1.58	0.79	2.15
	2	1.46	0.81	2.31
KM10	0	2.30	0.65	1.49
	1	1.51	0.81	2.25
	2	1.37	0.84	2.45
D10	0	1.55	0.81	2.22
	1	1.41	0.83	2.41
	2	1.30	0.86	2.61

Table 4.3: PLS organic C calibration validation statistics from models developed with KBr-diluted samples using three different spectral pretreatments.

Transformation of diluted sample spectra	# Local Cores	RMSD (g kg^{-1})	r^2	RPD
AB10	0	2.19	0.62	1.57
	1	1.84	0.72	1.86
	2	1.69	0.76	2.04
KM10	0	2.08	0.65	1.65
	1	1.73	0.75	1.97
	2	1.64	0.78	2.10
D10	0	2.06	0.65	1.67
	1	1.99	0.68	1.75
	2	1.74	0.76	1.99

Table 4.4: PLS inorganic C calibration validation statistics from models developed using neat samples and three different spectral pretreatments.

Transformation of undiluted sample spectra	# Local Cores	RMSD (g kg^{-1})	r^2	RPD
AB10	0	4.59	0.85	2.36
	1	2.41	0.95	4.33
	2	2.23	0.95	4.54
KM10	0	7.48	0.56	1.45
	1	2.91	0.93	3.59
	2	1.95	0.96	5.07
D10	0	4.55	0.91	2.38
	1	2.13	0.96	4.90
	2	2.22	0.95	4.56

Table 4.5: PLS inorganic C calibration validation statistics from models developed using KBr-diluted samples and three different spectral pretreatments.

Transformation of diluted sample spectra	# Local Cores	RMSD (g kg^{-1})	r^2	RPD
AB10	0	6.52	0.68	1.66
	1	4.22	0.84	2.48
	2	4.27	0.84	2.46
KM10	0	6.52	0.68	1.66
	1	5.53	0.75	1.89
	2	4.18	0.84	2.47
D10	0	6.17	0.74	1.75
	1	5.71	0.78	1.91
	2	4.00	0.86	2.63

Table 4.6: SEP values for organic C validation results from Regional and Local calibrations developed using three different spectral transformations of spectra from neat and KBr-diluted samples.

Organic Carbon SEP values (g kg^{-1})				
Preparation Method	# of Local Cores Added	Spectral Transformation		
		KM	AB	D
Neat	0	2.31	2.54	1.56
	1	1.52	1.58	1.42
	2	1.38	1.47	1.31
KBr-diluted	0	2.09	2.19	2.07
	1	1.74	1.85	1.99
	2	1.64	1.70	1.74

Table 4.7: SEP values for inorganic C validation results from Regional and Local calibrations developed using three different spectral transformations of spectra from neat and KBr-diluted samples.

Inorganic Carbon SEP values (g kg^{-1})				
Preparation Method	# of Local Cores Added	Spectral Transformation		
		KM	AB	D
Neat	0	7.51	4.61	6.34
	1	2.92	2.42	4.02
	2	1.96	2.24	2.82
KBr-diluted	0	6.54	6.54	6.19
	1	5.55	4.23	5.73
	2	4.20	4.29	4.01

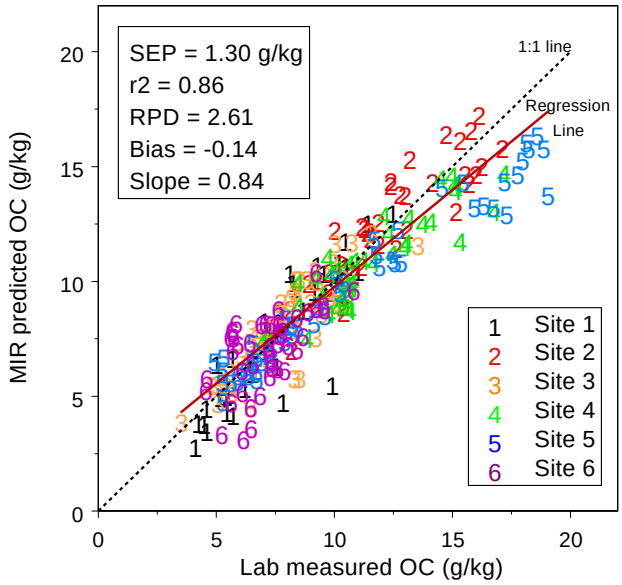

Figure 4.4. Validation results of the most accurate organic C calibration developed with first derivative transformation of Regional MIR spectral data + 2 Local cores.

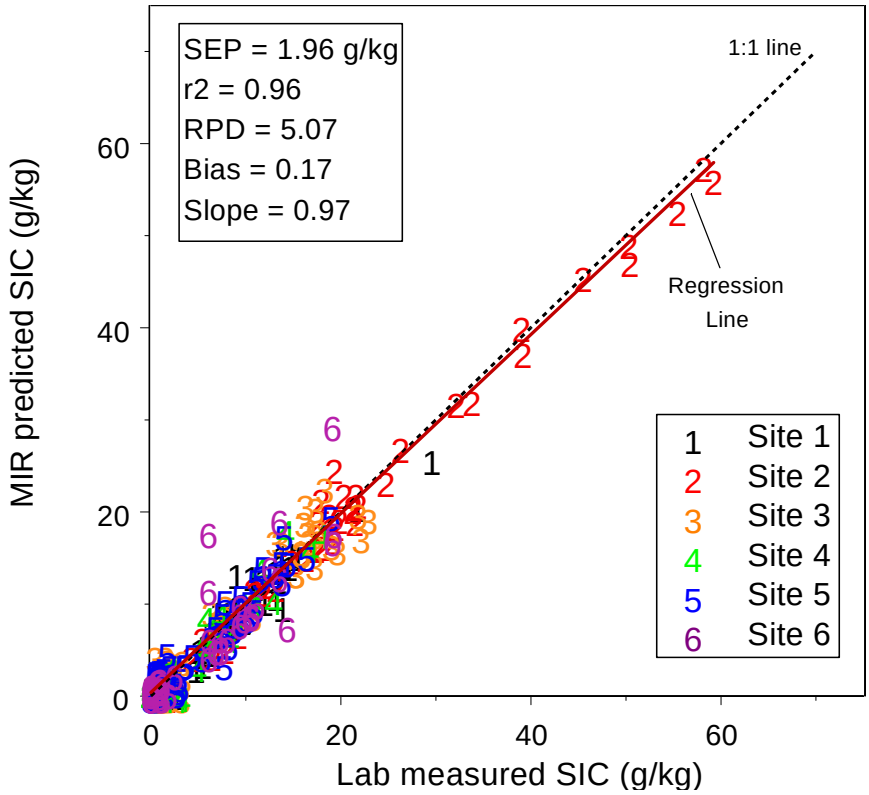

Figure 4.5. Validation results of the most accurate inorganic C calibration developed with Kubelka Munk transformed Regional MIR spectral data + 2 Local cores.

Table 4.8: Calibration statistics of the most accurate SOC and SIC models.

Variable	Most Accurate Calibration Model	RMSD (g kg^{-1})	r^2	RPD
SOC	D10 + 2 cores	0.86	0.92	4.03
SIC	KM10 + 2 cores	1.60	0.97	6.83

Loadings. Loadings were determined for both KBr-diluted and undiluted sample KM calibration models and the weighted average from the first ten components are displayed in Figs. 4.6 and 4.7. The weights from KBr diluted models show the most weight with the band at 700 cm^{-1}. Other influencing bands are found at 1060, 1480, 1860 and 2820 cm^{-1}. The weights from the neat sample calibration model show the most weight with the bands at 1060, 1320, 1480, 1680, and 1800 cm^{-1}. Other important bands are seen at 2610, 2820 and the region from 3700-3200 cm^{-1}.

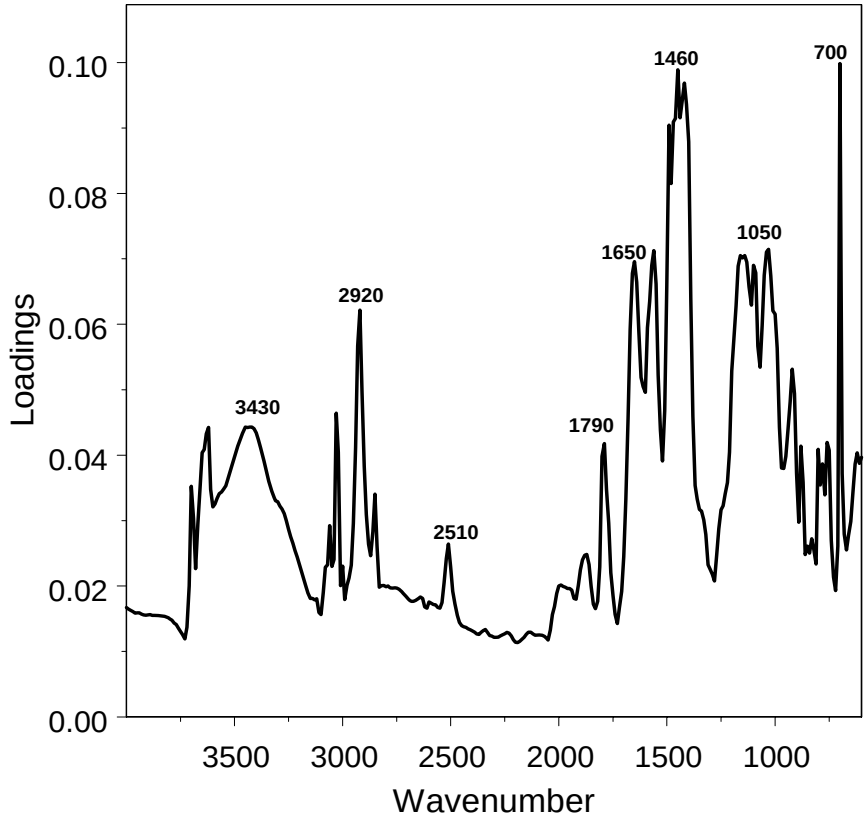

Figure 4.6. Weighted average loadings of the calibration model developed with Kubelka-Munk transformed spectral data from KBr-diluted samples with each line representing a component explaining the presented amount of variability.

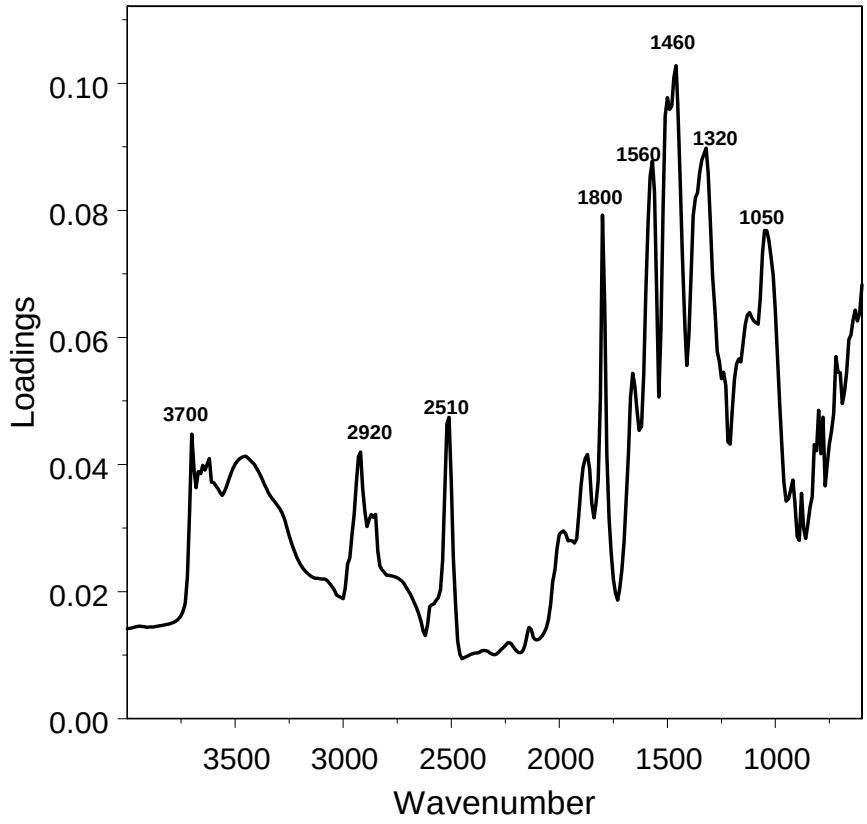

Figure 4.7. Weighted average loadings of the calibration model developed using Kubelka-Munk transformed spectral data from neat sample presentation with each line representing a component explaining the presented amount of variability.

Principal Component Analysis

The most telling display of spectral patterns came from the undiluted first derivative PCA diagram (Fig. 4.8). The first 5 components in this model explain 82% of the variability.

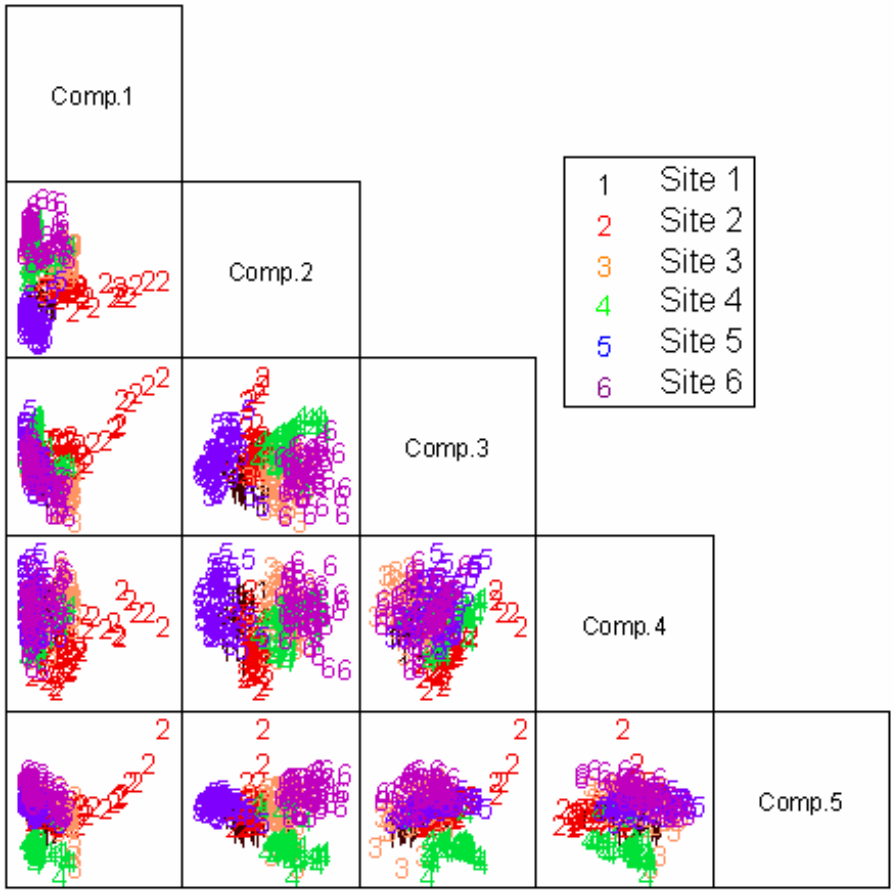

Figure 4.8. Scatterplot matrix diagram showing the first five principal components of the MIR first derivative soil spectra explaining 82% of the variability.

Spectral Comparison

Four samples with extreme C contents were compared spectrally with both KBr-diluted and neat preparation methods. Samples included were: sample 2 - containing the highest SIC content (SIC = 72.5 g kg^{-1}, SOC = 9 g kg^{-1}), sample 3 – containing the lowest

SOC content (SIC = 19.2 g kg^{-1}, SOC = 3.5 g kg^{-1}), sample 4 - without SIC but with relatively high SOC content (SIC = 0, SOC = 17.2 g kg^{-1}), and sample 5 - the highest SOC content (SIC = 1.0 g kg^{-1}, SOC = 19.1 g kg^{-1}). These samples were qualitatively examined together to decipher spectral influences due to $CaCO_3$ and SOM content as well as to determine the effect sample preparation has on different sample types.

KBr-Dilution. Diluted sample spectra show three distinct features that are indecipherable in undiluted sample spectra; a comparison of the four samples, both diluted and undiluted, can be seen in Figs. 4.9 and 4.10. An organic stretching region from 3100-2800 cm^{-1} is observed in diluted samples and appears stronger in the low SOC sample as it can be overlapped by CO_3 overtone absorption from 3000-2880 cm^{-1} (Nguyen, 1991). Diluted samples also show SIC content visible at 2510 cm^{-1} as the CO_3 overtone and combination vibration (Nguyen et al., 1991; Bertrand et al., 2002). This feature is shown in the low SOC content sample and more distinctly within the high SIC sample spectra. All diluted samples show identical doublet absorptions from 1200-980 cm^{-1}. This is an area of intense Al-OH and Si-O stretching, resulting from mineralogical absorbance (Nguyen et al., 1991; Bertrand et al., 2002) that is missing in undiluted samples.

In undiluted samples, there is an overall higher absorption in comparison to KBr-diluted samples throughout the MIR region. Higher SOC content gives higher absorbance throughout the MIR region for neat samples. Beginning at the far end of the MIR spectrum, diluted and undiluted samples have the same features. The C-H stretch at

3100-2800 cm^{-1} is not apparent in undiluted spectra and the carbonate feature here is only visible with a high SIC sample. The CO_3 absorption at 2500 cm^{-1} is much more dramatic in the undiluted spectra than in the diluted, even for low SIC content samples. For the high SOC content sample as well as the non-carbonaceous sample, there is a large unidentifiable peak between 1400 and 1250 cm^{-1} that could be a specular reflection artifact resulting from SOM features in this region (Avram and Mateescu, 1966; Conley, 1972; Griffith and Schnitzer, 1975; Baes and Bloom, 1989; Nguyen et al, 1991; Stevenson, 1994; Celi et al, 1997). This feature does not appear in higher SIC content samples, or within any of the diluted spectra. Finally, the doublet absorption from 1200-980 cm^{-1} resulting from clay mineralogy that is shown at for all samples in diluted spectra is unidentifiable in the undiluted spectra of all samples.

There are some consistencies between diluted and undiluted sample spectra. For both preparations, SOM combination bands at or near 3400 cm^{-1} (Nguyen, 1991) appear to be dominated by O-H vibration of structural water seen from 3600-3000 cm^{-1} (Ellerbrock and Gerke, 2004). A carbonate feature that is not visible in Fig. 4.7 because of low SIC content samples can be seen at 1800 cm^{-1} in Fig. 4.8 of both diluted and undiluted high SIC samples. Farther down the spectrum, there are SOM combination bands apparent at 1550-1600 cm^{-1} (Griffith and Schnitzer, 1975; Stevenson, 1982; MacCarthy and Rice, 1985; Nguyen et al, 1991; Celi et al., 1997; Alciaturi et al, 2001; Ellerbrock and Gerke, 2004). Both preparations show this feature but the diluted high SOC content sample shows the most distinct absorption. Finally, there is a considerable carbonate absorption feature for the low SOC and high SIC content samples at 1450 cm^{-1}

(Nguyen et al, 1991; Clark, 1999; Bertrand et al, 2002) that is visible with or without KBr-dilution.

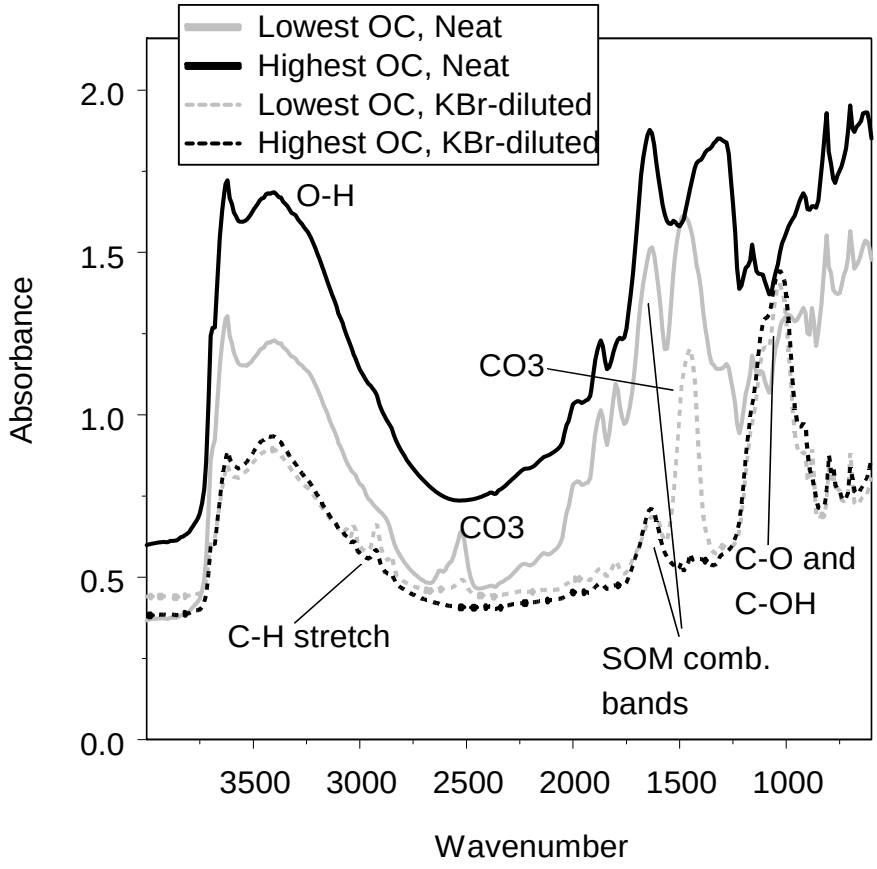

Figure 4.9. Spectral comparison of two samples, neat and diluted, containing both the highest and lowest organic C contents.

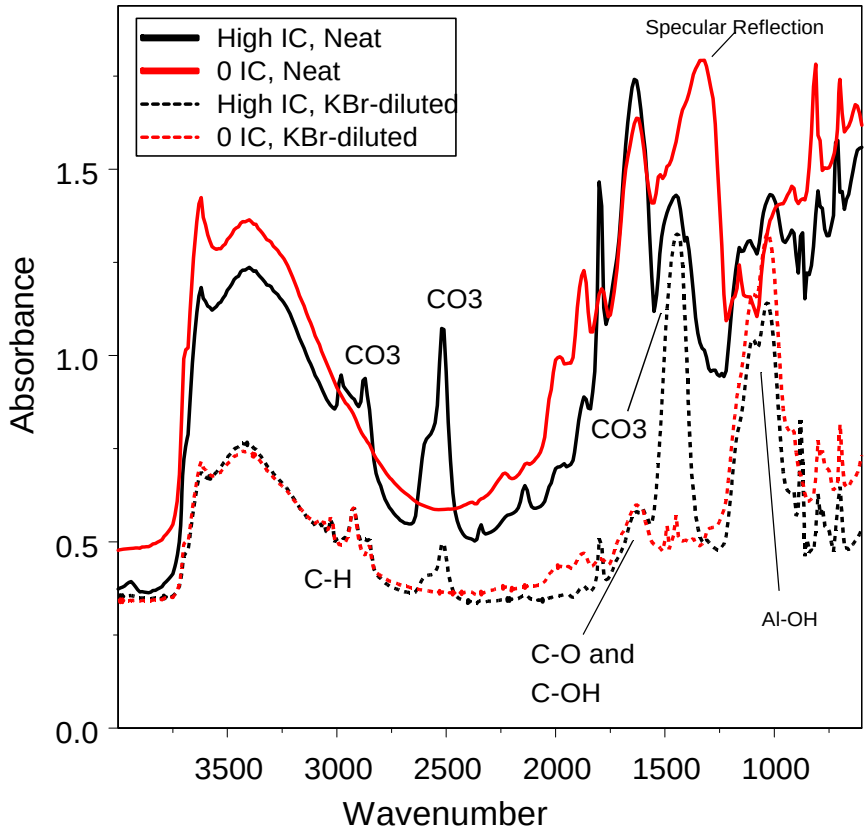

Figure 4.10. Spectral comparison of two samples, KBr-diluted and undiluted, containing the highest and lowest inorganic C contents.

Organic Matter Removal

The correlation between spectral difference of samples with and without SOM and SOC content was displayed to determine where SOC content relates to spectral SOM absorptions (Fig. 4.11). For KBr-diluted spectra, the different transformations appear to be grouped tighter as they exhibit equivalent correlation values throughout the spectrum (excluding the region between 3700 and 3200 cm^{-1} where lower correlation for KM and AB transformations is displayed). One of the higher correlations for both sample presentations and all transformations was at 1550 cm^{-1}. Other strong correlations were

found at 2920, 2850 and 1380 cm^{-1} for diluted samples and 3680, 2890, and 2520 cm^{-1} for undiluted samples.

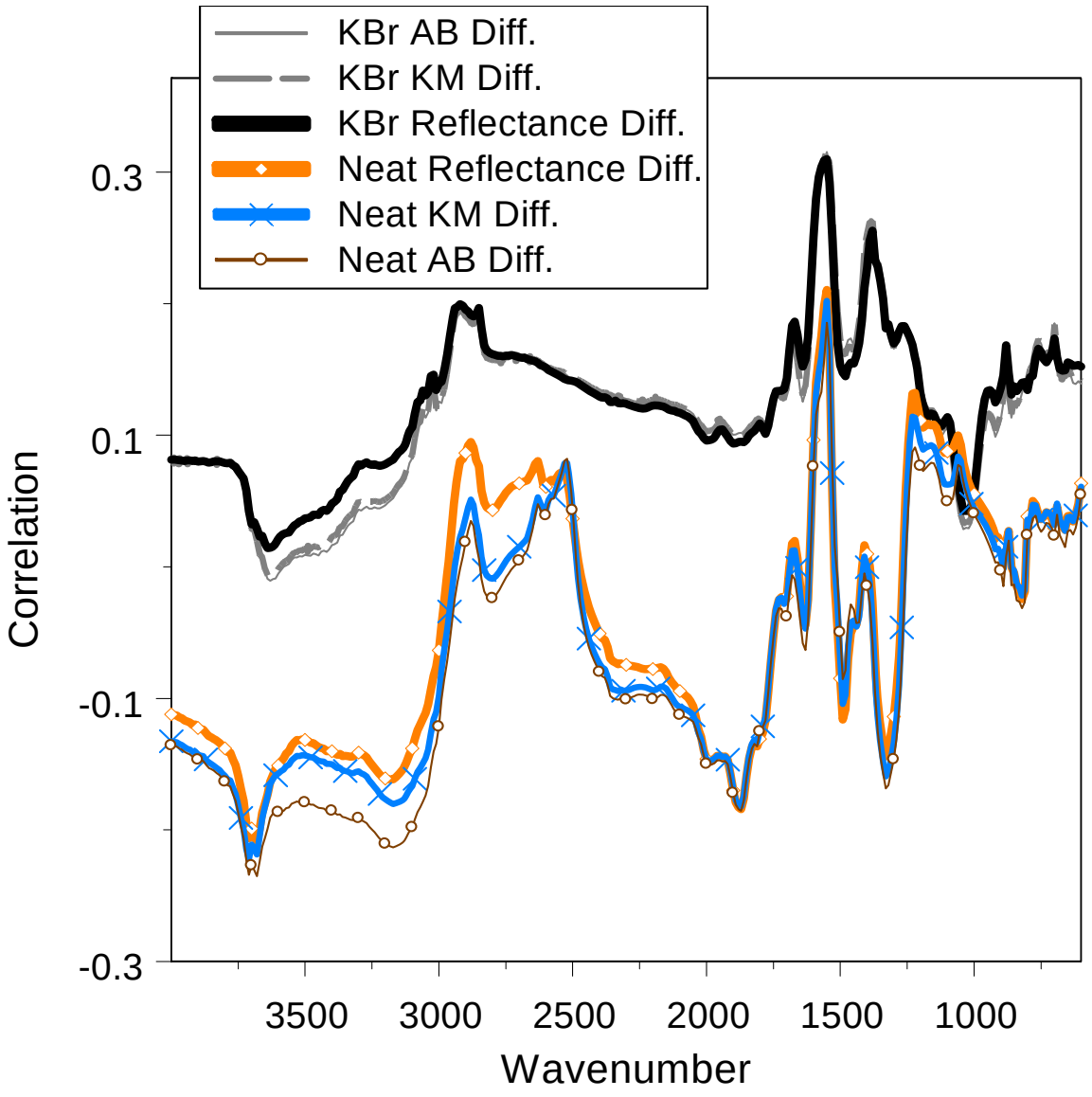

Figure 4.11. Correlation between the spectral difference ("Diff.") of samples with and without organic matter, within 3 spectral transformations, in both KBr diluted and neat preparations.

Discussion

PLS Regression Predictions

Prediction results were distinct between two different types of C. Inorganic C prediction results had higher r^2 and RMSD values than SOC prediction results. Inorganic C predictions also have much higher RPD values, most approaching 5, which is a good indication of accuracy. Validation site core addition affected SIC predictive accuracy more so than SOC predictions. With a greater SIC spread throughout the dataset, adding spectrally dissimilar cores from the validation site to the calibration model makes for an even more robust model specific to the site being predicted. Inorganic C predictions were accurate due to the distinct carbonate absorption features within the MIR portion of the spectrum. Based on literature values, carbonate features can be identified at 3000-2880, 2600-2500, 1830-1800, 1470-1410, and 1100-1000 cm^{-1} (Nguyen et al., 1991; Clark, 1999). These were all identified in our soils, through qualitative examination of soil spectra.

Organic C prediction results had lower RPD and r^2 values than those resulting from SIC predictions. McCarty et al. (2002) claim that SIC content can degrade SOC PLS predictions; this could explain lower RPD and r^2 values. But, SOC RMSD values were much lower overall showing less predictive deviation than SIC predictions. This was a result of a smaller range of SOC values within this dataset

Without the addition of local cores, KBr-diluted prediction results were more accurate than neat sample predictions for certain transformations; although, the addition

of local cores did improve predictive accuracy more so than dilution with KBr. A potential explanation for this is component intensity. Past studies have shown spectral absorptions to be characteristic of particular soil components by isolating the component from the soil and then diluting in KBr for spectral analysis. When analyzing a soil sample in its entirety with dilution, the concentration of the spectral influencing structure is much lower than it would be without dilution. It is questionable whether these absorptions are as intense within the resulting diluted soil spectra. The accuracy of predictive models, which is based on these spectral features, will be affected when soils are diluted.

Core addition to the calibration model increased predictive accuracy of SOC and SIC validations with both diluted and neat samples, more so with neat. The addition of spectrally dissimilar samples contributes to the robustness of the calibration model for these soils in particular, without degrading accuracy. This could be a new path within predictive model building: the addition of a limited amount of spectrally dissimilar samples to contribute to the overall applicability of the model without reducing accuracy.

Standard error of prediction values for both SOC and SIC were less than 2.0, with SIC values slightly higher. Though SEP values by MIR-PLS predictive techniques are more than triple the SEL values by traditional laboratory procedures, they are still low enough to claim semi-quantitative results approaching laboratory precision.

Principal Components Analysis

Principal Component Analysis results indicate that spectral similarity between samples is dominated by site. Even though all samples were collected from the same physiographic region, with similar characteristics at each site, samples were narrowly grouped according to site. This dominated spectral similarity, even over depth of the sample or soil characteristics such as carbon content or texture. Grouping by site displays that spectral similarity based on a limited locale is another influence on prediction model accuracy and could affect future applications of regional models. If spectral similarity of soils is this limited, the applicability of regional as well as global calibration models should be inspected further.

Spectral Comparisons

Dilution of the soil sample in a non-absorbing KBr matrix clearly created qualitative changes in the spectra. The most apparent were carbonate, clay, and SOM absorption features that can remain invisible or distorted in both KBr-diluted and undiluted spectra. For example, C-H stretching is apparent in diluted spectra from 3100-2800 cm^{-1} and is left unseen in undiluted spectra. The distinct mineralogical feature from 1200-1000 cm^{-1} is seen only in diluted spectra. In neat spectra, there could be carbonate interactions distorting this feature or even a complete band inversion that appears visible in low SIC content samples (Nguyen et al., 1991). Also in neat samples, carbonate features appear dominant, especially the feature at 2500 cm^{-1} that does not show up as intensely as in diluted samples.

The feature from 1400-1250 cm^{-1} is a dramatic absorption feature that does not appear in diluted spectra, or in high SIC content samples. This feature is very broad and appears to be a distortion. In high SIC content samples, the feature is narrowed and dominated by the carbonate fundamental absorption from 1500-1400 cm^{-1} (Nguyen et al, 1991; Clark, 1999; Bertrand et al, 2002). I believe this feature to be an artifact of specular reflection resulting from SOM absorption features within this vicinity such as the SOM combination bands found from 1400-1390 cm^{-1} (Griffith and Schnitzer, 1975; Nguyen et al, 1991; Stevenson, 1994; Celi et al, 1997) and C-O stretch and OH deformation of COOH and phenols from 1280-1200 cm^{-1} (Avram and Mateescu, 1966; Conley, 1972; Griffith and Schnitzer, 1975; Baes and Bloom, 1989; Stevenson, 1994; Celi et al, 1997). As these are the only reported features within this vicinity, in two very narrow regions, the broad feature apparent in our spectra must be a distortion of these absorptions. This SOM feature appears as a very small peak in the diluted spectra of non-carbonaceous samples at 1490 and 1480 cm^{-1}, possibly shifted CH_2 and CH_3 bends (Stevenson, 1994; Celi et al, 1997; Alciaturi et al, 2001) and obviously unaffected by specular distortions.

When SOM is removed from the samples and the spectral difference is correlated with SOC, more differences between diluted and neat preparations are noticed (Fig. 4.11). The central commonality observed is the high correlation at 1550 cm^{-1} which can be explained by organic stretching of either aromatic C=C (Baes and Bloom, 1989), CN amide or COO (Stevenson, 1994). Both diluted and neat spectra show correlation between 2850 and 2950 cm^{-1}, which is related to the CH stretching region found from

3000-2880 cm^{-1} (Eglington, 1970; Conley, 1972; Griffith and Schnitzer, 1975; Nguyen et al, 1991; Stevenson, 1994). KBr-diluted samples also showed correlation at 1380 cm^{-1}, a region of SOM combination bands: CH2 and CH3 bending, C-OH deformation of COOH and COO- symmetric stretch (Griffith and Schnitzer, 1975; Nguyen et al, 1991; Stevenson, 1994; Celi et al, 1997). The neat sample absorbance spectra show correlation at 3680 cm^{-1}, a result of a fundamental OH stretch found at 3650 cm^{-1} (Clark, 1990) relevant to structural water content within clay. Varying spectral differences and correlations convey more disparity between diluted and neat spectra but also key relations between spectral features and SOM content for both.

KBr dilution appears to make a qualitative improvement over neat sample spectra. It is important to recognize that specific absorption features appear differently, or only appear if certain concentrations are available, in both diluted and neat samples. If qualitative analysis is the goal at hand, diluted spectra should be examined first and foremost; but it would facilitate spectral interpretation to look at neat sample spectra as well.

Loadings

The important bands uncovered through analysis of loadings exhibited commonalities with bands correlated to SOC. KBr-dilute and neat sample preparation conveyed the highest loading at 1480 cm^{-1} with this peak running into 1550 cm^{-1}– the same band containing the highest correlation between spectral difference with organic removal and SOC for both neat and KBr diluted samples. As stated earlier, this is an area

of organic stretching and based on these results, it is obviously influencing spectral activity in both KBr-diluted and neat sample presentations. Diluted samples displayed another commonality between correlation and loadings at 2820 cm^{-1} established as a CH stretching region. Neat sample preparation gives less emphasis to this feature along with any other feature below 1800 cm^{-1}. The most loading is given to the feature at 700 cm^{-1} for KBr-diluted analysis which could relate to a carbonate or aromatic C-H bending feature found in this area.

Transformations

For SOC, best prediction results were with a first derivative transformation and addition of two Local cores. For SIC, best prediction results were with Kubelka-Munk spectral transformation with the addition of two Local cores. The transformations affected SOC and SIC predictions differently due to the differing absorption qualities of SOM constituents and carbonates. Organic constituents are much more difficult to identify in soil spectra; baseline removal or first derivative transformation would be the most productive way of enhancing these absorptions. Carbonates are lighter in color, with physical scattering characteristics, so reduction of the scattering affect with the KM transformation would aid in the clarification of these features. This calls attention to the specificity of different transformation methods, and their relation to what is being spectrally analyzed.

Conclusions

Mid-infrared diffuse reflectance spectroscopy modeled with PLS regression has proven to be an accurate method of soil SOC and SIC prediction for calcareous glacial till soils. The method provided quantitative results for both SIC and SOC, with SEP values of < 2.0 g kg-1 for the most accurate calibration models. The most accurate predictions were derived from Regional + Local calibrations. Different spectral transformations were compared and results were found to be dependent on predictor variable as well as sample preparation. For SOC, the most accurate prediction results were with first derivative transformation and addition of two Local cores. For SIC, the most accurate prediction results were with a Kubelka-Munk transformation and addition of two Local cores.

Neat samples were analyzed in comparison to KBr-diluted samples to determine the role specular reflection plays within predictive accuracy of soil C. It was determined that without the addition of Local cores, KBr-diluted and neat samples had comparable predictive accuracy. After the addition of Local cores, neat samples consistently provided more accurate predictions. Within the qualitative investigation, KBr-diluted and neat samples show different absorption features. Neat sample analysis did present a specular reflection artifact that apparently did not affect predictive accuracy after the addition of Local cores. For future qualitative examinations with the intention of soil component identification, both diluted and neat samples should be examined simultaneously as each preparation method reveals different component features.

The absorption feature at approximately 1500 cm^{-1} was found to be the most influential factor to SOC predictions from neat as well as KBr-diluted samples. This feature showed the most correlation for both sample presentations, as well as high loading for KBr-diluted and neat sample PLS calibrations. There was also correlation and loading given to non-organic constituents such as clay and carbonate features.

The Regional and Local calibration models were all independently validated, displaying potential application of the model within the north central Montana region. With complex mineralogy and a spectral fingerprint by site (determined through PCA), this model would likely fail if applied outside the region. Core addition was a successful method used to increase calibration model robustness without degrading predictive accuracy. Future applications of this model or other regional models might succeed within the region, and possibly outside of the region, with the addition of a limited number of spectrally dissimilar samples from the site being predicted. Past studies have shown higher predictive accuracies, but most have only conveyed calibration results or non-independent validation results. With these independently validated model results, we are now aware of the accuracy of MIR-DRS PLS regression when applied outside of the calibration dataset.

References

Alciaturi, C.E., Escobar, M.A., de La Cruz, C., and Vallejo, R., 2001. Determination of chemical properties of pyrolysis products from coals by diffuse-reflectance infrared spectroscopy and partial least squares. Analytica Chimica Acta, 436: 265-272.

Amundson, R., 2001. The carbon budget in soils. Annu. Rev. Earth Planet Sci., 29:535-562.

Avram, M. and Mateescu, G.H., 1966. Infrared spectroscopy: Applications in organic chemistry. Wiley-Interscience, New York, NY.

Baes, A.U. and Bloom P.R., 1989. Diffuse reflectance and transmission fourier transform infrared (DRIFT) spectroscopy of humic and fulvic acids. Soil Sci. Soc. Am. J. 53: 695-700.

Ben-Dor, E. and Banin, A., 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Sci. Soc. Am. J, 59: 364-372.

Bertrand, I., Janik, L.J., Holloway, R.E., Armstrong, R.D., and McLaughlin, M.J., 2002. The rapid assessment of concentrations and solid phase associations of macro- and micronutrients in alkaline soils by mid-infrared diffuse reflectance spectroscopy. Aust. J. Soil Res., 40: 1339-1356.

Brown, D.J., Brickelmeyer, R.S., and Miller, P.R., 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C predictions in Montana. Geoderma, 129: 251-267.

Celi, L., Schnitzer, M., and Negre, M., 1997. Analysis of carboxyl groups in soil humic acids by a wet chemical method, FTIR spectrophotometry, and solution-state C-13 NMR: a comparitive study. Soil Science, 162 (3): 189-200.

Clark, R.N., King, T.V., Klejwa, M., and Swayze, G., 1990. High Spectral Resolution Reflectance Spectroscopy of Minerals. J. of Geophysical Research, 95(B8): 12653-12680.

Clark, R.N., 1999. Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy. In: N. Rencz (Editor), Remote Sensing for the Earth Sciences: Manual of Remote Sensing, 3 ed. John Wiley & Sons, New York, pp. 3-52.

Coates, J., 1998. Review of IR Sampling Methods. In: Applied Spectroscopy – A Compact Reference for Practitioners. Workman, J. and Sprinsteen, A. (eds.). Academic Press, Chestnut Hill, MA, pp. 73-74.

Conley, R.T., 1972. Infrared Spectroscopy. Allyn and Bacon, Boston, MA.
Duckworth, J.H., 1998. Spectroscopic Quantitative Analysis. In: Workman, J. and Springsteen, A. (Eds.), Applied Spectroscopy – A Compact Reference for Practitioners, Academic Press, Chestnut Hill, MA, pp. 93-107.

Duckworth, J.H., 1998. Spectroscopic Quantitative Analysis. In: Workman, J. and Springsteen, A. (Eds.), Applied Spectroscopy – A Compact Reference for Practitioners, Academic Press, Chestnut Hill, MA, pp. 93-107.

Dunn, B.W., Beecher, H.G., Batten, G.D., Ciavarella, S., 2002. The potential of near-infrared reflectance spectroscopy for soil analysis – a case study from the Riverine Plain of south-eastern Australian. Australian Journal of Experimental Agriculture, 42: 607-614.

Ellerbrock, R.H. and Gerke, H.H., 2004. Characterizing organic matter of soil aggregate coatings and biopores by Fourier transform infrared spectroscopy. European J. of Soil Science, 55: 219-228.

Griffith, S.M. and Schnitzer, M., 1975. Analytical characteristics of humic and fulvic acids extracted from tropical volcanic soils. Soil Sci. Soc. Amer. Proc., 39: 861-866.

Islam, K., Singh, B., and McBratney, A., 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. Australian Journal of Soil Research, 41: 1101-1114.

Janik, L.J, and Skjemstad, J. O., 1995. Characterization and analysis of soils using mid-infrared partial least-squares. II. Correlations with some laboratory data. Aust. J. Soil Res., 33: 637-650.

Janik, L.J., Merry, R.H., Skjemstad, J.O., 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? Australian Journal of Experimental Agriculture, 38: 681-96.

MacCarthy, P. and Mark H.B., 1975. Direct measurement of infrared-spectra of humic substances in water by Fourier-transform infrared spectroscopy. J. of Ag. And Food Chem., 23 (3): 600-602.

MacCarthy, P. and Rice, J.A., 1985. Spectroscopic methods (other than NMR) for determining functionality in humic substances. In: Humic Substances in Soil Sediments and Water. G.R. Aiken et al. (eds.), John Wiley and Sons Inc., New York, pp 527-559.

Madari, B.E., Reeves, J.B., Coelho, M.R., Machado, P., De-Polli, H., 2005. Mid- and near-infrared spectroscopic determination of carbon in a diverse set of soils from the Brazilian National Soil Collection. Spectroscopy Letters, 38: 721-740.

Martens H and Naes T, 1987. Multivariate calibration by data compression. John Wiley and Sons: Chichester, UK.

McCarty, G.W., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Sci. Soc. Am. J, 66: 640-646.

Nguyen, T., Janik, L.J., Raupach, M., 1991. Diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy in soil studies. Aust. J. Soil Res, 29: 49-67.

Perkins, W.D. 1993. Sample handling in infrared spectroscopy – An Overview . p. 11-53. In P.B. Coleman (ed.) Practical sampling techniques for infrared analysis, CRC Press, Boca Raton, FL.

Post, W.M., Izaurralde, R.C., Mann, L.K., Bliss, N., 2001. Monitoring and verifying changes of organic carbon in soils. Climatic Change, 51: 73-99.

Reeves III, J.B. and Reeves, V.B., 2002. Effects of apodization function, zero filling, background spectra, and absorbance transformation on mid-infrared calibrations for feed composition. Spectroscopy Letters, 35 (5): 663-680.

Reeves III, J.B., McCarty, G.W. and Mimmo, T., 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. Environmental Pollution 116: S277-S284.

Reeves III, J.B., 2003. MIR DRS: Is dilution with KBr really necessary, and if so, when? American Laboratory, 2003: 24-28.

Rumpel, C., Janik, L.J., Skjemstad, J.O., and Kogel-Knabner, I., 2001. Quantification of carbon derived from lignite in soils using mid-infrared spectroscopy and partial least squares. Organic Geochemistry, 32: 831-839.

Schnitzer, M., 1969. Characterization of humic constituents by spectroscopy. In: McLaren, A.D. and Skujins, J. (eds.), Soil Biochemistry II, Marcel Dekker, New York, pp. 60-95.

Shepherd, K.D. and Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Sci. Soc. Am. J., 66: 988-998.

Siebelec, G., McCarty, G.W., Stuczynski, T.I., and Reeves, J.B., 2004. Near- and mid-infrared diffuse reflectance spectroscopy for measuring soil metal content. J. Environmental Quality, 33: 2056-2069.

Stevenson, F.J. and Goh, K.M., 1971. Infrared spectra of humic acids and related substances. Geochimica et Cosmochimica Acta 35 (5): 471.

Stevenson, F.J. and Goh, K.M., 1974. Infrared spectra of humic acids – elimination of interference due to hygroscopic moisture and structural changes accompanying heating with KBr. Soil Science, 117 (1): 34-41.

Stevenson, F.J., 1982. Nitrogen in Agricultural Soils. American Society of Agronomy, New York, pp. 23-30.

Stevenson, F.J., 1994. Humus Chemistry: Genesis, Compositions and Reactions. John Wiley and Sons Inc., New York.

Tan, K.H., 1977. Infrared-absorption similarities between hymatomelanic acid and methylated humic acid. Soil Sci. Soc. Am. J., 39 (1): 70-73.

Vinkler, P., Lakatos, B. and Meisel, J., 1976. Infrared spectroscopic investigations of humic substances and their metal complexes. Geoderma, 15 (3): 231-242.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., and Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma, 131: 59-75.

Workman, J., 2004. Near infrared spectrophotometers. In: Near-Infrared Spectroscopy in Agriculture. Roberts, C.A., Workman, J., and Reeves III, J.B. (eds.). American Society for Agronomy Inc., Madison WI, pp. 11-31.

CHAPTER 5

SYNTHESIS

Within this study, I compared different methods of soil C characterization using DRS with the Vis-NIR and MIR regions. For both regions, Regional and Local calibrations were compared as were different sample preparations. Global calibrations and different spectral pretreatments were also compared for the Vis-NIR and MIR regions, respectively. The dataset used included calcareous glacial till soils with complex mineralogy, providing a unique application of study results.

DRS proved to be an effective means of soil C quantification. Past research has demonstrated effectiveness through calibration model accuracy but with both calibration and validation results presented here, the method was deemed quantitative for both regions of the spectrum. Vis-NIR Regional + Local PLS validation results provided RMSD values of 1.64 g kg^{-1} for SOC and 2.81 g kg^{-1} for SIC. MIR Regional + Local PLS validation results provided RMSD values of 1.30 g kg^{-1} for SOC and 1.95 g kg^{-1} for SIC.

Calibration set size and place of origin were determined to be key factors within model development. For these calcareous glacial till soils, a distinct spectral fingerprint was displayed by site. Adding spectrally dissimilar cores from the validation site to the calibration set facilitated more accurate validations for Vis-NIR and MIR calibration models. In the Vis-NIR study, larger calibration sets were compared and found to be less accurate than the smaller regional + local cores. For relatively young, less weathered

soils such as these, the notion of "quality over quantity" applies to calibration set construction with a goal of accurate validations.

Sample preparation methods were examined within Vis-NIR model validations and found to affect predictive accuracy. *In-situ* analysis was simulated with field-moist and dry cores and degraded predictive accuracy. But, RMSD values for SOC validations using moist and dry cores were both <3.0 g kg^{-1} conveying the possibility for semi-quantitative analysis. For these soils, analysis in the laboratory of more homogenized soil samples leads to greatest predictive accuracy of soil C.

Different preparation methods, neat vs. KBr-diluted samples, were compared using the MIR region. KBr-diluted and neat samples gave comparable predictive accuracy without the addition of local cores. With addition of local cores, neat samples consistently produced more predictive accuracy. Conversely, qualitative examination of neat and diluted samples together showed that neat samples omit key absorption features as well as exhibit specular reflection artifacts.

Soil organic matter was investigated in both the Vis-NIR and MIR regions. Correlations between SOC and spectral difference after organic matter removal, as well as loadings of PLS-SOC calibrations, assisted in determination of relevant wavelength regions. Within Vis-NIR SOC PLS calibations, loadings implied importance of water and hydroxyl bands, conveying possible correlation to clay content. There are higher loadings as well as correlation to SOC within the visible region as well. This is a key finding as most studies place emphasis on vibrational absorptions found within higher wavelength regions. Rather, chief SOM features are in fact due to electronic absorptions.

In the MIR region, loadings and correlation suggest the band at 1550 cm^{-1} is a key feature for SOM spectral signatures of diluted and undiluted soil matrices. There were patterns recognized between SOC content and SOM features, as well as clay and carbonate features, displaying that directly as well as indirectly correlated constituents play a role in predictions of SOC.

In the comparison of the two studies, the MIR region provided more accurate SOC and SIC predictions than the Vis-NIR region. But, both regions convey comparable means to soil C predictions. With less expensive Vis-NIR instrumentation, quicker analysis and sample preparation, the Vis-NIR region could be established as the more efficient method for soil C measurements. Variability within a site as well as necessity of accuracy should be considered while making the choice between the two methods of analysis and sample presentation. If the site is likely to be variable itself, and accuracy within 10 g kg^{-1} is acceptable, Vis-NIR *in-situ* analysis should be looked into further. This method is by far the most convenient and could apply DRS as an efficient form of analysis for many new and varying objectives.