DEVELOPMENT OF FIELD TECHNIQUES TO PREDICT SOIL CARBON, SOIL NITROGEN AND ROOT DENSITY FROM SOIL SPECTRAL REFLECTANCE

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

Doctor of Philosophy

in

Soil Science

at Massey University, Palmerston North, New Zealand

Bambang Hari Kusumo

2009

Abstract

The objectives of this research were to develop and evaluate a field method for *in situ* measurement of soil properties using visible near-infrared reflectance spectroscopy (Vis-NIRS). A probe with an independent light source for acquiring soil reflectance spectra from soil cores was developed around an existing portable field spectrometer (ASD FieldSpecPro, Boulder, CO, USA; 350-2500 nm).

Initial experiments tested the ability of the acquired spectra to predict plant root density, an important property in soil carbon dynamics. Reflectance spectra were acquired from soil containing ryegrass roots (*Lolium multiflorum*) grown in Allophanic and Fluvial Recent soils in a glasshouse pot trial. Differences in root density were created by differential nitrogen and phosphorus fertilization. Partial least squares regression (PLSR) was used to calibrate spectral data (pre-processed by smoothing and transforming spectra to the first derivative) against laboratory-measured root density data (wet-sieve technique). The calibration model successfully predicted root densities (r^2 = 0.85, RPD = 2.63, RMSECV = 0.47 mg cm^{-3}) observed in the pots to a moderate level of accuracy. This soil reflectance probe was then tested using a soil coring system to acquire reflectance spectra from two soils under pasture (0-60 mm soil depths) that had contrasting root densities. The PLSR calibration models for predicting root density were more accurate when soil samples from the two soils were separated rather than grouped. A more accurate prediction was found in Allophanic soils (r^2 = 0.83, RPD = 2.44, RMSECV = 1.96 mg g^{-1}) than in Fluvial Recent soils (r^2 = 0.75, RPD = 1.98, RMSECV = 5.11 mg g^{-1}). The Vis-NIRS technique was then modified slightly to work on a soil corer that could be used to measure root contents from deeper soil profiles (15-600 mm depth) in arable land (90-day-old maize crop grown in Fluvial Recent soils). PLSR calibration models were constructed to predict the full range of maize root densities (r^2 = 0.83, RPD = 2.42, RMSECV = 1.21 mg cm^{-3}) and also soil carbon (C) and nitrogen (N) concentrations that had been determined in the laboratory (LECO FP-2000 CNS Analyser; Leco Corp., St Joseph, MI, USA).

Further studies concentrated on improving the Vis-NIRS technique for prediction of total C and N concentrations in differing soil types within different soil orders in the field. The soil coring method used in the maize studies was evaluated in permanent and recent pastoral soils (Pumice, Allophanic and Tephric Recent in the Taupo-Rotorua Volcanic Zone, North Island) with a wide range of soil organic matter contents resulting

from different times (1-5 years) since conversion from forest soils. Without any sample preparation, other than the soil surface left after coring, it was possible to predict soil C and N concentrations with moderate success (C prediction r^2 = 0.75, RMSEP = 1.23%, RPD = 1.97; N prediction r^2 = 0.80, RMSEP = 0.10%, RPD = 2.15) using a technique of acquiring soil reflectance spectra from the *horizontal* cross-section of a soil core (**H** method).

The soil probe was then modified to acquire spectra from the curved *vertical* wall of a soil core (**V** method), allowing the spectrometer's field of view to increase to record the reflectance features of the whole soil sample taken for laboratory analysis. Improved predictions of soil C and N concentrations were achieved with the **V** method of spectral acquisition (C prediction r^2 = 0.97, RMSECV = 0.21%, RPD = 5.80; N prediction r^2 = 0.96, RMSECV = 0.02%, RPD = 5.17) compared to the **H** method (C prediction r^2 = 0.95, RMSECV = 0.27%, RPD = 4.45; N prediction r^2 = 0.94, RMSECV = 0.03%, RPD = 4.25).

The **V** method was tested for temporal robustness by assessing its ability to predict soil C and N concentrations of Fluvial Recent soils under permanent pasture in different seasons. When principal component analysis (PCA) was used to ensure that the spectral dimensions (which were responsive to water content) of the data set used for developing the PLSR calibration model embraced those of the "unknown" soil samples, it was possible to predict soil C and N concentrations in "unknown" samples of widely different water contents (in May and November), with a high level of accuracy (C prediction r^2 = 0.97, RMSEP = 0.36%, RPD = 3.43; N prediction r^2 = 0.95, RMSEP = 0.03%, RPD = 3.44).

This study indicates that Vis-NIRS has considerable potential for rapid *in situ* assessment of soil C, N and root density. The results demonstrate that field root densities in pastoral and arable soil can be predicted independently from total soil C, which will allow researchers to predict C sequestration from root production. The recommended "**V**" technique can be used to assess spatial and temporal variability of soil carbon and nitrogen within soil profiles and across the landscape. It can also be used to assess the rate of C sequestration and organic matter synthesis via root density prediction. It reduces the time, labour and cost of conventional soil analysis and root density measurement.

Acknowledgements

I would like to sincerely thank my chief supervisor, Professor Mike J. Hedley, for his great supervision, brilliant ideas, guidance, advice, effort and patience throughout this study. My sincere thanks go also to my co-supervisors Carolyn B. Hedley and Mike P. Tuohy for their kindness, effort, advice, encouragement and constructive suggestions during my study. My special gratitude goes to my co-supervisor the late Dr. Greg C. Arnold who helped me with the statistical analysis used in this thesis. I am indebted to their invaluable help in the production of journal articles produced from this study.

Many thanks also to Dr. Alec McKay (President of New Zealand Society of Soil Science) who gave me a prestigious award (Summit Quinphos Postgraduate Bursary) in 2007. I really appreciate this. My special thanks to Andreas Hueni who created SpectraProc software for spectral pre-processing. Thanks also to Dr. Dave Scotter for editing the literature review, to Ted Pinkney for manufacturing the prototype soil probe, and to Martin Hawke for his help sample collecting in our Taupo _Rotorua project.

Many thanks to Massey University Staff; Assoc. Prof. P. Loganathan, Dr. Alan Palmer, Dr. Ian Yule, and Dr. Bob Stewart who provided useful consultations; and Prof. Vince Neal, Dr. David Horne, and Assoc. Prof. Marta Camps for their kindness.

Many thanks to the staff of the Fertilizer & Lime Research Centre and Soil & Earth Sciences for their assistance and friendship throughout my study: Moira Hubbard, Bob Toes, James Hanly, Ann West, Ian Furkert, Glenise Wallace, Ross Wallace, Mike Bretherton and Lance Currie.

Many thanks to the New Zealand government who provided me with an NZAID scholarship, and also to the International Student Officers of Massey University: Sue Flynn, Olive Pimentel and Sylvia Hooker, for their support during my study.

My appreciation goes to my friends, postgraduate students: Ieda Sanches, Anja Moebis, Matthew Irwin (NZPA staff), Janice Asing, Peter Bishop, Paramsothy Jeyakumar and Indika Herath, for their friendship.

Last, but not least, my immense gratitude to my parents: the late R. Soerip Padmonobo and Siti Zaenab and my parents-in-law: the late H. Masnin Junaidi and the late Hj. B. Sasih for their support and prayers. To my lovely wife, Agustin Rahmawati, and my lovely children, Muhammad Rahman Haryokusumo and Muhammad Farras Hadikusumo, you all strengthen me, and bring and share happiness.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANN	artificial neural networks
ASD	Analytical Spectral Devices
AVIRIS	airborne visible infrared imaging spectrometer
CEC	cation exchange capacity
cf.	compare or consult
DM	dry matter
DT	de-trending
EM	electromagnetic
EMSC	extended multiplicative scatter correction
FIR	far infrared
FSL	fine sandy loam
FT	Fourier transforms
GIS	geographic information system
GPS	geographical positioning system
H method	horizontal method
INS	inelastic neutron scattering
IR	infrared
LAI	leaf area index
LIBS	laser induced breakdown spectroscopy
MeV	multiple of electron volt or one million electron volts
MLR	multiple linear regression
MSC	multiplicative scatter correction
NIR	near infrared
NIRS	near infrared spectroscopy
OM	organic matter
PAR	photosynthetically active radiation
PC	principal component
PCA	principal component analysis
PCR	principal component regression
PLS	partial least squares
PLSR	partial least squares regression

PRESS	predicted residual error sum of square
PVC	polyvinyl chloride
RD	root density
RER	ratio error range
RMSE	root mean square error
RMSECV	root mean square error of cross validation
RMSEP	root mean square error of prediction
RMSEV	root mean square error of validation
RPD	ratio prediction to deviation
RR	ridge regression
SEC	standard error of calibration
SEP	standard error of prediction
SIS	spectral interference subtraction
SL	silt loam
SMLR	stepwise multiple linear regression
SNV	standard normal variate
SNVD	standard normal variate and detrend
SOC	soil organic carbon
SOM	soil organic matter
SWIR	short wave infrared
SW-NIR	short wave near infrared
TIR	thermal infrared
UV	ultra violet
VI	vegetation indices
VIS	visible
Vis	visible
Vis-NIRS	visible near infrared spectroscopy
V method	vertical method
VNIR	visible near infrared
vs	versus
VSS	variable subset selection
WC	water content
yr	year

CHAPTER 1

General Introduction

1.1. Background

Soil carbon in its organic form plays a vital role in soil fertility, productivity and water retention (Baldock and Nelson 2000). Soil organic matter (SOM) also provides a primary source of energy for soil flora and fauna, and an important source of nitrogen, sulphur and phosphorus (Halvin *et al.* 1999).

Globally, soil contains about 1500 gigatons (Gt) of carbon, about three times the amount found in vegetation and about twice the atmospheric pool (760 Gt) (Smith 2008). Interest in soil carbon is not only soil fertility-related but there is now major interest in its role in the global C cycle. Soil organic matter has the potential to be a C sink for atmospheric carbon dioxide (CO_2), which continues to increase and is implicated in global warming. This potential is not currently fully exploited and the current loss of C from soil is estimated to be 1.6 ± 0.8 Gt yr^{-1} (Smith 2008). Since the Industrial Revolution which began around 1850, with extensive land use changes and attendant agricultural practices, soil has lost approximately 78 ± 12 Gt C to the atmosphere. This is 57 % of the total (136 ± 55 Gt) terrestrial emission, compared with 270 ± 30 Gt from fossil fuel combustion (Lal 2007; Lal *et al.* 2007). Land use changes such as converting forest, grassland, or native ecosystems, to croplands, and draining, cultivating and liming highly organic soils cause CO_2 to be emitted to the atmosphere (Smith 2008). Implementation of soil conservation practices such as no-till and reduced tillage with crop residue incorporation may be options to slow the loss of terrestrial CO_2 to the atmosphere (Paustian *et al.* 2000; Smith 2004). However in some soil-climate conditions, these conservation tillage practices may only extend the half life of soil C and not affect the total amount of soil C stored long-term (Poirier *et al.* 2009).

Cultivating deeply rooting plants is an alternative option to sequester C in deep soils (Smith 2008). Organic carbon stored deep in the soil profile can reduce its turn-over time and slow down its return to the atmosphere. Turn-over times of 1000-2000 and 9000-13000 years can be achieved if it is stored below 20 cm and 2 m, respectively (Follet *et al.* 2003). In long-term trials on the effects of fertilizer on soil carbon storage

under cereal crops, Solberg *et al*. (1998) found that cereal roots and straw incorporation contributed to increases in the soil carbon pool. These trials compared incorporation with no incorporation of straw residue and concluded that increases in SOM with no straw incorporation point to the influence of cereal roots in increasing soil C.

The amount of C sequestered is also affected by land use change (e.g. conversion of arable land to grassland) and grassland management (Conant *et al.* 2001; Soussana *et al.* 2004). Nie *et al*. (1997) reported higher root biomass under grazed than fallowed pasture. Plant-soil associations that develop greater root biomass generally have high soil C values (Saggar *et al.* 1999). Grassland plays an important role in C sequestration through the accumulation of biomass mostly below ground (Mannetje 2007). Introduction of deep-rooted grasses into field crop rotations may sequester significant amounts of C deeper in the soil profile (Boddey *et al.* 2002; Fisher *et al.* 1994). Fisher *et al*. (1994) found about half of the 237 ton ha^{-1} of C stored under a 6 year-old *Andropogon gayanus – Stylosanthes capitata* pasture of the South American savannas, in the 40-100 cm soil layer. Large deposition of C at depth was explained by the massive root system and deep-rootedness of the grasses. Compared to unimproved savanna which held 197 ton ha^{-1} C, Fisher *et al*. (1994) reported higher C under pasture of *Brachiaria humicidicola* alone (223 ton ha^{-1}) and *Brachiaria humicidicola – Arachis pintoi* (268 ton ha^{-1}). Deeper-rooting plants may have other benefits. Deeper denser root systems will be able to recapture nitrate that would have leached past shallower root systems (Bowman *et al.* 1998; Crush *et al.* 2007; Dunbabin *et al.* 2003; Thorup-Kristensen 2006). Deeper-rooting plants can also increase water use efficiency by optimising the use of subsoil water which can minimize irrigation frequency (Hedley and Yule 2008). Plants with deeper and denser root systems that have improved nutrient acquisition are required to drive the second “green revolution” in low fertility soils and help the farmer with little access to fertilizer (Lynch 2007).

Measuring root biomass or root biomass changes could provide early indications of a plant-soil system with the potential to sequester soil C. Nevertheless, measuring the root content using conventional methods (e.g. washing technique) is a tedious procedure. Measuring root density change with soil depth, such as the profile wall (Böhm 1979) and soil corer methods (Escamilla *et al.* 1991) followed by root length measurement by the line intersect method (Newman 1966), needs separation of roots

and soil, which is also a tiresome procedure. Direct discrete image analysis has been shown to be more accurate than the traditional line intersect method, and this automated method can be used for root length and diameter measurement (Lebowitz 1988). The combination of digital image analysis and soil coring used by Kokko *et al.* (1993) for total root surface area measurement, has been claimed to be a precise, repeatable and rapid method. However, these procedures require root separation from soil before the root image can be captured and analysed. A technique using a combination of digital photography following disintegration of roots and soil with an air-knife (high velocity stream of air) is claimed to speed up root separation and measurement (Bingham *et al.* 2005). Flatbed scanners used by Pan and Bolton (1991) have been proposed for speeding up the counting of separated roots. Similarly, digital image acquisition methods have been proposed for analysing root systems in soil profile walls (Ortiz-Ribbing and Eastburn 2003), but these techniques still require soil profile pits to be dug.

Measuring soil carbon using traditional methods (e.g. Walkley and Black, dry combustion) is time-consuming and labour intensive, besides there are some limitations. The Walkley and Black procedure (Walkley and Black 1934) that has been used for many years has limitations because it provides an incomplete measure of soil organic C requiring correction factors, which can vary noticeably with soil depth and soil type (Rosell *et al.* 2001). Dry combustion techniques include estimation of organic plus inorganic C. It is necessary to ensure that all carbonates and charcoal C have been removed prior to determination if total organic carbon is to be quantified (Nelson and Sommer 1996). The weight loss-on-ignition method tends to overestimate soil organic matter due to accompanying loss of inorganic constituents mainly hydrated clays affected by high temperature during heating (Cambardella *et al.* 2001). In addition, all these techniques can be costly when used for C measurement surveys of large land areas.

Near infrared reflectance spectroscopy (NIRS), which is a rapid and inexpensive technique, has been used in the laboratory to predict a variety of soil properties including soil C (Ben-Dor and Banin 1995; Chang and Laird 2002). It has been used successfully for measurement of total C, organic C, total nitrogen (N), cation exchange capacity (CEC) and moisture (Malley and Martin 2003). NIRS is based on the interaction of near-infrared radiation with soil constituents particularly the covalent

bonds of small atoms such as O, C, H and N, abundant in organic matter. It has the potential to be a rapid technique if it is properly calibrated. As portable spectrometers become available, attempts have been made to implement this technique for *in situ* and *on-the-go* (mobile) soil measurements. Most results to date have been variable due to field conditions e.g. variable soil water content (Dematthê *et al.* 2006), variable soil aggregate size (Udelhoven *et al.* 2003), coarse materials (stone and plant debris) (Mouazen *et al.* 2005), and variable distance between the soil surface and sensor head during reflectance measurement (Shibusawa *et al.* 2005; Sudduth and Hummel 1993). If field NIRS techniques can achieve improved accuracy, they will be very useful for determining spatial and temporal variability of soil C. Such techniques may provide a rapid and low cost technique for soil C assessment, soil C accounting and monitoring soil C change with land use management changes. In addition, although NIRS has been used to quantify and qualify properties of organic materials such as forage and feedstuffs (Roberts *et al.* 2004), oilseeds and grain crops (Delwiche 2004; Dyer 2004), fruits and vegetables (Slaughter and Abbott 2004) and animal by-products (Cozzolino and Murray 2004), it has not been widely used for crop root measurement. However, NIRS has been used to determine the content of *Ericales* roots and other physical characteristics and macrofossil components of peat (McTiernan *et al.* 1998).

NIRS has the potential to become a rapid and inexpensive technique for *in situ* measurement of soil properties. For this potential to be realised there needs to be robust calibration of soil spectral data against a reference set of soil property measurements. Soil coring techniques, normally used to collect samples from the soil profile for laboratory analysis, could be combined with a probe for *in situ* NIRS measurement.

This thesis reports the development and evaluation of soil coring techniques for acquiring Vis-NIRS data from soil cores in the field for the prediction of field root density and soil C and N concentrations.

1.2. Research Objectives

There are 6 main objectives of this thesis, besides specific objectives described later in each chapter. The main objectives are:

- To examine whether Vis-NIRS can be used to predict root density (glasshouse evaluation); Chapter 3.

- To investigate the ability of a Vis-NIRS technique to predict root density in the field under pasture; Chapter 4.
- To investigate the ability of a Vis-NIRS technique to predict maize root density and soil C and N concentrations in arable land; Chapter 5.
- To test the accuracy of the Vis-NIRS technique to predict soil C and N from various types of New Zealand soils; Chapter 6.
- To investigate methods for acquiring soil reflectance that can improve the prediction accuracy of soil C and N; Chapter 7.
- To examine the temporal robustness of calibration models in predicting soil C and N as seasonal soil moisture conditions vary (dry and wet month in autumn and spring); Chapter 8.

1.3. Thesis Structure

This thesis is divided into 9 chapters. The first and second chapters are general introduction and review of literature, respectively. This is followed by 6 main chapters. Each of the main chapters is a "stand alone" manuscript containing introduction, materials and methods, results and discussion, and conclusions. The last chapter is an overall summary and recommendations for future work. A brief description of each chapter is explained as follows:

CHAPTER 1 This chapter describes the background and potential for using NIRS for *in situ* measurement of root density, soil C and soil N. This chapter also highlights the main objectives and the structure of this thesis.

CHAPTER 2 Provides a review of literature which mainly describes the principle of near infrared reflectance, factors affecting soil spectral reflectance, spectral pre-processing techniques and multivariate analysis of spectral and analyte data. Previous research on the use of Vis-NIRS for C and N prediction and emerging technologies for *in situ* soil C measurement are also discussed. Finally the advantages and disadvantages of using NIRS are discussed with respect to the focus of this study.

CHAPTER 3 Describes an assessment of whether the Vis-NIRS technique can be used to predict root density. The assessment was carried out using various ryegrass root densities developed in a glasshouse study using

two contrasting soil types (Allophanic and Fluvial Recent soils).

CAPTER 4 Examines the ability of an *in situ* Vis-NIRS technique to predict root densities in permanent pasture soils from calibration models developed from field and glasshouse (Chapter 3) root density data. The ability of this technique to predict soil C and N is also presented.

CHAPTER 5 A study of the ability of the Vis-NIRS technique (developed in Chapter 4) to predict root densities in the soil profile under a 90-day-old crop of maize. This chapter also describes the prediction of soil profile C and N concentrations.

CHAPTER 6 This chapter investigates the accuracy of the Vis-NIRS technique developed in Chapter 5 for total soil C and N prediction in a wider range of soils. The assessment was conducted in the Central Volcanic Zone of the North Island, New Zealand using three soil orders; Allophanic, Pumice and Recent. The large number of soil samples collected from these soils allowed different techniques for selecting calibration and validation data sets to be compared.

CHAPTER 7 This chapter evaluates two different techniques of acquiring soil spectral reflectance (first, from a flat horizontal cross section of a soil core – **H** method; second, from a curved vertical wall of a cylindrical soil core – **V** method) for predicting soil C and N. The evaluation was carried out on Fluvial Recent soils under permanent pasture.

CHAPTER 8 This chapter examines the temporal robustness of Vis-NIR soil C and N calibration models to changing soil moisture conditions (dry and wetter months, autumn and spring).

CHAPTER 9 Comprises a summary and research outcomes and recommendations for future work.

Some repetition occurs in sections of ‘materials and methods’ and ‘results and discussion’ because each of the main chapters is a “stand alone” manuscript. The title of each main chapter is the same or similar to the title of a manuscript sent for journal publication.

CHAPTER 2

Literature Review

2.1. Introduction

Near infrared spectroscopy (NIRS) has become an extremely important analytical technique in recent years. It is used to analyze a wide range of samples from gas, liquid, to solid (Stuart 2004). It is used to measure organic composition and functional properties in e.g. crops, food, animal feed, pharmaceuticals, polymers, textiles, brewing materials, petroleum hydrocarbons, pulp and paper, as well as for medical analysis, and environmental analysis (e.g. soil, aquatic sediments, bio-solids such as manures and compost) (Malley and Martin 2003). In recent years great advances have been made in using near infrared technology to investigate soil properties (Malley *et al.* 2004).

One of the earliest NIRS successes in soil science was the successful determination of soil water content (Bowers and Hanks 1965; Dalal and Henry 1986; Kano *et al.* 1985; Slaughter *et al.* 2001). Bower and Hanks (1965) were also able to demonstrate that organic matter and particle size influenced soil reflectance. Then Dalal and Henry (1986) found that NIRS can predict organic matter and total organic nitrogen content. The use of NIRS techniques was then extended to measure other soil physical and chemical properties (Ben-Dor and Banin 1995; Chang *et al.* 2001; Cozzolino and Moron 2006; Odlare *et al.* 2005; Reeves III and McCarty 2001; Shepherd and Walsh 2002; Stenberg *et al.* 1995). The success of the technique is not only in the laboratory using static equipment, but also in the field (Mouazen *et al.* 2005; Mouazen *et al.* 2007; Shonk *et al.* 1991; Sudduth and Hummel 1993), even though it has reduced accuracy there.

Successes have also been reported in plant analysis. Research by Karl Norris in the 1960s led to the development of an instrument which could record NIR spectra and be used to determine moisture, oil, protein and starch in biological samples (Davies 1996; Norris 1996). Phil Williams in the early 1970s recognized the potential use of NIRS to segregate wheat grain according to protein content (Williams 1992; 1995). It was also used to analyze pasture and forage samples for crude protein, neutral detergent fibre, acid detergent fibre, lignin and *in vitro* dry matter digestibility (Norris *et al.* 1976) and

some minerals in forages (Shenk *et al.* 1981), tobacco (McClure 1985) and pasture grasses (Clark *et al.* 1989; Clark *et al.* 1987).

Remotely sensed (i.e. by satellites or aeroplanes) NIRS has been widely used to assess vegetation. Solomonson *et al*. (1980) reviewed several geobotanical studies and concluded that there was a relationship between leaf reflectance and metal concentrations which varied with wavelength, metal species and tree species. Using the 512-band MARK II spectroradiometer, Collins *et al*. (1983) observed a blue shift due to apparent reduced chlorophyll absorption associated with mineral deposits. It is well known that vegetation indices (VI) can be used (1) to assess biological properties of plants, plant biomass and leaf area index (LAI), and (2) as a mapping device; to assist image classification, to separate vegetated and non-vegetated area, to distinguish between types and densities of vegetation, and to monitor seasonal variations of vegetation (Campbell 1996).

An NIR reflectance spectrum represents a composite of all the optical information of the outer layer (a few millimetres thick depending on the surface and structure composition) of a substance. This complex spectrum, which is rich in information, can be interpreted on a basis of chemical and physical composition (Workman and Shenk 2004). Some parts of the chemical structure (e.g. C-H bonds) can contribute at several wavelengths in the overtone and combination bands, making the NIR spectrum highly repetitive. This information was virtually impossible to understand before modern software and multivariate statistics became available (Workman and Shenk 2004). Using multivariate analysis, the spectral and reference data can be interpreted to characterize and quantify soil constituents. Multivariate analysis is defined as “those statistical methods concerned with either the analysis of data with many variables or the display of results in many dimensions, or both” (Gower 1982).

Soils are complex mixtures of many different constituents. Traditionally such mixtures had to be subjected to extensive laboratory fractionation before the individual constituents could be characterized and quantified. However, with modern multi-variable measurement techniques and multivariate data analysis, it is possible to “unscramble the scrambled soil” without laboratory fractionation, because spectral and soil analysis data have been found to correlate very well using additive multivariate statistical models. This allows several soil properties to be measured simultaneously.

The focus of this chapter is to review the literature related to the use of visible - near infrared (Vis-NIR) reflectance spectroscopy for the measurement of soil properties. The review includes (1) explanations of terminology used for electromagnetic waves, (2) the principal of Vis-NIR reflectance and causes of light absorption, (3) factors affecting soil reflectance, (4) pre-processing spectral data and data analysis, (5) success of Vis-NIRS for soil properties measurement, (6) emerging technologies for *in situ* measurement of soil C, (7) the development of *in situ* method of Vis-NIRS, and (8) advantages and disadvantages of NIRS techniques.

2.2. Spectral Reflectance and Soil Properties

2.2.1. Terminology Used for Electromagnetic Waves

NIR refers, by convention, to the region of the electromagnetic spectrum from 780 to 2500 nm (12 820.5–4000 cm^{-1} of wave numbers), between the visible and infrared region (Malley and Martin 2003; Miller 2001; Murray and Cowe 2004; Workman and Shenk 2004). The term "near infrared" is used because the technology that exploits the neighboring spectral region of 2500–25000 nm (or 4000-400 cm) has been called "*infrared (IR) spectroscopy*" or "*mid-infrared spectroscopy*" (Miller 2001). Other possible names of the near infrared spectroscopy are "*far-visible spectroscopy*" or "*overtone vibrational spectroscopy*" (Miller 2001), but these names are not well-known. In the following figure, the visible and infrared portions of the electromagnetic spectrum are shown.

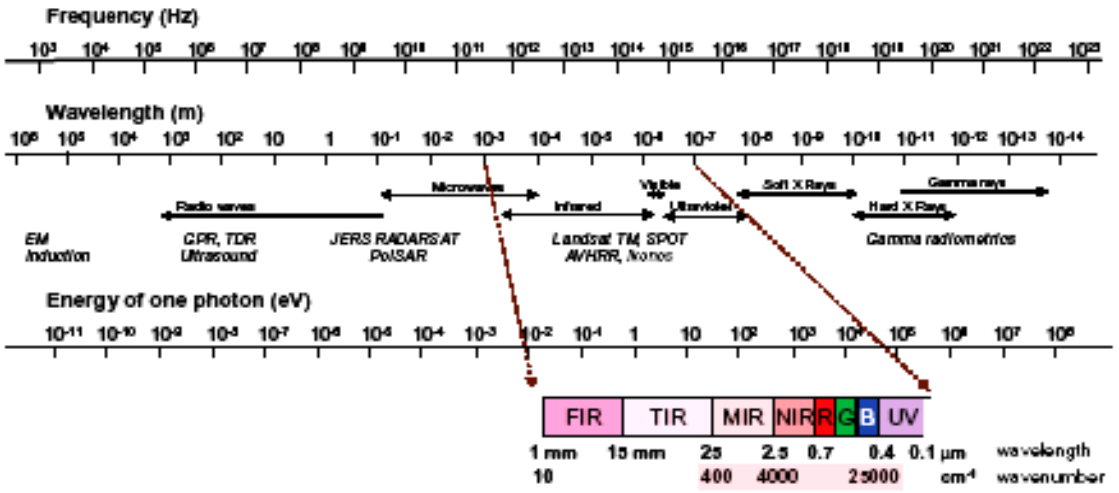


Figure 2.1 The electromagnetic (EM) spectrum highlighting the visible and infrared portions (McBratney *et al*. 2003; Viscarra Rossel *et al*. 2006).

Terminology for the different near infrared regions varies between publications and research disciplines. In remote sensing, the electromagnetic spectrum between 700 and 1300 nm is called near infrared, whilst the regions from 1300 to 3000 nm and from 3000 to 14000 nm are named mid infrared and thermal infrared respectively (Lillesand *et al.* 2004). Campbell (1996) divided infrared radiation into three regions; near infrared (720-1300 nm), mid infrared (1300-3000 nm) and far infrared (7000–1000000 nm). The ASD technical guide (1999) refers to the near infrared region as 750-2500 nm which is divided into 750-1100 nm (short wave near infrared (SW-NIR)), 1000–1800 nm (NIR1 or SWIR1) and 1800-2500 nm (NIR2 or SWIR2). The electromagnetic spectrum between 1000–2500 nm is called conventional near infrared. Ben-Dor *et al.* (1999a) divided bands related to soil reflectance into two regions; visible near-infrared (VNIR; 400-1100 nm) and shortwave-infrared (SWIR; 1100-2500 nm). The visible part of electromagnetic radiation (400-700 nm) consists of blue (400-500 nm), green (500-600 nm) and red (600-700 nm). Waveband ranges for ultra violet, visible and near-infrared according to the ASD technical guide (1999) are presented in the Table 2.1.

Ben-Dor *et al.* (1999a) stated that the VNIR-SWIR region (400-2500 nm) covers the basic processes involved in the interaction between soil and electromagnetic radiation, even though the thermal infrared regions (3000-5000 nm and 8000–12000 nm) also contain some diagnostic information about soil material.

Table 2.1 Waveband range of optical electromagnetic radiation (Analytical Spectral Devices 1999).

Region Names in Optical Electromagnetic Radiation			Wavelength (nm or µ)
Ultra-Violet (UV)	Far UV		1-200 nm
	UV C		200-280 nm
	UV B		280-315 nm
	UV A		315-400 nm
Visible (VIS)	Photosynthetically Active Radiation (PAR)	Blue light	400-525 nm
		Green Blue	525-605 nm
		Yellow Light	605-655 nm
		Red Light	655-725 nm
		Far Red	725-750 nm
Infrared (IR)	Near-IR	Short Wave Near Infrared (SW-NIR)	750-1100 nm
		Typical 1st NIR region detector (NIR1) or (SWIR1)	1000-1800 nm
		Typical 2nd NIR region detector (NIR2) or (SWIR2)	1800-2500 nm
		"Conventional" Near Infrared (NIR)	1000-2500 nm or 1.0-2.5 µm
	Mid-IR		2.5-50 µm
		Thermal (emitted)	8-15 µm
	Far-IR		50-1000 µm

2.2.2. Near Infrared Reflectance and Causes of Light Absorption

Near infrared reflectance spectroscopy which was originally used as an analytical chemistry method is now also widely used in remote sensing. The technology is based on interpreting the interaction between electromagnetic radiation and material (Miller 2001) which is controlled by specific vibrations of the atoms in a molecule (Stuart 2004). When radiation interacts with a material surface, a portion of the radiation which corresponds to particular bonds is absorbed by the material, while the remaining which does not interact with the bonds is either transmitted through the bulk of material or reflected from the surface.

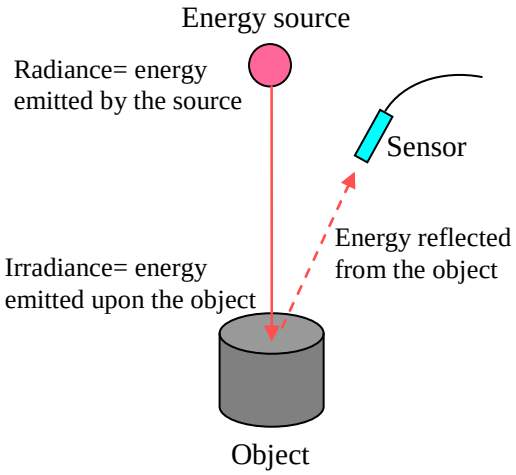


Figure 2.2 Interaction between energy source, object and spectral sensor.

The ratio of the reflected radiation to the total radiation falling upon the surface is defined as *reflectance* (Chandra and Gosh 2006; Lillesand *et al.* 2004), which is not an intrinsic material property, but rather the result of a measurement relating to the above mentioned ratio (Baumgardner *et al.* 1985). The *spectral reflectance* is mathematically defined as:

$$\rho_\lambda = \frac{E_R(\lambda)}{E_I(\lambda)} x100 = \frac{energy\ of\ \ wavelength\ \lambda\ reflected\ \ from\ the\ object}{energy\ of\ \ wavelength\ \lambda\ incident\ upon\ the\ object} x\,100$$

where ρ_λ is the spectral reflectance expressed as a percentage.

A soil reflectance spectrum can be defined as a set of data or a graph providing the relative intensity of reflected radiation as a function of wavelength (Ben-Dor *et al.* 1999a). The reflected intensity is reported relative to the intensity of the illuminating radiation. Practically, soil reflectance can be determined as the ratio of energy reflected from the soil surface to the energy reflected by a bright diffuse reference material (Ben-Dor *et al.* 1999a). A barium sulfate ($BaSO_4$) coated disc is commonly used as the diffuse reference material.

NIRS works based on the absorption of light in the near infrared region (780-2500 nm) by overtones and combination bands of covalent bonds between small atoms particularly C-H, N-H, and O-H (Malley and Martin 2003; Workman and Shenk 2004). Groups such as C-O, C-C, and C-N do not show absorption spectra in the NIR, except

under unusual situations where the weak absorbances are not dominated by the much stronger absorbers, i.e. C-H, N-H and O-H bonds (Kramer *et al.* 2004). All bonds have specific vibrational frequencies, and NIR absorption can be used to describe (1) the location of absorption in terms of wavelength, (2) the amplitude of the absorption peak (relative intensity), and (3) the width of the peak describing its intensity (bandwidth) (Workman and Shenk 2004).

A molecule can absorb incoming infrared radiation if the incoming radiation has the same frequency as one of the fundamental modes of vibration of the molecule (Stuart 2004). Also, the molecule must possess a specific feature, for instance an electric dipole moment must change during the vibrations (Anderson *et al.* 2004; Clark *et al.* 1990; Stuart 2004). A molecule can absorb photons if there is precisely enough energy provided to promote the atom or molecule from one energy stage to a higher state, or when an atom or a molecule makes transition between energy states (especially from a lower to higher state) (Ben-Dor *et al.* 1999a). The photon energy absorbed will raise the rotational, vibrational or electronic energy of the molecule by a discrete amount (Miller 2001; Murray and Williams 1987). Also, the intensity of energy absorbed decreases as each overtone increases (Murray and Williams 1987).

As NIR spectra are primarily the result of overtones and combination bands, they are complex and not directly interpretable compared to other spectral regions occurring in the mid and far-infrared region, which are mostly fundamental bands (Sandorfy *et al.* 2007; Workman and Shenk 2004). Compared to the medium infrared region (MIR), the NIR region is dominated by broader signals, rather than sharp peaks, due to additive effects of two or more bonds (combinations of absorbances) at each wavelength (Workman and Shenk 2004). The fundamentals absorption is the most intense absorption of energy and occurs in the mid-infrared region (2500–15000 nm). Each higher overtone and combination band is typically 10-100 times weaker than the fundamental bands (Sandorfy *et al.* 2007), consequently the spectrum can be quite complex (Clark 1999). However, it is measured easily by reflectance spectroscopy (Clark 1999).

Traditionally, frequencies of fundamental vibrations are labelled with the Greek letter ν and a subscript (e.g. ν_1, ν_2, ν_3) while overtones can be written as double or triple

times ν (e.g. $2\nu_1$, $2\nu_1$, $3\nu_2$), and combination bands are labelled as $\nu_1+\nu_2$, $\nu_2+\nu_3$, $\nu_1+\nu_2+\nu_3$, and so on (Herzberg 1945).

Vibration of the atoms of a molecule can involve either a change in bond length (*stretching*) or bond angle (*bending*) (Stuart 2004). Stretching vibration consists of symmetric and asymmetric stretching, while bending vibration can be divided into deformation, rocking, wagging and twisting (Stuart 2004). Symmetric vibration is generally weaker than asymmetric vibration, because symmetrical molecules have fewer "infrared active" vibrations than asymmetrical ones (Stuart 2004). Near infrared radiation gives information about materials based on the stretching and bending of CH, OH and NH bonds within the compounds (Workman and Shenk 2004).

Another type of absorption is called *electronic* absorption (Clark 1999), involving electrons. Absorption of a photon by an electron causes the electron to jump up to a higher energy level which is translated into different wavelengths (lower wavelength). Photon is light properties of both waves and particles which can be as "carriers" and "transfers" of energy. Emission of the photon occurs as a result of a change in energy state from high to low (Clark 1999). The bands associated with the electron transition generally occur in the *ultraviolet and visible portions* of the spectrum (Ben-Dor *et al.* 1999a). Compared to electronic absorption, v*ibrational* absorption is associated with infrared bands and is more complicated (Ben-Dor *et al.* 1999a), because it still involves the absorption of energy (light) by electrons. As an electron moves about the atom(s), the bonded atom is drawn or repulsed, creating a vibration motion (frequency) that can only absorb discrete energy amounts; the same frequency as the vibration motion. As the molecule vibrational energy level increases, the energy to be absorbed also increases.

2.2.3. Soil Chromophores Affecting Soil Reflectance

The term "soil chromophore" is used by Ben-Dor *et al.* (1999a) to describe chemical and physical parameters of a substance that significantly affect the shape and nature of its reflectance spectrum. The soil spectrum *per se* is a complex mixture of various spectral components (such as minerals, organic matter, salts, water) each of which has unique electronic transitions (caused by its atoms) and vibrational stretching

and bending (caused by structural groups of atoms), leading to specific spectral signatures at particular wavelengths.

There are a number of soil chromophores affecting soil reflectance, e.g. moisture, organic matter, particle size, iron oxides, mineral composition, soluble salts, parent material, and cation exchange capacity (Baumgardner *et al.* 1985). Ben-Dor *et al.* (1999a) categorized soil chromophores into threee major groups: [1] minerals (mostly clay and iron oxides), [2] organic matter (living and decomposed) and [3] water (solid, liquid and gas phases). Later, Ben-Dor (2002) described the six most common chromophores (clay mineral, carbonates, organic matter, water, iron and soil salinity) in the soil environment and their relationship with the Vis-NIR-SWIR spectral region.

Soil reflectance can be acquired in the laboratory, in the field, and via remote sensing (Ben-Dor *et al.* 1999a). In the laboratory, under controlled condition, it is relatively easy to achieve a high signal to noise ratio. However, in the field, some problems may occur due to e.g. variation of viewing angle, illumination change, and soil roughness (Ben-Dor *et al.* 1999a). Factors such as physical surface condition, surface cover, and surface expression of subsoil characteristics can also affect soil reflectance (Baumgardner *et al.* 1985). Also soil reflectance acquired from air-borne satellites are affected by atmospheric attenuation (water vapor, atmospheric gasses, and aerosols) (Ben-Dor *et al.* 1999a).

2.2.3.1. Water and Hydroxyl

Water and hydroxyl bonds have an overriding effect on diagnostic Vis-NIR absorption by minerals (Clark 1999). The hydroxyl (OH^{-1}) bond is the strongest absorber in the NIR region and usually predicted reasonably well in soil (Malley *et al.* 2004). Water (H_2O) with 3 atoms in its polyatomic molecules (N) will have 3N degrees of translational freedom (Stuart 2004) and so there are 3N-6 = 3 fundamental vibrations (Clark 1999). Fundamental vibrations of water in the vapor phase occur at: 2738 nm (ν_1, symmetric O-H stretch), 6270 nm (ν_2, H-O-H bend), and 2663 nm (ν_2, asymmetric O-H stretch), and in the liquid phase at: 3106 nm (ν_1), 6079 nm (ν_2) and 2903 nm (ν_3) (Clark 1999). First overtones of the O-H stretches occur at about 1400 nm, and the combination of the H-O-H bend with O-H stretches occurs near 1900 nm (Clark 1999). So, a mineral which has a 1900 nm absorption band contains water, but a mineral which

has 1400 nm band and has no 1900 nm band indicates only hydroxyl is present (Baumgardner *et al.* 1985; Clark 1999). Water and the hydroxyl ion will cause decreased reflectance of the 1400 nm band. O-H absorption in OH–bearing minerals can also take place typically near 2700 to 2800 nm and anywhere between 2670 and 3450 nm. Metal-OH bend occurs near 10000 nm and combination metal OH bend plus OH stretch occurs near 2200 to 2300 nm (Clark *et al.* 1990).

The prominent features of water bands may occur at different wavebands in the clay minerals montmorillonite and kaolinite. Montmorillonites exhibit very strong water absorption bands at 1400 and 1900 nm (Bowers and Hanks 1965; Hunt and Salisbury 1970; 1971). Kaolinites exhibit very strong hydroxyl absorption bands at 1400 and 2200 nm and a weaker water absorption band at 1900 nm (Clark *et al.* 1990; Hunt and Salisbury 1970; 1971). The band at 1900 nm is the combination of the v_2 and v_3 fundamental bands and the most sensitive to water as well as the best measurement of soil moisture content (Bowers and Hanks 1965).

Soil water content affects soil reflectance. The influence of water on soil reflectance was investigated several decades ago (Baumgardner *et al.* 1985; Bowers and Hanks 1965; Hunt and Salisbury 1971). Generally the higher the soil water content, the lower the reflectance (Baumgardner *et al.* 1985; Bowers and Hanks 1965), as depicted in Figure 2.3. As can be seen the strong absorption features of water occur at 1400 and 1900 nm. More recently Mouazen *et al*. (2005) reported decreased soil reflectance as water content increased from 0.0, 0.05, 0.10, 0.15, 0.20 to 0.25 in a laboratory investigation. Similarly Bogrekci and Lee (2004) found that as soil moisture content increased, soil reflectance decreased. All these studies indicate that water is a strong light absorber. However in other cases, the addition of water in endmember minerals of montmorillonite and kaolinite did not always reduce reflectance in all bands, but tended to decrease reflectance in particular band regions (Clark 1981). “An endmember in mineralogy is a mineral that is at the extreme end of a mineral series in terms of purity” (www.wikipedia.org).

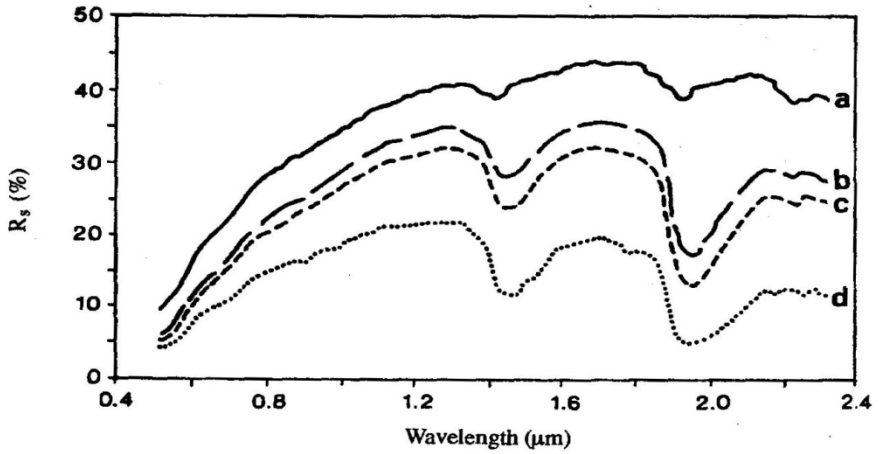


Figure 2.3 Spectra of Typic Hapludalf soil at four different moisture tensions: a (oven dry), b (15 bar, 1500 kPa), c (0.3 bar, 30 kPa) and d (0.1 bar, 10 kPa) (Baumgardner *et al*. 1985).

The shape of soil reflectance is not only affected by the presence of strong water absorption bands at 1450 and 1950 nm, but also occasionally by weaker water absorption bands at 970, 1200 and 1770 nm (Baumgardner *et al.* 1985), and at 2200 nm (Clark *et al.* 1990; Hunt and Salisbury 1970) (see Figure 2.4). The 970, 1200 and 1770 nm bands are overtones and combinations of three fundamental vibration frequencies of the water molecule which occur beyond 2500 nm (Bowers and Smith 1972). These bands were found corresponding to spectra of water of a few millimeters in thickness (Lindberg and Snyder 1972). The 2200 nm band is a vibrational mode of the hydroxyl ion (Hunt and Salisbury 1970). Using smectite minerals, Cariati *et al*. (1983) showed shifts in the OH absorption features at 1400, 1900 and 2200 nm. The OH absorption features can occur at around 950 nm (very weak), 1200 nm (weak), 1400 nm (strong) and 1900 nm (very strong) for a mixed system of water and minerals (Ben-Dor 2002).

Sharply defined bands of water at 1450 and 1950 nm indicate that water molecules are in a well-defined location (ordered sites), while broad bands at these wavelengths indicate that they are relatively unordered (Hunt and Salisbury 1970), which is frequently the case of naturally occurring soils (Baumgardner *et al.* 1985) (Figure 2.4).

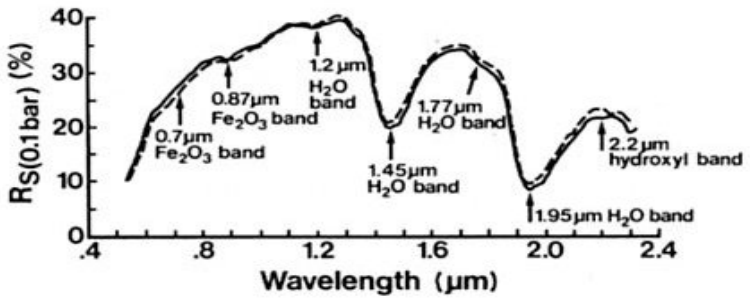


Figure 2.4 Spectral curves duplicate samples of a highly reflected soil annotated with prominent iron and water absorption bands; ____ (sample 1), ------- (sample 2) (Baumgardner *et al*. 1985).

Successful soil moisture content measurement in the laboratory using NIRS has been reported by many workers (Bowers and Hanks 1965; Dalal and Henry 1986; Kano *et al.* 1985; Slaughter *et al.* 2001). More recently, mobile NIR spectroscopy has been used in the field to estimate of soil moisture content (Mouazen *et al.* 2005; Mouazen *et al*. 2007).

Water may affect the ability of NIRS to measure other soil properties. Bogrekci and Lee (2004) investigated the effects of three soil moisture contents (4%, 8%, and 12%) on the reflectance spectra of sandy soils with different phosphorus (P) concentrations using UV-Vis-NIR (225-2550 nm) reflectance spectroscopy. The results showed that removing the moisture effect from the spectral signal considerably improved soil P predictions. Barthes *et al*. (2006) found better prediction accuracy of total soil carbon and nitrogen with oven-dried (40^{o}C during 24 h) than with air-dried samples. Malley *et al*. (2002) also reported better predictions of organic matter, NH_4-N, K, SO_4-S and Ca content with dry soils (dried at 40^{o}C) than with field moist soils. In contrast, Fystro (2002) found organic C, total N and their potential mineralization were predicted better using NIRS of thawed-moist grassland soils than dried samples, but no explanation was proposed. Similarly, Malley *et al*. (2002) found better prediction for total nitrogen in moist soils than in dry soils. However, Chang *et al*. (2005) found reasonable accuracy for both air-dried and moist soils predicting total C, organic C, total N, CEC, % clay and moisture, although the predictions were slightly more accurate for air-dried soils than for moist soils. They state that under field conditions soil water might adversely affect the ability of the NIR technique to predict soil properties because strong water absorption bands frequently mask the peaks associated with organic functional groups.

Other studies have reported similar prediction accuracy for total carbon with dried and field-moist unground samples (Chang *et al.* 2005; Reeves III *et al.* 2002).

2.2.3.2. Organic Matter

Organic matter and its constituents play a major role in the biological, chemical and physical characteristics of soil and it has a strong effect on soil reflectance. Aber *et al*. (1990) explain that organic matter and its decomposition products affect the reflectance properties of mineral soil. In general, as organic matter increases, overall soil reflectance decreases in the band region of 400-2500 nm (Hoffer and Johannsen 1969). Bowers and Hanks (1965) found greater reflectance over 400-2500 nm for a soil oxidised using 30% H_2O_2 compared to a non-oxidized one. Similarly, Demattê *et al*. (2003) showed higher reflectance intensity throughout the band of 450-2500 nm after organic matter removal using 30% H_2O_2. Mathew *et al*. (1973b) found higher reflectance at 500-1200 nm for soil treated to remove organic matter using 10% H_2O_2 compared to the non-treated sample, but soil reflectance with no-H_2O_2 treatment was higher in the 1200-2600 nm band. Stoner and Baumgardner (1981) give five soil spectral curves for soils containing different amounts of organic matter and iron; *A = Organic dominated* (OM = 5.61%, Iron oxides = 0.76%, moisture at 0.1 bar = 41.1%, color = black), *B = minimally altered* (OM = 0.59%, Iron oxides = 0.03%, moisture at 0.1 bar = 17.0%, brown), *C = Iron affected* (OM=1.84%, Iron oxides = 3.68%, moisture at 0.1 bar = 28.2%, strong brown), *D = Organic affected* (OM = 3.3%, Iron oxides = 0.81%, moisture at 0.1 bar = 27.3%, dark brown), and *E = Iron dominated* (OM = 2.28%, Iron oxides = 25.6%, moisture at 0.1 bar = 33.1%, dark red) (Figure 2.5). It can be seen that the soils with the higher OM tend to have lower reflectance.

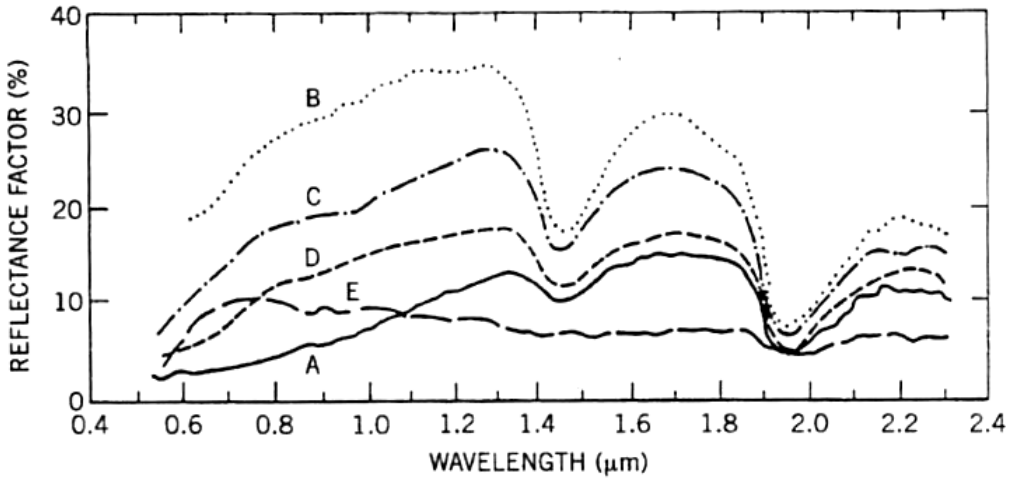


Figure 2.5 Soil spectral curves of five mineral soils; A = Organic dominated, B = minimally altered, C = Iron affected, D = Organic affected, and E = Iron dominated (Stoner and Baumgardner 1981). See text for details.

More recently Demattê *et al*. (2004a) demonstrated lower reflectance intensity throughout the entire spectrum due to higher organic matter content (Figure 2.6 a and d). As OM absorbs energy, it promotes lower reflectance intensity (Henderson *et al.* 1992). Demattê *et al*. (2004a) also showed the increase of reflectance intensity in the band of 550-1900 nm from the soil surface (depth A) to the deepest soil sample (depth B and C) as a result of the lower organic matter content in the subsurface layer (Figure 2.6 b). In Figure 2.6 c, it can be seen that spectrum concavity at 850 nm occurs as a result of iron content (Demattê *et al.* 2004a). Lower iron content made the curve concavity weaker. Demattê and Garcia (1999) studied the effect of the form of iron on the shape of the spectral reflectance curve; they found that crystalline iron promoted the concavity, whereas amorphous iron did not. Interestingly, the presence of OM on the surface soil masked the concavity caused by the crystalline iron, and decreased spectral reflectance intensity (Figure 2.6 c).

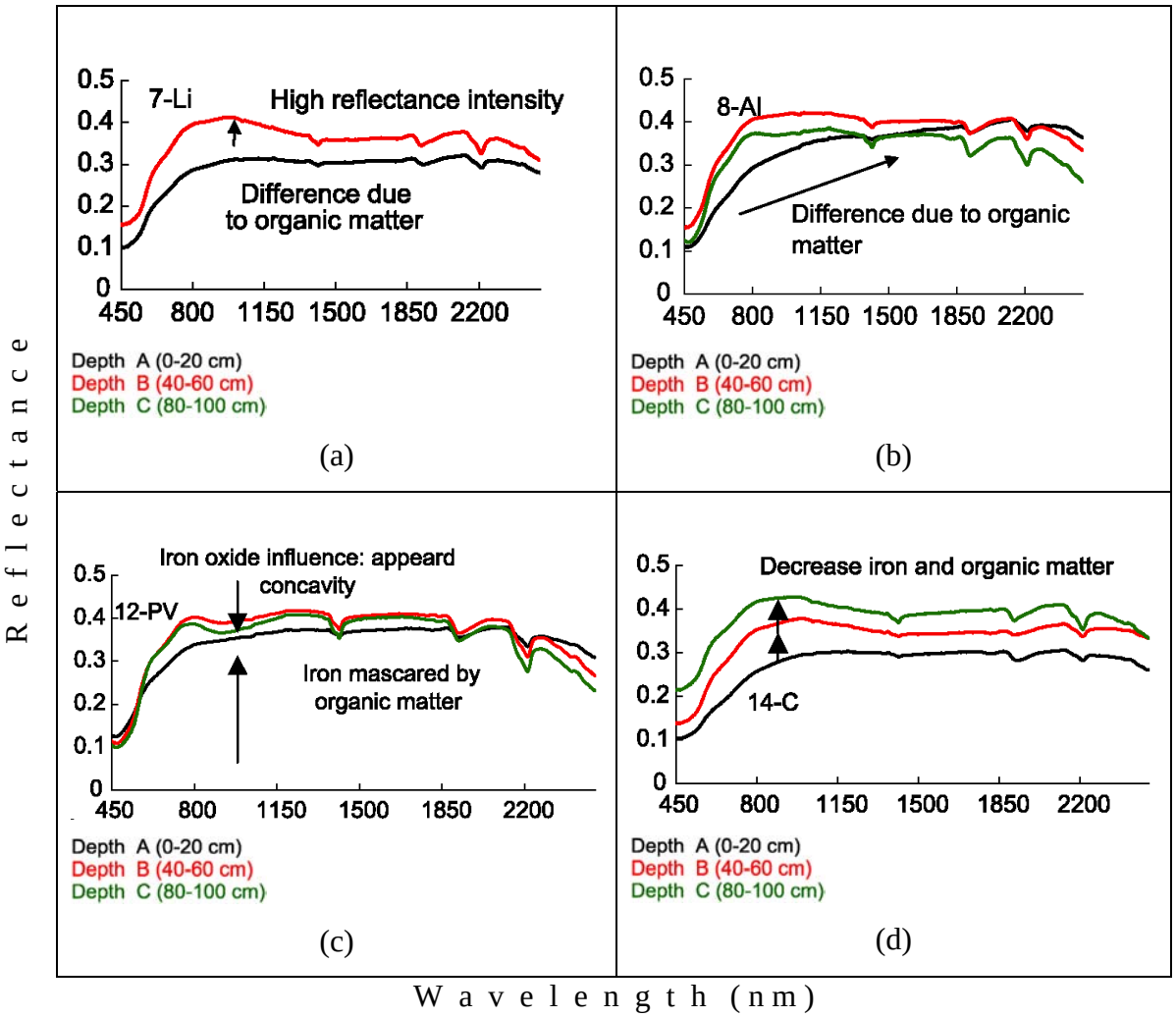


Figure 2.6 Mean spectral reflectance curves at three depths with different OM and Iron content of four soil types; (a) lithic distrochrept – Li, (b) typic fluvent – Al, (c) typic paleudult – PV and (d) typic distrochrept – C (Demattê *et al.* 2004a).

Age or different decomposition stages of organic matter influence soil reflectance patterns. Stoner and Baumgardner (1981) showed different spectral patterns for three organic soils with different decomposition stages; fibric – slightly decomposed, hemic – intermediately decomposed, and sapric – completely decomposed (Figure 2.7). Interestingly, fabric material which had the highest OM content (84.8%), compared to hemic (54.4%) and sapric (76.4%), shows the highest reflectance. High reflectance in the NIR region of fabric soil material resembles the reflectance of senesced leaves (Gausman *et al.* 1975), which is related to tissue morphology containing a high number of air voids; as interfacing with more air-cells increases reflection (Baumgardner *et al.* 1985).

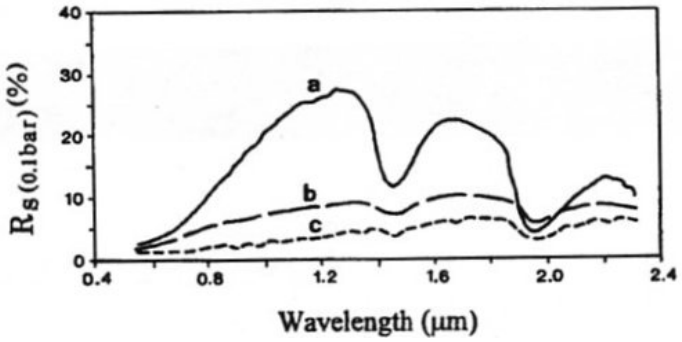


Figure 2.7 Spectral curves of three organic soils displaying different levels of decomposition; fibric (a), hemic (b) and sapric (c); Stoner and Baumgardner (1981).

Ben-Dor *et al.* (1997) also found significant changes in Vis-NIR-SWIR spectra (400-2500 nm) occurred with increasing age of organic matter (Figure 2.8). Using separated cattle manure (CSM) and grape marc (CGM) organic matter, they investigated the decomposition process over 392 days and found significant changes across the entire spectrum of the Vis-NIR-SWIR region. It can be seen that numerous absorption features occur that relate to the high number of organic matter functional groups. The most significant change occurred in the slope value over the Vis-NIR region (400-1100 nm). The slope tended to decrease with increasing decomposition age. The slope change during the initial decomposition stage (0-60 days) was strongly affected by decomposition activity. So caution is needed when using the slope at 400-1100 nm as an indicator of organic matter content, because the slope might be changed due to decomposition activity (age factors), and not because of a change in organic matter content. In practice, the different ages of decomposing plant debris mixed with soils should also be considered when measuring soil organic matter using reflectance spectroscopy.

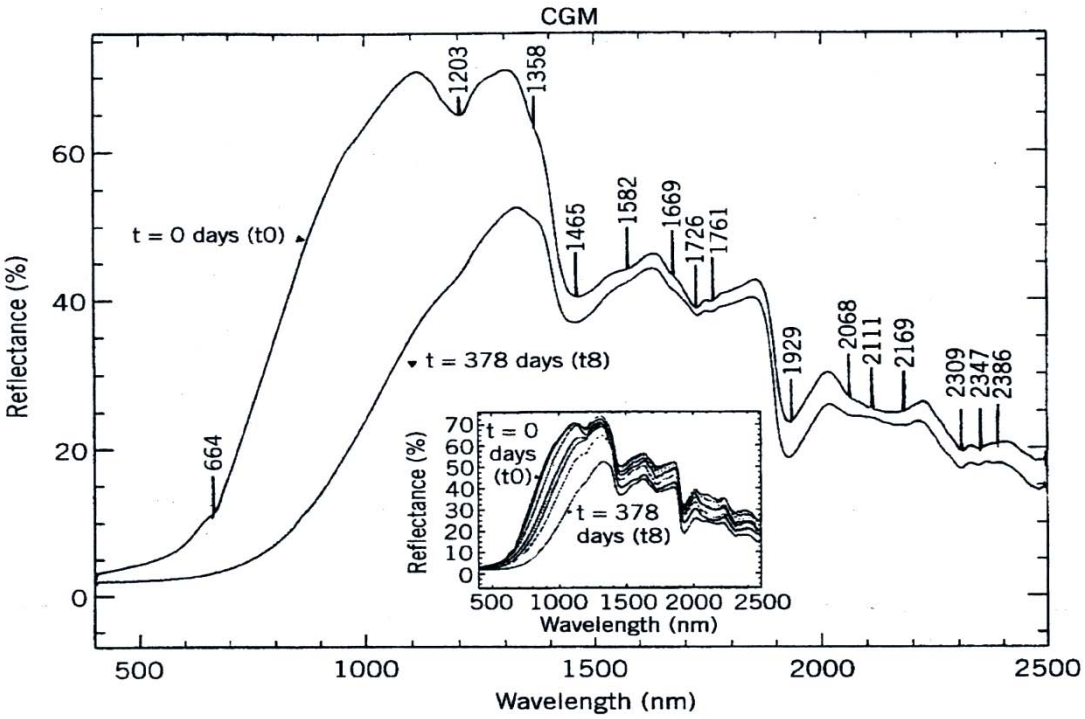


Figure 2.8 The reflectance spectra of CGM that represents two extreme composting stages t_0=0 days and t_{378}=378 days. The inset shows the spectral intermediate decomposition stages (Ben-Dor *et al.* 1997).

Baumgardner *et al*. (1970) found a negligible effect on soil reflectance when the organic matter decreased below 2.0%, however, when organic matter content exceeded 2.0%, it considerably influenced soil spectral properties. Montgomery (1976) cited by Baumgardner *et al*. (1985) found that organic matter did not mask the contribution of other soil properties to soil reflectance if organic matter content was as high as 9%. However, He *et al*. (2005) were able to accurately predict organic matter content using NIR-PLSR technique in soils with organic matter ranged between 1.06 and 1.65%. As stated by Ben-Dor (2002), accurate assessment of organic matter in soils can be obtained even though mineral soil consists of relatively low organic matter content (around 0-4%), but it requires high spectral resolution data across the whole Vis-NIR-SWIR region (400-2500 nm).

Since organic matter has spectral activity in the whole Vis-NIR-SWIR region (400-2500 nm), it is understandable that different conclusions have been reached by different workers as to the best band for measuring organic matter in soils. Krishnan *et al*. (1980) found 623.6 and 564.4 nm as the optimal wavelengths for predicting organic matter content, and they did not find an absorption peak of organic matter (on the sample used)

in the region of 800-2400 nm. Mathew *et al*. (1973a) found the best correlation between organic matter and its reflectance in the band range of 500-1200 nm, while Beck *et al*. (1976) recommended the 900 to 1220 nm band as the best for mapping organic carbon in soils. Stoner (1979) cited by Baumgardner *et al*. (1985) found that the strongest correlation between organic matter and its spectra occurred at visible wavelengths. Similarly, Vinogradov (1981) developed an exponential model in order to predict humus content in the upper horizon of ploughed forest soils by using bands ranging from 600 to 700 nm. However, reflectance in the SWIR region (1702-2052 nm) was used by Dalal and Henry (1986) and they were able to predict both organic matter and total organic nitrogen content in Australian soils using those wavelengths. Likewise, Morra *et al*. (1991) identified the SWIR region (1726 and 2426 nm) as suitable for assessing organic matter composition. Chang and Laird (2002) showed good prediction of organic C, inorganic C, total C and total N using the 1100-2498 nm segment of the 400-2498 nm band. Cozzolino and Moron (2006) were able to accurately predict soil organic C in different particle size fractions using the entire Vis-NIR bands (400-2500 nm). Using the whole Vis-NIR bands (400-2500 nm), Fystro (2002) successfully predicted organic C and its potential mineralization.

Stuart (2004) stated that organic molecules may be investigated by using NIR reflectance because organic matter may contain C-H, O-H, and N-H bonds which fall in the near infrared region. The common near infrared bands of organic compounds are presented in Table 2.2.

Table 2.2 Common near-infrared bands of organic compounds.

Wavelength (nm)	Assignment
2200-2450	Combination C-H stretching
2000-2200	Combination N-H stretching, combination O-H stretching
1650-1800	First overtone C-H stretching
1400-1500	First overtone N-H stretching, first overtone O-H stretching
1300-1420	Combination C-H stretching
1100-1225	Second overtone C-H stretching
950-1100	Second overtone N-H stretching, second overtone O-H stretching
850-950	Third overtone C-H stretching
775-850	Third overtone N-H stretching

Source: Stuart, 2004.

2.2.3.3. Clay and Non-Clay Minerals

Clay minerals (*phyllosilicate* minerals) have a major influence on soil reflectance (Ben-Dor *et al.* 1999a). The structure of clay minerals comprises two basic units; (1) the Si tetrahedron formed by a Si^{4+} ion surrounded by four O^{2-} ions in tetrahedral configuration, and (2) the Al octahedron formed by an Al^{3+} ion surrounded by four O^{2-} and two OH^{-} ions in the octahedral configuration. Those units may be joined by adjacent Si tetrahedral sharing all three basal corners and by Al octahedrons sharing edges to form layer type of 1:1 (one tetrahedral and one octahedral sheet) or 2:1 (two octahedral sheets and one tetrahedral sheet). The only functional mineral group active in the VNIR-SWIR region is hydroxide (OH) (Hunt 1979). This OH group can be found as a part of the mineral structure (mostly in octahedral sheet, termed *lattice water*) or in water molecules attached directly and indirectly to the mineral surface (termed *adsorbed water*) (Hunt and Salisbury 1971). Hunt and Salisbury (1970) found three major regions that are generally active for clay minerals. For smectite, they are around 1300-1400 nm, 1800-1900 nm, and 2200-2500 nm, as discussed in Section 2.2.3.1. In Ca-montmorillonite, the lattice OH features were found at 1410 nm and at 2206 nm, while OH features of free water in Ca-montmorillonite were found at 1456 nm and at 1910 nm and at 1978 nm. Ben-Dor *et al*. (1999a) pointed out that those band positions can change slightly for different smectites, depending on their chemical composition and surface activity.

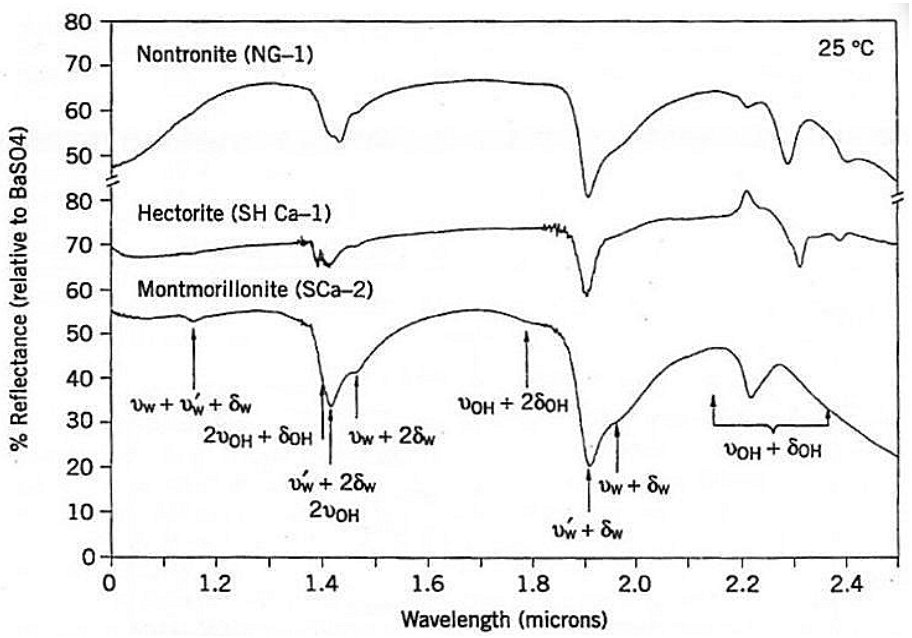


Figure 2.9 Reflectance spectra of three pure smectite end members in the NIR/SWIR region at room temperature (25°C) (Ben-Dor *et al*. 1999a).

Kaolonite and illite are also NIR-spectrally active minerals because they contain octahedral OH sheets. In the case of kaolinite, the lattice OH signals at around 1400 and 2200 nm are relatively strong, whereas the signal at around 1900 nm is very week (Grove *et al.* 1992). This is because of the relatively low surface area and adsorbed water content of this mineral. In gibbsite the signal at 1400 nm is even stronger than in kaolinite. Non-smectite layer-clay minerals (kaolinite, chlorite, gibbsite, vermiculite and illite) are shown in Figure 2.10 below.

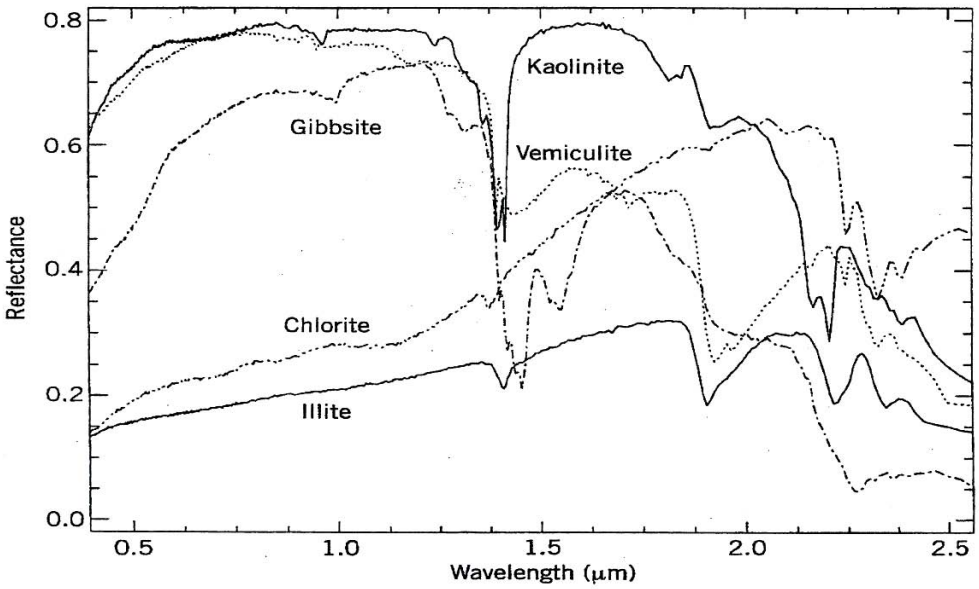


Figure 2.10 Reflectance spectra of representative pure non-smectite clay minerals (Grove *et al.* 1992).

Formation of either 1:1 or 2:1 layer type minerals is related to the degree of weathering and the chemical nature of the soil environment (Ben-Dor *et al.* 1999a). Minerals of type 1:1 (kaolinite), which may be developed from 2:1 minerals, are predominant in highly weathered soil. They usually occur in warm temperate soils with suitable ratio of silica and alumina accumulation. They may turn to gibbsite under acid condition. Minerals type 2:1, such as smectite, are usually formed at the middle stage of the weathering process. Smectite is the most active clay mineral due to its high specific surface area and electrochemical reactivity. Smectite and illite and have a tendency to be formed in cold regions. Illite and vermiculite are usually present in younger materials. Concavity alteration at a band around 850 nm is due to the influence of iron content and can be used to determine the extent of weathering in soils (Demattê *et al.* 2004a) (Figure 2.11). A concave shape in the 850 nm band indicates the presence of crystalline iron and strong weathering, while the absence of a concave shape in this band indicates less weathered soils. However, this concave shape may also be masked due to high organic matter content (Demattê *et al.* 2004a). It can also be seen from Figure 2.11 that a stronger absorption band at around 1900 nm indicates the presence of a 2:1 clay mineral type. In contrast, if absorption at 1900 nm band is weak, this indicates the presence of 1:1 mineral type. The presence of kaolinite can also be seen at 2200 nm band. The presence of gibbsite is indicated at 2265 nm (Demattê *et al.* 2004a).

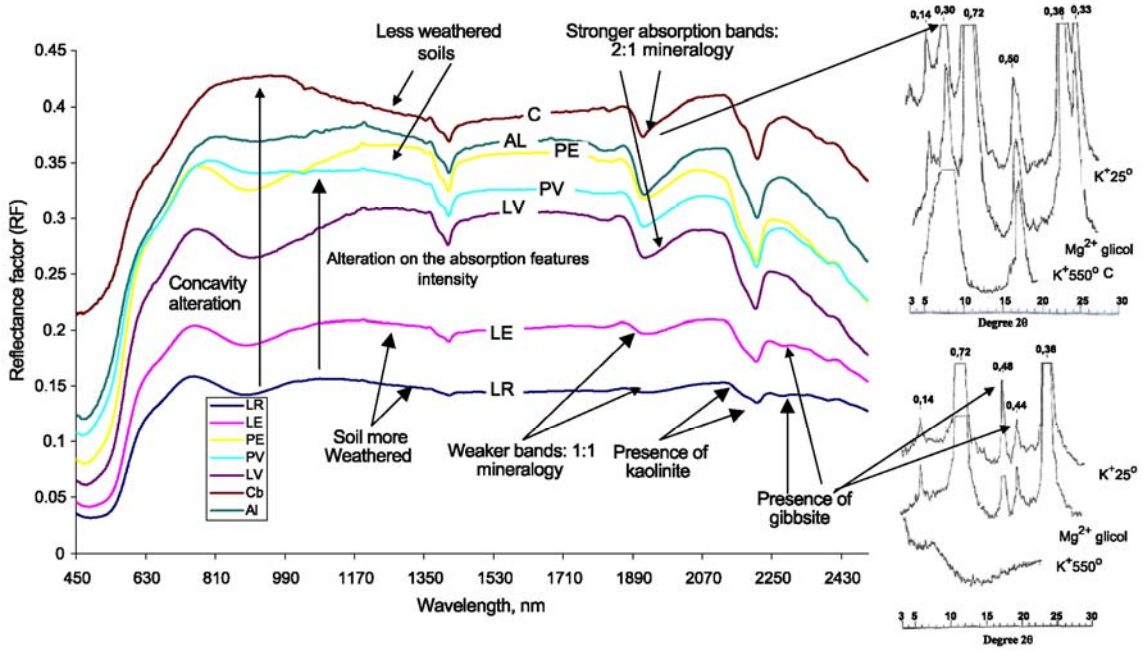


Figure 2.11 Comparison of mean spectral reflectance curve from 80 to 100 cm depth of seven soils (typic distrochrept – C; typic fluvent – Al; rhodic paleudult – PE; typic paleudult – PV; typic haplorthox – LV; typic haplorthox – LE; typic haplorthox – LR) and diffractograms showing presence of minerals (Demattê *et al.* 2004a).

Some common non-clay minerals (grouped into five groups; silicates, phosphates, oxides and hydroxides, carbonates, and sulfides and sulfates) can be found in soil (Ben-Dor *et al.* 1999a). Quartz is spectrally inactive in the Vis-NIR-SWIR region and diminishes other spectral features in the soil mixture, whereas oxides such as iron oxide (hematite) are strongly spectrally active, mostly in the visible region due to the crystal field and the charge transfer mechanism (Ben-Dor *et al.* 1999a). Phosphate and sulfate minerals can be found in soils in small quantities as apatite and gypsum, where each of them has unique spectral features. The reflectance spectra of representative pure non-clay minerals in soils are presented in Figure 2.12 below.

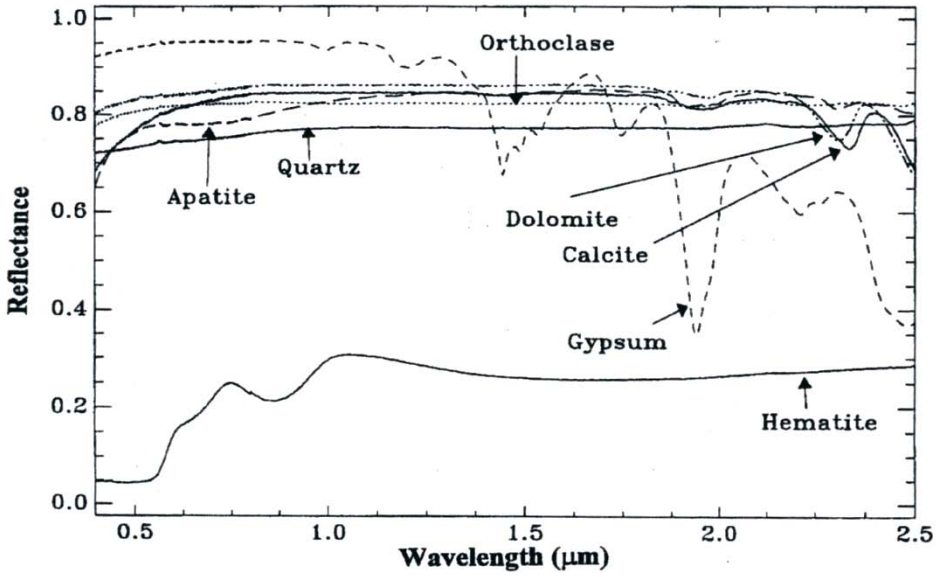


Figure 2.12 Reflectance spectra of representative pure non-clay minerals in soil (Grove *et al.* 1992).

2.2.3.4. Carbonates

Carbonates can also display diagnostic vibrational absorption bands due to the planar CO_3^{2-} ion (Clark 1999). Fundamental vibrations of this carbonate ion occur at 9407 nm (ν_1, symmetric stretch), 11 400 nm (ν_2, out-of-plane bend), 7067 nm (ν_3, asymmetric stretch), and 14 700 nm (ν_4, in-plane bend). Its strongest overtone and combination bands occur in the near infrared at 2500-2550 nm ($\nu_1+2\nu_3$) and 2300-2350 nm ($3\nu_3$) and its three weaker overtone and combination bands occur near 2120-2160 nm ($\nu_1+2\nu_3+\nu_4$ or $3\nu_1+2\nu_4$), 1970-2000 nm ($2\nu_1+2\nu_3$), and 1850-1870 nm ($\nu_1+3\nu_3$). Ben-Dor and Banin (1990b) correlated reflectance spectra and carbonate concentration and found three wavelengths the best predictors of calcite; 1800, 2350 and 2360 nm.

Carbonates, especially calcite and dolomite, are found in soil formed from calcareous parent materials. Their fine particle size means they can play a major role in soil chemical processes in the root zone. High concentrations of carbonates may cause iron fixation in soil, and hinder chlorophyll production, whereas, the absence of carbonates in soil may negatively affect soil buffering capacity (Ben-Dor *et al.* 1999a).

2.2.3.5. Particle Size

Particle size influences spectral reflectance of soils. Using a textural sample range of coarse silty to very coarse sandy bentonite (0.059-1.495 mm particle size range) and

kaolinite (0.022-2.68 mm), Bower and Hanks (1965) showed the rapid exponential increase in reflectance around 400-1000 nm with decreasing particle size in both minerals. They noticed considerable increases occur at sizes less than 0.4 mm diameter. They also observed that as the particle size decreases, the surface apparently became smoother. So, they suggested that the reflectance changes are a function of surface roughness. They believed that lower reflectance of larger particle sizes was due to trapped light within the large pores. Baumgardner *et al*. (1985) also stated that low reflectance can be caused by trapping most of the incident light into inter-aggregate spaces of coarse aggregates having an irregular shape, as shown in Figure 2.13.

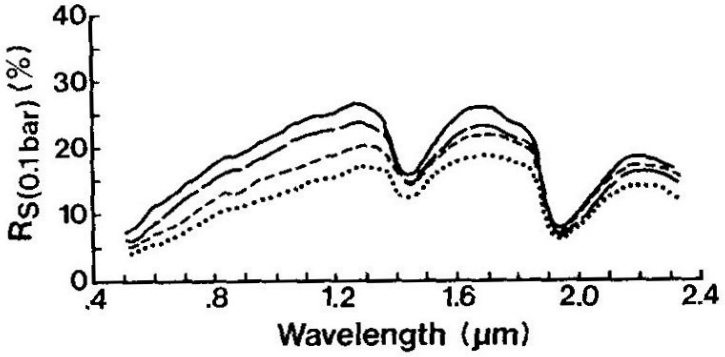


Figure 2.13 Reflectance of different particle sizes; ___ fine sands, _ _ _ fine loamy sands, ---- loamy sands, ….. loamy coarse sands (Baumgardner *et al*. 1985).

NIR spectroscopy has been used for approximate soil texture analysis (Ben-Dor and Banin 1995b; Chang *et al.* 2001; Cozzolino and Moron 2003). Sorensen and Dalsgaard (2005) demonstrated a relationship between prediction error and clay content in their calibration set. By using PLSR they found lower prediction error (1.7-1.9%) when the percent clay was 1-26%, and a larger prediction error (3.4%) in the range 3-74% clay. They concluded that the NIR technique may be useful in predicting not only the clay fraction but also the silt and sand fractions.

Soil particle size can affect the NIRS prediction accuracy of other soil properties. Bogrekci and Lee (2005) investigated the effects of soil particle size on the prediction of phosphorus concentration using UV-Vis-NIR reflectance spectra of sandy soils. Pure sandy soil was graded into three particle sizes; 125 μm (fine), 250 μm (medium), and 600 μm (coarse). Phosphorus application rates were 0, 12.5, 62.5, 175, 375, 750, and 1000 mg kg-1. Results showed that the prediction of P using individual calibration models for each soil particle size produced lower prediction errors. Better results (lower

SEP) were found by removing the influence of particle size effect, compared to non-removal of particle size effect. Using particle size < 8 mm and < 2 mm, Chang *et al*. (2001) found greater variation in surface physical properties (such as size and shape of soil aggregate) of samples with particle size < 8 mm. In order to get a more accurate result, they suggest that samples should have similar particle size. Barthes *et al*. (2006) found more accurate prediction of total carbon and nitrogen content of soil using oven-dried finely ground samples of less than 0.2 mm compared to oven-dried 2 mm sieving samples. The prediction accuracy that they found was more affected by grinding than by drying. The positive effect of fine grinding on prediction accuracy of total carbon and nitrogen has also been reported by other workers in clayey soils and several soil types (Dalal and Henry 1986; Reeves III et al. 2002; Russel 2003). In contrast, Fystro (2002) reported lower prediction accuracy with finer samples (0.5 mm ground) than with coarser samples (4 mm sieved) in coarse-textured soils. These apparently contradictory results might be explained by different light transmission in clayey and coarse-textured soils due to the different size and arrangement of soil particles (Chang *et al*. 2001). Barthes *et al*. (2006) believed that aggregate destruction upon grinding might increase the prediction accuracy, because the wide range of well developed heterogeneous aggregates in many clayey soils might affect reflectance adversely. In contrast, in coarse-textured soils with little aggregate development, however, grinding might break quartz (or other primary minerals) up which might alter sample composition and lead to the peeling of coatings, which might create a more heterogeneous particle size distribution (Barthes *et al*. 2006).

2.2.3.6. Iron

Iron affects the shape of soil reflectance in the Vis and NIR region as a result of electronic transition of iron cations (3+, 2+), either as the main constituent (e.g. as iron oxides) or as impurities (e.g. as iron smectite) (Ben-Dor 2002). Added colours (red, yellow, or brown) produced by Fe^{3+} oxides are because of selective colour absorption in the VIS region caused by transitions in the electron shell. Hunt (1971a) showed that the Fe^{2+} (ferrous) and Fe^{3+} (ferric) ion are spectrally active in the Vis-NIR region; the ferrous ion at 430, 450, 510, 550 and 1000 nm and the ferric ion at 400, 700 and 870 nm. It is well known that soil colour can be significantly changed with a small amount of iron oxides. Hematite soils are reddish, and goethite soils are yellowish brown (Ben-

Dor 2002). Baumgardner *et al*. (1985) demonstrated soil spectra with various amount of total Fe_2O_3 (Figure 2.14).

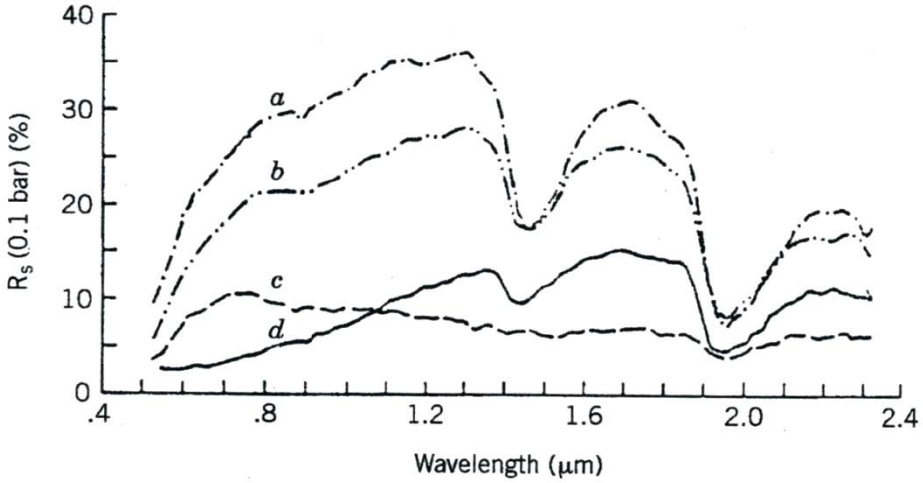


Figure 2.14 Reflectance spectra of soils with different textures but exhibiting iron absorption bands. a. fine sand, 0.20% Fe_2O_3, b. sandy loam, 0.64% Fe_2O_3, c. silty clay loam, 0.76% Fe_2O_3, d. clay, 25.6% Fe_2O_3 (Baumgardner *et al.* 1985).

More recently Demattê *et al*. (2004a) demonstrated different reflectance shapes due to different iron contents (Figure 2.15). High Fe_2O_3 content in soil LR (188.2 g kg^{-1}) and LE (122.2 g kg^{-1}) gave these two soils a very low reflectance factor (between 0.03 and 0.15) compared to the others; LV (92.3 g kg^{-1}), PE (85.4 g kg^{-1}), PV (50.5 g kg^{-1}), and Al (39.6 g kg^{-1}). The presence of magnetite which is a very common iron oxide in soils influenced the lower reflectance intensity. The presence of iron oxides can also be seen by the presence of a concave shape between 450 and 850 nm (Figure 2.15). In addition, soils richer in goethite showed narrower concavity shape and less absorption in the band around 450 nm compared with soils rich in hematite which had wider concavity and greater absorption in this band (Demattê *et al.* 2004a). Hematite (α-Fe_2O_3) has Fe^{3+} ions in octahedral coordination with oxygen. While, geothite (α-FeOOH) also has Fe^{3+} ions in octahedral coordination, and so different locations of distortions along with the oxygen ligand (OH) result in the main absorption feature at around 900 nm (Ben-Dor 2002).

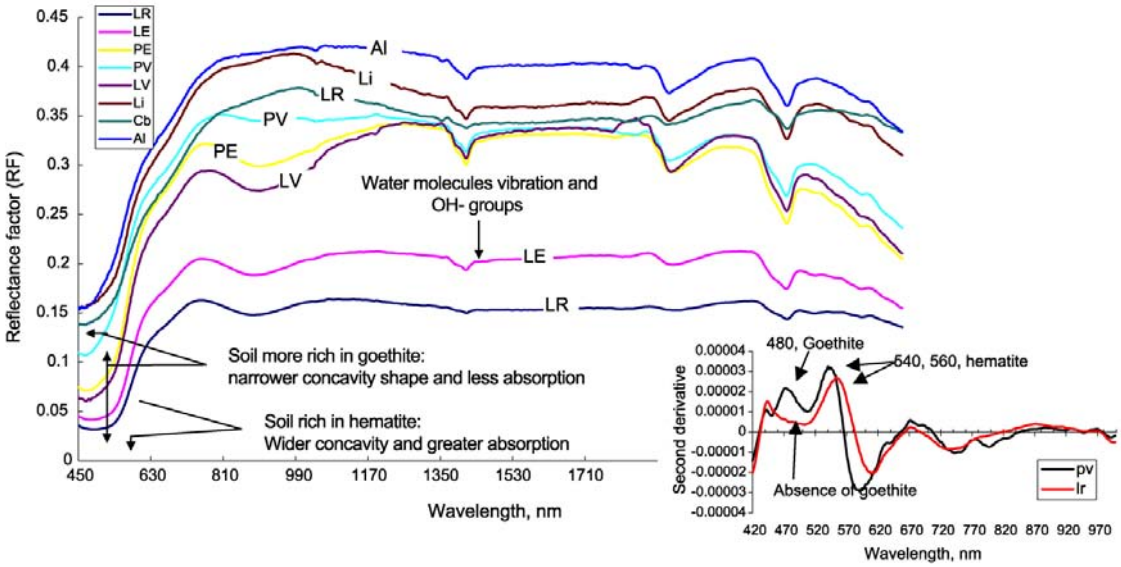


Figure 2.15 Comparison of mean spectral reflectance obtained from 40 to 60 cm depth of eight soils (typic fluvent – Al; lithic distrochrept – Li ; typic haplorthox – LR; typic paleudult – PV; rhodic paleudult – PE; typic haplorthox – LV; typic haplorthox – LE; typic distrochrept – Cb) and second derivative (Demattê *et al.* 2004a).

Iron-related features have been found at 640 nm (Obukhov and Orlov 1964), near 900 nm (Grove *et al.* 1992), and at 425, 1025, 1075 nm (Ben-Dor and Banin 1995), as well as at 1266.1, 2259.8, 2294.9 nm (Ben-Dor and Banin 1990a). Due to the complexity of the iron component and its interaction with other soil components, high spectral resolution is required to distinguish it (Ben-Dor 2002). Grove *et al*. (1992) demonstrated the characteristically different spectral shapes of four common iron oxides found in soils (Figure 2.16).

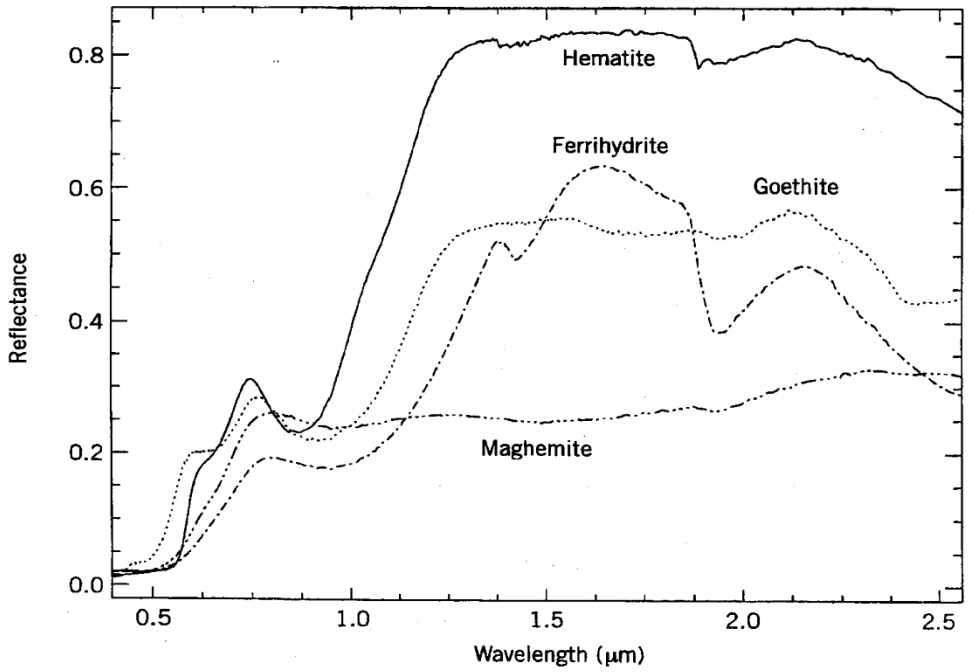


Figure 2.16 Reflectance spectra of representative iron oxides in soils (Grove *et al.* 1992)

2.2.3.7. Soil Salinity

Salt is almost spectrally featureless (Hunt *et al.* 1971c). However, its existence in soils can be detected indirectly using spectroscopy due to its interaction to other soil components (e.g. organic matter, particle size distribution, EC, water, soil structure, iron) (Ben-Dor 2002). Because salt is not a direct chromophore, soil salinity can only be characterised in soils using high spectral resolution data (Hirschfeld 1985). Hirschfeld (1985) found subtle changes of the hydrogen bond in water affected by salt. Salts may be detectable in high-moisture samples due to their ability to change H bonding, resulting in band shifts (Burns and Cziurczak 1992). Although salt is spectrally featureless, Hick and Russell (1990) believe that particular wavelengths in the Vis-NIR-SWIR region can provide information about saline-affected areas. Csillag *et al.* (1993) found six channels provided information about salinity status. Rao *et al.* (1995) showed spectral reflectance variation from salt-affected soils. More recently Ben Dor *et al.* (1999b) produced a soil map using an HSR airborne sensor, in which soil salinity emerged via soil properties such as electrical conductivity (EC) and pH.

Salt-affected soils can also be indirectly detected from vegetation reflectance. Cotton leaves grown in saline soils had higher chlorophyll than that in low-salt soil (Gausman *et al.* 1970). Wang and Shannon (2001) reported soya bean canopy

reflectance in the NIR range was significantly reduced by salinity. The reduction was attributed to increases in specific leaf mass caused by salinity, and can be delineated with the simple ratio vegetation index (SRVI), with 660 and 830 nm as the most sensitive waveband combination. The spectral reflectance of a *Spartina alterniflora* canopy showed a negative correlation between soil salinity and spectral vegetation indices (Hardisky *et al.* 1983). However, Poss *et al.* (2006) found that vegetation indices (VI) were not strongly correlated with leaf area index changes attributed to water and salinity stress treatments for both alfalfa and wheatgrass.

2.2.3.8. Other (Parent) Materials

There are other materials, not already discussed, which also influence the shape of soil reflectance spectra. Soil reflectance variations have been reported which related to soil type and parent material (Wu and Wang 1991). Mathews *et al.* (1973b) showed contrasting reflectance curves for soils developed from limestone (see Section 2.2.3.4), shale, sandstone and glacial deposits. Soils developed from these parent materials were successfully separated with a high degree of accuracy (Mathews *et al.* 1973a). Reflectance intensities of igneous rocks as reported by Hunt *et al.* (1973b; 1973c; 1974) decreased from acid to intermediate to basic. Ultra basics igneous rocks showed a well-defined ferrous iron effect in a band near 1000 nm (Hunt *et al.* 1974). It was reported by (Demattê *et al.* 2004a) that parent material trends along toposequences and weathering sequences modified soil characteristics and respective spectral reflectance data. Qing *et al.* (2004) showed that the reflectance of blue clayed paddy soil derived from sediments is higher than that of red paddy soil derived from the Quaternary red earth, probably due to different clay mineral assemblages and/or iron oxides as discussed in Sections 2.2.3.3 and 2.2.3.6. YingFeng and TengHui (1995) found different reflectance curves from different parent materials, including flat latosols developed from basalt, sloping type of paddy soils, and steep type of red and yellow soils. They believed that these spectral differences were caused by natural zonal factors such as slope. They are also probably formed by different aeration conditions leading to different oxidation and reduction processes, resulting in different patterns of spectral reflectance.

2.3. Chemometrics for Data Analysis

Chemometrics is the use of mathematics or statistics for analysis of chemical data. Data obtained in association with spectroscopy-measurement techniques can be divided into *spectral* and *analytical* data. Spectral data contain a great deal of physical and chemical information which cannot be extracted straightforwardly for two reasons, one intrinsic and the other practical (Ozaki *et al.* 2007). The intrinsic reason is because NIR spectra consist of a number of bands arising from overtones and combination modes overlapping with each other causing multicollinearity. The practical reason is appears because NIR spectroscopy often involves “real-world” samples which may produce relatively poor signal-to-noise ratios, baseline fluctuations, and severe overlapping bands due to the various components present. To overcome those two difficulties, spectral processing is needed. Whilst, analytical data are data collected using reliable techniques mostly involving chemical substances.

Chemometrics, particularly multivariate data analysis, is useful for the handling and processing the large data sets (spectral and analytical data) produced by spectral reflectance studies. It typically includes the use of spectral data processing software plus statistics software. It can extract rich and complex information from NIR spectra and overcome multicollinearity problems. The well-known multivariate data analysis methods are for example principal component analysis (PCA), principal component regression (PCR), and partial least square regression (PLSR). There is a basic assumption that any soil spectrum is a linear combination of the spectral signatures of its components (Ben-Dor 2002). Based on this assumption, reflectance spectroscopy usually utilizes an empirical, multivariate linear model to quantitatively predict various soil properties from its spectral data.

2.3.1. Spectral Pre-Processing

The spectral data should be pre-processed before any statistical analysis is carried out. This is due to by the fact that pre-processing spectral data may improve NIRS prediction accuracy (Barnes *et al.* 1989). Many available software packages e.g. GRAMS, MATLAB, Unscrambler, Pirouette, Vision, Win-DAS, Win-ISI (Westerhaus *et al.* 2004) and more recently ParleS (Viscarra Rossel 2008) can be used to pre-process the spectral data. Hueni (2006) created a software package called SpectraProc that can

also be used for spectral data pre-processing, discriminant analysis, PCA and data base structuring. All processing steps of the SpectraProc software including data reduction can work on high resolution (hyperspectral) 1 nm spaced spectral data. The stages following data reduction are then conducted using data spaced at 5-10 nm, depending on the reduction parameters. The processing steps in this review will focus on the following operations: waveband filtering, spectral smoothing, data reduction, derivative calculation and averaging (Figure 2.17).

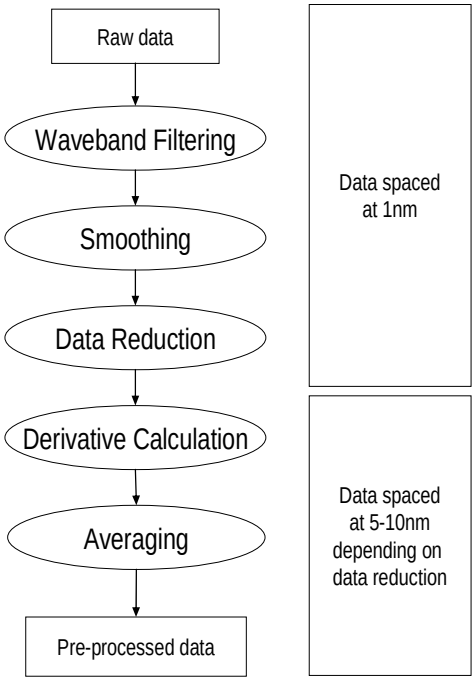


Figure 2.17 Spectral pre-processing steps.

2.3.1.1. Waveband Filtering

Waveband filtering is applied to hyperspectral data because a low signal to noise ratio can be found due to atmospheric interference; e.g. from water vapour at 1350-1440 nm, 1790-1990 nm and 2360-2500 nm (Thenkabail *et al.* 2004). By decreasing the distance between an object and the spectroscopy sensor, this interference can be minimized; e.g. by using a contact probe to collect leaf or soil reflectance. However, even though the sensor is placed close to the object, noisy regions may occur at the edges of a spectrum due to the low intensity of incoming light. Kooistra *et al.* (2003) removed spectral data beyond 2400 nm because they were too noisy due to a low level of incoming radiation. Mouazen *et al.* (2005) removed the edges of noisy regions, so

spectral data collected at 306.5-1710.9 nm were reduced to 401.4-1699 nm. Removing the noisy spectra is termed *waveband filtering*. An example of noise at the edges of spectra (Mouazen *et al.* 2006) is shown in Figure 2.18.

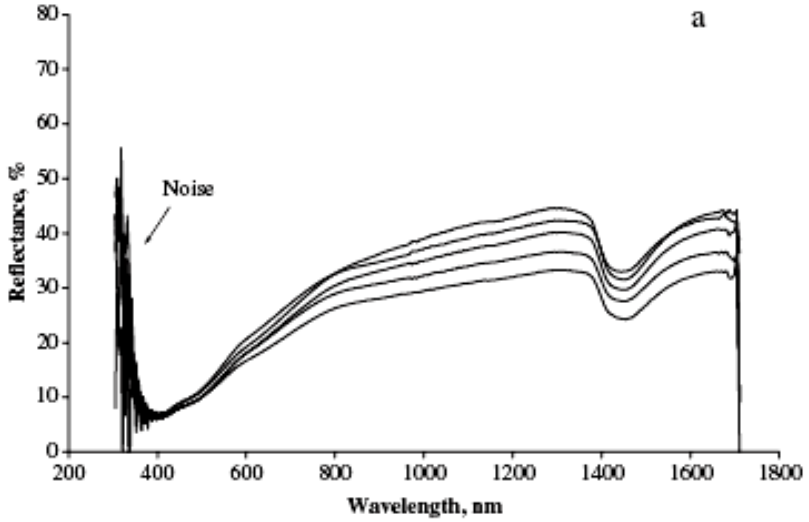


Figure 2.18 Edges of a spectrum showing extreme noise (Mouazen *et al*. 2006b).

2.3.1.2. Spectral Smoothing

Smoothing is a very important second step after waveband filtering, because the output from any radiometer contains noise (Milman 1999).

Noise is unwanted signals generated by the instrument used to make a measurement (Milman 1999). According to McClure (2001) photometric noise in a computerised spectrophotometer is a measure of total system noise inherent in detectors, amplifiers, and A/D converter. Noise can be caused by internal and external factors (McClure 2001). Internally, noise may be produced by the instrument due to the internal heating of the instrument's circuit, while examples of external factors are atmospheric humidity, light leakage and temperature change (McClure 2001). Noise is differentiated into two types; thermal noise and shot noise (Milman 1999). The increased temperature of the resistor in the electrical circuit of the radiometer will emit radiation and produce thermal noise. The predominant source of noise in the visible and infrared radiometer is called shot noise. Shot noise can also be generated because the reflected spectra is variable caused by external factors associated with illumination of the object. Williams and Norris (2001) classified noise into two; random noise and systematic noise. Random noise is caused by the radiometer's detector and electronic amplifiers, whereas

systematic noise is generated by external factors such as a worn bearing, changes in the humidity of the instrument, electrical interference, light leaks, inadequate power supply, and change in temperature. The systematic noise may cause more serious problems because it causes errors in the accuracy of all constituents (McClure 2001).

There are two common smoothing methods used in reducing noise, either *running mean* (*moving average* or *boxcar smooth*) or a *Savitzky-Golay filter* (Adam 2004), although several other methods, e.g. exponential averaging (Massart *et al.* 1988), binomial, Fourier, and Gaussian (Hruschka 2001) smoothing, can also be used. Wavelet Transformation is another smoothing technique (Schmidt and Skidmore 2004). They found this technique (Discrete Wavelet Transformation – DWT) more superior for noise reduction of vegetation spectra compared to other smoothing techniques (Mean, Median, Savitzky-Golay, Non-decimated DWT and Cubic Spline).

A running mean is calculated by replacing the value at each point by the mean of the values in a wavelength interval surrounding it. The window is moved one data point by dropping the last point and adding the next to the window (Massart *et al.* 1988). The greater the number of points averaged, the greater the degree of smoothing, which might cause signal distortion and subsequent loss of information (Adam 2004).

In contrast, a Savitzky-Golay filter works by fitting the spectrum in a wavelength interval with a polynomial by a least square method (Savitzky and Golay 1964). It is one of the most popular smoothing functions applied to hyperspectral data (Tsai and Philphot 1998). It is a technique for smoothing the primary data, and obtaining smoothed derivatives, using convolution arrays derived from the coefficient of least-squares-fit formulae (Madden 1978). A symmetric convolution function is moved over the data points and the midpoint of this moving window is the data point to be smoothed (Savitzky and Golay 1964). Convolution is the mathematical process of implementing a moving average technique (Adam 2004). An advantage of symmetric convolutes is the avoidance of a shift of the peak position and the preservation of the peak symmetry (Massart *et al.* 1988). The convolution process is depicted in the following equation:

$$y_j^* = \frac{\sum_{i=-m}^{i=+m} C_i Y_{j+i}}{N}$$

Where:

y_j^* = smoothed data point

C_i = convolution coefficient

Y_{j+i} = original data point

N = moving window size (-m…+m)

The Savitzky-Golay method performs well and involves relatively simple algorithms to remove noise (Tsai and Philphot 1998). The amount of smoothing is controlled by two parameters: the polynomial order and the number of points used to compute each smoothed output value (Savitzky and Golay 1964). A polynomial of a certain order is fitted to *N* data points, where *N* is defined by the filter size. An advantage of this filter is the ability to calculate smoothed derivative data in one operation. The zero-order derivative is the Savitzky-Golay smoothed data. The figure below depicts a raw and Savitzky-Golay smoothed soil spectrum.

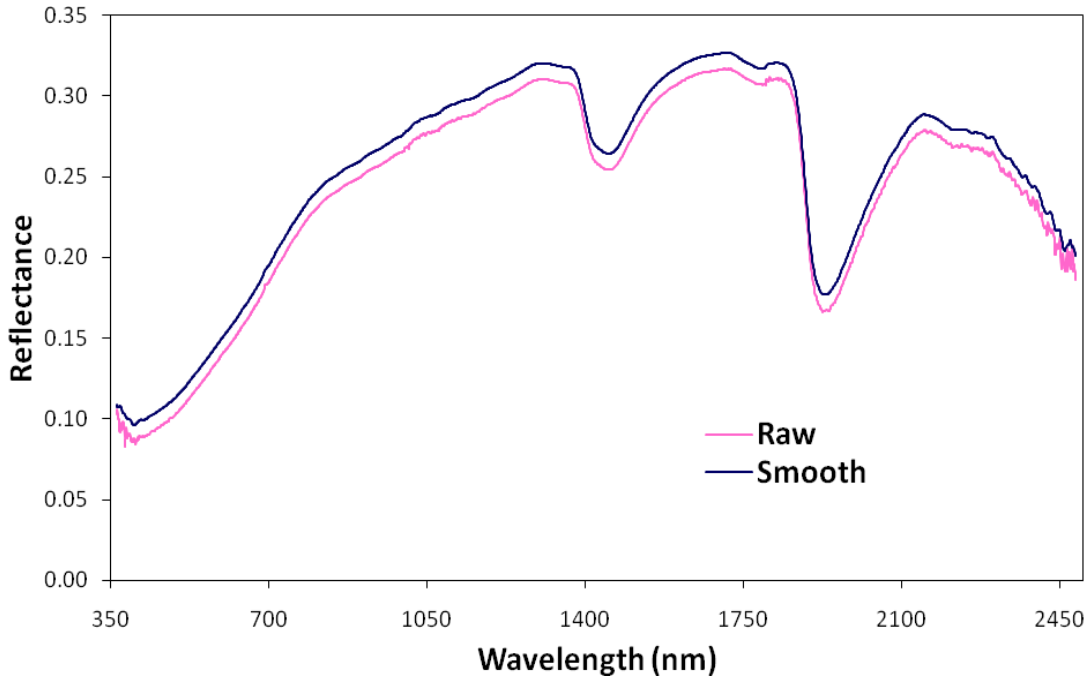


Figure 2.19 Example of raw and smooth spectrum by Savitzky-Golay filter (offset by 0.01 to enable a visual comparison). The example was processed using SpectraProc software (Hueni and Tuohy 2006) from reflectance of a Fluvial Recent soil in Chapter 7.

2.3.1.3. Spectral Data Reduction

The raw ASD spectral data contain complex information in tightly packed spectral regions (e.g. the NIR spectral region), and one chromophore (for example the C-H bond) can influence many parts of the spectrum as overtone and combination bands. This makes the NIR spectrum rich in information, but highly repetitive or heavily over-sampled with a high degree of correlation between many neighbouring bands. This data redundancy can be reduced by data reduction.

One of the simplest methods of data reduction is by averaging several adjacent spectral points. Chang and Laird (2002) collected spectral samples from 1100-1498 nm with a 2-nm interval and reduced spectral data into 140 new data by averaging 5 adjacent spectral points, so one data point represents a 10-nm interval. Tomasson *et al*. (2001) reduced the number of spectral data by averaging every 50-nm band between 250-2500 nm. Henderson *et al*. (1992) averaged every 10 contiguous 1-nm bands, resulting in 210 data points from the 400-2500 nm data collected. McCarty and Reeves (2006) averaged every four data points from 1100 to 2498 nm, resulting in 175 data points for calibration development.

Other methods of data reduction are (a) Hyperion sensor synthesizing and (b) down-sampling by factors.

The Hyperion sensor, as flown on the Earth Observing-1 (EO-1) satellite, captures data from 400 to 2500 nm with bandwidths of 10 nm (Liao *et al.* 2000; Pearlman *et al.* 2003). By reducing the ASD data to Hyperion-like bands, the dimensionality is reduced by approximately a factor of 10. Also the band synthesizing process achieves a smoothing of the data due to the response function of the simulated sensor elements. Hyperion synthesizing is an option available that can be used as a method for data reduction and implicit smoothing. The above-mentioned spectral range and resolution of the Hyperion sensor is a generalization. Waveband centres do not lie at whole number frequencies, bandwidths are not sharply defined and the sensitivity of the sensor is not uniform over the bandwidth. The actual Hyperion sensor has some uncalibrated bands; of 242 Hyperion bands, typically 198 bands were provided in the calibrated data (Pearlman *et al.* 2003). The process of spectral band synthesis is based partly on the algorithms used by Zanoni *et al.* (2002). The spectral response function of the sensor

elements is well approximated with a Gaussian function (Zanoni *et al.* 2002). Hyperion wavebands are calculated by:

$$r_{Hyp\_j} = \frac{\sum_{i=\mu-range}^{\mu+range} c_i \cdot r_i}{\sum_{i=\mu-range}^{\mu+range} c_i}$$

where

r_{Hyp_j} = the synthesized reflectance value of the j-th Hyperion band.

μ = the mean value over the band.

c_i = the coefficient determined by the Gaussian function f(x) for x = wavelength of i-th ASD band.

r_i = reflectance value of i-th ASD band.

$_{range}$ = defines the range of values symmetrically to the middle ASD waveband. The middle ASD waveband is the one closest to the average wavelength of the j-th Hyperion band.

Data reduction by downsampling is a straightforward process. According to a given factor n, every n^{th} waveband is chosen and the in-between bands discarded. Shepherd and Walsh (2002) and Shepherd *et al*. (2003; 2005) selected every 10^{th} nm value from 350-2500 nm spectral data in order to reduce and match the volume of data more closely to the spectral resolution of the instrument (3-10 nm). Using the 10-nm space data, they did not find loss in prediction performance compared with the 1-nm space (Shepherd *et al.* 2005). Signal processing theory states that a downsampling process must be preceded by a smoothing filter. The combination of smoothing and downsampling is termed *decimation* (Fliege 1994).

2.3.1.4. Derivative Calculation and Other Spectral Manipulations

Transforming reflectance into derivative form is one of the common pre-processing algorithms used in spectral manipulation in order to remove specific interferences. The interferences may be (1) light scattering caused by particle size distribution and alignment with the incident beam of light, (2) path length differences due to different thicknesses of sample, and (3) spectral baseline differences due to e.g. detector drift, changing environmental conditions such a temperature, spectrometer purge and sampling accessories (Duckworth 2004). First, second, and third derivatives, and higher-order derivatives, can be found.

The use of derivatives is an approach which addresses two of the basic problems related to NIR spectra: *large baseline variations* and *overlapping peaks* (Hruschka 2001). It is one of the best methods for removing the baseline effect (Duckworth 2004). The first derivative is simply a measure of the slope of spectral curve at every point. The first derivative implies normalisation of spectra because it reduces a constant baseline to zero (Hruschka 2001). Normalisation can help to suppress unwanted sources of variability resulted from, for example, radiation scattering.

The second derivative is the change in the slope of the original curve. It can be used to separate overlapping absorption bands because it is a measure of the curvature and has the same sign as the curvature of a spectrum (Hruschka 2001). Second derivatives have successfully resolve overlapping bands in Fe oxide mineral-related studies; thus determining goethite and hematite content in soils (Scheinost *et al.* 1998), and identifying the nature of Fe-oxides (Malengreau *et al.* 1996). Using the second derivative with PLSR has also been successful in detecting and quantifying swelling clays (Goetz *et al.* 2001). The second derivative can also compensate for the variation in light scattering for reflectance and for variation in transmission measurements, and it can also provide adequate baseline correction (Brimmer *et al.* 2001). Even though higher derivatives will resolve overlapping absorption bands better than lower derivatives, they are more sensitive to noise than the lower ones. Derivatives may (1) have side lobes, which can give false information about an absorption band, (2) increase sensitivity to spectral noise, and (3) increase wavelength instability (Hruschka 2001). To reduce the problem of noise enhancement, smoothing is usually first applied to the spectral data.

There are many ways to calculate derivatives; e.g. simple forward differences, finite-differences (gap-method), Fourier Transforms (FT), and the Savitzky-Golay method. Two of these methods (*finite differences* and the *Savitzky-Golay*) are commonly used. Both methods use information from a localized segment at a particular wavelength rather than the difference between adjacent points (Duckworth 2004). As already stated, smoothing spectral data before calculating derivatives (when using finite difference method) can be applied in order to reduce noise enhancement. The first derivative by the finite difference method is calculated as:

$$\rho'(b_i b_{i+1}) = \frac{\rho(b_{i+1}) - \rho(b_i)}{\lambda(b_{i+1}) - \lambda(b_i)}$$

where:

$\rho(b_i)$ = reflectance of band i

$\lambda(b_i)$ = wavelength of band i

$\rho'(b_i b_{i+1})$ = first derivative of linear curve segment between reflectance of band i and band i+1

The finite-difference method is easier to calculate, but is more sensitive to noise than the Savitzky-Golay method, because the latter includes a polynomial smoothing. The first derivative calculation by Savitzky-Golay method uses a convolution function (Duckworth 2004) thus the number of data points (filter size) in the function must be specific. The derivative is taken of the polynomial which fits the spectrum. If the segment is too small, the result may not be better than using a simple difference method. But if the segment is too large, the Savitsky-Golay method will be too smoothed out with the loss of important information. The number of points required to cover the full width at one-half the height of the largest absorbing band in the spectrum is a good general rule to use in determining the appropriate size of the spectral segment (Duckworth 2004). In addition, the Savitzky-Golay method is particularly useful when dealing with a very sharp absorption band, with high noise spectra, and in cases where the value of adjacent measured wavelengths is desired, and in detecting very small wavelength shifts (Hruschka 2001; Martens and Naes 2001). In conclusion Savitzky-Golay coefficients allow useful and rapid calculation of smoothed derivatives, and thus a common method for the determination of derivatives (Brereton 2003).

Derivative transformation may present difficulties for spectral interpretation. It makes it more difficult to visually interpret the residual spectrum, as a result it is difficult to locate impure spectral absorbances (Duckworth 2004). However the effective ability of derivatives to remove baseline shifts, normally outweighs these disadvantages. Thus derivative transformation of spectral data has been used successfully by many workers to predict e.g. organic C (Waes *et al.* 2005), organic C and inorganic C (Brown *et al.* 2005), C mineralization rate (Matuo *et al.* 2006), total N and organic C (Reeves III and McCarty 2001).

Other spectral manipulations are Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) transformation and De-Trending (DT). MSC can be used to remove the multiplicative scatter effect (change in NIR absorption/reflectance intensity caused by the change of light scattering of a sample) (Geladi *et al.* 1985). MSC can also be used to correct spectra with intermediate path lengths (Duckworth 2004). It is only applicable to the spectra that have responses fairly linear with concentration. It works well with the spectra of samples that are chemically similar; not on data which varies due to a wide range of variability in the sample composition (Duckworth 2004). Martens and Stark (1991) created the Extended Multiplicative Scatter Correction (EMSC) method which is designed to improve the separation of light scattering and light absorbance. They also created the Spectral Interference Subtraction (SIS) method for elimination of interferences with known spectral effects. Another spectral manipulation called SNV transformation and de-trending can be used to remove the major effects of light scattering from the spectra (Barnes *et al.* 1989). It is interesting to note that Barnes *et al*. (1989) found similar and satisfactory results using SNV, DT, and derivative transformation, with all three of them being better than non pre-processed spectra. Waes *et al*. (2005) obtained the best calibration equation without scatter correction, but using the first derivative of the spectra. Also Moron and Cozzolino (2003) obtained good predictions of silt, clay, sand, Cu, Fe, Mn, and Zn without applying scatter correction on the spectral data, but using the second derivative and SNVD (standard normal variate and detrend).

In summary, it seems that there is no single spectral manipulation that is best for all data. The nature of the data influences the choice or the appropriate kinds and number of spectral pre-processing steps.

2.3.1.5. Averaging Spectral Measurements

Spectroscopy-related studies facilitate the collection of rapid, multiple scans per sample, providing a more robust sample spectral analysis. In the laboratory this is usually done by rotating the sample, so reflectance of the sample can be viewed from many different angles. The goal of averaging is to reduce uncertainty in measurements and to reduce the effect of noise, as a result averaging produces smoother data set (Esbensen *et al.* 2006).

The recommended optimum number of scans per sample varies among publications and may affect prediction accuracy. Studies conducted by Sudduth and Hummel (1993) indicated that less than 10 scans per sample did not allow the sensor to meet its design objective of a standard error of prediction (SEP) of less than 0.29% organic carbon. So, 10 or more scans resulted in a better SEP. The best predictor of OM was achieved with 100 scans, but little improvement was seen by increasing the number of scans from 60 to 100. He *et al*. (2005) collected 60 spectra per soil sample from 3 angles; by rotating a Petri dish they collected 20 scans every 120^{o}. Chang and Laird (2002) averaged 25 scans per sample. Martin et al. (2002) scanned each soil sample only three times to provide three "views" of the sample. The average number of a number of readings can be calculated using the equation below:

$$mean = \frac{\sum \rho(\lambda)}{N}$$

where $\rho(\lambda)$ = reflectance at wavelength λ, and N = number of spectra per sample. Even though little improvement was achieved by increasing the number of scans from 60 to 100 (Sudduth and Hummel 1993), 100 scans may be better to use, because a higher number of scans may produce smoother spectra.

2.3.2. Data Processing

2.3.2.1. Partial Least Square Regression and Other Multivariate Analysis Methods

Multiple linear regression (MLR) was one of the earliest methods used to deal with multivariable data, but it was found that it could not cope with multicollinearity. To overcome this problem several statistical methods can be used; e.g. variable subset selection (VSS), ridge regression (RR), principal component regression (PCR), and partial least square regression (PLSR). Cheng and Wu (2006) compared these methods and reported that PLSR had performed best. PLSR is generally characterized by high computational and statistical efficiency and offers great flexibility and versatility in its handling of analysis problems (Boulesteix and Strimmer 2007).

Partial least squares regression was first developed by Herman Wold in the 1960's and 1970's to address problems in economic path modelling and was subsequently adapted by his son, Svante Wold, and many others in the 1980's to address problems in chemometric and spectrometric modelling, and now it has become the most popular regression method in chemometrics (Cheng and Wu 2006). This popularity is related to the ability of this regression technique to deal with (1) a greater number of independent variables (descriptors or explanatory variables) than dependent variables (data points or observations), and (2) the existence of correlations among explanatory variables (Boulesteix and Strimmer 2007).

Compared to other multivariate analysis, MLR is the simplest method to use. It is applicable in situations when there are many more samples than independent variables and when the independent variables have low noise and low inter-correlation. However, when the number of explanatory variables is large compared to the number of observations, multicollinearity (inter-correlation) often exists and causes difficulty estimating parameters. Multicollinearity can create grave uncertainties in the modelling and interpretation and corresponding decrease in prediction ability (Martens and Martens 1986). It can cause large variances of parameter estimates leading to less precise estimation. This problem must be resolved when using least squares methods (such as ridge regression, principal component regression and partial least squares regression), due to the assumption of these methods to the explanatory variables not

being correlated with each other (Adnan *et al.* 2006). MLR is not applicable when dealing with high dimensional data because this requires special techniques such as variable selection or dimensional reduction which are not possessed by MLR analysis. Also MLR requires selection of a reduced number of "the best X-variables" or "the best wavelengths" of the spectral data which must be less than the number of response variables. Miller and Miller (2005) stated that rather than select a few of independent variables, it is better to reduce their number by using PCR or PLSR. Moreover, both of these methods give satisfactory results when multicollinearity occurs in the independent variables (Miller and Miller 2005).

Martens and Martens (1986) compared the concepts behind MLR and PLSR, as summarised in the figure below. In MLR, all X-variables simultaneously model the Y-variable (solid arrows). While in PLSR1, Y is used to guide the extraction of latent variables (T) from X-variables (dotted arrows). Then, these latent variables (T) are in turn used for modelling both X and Y (solid arrows).

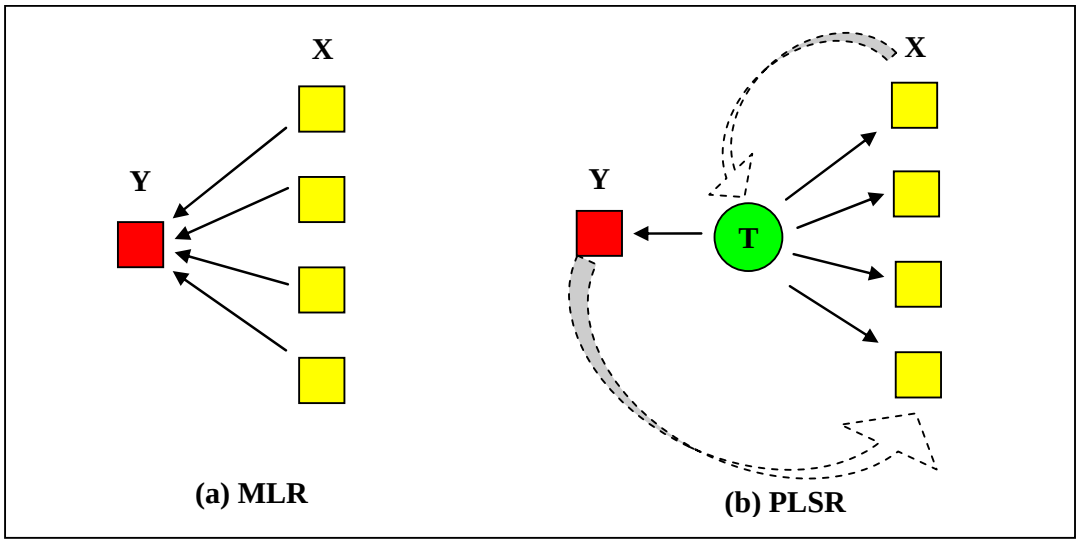


Figure 2.20 Conceptual difference between (a) MLR and (b) PLSR (Martens and Martens 1986).

Stepwise multiple linear regression (SMLR) can be used with sets of independent variables that are highly correlated. This method is a family of procedures that can work by selecting the "best X-variables". A search algorithm is used to find the subset of independent variables that provides the best description of the dependent variable. However, "this may decrease the interpretability and represents a waste of information, unless one explicitly wants to find such a representative subset of X-variables to

describe Y" (Martens and Martens 1986). If there are few samples, SMLR may cause over-fitting of the data, and the ability to predict other than the calibration samples may be reduced or lost. Hruschka (1987) suggested that a minimum of ten samples were required for each estimated regression constant and for each varied parameter (such as wavelength).

Principal component analysis (PCA) is a multivariate analysis used for data reduction. It works by reducing data dimensionality and creating new latent variables (or factors) which are uncorrelated. So it works well when there is correlation present between independent variables. In contrast, it is not a useful technique if the independent variables are uncorrelated (Miller and Miller 2005). PCA transforms the original dataset of variables into a smaller set of linear combinations of original variables called components that represent most of the variance of the original set (Dillon and Goldstein 1984; Miller and Miller 2005). Latent variables or principal components (Z1, Z2,…, Zn) are linear combinations of the original variables (X1, X2,…, Xn):

$$Z_1 = a_{11}X_1 + a_{12}X_2 + ... + a_{1n}X_n$$

$$Z_2 = a_{21}X_1 + a_{22}X_2 + ... + a_{2n}X_n$$

etc.

where: a = coefficient

The first principal component (PC1) accounts for most variation in the new data set and the second component (PC2) the second most variation, and so on. The new latent variables, unlike the original variables, are uncorrelated with each other (Miller and Miller 2005). A score plot of the first two principal components can be used to explain the differences/similarities of constituents in terms of chemical structure (e.g in soils). Sometimes it may be used for physical interpretation (Miller and Miller 2005). It may also be used for outlier detection (Cozzolino and Moron 2006). Points which are close to each other on the score plot reveal greater spectral similarities, and points which are far away from the rest of the data may be outliers (Esbensen *et al.* 2006). PCA was first applied to soil spectral reflectance analysis by Condit (1970).

Principal component regression (PCR) is an example of a latent variables regression method which can cope with multicollinearity (Miller and Miller 2005). This

method work well if there is a considerable degree of correlation between independent variables. This method consists of two multivariate analyses; PCA and MLR. The basic principal of PCR is to reduce the number of original independent variables into a smaller set of orthogonal components which represents most of the variability in the original data, contains a new set data which are not correlated one each other, and contains a reduced amount of random measurement noise (Adam 2004). PCR works first by reducing the dimensionality of the data by the use of principal component analysis (PCA), and then independently performing a MLR to relate the PCA factors to the independent variable (Miller and Miller 2005).

Partial least squares regression (PLSR), which also involves latent components, is similar to PCR but the components extracted from the data are a function of the values of both independent and dependent variables. It differs from PCR by including the dependent variable in the data compression and decomposition operation, i.e. both Y and X data are actively used in data analysis (Adam 2004). Unlike PCR, the latent components obtained by PLS are chosen with the response of variable of the regression taken in mind (Boulesteix and Strimmer 2007). In PLS, independent variables that show high correlation with the response variables are given extra weight because they will be more effective at prediction (Miller and Miller 2005). Moreover, dimension reduction and regression are performed simultaneously (Boulesteix and Strimmer 2007), so that small but relevant differences that might be ignored in a PCR are included in the solution. The optimum number of components to use in the analysis can be found using cross-validation**.** Since PLS components are selected to maximize both description of the independent variables and correlation with the dependent variables, the relevant predictive information tends to be concentrated in fewer components, making this method more efficient than PCR (Martens and Naes, 1987). Martens *et al*. (1983) stated that PLS should yield better results than other methods when the number of samples is low and/or when the calibration data are noisy. The table below gives a comparison between MLR and PCR and PLSR.

Table 2.3 Comparison of multiple-wavelength and full-spectrum modelling methods (Brimmer *et al.* 2001).

Specification	Characteristic of methods	
	Multiple-wavelength methods (e.g. MLR)	Full-spectrum methods (e.g. PLSR and PCR)
For characterizing	Simple chemical system	Complex chemical system
Interpretation of spectral information	Easier to interpret	More difficult to interpret
Dataset requirement for calibration	Smaller dataset	Large dataset
Transferring the methods	More straightforward	More difficult to transfer
Sensitivity to un-modelled matrix variation	Less sensitive	More sensitive

In terms of the number of response variables (dependent variables), there are two kinds of PLSR; PLSR1 and PLSR2. PLSR1 processes only one response variable, whereas PLSR2 processes more than one response variable in one operation. In PLSR1, each response variable is treated separately, while in PLSR2, response variables are treated collectively. PLSR2 is only used when the response variables are correlated with each other (Miller and Miller 2005).

Specifically, PLSR has many advantages over other computing approaches because: (1) it selects variables automatically, (2) it can perform a variety of tasks including classification, (3) it is very efficient statistically, and (4) it is computationally very fast and therefore suitable for application to large data sets (Boulesteix and Strimmer 2007).

The superiority of PLSR to other statistical methods has been reported by a number of other workers. Sudduth and Hummel (1991) stated that PLSR has the most promise for prediction of soil organic matter content. They found that PLSR resulted in better calibration than SMLR and PCR. Bolster et al. (1996) showed that PLSR was more accurate than SMLR in predicting the chemical composition of tree foliage. Bogrekci and Lee (2005) found that PLSR was better than both to MLR and stepwise MLR for predicting soil P concentration.

2.3.2.2. Calibration, Validation and Cross-Validation

2.3.2.2.1. Calibration

"Calibration concerns how to find the mathematical formula that optimally converts, for example, instrument data X to results Y" (Martens and Martens 1986). The purpose of the calibration model is to relate the measured sample property to the spectral data collected from that sample (Westerhaus *et al.* 2004). The samples used to build the calibration have to (1) comprise a wide range of analyte values and also spectral responses, (2) have a wide variety of sample types and large constituent range, and (3) represent the target population (Malley and Martin 2003; Westerhaus *et al.* 2004). The calibration set should represent not only the chemical characteristics of the constituent of interest anticipated in the future samples to be predicted by the calibration, but also the physical characteristics (Brimmer *et al.* 2001; Malley and Martin 2003). Samples in the calibration set are also called the *training set*.

Various statistical methods are used to develop calibration models. After pre-treatment procedures are applied to the spectral data, a range of statistical methods can be used to relate pre-processed spectral data and analyte values. Statistical methods vary from relatively straightforward methods, such MLR, to more complex methods such as SMLR, PCR, PLSR, and Artificial Neural Networks (ANN) (Westerhaus *et al.* 2004).

There are many ways to select the samples for the calibration set. Westerhaus *et al*. (2004) described seven methods for determining how samples are selected. These are (1) random selection, (2) manual selection, (3) spectral subtraction, (4) stratified selection, (5) discriminant analysis using wavelength selection, (6) discriminant analysis using PCA selection, and (7) correlation matching. In general, the purpose of the sample selection is (a) to reduce the number of samples needed to develop a calibration at a practical level and (b) to obtain a data set that represents the population to be analysed by NIR spectroscopy (Westerhaus *et al.* 2004). The samples should cover the range of variability anticipated in future samples (Hruschka 2001).

Islam *et al*. (2003) used a random technique to separate the calibration and validation set. As reported by Brown *et al*. (2005), random selection may overestimate predictive accuracy and seems not appropriate because different selections led to different results. McCarty *et al*. (2002) found quite good prediction (with little bias) of

total C, inorganic C, organic C and organic C (acid) using validation set separated from one third samples of calibration set. There were 273 samples collected samples from 14 geographically diverse locations in the central of USA. But when they used an independent validation set composed solely of samples from the Nebraska sampling site, they found higher bias. Experience from previous study by inclusion of just a few samples from new sampling location into the calibration set can readily correct bias for similar location (McCarty and Reeves III 2000). As stated by McCarty *et al*. (2002), "the robustness of calibration set can depend on the extent of the property domain for the calibration set". Williams (2001) described a selection procedure for separating calibration and validation sets by sorting samples using reference data and transferring the first three samples into the calibration set; leaving the next two samples at the validation set and then again taking the next three (6-8) sample into the calibration set; leaving the next two (9-10) in the validation set, and then taking again the next three (11-13) samples for the calibration set, and so on. If the number of samples is large, Williams (2001) suggested the ratio of calibration and validation set should be 3:1. Martin *et al*. (2002) determined the calibration and validation set by sorting the data (reference data) from the lowest to the highest value and dividing them into two sets by putting every other spectrum into set A (the calibration set), and the remaining spectra into set B (the validation set), so the ratio of calibration set and validation set is 1:1. Each set therefore represented the full range of constituent concentrations. This method was also used by Malley *et al*. (2002) and they obtained useful and successful prediction of various chemical properties of soil and hog manure. These selection methods for the calibration and validation sets require prior knowledge of reference values. Selection can also be based on spectral data. Valdes *et al*. (2006) used Mahalanobis (*H*) distance to select samples for the calibration and validation sets. Samples with a standardised *H* distance of a minimum 0.6 from their nearest neighbour were chosen for the calibration set, and the rest of as the validation set. So the calibration set was selected on the basis of sample dissimilarities. This method is based on the ideas of Shenk and Westerhaus (1991a; b). Chang and Laird (2001) selected a calibration set from all of those in the data base (including the sample being tested) based on the shortest squared Euclidean distance to the derivative reflectance spectra for the test sample. So, in this case the calibration set was selected on a basis of similarity between samples. Due to lack reports dealing with comparative accuracy of those selection methods, this needs further investigation.

There is no fixed minimum number sample in the calibration sets used by researchers. Shepherd and Walsh (2002) evaluated the size of the calibration set from its predictive performance and found the performance decreased only gradually as the size of the calibration set was reduced from 800 to 200. But it then decreased rapidly when the calibration set contained less than 100-200 samples. Malley and Martin (2003) suggest at least 50 samples are needed for a calibration set, with preferably a minimum of 100 to 150 samples. However, some workers have used less than 50 samples; e.g. 41 samples (Daniel *et al.* 2003), 25 samples (Russel 2003) and 15 samples (Shibusawa *et al.* 2001). Thus it seems difficult to make a general statement about the minimum.

When using PLSR to develop calibration models, the optimum number of components (latent variables) should be determined. If the number of components used is more than the optimum number, the error of prediction may increase and the calibration models may be overfitting (Christy and Kvalheim 2007). The optimum number of components is obtained when the model produces the lowest predicted residual error sum of square (PRESS). This can be found by running leave-one-out cross-validation (Miller and Miller 2005; MINITAB Inc. 2003).

To calibrate spectral data with more than one response variable (dependent variable or Y-variable), we may treat the response variables separately or collectively. To calibrate spectral data against one Y-variable value is called PLSR1 (Miller and Miller 2005). "PLSR1 regression relates a single **Y**-variable to a block of **X**-variables x_1, x_2, x_3, x_4,…, x_K onto a new factors t_1, t_2,…,t_A ($A<K$) and using these estimated factors as regressors for the variable **y**" (Martens and Martens 1986). While, to treat more than one Y-variable collectively is called PLS2. It is usually used when the response variables are correlated with each other (Miller and Miller 2005).

There can be many factors causing error in developing calibration model. In Table 2.4 below, Mark and Workman (2003) summarised the sources of major calibration error. Furthermore, they emphasized the three commonest error sources in any calibration: (1) reference laboratory error (stochastic error source), (2) different aliquot error (non-homogeneity of sample - stochastic error source), and (3) non-representative sampling in the training set or calibration set population.

Table 2.4 The sources of major calibration error (Mark and Workman 2003).

Source of error
• Sample non-homogeneity • Laboratory error • Physical variation in sample • Chemical variation in sample with time • Population sampling error • Non-Beer's Law relationship (non-linearity) • Spectroscopy does not equal wet chemistry • Instrument noise • Integrated circuit problem • Optical polarization • Sample presentation extremely variable • Calibration modelling incorrect • Poor calibration transfer • Outlier samples still in the calibration set • Transcription error

2.3.2.2.2. Validation

Comparison of NIR measurements and reference data in the new sample set is called *validation* (or verification) of the calibration. As stated by Martens and Martens (1986), "validation is a procedure to ensure that the model and its estimated parameters really have predictive ability for future, unknown samples of the same general kind". Validation provides a basis for calculation of the true measurement error. "It is an important step in the data analysis and guards against over-fitting" (Martens and Martens 1986).

2.3.2.2.3. Cross-Validation

"Cross-validation is a procedure that repeats the modelling several times, each time using only some of the training samples in the model estimation and the rest as test samples from which y is predicted from X" (Martens and Martens 1986). It involves eliminating the first sample, or block of samples, from the calibration model development, and predict the eliminated samples with the model developed; this procedure is repeated for all samples or block of samples (Williams 2001). Cross-

validation can be carried out by dividing the sample into equal "blocks" and eliminating samples one block at a time. However, the most effective way is to eliminate one sample at a time (called one-leave-out cross validation) (Williams 2001). Cross-validation is applied when the sample set is small. It is also applicable to large populations, but the computation may be time consuming. Cross-validation is used for model fitting and validation (Martens and Martens 1986).

2.3.2.3. Waveband Region Selection

New instruments with high resolution and broader bands (350-2500 nm e.g. ASD, FieldSpecPro, Boulder, CO, USA) have recently been developed, so people can now collect spectral reflectance of material from the ultra violet, visible, to near-infrared regions in one scanning. Various predictions of accuracy have been reported by researchers using full bands or part of the bands. Some workers demonstrated good prediction accuracy by using all bands (350-2500 nm) (Ludwig *et al.* 2002; Matuo *et al.* 2006), but some others reported better accuracy by using only part of the band regions (Chang *et al.* 2005; Chang *et al.* 2001; Reeves III and McCarty 2001). This indicates that waveband region selection is important in developing a good calibration model. Matuo *et al*. (2006) were successful in predicting C mineralization rate using bands from 350 to 2500 nm. Ludwig *et al*. (2002) successfully predicted the contents of C, N, and cumulative respiration and N mineralized after 264 days using Vis-NIR region (400-2500 nm). Chang *et al*. (2001) found that rather than using the full wavelength band (400-2498 nm) they preferred using bands between 1300-2498 nm and successfully predicted total C, total N, moisture, CEC, 1.5MPa water, basal respiration rate, sand, silt and Mechlich III extractable Ca. They did not use spectra from 400 to 1300 nm because it was found to reduce the accuracy of predicted soil properties. Similarly, Chang *et al*. (2005) collected soils spectra from 400 to 2498 nm with an interval of 2-nm to successfully predict total C, organic C, total N and moisture content. They did this by developing calibrations with PLSR using the first derivative of wavelengths between 1100 and 2498 nm. Reeves and McCarty (2001) determined soil constituents (total N, organic C, active N, biomass and mineralizable N and pH) using NIR reflectance spectroscopy and found out optimal calibrations using data from 1100-2300 nm with every 20 data points averaged from 1900 to 2300 nm. Comparing NIR and MIR techniques to measure soil properties, McCarty and Reeves (2006) used NIR

bands between 1100-2498 nm, rather than the full recorded bands of 400-2498 nm, to successfully predict organic carbon, total N and soil texture. They reported that no advantage was obtained by including bands 400-1098 nm. Locher *et al*. (2005) developed calibration models using a spectral range of 1333-2532 nm and successfully estimated legume content of multispecies legume-grass mixtures. Börjesson *et al*. (1999) found improvement of prediction accuracy by removing the water peak region around 1900 nm. Goetz *et al*. (2001) used spectral reflectance from 1800-2400 nm to distinguish smectites from illites and kaolinite and found that the spectral region of 1800-2000 nm and 2150-2250 nm contained the most relevant information for detection and quantification of smectite content.

It is interesting to note that the success of soil property prediction for particular soil constituents, for example carbon and nitrogen, can be obtained by using either complete bands of UV-Vis-NIR (350-2500 nm) (Ludwig *et al.* 2002) or part of the bands; 1300-2498 nm (Chang *et al.* 2001), 1100-2498 nm (Chang *et al.* 2005), and 1100-2300 nm (Reeves III and McCarty 2001). This may be explained because the signal of C and N may be found in the visible and/or NIR region. Colour, which is often used in the field to assess organic matter content (Krishnan *et al.* 1981), may be consistently associated with the change of soil C and N. As described by Stuart (2004) overtones and combination bands of organic matter compounds can be found scattered within the NIR region (section 2.2.3.2, Table 2.2).

2.3.2.4. Outlier Detection

Detecting outliers is necessary for both calibration development and for monitoring NIR methods (Martens and Naes 2001; Westerhaus *et al.* 2004). The two main types of outliers are: *spectral outlier* and *analytical outlier* (McClure 2001). For identifying spectral outliers, Mahalonobis distance (*H* statistics) and spectral residuals (Westerhaus *et al.* 2004) can be used. If a sample lies outside the D^2_{max} (the square of the maximum Mahalonobis distance) it is considered to be suspect (Westerhaus *et al.* 2004). For identifying an analytical outlier, it is detected by observing the difference between measured analytical values and predicted values. If the difference is large, it is considered as outlier, as discussed below.

Many criteria have been used to describe outliers, and these evaluation criteria are often subjective (Westerhaus *et al.* 2004). Mark and Workman (2003) defined an outlier as a sample spectrum with a distance of greater than three times the Mahalanobis distance from the centre of the spectral data. Outliers can also be defined as the points which are far away from the line of normal probability plot, or the points which have a standardized residual greater than ±2.0 (MINITAB Inc. 2003). GRAMS' software used by Mimmo *et al*. (2002) classify outliers as having predicted values greater than twice the standard deviation of the actual values. Cozzolino and Moron (2006) defined data points as outliers if the residuals from reference analysis were greater than 2.5 times the standard error of calibration (SEC). They also used PCA to detect spectral outliers; the points which are scattered far away from most of the data on the score plot of the first two principal components (PC). Williams and Norris (2001) define the outlier as "a result that differs from the rest of the population by more than three times the SEP (standard error of prediction)". Islam *et al*. (2003) also identify a value as an outlier if the difference between measured and predicted value is larger than three times the SEC or SEP. Drapper and Smith (1998) and Westerhaus *et al*. (2004) defined as outliers the samples whose absolute spectral residual value are greater than 3 or 4 standard deviations from the mean residual value. While there is no agreed upon definition, this definition, or one similar to it, seems typical.

There is also no agreed standard value in terms of the number of steps taken to eliminate outliers and what process to use. Elimination can be carried out in two steps; first before the calibration and validation sets are determined, and second during the validation process (Cozzolino and Moron 2003; 2006; Moron and Cozzolino 2003; 2004). The first step is done in order to supply data with no outliers into both calibration and validation sets. Removal of outliers in this step is usually based on spectral data. For example, when the samples have Mahalanobis distance (*H statistics*) more than 3, they are not included in both calibration and validation sets. The second step of outlier removal can be carried out in the process of validation, for example when the samples have prediction values more than 2.5 times the standard error of calibration. Two passes of outlier elimination similar to the above mentioned steps were also used by some other workers (Fystro 2002; Urselmans *et al.* 2006). Fystro (2002) firstly removed 5 outliers of 80 samples using *standardised H distance* ($H > 3$) which is based on the spectral data, and secondly removed again outliers based on large residual values (*t value* > 2.5).

In the second step, using a cross-validation procedure he removed the outlier (*t-outliers*) maximally twice in order to reduce samples with possible inaccurate reference values or spectral data (Fystro 2002). Shenk and Westerhaus (1991a), on the other hand, omitted samples with more than 2.5 residual values (*t-outliers*) and performed cross-validation again a maximum of three times. They did this in order to establish a calibration set.

As stated by William and Norris (2001) outliers are the samples that cannot be recognised by the instrument during calibration development, and as a result the instrument is forced to "guess" the composition and functionality of those samples. The number of outliers usually occur irregularly and at up to 2% of the total population (Williams and Norris 2001). Theoretically, PLS regression should afford better protection against outliers, because it is not heavily dependent upon reflectance at individual wavelengths. In practise, however, outliers still occur even with PLR calibrations (McClure 2001).

Mark and Workman (2003) described three major reasons for outliers. First, an outlier may be a spectrally perfect normal sample, but its reference laboratory value may have a gross error. Second, the outlier is not part of the target population. Third, an outlier is bad spectral data. In addition, McClure (2001) described factors affecting outliers as presented in the table below.

Table 2.5 Factors causing outliers (McClure 2001).

- Chemical composition
- Interaction among constituents
- Wavelength selection
- Moisture status
- Physical texture
- Particle size and shape
- Bulk density
- Sample preparation technique
- Sample orientation in cell
- Sample temperature
- Number of wavelengths use in calibration
- Mathematical treatment of log (1/R) data

William and Norris (2001) make four recommendations for dealing with outliers: (1) Ensure the outlier is an outlier by finding the cause of it to be different and whether or not it is likely to occur again. (2) If the same outlier is likely to occur in the future, collect most of them and add them to the calibration/validation set, then evaluate the degree to which the new calibration has accommodated the suspected outliers. (3) If outliers are unlikely to reappear and the sample set is sufficiently large enough to represent everything that will appear, determine the standard error of prediction (SEP) and other statistics with the outlier eliminated. There is no need to eliminate outliers if the statistics for the prediction or validation set are acceptable. (4) For long-term practical analysis, outliers should not be removed, because they may contain beneficial variance, unless the outliers do definitely not belong to the calibration set (e.g. a sample of ground grain which is not deliberately put into the calibration set of forages should be removed).

It is interesting to note that outlier problems relate to the type of NIR spectroscopy used. Reeves and McCarty (2001) compared spectroscopy using a spinning cup and a fibre optic probe in determining various constituents (total N, organic C, active N, biomass and mineralizable N and pH) in agricultural soils. They found more problems with outliers when using a fibre-optic probe than using a spinning cup. The reason for this was unknown (Reeves III and McCarty 2001). Similarly, better prediction accuracy of organic-C using rotating sample cup than fibre-optic probe was reported by Reeves III *et al*. (2002). They suggested this due to the smaller sample scanned by the probe and the problem created by the probe itself; spectral region beyond 2100 nm was very noisy.

2.3.2.5. Statistical Terms for the Evaluation of Accuracy and Precision

The ability of NIRS techniques to predict soil constituents has been evaluated using the following statistical parameters; coefficient determination (r^2), SEC, SEP (RMSECV or RMSEP), RPD and RER (Malley *et al.* 2004; Williams 2001). Bias (mean difference between the reference data and NIRS-predicted data) and slope (regression coefficient, *b*, between reference data and NIRS-predicted data) can also be used as the accuracy of evaluation (Williams 2001). Slope of 1.0 is excellent, indicating identical rate of change of reference and predicted data, whilst zero bias is difficult to achieve, as Williams (2001) stated that "*bias cannot be ignored*".

Coefficient determination (r^2 calibration) measures the proportion of the total variation accounted for by the model. Predicted coefficient determination (r^2 cross-validation and r^2 validation) is a measurement of the proportion of total variation accounted for by the model when using cross-validation and or validation analysis. Root mean square error of calibration (RMSEC) is the standard deviation of the difference between the measured and the estimated values for samples in the calibration set. RMSEC is also called standard error of calibration (SEC). RMSEC is calculated using the following equation (Brereton 2003):

$$RMSEC = \sqrt{\frac{\sum (y_m - y_p)^2}{N}}$$

where y_m are measured values from the laboratory measurements, y_p are predicted values derived from the spectral data using a PLS regression equation, and *N* is the number of samples.

The standard error of prediction (SEP) is the standard deviation of the difference between the measured and the estimated values for samples in the validation set. The SEP calculation incorporates a correction for bias (Williams 2001);

$$SEP = \sqrt{\frac{\sum (y_m - y_v)^2 - \left\{\left[\sum (y_m - y_v)\right]^2 / N\right\}}{N-1}}$$

where y_m are the measured values using the laboratory method, y_v are predicted values, and *N* is number of samples. It can be simplified using RMSECV or RMSEP without bias inclusion, and this has been suggested. RMSECV, which can also be referred to as the standard error of cross validation (SECV), is calculated using the following equation (Brereton 2003):

$$RMSECV = \sqrt{\frac{\sum (y_m - y_{cv})^2}{N}}$$

where y_m are the measured values using the laboratory method, y_{cv} are predicted values using cross-validation in PLS and *N* is the number of samples. Cross-validation is usually carried out when the number of samples is not large enough to be allocated in a separate validation set (Williams 2001). If the sample is enough for both calibration

and validation sets to be obtained, RMSE is usually calculated from the validation set, called the RMSEP. It is found as (Brereton 2003):

$$RMSEP = \sqrt{\frac{\sum (y_m - y_v)^2}{N}}$$

where again y_m are the measured values found using the laboratory method, y_v are predicted values obtained using validation set in PLS, and *N* is the number of samples. RMSECV and RMSEP have the same units as the original measurement units: ml if y is in ml; the unit is % if y is % (Esbensen *et al.* 2006).

RPD (ratio prediction to deviation) is the ratio of the standard error of prediction to the standard deviation of reference data in the validation set; it is calculated as SD/SEP (Williams 2001). However, some other researchers (e.g. Cozzolino *et al*. 2005; Mouazen *et al.* 2006a; Saeys *et al*. 2005) simplify the RPD by using the equation of SD/RMSECV or SD/RMSEP. RER (ratio error range), which is the range of reference values divided by SEP (Starr *et al*. 1981; Malley *et al.* 2004), can also be simplified as the ratio of the range of measured values of soil properties to the RMSECV or RMSEP.

$$RPD = \frac{STDEV(y_m)}{RMSECV} \quad \text{, or} \quad RPD = \frac{STDEV(y_m)}{RMSEP}$$

$$RER = \frac{Max(y_m) - Min(y_m)}{RMSECV} \quad \text{, or} \quad RER = \frac{Max(y_m) - Min(y_m)}{RMSEP}$$

Good prediction models will have (1) high values of r^2 calibration, r^2 cross-validation, r^2 validation, RPD and RER, and (2) low values of SEC, SECV, and SEP (Williams 2001). High values for the RPD indicate efficient NIR prediction (Malley *et al.* 2004; Williams 2001). An high coefficient determination only may be misleading, unless other predictive performance values are present (Williams 2001); it should be followed by a high RPD. An RER of $\geq$ 10 for example, indicates that the NIR reflectance instrument can predict the data at better than 10% of the mean (Starr *et al.* 1981). The SEC is not a good predictor of accuracy because it only describes how well the reference values are fitted by the calculated regression line; in contrast, SECV and SEP are the true estimate of prediction capability because they are calculated on

independent samples (Shenk and Westerhaus 1996). SEP and RMSE are "amalgamations of the combined errors of both reference and NIR reflectance test and the sampling, sample preparation, and variability factors inherent in testing an array of randomly distributed unknown samples" (Williams 2001). In MINITAB 14, predicted residual error sum of squares (PRESS) is used as an indicator of the predictive power of the model. This value is calculated from the sum of the squares of differences between the actual values (e.g. laboratory analysis) and predicted values. The closer PRESS value to zero, the better the predictive power of the model (Miller and Miller 2005).

There is no one standard value used to assess NIR calibration performances. In soil-related studies several workers have proposed categories for calibration success (Chang *et al.* 2001; Dunn *et al.* 2002) (Table 2.6). The categorisations proposed are based on the NIR measurements conducted in a controlled environment, in a laboratory, and using physically homogenous samples (e.g. dried ground soils). Malley *et al*. (2004) proposed guidelines for evaluating calibrations of environmental samples such as soil, sediments, animal manure, and compost. For on-site measurement of manure, Saeys *et al*. (2005) proposed the categorisation of accuracy based on RPD; < 1.5 not usable, between 1.5-2.0 possible to distinguish between low and high, 2.0-2.5 approximate quantitative prediction possible, 2.5-3.0 good prediction, and above 3.0 excellent.

Table 2.6 Proposed guidelines for NIR calibration of soil and environmental samples.

----------Soils ----------		Environmental samples (soil, sediment, animal manure, compost)
Dunn *et al*. (2002)	**Chang *et al*. (2001)**	**Malley *et al*. (2004)**
RPD < 1.6; poor	RPD < 1.4, r^2 < 0.5; C Not useful	RPD 1.75-2.25, RER 8-10, r^2 0.70-0.80; Moderately useful
RPD 1.6-2.0; acceptable	RPD 1.4-2.0, r^2 0.5-0.8; B Moderately useful	RPD 2.25-3, RER 10-15, r^2 0.80-0.90; Moderately successful
RPD > 2.0; excellent	RPD > 2.0, r^2 0.8-1.0; A Useful	RPD 3-4, RER 15-20, r^2 0.90-0.95; Successful
		RPD > 4, RER > 20, r^2 > 0.95; Excellent

For agricultural commodities, Williams (2001) categorised the success of NIR calibration based on RPD, RER, and r^2 (Table 2.7 and 2.8). These requirements are much higher than for soils and environmental samples, because rigorous guidelines are a requirement of food quality and other commercial agricultural products. For agricultural commodities, predictive values with $r^2 > 0.95$, RPD > 5 and RER > 20 are usually required for a successful calibration (Malley *et al.* 2002).

Table 2.7 Significance of values of RPD and RER statistics in commercial application (Williams 2001).

RPD value	RER Value	Classification	Application
0.0-2.3	Up to 6	Very poor	Not recommended
2.4-3.0	7-12	Poor	Very rough screening
3.1-4.9	13-20	Fair	Screening
5.0-6.4	21-30	Good	Quality control
6.5-8.0	31-40	Very good	Process control
8.1+	41+	Excellent	Any application

Table 2.8 Guidelines for interpretation of coefficient correlation and determination (Williams 2001).

Value of *r*	r^2	Interpretation
Up to ± 0.5	Up to 0.25	Not usable in NIR calibration.
±0.51-0.70	0.26-0.49	Poor correlation: reason should be researched.
±0.71-0.80	0.50-0.64	OK for rough screening. >50% of variance in Y accounted for by X.
±0.81-0.90	0.66-0.81	OK for screening and some other "approximate" calibrations.
±0.91-0.95	0.83-0.90	Usable with caution for most applications, including research.
±0.96-0.98	0.92-0.96	Usable for most applications, including quality assurance.
±0.99 +	0.98 +	Usable in any application

2.3.2.6. Sources of Error in NIRS Measurement

There are three broad error categories in NIR measurement; sampling error, reference method error, and NIR method error (Hruschka 2001). Importantly, Hruschka (2001) points out that errors are not necessarily due to imperfect human or machine

activities, but rather the result of differences between different measurement methods. For example, the NIR method measures molecular bonding when measuring protein, whereas the Kjeldahl method determines total nitrogen content, which is not perfectly correlated with molecular bonding. Thus, to some extent, the word "*error*" might be replaced by the term "*difference*".

Sampling error which is usually the largest *difference* (error) between reference method and NIR measurement can be caused by: (1) different sub-samples used in the reference and NIR method, (2) different amount of sample used by various reference methods when making the same analysis, and (3) repacking error and different portions of samples being scanned by NIR measurements (Hruschka 2001). Error in the reference method cannot be avoided because each reference method has measurement error. This is usually shown by the standard deviation of replicate samples. Therefore error in NIR methods can be caused by spectral measurement errors, lack of intrinsic correlation between reference and spectral data, and poor spectral processing (Hruschka 2001).

2.4. Success of Reflectance Spectroscopy for Soil Properties (C and N) Measurement

Many laboratory studies have been conducted using visible and NIR reflectance to predict soil properties. In the laboratory, samples are usually dried and sieved in order to obtain physically homogenous range and remove the effect of water, and also samples are ground in order to minimize particle size variation. In addition, the distance between sample and sensor is usually standardised. Using static NIR spectroscopy, soil moisture content was one of the earliest soil properties that was successfully measured (Bowers and Hanks 1965). Subsequently NIR-measured soil moisture has been successfully determined by many other workers (Dalal and Henry 1986; Kano *et al.* 1985; Slaughter *et al.* 2001). The use of static NIR spectroscopy was later extended for measurement of other soil chemical and physical properties such as total C, organic C, total N, NH_4-N, NO_3-N, CEC, P, Ca, K, Mg, pH, soil particle size and moisture (Ben-Dor and Banin 1995; Chang *et al.* 2001; Cozzolino and Moron 2006; Islam *et al.* 2003; Odlare *et al.* 2005; Reeves III and McCarty 2001; Sheperd and Walsh 2002; Stenberg *et al.* 1995). Despite different prediction accuracies, the results obtained varied from poor to excellent, Malley and Martin (2003) stated that NIR spectroscopy is mostly able to

predict total C, organic C, total N, CEC and moisture. These soil properties have NIR absorbers such as O-H, N-H, and C-H (Williams and Norris 2001). Mineral constituents (i.e. K, Ca, Mg) may be predictable if they are bound to, correlate with, organic compounds (Burns and Cziurczak 1992).

The success or otherwise of using soil reflectance technique for measuring organic matter has been partly discussed in the section 2.3.2.2. Some other aspects are discussed in this section. Early studies showed good correlations between SOM and both visible (Vis) and near infrared (NIR) reflectance (Krishnan *et al.* 1980; Stoner and Baumgardner 1981). Later laboratory tests by Sudduth and Hummel (1996) and Hummel *et al*. (2001) have shown that NIR soil reflectance technique can be used to estimate soil OM, soil moisture, and CEC in soils from a wide geographic area of the US Corn Belt for both surface and subsurface soils. Shonk *et al*. (1991) developed a shank-mounted sensor that measured soil reflectance at 660 nm and then correlated this measurement with organic matter for moist soil (r^2 = 0.71). They used this "organic matter sensor" to control soil-applied herbicide rates in real time based on the variability of OM. In spite of this success, a single wavelength sensor has limitations in more widespread use for soil-specific calibration, because the calibration model developed may be suitable for just one soil type. Reeves III *et al*. (2002) found great potential in determining organic carbon using both near- and mid-infrared spectroscopy. Similarly, Fystro (2002) reported potential ability to predict organic carbon (r^2=0.65) and total nitrogen (r^2 = 0.87) using Vis-NIR spectroscopy. However, Viscarra Rossel and McBratney (1998) found that the laboratory NIRS technique could predict clay and moisture content but not organic matter.

While many workers have reported that they can successfully predict both soil C and N using the NIR technique, whether it can predict C and N independently, or this success is based on the correlation between soil C and N, is open to question (Martin *et al.* 2002). Chang *et al*. (2001) were able to successfully predict total C (r^2 0.87, RPD 2.79) and total N (r^2 0.85, RPD 2.52), but total C and total N were highly correlated (*r* 0.95). A high correlation between C and N in the reference data has also been shown by other workers; *r* between organic C and total N 0.94 (Malley *et al.* 2000); *r* between C and N ranged from 0.966 to 0.989 (Reeves III and McCarty 2000). In contrast, low correlation of C and N reference data (*r* 0.667) was found by Martin *et al*. (2002). They used this low correlation to explain why they were able to predict C organic (r^2 0.785

and 0.783, RPD 2.16), but not for N organic (r^2 0.326 and 0.366, RPD 1.22 and 1.26). These findings support the hypothesis that successful prediction of both C and N is based on the correlation between these soil constituents. In contrast to this, Chang and Laird (2002) proved that C and N were able to be predicted independently using the NIR technique. They found good prediction of organic C, inorganic C, Total C, Total N, total C : total N ratio, and organic C : total N ratio with high r^2 and RPD values in the cross-validation and validation sets, even though they found low correlations between those constituents with r^2 ranging from 0.02-0.61. In contrast the r^2 and RPD in cross-validation were 0.86-0.97 and 2.5-5.4 respectively, and in the validation set were 0.85-0.96 and 2.8-5.5. More accurate NIR prediction than autocorrelation of these soil properties supports the existence of independent signal bands for C and N to the incident light, even though specific absorption peaks for N containing functional groups are difficult to assign (Chang and Laird 2002). Martin *et al.* (2002) summarised that if the correlation of reference data of C and N is high, N may be predicted based on its correlation with C, but if the correlation is low, N is predicted based on N absorbers.

2.5. Emerging Technologies for *in situ* Measurement of Soil Carbon

Recently efforts have been focused on measuring soil carbon *in situ*, as laboratory based methods (e.g. dry/wet combustion) are labour intensive, time consuming and costly, which limits their applicability for measuring the spatial and temporal variability of C over large land areas. NIRS is not the only method developed for efficient field measurement. The other methods include Laser Induced Breakdown Spectroscopy (LIBS), Inelastic Neutron Scattering (INS) and remote sensing (Gehl and Rice 2007).

The LIBS method is based on atomic emission spectroscopy (Gehl and Rice 2007). An intense laser pulse is focused on a sample, forming a microplasma that emits light characteristic of the compositional elements of the sample, which is then recorded and spectrally determined by a time-gated sensor to detect the concentration of elements according to their unique spectral characteristics (Gehl and Rice 2007). This technique has been used to measure soil carbon (Cremers *et al.* 2001; Ebinger *et al.* 2003), soil nitrogen (Harris *et al.* 2004) and other soil properties including metal ions (Mosier-Boss *et al.* 2002; Yamamoto *et al.* 1996). The accuracy of this technique to measure soil carbon when calibrated against the dry combustion technique has been reported by

Cremers *et al.* (2001) who found an adjusted r^2 calibration of 0.96. Ebinger *et al.* (2003) also reported high correlation (r^2 calibration 0.99 and r^2 validation 0.95) between C measured by dry combustion and by LIBS. For field measurement, this technique is still under development. The drawback of using the signal at 248 nm to detect C concentration is interference with Fe; this can be overcome by using 193 nm for C analysis (Ebinger *et al.* 2003). However some current issues still need to be explored further; such as the development of more general calibration curves for various soil types, and evaluation of the effect on LIBS accuracy of fine roots, different decompositional stages of organic matter, carbonates and water content (Gehl and Rice 2007).

The INS method is based on inelastic scattering of 14 MeV neutrons from C nuclei in soil and measurement of the resulting 4.44 MeV gamma ray emission (Gehl and Rice 2007; Wielopolski *et al.* 2000). A Deuterium-Tritium (d, t) neutron generator placed above the ground produces the 14 MeV neutrons and the subsequent 4.44 MeV gamma ray emission is detected by NaI detectors, then the spectrum is analysed for the C peak intensity that is proportional to the soil C concentration (Gehl and Rice 2007; Wielopolski *et al.* 2008). Wielopolski *et al.* (2003) reported measurement error of 5-12% and minimum detection limit (MDL) of 0.018 g C cm^{-3} when they tested the method in static mode at three different field sites (pine stand, oak forest and sandy patch) to verify the calibration curve developed using C concentration in the sandpit. Wielopolski *et al.* (2008) stated that the INS method can be used in continuous-scanning modes (mobile), although the results of this use have not been reported. They claimed that INS is an *in situ* non-destructive technique, as it can measure soil C to approximately 30 cm depth without digging or coring for soil samples, except for developing or verifying calibration curves. Some issues that they (Wielopolski *et al.* 2008) mentioned will need to be further explored are e.g. the universality of the calibration line for various soil types, confirmation of the soil volume measured by the instrument, and evaluation of the minimum detectable change of soil C.

Remote sensing techniques have been used for a long time to indirectly monitor soil C through assessment of land cover, land management practices, cropping and tillage practices, net plant productivity and plant residue, as these features affect soil C dynamics (Gehl and Rice 2007). Direct measurement of soil C using remote sensing was conducted by relating soil colour (visible reflectance) with soil organic carbon

(SOC) calculated within a landscape (Fernandez *et al.* 1988). As reported by Henderson *et al.* (1992), a high correlation between colour and SOC can be obtained on soil formed from similar parent materials. However problems may occur because soil colour and reflectance properties of surface soil are a function of many factors in addition to organic matter, such as soil moisture, iron content, texture, parent material, chemical soil compositions and surface condition (Baumgardner *et al.* 1985; Ben-Dor 2002; Stoner and Baumgardner 1981). Using medium-infrared (MIR) bands of 1955-2495 nm rather than visible bands, Henderson *et al.* (1992) found them strongly correlated with SOC for soils formed from different parent materials. So far, there has been limited research on the direct quantification of SOC or SOM. Hierarchical foreground and background analysis and AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) were used by Palacios-Orueta *et al.* (1999) to predict SOM and Fe contents in California. Using bare-soil aerial photograph (colour slides), Chen *et al.* (2000) mapped surface SOC for a 115 ha field in Georgia. Using Landsat imagery and ground truth site information, Lacelle *et al.* (2001) determined C stocks and variations in existing estimations for North America. However, direct measurement of C on surface soils using remote sensing can be restricted by e.g. surface residues, land cover/vegetation and cloudy sky (a frequent occurrence in New Zealand).

2.6. Use of the NIRS Techniques for *in situ* Measurement and its Limitations

The present challenge for using the Vis-NIR reflectance spectroscopy technique is its use for *in situ* field measurements. There are only a limited number of publications related to the *in situ* measurement of soil properties using NIRS (Mouazen *et al.* 2005; Shibusawa *et al.* 2005; Shonk *et al.* 1991; Sudduth and Hummel 1993) and these publications outline several difficulties associated with field assessments. For example, the soil surface presented may contain gravel, stones and plant debris which will likely reduce the prediction accuracy of the Vis-NIRS techniques (Mouazen *et al.* 2005; Mouazen *et al.* 2007). In addition, much of the delicate and expensive instrumentation is not suitable for work in the field.

Numerous attempts to develop on-the-go soil sensors have been reported and reviewed (Hummel *et al.* 1996; Sudduth *et al.* 1997). Further development of sensor techniques is expected to increase the effectiveness of precision agriculture (Adamchuk

et al. 2004). Shonk *et al*. (1991) found that soil moisture content and surface preparation significantly influenced output of the sensor. Shonk *et al.* (1991) found that they could predict organic matter for one soil landscape but were unable to extrapolate to a larger geographic area. This may be because the calibration model developed was only robust for only one soil type.

When NIRS is combined with GPS and GIS, the technique can be used to assess spatial and temporal variability of soil properties in the landscape on the basis of the whole integrated soil-plant-landscape system, providing information for site specific management (Malley *et al.* 2004). Recent precision agriculture developments have focused on development of sensors (electrical and electromagnetic, optical and radiometric, mechanical, acoustic, pneumatic, and electrochemical sensor) for on-line measurement of soil properties (Adamchuk *et al.* 2004). A combination of reflectance spectroscopy with GPS and electromagnetic induction (EM) surveys makes it possible to assess individual soil properties such as C, N and water content, and to assess soil variability on the basis of soil texture (Hedley *et al.* 2007). This can then be applied to e.g. monitoring carbon sequestration and nitrogen dynamics and for the purpose of irrigation management in the landscape (Hedley *et al.* 2007). Therefore, the development of field reflectance spectroscopy together with GPS has the potential to map the high spatial variability of soil properties (such as soil organic matter, nitrogen, texture, CEC) on a large scale. The map can then be used as a land management tool for e.g. soil fertility assessment, fertilizer placement, effective irrigation, and herbicide placement. In addition, NIRS may provide a rapid and affordable way of monitoring the spatial and temporal variability of carbon in landscapes caused by land use change, so the best land-use practices can be adopted. Soil carbon sequestration would help to offset greenhouse gas emissions (Lal *et al.* 2007).

High soil water content may mask other soil properties and lead to difficulty in the detection of other soil chromophores (Malley *et al.* 2004). Therefore in routine laboratory analysis soil samples are dried before being measured using NIRS. However the possibility of *in situ* NIR measurement using field-moist, “as-is” soils, was investigated by Chang *et al*. (2005). They successfully predicted total C, organic C, total N and moisture content from field-moist soils (ranged from 28 to 254 g kg^{-1}). They concluded that the NIR-PLSR technique can predict soil properties with field moist conditions with acceptable accuracy, as long as diverse soil samples from the same

region are included in the calibration, even though they found slightly better result when they dried the soil samples. Using field moist samples taken into laboratory for NIRS measurement, van Vuuren *et al*. (2007) were able to predict soil organic matter, total N, N mineralization, pH, P, Ca, Mg, K, and Na with acceptable accuracy. Fystro (2002), on the other hand, found thawed moist soils were more accurately predicted than dried samples, and no benefit was gained from the grinding of sieved (4 mm) and dried samples. The reason for this was not reported, also the water content range of the thawed moist samples was not stated. Mouazen *et al*. (2007) were able also to predict total C and organic C from field moist soils, but with moderate accuracy. It seems that moist condition may not always become a major impediment for field use of NIRS.

One of the problems faced when using field spectroscopy is a low signal to noise ratio due to atmospheric interference e.g. from water vapour; at 1350-1440 nm, 1790-1990 nm and 2360-2500 nm (Thenkabail *et al.* 2004), particularly when the distance of object (sample) is great, e.g. with remote sensing by aeroplane or satellite. This interference can be removed by decreasing the distance between sample (e.g. soil) and spectroscopy sensor, i.e. using a ground-based sensor.

However, close-up scanning then creates another problem related to the precise soil-to-sensor distance. Shibusawa *et al*. (2003) found inaccuracy measurements due to variable distance between the sample and sensor head. Sudduth and Hummel (1993) found inaccurate estimations of soil properties (OM, CEC and water content) associated with variation in the soil-to-sensor distance during scanning when developing an on-line measurement technique. To reduce this variability, they normalized the reflectance signal to a mean value for each sample curve. This normalization procedure corrected for the height variability while maintaining significant inter-soil variability. Mouazen *et al*. (2005) also found less accuracy in the measurement of water content using on-line NIR measurement in the field than the laboratory due to variations in the distance between the soil and optical sensor. However, by using their modified method (field spectroscopy with the sensor attached to a medium-deep sub-soiler fixed to a frame mounted on the three point hitch of a tractor), they claimed that they could minimize the effect of the soil-to-sensor optical unit distance variation. More recently, using their mobile NIRS technique, Mouazen *et al*. (2007) found good measurement of moisture content (RPD=3). Even though cross-validation measurements for C-organic (r^2=0.74, RPD=1.97), C-total (r^2=0.73, RPD=1.92), pH (r^2=0.71, RPD=2.14), P-available

(r^2=0.69, RPD=1.80), and P-extractable (r^2=0.73, RPD=1.94) are acceptable, they stated that it is "only possible to provide quantitative approximations" for these parameters.

Movement of the sensor relative to the sample (travel speed) can affect the NIRS prediction accuracy. Sudduth and Hummel (1993) ran the tractor with an average speed of approximately 0.65 m/s (2340 m hour^{-1}). Using this speed they could not get good organic matter estimation, compared to the static measurement in laboratory. Mouazen *et al*. (2007) used a travel speed of 1500 m hour^{-1}. In their previous study Mouazen *et al*. (2005) used travel speed at 1200 m hour^{-1}, but still found less accuracy when on-line rather than stationary measurement was carried out. Sudduth and Hummel (1993) concluded that movement of sample relative to the sensor during scanning seemed to be a major contributor to the lack of predictive capability.

Udelhoven *et al*. (2003) found unsatisfactory predictions in *in situ* assessment of all soil parameters considered. One of the reasons given for this was the effect of soil aggregation. Trapping most of the incident light into interaggregate spaces of coarse aggregates having irregular shape may extinguish the light (Baumgardner *et al.* 1985). Other reasons were (1) the distance between the sensor and soil surface (about 10 cm), resulting in small surface variations and micro shadows which could not be eliminated by simple spectra pre-treatment such as min-max normalization, and (2) sample differences in water and nutrient distributions (Udelhoven *et al.* 2003).

2.7 Advantages and Disadvantages of NIRS Techniques

When used with soils, NIRS techniques have both advantages and disadvantages (Malley and Martin 2003). The advantages are as follows:

- This technique is quick and can be continuous.
- The accuracy is equivalent to any reference methods.
- The precision can be superior to reference methods.
- There is little or no need for sample preparation.
- It has a low labour cost, or low cost per test.
- There is no need for chemicals (reagents), except for reference analysis.
- The technique can simultaneously determine numerous constituents or functional parameters.

- It allows living material to be analysed due to its low light energy.
- The instruments are small.
- Field-portable instruments are available.
- The instruments are easy to maintain.
- The instruments can be controlled remotely.
- The output can be combined with other data streams for mapping and decision-making.

The disadvantages of soil analysis using NIRS are:

- The technique requires separate calibration for each constituent and functional parameter.
- It may need calibration from different locations, soil series or textural classes.
- The accuracy and precision need to be checked continuously using reference analysis.
- The availability of training in NIRS implementation is poor.
- The cost of the instrument is relatively expensive, although that cost is quickly offset by savings in labour, time, chemicals and maintenance.

2.8. Focus of Study

From the review above, it can be summarised that NIRS techniques can be used to quantify soil constituents that are "near-infrared active" constituents (covalent bonds of small atoms, particularly, O, H, C and N), and other soil constituents which are featureless in the near-infrared but well correlated to one or more of the "near-infrared active" soil components. Field root density measurements using NIRS have not been reported, although roots are "near-infrared active". The only reported laboratory root density measurements using NIRS were carried out on peat samples to quantify *Ericales* roots for palaeoecological studies (McTiernan *et al.* 1998). Field measurements have not been made of root density in pastoral and arable soils using NIRS. Root density measurement using conventional techniques (e.g. wet sieving prior to counting or weighing) is tedious, and requires separating the roots from the soil.

Related to the identified gaps in the use of NIRS, the first study in this thesis aims to develop a method for the non-destructive measurement of root density in cores of

intact soil using Vis-NIRS. Root density measurements of ryegrass grown in a pot trial are reported, assessing the ability of Vis-NIRS in predicting root density. For this study, development of a purpose-built soil probe suitable for close distance spectral acquisition was needed.

Assessment of root density prediction using Vis-NIRS from a pot trial is followed by root density prediction in the field, in permanent pasture. Because many soil chromophores e.g. water, iron oxides, decomposed organic matter and soil aggregates may influence soil reflectance of field soils, the ability of Vis-NIRS to predict root content from spectral reflectance needs to be examined. The utility of calibration models developed from glasshouse data to predict field root density is also a focus of this study. In addition, as soil organic C is partly formed from decomposed plant root, whether Vis-NIRS predicts roots independently from soil organic matter needs to be evaluated, in order to identify potential C sequestration from root production. This study required a purpose-built soil probe that allowed reflectance spectra to be acquired from a soil corer.

Deep rooting plants are useful in e.g. extracting subsoil water and nitrate leached past shallower roots, and in sequestrating carbon deeper down in the soil which will prolong its residence time. As part of studies of such deep rooting plants, field assessment of root density using Vis-NIRS was extended to arable land and maize root density assessment down to 600 mm depth. This also involved the use of a soil probe attached to a soil corer.

There is no published literature on the use of NIRS for *in situ* field measurements of C and N in New Zealand soils. Hill Laboratories in New Zealand has used NIRS techniques, but only using dried ground soil samples in the laboratory, although the results have not yet been published. Moreover, there are limited numbers of publications assessing the ability of NIRS techniques in real-field soil conditions where there are variations in other soil chromophores e.g. water, clay and non-clay mineral types, iron oxides and parent materials. To address these concerns, the next objective of this thesis was to assess *in situ* “Vis-NIRS-soil core” technique in quantifying soil C and N from various New Zealand soils e.g. Allophanic, Pumice and Tephric and Recent soils.

Another area that needs further attention is the basis for selection of the spectral data samples (out of a larger population of spectra data) in which the reference property will be analysed for use in the chemometric calibration procedure. There are limited numbers of publications using the spectral data as a basis for this selection of the calibration and validation sets. This is an important piece of work because if accurate predictions of soil properties are to be achieved from spectral reflectance, the chemometric calibration may have to be restricted to unknown sample with similar spectral attributes. So the influence of sample selection techniques for the calibration and validation sets on the NIRS prediction accuracy is another focus of this study.

Even though excellent prediction of soil C and N using NIRS can be achieved (Chang and Laird 2002; Moron and Cozzolino 2004), this was carried out using dried ground sieved soil samples. Excellent accuracy for soil C and N measurement in the field has not been reported, but may be attempted by developing appropriate field techniques. This study is focused on developing and testing such spectral acquisition techniques. Spectral reflectance acquired from flat horizontal cross-sections of soil cores are compared with spectra collected from the outer surface of the curved, vertical wall of cylindrical soil cores. Being able to predict soil C concentrations to depth will be important because the IPCC (Intergovernmental Panel on Climate Change) protocol for soil C accounting recommends C measurement to 300 mm depth. So, accuracy of the two techniques in predicting soil C and N will be assessed from soil samples collected from 15 mm to 315 mm depth.

It has been explained in the Sections 2.2.3.1 and 2.2.3.6 that water and iron oxides considerably affect soil spectral reflectance (Baumgardner *et al.* 1985; Ben-Dor 2002). Seasonal fluctuation of soil moisture deficit together with fluctuation of water table, reduction/oxidation reactions and the iron oxides formed also influence the spectral reflectance which may in turn influence the accuracy of field Vis-NIRS measurement. The objective of the last study reported in this thesis was to assess the temporal robustness of calibration models used for predicting soil C and N. The robustness of the model was evaluated using data collected under different field moisture conditions (in dry and slightly wet months). The best spectral acquisition technique from the earlier studies was used to collect spectral reflectance from 15 to 315 mm soil depth.

CHAPTER 3

The Use of Vis-NIR Spectral Reflectance for Determining Root Density: Evaluation of Ryegrass Roots in a Glasshouse Trial

---

The review of literature identified that in response to global warming researchers are developing an increasing interest in the potential for greater sequestration of soil C via deeper plant root systems. Studies of root system turnover and C sequestration will require improved and more rapid techniques for root density measurement. This chapter develops and tests a soil probe for acquiring Vis-NIR reflectance spectra from soils to be used for root density prediction. This chapter also covers the spectral data processing that is necessary for the acquired spectra to be calibrated against reference measurements of root density.

A paper from this study has been published: B.H. Kusumo, M.J. Hedley, C.B. Hedley, A Hueni, M.P. Tuohy and G.C. Arnold. 2009. The use of Vis-NIR spectral reflectance for determining root density: evaluation of ryegrass roots in a glasshouse trial. *European Journal of Soil Science,* 60: 22-32.

3.1. Introduction

The study of plant root structure, function and turnover is expected to receive renewed interest from researchers investigating solutions to key global environmental problems. For example, to reduce concentrations of atmospheric carbon dioxide (CO_2), deeper rooting crops and trees with slower root turnover times are required to increase the rate of CO_2 sequestration (Guo *et al.* 2005; Lal *et al.* 2007; Rees *et al.* 2005). In long-term studies of soil organic matter (SOM) in soils cultivated for cereals, Solberg *et al*. (1998) found that increases in SOM with increased N fertilizer could only be accounted for by increased root mass and not increased straw residue alone. The use of deeper rooting crop and pasture species has also been proposed as a method to recapture nitrate at depth and reduce nitrate leaching (Bowman *et al.* 1998; Crush *et al.* 2007; Dunbabin *et al.* 2003) in nitrate sensitive catchments. In addition, plants with root systems that have improved phosphorus (P) foraging abilities in surface soils are required to drive the second ’green revolution’ in low fertility soils of developing

countries (Lynch 2007). After plant breeders have selected cultivars with the desired root traits (e.g. deeper rooting), proof of concept field experiments will be required to determine how successfully the traits are expressed under field soil conditions. At both the cultivar selection and field evaluation stages, it would be advantageous to researchers if a technique for rapidly assessing root density was available.

Rapid techniques for measuring root density without separating root and soil are not common. Two existing methods of measuring root density change with soil depth, such as the profile wall (Böhm 1979) and soil corer methods (Escamilla *et al.* 1991), need separation of roots and soil, which is a tiresome procedure, followed by root length measurement by the line intersect method (Newman 1966). An automated method, direct discrete image analysis, can be used to measure root length and diameter and has been shown to be more accurate than traditional line intersect method (Lebowitz 1988). The combination of soil coring and digital image analysis, used by Kokko *et al.* (1993) for total root surface area measurement, has been claimed to be a precise, repeatable and rapid method. However, these are still tedious procedures because roots must be separated from soil before the root image can be captured and analysed. A new technique using a combination of digital photography following disintegration of roots and soil with an air-knife (high velocity stream of air) is claimed to speed up root separation and measurement (Bingham *et al.* 2005). Flatbed scanners have been proposed for speeding up the counting of separated roots (Pan and Bolton 1991). Similarly digital image acquisition methods have been proposed for analysing root systems in soil profile walls (Ortiz-Ribbing and Eastburn 2003), but both techniques still require soil profile pits to be dug.

Near infrared reflectance spectroscopy (NIRS), which is a rapid and non-destructive technique, has been used to predict a variety of soil properties (Ben-Dor and Banin 1995; Chang and Laird 2002) including total carbon (C), organic C, total nitrogen (N), cation exchange capacity (CEC) and moisture (Malley and Martin 2003). NIRS has not been specifically evaluated for root density measurement; however, McTiernan *et al*. (1998) used NIRS to determine the content of *Ericales* roots and other physical characteristics and macrofossil components of peat.

Soil organic matter content has been successfully predicted by NIRS through absorption by, particularly, O-H, N-H, and C-H bonds in the NIR wavelength range.

This interaction allows NIRS to be used to measure the organic composition and functional properties of a wide range of agricultural products such as forage and feedstuffs (Roberts *et al.* 2004), oilseeds and grain crops (Delwiche 2004; Dyer 2004), fruits and vegetables (Slaughter and Abbott 2004) and animal by-products (Cozzolino and Murray 2004). Given the sensitivity of NIRS to organic compounds, our hypothesis is that a portable NIR field spectrometer could be used to report on plant root density.

The present paper reports on the development of a soil probe to acquire reflectance spectra from soil surfaces and the development of techniques to predict root density from the acquired spectra.

3.2. Materials and Methods

3.2.1. Glasshouse Study to Prepare Soils with Different Root Densities

Topsoils (0-10 cm depth) were collected from two soil types under permanent pasture: (1) Ramiha silt loam (Andic Dystrochrept, USDA NRCS classification); (Typic Allophanic Brown soil, New Zealand classification (Hewitt 1998)) derived from mixtures of greywacke of loess and andesitic ash, and (2) Manawatu fine sandy loam (Dystric Fluventic Eutrochrept, USDA NRCS classification) (Weathered Fluvial Recent soil, New Zealand classification (Hewitt 1998)) a recent soil derived from greywacke alluvium (Cowie 1978). Soils were air dried and sieved to less than 5 mm aggregate size, and coarse pasture roots were removed and each soil was remixed thoroughly to give homogenous samples.

The soils had markedly different soil chemical and physical characteristics (Table 3.1). The Ramiha soil, containing allophane, had a higher organic matter content and soil P retention than the Manawatu soil but had a lower Olsen extractable P status and bulk density.

Sixty 450 g subsamples of each air-dried soil were weighed into plastic bags. One of 20 combinations of N and P fertilizer treatments (N_0P_0, N_0P_1 --- N_3P_4) were applied and mixed thoroughly with the soil in each bag. Each treatment was replicated 3 times for each soil. N treatments were 0 (N_0), 111 (N_1), 222 (N_2) and 444 (N_3) mg N kg^{-1} soil and P treatments were 0 (P_0), 111 (P_1), 222 (P_2), 333 (P_3), and 444 (P_4) mg P kg^{-1} soil. Urea and $Ca(H_2PO_4)_2.H_2O$ were used as the N and P sources, and 222 mg K_2SO_4

(containing 100 mg K and 41 mg S) was added as a basal treatment. Then, 40 seeds of Moata ryegrass (*Lolium multiflorum* Lam., commonly used as a winter growing pasture in New Zealand) were mixed with the surface soil of each pot. After germination, the 15 strongest plants were grown for 72 days. Pots were watered regularly to a weight that represented 80% of water capacity of the soil in the pot. At 72 days selected pots were taken for root density measurements.

Table 3.1 Chemical and physical characteristics of the two soil types.

Characteristics[1]	**Ramiha**	**Manawatu**
C (% of dry weight)	6.3	3.0
N (% of dry weight)	0.57	0.29
P (% of dry weight)	0.09	0.88
P retention (% of total P)	62	14
Ca (me 100 g^{-1})	1.66	6.27
Na (me 100 g^{-1})	0.22	0.48
K (me 100 g^{-1})	0.21	1.05
Mg (me 100 g^{-1})	0.50	1.70
CEC (me 100 g^{-1})	26.58	28.49
Base saturation (%)	10	33
Olsen extractable P (mg P l^{-1})	12	30
pH	5.26	4.86
Bulk density (g cm^{-3})	0.76	0.92
Texture	Silt loam	Fine sandy loam

[[1]Analytical methods in Blakemore *et al.* (1987)]

3.2.2. Contact Probe Modification and Spectra Acquisition

A prototype soil reflectance probe was developed which was based on a plant contact probe (ASD FieldSpecPro, Boulder, CO.); an internal light source was replaced by a greater intensity parabolic-reflector halogen lamp (4.5 watt). A round casing was developed to avoid direct contact of the quartz probe window with the soil, to exclude the ambient light, and to provide a fixed distance (30.5 mm) between the object (soil surface) and the probe window (Figure 3.1c). The object was rotated through 360^{o}, to give a field of view of 561 mm^2 (Figure 3.1c).

Soil reflectance measurements were recorded from a flat-sectioned, horizontal soil slice (1.3 cm depth). The instrument records spectra from 350-2500 nm with a sampling interval of 1.4 nm for the region 350-1000 nm and 2 nm for the region 1000-2500 nm. The data processing software associated with the FieldSpecPro spectroradiometer interpolates the 1.4- and 2-nm-spaced data to produce 1-nm spaced data.

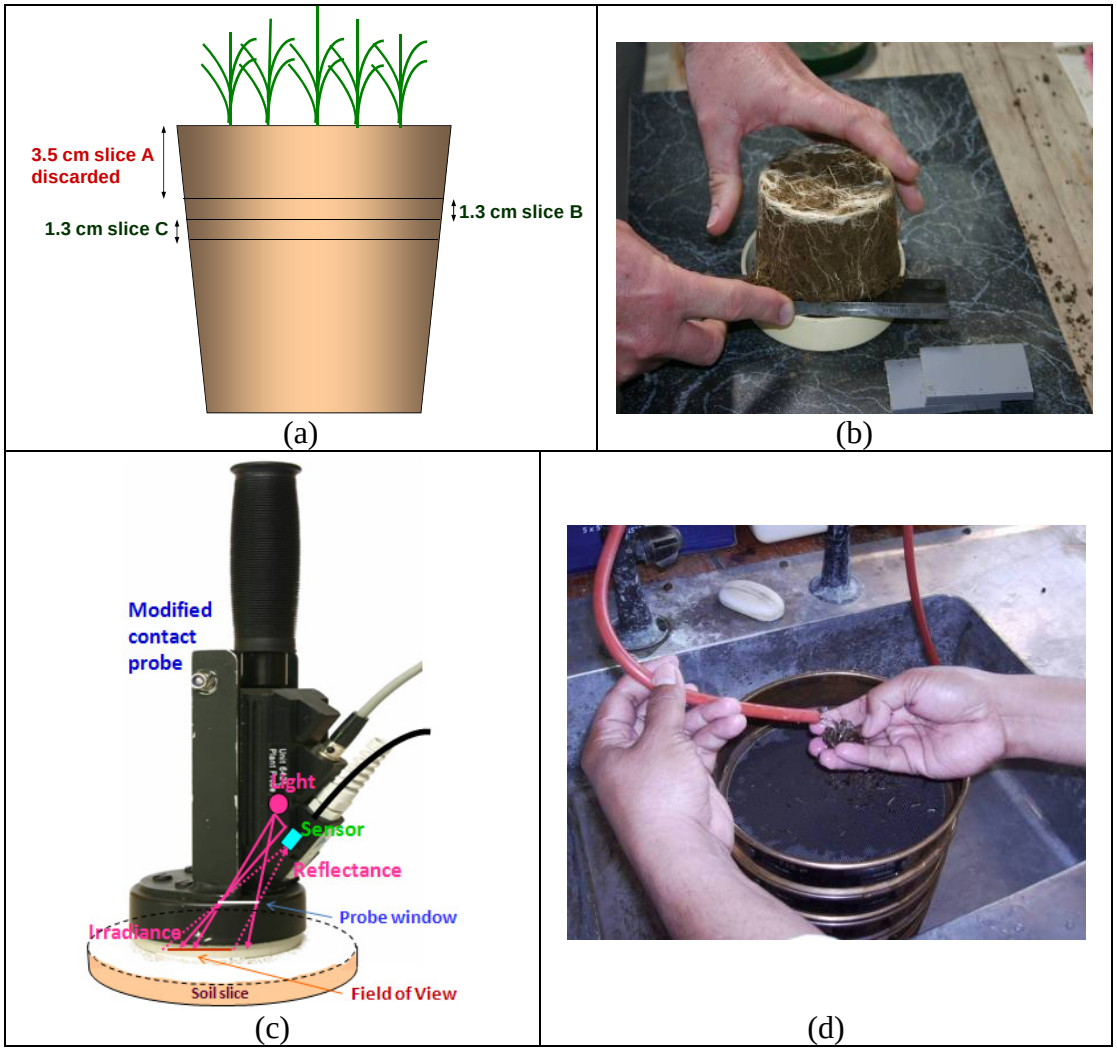


Figure 3.1 (a) Soil removed from pot, (b) sliced with a spacer ring and microtome blade, (c) spectral reflectance and (d) roots separated from soil with tap water.

3.2.3. Root Density and Spectral Measurement

Out of a total of 120 pots, 57 were selected on the basis of large shoot yield differences (0.81 to 2.91 g DM pot^{-1}) resulting from the range of N and P treatment combinations. Thirty samples were selected from the Ramiha soil and 27 samples from the Manawatu soil. It was expected that greater shoot growth would be associated with greater root growth.

The top 3.5 cm section (slice A) of soil from each pot was sliced using a microtome blade and discarded (Figure 3.1a and 3.1b). Then the reflectance was recorded at the freshly sliced surface (slice B) using the modified contact probe attached to the portable spectroradiometer (ASD FieldSpec Pro, Boulder, CO.). From immediately below the freshly cut surface, a 1.3 cm thick soil slice (slice B) was harvested and weighed. Roots were separated from the soil slice B by washing the soil slice with tap water through a sieve stack starting at 1000 μm followed by 500 μm, 400 μm, 350 μm and 300 μm diameters (Figure 3.1d). The root mass retained on the sieves was bulked and dried in an oven at 50^{o}C for 3 days.

A third 1.3 cm soil slice (slice C) was taken, weighed wet and then weighed after drying to constant weight in an oven at 105^{o}C. The average wet bulk densities for the Ramiha and Manawatu potted soils were calculated from the C slices. This wet bulk density was used to convert the wet weight of slice B into a volume (cm^{3}). The root mass in slice B was expressed as a density, mg dry root cm^{-3} soil.

3.2.4. Spectral Measurement of Standard Root Contents

Air-dry soil (sieved < 2mm, Ramiha) and oven dry root (50^{o}C until constant weight) were prepared and mixed to give percentages of root content of 0, 1, 5, 10, 25 and 100%. Air-dried sieved soil was used to minimize the effect of variable water content on the spectral reflectance and to provide a standard basal root content. Surfaces of the air-dry soil/root mixtures were prepared by pressing samples (1-cm thick) into a plastic petri dish (10 cm diameter). The spectral reflectance of the flat surface was recorded using the modified contact probe attached to the FieldSpecPro spectroradiometer.

3.2.5. Continuum Removal

Continuum removal was applied to the smoothed spectral data of the standard root contents. This is a procedure that facilitates making spectral curves distinction or easier to compare (Clark and Roush 1984). The depth of an absorption band, *D*, is usually defined relative to the continuum, *Rc*:

$$D = 1 - \frac{Rb}{Rc}$$

where *Rb* is the reflectance at the bottom (trough center point) of a band and *Rc* is the reflectance of the continuum at the same wavelength as *Rb* (Clark and Roush 1984). The depth of an absorption is related to the amount of the absorber. This approach has a powerful tool for enhancement and separation of small but often significant differences of bands of particular functional groups.

3.2.6. Spectral Pre-Processing and Data Analysis

Ten replicate reflectance spectra were acquired from each soil slice using the purpose-built contact probe (Figure 3.1c). Before statistical analysis, reflectance spectra were pre-processed using SpectraProc V 1.1 software (Hueni and Tuohy 2006). After a number of iterations the following pre-processing steps were adopted as standard: elimination of noisy data at wavebands 350-470 and 2440-2500 nm, followed by spectral smoothing using a Savitzky-Golay 4^{th} polynomial order filter (Shepherd and Walsh 2002) with a window size of 33 nm. After smoothing, the data were reduced by taking every 10^{th} waveband, the first derivative calculated and finally the ten replicate first derivatives were averaged. The first derivative data were imported to Minitab 14 (MINITAB Inc. 2003) for principal component analysis (PCA) and partial least squares regression (PLSR) analysis against the reference analytical data (root density). A PCA score plot was used to observe the pattern of sample scattering. During PLSR processing, samples which had a standardized residual of > 2.0 were removed as outliers (MINITAB Inc. 2003). The accuracy of the models was tested internally using a leave-one-out cross-validation method.

3.2.7. Regression Model Accuracy

The ability of the PLSR model to predict root density was assessed using the following statistics. RMSE (root mean square error), which is the standard deviation of the difference between the measured and the predicted root density was calculated: that from calibration data is called root mean square error of calibration (RMSEC) and from cross-validation is called root mean square error of cross-validation (RMSECV);

$$RMSECV = \sqrt{\frac{\sum (y_m - y_{cv})^2}{N}}$$

where y_m is the measured laboratory value, y_{cv} is the value predicted from the PLSR model, and N is the number of samples. Coefficient of determination (r^2) is the proportion of variability in a data set (measured and predicted value) that is accounted for by a statistical model. RPD (ratio of prediction to deviation) is the ratio of the standard deviation of the measured soil properties to the RMSECV. RER (ratio error range) is the ratio of the range of measured values of root density to the RMSECV.

$$RPD = \frac{STDEV(y_m)}{RMSECV}$$

$$RER = \frac{Max(y_m) - Min(y_m)}{RMSECV}$$

The best prediction model is shown by the largest RPD, RER, r^2 and the smallest RMSECV.

3.3. Results and Discussion

3.3.1. Grass Root Density Response to Fertilizer Treatments

A large variation in root densities was clearly evident when the pots of each soil type were destructively sampled (Table 3.2). Pots of Manawatu soil contained greater ryegrass root densities than pots of Ramiha soil. The different fertilizer N and P treatments had caused marked differences in root biomass and root density (Figure 3.2). In general, root density responded to both fertilizer N and P application to the Ramiha soil, which had low initial Olsen P status and a large % P retention. In the Manawatu soil, which had high initial Olsen P status and small % P retention, root density responded mostly to N application.

Table 3.2 Root densities and water contents of selected soil samples.

Soil properties	Samples from both soil types				Ramiha		Manawatu	
	Min	Max	St.Dev	Mean	Range	Mean	Range	Mean
Root density (mg cm^{-3})	0.46	5.02	1.27	2.21	0.46-3.84	1.25	1.76-5.02	3.27
Water content (%)	38.3	69.5	8.2	52.5	52.0-69.5	59.4	38.3-52.4	44.9

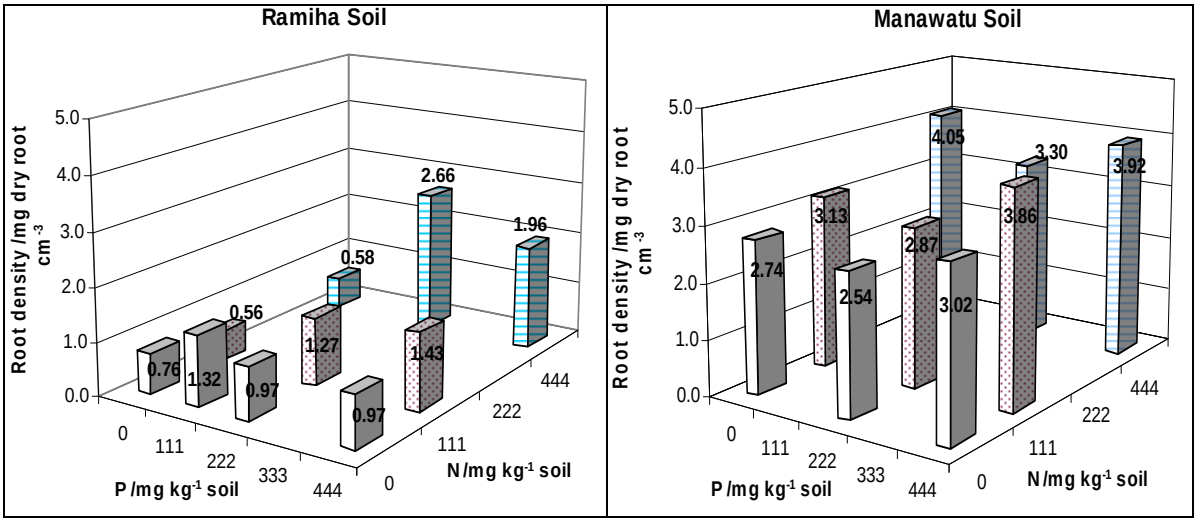


Figure 3.2 An illustration of the range of ryegrass root densities grown on Ramiha (left) and Manawatu soils (right) fertilized with different combinations of N and P (mg kg^{-1} soil). Results from selected pots only.

3.3.2. Spectral Reflectance Properties of the Two Soils (Ramiha and Manawatu)

The averaged spectral reflectance acquired from pots of Ramiha and Manawatu soils with a wide range of root densities are shown in Figure 3.3. The difference in the two spectra reflects differences not only in root content but also in content of other chromophores (iron oxides, water, clay and non-clay minerals, decomposed soil organic matter, parent material, etc.). The water content of the soil slices ranged from 38.3-69.5% for both Ramiha and Manawatu soil (Table 3.2). Strong light absorption occurs at bands close to 1455 nm and 1930 nm that are caused by water and O-H absorption. Bands close to 1400 nm are the first overtones of O-H stretching bond, and bands close to 1900 nm are the combination of H-O-H bend with O-H stretching (Clark 1999). The Ramiha soil reflects more light in the 655- 750 nm range, which is consistent with it being an older soil from a higher rainfall regime (more weathered) with a greater abundance of iron oxide and allophanic type clays than the Manawatu Fluvial soil. The Ramiha soil shows a slightly concave shape at bands near 900 nm (Ramiha spectrum, Figure 3.3), probably attributed to crystalline iron content (Demattê *et al.* 2004a; Demattê and Garcia 1999). Large initial organic matter contents in this soil (6.3% C) and additional root mass during the pot trial cannot completely mask the effect of iron content at the band close to 900 nm. The sharper angle of the Ramiha spectrum at ~ 780

nm than that of the Manawatu spectrum is consistent with moderately weathered soil. No concave shape (close to 900 nm) is found in Manawatu alluvial soil, probably because of the smaller iron content of this soil.

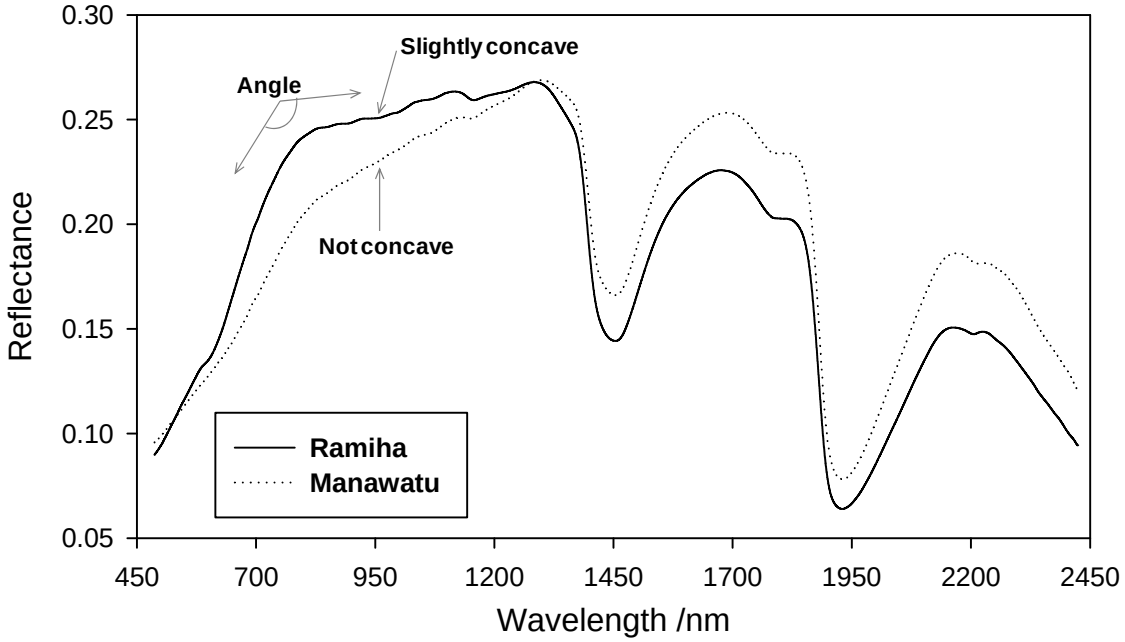
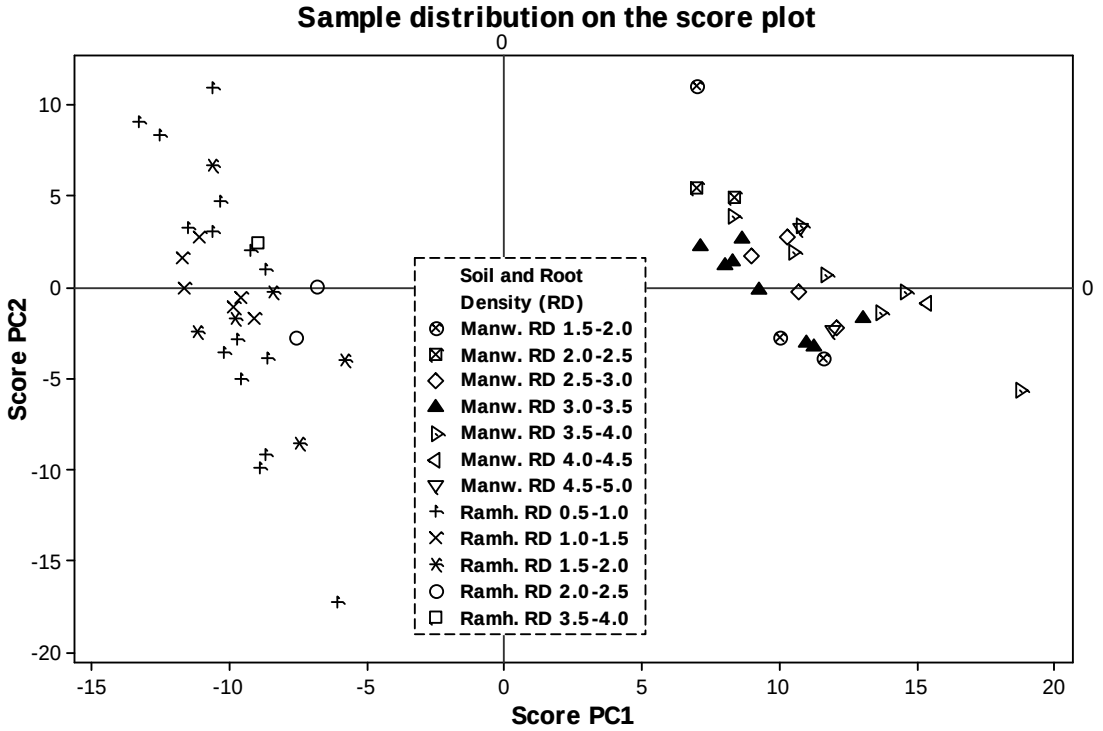


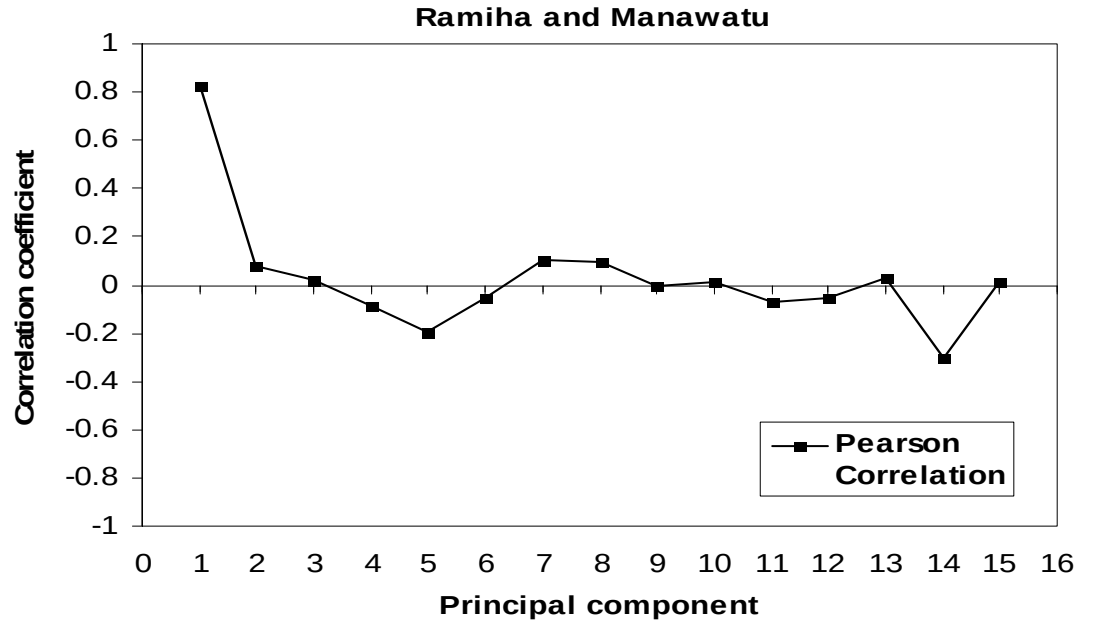
Figure 3.3 The average spectral reflectance for all Ramiha (average root density 1.25 mg cm^{-3} and water content 59.4%) and Manawatu soil samples (average root density 3.27 mg cm^{-3} and water content 44.9%).

The differences in spectral reflectance between soils (including difference in root densities) can be illustrated by using principal component analysis (PCA) of the overall variance in the spectra acquired from both soils (Figure 3.4a). The first PC accounts for 58.4% of the variation of the full data set and the second PC explains a further 12.9%. Figure 4a therefore displays a total of 71.3% of the variance of the first two PCs of the spectral data. The symbols in Figure 3.4a represent soil type and different narrow ranges of root density. Using a score plot of the first two principal components of the PCA, Ramiha and Manawatu samples can be easily differentiated. The major chromophores that may contribute to soil type cluster differences are parent material, non-crystalline mineral (e.g. allophane), SOM (quality and quantity), water and iron. Some samples from the same soil type with similar root density are closely distributed, but some others (especially in Ramiha soil) are separately dispersed although the root contents are similar. This indicates that the spectral variance of samples in the score plot is governed by chromophores other than root (e.g. water content). The range of water content is 38.3-69.5% with standard deviation 8.2%. PC1, which accounts for more than half of

the spectral variance (58.4%) was also well correlated to the root density (r = 0.83) (Figure 3.4b). This reflects the difference in root density between the two soils (see later).



(a)

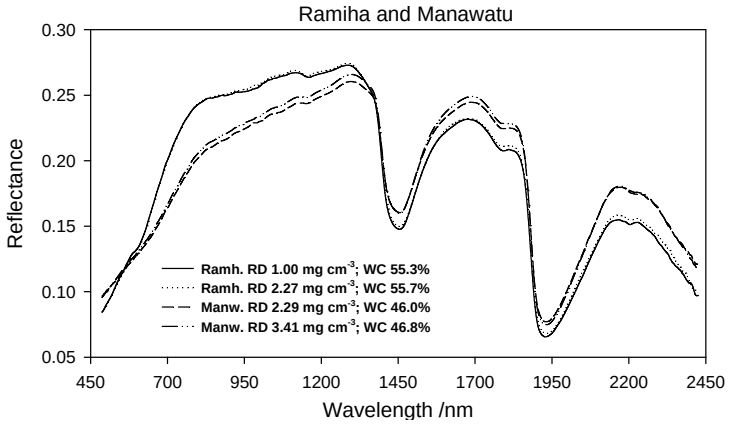


(b)

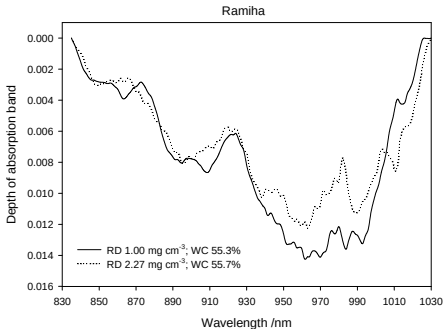
Figure 3.4 (a) The score plot of first and second component from the PCA of the first derivative of the reflectance spectra acquired from Ramiha and Manawatu soils and (b) relationship between root density and each of the principal components.

3.3.3. Spectral Reflectance of Samples with Different Root Density

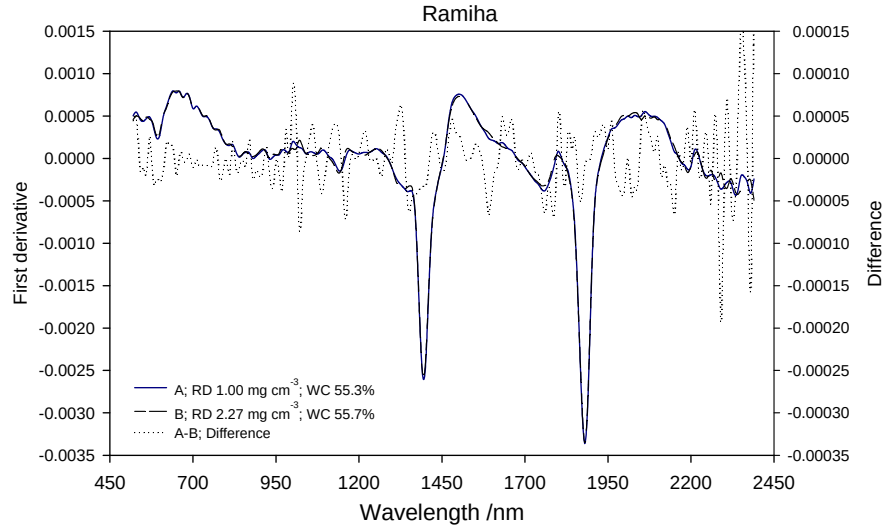
Reflectance spectra (1-nm spaced, smoothed) acquired from Ramiha soil with low (1.00 mg cm^{-3}) and high (2.27 mg cm^{-3}) root densities but with similar water content (55.3 and 55.7%, respectively) are shown in Figure 3.5a. Visually small differences in the spectral reflectance can be noticed (Figure 3.5a) as root density changes. However, continuum removal (Clark and Roush 1984) between wavelengths of 835-1030 nm and magnification of scale show how increased root density increases reflectance especially at bands near 950-990 nm (Figure 3.5b). Stronger absorption with less root-mass is perhaps caused by iron content (Demattê *et al.* 2004a; Demattê and Garcia 1999). Differences in the first derivative of reflectance (10-nm spaced spectral data, Figure 3.5c and 3.5d) show differences between low and high root density in two soils more clearly. Differences between the first derivatives are spread across all wavelength bands (Figure 3.5c and 3.5d). As expected, this variation in the first derivative of spectral reflectance is represented by the bands of greatest importance (largest coefficient) selected for the PLSR regression model built to predict root density in the potted soils (Figure 3.8).



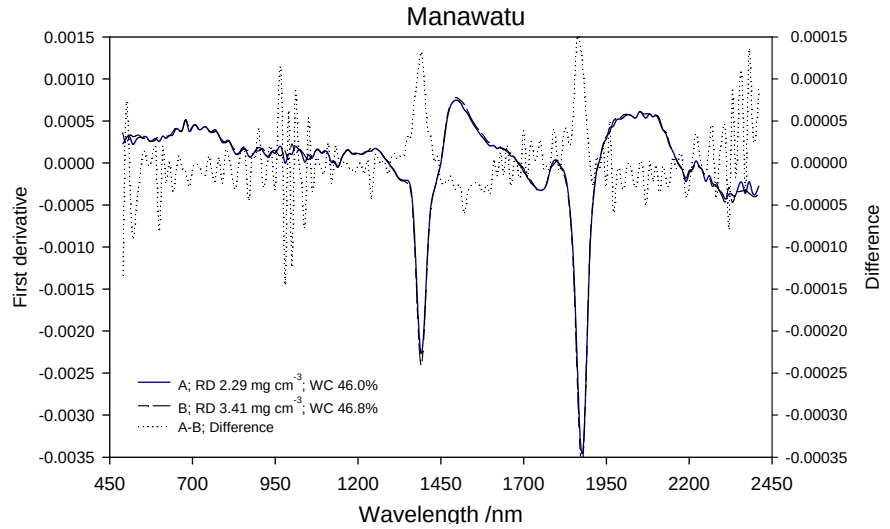
(a)



(b)



(c)



(d)

Figure 3.5 (a) The spectral reflectance from examples of low and high root density in Ramiha (1.00 and 2.27 mg root cm^{-3} soil; water content 55.3 and 55.7%; respectively) and Manawatu soils (2.29 and 3.41 mg root cm^{-3} soil; water content 46.0 and 46.8%; respectively), (b) continuum removal of band between 835-1030 nm in Ramiha samples, and the first derivative of those spectra from (c) Ramiha and (d) Manawatu samples.

The effect of root density variation on the reflectance spectra is more apparent when root density standards were prepared in the laboratory (Figure 3.6) by mixing air-dry soil (Ramiha) with different amounts of ground dry grass root. The slight concave shape near 900 nm is still clear in the spectra reflected from soil with low root contents (0, 1 and 5 % by weight root), but is absent when the root content is 10 % or larger. In addition, as the root content increases, the angle close to 780 nm becomes more obtuse. The spectra of soil with more roots tend to show increased reflectance at the near infrared band 780-2500 nm which may be related to the increased number of root-borne air voids. Air-root cell interfaces cause increased reflection (Baumgardner *et al.* 1985). The effect of additional root density is reflected in increases in slope between 830-940 nm (Figure 3.6a) and the relationship between slope and root mass is non-linear (Figure 3.6b). More obvious effects of root mass variation can be seen from continuum removal (Clark & Roush, 1984) of bands between 2035 and 2130 nm (Figure 3.6c); the larger the root mass, the deeper the absorption-band depth in this region. The strongest absorption band this region seems to occur at bands close to 2080 nm which was reported by Curran (1989) as the O-H stretch/O-H deformation from sugar and starch. The reverse trend occurs because of additional root content at bands 2175-2215 nm (Figure 3.6d); the larger the root content, the shallower the absorption-band depth. The band close to 2200 nm is well known as the clay lattice OH absorption band (Ben-Dor *et al.* 1999a), therefore it could be expected that as root mass increases the exposure of OH in clay lattice decreases.

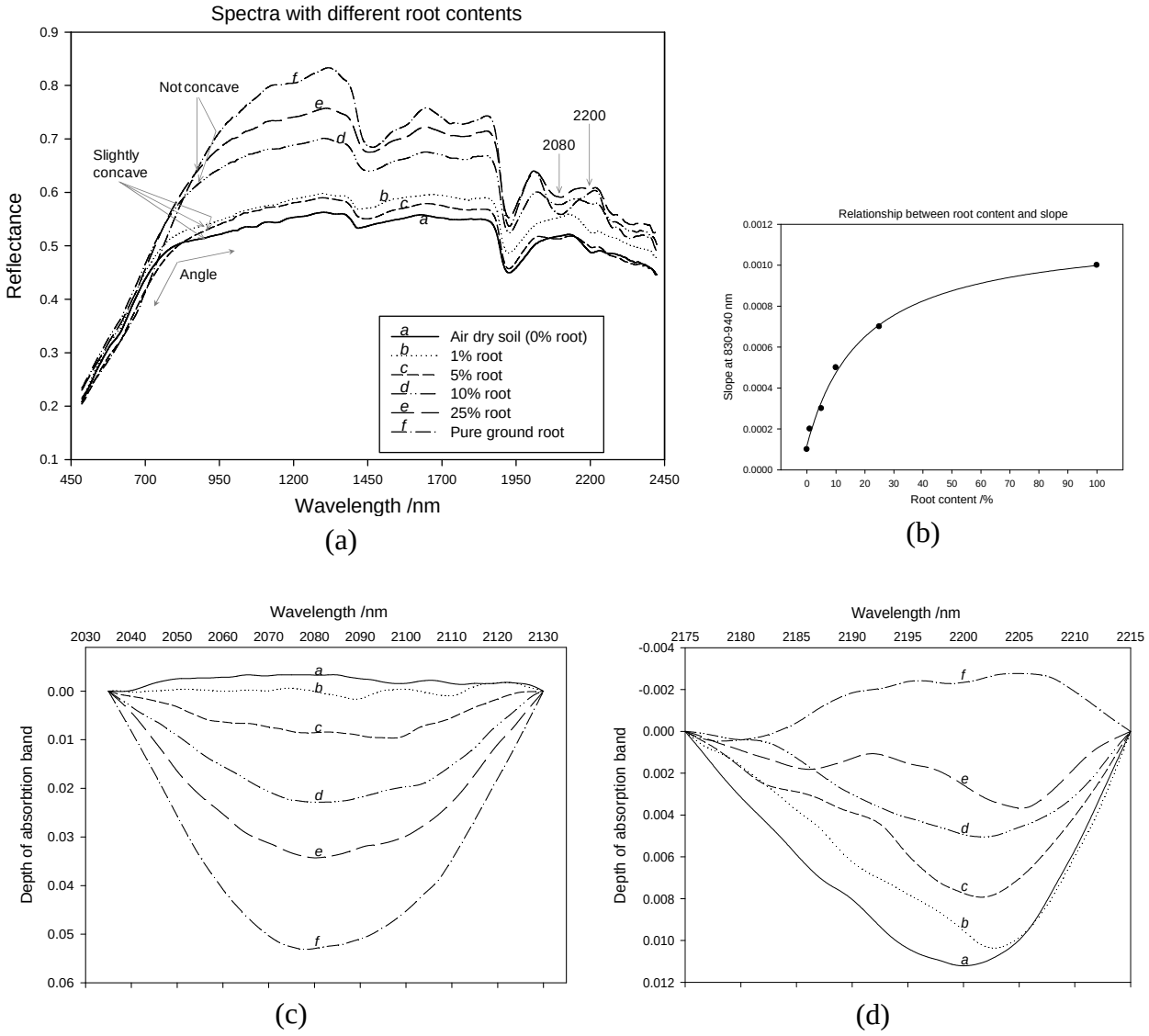


Figure 3.6 (a) Spectral reflectance of air dry soil (Ramiha) mixed with different dry root contents, (b) relationship between slope (bands 830-940 nm) and root contents, and continuum removal of bands between (c) 2035-2130 nm and (d) 2175-2215 nm.

3.3.4. Can NIRS Technique Predict Root Density?

The separation of Ramiha and Manawatu soils in the PCA score plot (Figure 3.4a) shows that the NIRS technique responds well to differences between soils, of which root density is part. To better measure the response of NIRS to root density independently of other chromophores, PLSR regression models to predict root density were fitted to each soil separately. The validity of each model was tested internally by using leave-one-out cross-validation. Leave-one-out cross-validation involves simply leaving out one sample and predicting it by using the remaining samples. This procedure is repeated for

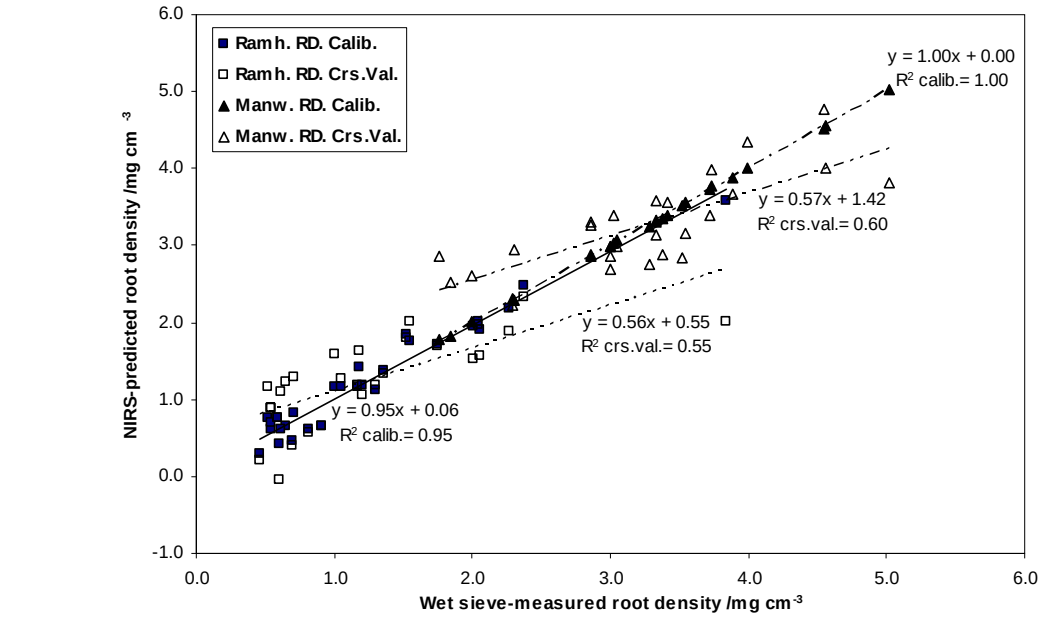
all samples. This process is usually carried out when small numbers of samples (e.g. < 60) do not permit separate calibration and validation data sets (Williams 2001). Separating the soil clusters creates two regression models (Figure 3.7a) with much poorer prediction accuracy (i.e. smaller r^2 and RPDs) (Table 3.3), compared with when both soils are included in constructing the PLSR regression model. The improved prediction accuracy when both soils are included results partly from the increased range of root densities in the observations, partly from the additional differences between soil types unrelated to root density and partly from the greater sample size.

Table 3.3 Calibration and prediction values of root density.

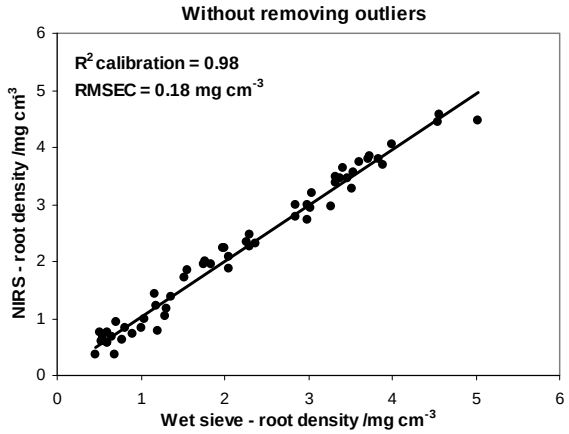
Source of samples	n	Comp	--Calibration--		Prediction values of root density ----------Cross-validation----------					
			r^2	RMSEC	r^2	RMSECV	RPD	RER	Bias	Slope
Ramiha & Manawatu	57#	8	0.98	0.18	0.78	0.60	2.14	7.65	-0.0017	0.85
Ramiha & Manawatu	55##	9	0.99	0.11	0.85	0.47	2.63	8.77	-0.0127	0.87
Ramiha	28##	4	0.95	0.16	0.55	0.51	1.52	8.68	-0.0078	0.56
Manawatu	25##	9	1.00	0.02	0.60	0.51	1.62	6.39	0.0126	0.57

Note: n = number of samples; Comp = component (factor or latent variable); RMSEC and RMSECV in mg cm^{-3}; # without removing outliers; ## with removing outliers; Bias = mean difference between measured and predicted root density; Slope = regression coefficient between measured and predicted root density

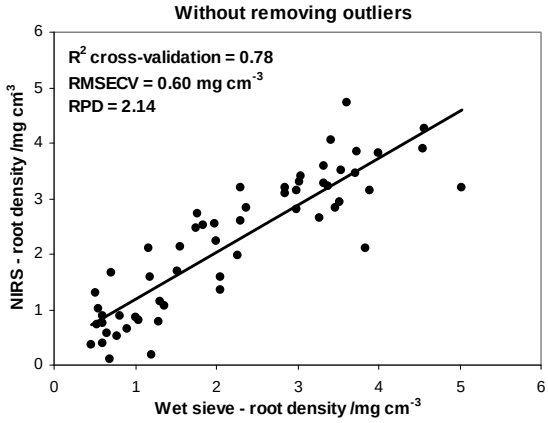
The pooled soil PLSR model relating spectral data (first derivative of 10-nm spaced smoothed reflectance) to measured root density (Figure 3.7c, cross-validation) predicted an acceptable quantitative estimate of root density across both soils (r^2 = 0.78 and RPD = 2.14; Table 3.3). Improved calibrations between the first derivative of spectral reflectance and root density measured by wet-sieving were achieved (larger r^2 and smaller RMSEC) when two outliers were removed (Figure 3.7d). Removing outliers also improves the root density prediction as shown by using cross-validation (r^2 = 0.85 and RPD = 2.63; Table 3.3; Figure 3.7e). Successful prediction of root density indicates that the NIRS-PLSR technique can extract important information from reflectance spectra which is very well correlated with the variance of the root density. This technique will allow root density to be predicted without separating roots from soil.



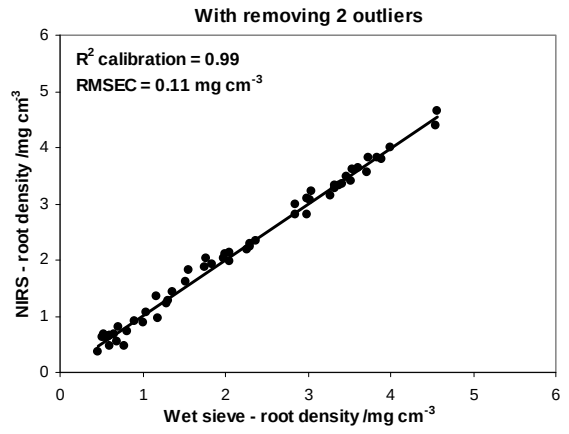
(a)



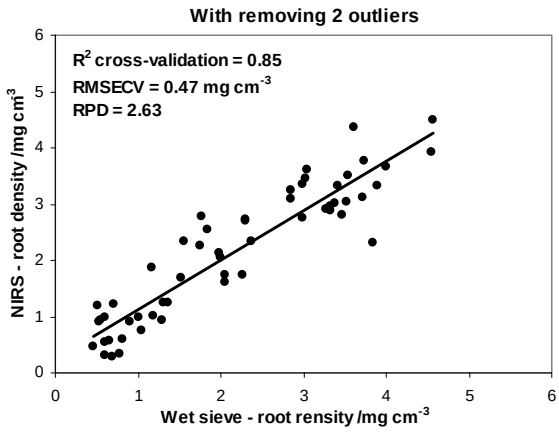
(b)



(c)



(d)



(e)

Figure 3.7 (a) Partial least squares regression calibration and cross-validation of spectral and root-density data in each soil type plotted on the same graph, (b, d) calibration of the spectral and root-density data using both soils, and (c, e) prediction of root density from both soils by leave-one-out cross-validation.

3.3.5. Important Wavebands Explaining the Variance of Root Density

The PLSR model for predicting root density from the first derivative of the smoothed reflectance spectra is presented graphically in Figure 3.8. The size of the coefficient (positive or negative) represents the importance of the band in terms of explaining the variance observed in root density. Important bands (bands with large value coefficients) in the visible region (400-700 nm) occur close to 600 nm (yellow-orange) and 750 nm (far red). This indicates that variability in root density is strongly represented by the apparent root colour. Williams (2001) reported starch bands at 614 and 758 nm, and cellulose contributing also at 758 nm. Important bands occur also in the NIR region (780-2500 nm), and more obviously at bands between 1000-1200 nm, 1460-1960 nm and 2100-2400 nm (Figure 3.8). These are attributed to, particularly, C-H, O-H and N-H bonds; vibration of those bonds absorbs energy (light) from the NIR region. Light absorptions in these bands were highly correlated with root density as shown by larger regression coefficients (Figure 3.8).

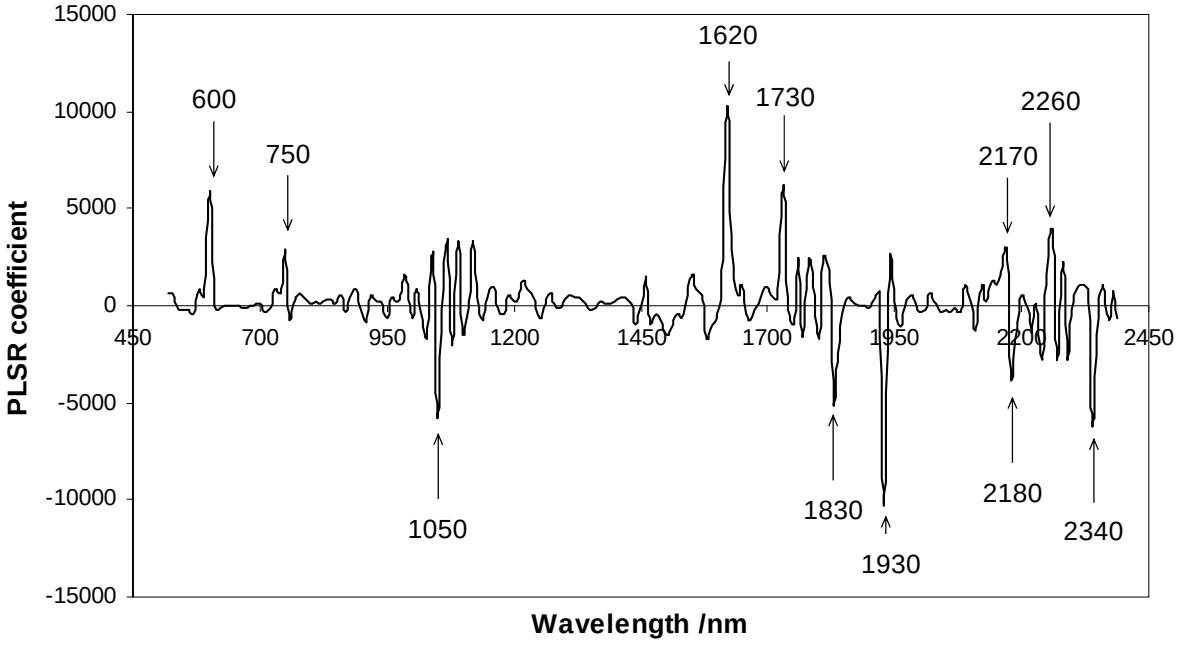


Figure 3.8 Coefficients of partial least squares regression model explaining the variance of root density.

A number of other researchers, Williams (2001), Ben-Dor *et. al*. (1997), Curran *et al.* (1992), Elvidge (1990), McLellan *et al.* (1991a), McLellan *et al*. (1991b) have reported that bands between ~600 and ~2500 nm are absorption bands of the main

constituents of plant material such as protein, cellulose, starch, oil, lignin, pectin, glucan, humic acid, wax, tannin and water.

3.4. Conclusions

The development of a self-illuminated soil probe permits the acquisition of reflectance spectra, in the Vis-NIR range, from samples of soil. The reflectance spectra from two soil types were clearly modified as root density increased, irrespective of chromophores other than roots, which dominated the reflectance spectra between the two soils used. Partial least squares regression of the first derivative of the 10 nm spaced spectral data against measured root densities produced calibration models that allowed quantitative estimates of root densities to be predicted. Once calibrated, this technique has the potential to allow rapid measurement of root density without further separating root from soil, which remains a tedious procedure. This study indicates that further field evaluation of the prediction of root density using reflectance spectroscopy is warranted. The ability to predict root densities in soil profiles will be an important asset in evaluating plant species for their potential to sequester CO_2 into soil organic matter.

CHAPTER 4

Predicting Pasture Root Density from Soil Spectral Reflectance: Field Measurement

The successful methods developed for root density prediction in the pot trial described in Chapter 3 are now evaluated for predicting pasture root density in the field. The method is also extended to the prediction of soil C and N concentrations.

Some results of this study have been published in the workshop proceeding of Designing Sustainable Farms: Critical Aspects of Soil and Water Management, 8th and 9th February 2007, Fertilizer & Lime Research Centre Massey University, Palmerston North, New Zealand. A paper from complete results of this study has been submitted to *European Journal of Soil Science*, on January 2009 (under reviewed).

4.1. Introduction

The potential to predict root density from the visible-near infrared (Vis-NIRS) reflectance spectra of soil was demonstrated by Kusumo *et al.* (2009a) using a glasshouse ryegrass pot trial. In this paper, the method is modified to acquire reflectance spectra from soil cores taken in the field. This modified method is evaluated as a rapid technique for predicting root density in field soils.

In our glasshouse study (Kusumo *et al.* 2009a), soil chromophores [e.g. soil organic matter (SOM), iron oxides, clay mineralogy, soil texture and aggregation] other than roots were homogenised by using sieved and well-mixed soils in pots. Under field conditions, these chromophores change significantly with distance and soil depth (Demattê *et al.* 2004a). Even in the same soil type, the dynamic nature of soil chromophores will influence the range of information acquired in the reflectance spectra and possibly influence the prediction accuracy for root density. Therefore, studies are required to confirm the potential for field-acquired reflectance spectra to predict root densities in pasture soils. If field root density can be predicted using Vis-NIR spectroscopy, the technique would be extremely valuable in support of studies measuring the potential for carbon (C) sequestration from fine root turnover, or, studies evaluating new deep rooting cultivars being developed for more efficient nitrogen (N)

and water use (Crush *et al.* 2007; Dunbabin *et al.* 2003). Thus, the primary objective of this study is to test whether Vis-NIR spectroscopy can be used to predict root density in the field.

Roots contribute to the SOM content via growth and decomposition (Trujillo *et al.* 2006). The amount of SOM in pastoral soil is closely related to net productivity of roots, which is determined by previous soil and pasture management history (Cullen *et al.* 2006; Nie *et al.* 1997) and type and stage of growth of the pasture (Fisher *et al.* 2007). Land slope and aspect may influence availability of water and soil fertility, which in turn influence pasture species and root density (Saggar *et al.* 1999) and soil organic matter content.

Vis-NIR spectroscopy can quantify both root density (Kusumo *et al.* 2009a) and soil C and N (e.g. Chang and Laird 2002; Kusumo *et al.* 2008a; Moron and Cozzolino 2002). If this technique can be used to quantify root density independently from soil C and N content, it will be particularly useful for monitoring SOM dynamics.

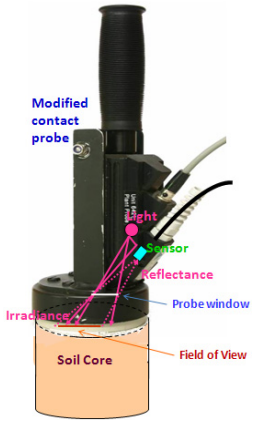
Vis-NIR spectroscopy does not provide a direct independent measure of root density, nor any soil property for that matter. Acquired spectral properties must be calibrated against standard reference measurements e.g. soil C content, or in this case, root density measured by separating roots from soil by wet sieving. Other researchers have applied calibration models developed from soil samples studied in controlled conditions (e.g. in the laboratory) to predict soil constituents of interest in field samples (e.g. Mouazen *et al.* 2007). This technique may result in accurate prediction if the spectral range of samples from the controlled condition (or calibration set) embraces the same range as the field samples. Often, differences in soil moisture and structural condition confound such an application. Therefore a secondary objective of this study is to test whether a calibration model developed using soil spectral and root density data from glasshouse grown ryegrass plants (Kusumo *et al.* 2009a) could be used to predict field root density in pastoral soils.

4.2. Materials and Methods

4.2.1. Site Locations and Soil Property Measurement

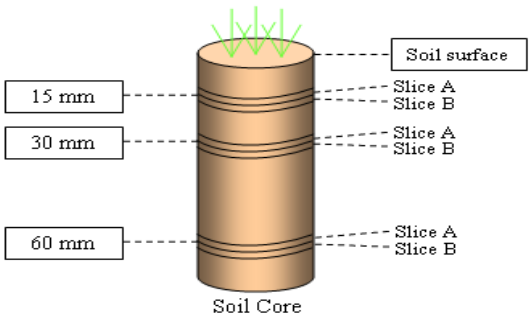
The root density assessment was evaluated at two permanent pasture sites: the first on Ramiha silt loam (Allophanic soil, Andic Dystrochrept derived from loess and andesitic ash) and the second on Manawatu fine sandy loam (Fluvial Recent soil, Dystric Fluventic Eutrochrept derived from greywacke alluvium) in the Manawatu region, New Zealand (Hewitt 1998). The permanent ryegrass (*Lolium perenne*) white clover (*Trifolium repens*) dominant pastures at both sites had been present for more than 20 years. A total of 18 soil cores, 10 m apart from each other, were collected from each site at 3 depths (15, 30 and 60 mm), providing 54 samples at each site and a total of 108 soil samples. Soil reflectance spectra were acquired at three horizontal surfaces (15, 30 and 60 mm) cut through an 80 mm diameter soil core (Figure 4.1c - 1e). A 3 mm soil slice (slice A), was collected, at each surface, and stored moist at 4^{o}C for no more than 3 days, before root densities were measured (Figure 4.1f) using the wet sieve method (Kusumo *et al.* 2008a). Root density was expressed as mg dry root g^{-1} dry soil.

A further 3 mm soil slice (slice B) was collected and used for determining dry soil weight and soil moisture content (see Figure 4.1c). Water content of subsample slice B was determined using the gravimetric method of drying at 105^{o}C in an oven until reaching a constant weight. For total C and N analysis, 42 and 36 samples were analysed from Ramiha and Manawatu soil respectively. Total C and N were measured from an air-dry subsample of slice B using a Leco FP-2000 CNS analyser (LECO Corp., St Joseph, MI, USA). Chemical (pH H_2O with ratio of soil and water 1 : 2.5, P-retention, Olsen-P, cation exchange capacity – CEC, sulphate, exchangeable-K -Ca -Mg and -Na, total C, total N) and physical properties (soil texture) of the two soils were also determined using three soil samples, each a composite of six replicate soil cores taken from the 0 – 100 mm soil depth (Blakemore *et al.* 1987). Bulk density was measured using separate soil cores.



(a)

(b)



(c)

(d)

(e)

(f)

Figure 4.1 (a) Modified soil probe was used (e) to collect reflectance spectra from (b and c) a soil core, and then (d) the soil slice (f) was washed to separate root from soil using wet sieve method.

4.2.2. Developing Calibration Model

Diffuse spectral reflectance of each freshly cut soil surface was recorded using a purpose-built soil probe (Figure 4.1a-e). The spectral data were pre-processed (for details see Kusumo *et al*., 2009a) and first derivatives of 5-nm spaced data calculated using SpectraProc V 1.1 software (Hueni and Tuohy 2006). The first derivative data were imported to Minitab 14 (MINITAB Inc. 2003) for partial least squares regression (PLSR) analysis. Calibration models were developed by using PLSR to fit the reference data (root density, soil C and N) to pre-processed spectral data. The resulting regression models were then used to predict root density, soil C and N in unknown samples. The accuracy of the models was tested internally using a leave-one-out cross-validation procedure and externally from separate validation sample set. Separate calibration and validation sets were determined by ranking the soil samples from the lowest to the highest root density, and odd and even ranked numbers were allocated to calibration and validation set, respectively. This resulted in 1:1 ratio of calibration and validation set. The number of principal components (latent variables or factors) used to develop calibration models were those which produce the lowest PRESS (predicted residual error sum of squares) in the leave-one-out cross-validation procedure (Miller and Miller 2005). During PLSR processing, samples which had a standardized residual > 2.0 were removed as outliers (MINITAB Inc. 2003) from the calibration and validation sets.

4.2.3. Principal Component Analysis

Prior to PLSR analysis, a principal component analysis (PCA) was conducted on the first derivative of the spectral data. A score plot of the first two components (PC1 and PC2) was used to explore the pattern of spectral differences between the Ramiha and Manawatu samples.

4.2.4. Predicting Field Root Density Using the Calibration Model Constructed from Glasshouse Data

Different root densities were created (Kusumo *et al.* 2009a) by differential nitrogen and phosphorus fertilization of ryegrass (*Lolium multiflorum* Lam.) grown in pots of Ramiha and Manawatu soil in a glasshouse. A PLSR calibration model was developed from the first derivative spectra and root density data from the glasshouse grown ryegrass plants. This calibration model was used to predict root density in this study.

The root density unit used in the glasshouse study was mg dry root cm^{-3} and using pot bulk densities was converted to mg dry root g^{-1} dry soil to make it compatible with the measurements made in the field study.

4.2.5. Regression Model Accuracy

The ability of PLSR models to predict soil properties was assessed using the following statistics. RMSE (root mean square error) is the standard deviation of the difference between the measured and the predicted soil property values. RMSE which is calculated from cross-validation is called RMSECV, and from validation data is RMSEP. RPD (ratio of prediction to deviation) is the ratio of the standard deviation of measured value of soil properties to the RMSE. RER (ratio error range) is the ratio of the range of measured values of soil properties to the RMSE. The best prediction model is shown by the highest RPD, RER, r^2 and the lowest RMSECV or RMSEP (for more detailed explanation see Kusumo *et al*., 2009a).

4.3. Results and Discussion

4.3.1. Summary of Pasture Root Density, Soil Chemical and Physical Data

The range of measured pasture root densities and soil chemical and physical properties of Ramiha silt loam and Manawatu fine sandy loam soil are presented in Table 4.1. The mean of root density at each depth in Ramiha soil is less than in Manawatu soil, except at the 60 mm depth. Differences in root density may result from the different characteristics of each soil. For example, the Manawatu soil has a fine sandy loam texture with high available phosphate status (Olsen-P, Table 4.1), low P retention and greater concentration of exchangeable cations than the silt loam textured, high P retention, Ramiha soil. Mechanical impedance to root growth is unlikely in both soils, which have bulk densities below the critical level (1.4 g cm^{-3}) that restricts root growth.

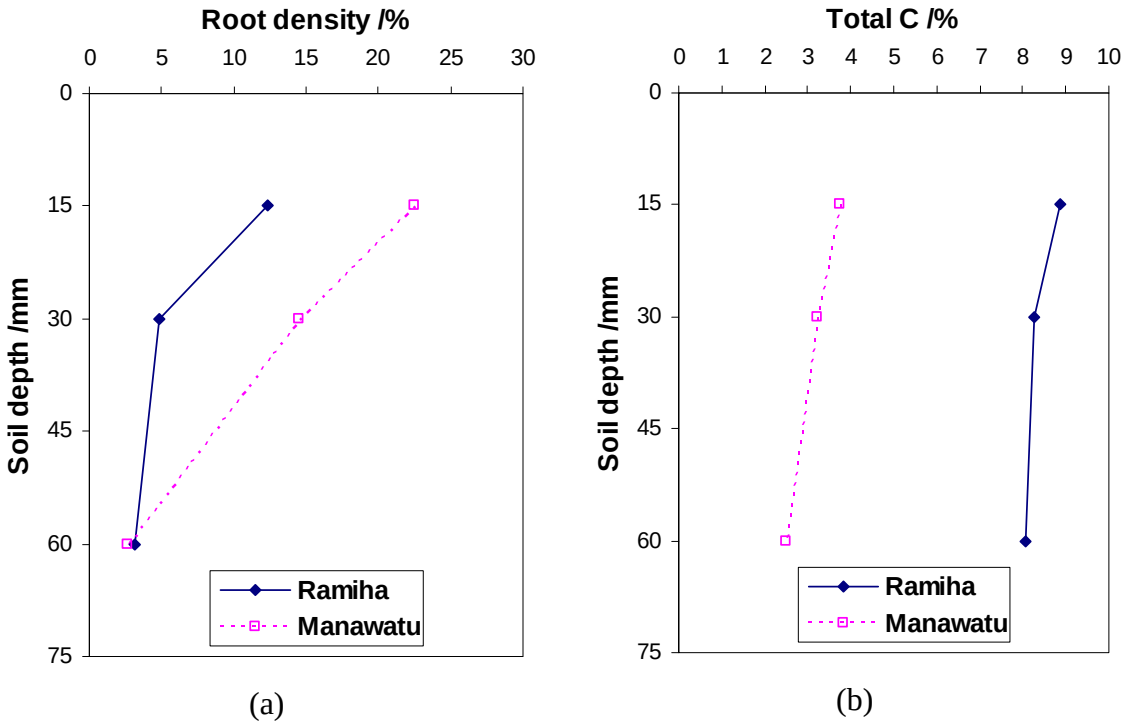
Table 4.1 Root density, chemical and physical properties of Ramiha and Manawatu soil.

Properties	Ramiha	Manawatu
Root density (mg g^{-1} of 54 samples of each Ramiha and Manawatu)		
Minimum	2.27	1.53
Maximum	23.10	37.03
Mean	6.82	13.21
Variance	23.61	99.06
Mean root density (mg g^{-1} each depth)		
at 15 mm depth	12.39	22.51
at 30 mm depth	4.85	14.49
at 60 mm depth	3.20	2.63
Total C (% of 42 Ramiha and 36 Manawatu samples)		
Minimum	6.19	2.13
Maximum	10.19	4.47
Mean	8.39	3.17
Variance	0.82	0.35
Total N (% of 42 Ramiha and 36 Manawatu samples)		
Minimum	0.51	0.24
Maximum	0.73	0.43
Mean	0.61	0.32
Variance	0.0025	0.0021
Water content (% of 54 samples of each Ramiha and Manawatu)		
Minimum	50.70	24.30
Maximum	70.04	38.60
Mean	57.11	31.24
Variance	28.24	14.88
Soil (0-100 mm)		
pH	4.7	4.9
P-Olsen (μgP g^{-1})	18.9	71.1
P retention (%)	73	14
CEC (me 100 g^{-1})	29.5	15.0
SO4 (μg S g^{-1})	85.0	38.9
K (me 100 g^{-1})	0.18	0.55
Ca (me 100 g^{-1})	4.7	5.9
Mg (me 100 g^{-1})	0.41	1.33
Na (me 100 g^{-1})	0.07	0.09
Total C (%)	7.38	3.18
Total N (%)	0.59	0.31
Bulk density (g cm^{-3})	0.89	1.18
Texture	Silt loam	Fine sandy loam

Root density, collected in summer (December/January), decreases significantly with depth, both in Ramiha and Manawatu soil (Figure 4.2a); however at the 30 and 60 mm depths in Ramiha soil, root densities were not significantly different.

Decreasing amounts of soil C with depth were also found in both soils (Figure 4.2b). Interestingly, Ramiha soil with lower root densities contains larger amounts of soil C. This is probably because SOM decomposition is inhibited by complexation with

the amorphous clay mineral allophane (Boudot *et al.* 1988) which is abundant in this soil (Theng *et al.* 1986). The finer texture of the Ramiha soil also presents more surface area for clay-organo complexes to be formed (Theng *et al.* 1986) and protect the organic matter from decomposition in smaller pores or soil aggregates (Goldchin *et al.* 1994; Six *et al.* 2000; Tisdale and Oades 1982).



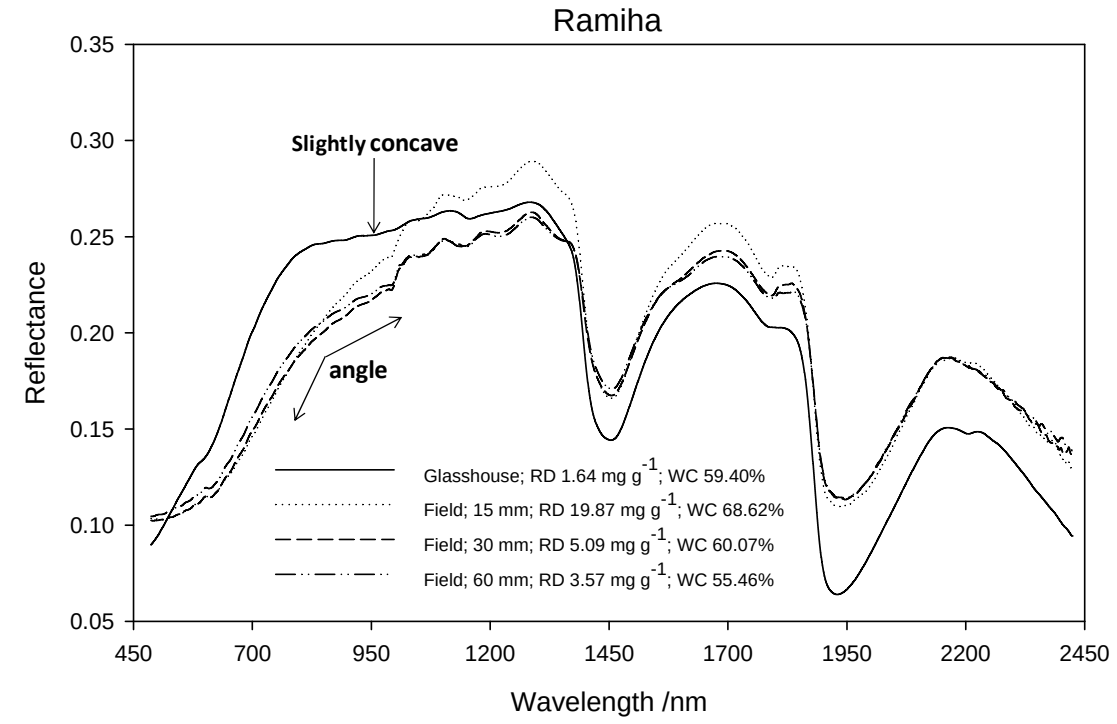
(a) (b)

Figure 4.2 (a) Root density and (b) total C at each depth in the Ramiha and Manawatu soil.

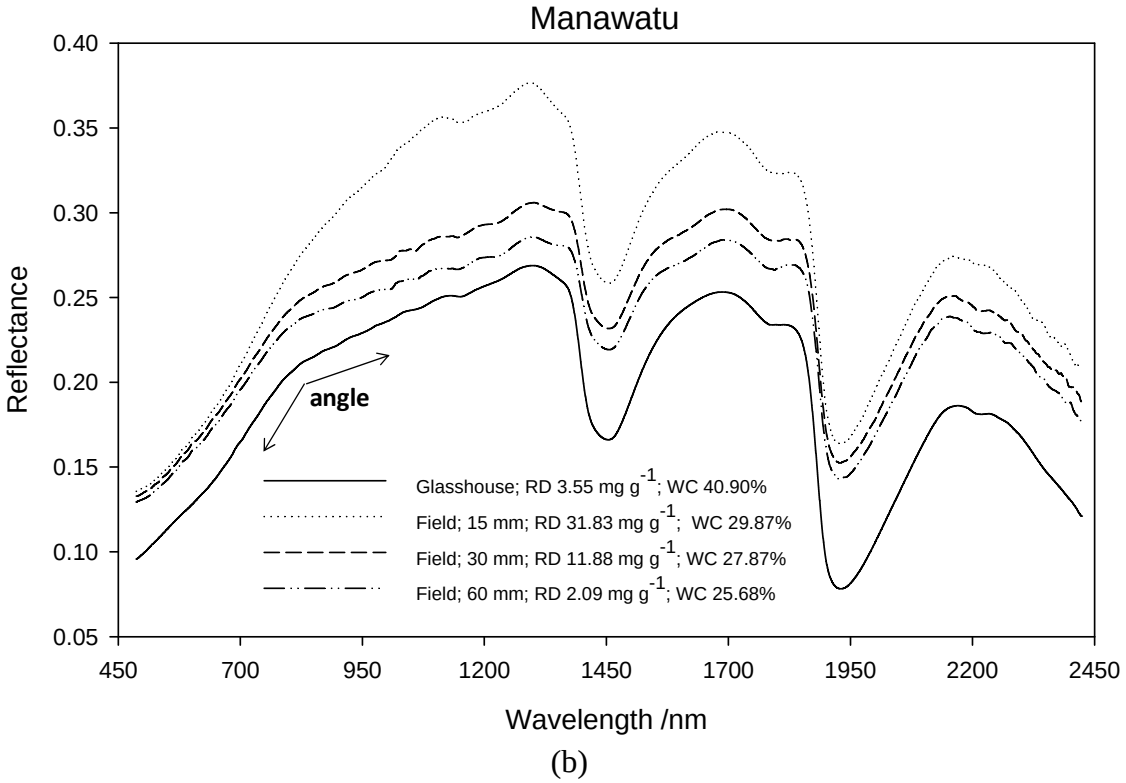
If the average amount of root mass in the 0-60 mm depth is expressed per hectare, the total root dry mass will be 3642 and 9353 kg DM ha^{-1} in Ramiha (Andic Dystrochrept) and Manawatu soil (Dystric Fluventic Eutrochrept), respectively. These values for root biomass are consistent with other values reported for grazed pastures in this region of New Zealand. Nie *et al*. (1997) reported winter (June) root masses of 3650 and 2100 kg DM ha^{-1} (0-600 mm depth) for grazed and fallowed pasture on Yellow-grey/Yellow-brown Earth intergrades (Typic Dystrochrepts) in New Zealand. Saggar *et al*. (1999) reported the average standing pasture root biomass at low, medium and steep slopes was 11 330, 13 310 and 12 210 kg DM ha^{-1} (0-100 mm) respectively, collected in spring from the same location as Nie *et al*. (1997).

4.3.2. Spectral Reflectance of Ramiha and Manawatu Soil

The reflectance spectrum collected from the glasshouse trial (Kusumo *et al.* 2009a) and the pattern of reflectance spectra changing with soil depth for each soil are presented in Figure 4.3. Spectra are influenced not only by root but also by other soil chromophores e.g. parent material, clay and non-clay minerals, iron oxides, water, and soil organic matter (Baumgardner *et al.* 1985; Ben-Dor 2002). In the Ramiha soil (Figure 4.3a) the glasshouse-collected spectrum with low root density depicts a much sharper angle at around 780 nm due to increased reflectance around the yellow-red band region (600-750 nm), probably caused by reflected colour from iron oxides. The effect of iron oxides can also be seen at the band around 950 nm; the spectrum is slightly concave as the iron oxides absorb more NIR light in this band region (Demattê *et al.* 2004a; Demattê and Garcia 1999). In the field soil spectra, high root content masks the effect of iron oxides, and the angle near 780 nm becomes more obtuse (Figure 4.3a). No obvious features of iron oxides were found in the Manawatu soil due to its naturally low iron oxide content (Figure 4.3b). For the Manawatu soil, the shapes of glasshouse-soil spectrum and field-soil spectra (low-root density) are quite similar, and account for the close spatial distribution of glasshouse- and field-Manawatu samples on the PCA score plot (Figure 4.10). The effect of root on the spectral reflectance can be clearly seen from the surface samples (15 mm depth), which have a more obtuse angle and tend to have higher reflectance in the near infrared region 780-2500 nm, and is more pronounced in the Manawatu (Figure 4.3b) than in the Ramiha soil (Figure 4.3a). Spectra from mixtures of dry soil and ground root (Kusumo *et al.* 2009a) also showed higher reflectance with increasing root content. Higher reflectance in the near infrared region is probably due to higher air voids contained in the root and air-root cell interfaces which cause increased reflection (Baumgardner *et al.* 1985). In addition, spectral reflectance from the 15 mm depth sample tend to resemble the spectral reflectance of senescent leaves (Gausman *et al.* 1975). They also resemble minimally decomposed soil material (fibric) reported by Stoner and Baumgardner (1981).



(a)



(b)

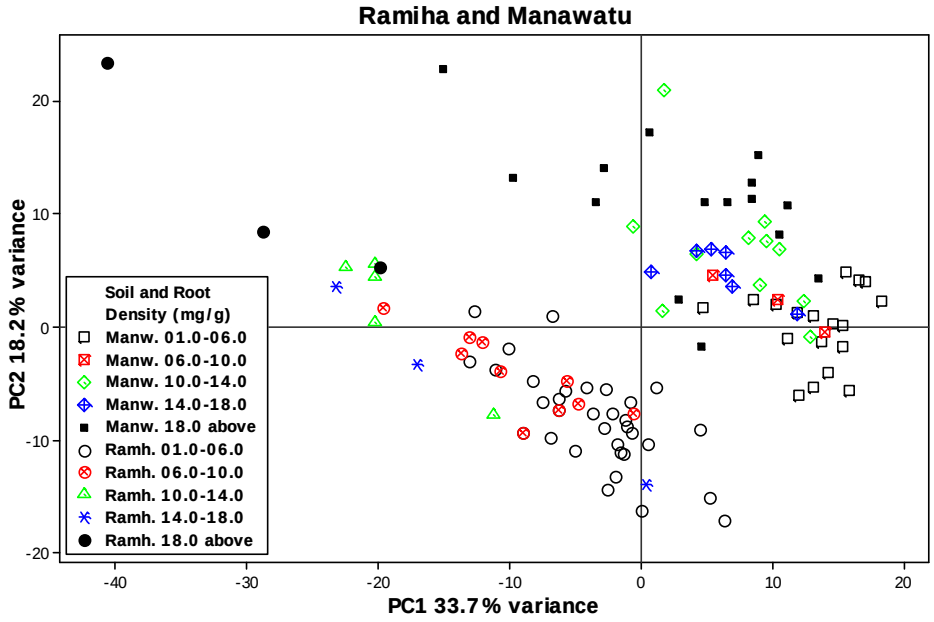
Figure 4.3 Spectra acquired from soils with different root densities from the glasshouse trial (Kusumo *et al.* 2009a) and from soil cores collected at different depths in the field; (a) Ramiha and (b) Manawatu soil. RD and WC are root density and water content, respectively.

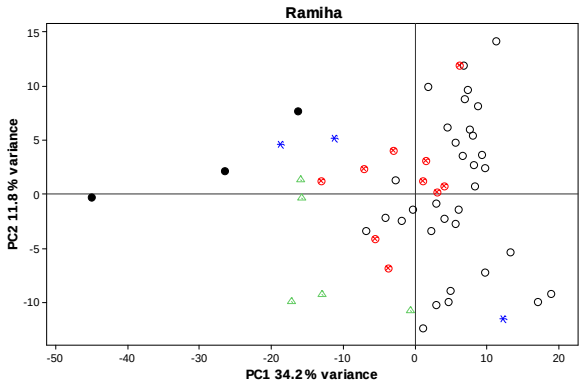
4.3.3. Principal Component Analysis of the Spectral Data

A principal component analysis was used to transform the spectral data set (first derivative) into a smaller set of linear combinations, or principal components, that represent as much of the variance of the spectral data as possible (Dillon and Goldstein 1984; Miller and Miller 2005). The first two principal components (PC), therefore, represent the largest variance of the spectral data possible for 2 dimensions. The score plot of the first two PCs (Figure 4.4a) calculated from the combined soils clearly differentiates between the spectral properties of the Ramiha and Manawatu. Soil chromophores, other than root, such as clay types, non-clay minerals, iron oxides, soil moisture and organic matter may contribute to this obvious discrimination. PC1, which contains the largest spectral variance (33.7%), is not correlated with changes in root density (Figure 4.4d; r = -0.13), but samples with similar root densities tend have a vertical distribution in the score plot, showing that PC2 is correlated with density (Figure 4.4a). This is confirmed when the components are calculated separately for the two soils (Figure 4.4b and 4.4c). By separating the soils (i.e. removing the variance due to soil type), root density becomes a source of variance that is more highly correlated with the first principal component (PC1).

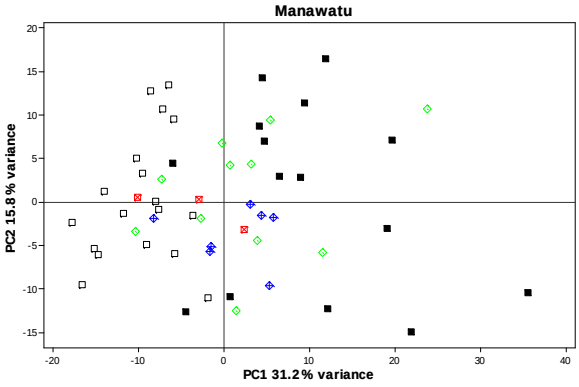
If the soil property of interest causes the major variance in the spectral data, then the PCA of spectral data followed by a PC score plot analysis can be useful for choosing the sample population to be used in the PLSR model calibration. However, such preliminary analysis is of little use if the soil property of interest causes only small variance in the spectral data. PLSR analysis, however, selects the component that is most strongly correlated to a soil property providing useful predictions even if that component is not a large part of total soil variance. In the case of the data presented in Figure 4.4a, b and c, the PCA analysis and the subsequent correlation (Figure 4.4d) of principal components against root densities suggest that variance due to root density could be allocated to PC1 if soils are separated, and allocated to PC2 if soils are grouped.

(a)





(b)



(c)

(d)

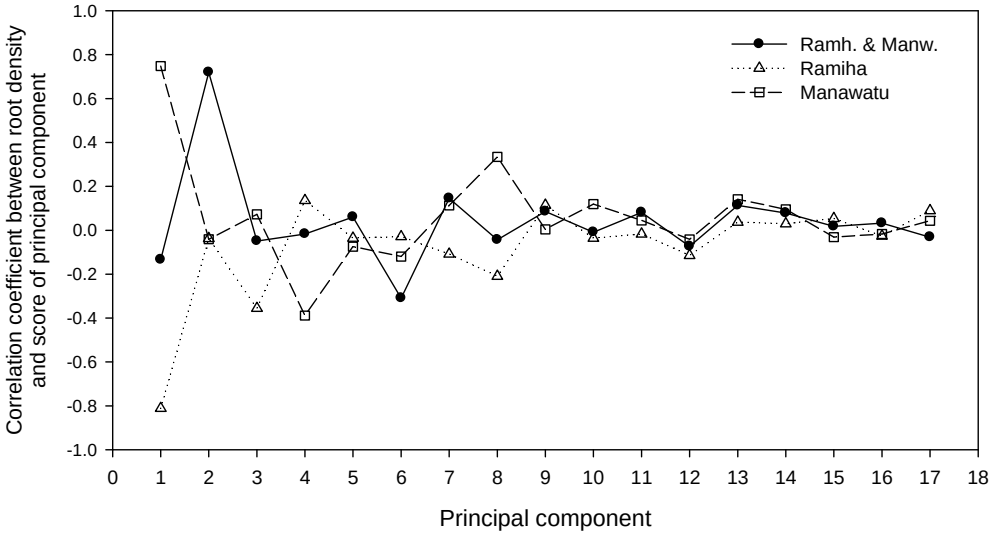


Figure 4.4 A score plot of the first and second components from the PCA of the first derivative of the reflectance spectra acquired from (a) combined and (b) separated Ramiha and (c) Manawatu soil and (d) coefficients of correlation between root density and the principal components when the two soils were combined and separated.

4.3.4. Can Pasture Root Density in the Field be Predicted from Spectral Reflectance?

Figure 4.5 illustrates the relationship between measured root density (wet-sieve) and predicted root density (Vis-NIRS technique, see also Table 4.2). PLSR of first derivative spectra and reference data produces good calibration models which are shown by a high coefficient of determination (r^2 > 0.99) and small error (RMSEC = 0.43 and 0.93 mg g^{-1}) obtained for each soil type, Ramiha and Manawatu. Weaker calibration relationships (r^2 = 0.86; RMSEC = 3.09 mg g^{-1}) are found when samples from both soil types are amalgamated. Obvious separation between Ramiha and Manawatu soil on the score plot (Figure 4.4a) and better correlation between root density and scores of PC1 when the two soils are separated (Figure 4.4d) support a separate development of the calibration model. Due to the small number of samples involved (less than 60 samples), leave-one-out cross-validation procedure was initially used to test the model (Williams 2001). More accurate cross-validation predictions are found also when samples from each soil type are separated rather than grouped (Table 4.2). More accurate root density prediction was found for Ramiha (RPD = 2.42; r^2 cross-validation = 0.83) than for Manawatu soil (RPD = 1.98; r^2 cross-validation = 0.75). Malley *et al*. (2004) categorised prediction accuracy into moderately useful (in Manawatu soil) when RPD 1.75-2.25 (with RER 8-10 and r^2 0.70-0.80) and moderately successful (in Ramiha soil) when RPD 2.25-3 (with RER 15-20 and r^2 0.80-0.90).

Table 4.2 The partial least squares regression calibration and cross-validation of spectral and root-density data when observations from Ramiha and Manawatu field pasture soils are combined and separated.

Soil Type	Allocated samples	Outliers	--Calibration--			---------Leave-one-out cross-validation--------						
			n	r^2	RMSEC	Comp	r^2	RMSECV	RPD	RER	Bias	Slope
Ramiha & Manawatu	108	2	106	0.86	3.09	4	0.66	4.80	1.71	7.39	0.063	0.70
Ramiha	54	2	52	0.99	0.43	6	0.83	1.96	2.44	10.65	-0.028	0.87
Manawatu	54	3	51	0.99	0.93	6	0.75	5.11	1.98	6.95	0.011	0.81

Note: n = number of samples; comp = components (factors or latent variables); RMSEC and RMSECV in mg dry root g^{-1} soil; Bias = mean difference between measured and predicted root density; Slope = regression coefficient between measured and predicted root density

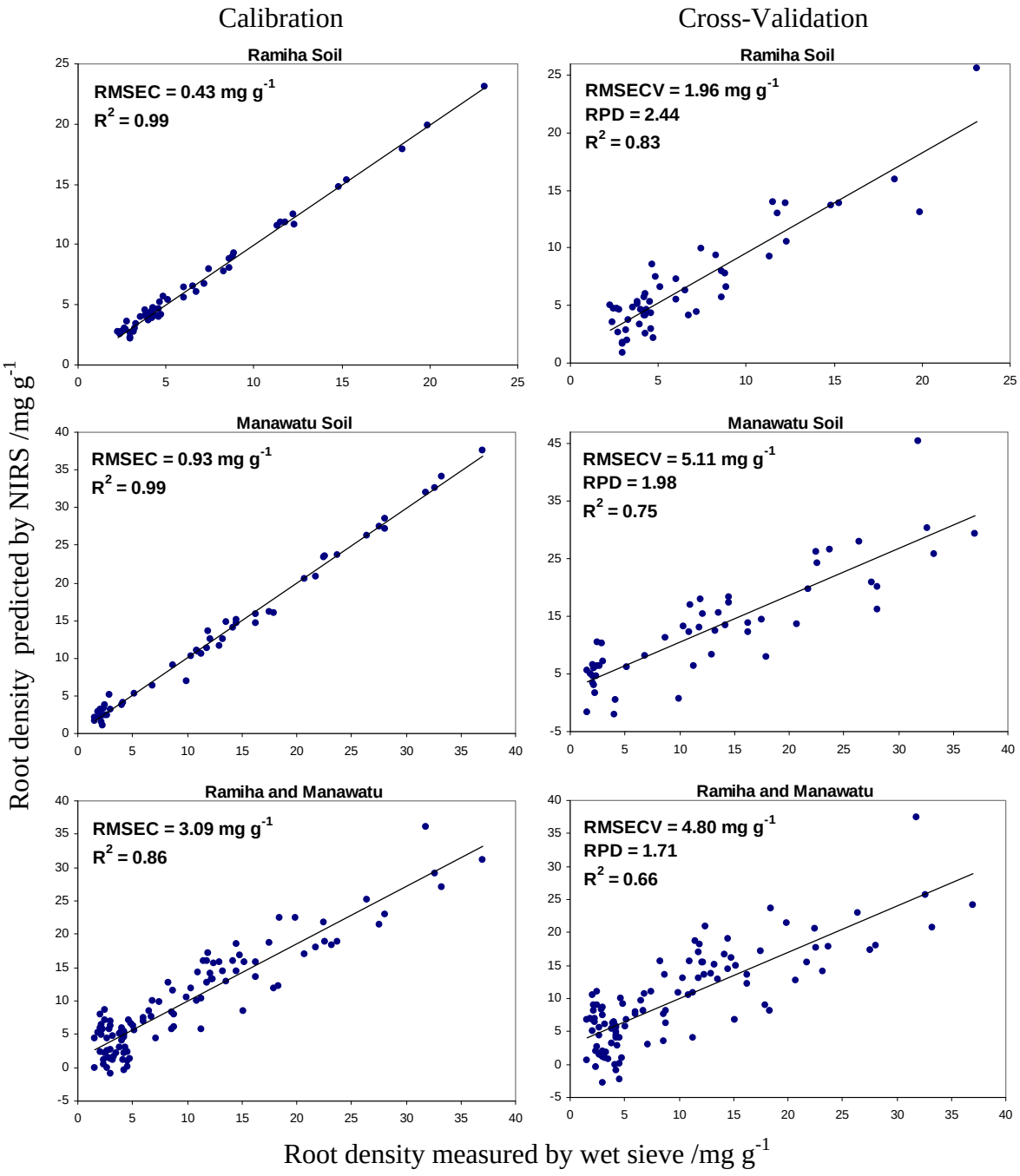


Figure 4.5 Relationship between root density measured by wet sieve and predicted from soil spectral reflectance for PLSR calibration (left) and cross-validation data (right) of separated (above and middle) and combined (below) Ramiha and Manawatu soil data.

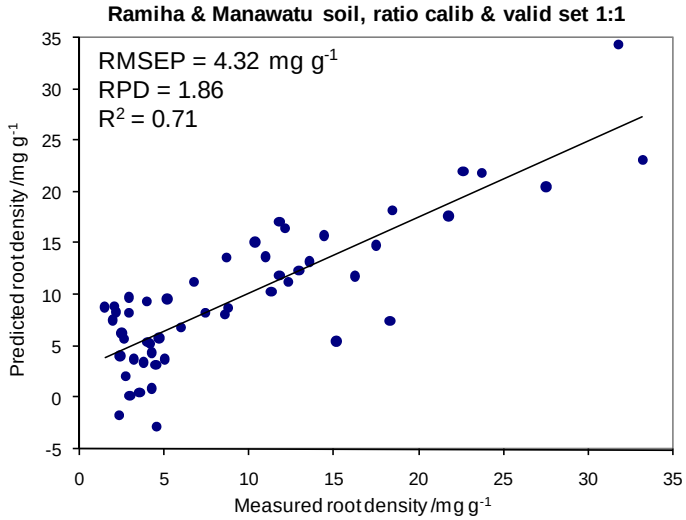
Some researchers do not favour cross-validation as a preferred option for testing the accuracy of a calibration model, preferring external validation with separate calibration and validation sets. Although producing separate calibration and validation sets for the Ramiha and Manawatu soils (Table 4.3; Figure 4.6), involves very low sample numbers (25-26), we still find that the calibration models gave “moderately” accurate prediction

of root density (Figure 4.6b and c). These results show that Vis-NIR spectroscopy has the potential to be used for prediction of pasture root density in the field. One could imagine that the technique could be used in conjunction with a standard root measurement technique (such as wet sieving) to reduce the numbers of samples and time spent processing soil by the standard technique.

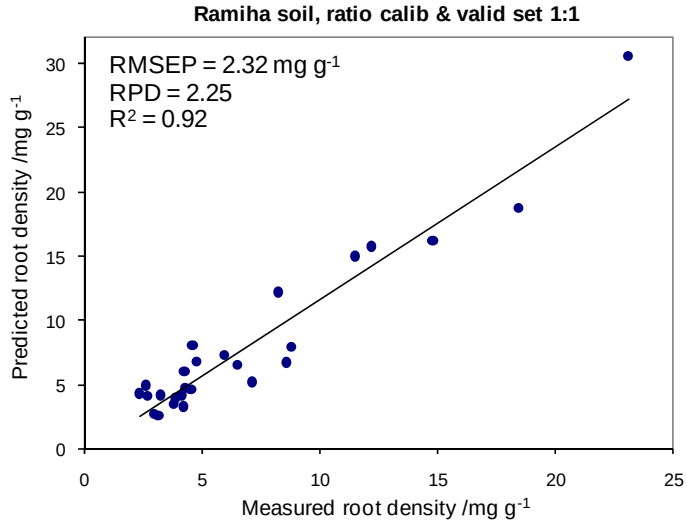
Table 4.3 Root density prediction using separate calibration and validation set (ratio 1:1) of the amalgamated and separated Ramiha and Manawatu soils.

Soil Type	Allocated samples	Outliers	Calib set	Valid set	Validation values						
					Comp	r^2	RMSEP	RPD	RER	Bias	Slope
Ramiha& Manawatu	108	2	54	52	4	0.71	4.32	1.86	7.34	0.29	0.74
Ramiha	54	2	26	26	3	0.92	2.32	2.25	8.94	1.06	1.19
Manawatu	54	3	26	25	4	0.80	5.02	1.98	6.32	2.40	0.88

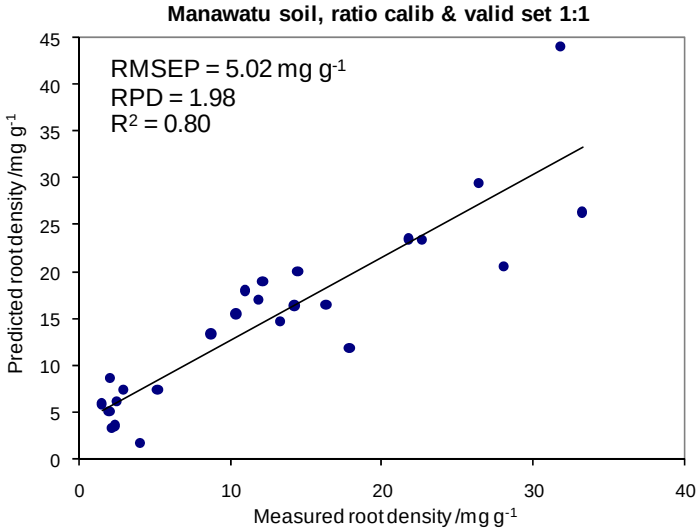
Note: Comp = components (factors or latent variables); Bias = mean difference between measured and predicted root density; Slope = regression coefficient between measured and predicted root density



(a)



(b)



(c)

Figure 4.6 Relationship between root density measured by wet sieve and predicted by NIRS of (a) combined and (b) separated Ramiha and (c) Manawatu soil data.

4.3.5. Is Root Density Predicted Independently from Soil C?

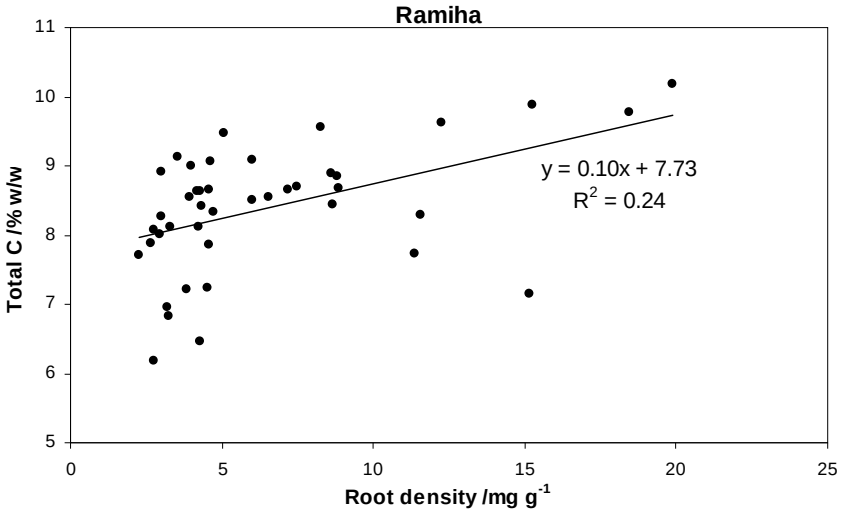
The interaction of NIR light with organic materials containing bonds of small atoms such as N-H, C-H, and O-H allows spectral reflectance to be used to predict organic materials including roots (Kusumo *et al.* 2009a). Spectral reflectance collected from pasture soils reflects root attributes and decomposed soil organic matter, both of which decrease with soil depth (Figure 4.2). Both root density (r^2 = 0.83, Table 4.2) and soil C (r^2 = 0.86, Table 4.4) can be predicted with high accuracy using this NIRS technique (Table 4.2 and 4.4) in the Ramiha soil. Poor correlations exist, however, between measured root density and total C of the Ramiha soil (r^2 = 0.24, Figure 4.7a), indicating that root density and soil C are predicted independently by the PLSR of spectral data. Closer examination of the PLSR model coefficients, show that different wavebands are given different levels of importance in the regression models used to predict root density and soil C (Figure 4.7c). In addition, soil N content is also predicted with reasonable accuracy (r^2 = 0.72, Table 4.4) in the Ramiha soil, despite a poor correlation between LECO-measured soil C and N (r^2 = 0.38). Independent prediction between soil C and N is reported also by Chang and Laird (2002). In the Manawatu soil, however, strong correlations exist between root density and soil C (r^2 = 0.73) and between soil C and N (r^2 = 0.83). Because they are correlated in the data, it is not possible to show that root density can be predicted independently of change in soil organic matter (Table 4.2 and 4.4).

Table 4.4 The partial least squares regression calibration and cross-validation of spectral and soil C and N data from Ramiha and Manawatu field pastoral soils.

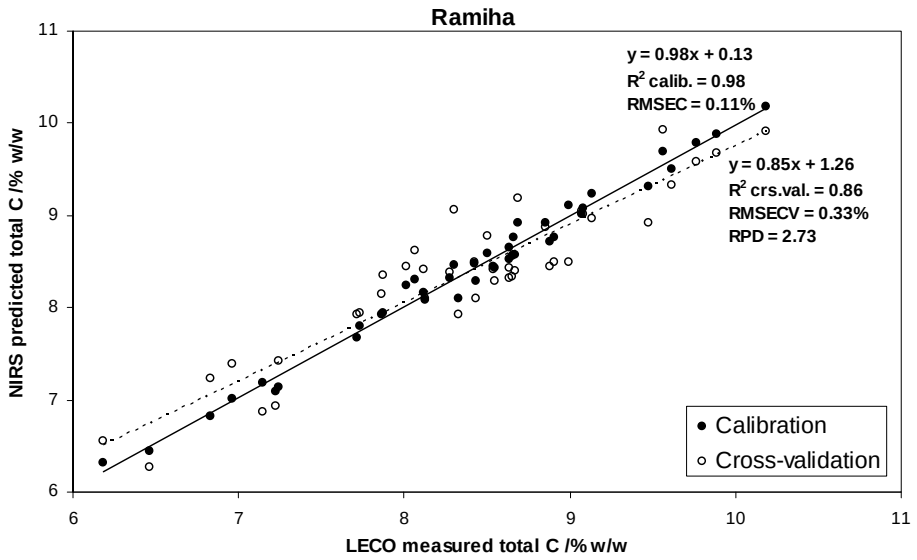
Soil type	Allocated samples	Soil property	Calibration		----------Leave-one-out cross-validation----------						
			r^2	RMSEC	Comp	r^2	RMSECV	RPD	RER	Bias	Slope
Ramiha	42	C	0.98	0.11	4	0.86	0.33	2.73	12.06	-0.010	0.85
Ramiha	42	N	1.00	0.0009	9	0.72	0.03	1.90	8.36	-0.0002	0.67
Manawatu	36	C	0.99	0.04	5	0.86	0.22	2.64	10.41	-0.006	0.77
Manawatu	36	N	1.00	0.002	6	0.80	0.02	2.24	9.72	-0.0001	0.72

Note: Comp = components (factors or latent variables); RMSEC and RMSECV in %; Bias = mean difference between measured and predicted soil C and N; Slope = regression coefficient between measured and predicted soil C and N

a



b



c

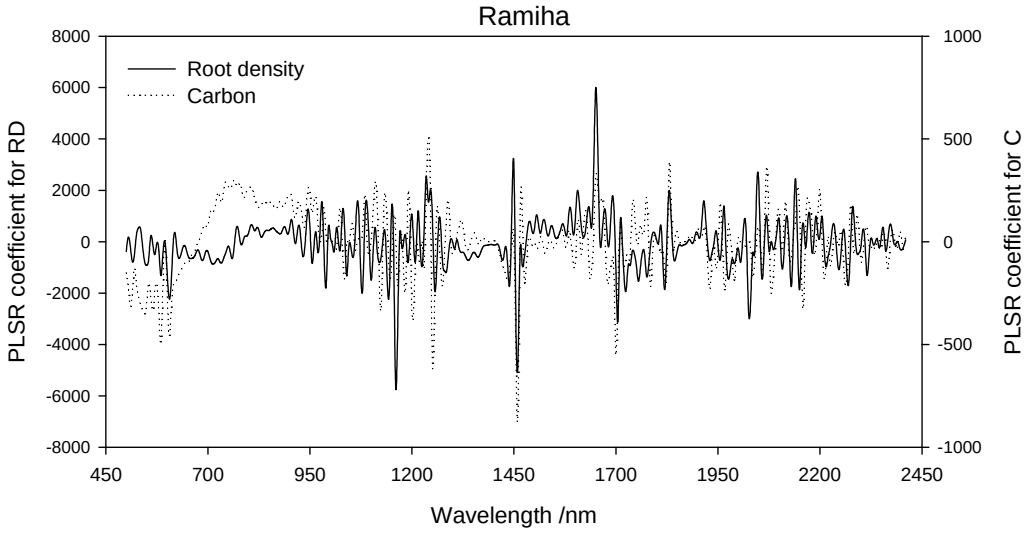


Figure 4.7 Relationships between (a) measured root density and soil C, (b) measured soil C and predicted soil C, and (c) the coefficients of the PLSR models used to predict root density (RD) and % C of the Ramiha soil in the field.

4.3.6. PLSR Coefficient Reflecting Important Bands

Coefficients resulting from PLSR models for predicting root density in the pastures growing on Ramiha and Manawatu soils are presented graphically in Figure 4.8. The size of the coefficient (negative or positive) in each model represents the importance of each wavelength band in explaining the variation in root density. Both calibration models show differences in selection of wavelengths that vary with root density. These differences between soils are expected because parts of the spectra are unique to each soil and in addition roots, an "intrusive" feature of soils, will influence these unique spectral regions. Other regions in the spectra are derived directly from the root material. Interestingly, both models select the visible band at 600 nm (yellow-orange) as an important band, which is similar to a previous study conducted in the glasshouse (Kusumo *et al.* 2009a), where it was associated with reflectance from grass roots. Some other important bands in the near infrared region (see Figure 4.7), particularly at 1620, 1830 and 1930 nm, were important in relationships describing ryegrass root density in the glass house study (Kusumo *et al.* 2009a).

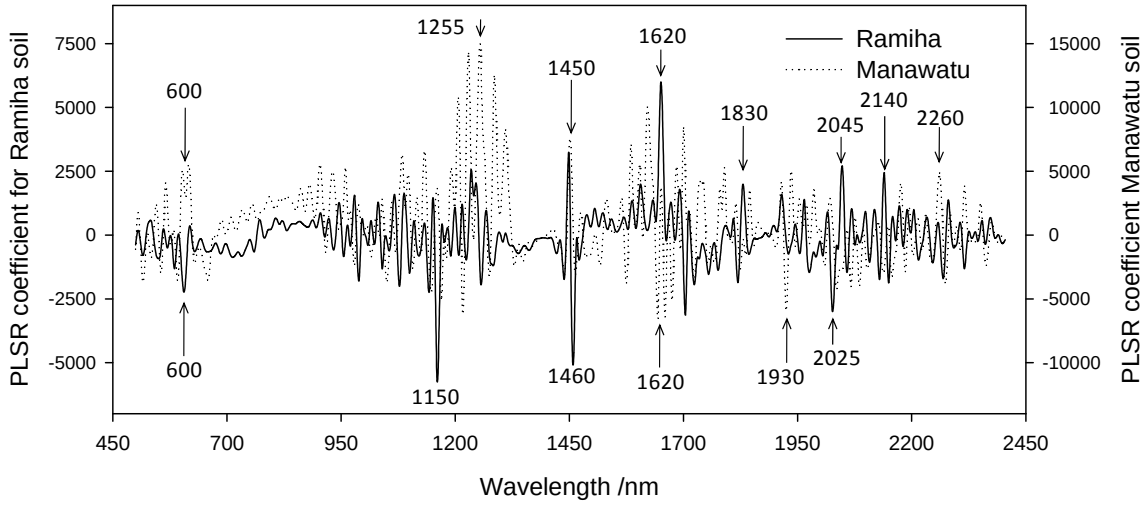


Figure 4.8 Coefficients of the PLSR models for predicting root density in Ramiha and Manawatu pastoral soils.

4.3.7. Can the Glasshouse-Prepared Calibration Model be used to Predict Field Root Density?

The best calibration model that was developed from the glasshouse study (Kusumo *et al.* 2009a) was used to predict root density from acquired spectral data in this study. The results demonstrate that root density was poorly predicted, which is shown by a

very small coefficient of determination (r^2 = 0.35), large error (RMSEP = 10.32 mg g^{-1}) and very poor RPD (0.82) (see Figure 4.9). The glasshouse calibration model tends to underestimate field root density values, which is shown by lower root density prediction (Figure 4.9). The poor prediction of the glasshouse calibration may be explained partly because the range and variance of the glasshouse root densities (range 0.60-5.46 mg g^{-1}, variance 1.78) do not cover the field observations (range 1.53-37.03 mg g^{-1}, variance 71.08). Glasshouse data have a smaller range and variance compared to field root density data. It is well known that the range of the calibration set should cover the range of the validation set (Malley and Martin 2003; Westerhaus *et al.* 2004; Williams 2001) to get the best calibration.

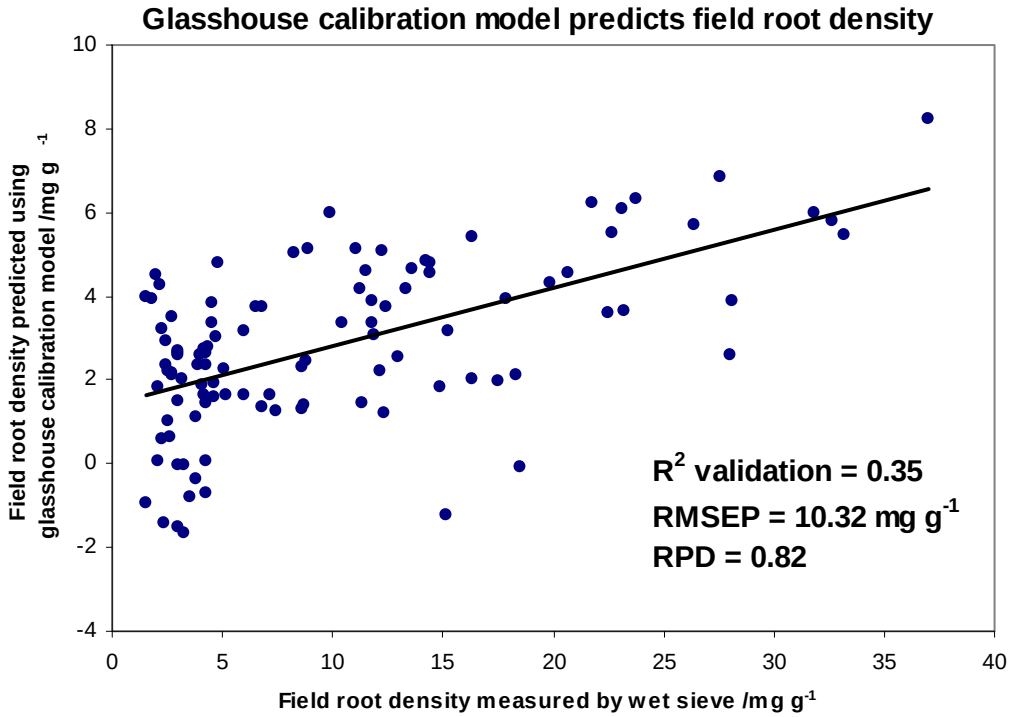


Figure 4.9 Field root density predicted by the glasshouse PLSR calibration model compared with the measured field root density for Ramiha and Manawatu soil.

The PC1 and PC2 score plot of the PCA of the spectra acquired from the field and glasshouse data (Figure 4.10) shows that the spectral properties of the glasshouse calibration set (full and empty circles) do not embrace the spectral variance of the validation set, the field spectral data (full and empty diamonds). The scattered distribution of the glasshouse calibration set (circles) occupies a different spatial distribution in the PC1 and PC2 score plot to the field validation set (diamonds) (Miller and Miller 2005). Ideally, the scattered distribution of the validation set should also be

within the distribution of the calibration set. Better prediction may be obtained if the distribution patterns between calibration and validation set on the score plot were more similar (Esbensen *et al.* 2006), and both sets should embrace the same variance dimensions (Williams 2001).

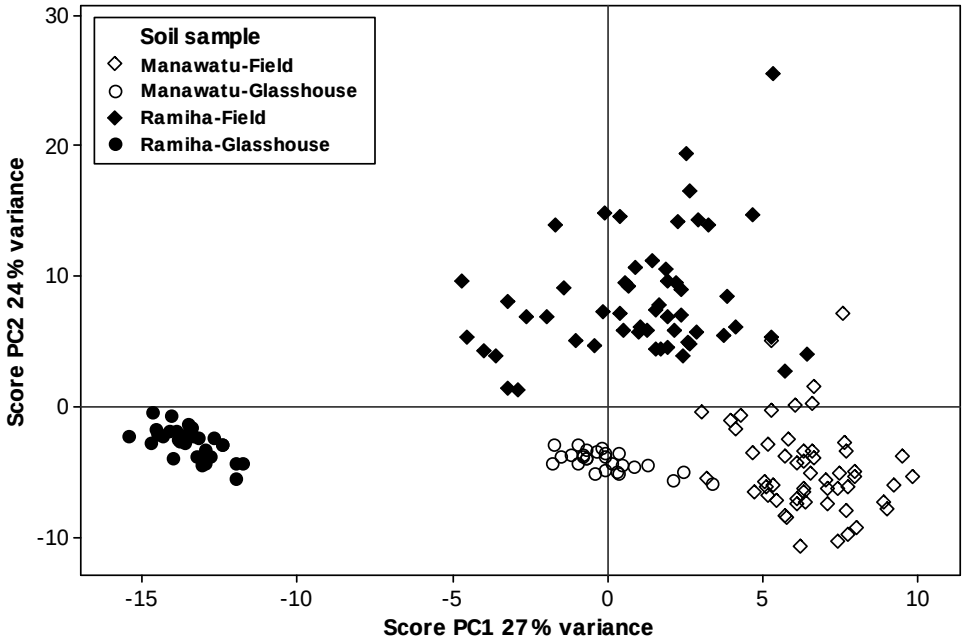


Figure 4.10 A score plot derived from the principle component analysis of the spectral reflectance data obtained from the glasshouse study and field study of Ramiha and Manawatu soil.

The inability of the glasshouse data to cover the same range of measured root densities and spectral reflectance characteristics implies that the most robust regression models for predicting pasture root density in the field should be developed from the field study rather than from the glasshouse samples, except in cases where similar glasshouse and field sample variance dimensions can be created. By creating a calibration model using a subset of the field samples and then predicting the remaining samples, can reduce the number of sample that must be analysed especially when dealing with a large number of samples (e.g. for mapping purpose).

4.4. Conclusions

A method has been developed that uses diffuse Vis-NIR spectral reflectance acquired from soil cores to predict field pasture root density. This technique requires calibration of spectral reflectance against measured root densities in a sub-sample of the

observations. The prediction of root density within the larger sample population can be achieved by applying the calibration model to the spectral reflectance recorded from the remaining samples. Improved regression models for predicting field root density can be produced by selecting calibration data from the field data source with similar spectral attributes to the validation set. For example, a calibration model developed from spectral reflectance and root densities measured in glasshouse soils was not able to predict field pasture root density. Improved prediction of root density was achieved using the field calibration data and by building separate calibration models for each of two contrasting soil types, rather than treating spectral data from two soils as a single population. Root density was predicted independently from soil C, which suggests that soil spectral reflectance will be particularly useful in studies examining potential rates of soil organic matter synthesis via root production. Further research is required to improve the accuracy of this technique. Improvement may be possible by ensuring that the soil sample in the field of view of the spectrometer is the same sample of soil used for the reference measurement (root density or C and N concentration).

CHAPTER 5

Predicting Maize Root Density, Soil Carbon and Nitrogen from Soil Vis-NIR Spectral Reflectance

In this Chapter, the field technique for using NIRS to predict pasture root density (Chapter 4) is further tested for its ability to predict maize root density and soil carbon and nitrogen concentrations to 60 cm depth.

Results of this study have been orally presented at the Joint Conference of the Australia and New Zealand Societies of Soil Science in conjunction with the International Year of Planet Earth, 1-5 December 2008 at Massey University, Palmerston North, New Zealand. A paper summarizing the results of this study will be submitted to the scientific journal *Plant and Soil*, in 2009.

5.1. Introduction

The patterns of carbon (C) sequestration in soils correlate well to plant root density and turnover times (Rees *et al.* 2005). Deeper root systems have the potential to sequester carbon (Smith 2004) deeper in the soil profile, where root turnover times can be slower. Organic carbon stored deeper in soil can slow the return of CO_2 to the atmosphere. Soil organic carbon stored deeper than 0.2 m and below 2 m can have residence times of 1000-2000 and 9000-13 000 years, respectively (Follet *et al.* 2003). Moreover, plants with deeper denser root systems will be able to recapture nitrate that would have leached past shallower root systems (Bownman *et al.* 1998; Crush *et al.* 2007; Dunbabin *et al.* 2003). Deeper rooting plants have the potential to increase water use efficiency by optimising the use of subsoil water which can minimize irrigation frequency (Hedley and Yule 2008). Plants with deeper and denser root systems that have improved nutrient acquisition are required to drive the second “green revolution” in low fertility soils and help the farmer with little access to fertilizer (Lynch 2007). To exploit these opportunities crop cultivars that express deeper and denser rooting characteristics will need to be identified and their potential to achieve the afore mentioned benefits evaluated in field trials.

Soil spectral reflectance in the range 400-2500 nm (Visible near-infrared, Vis-NIR) contains complex information of soil chemical and physical properties. Using sophisticated software (multivariate analysis), the rich spectral data can be calibrated against soil properties measured using standard methods (Workman and Shenk 2004). Vis-NIR reflectance spectroscopy has been successfully used to predict soil C (Chang and Laird 2002; Kusumo *et al.* 2008a), ryegrass root density in a glasshouse experiment (Kusumo *et al.* 2009a) and in field trials (Kusumo *et al.* 2009b), despite variations in soil texture, structure (Udelhoven *et al.* 2003) and other chromophores (e.g. clay and non-clay contents, iron oxides, water, decomposed organic matter) (Baumgardner *et al.* 1985; Ben-Dor 2002). As far we are aware Vis-NIRS techniques have not been applied to measuring crop root density in arable soils. Current common methods of measuring root density are the profile wall (Böhm 1979) and soil corer methods (Escamilla *et al.* 1991). Both require separation of roots and soil, which is a tiresome procedure, followed by root length measurement by the line intersect method (Newman 1966).

The rich spectral data allows the reflectance (Vis-NIR) technique to simultaneously quantify several soil properties, such as total C, organic C, total N, CEC and moisture (Malley and Martin 2003). Soil constituents that have direct active interactions with light of near-infrared wavelengths, are primarily water, organic matter, hydrous iron and aluminium oxides, and clays (Ben-Dor 2002). These chromophores can be masked by other “intrusive” soil properties such as roots. Masking this change, changes the reflectance spectra of a soil sample, may therefore be correlated to the change in the intrusive property, such as root density. In a previous study, Kusumo et al. (2009b) demonstrated that variations pasture root density could be predicted independently from variations in soil C in a Allophanic soil, but not in a Fluvial Recent soil, because in Fluvial Recent soil, measured root densities were co-correlated with measured soil C values.

In this paper we examine whether soil spectral reflectance can be used in the field to predict changes in maize root density, and whether this change is predicted independently from changes in soil C and N concentrations.

5.2. Materials and Methods

5.2.1. Contact Probe Modification

A prototype of a soil probe was developed (Kusumo *et al.* 2008a), based on the plant contact probe supplied by Analytical Spectral Devices (ASD, Boulder, Colorado, USA) with an internal light source replaced by a stronger intensity of parabolic reflector halogen lamp (Welch Allyn^{tm} 4.5 watt). A round casing was developed to avoid ambient light and direct contact of the quartz probe window with the soil, and to provide a fixed distance (30.5 mm) between soil surface and the probe window. The soil core was rotated 360^{o}, to give a field of view of 561 mm^{2} (Figure 5.1b).

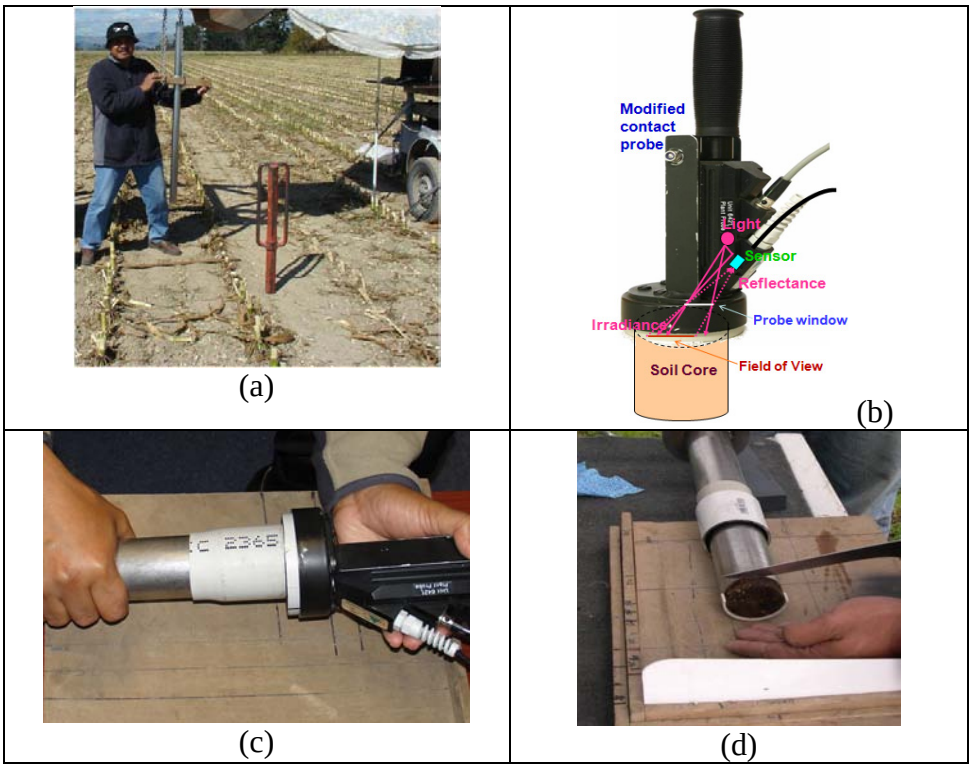


Figure 5.1 (a) Diffuse reflectance were collected from (a) soil core using (b, c) modified contact probe, then (d) the soil was sliced for laboratory analysis.

5.2.2. Site Locations, and Plant and Soil Property Measurement

On the day of harvest, soil cores (18 x 600 mm depth by 46 mm diameter) were collected from two sites within a field of 90-day-old maize (*Zea mays*, grown for silage) in Kairanga, Manawatu, New Zealand. At one site, the soil was dominantly silt loam and at the second site fine sandy loam (Gley Soils, Hewitt 1998). At each site, three replicate soil cores were taken at 0, 15 and 30 cm distance from the maize stem towards

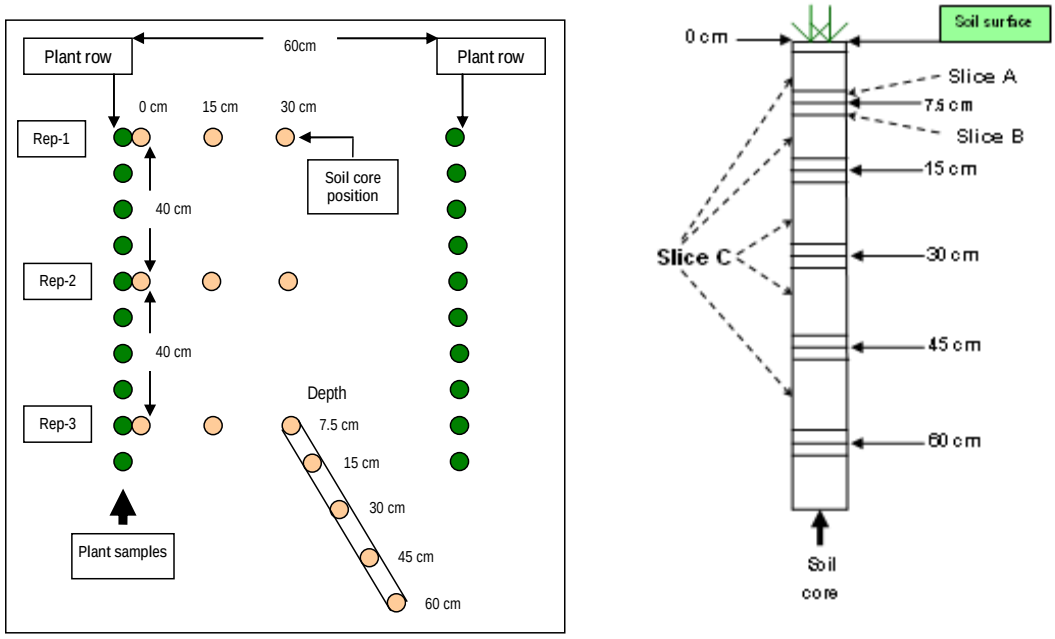
the centre of the 60 cm row (see Figure 5.2a). The distance between each replicate was 40 cm. A soil core was sectioned at 5 depths (7.5, 15, 30, 45, and 60 cm) and at each depth the soil reflectance spectra was acquired *in situ* from the freshly cut surface using a purpose built soil probe attached by fibre optic cable to a field spectroradiometer (ASD FieldSpecPro, Boulder, Colorado, USA). A 1.5 cm soil slice (slice A) was taken above this surface to obtain root mass reference data (using wet sieve laboratory root measurement). A total of 90 samples were taken (Figure 5.2b). Root was separated from soil in slice A by washing it with tap water on the following sieve sizes: 710 μm, 500 μm, 355 μm, 250 μm, and 63 μm diameter. The root retained on the sieves of 710 μm, 500 μm and 355 μm was bulked and dried at 50^{o}C oven for 3 days. Root density was presented in mg dry root / cm^{3}.

Another 1.5 cm soil slice (slice B) was for total soil C and N determinations (LECO FP-2000 CSN Analyser, combustion method) (Blakemore *et al.* 1987), and water content measurement (Figure 5.2b). Total C and N concentrations (LECO-C and LECO-N in figure legends) were measured using 2-mm sieved air-dry soils from 40 soil slices selected from replicates 1, 2 and 3 at 30 cm distance from the stem, and soil cores at replicate 2 at 15 cm distance from the stem. Water content was determined gravimetrically by drying to constant weight at 105^{o}C.

A soil slice C (adjacent to slice A; between each depth) (Figure 5.2b) was used to measure soil bulk density.

Chemical (pH H_2O with ratio of soil and water 1 : 2.5, P-Olsen, P-retention, CEC, SO_4, K, Ca, Mg, Na) and physical properties (soil texture) of the two soils were each determined using three soil samples, each a composite of three replicate soil cores taken from the 0-10 cm soil depth (Blackmore *et al*. 1987).

Biomass data of stem and cob were collected one day before harvesting and the fresh weight of each stem and cob was measured. Stem and cob dry matter values were obtained by drying them at 60^{o}C for 12 days.



(a) (b)

Figure 5.2 (a) Position of soil core and (b) soil slice.

5.2.3. Developing Calibration Model

The spectroradiometer records spectra from 350-2500 nm with 1-nm spaced data. An average spectrum for each soil sample was acquired, which was the average of 10 replicate scans. Then, the spectral data were pre-processed (SpectraProc V 1.1, (Hueni and Tuohy 2006) by first eliminating the noisy data (350-470 nm and 2440-2500 nm), smoothing using Savitzky-Golay filter with window size 33 nm and polynomial order 4, reducing the data by taking every 5^{th} waveband, transforming them into a first derivative and averaging the ten replicates recorded. The processed data were imported to Minitab 14 (MINITAB Inc. 2003) for partial least squares regression (PLSR) analysis. Calibration models were developed by PLSR-1 using processed spectral data and each of the reference data (root density, soil C and N). The PLSR calibration models were used to predict root density, soil C and N concentrations from unknown samples. The accuracy of the models was tested internally using leave-one-out cross-validation procedure. An external validation test was conducted only on root density prediction by allocating samples into separate calibration and validation sets with a 1:1 ratio. Separating these two sets was carried out by ranking the measured root density from the lowest to the highest root density, and odd and even ranked numbers were allocated to calibration and validation set, respectively. MINITAB automatically selected the optimum principal components (latent variables or factors) that produced

the lowest PRESS (predicted residual error sum of square) using a cross-validation procedure (Miller and Miller 2005). During PLSR processing, samples which had a standardized residual > 2.0 were removed as outliers from the calibration and validation set (MINITAB Inc. 2003).

5.2.4. Principal Component Analysis

Prior to PLSR analysis, principal component analysis (PCA) was carried out to describe the pattern of variability in the first derivative of the spectral data. A score plot of the first two components which account for the highest variance of spectral data was used to illustrate the pattern of sample scattering.

5.2.5. Regression Model Accuracy

The ability of the PLSR model to predict soil properties was assessed using the following statistics. RMSE (root mean square error) is the standard deviation of the difference between the measured and the predicted soil property values. RMSE which is calculated from cross-validation and separate validation set is called RMSECV and RMSEP, respectively. RPD (ratio of prediction to deviation) is the ratio of the standard deviation of the measured value of soil properties to the RMSE. RER (ratio error range) is the ratio of the range of measured values of soil properties to the RMSE. The best prediction model is shown by the largest RPD, RER, r^2 and the smallest error (RMSECV or RMSEP) (Kusumo *et al.* 2008a).

5.3. Results and Discussion

5.3.1. Summary of Soil Properties

The summary of laboratory analyses of root density, soil C and N concentrations and other chemical and physical properties are presented in Table 5.1. Higher root densities were found in the fine sandy loam. In contrast, total C and N concentrations are higher in the silt loam. Total C and N concentrations in these cultivated soils were lower compared to the same soil type profile under permanent pasture, where soil C and N concentrations ranged from a maximum of 5.20% (at topsoil) to a minimum of 0.50% (at 30 cm depth), respectively (Kusumo *et al.* 2008b). Cultivation often causes soil organic matter (SOM) decline (Shepherd *et al.* 2001); partly because aggregate

disruption allows microbial decay of previously aggregate-protected organic matter and annual C input to soil is lower with crops than pasture (Paustian *et al.* 1997; Tisdale and Oades 1982).

Table 5.1 Laboratory analysis of root density, soil C and N, and chemical and physical properties of soil samples (fine sandy loam and silt loam).

Soil Properties	Fine sandy loam (FSL)	Silt loam (SL)
Root density (mg/cm^3) of all 90 samples		
Minimum	0.4	0.3
Maximum	17.3	10.9
Median	1.5	0.9
Mean	4.5	2.4
Variance	26.6	6.6
Total Carbon (%) of 40 selected samples		
Minimum	0.07	0.29
Maximum	3.07	3.50
Median	0.87	1.75
Mean	1.31	1.94
Variance	1.28	1.85
Total Nitrogen (%) of 40 selected samples		
Minimum	0.01	0.04
Maximum	0.26	0.33
Median	0.08	0.19
Mean	0.11	0.19
Variance	0.008	0.015
Water content (%) of all 90 samples		
Minimum	8.1	15.6
Maximum	29.7	40.3
Median	15.5	27.3
Mean	16.3	26.4
Soil (0-15 cm)		
pH	4.7	5.2
P-Olsen (μgP/g)	22	34
P retention (%)	24	31
CEC (me/100g)	19	29
SO_4 (μgS/g)	47	134
K (me/100g)	0.26	0.29
Ca (me/100g)	7.5	14.9
Mg (me/100g)	0.82	2.83
Na (me/100g)	0.15	0.24
Bulk density (g/cm^3)	1.36	1.15
Texture	Fine sandy loam	Silt loam

5.3.2. Root Density and Maize Biomass in Each Soil Type

Root density and plant biomass of each soil type are depicted in Figure 5.3. Root density is higher in the top soil (15 cm depth) and tends to decrease with depth (Figure 5.3a). It also decreases with distance from the plant stem (in the silt loam soil) (Figure 5.3b). In the fine sandy loam, highest densities of root are found 15 cm distance from the plant stem (Figure 5.3b). The relationship between root density and soil depth and distance from stem is presented in Figure 5.3c and 5.3d.

Fresh and dry stem weights were higher in the fine sandy loam than that in the silt loam (Figure 5.3e). This trend was reflected in fresh and dry cob weights, but dry cob weights were not significantly different (Figure 5.3f). The higher root density in fine sandy loam (Figure 5.3a) may support greater absorption of nutrients and water allowing greater plant biomass to develop. In this fluvial landscape the fine sandy loams are associated with old levees of meandering rivers and streams (Cowie 1978) and the silt loams occur in the slightly lower (0.5-1 m) poorly drained areas. Poor plant growth and root growth on the silt loams is often associated with poor seedling establishment when the soils are at field capacity in spring. In addition lower clay content in the fine sandy loam may create lower soil strength (Mathers *et al.* 1966) with less mechanical impedance to root growth (Unger and Kaspar 1994).

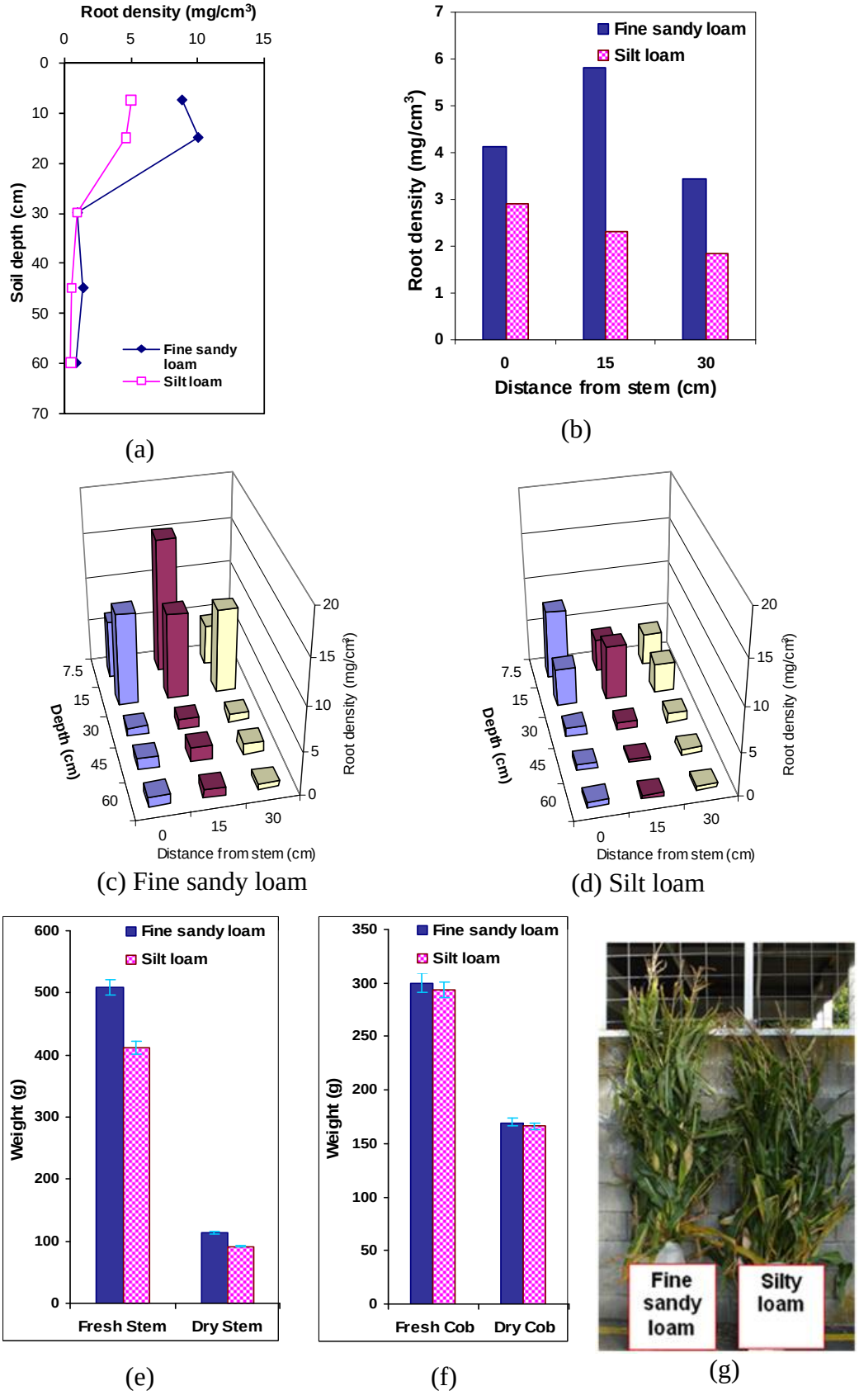
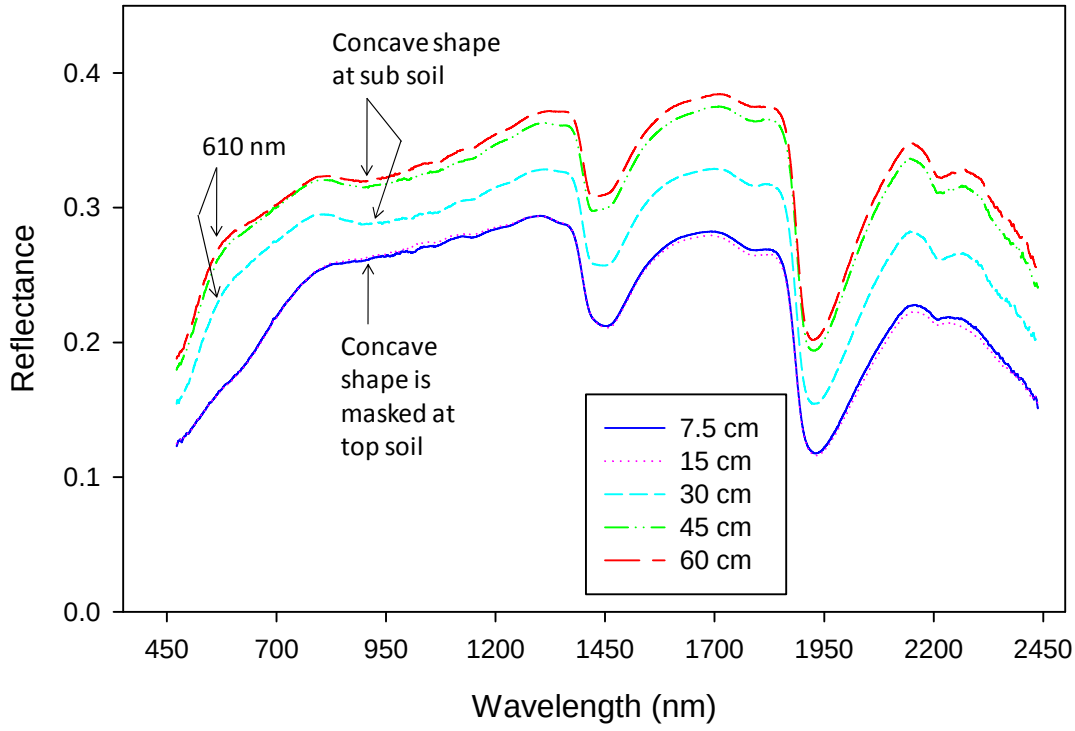


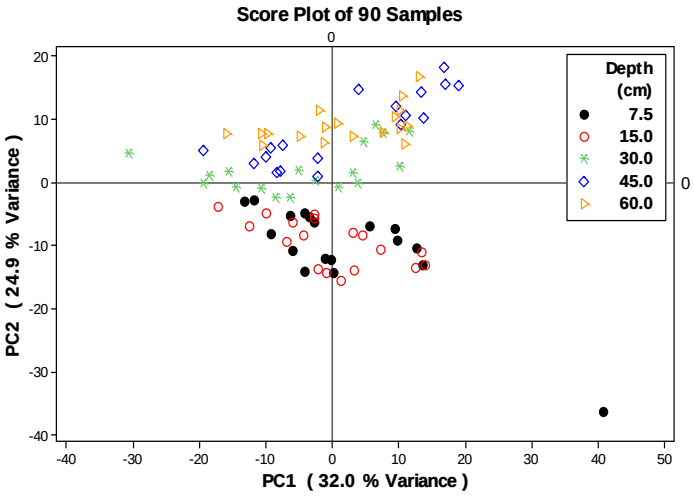
Figure 5.3 Root density and maize biomass in different soil types

5.3.3. Spectral Differences with Depths

The reflectance spectra acquired (same soil core) from the 7.5-60 cm depth of the fine sandy loam and the PCA score plot explaining the variation in first derivative of the spectra from all 90 samples are shown in Figure 5.4a and 5.4b, respectively. The spectra are not only influenced by root content and soil organic matter, but also by other soil chromophores such as clay and non-clay mineral content, water, iron, salt, soil texture, quality and quantity of organic matter and parent material (Baumgardner *et al.* 1985; Ben-Dor 2002). Spectral reflectance of subsoil samples (30, 40 and 60 cm depth, Figure 5.4a) obviously show a concave shape at the band around 900 nm, which is probably caused by absorbance due to iron oxides (Demattê *et al.* 2004a; Demattê and Garcia 1999). The presence of iron oxides is also shown by the increased reflectance at the visible band around 590-610 nm (yellow-orange) (Figure 5.4a) and the mottled nature of subsoil samples in Figure 5.4c. The spectra from the top soil samples (7.5 and 15 cm) show a less marked concave shape at band ~ 900 nm; where the iron oxides are probably masked by a high organic matter content (Demattê *et al.* 2004a) present as soil C and root.



(a)



(b)

(c)

Figure 5.4 (a) The reflectance spectra from different soil depths in a single core and (b) the PCA score plot illustrating the effect of soil depth on the variation in first derivative of the spectra from all 90 samples (c) the appearance of the fine sandy and silt loam soil.

Spectra from the top soil (7.5 and 15 cm depth) are clearly quite different from the spectra from the subsoil (Figure 5.4b). Topsoil samples from 7.5 cm and 15 cm depth occur and overlap at the same convex hull (Figure 5.4b) indicating that they may possess similar spectral characteristics. They also have similar laboratory-measured (LECO) soil C and N concentrations (Figure 5.6a and 6.6b). This spectral similarity

arises because samples from 7.5 and 15 cm (topsoil) are from the cultivation zone which could be considered to be well mixed. Although the subsoil samples were not cultivated, samples from each depth (30, 45 and 60 cm) are not clearly separated and some are overlapping, indicating quite homogenous subsoil with little vertical variation (Hewitt 1998). One sample at PC1 and PC2 co-ordinates of 41 and -37, which was expected to be an outlier, is actually the sample that has the highest root content.

5.3.4. Soil Properties Prediction from Spectral Reflectance

The predicted values for root density and soil C and N concentrations are presented in Table 5.2. Root density was predicted with moderate accuracy (RPD 1.75) (Malley *et al.* 2004) when all samples were included, without removing outliers (Table 5.2). However, improved prediction (RPD 2.42) was obtained by removing these outliers (Table 5.2). Malley *et al*. (2004) categorised prediction accuracy as moderately accurate (successful) when the RPD is between 2.25 and 3 (with RER 10-15 and r^2 0.80-0.90). Successful prediction of root density (r^2 0.90; RPD 2.68) was also obtained on the separate validation set (Table 5.2). As RER can be influenced by the value of a single sample with extreme concentration (Williams 2001), it has limited use as a meaningful tool of accuracy. Williams (2001) stated that r^2, bias and RPD are the most meaningful statistics for “instant” appraisal of the accuracy of a predictive model.

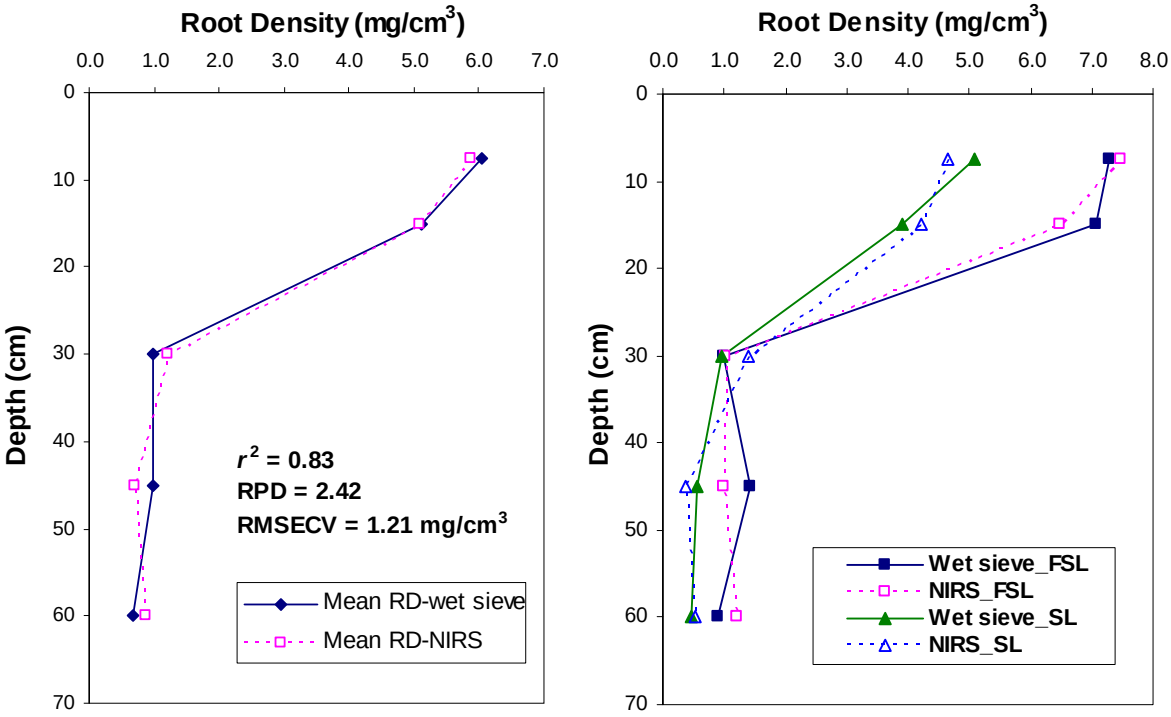
Table 5.2 Predicted values of root density, carbon and nitrogen.

Soil property	Allocated samples	n	Prediction values (leave-one-out cross-validation)						
			Comp	r^2	RMSECV	RPD	RER	Bias	Slope
Root density	90	90[#]	1	0.67	2.38	1.75	7.15	-0.002	0.66
Root density	90	83[##]	5	0.83	1.21	2.42	14.12	-0.009	0.87
Total C	40	38[##]	4	0.86	0.48	2.66	7.16	-0.0002	0.89
Total N	40	38[##]	3	0.81	0.05	2.32	6.55	-0.001	0.80
			Prediction values (separate validation set)						
			Comp	r^2	RMSEP	RPD	RER	Bias	Slope
Root density	83[##]	Cal:Val, 41: 42	4	0.90	1.21	2.68	14.05	-0.007	1.11

Note: n = number of samples; comp = components (factors or latent variables); [#] no outlier removal; [##] with removing outliers; root density in mg/cm^3; total C and N in %; cal = calibration set; val = validation set; Bias = mean difference between measured and predicted soil properties; Slope = regression coefficient between measured and predicted soil properties

For C and N estimation, the PLSR calibration models can predict these soil properties with moderate accuracy (Table 5.2).

Mean root density of each depth (excluding outliers) which was measured by wet sieving and predicted by NIRS is presented in Figure 5.5a. Despite the moderate accuracy of root density prediction, our NIRS technique facilitates a rapid and non-destructive root density measurement within the soil profile. This can be a useful technique in providing information about the potential role of deep root in synthesizing organic matter and sequestering atmospheric CO_2 (Rees *et al.* 2005).



(a) (b)

Figure 5.5 Measured and predicted root density with depth; (a) mean of both soil types, and (b) means within each soil type.

It is interesting to note that, the fine sandy loam soil has greater root content (Figure 5.5b) and lower C and N content, compared with the silt loam soil (Figure 5.6c and 5.6d). In this mixed permanent pasture/cereals farming system, the C and N content of the soils is only partially influenced by maize root turnover and the residual organic matter from long-term permanent pasture needs to be considered. The higher C and N content in the silt loam soil is probably due to the organic matter having greater aggregate-protection from decomposers with more C and N associated with clays (Theng *et al.* 1986) or within small pores (Elliot and Coleman 1988) and soil aggregates (Goldchin *et al.* 1994). In contrast, the greater aeration in sandy soil may accelerate organic matter mineralisation (Hassink 1992).

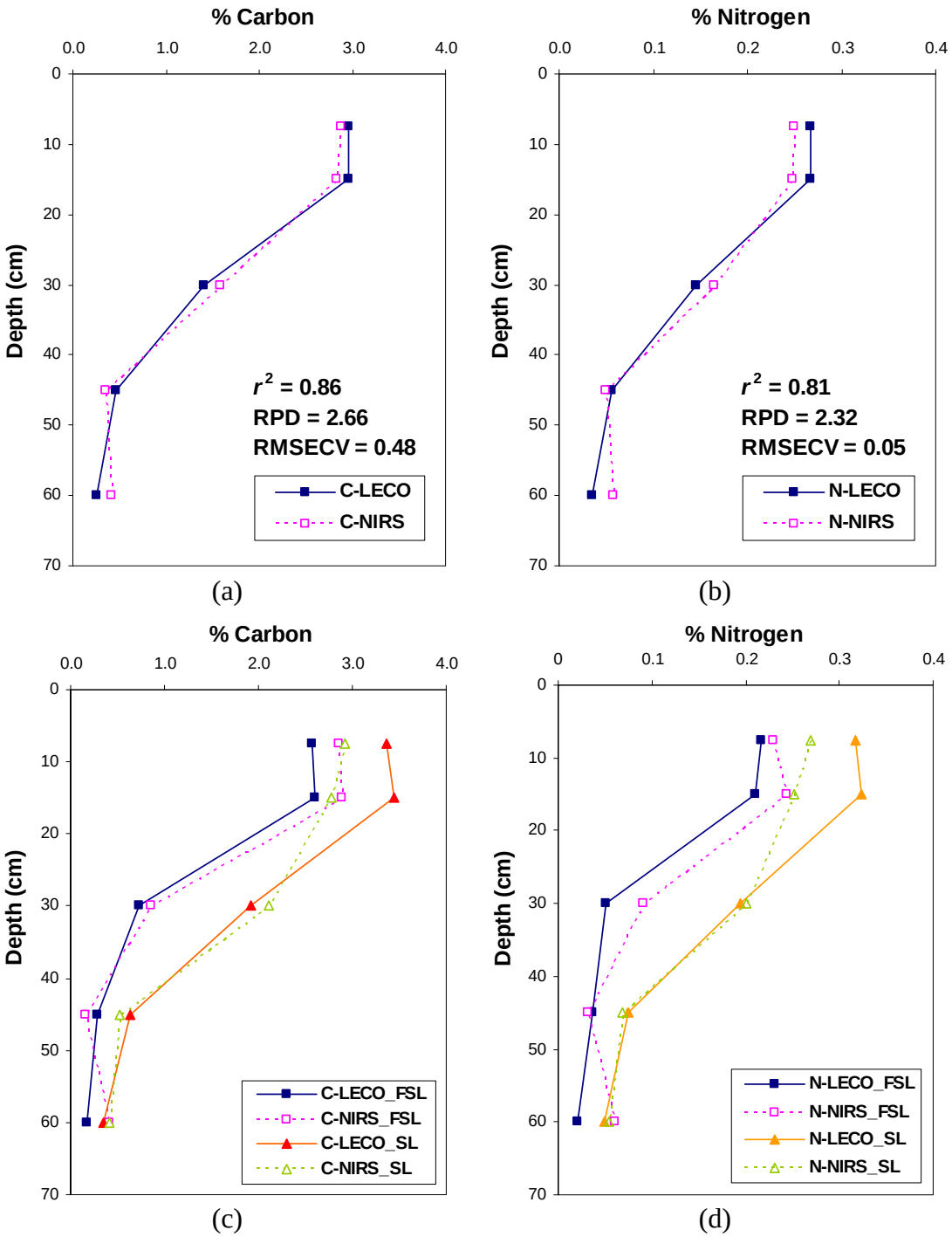


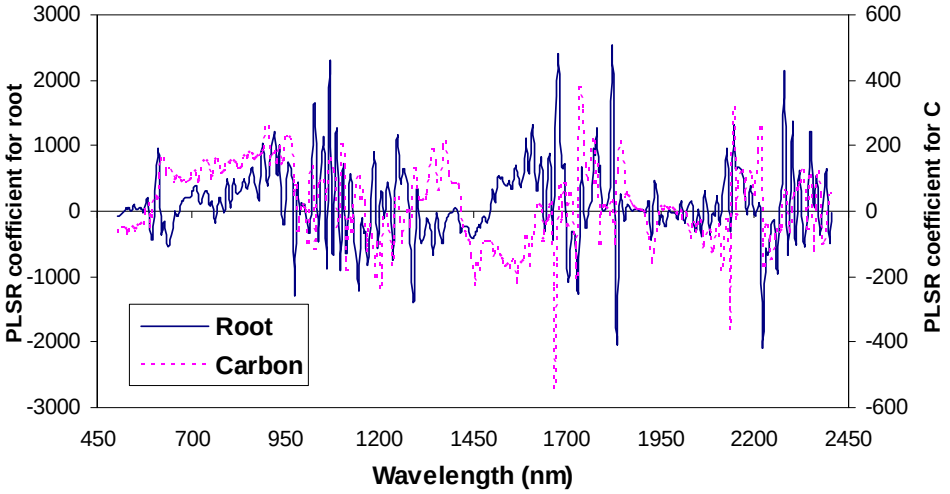
Figure 5.6 The variation in laboratory-measured (LECO) and NIRS-predicted C and N with soil depth.

5.3.5. Is Root Density Predicted Independently from Soil C (or N)?

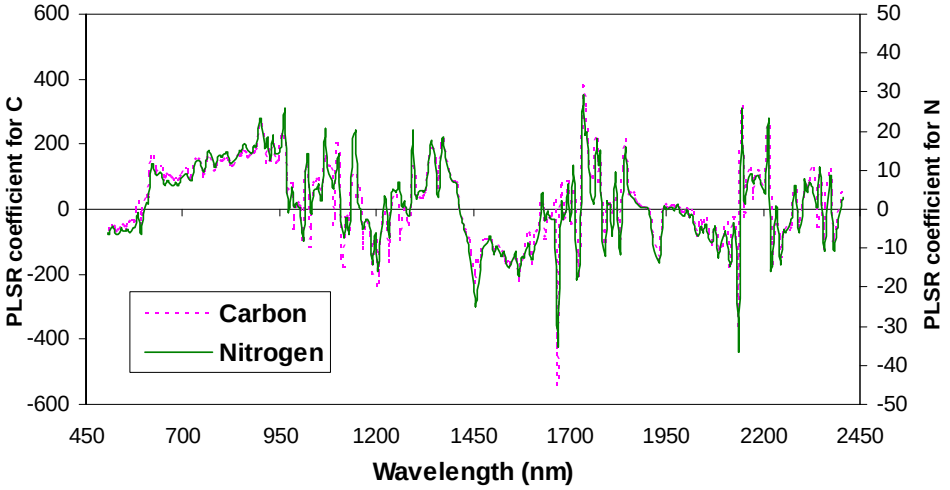
Estimation of soil C and N concentrations and root density by an indirect measure such as soil reflectance can be confounded if chromophores related to one property (root density) cause changes in spectral reflectance bands similar to another property (soil

organic matter content i.e. soil C). In such a case, increases in soil organic matter, without increases in root density, may lead to the over-prediction of root density from reflectance data. When Vis-NIR reflectance is to be used to predict soil properties the form of the regression model and co-correlation of reference and predicted soil properties should be examined. The coefficients for the PLSR model for predicting root density, and soil C and N from the first derivative of the reflectance spectra is presented graphically in Figure 5.7a and 5.7b. The size of the coefficient (positive or negative) represents the importance of a band (5 nm) in terms of explaining the variance observed in reference (measured) soil property. Similarly important bands for root density, C and N prediction occur in the visible region around 610 nm (Figure 5.7a and 5.7b). Williams (2001) reported that changes in reflectance at 614 nm were related to starch content.

Differences in the regression model coefficients for root density and soil C (or N) indicate that the important wavelengths selected for root density were different from those chosen for soil C (or N) prediction (Figure 5.7a and 5.7b). Good prediction accuracy of root density, C and N (Table 5.2) with r^2 between 0.81-0.86, but low correlation between reference (measured) root density and soil C (r = 0.60) or soil N (r 0.53) (Figure 5.8b and 5.8c) suggest that a significant proportion of chromophores useful in the prediction of root density are independent of change in soil C (or N). High accuracy of NIRS prediction of soil C but low autocorrelation between reference values of C and N, allowed Chang and Laird (2002) to conclude that they had achieved independent prediction of C and N concentration.

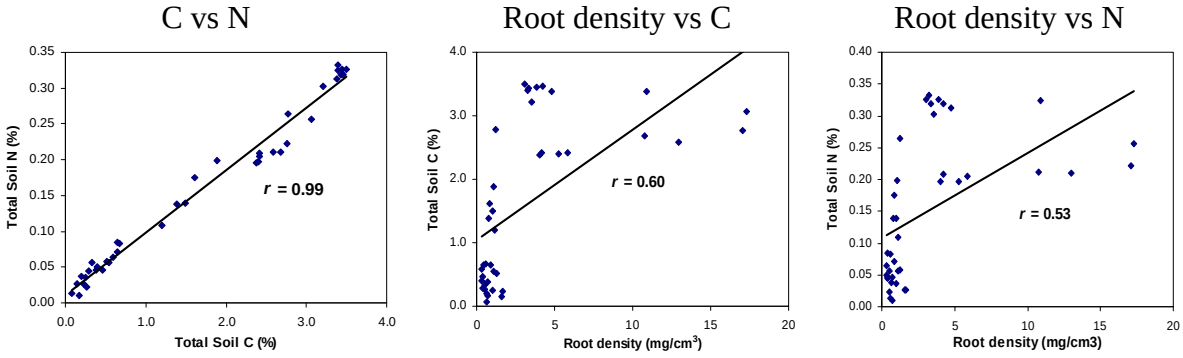


(a)



(b)

Figure 5.7 PLSR Coefficient of root density, C and N prediction.



(a) (b) (c)

Figure 5.8 Autocorrelation between laboratory-measured soil properties.

It is interesting to note that in our case with these fluvial recent soils, we are unable to prove independence for soil C and N prediction because the PLSR models select the same important wavebands for their prediction (Figure 5.7b), which is to be expected when the laboratory measured values are so highly correlated (Figure 5.8a). Others have found this high degree of correlation between soil C and N confounds the ability to distinguish independent C and N prediction by NIRS (Martin *et al.* 2002). For example, Chang e*t al*. (2001) successfully predicted total C (r^2 0.87; RPD 2.79) and total N (r^2 0.85; RPD 2.52), but total C and N were highly correlated (r 0.95). High correlation between C and N was also reported by other workers; r between organic C and total N was 0.94 (Malley *et al.* 2000); r between C and N was 0.966-0.989 (Reeves III and McCarty 2000). Martin *et al.* (2002) mentioned a low correlation (r 0.667) between C and organic N as the reason why they could predict organic C well (r^2 0.785 and 0.783; RPD 2.16) but not organic N (r^2 0.326 and 0.366; RPD 1.22 and 1.26). In these and other studies NIRS predicts C and N based on their autocorrelation. Chang and Laird (2002), however, were able to predict C and N, even though the correlation between C and N was low. In their case, the PLSR technique must have selected one, or a series of wavebands, that responded more independently to organic N compounds, although specific NIR absorption peaks of N containing functional groups were difficult to assign (Chang and Laird 2002). Martin *et al.* (2002) suggested that if the correlation of C and N is high, N may be predicted based on its correlation with C, but if the correlation is low, N should be predicted based on wavebands sensitive to N functional groups.

To avoid inaccurate prediction of a soil property that is strongly correlated with Vis-NIR reflectance spectra but "shares" important wavebands with another variable soil property (e.g. soil C and N content) the following procedural protocols should be adhered to. Firstly, the distribution of variances in the spectral data are examined in a PCA score plot (the first two PCs) to ensure that the spectral attributes of the calibration set [the samples taken for reference soil property measurement and used to build (calibrate) the PLSR model] embrace those of the population to be assigned for both predicted properties (e.g. soil C and N) (e.g. cf. Figure 5.4b and other examples Kusumo *et al.* 2008a, 2009a and Chapter 4). For further discussion of the use of PCA score plot for this purpose, see Chapter 6 and 8. Then check that the PLSR model accurately predicts the soil property concerned (e.g. soil C and N, Table 5.2). If two predicted soil properties are co-correlated and the important coefficients in the PLSR model have

similar waveband selections, then the reference data for the two soil properties must also be very highly correlated if unknown values are to be predicted accurately (Figure 5.8a). Unknown values should only be predicted from population samples that fall within the range of spectral attributes of the calibration set.

5.4. Conclusions

This study indicates that rapid field assessment of maize root density, soil C and N concentrations using Vis-NIR diffuse spectral reflectance has considerable potential. This technique facilitates *in situ* measurement of these soil properties to depth in the soil profile. In this study, the PLSR regression models for predicting root density and soil C (or N) from Vis-NIR reflectance spectra were independent. PLSR regression models for predicting soil C and soil N from Vis-NIR reflectance spectra were not independent. Strong co-correlation of the measured soil C and N concentrations in the calibration data allowed accurate prediction of the soil C and N concentrations in the unknown population (the validation set).

The independent prediction of root density from soil C (organic matter) requires testing in other soils, particularly long-term arable soils in which the organic matter distribution in the soil profile may be co-correlated to crop root density.

CHAPTER 6

The Use of Diffuse Reflectance Spectroscopy for *in situ* Carbon and Nitrogen Analysis of Pastoral Soils

Success of soil C and N prediction using "Vis-NIRS-soil core" technique in Fluvial Recent soil (Chapter 5) needed further evaluation in other soil types. This chapter describes the study of soil C and N prediction using this technique on a range of pasture soils, recently converted from forest, on Allophanic, Pumice and Thepric Recent soils in the central North Island New Zealand. Much of the experimentation in this chapter is focused on methods for selecting spectral data suitable for building partial least squares regression calibration models for improved prediction of soil properties.

A paper from this study has been published; B.H. Kusumo, C.B. Hedley, M.J. Hedley, A. Hueni, M.P. Tuohy and G.C. Arnold. 2008. The use of diffuse reflectance spectroscopy for in situ carbon and nitrogen analysis of pastoral soils. *Australian Journal of Soil Research*, 46:623-635.

6.1. Introduction

Global warming as a consequence of rising atmospheric CO_2 concentrations has focussed the attention of researchers on the dynamics of soil organic carbon (SOC) (Post *et al.* 2001). Global soil carbon (C) stocks are estimated to be 1500 Pg C which is approximately three times the stocks in standing global vegetation (Lal 2003; Post *et al.* 2001). Since the Industrial Revolution, it is estimated that 136±55 Pg C have been lost from soil via respiration of soil organisms, leaching, and soil erosion (Lal 2003). Accelerated respiration resulting from land-use change probably accounts for the largest loss (Rees *et al.* 2005). Soil C turnover time is sensitive to site and climate variables, land-use change, soil cultivation (Tan *et al.* 2004) and soil and vegetation management (Rees *et al.* 2005). To reduce atmospheric CO_2 concentrations, changes in land management practices are proposed as a way of increasing soil C sequestration rates. To evaluate the effectiveness of these C sequestration strategies rapid methods of

measuring the spatial and temporal variability in soil C stocks are required (Rees *et al.* 2005) for field survey.

Visible near-infrared (Vis-NIR) reflectance spectroscopy has been extensively used for soil C and N measurement. It is a rapid and non-destructive method. Most published reports of NIRS-predicted C and N are of laboratory methods using homogenous samples prepared by drying, grinding and sieving soils to eliminate the effect of water and heterogeneous surface structure on sample reflectance (Malley *et al.* 2004). Some researchers, however, have found it unnecessary to prepare dried and homogenised samples to obtain useful predictions of soil C and N content from reflectance spectra (Fystro 2002). There is a significant history of development of laboratory methods using near infrared reflectance spectroscopy (NIRS) to observe change in soil organic matter (Bowers and Hanks 1965). NIRS was then extended to predict other soil chemical and physical properties such as total C, organic C, total N, NH_4-N, NO_3-N, CEC, P, K, Ca, Mg, pH, moisture, and soil particle size (Ben-Dor and Banin 1995; Chang *et al.* 2001; Cozzolino and Moron 2006; Islam *et al.* 2003; Reeves III and McCarty 2001; Shepherd and Walsh 2002). Despite prediction accuracies for a wide range of soil properties ranging from poor to excellent (Viscarra Rossel *et al.* 2006), Malley and Martin (2003) stated that NIRS is mostly successful for predicting total C, organic C, total N, CEC and moisture. Prediction accuracy depends upon sample preparation and spectral pre-treatment selected before running a multivariate statistical analysis (Chang and Laird 2002). The ability of multivariate regression analysis to accurately predict the chemical properties of unknown soils (validation sets) relies heavily on whether the spectral and chemical properties of the unknown soil fall within the range of the calibration set (Brimmer *et al.* 2001; Williams 2001). In the development of calibration models, it is common that the chemical properties of soils used for validation are known. In this case researchers allocate soils with the same range of chemical properties to the calibration and validation sets (Martin *et al.* 2002). In the practical use, however, only spectral data and location of a soil sample may be known, thus location or spectral analysis will be required to select appropriate calibration data for the unknown sample. In this paper as part of the development of measurement methods we evaluate a number of techniques for allocating data to calibration sets.

There are only a limited number of publications related to *in situ* soil property prediction using NIRS, and only a few of these are mobile (on-the-go) methods (Mouazen *et al.* 2005; Shibusawa *et al.* 2005; Shonk *et al.* 1991; Stenberg *et al.* 2007; Sudduth and Hummel 1993). These authors outline several difficulties associated with field assessments. For example, much of the delicate and expensive instrumentation is not suitable for working in the field. In addition, the soil surface presented to the spectroradiometer may contain gravel, stones, plant debris (Mouazen *et al.* 2007) or various sized soil aggregates (Udelhoven *et al.* 2003) which will likely reduce the prediction accuracy of the reflectance technique. In laboratory methods the distance between sample and sensor is kept constant (Shepherd and Walsh 2002) to minimise variation in reflectance. Inaccuracy during on-the-go field measurements can be caused by variable distance between sample and sensor head (Shibusawa *et al.* 2005; Stenberg *et al.* 2007; Sudduth and Hummel 1993).

Current NIRS applications for 'on the go' *in situ* field measurement were developed for cultivated and pasture soils (Mouazen *et al.* 2007; Shibusawa *et al.* 2005). These methods involve the use of tillage or subsoiler equipment, which may not always be suitable for use in pastoral or forest soils if damage to the sward and trees is to be avoided. In this paper we investigate the potential use of a diffuse reflectance (range 350 - 2500 nm), acquired from the flat sectioned horizontal soil surface of a soil core, for *in situ* C and N analysis of pastoral soils.

6.2. Materials and Methods

6.2.1. Spectrophotometer and Optical Probe

A prototype of a soil probe was developed, based on the commercially available plant contact probe supplied by Analytical Spectral Devices, USA. A high-intensity light source (4.5 watt) of halogen lamp was fitted to the probe. To avoid direct contact of the quartz probe window with the soil, a round casing was developed to provide a fixed distance between the soil surface and the probe window. This modification also excludes the ambient light (see Figure 6.1). The object/probe window distance could be easily adjusted by changing a plastic (PVC) spacer ring (inner diameter = 75 mm) that fitted into the casing. For this study a distance of 30.5 mm between the soil surface and probe window was used.

6.2.2. Soil Samples and Site Locations

Soil diffuse reflectance spectra were acquired *in situ* at two times (May and October 2006) from 7 pasture field sites under permanent pasture and from 1-year, 3-years and 5-years pine-to-pasture conversions from the Taupo - Rotorua Volcanic Region of New Zealand. The first three sites, near Atiamuri, were 1-year conversion (38^o 19.59 S, 176^o 2.48 E), 5-year conversion (38^o 20.05 S, 176^o 3.22 E) and permanent pasture (38^o 19.43 S, 176^o 2.84 E) sites, mapped as Taupo sandy silts (Vucetich and Wells 1978) and classified as Pumice Soils (Hewitt 1998). Two sites, 3-years conversion (38^o 8.89 S, 175^o 46.75 E) and permanent pasture (38^o 10.71 S, 175^o 49.65 E), around Lichfield, Tokoroa, are classified as Allophanic Soil (Hewitt 1998). The last two sites, 5-years conversion (38^o 0.14 S, 176^o 41.56 E) and permanent pasture (37^o 59.75 S, 176^o 4.14 E) at Manawahe, are classified as Tephric Recent Soil (Hewitt 1998). Thirty soil samples were taken at each site, from 5 positions at 15-m intervals along each of three 75-m transects that were 20 m apart, and at 2 depths. Soil cores were collected with a corer (46 mm diameter), extruded and sliced (15 mm) at 37.5 and 112.5 mm. The diffuse reflectance of the sliced surface was recorded and then a 15-mm soil slice was taken for further laboratory measurements (Figure 6.1).

6.2.3. Measurement of Soil Properties

Soil water content was determined by measuring the difference between the weight of field-moist soil and air-dry soil (dried at 35^oC). Air dry soils were then divided equally into two subsamples. The first subsample was ground into a fine powder using a ring grinder for total C and N analysis on a 1-g sample using a LECO FP-2000 CNS Analyser dry combustion method (Blakemore *et al.* 1987). The second subsample was kept for further analysis.

6.2.4. Reflectance Measurement and Spectral Pre-Processing

The diffuse reflectance spectra of each freshly cut surface were recorded from a flat sectioned horizontal soil surface of a soil core (46 mm diameter) using the purpose-built contact probe attached by fibre optic cable to an ASD FieldSpecPro spectroradiometer (Analytical Spectral Devices) (Figure 6.1). The instrument records spectra with sampling interval of 1.4 nm for the region 350-1000 nm and 2 nm for the region 1000-2500 nm. The data processing software associated with the ASD

FieldSpecPro spectroradiometer then interpolates this 1.4- and 2-nm-spaced data to produce 1-nm-spaced data. During reflectance recording, the soil core was rotated 360^{o} with the speed of 18^{o}/s in order to scan the whole target area. Thus the scanned target area summed up to 561 mm^{2}. The instrument was set to average ten readings internally for each spectrum saved. In this manner ten averaged spectra were collected per soil sample. A Spectralon® reference panel was used as white reference to calibrate the equipment between each soil sample recording.

Figure 6.1 Using the soil probe to make reflectance measurement on sectioned soil cores.

Wavebands exhibiting obvious noise (350-470 nm and 2440-2500 nm) were excluded from further processing. A Savitzky-Golay filter with window sizes of 31 and polynomial orders of 4 was applied to the spectra, then the data was reduced to 381 spectral bands by down-sampling to every 5th waveband. The first derivative was calculated using the Savitzky-Golay algorithm with window sizes of 7 and polynomial orders of 5. These derivatives were then averaged for every soil sample. All spectral processing steps were carried out in SpectraProc which is software created for spectral pre-processing, discriminant analysis, PCA and database structure (Hueni and Tuohy 2006).

6.2.5. Development of Calibration Models

Calibration models for the relationship between the spectral first derivative and total soil C and total soil N values were developed using partial least squares regression

(PLSR) in MINITAB 14 (MINITAB Inc. 2003). Using PLSR-1, the chemical analysis data (single Y-variable) (total C or N) was regressed against the pre-processed soil reflectance data. The number of factors (components or latent variables) used in the calibration model were the number that minimised the PRESS (predicted residual error sum of square) in the leave-one- out cross validation procedure (Miller and Miller 2005). The number of components obtained was between 3 and 8 (see Table 6.3 and 6.4). Calibration models were then used to predict the C and N concentrations of a separate population of soil samples (the validation set).

To provide a large pool of data from which to draw the calibration and validation datasets, soil spectral and reference data from May and October (210 samples) were amalgamated. A preliminary PLSR analysis of the laboratory-measured C or N against the spectral data using all the samples identified 10 (C values) and 11 (N values) samples that had cross-validated standardised residuals outside ±2 (MINITAB Inc. 2003). Although this is about the number that should be expected from normally distributed data, they were removed from subsequent analysis to avoid the comparison between methods being influenced by a few extreme values.

The remaining samples were separated into calibration (A) and validation (B) sets using four methods. These were:

(I) Prior knowledge of pedological range. Near neighbour sample pairs (soil type, location, transect, site and depth) were alternately allocated to set A and B.

(II) Prior knowledge of sample chemical analysis range. All soil samples were ranked from the lowest to the highest C (or N) content and odd and even ranked numbers were allocated to set A and B, respectively.

(III) No prior knowledge of samples, Euclidean distance analysis, whole pool. A principal component analysis (PCA) of the pre-processed spectral data of total data pool was conducted and the scores of the first two PCs were plotted against each other. Standardised Euclidean distances of all sample points from the centroid were calculated (Miller and Miller 2005) and the samples were ranked from the lowest to the highest standardised Euclidean

distance. Odd and even ranked numbers were allocated to set A and B, respectively.

(IV) No prior knowledge of samples, Euclidean distance analysis by quadrant. PC1 and PC2 scores were plotted and each sample point was ranked by standardised Euclidean distance within its quadrant. Odd and even ranked numbers from all quadrants were allocated to set A and B, respectively.

For all selection methods, the stability of the prediction model was evaluated by also using set B as the calibration set and set A as the validation set, a "vice versa test". Selection methods of II and III were further evaluated by varying the ratio of samples allocated to the calibration and validation sets by 2:1 and 3:1.

6.2.6. Regression Model Accuracy

The ability of the PLSR model to predict soil properties was assessed using the following statistics: RMSE (root mean square error), which is the standard deviation of the difference between the measured and the predicted values of soil properties; RMSEP (root mean square error of prediction) is calculated from the validation data (Esbensen *et al.* 2006):

$$\text{RMSEP} = \sqrt{\frac{\sum (y_m - y_v)^2}{N}}$$

where y_m is the measured laboratory value, and y_v is the predicted value from the PLSR model, and N is the number of samples; r^2, which is the proportion of variance in y_m accounted for by the PLSR model predicted values (y_v); RPD (ratio of prediction to deviation), which is the ratio of the standard deviation of measured values of soil properties to the RMSE (this shows how much more accurate, as measured by the standard error, a prediction from the model is than simply quoting the overall mean); RER (ratio error range), which is the ratio of the range of measured values of soil properties to the RMSE:

$$\text{RPD} = \frac{\text{STDEV}(y_m)}{\text{RMSEP}} \quad , \quad \text{RER} = \frac{\text{Max}(y_m) - \text{Min}(y_m)}{\text{RMSEP}}$$

The best prediction model is shown by the highest RPD, RER, r^2 and the lowest RMSEP.

6.3. Results and Discussion

6.3.1. Summary of Soil Properties

The results of laboratory analysis of soil properties (excluding outliers) are presented in Table 6.1. The wide range of measured soil total C and N contents is consistent with the wide range of organic matter contents that have been reported for New Zealand Pumice Soils in previous studies (Jackman 1960). Organic matter sourced from decaying roots and litter can accumulate rapidly in raw Pumice Soils because of organic matter sorption by non-crystalline minerals (allophane, imogolite, and ferrihydrite) especially allophane and metal-humus complexes, which slows organic matter decomposition (Boudot *et al.* 1988). It is interesting to note that C : N ratios of the samples are noticeably high (range 9-30 with the average of 15). These values are higher than would be expected for well drained soils and perhaps indicate the reduced rate of decomposition of organic matter (Daly and Rijkse 1974), or the presence of residual forest trash in the areas recently (1-5 years) converted to pasture. Gravimetric water content of the samples varies between 11 and 82%.

Table1 6.1 Soil properties (excluding outliers).

Soil Properties	Range		Median	Average	Variance
	Minimum	Maximum			
Total Carbon (%)	0.255	11.210	4.301	4.423	5.9282
Total Nitrogen (%)	0.018	1.012	0.277	0.316	0.0414

6.3.2. Explanation of Soil Reflectance Spectra Shape

Reflectance and the first derivative of 1-nm spaced smoothed spectra from soil samples selected for low and high C content are illustrated in Figure 6.2. Samples with low total C (number 1, 2 and 6) represent samples from the 112.5 mm depth and show a concave shape (reflectance) and negative first derivative at bands around 800-970 nm. The concavity is more obvious at sample number 1 and 2, perhaps due to increased crystalline iron content (Demattê and Garcia 1999). Interestingly, the presence of high

organic matter in the surface layer masked the effect of iron (Demattê *et al.* 2004a) and the concavity was unclear or disappeared. This occurs with sample number 3, 4, and 5 (Figure 6.2), which represent samples from 37.5 mm depth containing total C of 10.12%, 11.21% and 7.31% respectively. Their spectrum shapes resemble spectra of minimally decomposed organic soil (fibric) as reported by Stoner and Baumgardner (1981).

Strong absorption at bands around 1465 and 1945 nm are due to water absorption (Figure 6.2). First overtones of the O-H bond stretching occur at about 1400 nm, and the combination of the H-O-H bend with O-H stretching occurs near 1900 nm (Clark 1999). Water and the OH^- ion will cause decreased reflectance at 1400 nm (Clark 1999). So a mineral with absorption at 1900 nm contains water, but a mineral with absorption at the 1400 nm and none at the 1900 nm absorption bands indicates that only hydroxyl is present (Baumgardner *et al.* 1985; Clark 1999).

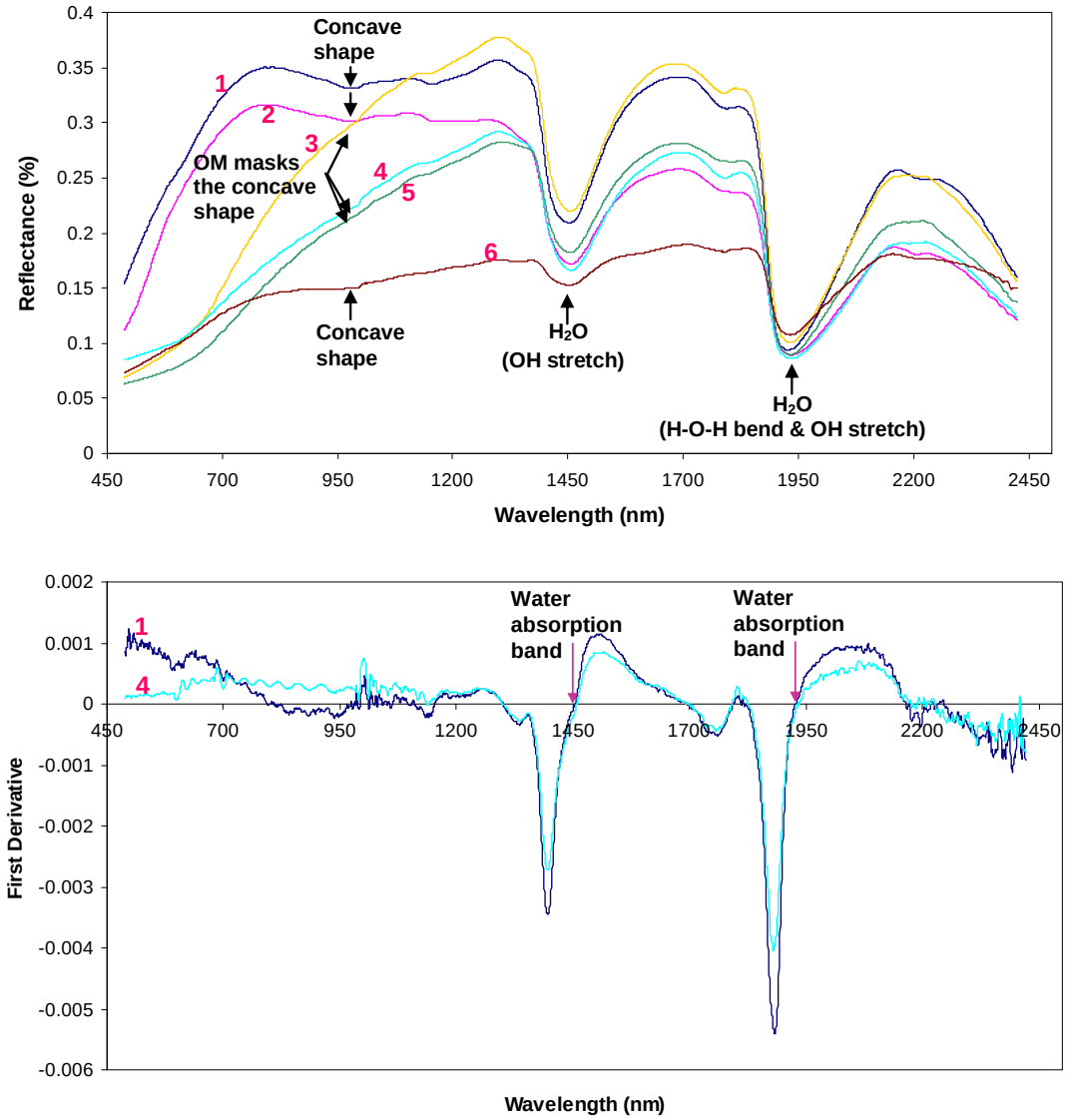


Figure 6.2 Reflectance spectra (top) and first derivative (bottom) from selected soil samples with different carbon and water contents; (1) C 0.26%, WC 38%; (2) C 1.41%, WC 68%; (3) C 10.12%, WC 56%; (4) C 11.21%, WC 78%; (5) C 7.31%, WC 36%; (6) C 0.82%, WC 11%.

6.3.3. Selection of Spectral and Chemical Data for Calibrating Regression Models to Predict Soil C and N

The ability of calibration sets to accurately predict the chemical properties of unknown soils (validation sets) relies heavily on whether the spectral and chemical properties of the unknown soil fall within the range of the calibration set (Brimmer *et al.* 2001; Esbensen *et al.* 2006; Williams 2001). A score plot of the first two principal components (PC1 and PC2, Figure 6.4), which explain most (68%) of the variation in the spectral data, allows the researcher to visualise the distribution of the whole data set

(Figure 6.4) and influence of each spectral data sample (Esbensen *et al.* 2006; Miller and Miller 2005). Positive scores of both PC1 and PC2 lie in quadrant 1 (top right), positive scores of PC1 and negative scores of PC2 lie in quadrant 2, etc. To examine the effect of selecting calibration data that do not represent the spectral attributes of the values to be predicted (validation set), spectral and chemical reference data were used from 3 of the 4 quadrants to construct PLSR calibrations for predicting the soil C and N values for the spectral data in the remaining quadrant (see Table 6.2 and Figure 6.3). The prediction values for C are poor (Table 6.2); all RPDs are lower than 1.5 and coefficient determinations (r^2) lower than 0.6. The best accuracy of prediction values for N (RPD 1.98 and 1.80; r^2 0.74 and 0.70) also do not reach the desired level of quantitative accuracy [r^2 > 0.8 and RPD >2 (Chang *et al.* 2001; Dunn *et al.* 2002)].

Table 6.2 Prediction accuracy of soil C and N values when the calibration data for the PLSR model (spectral plus measured soil C and N) are derived from different quadrants in Figure 6.3 to those being predicted (validation set).

Calibration set		Validation set		Prediction values for C			
Quadrant	n	Quadrant	n	r^2	RMSEP	RPD	RER
2, 3, 4	151	1	49	0.57	1.17 %	1.06	4.37
1, 3, 4	146	2	54	0.55	1.53 %	1.45	6.29
1, 2, 4	149	3	51	0.27	1.74 %	1.14	5.03
1, 2, 3	154	4	46	0.60	1.00 %	1.47	7.21
				Prediction values for N			
2, 3, 4	150	1	49	0.68	0.12 %	0.72	2.93
1, 3, 4	147	2	52	0.74	0.11 %	1.98	8.19
1, 2, 4	148	3	51	0.70	0.10 %	1.80	8.20
1, 2, 3	152	4	47	0.56	0.08 %	1.48	6.30

Note: Quadrant 1 (positive scores of PC1, positive scores of PC2); Quadrant 2 (positive scores of PC1, negative scores of PC2); Quadrant 3 (negative scores of PC1, negative scores of PC2); Quadrant 4 (negative scores of PC1, positive scores of PC2); n = number of samples

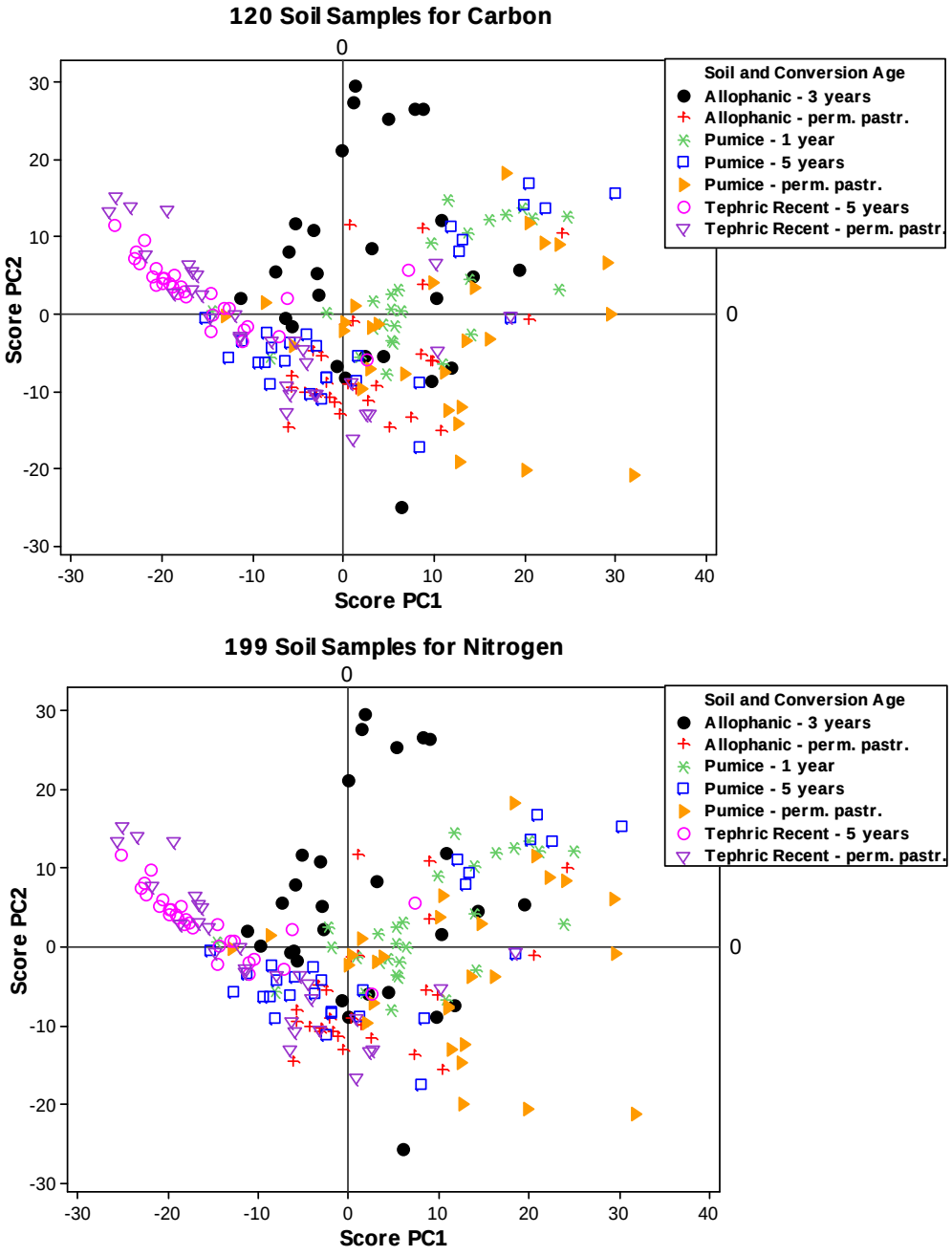


Figure 6.3 PCA score plot of soil samples from 7 sites based on soil type and conversion age for carbon (above) and nitrogen (below).

In the next section we examine 4 techniques that can be used to select data for the calibration of the PLSR prediction models (calibration set) that represent the range of the spectral and chemical properties of the values to be predicted (the validation set). With each selection method the distribution of spectral data in the calibration and validation sets is visualized using the PCA score plots (Figures 6.4).

The accuracy of predicting soil C and N using the different methods of selection for the calibration and validation sets are presented in Table 6.3. The ratio of allocating data to the calibration or validation set was 1:1. The accuracy of prediction was tested firstly using set A as a calibration set and set B as a validation set, and then vice versa. Using this technique, both sample groups have the opportunity to build the regression model and to test the accuracy and stability of the model. In general, the prediction quality for total soil N is better than total soil C, as shown by lower RMSEP and higher coefficient of determination (r^2), RPD and RER.

The first selection method (I) allocates samples of similar soil type and location (including transect, site and depth) equally to set A and B. This method produced a relatively stable regression model and prediction accuracy for values of r^2, RMSEP, RPD and RER (Table 6.3) in predicting N (compared to C) when the model is tested using a "vice versa test" (A to B and B to A test). The slightly improved prediction of soil C (RPD 2.01, selection method I, Table 6.3) when the B set is used for calibration rather than set A (RPD 1.78) is associated with the B data set having better coverage of the A data set (Figure 6.4a). For prediction of soil N, both datasets seem to have similar coverage (Figure 6.4b) and A or B sample set produces good calibration models (RPD's 2.31 – 2.54, with $r^2 > 0.81$), but set A results in slightly better N prediction (RPD 2.51 with r^2 0.86).

Table 6.3 Prediction accuracy of soil C and N values from PLSR calibration models using different methods (I to IV) to select the calibration and validation sets (ratio 1:1).

Selection method	Statistical parameter	Total C		Total N	
		A to B	B to A	A to B	B to A
I	n in set A	100	100	100	100
	n in set B	100	100	99	99
	Variance	A: 5.914	B: 5.889	A: 0.045	B: 0.038
	% range of recovery	A: 100	B: 85.5	A: 95.8	B: 80.4
	Component	7	3	5	6
	r^2-validation	0.69	0.75	0.86	0.81
	RMSEP (%)	1.36	1.21	0.08	0.09
	RPD	1.78	2.01	2.54	2.31
	RER	6.88	9.07	10.44	10.43
	Slope	0.78	0.79	0.95	0.80
II	n in set A	100	100	100	100
	n in set B	100	100	99	99
	Variance	A: 5.932	B: 5.983	A: 0.043	B: 0.040
	% range of recovery	A: 96.9	B: 99.0	A: 100	B: 90.0
	Component	4	5	8	6
	r^2-validation	0.73	0.74	0.82	0.84
	RMSEP (%)	1.28	1.27	0.09	0.08
	RPD	1.92	1.91	2.20	2.45
	RER	8.50	8.34	9.82	11.73
	Slope	0.69	0.82	0.93	0.86
III	n in set A	100	100	100	100
	n in set B	100	100	99	99
	Variance	A: 5.256	B: 6.397	A: 0.048	B: 0.035
	% range of recovery	A: 96.9	B: 99.0	A: 100	B: 90.0
	Component	5	5	7	6
	r^2-validation	0.69	0.76	0.82	0.79
	RMSEP (%)	1.40	1.16	0.08	0.10
	RPD	1.80	1.97	2.28	2.13
	RER	7.61	9.12	10.27	9.64
	Slope	0.73	0.87	0.92	0.72
IV	n set A	100	100	100	100
	n set B	100	100	99	99
	Variance	A: 5.391	B: 6.491	A: 0.039	B: 0.045
	% range of recovery	A: 96.9	B: 96.5	A: 90.5	B: 98.7
	Component	5	8	7	7
	r^2-validation	0.67	0.74	0.76	0.82
	RMSEP (%)	1.47	1.32	0.11	0.09
	RPD	1.74	1.76	1.99	2.27
	RER	7.21	8.04	9.23	10.39
	Slope	0.66	0.93	0.75	0.94

Note: n = Number of samples; A = Sample Set A; B = Sample Set B; Slope = regression coefficient between measured and predicted soil C and N

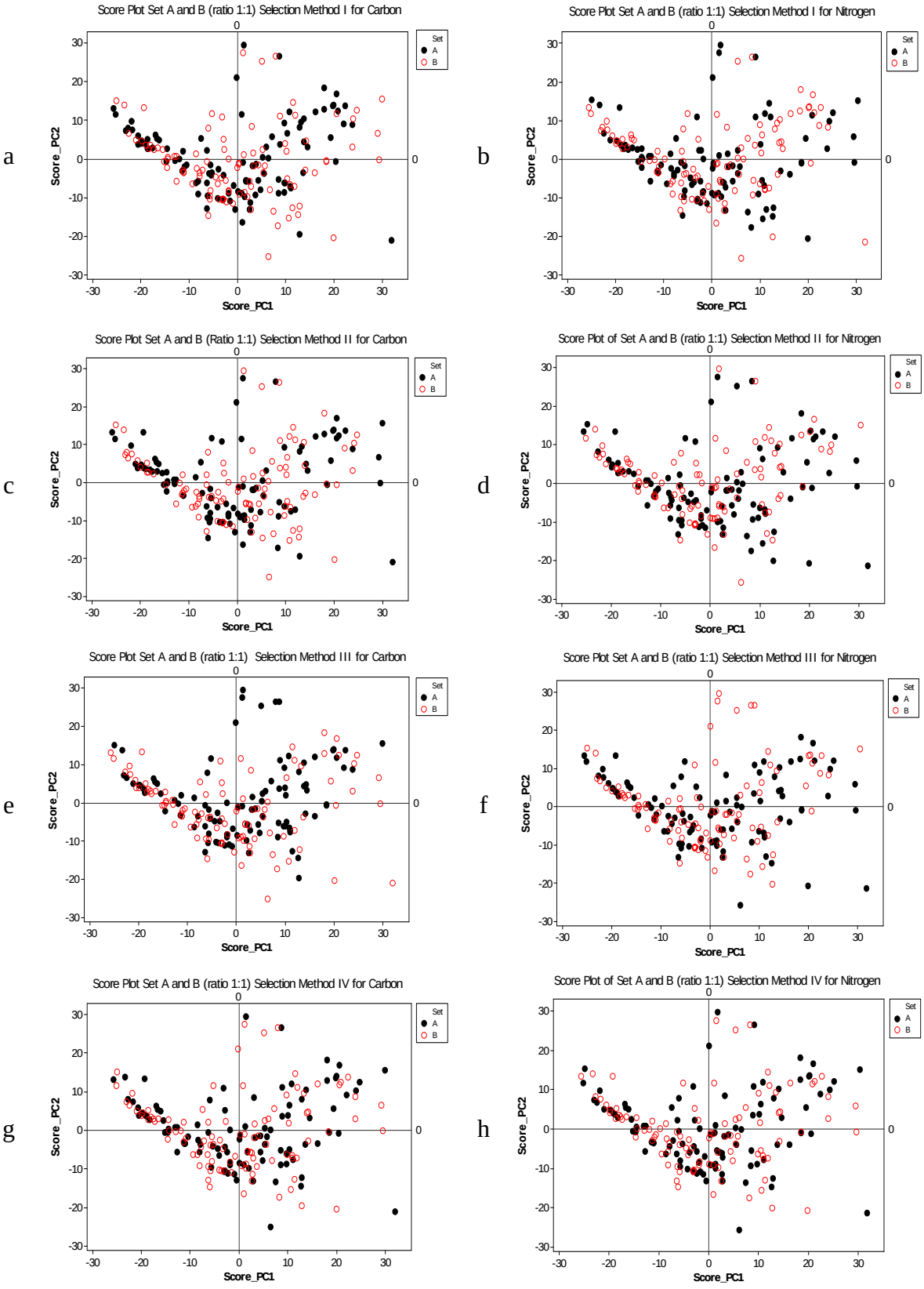


Figure 6.4 PCA score plot of set A and B selected by method I-IV for carbon and nitrogen.

The advantage of selection method I is that calibration and validation data sets can easily be chosen from knowledge of the range and type of soil samples taken. The practical use of this method may be in limiting the number of expensive chemical analyses that are required by analysing half the samples rapidly with spectral reflectance. The limitation of selection method I is that it cannot be implemented when the sample history is unknown.

The second selection method (II) was carried out by ranking the samples from the lowest to the highest C (or N) content and equally dividing them by allocating every odd rank to the calibration set and even rank to the validation set. This selection technique is normally used only in method development, and assures that both calibration and validation sets have similar reference-data range (Esbensen *et al.* 2006) and a wide range of analyte values (Westerhaus *et al.* 2004). In our datasets small differences of prediction accuracy values (Table 6.3, cf. r^2, RMSEP, RPD and RER for A to B and B to A test - *vice versa* test) illustrate that relatively stable regression models are produced for both C and N prediction. Martin *et al*. (2002) also used the *vice versa* test to select stable regression models to predict soil organic C content.

The selection of data for calibration and validation sets in method II is very simply based on the range of chemical reference values. Compared to the selection methods I, III and IV, selection method II produced the most stable model for C and N (Table 3.3, very similar values for r^2, RMSEP, RPD and RER for A to B and B to A, *vice versa* test). The distribution of sample sets A and B in the PC1 and PC2 score plots (Figures 6.4c and 6.4d) show very similar distributions in both sets. As mentioned by Esbensen *et al*. (2006), calibration and validation data sets should have the same general pattern in the score plot, although the two sets must not be too similar. The disadvantage of selection method II is that the chemical reference analysis of all samples must be known and therefore it can only be used in method development.

Unlike methods I or II which either require prior knowledge of soil sample characteristics or the values of reference chemical data (Y-variables), selection method III is an independent method based only on the PCA analysis of the spectral reflectance data (X-variables). Selection method III involved calculating standardised Euclidean distances using scores of PC1 and PC2 of the PCA as the coordinates (Figure 6.4e and

6.4f). The standardised Euclidean distances from the centroid were then ranked from the lowest to the highest, and then odd rank numbers were allocated to the calibration set and even numbers to the validation set. As expected, method III produces very similar spread and ranges of data for the calibration set A and validation set B (Esbensen *et al.* 2006). For carbon, calibration set B gives better coverage of scores (Figure 6.4e) and the slightly more accurate regression model (Table 6.3).

In an attempt to make the sample distribution more equal, method IV calculated Euclidean distance of subsamples of data restricted to each quadrant of the PC1 and PC2 score. The data in each quadrant was ranked from lowest to the highest standardised Euclidean distance from the centroid and then samples with odd numbered ranks in each quadrant were allocated to the calibration set and even numbers to the validation set. The results of this selection are depicted in Figure 6.4g and 6.4h. Compared to the distribution of data in Figure 6.4e and 6.4f, it appears that the distribution of each A and B sets are more evenly distributed.

Even though set A and B were more evenly distributed in the PC1 and PC2 score plot (Figure 6.4g and 6.4h), this did not improve the accuracy or stability of the prediction models for C (RMSEP = 1.47% and RPD = 1.74) or N (RMSEP = 0.09% and RPD = 2.27) (Table 6.3) when compared to selection method III.

Graphs of total C and N measured by LECO and values predicted using reflectance spectra and calibration set selection methods I, II, III and IV are shown in Figure 6.5. The best of these prediction models allow soil C in the 0-11% range to be predicted with an error (RMSEP) of 1.16% and soil N in the range 0-1.0% to be predicted with an error of 0.08% (Table 6.3).

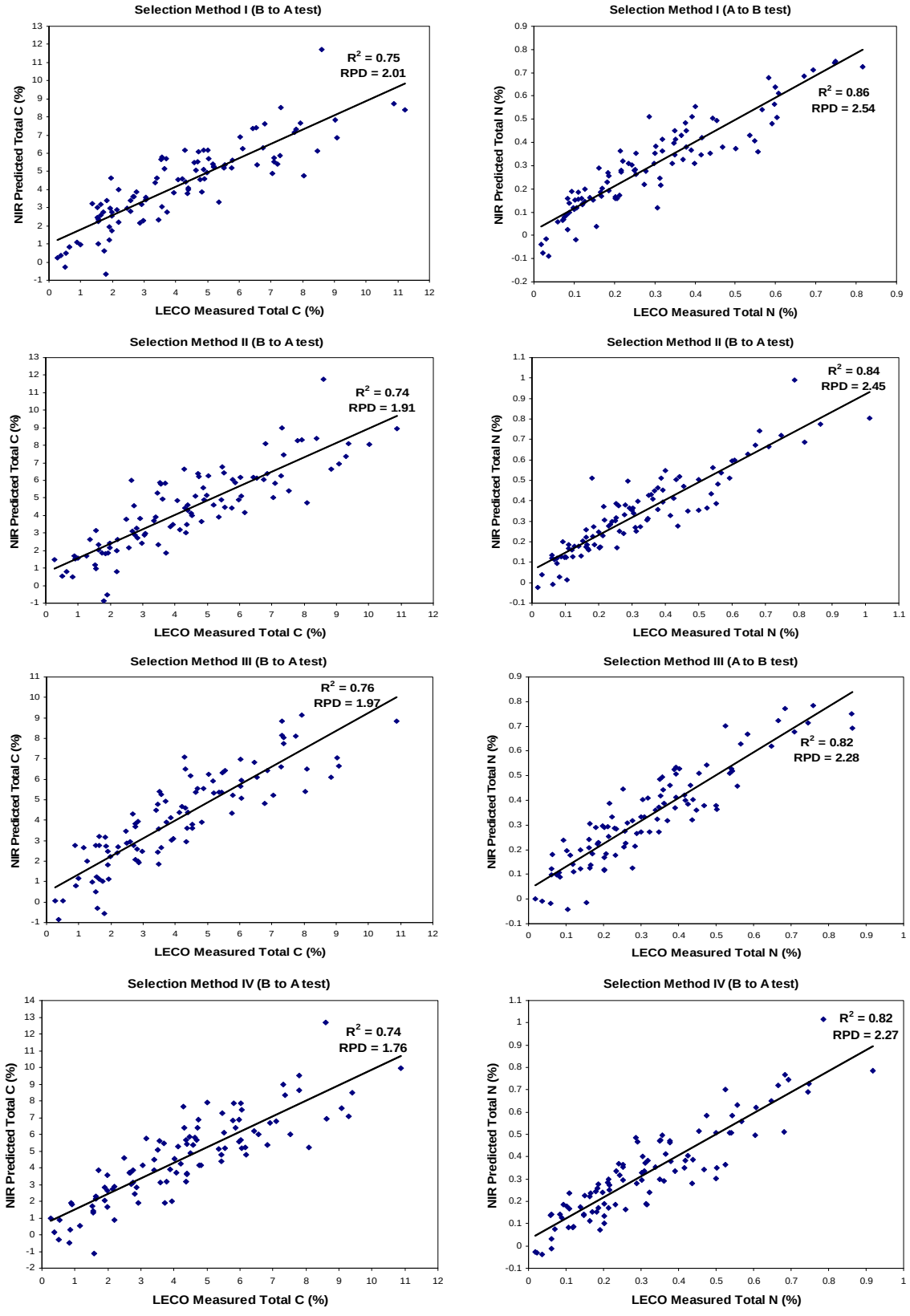


Figure 6.5 Linear regression relationships between measured (LECO) and predicted (reflectance PLSR calibration models) soil C and N produced by calibration set selection methods I-IV.

Other researchers (Malley and Martin 2003; Westerhaus *et al.* 2004) have stated that increasing the number of samples in the calibration set will increase the probability that the range of analyte values and spectral responses covered by the calibration set and may improve the accuracy of the regression model. After ranking the sample data using selection methods II and III, a larger number of samples were allocated to the calibration set than the validation set (at ratios of 2:1 and 3:1, respectively) by selecting the first 2 (or 3) rank numbers and allocating them to the calibration set and the 3rd (or 4th) ranked value to the validation set. The parameters indicating prediction accuracy for the respective regression models are presented in Table 6.4. There was little or no improvement in the accuracy of prediction of soil C or N (cf. values of r^2, RMSEP, RPD and RER for selection methods II and III in Table 6.4 with the similar values in Table 6.3).

Table 6.4 Prediction accuracy of soil C and N values from PLSR calibration models using different methods (II and III) to select the calibration and validation sets (ratio 2:1 and 3:1).

Selection method	Statistical parameter	Total C		Total N	
		Ratio 2:1 A to B	Ratio 3:1 A to B	Ratio 2:1 A to B	Ratio 3:1 A to B
II	n in set A	134	150	133	150
	n in set B	66	50	66	49
	Component	5	5	6	8
	r^2-validation	0.66	0.74	0.83	0.86
	RMSEP (%)	1.38	1.28	0.09	0.07
	RPD	1.73	1.94	2.27	2.66
	RER	6.99	8.34	9.92	11.15
	Slope	0.65	0.74	0.78	0.93
III	n in set A	134	150	133	150
	n in set B	66	50	66	49
	Component	5	6	6	6
	r^2-validation	0.74	0.72	0.86	0.81
	RMSEP (%)	1.41	1.42	0.08	0.09
	RPD	1.83	1.89	2.61	2.24
	RER	7.35	7.22	10.50	8.95
	Slope	0.65	0.72	0.83	0.89

Note: n = number of samples; Slope = regression coefficient between measured and predicted soil C and N

The explanation for the lack of improvement in the prediction accuracy of the model probably rests with the fact that selection methods II and III already ensure that sample analyte values, and spectral responses are already ranked for range before the data are allocated to the calibration and validation sets.

6.3.4. Important Wavebands for Explaining the Variance in Measured Soil C and N

The partial least squares regression coefficients for predicting soil C and N from the first derivative of the smoothed reflectance spectra produced using selection method II and III are presented graphically in Figures 6.6 and 6.7, respectively. Figures 6.6 and 6.7 show the PLS regression model coefficients for every 5 nm wavelength band. The size of the coefficient (negative or positive) represents the importance of the band in terms of the explanation of variance in soil C or N analysis. The similar patterns (cf. solid *v.* dashed lines) suggest that both data sets build very similar regression models to predict soil C (Figure 6.6) and N (Figure 6.7). Although in some bands the coefficient size may change, most of the important bands remain the same. Interestingly, both models for C and N have important bands at wavelengths between 1455-1460 nm and 1925-1935 nm. These are water absorption bands (Figure 6.2) associated with soil water that varies with soil organic matter content. Both models have small coefficients in bands of visible wavelength (475-750 nm), indicating that the visible wavelengths are less important in predicting soil C and N content in this sample of soils, which contrasts with previous findings that reported visible wavelengths 600-700 nm (Vinogradov 1981) 564.4 and 623.6 nm (Krishnan *et al.* 1980) were important for organic matter prediction; see also Baumgardner *et al*. (1985). Large variations in soil colour of this sample population which can cause considerable distortion of low wavelength ranges (e.g., 400-650 nm) (Williams and Norris 2001) may be the reason for the less importance of the visible band. Low coefficients can also be found at the bands between 750-850 nm.

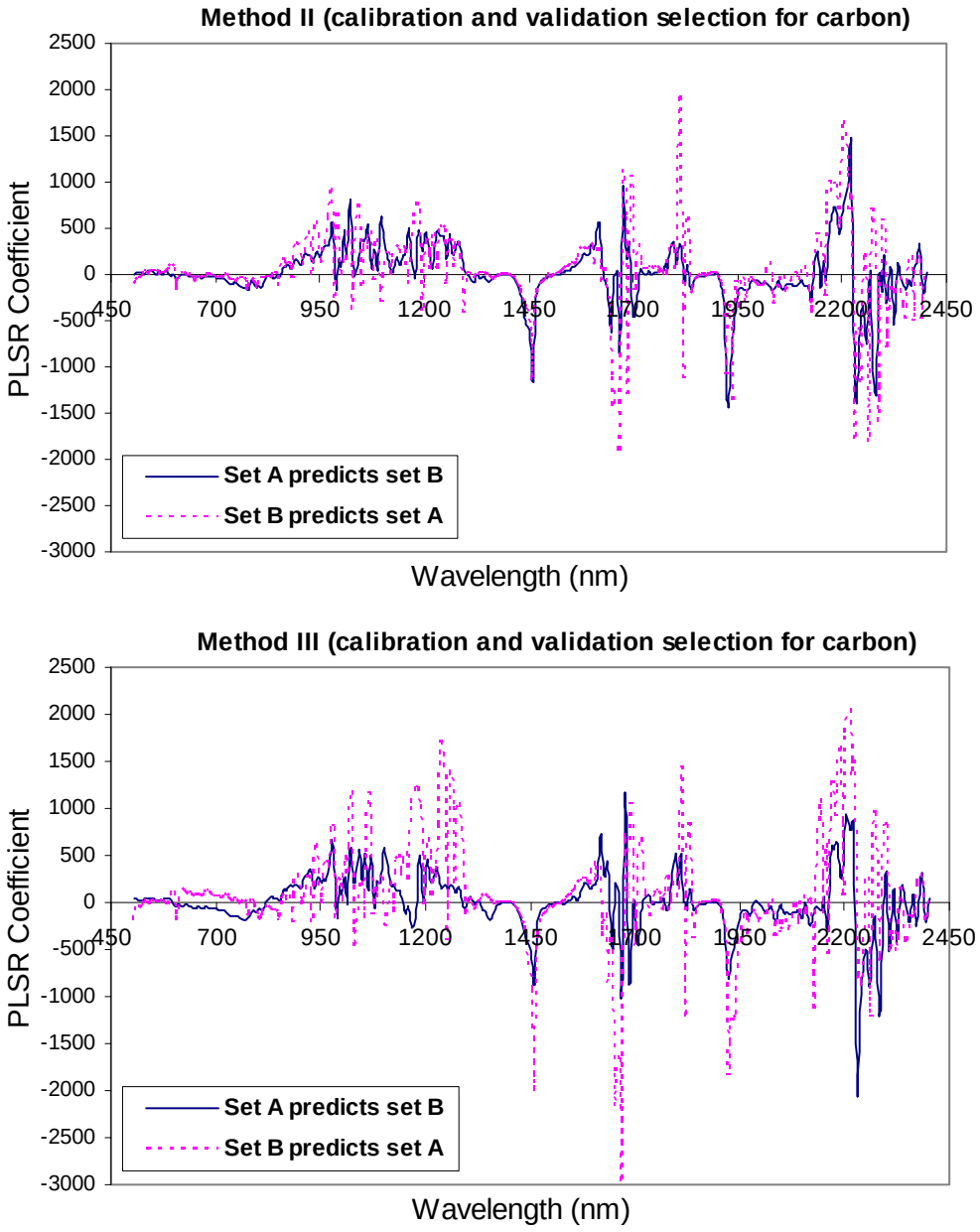


Figure 6.6 PLSR coefficient for C, selection method II and III, with "vice versa test" applied.

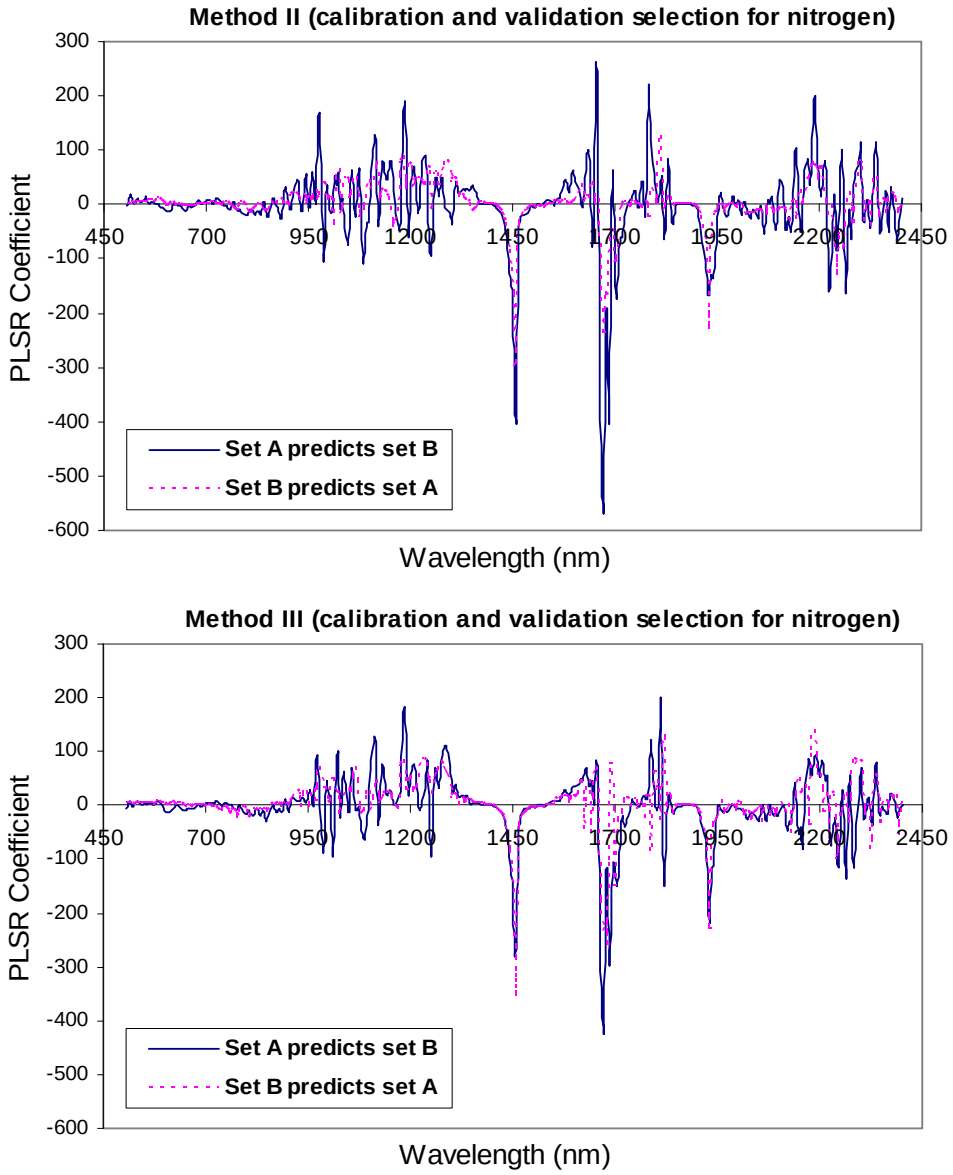


Figure 6.7 PLSR coefficient for N, selection method II and III, with "vice versa test" applied.

Larger coefficients are attributed to a significant number of wavelengths from 850 to 2405 nm (Figure 6.6 and 6.7). The importance of this band range is in agreement with previous workers who found wavelengths between 900 and 1220 nm were the best for mapping soil organic C (Beck *et al.* 1976) and between 1702-2052 nm for predicting organic matter and total organic N (Dalal and Henry 1986) and 1726 and 2426 nm for organic matter identification (Morra *et al.* 1991). Chang and Laird (2002) and Chang *et al*. (2005) used wavelengths between 1100-2498 nm to predict organic C, inorganic C,

total C, total N and moisture content. Chang *et al.* (2001) found wavelengths 1300-2498 nm successfully predicted total C, total N, and other soil properties (moisture, CEC, 1.5MPa water, basal respiration rate, sand, silt and Mechlich III extractable Ca). They did not use reflectance spectra from 400 to 1300 nm because it was found to reduce the accuracy of predicted soil properties. Similarly, Reeves and McCarty (2001) found optimal calibrations using reflectance spectra between 1100-2300 nm to predict total N, organic C, active N, biomass and mineralizable N and pH. McCarty and Reeves (2006) used NIR bands between 1100-2498 nm, rather than the full recorded range of 400-2498 nm, to successfully predict organic C, total N and soil texture. They reported that no advantage was obtained by including bands 400-1098 nm.

Wavelengths in the region 475-775 nm had low PLSR coefficients, but when they were cut from the spectral data, the clipped 775-2435 nm range did not produce a useful improvement in the accuracy of C and N prediction by the regression model. Therefore subsequent analyses were conducted with the 475-2435 nm spectral data range.

6.3.5. Comparison of Regression Model Prediction Accuracy with other Published Results

Statistics indicating the accuracy of prediction of soil C and N using soil spectral reflectance data from previous publications are compared with this study (Table 6.5 and 6.6, respectively). The first point to note is that all soil C and N prediction models have been calibrated by making both the spectral and chemical measurements in the laboratory. A number of authors have proposed categorisation of prediction accuracy e.g. (Chang *et al.* 2001; Dunn *et al.* 2002; Malley *et al.* 2004; Malley *et al.* 1999). Malley *et al.* (2004) categorised prediction accuracy into successful when the RPD 3-4 (with RER 15-20, r^2 0.90-0.95) and excellent when the RPD > 4 (with RER > 20, r^2 > 0.95). They categorised prediction accuracy into moderately useful and moderately successful when the RPD is 1.75-2.25 (with RER 8-10, r^2 0.70-0.80) and 2.25-3 (with RER 10-15, r^2 0.80-0.90) respectively, which were achieved by the majority of laboratory determined models presented in Table 6.5 and 6.6.

Table 6.5 A review of the prediction accuracy for using near-infrared spectroscopy to estimate soil C concentrations.

Soil properties predicted	Sample preparation	Method	Validation prediction accuracy				References
			RPD	RER	r^2	SEP*	
Organic matter	1, 3c	Lab.	-	-	0.55	1.34 %	(Ben-Dor and Banin 1995)
C Total	1, 3a	Lab.	2.79	-	0.87 cs-val	7.86 g/kg *cs-val*	(Chang *et al.* 2001)
C Total	1, 3e	Lab.	-	-	0.86 0.96	-	(McCarty *et al.* 2002)
C Total	1, ground to fine powder	Lab.	4.4	-	0.91	6.53 g/kg	(Chang and Laird 2002)
C Organic	1, 3c (but 97% of sample < 0.5 mm)	Lab.	1.7	8.8	0.66	0.25%	(Dunn *et al.* 2002)
C Organic	1, 3c	Lab.	2.16	10.87	0.79	3.32 mg/g	(Martin *et al.* 2002)
C Organic	1, 3d	Lab.	2.2	-	0.80	6.8 g/kg	(Fystro 2002)
	2, 4	Lab.	2.7	-	0.87	5.7 g/kg	
C Organic	1, 3c	Lab.	-	-	0.91 *calib*	2.2 g/kg *calib*	(Shepherd and Walsh 2002)
C Organic	1, 3c	Lab.	3.1-3.3	-	-	4.9-5.0 g/kg	(Moron and Cozzolino 2002)
C Organic	1, 3c	Lab.	1.7	8.6	0.76	0.44 %	(Islam *et al.* 2003)
C Organic	1, 3c	Lab.	3.0	-	-	3.9 g/kg	(Moron and Cozzolino 2004)
C Organic	1, 3e	Lab.	2.0	-	0.76	0.50 %	(Pirie *et al.* 2005)
C Total	2, 3c	Lab.	2.20	-	0.85	4.26 g/kg	(Chang *et al.* 2005)
C Organic	1, 3c	Lab.	1.73-2.62	-	0.68-0.77	1.09-1.47 g/kg	(Brown *et al.* 2005)
			0.48-2.09	-	0.39-0.85	0.96-3.99 g/kg	
C Total	1, ground	Lab.	-	-	0.88	1.6 g/kg	(McCarty and Reeves III 2006)
C Organic	1, 3c	Lab.	-	-	0.60 cs-val	0.18 dag/kg *cs-val*	(Viscarra Rossel *et al.* 2006)
Organic Matter	1, 3c	Lab.	-	-	0.93	0.06 %	(He *et al.* 2007)
C Total	1, 3c	Lab.& Field	1.92	-	0.73	0.27 %	(Mouazen *et al.* 2007)
C Total	Soil core, 2,4	Field	2.01	9.07	0.75	1.21 %	Kusumo *et al.* in this study

Note: SEP*s of those studies were reported as SEP (standard error of prediction) or other equivalent statistics such as RMSEP or RMSECV (root mean square error of cross-validation) or RMSEC (root mean square error of calibration), 1=Drying, 2=Moist, 3a=Sieving < 8 mm, 3b=sieving < 4 mm, 3c=sieving < 2 mm, 3d=Sieving < 0.5 mm, 3e=Sieving < 0.2 mm, 4=No Sieving, Lab.=Laboratory, *cs-val* =cross-validation, *calib*=calibration.

Table 6.6 A review of the prediction accuracy for using near-infrared spectroscopy to estimate soil N concentrations.

Soil properties predicted	Sample preparation	Method	Validation prediction accuracy				References
			RPD	RER	r^2	SEP*	
N Total	1, 3a	Lab.	2.52	-	0.85 cs-val	0.62 g/kg *cs-val*	(Chang *et al.* 2001)
N Total	1, ground to fine powder	Lab.	2.8	-	0.86	0.36 g/kg	(Chang and Laird 2002)
N Organic	1, 3c	Lab.	1.26	6.2	0.37	0.53 mg/g	(Martin *et al.* 2002)
N Total	1, 3c	Lab.	2.6-3.1	-	-	0.4-0.5 g/kg	(Moron and Cozzolino 2002)
N Total	1, 3d	Lab.	1.8	-	0.70	0.58 g/kg	(Fystro 2002)
	2, 4	Lab.	2.1	-	0.80	0.49 g/kg	
N Total	1, 3c	Lab.	4.0	-	-	0.3 g/kg	(Moron and Cozzolino 2004)
N Total	2, 3c	Lab.	3.45	-	0.92	0.23 g/kg	(Chang *et al.* 2005)
N Total	1, ground	Lab.	-	-	0.85	0.16 g/kg	(McCarty and Reeves III 2006)
N Total	1, 3c	Lab.	-	-	0.93	3.28 mg/kg	(He *et al.* 2007)
N Total	Soil core, 2,4	Field	2.66	11.15	0.86	0.07 %	Kusumo *et al*. in this study

Note: SEP*s of those studies were reported as SEP (standard error of prediction) or other equivalent statistics such as RMSEP or RMSECV (root mean square error of cross-validation) or RMSEC (root mean square error of calibration), 1=Drying, 2=Moist, 3a=Sieving < 8 mm, 3b=sieving < 4 mm, 3c=sieving < 2 mm, 3d=Sieving < 0.5 mm, 3e=Sieving < 0.2 mm, 4=No Sieving, Lab.=Laboratory, *cs-val* =cross-validation, *calib*=calibration

Our best regression models (Table 6.5 and 6.6) are able to predict soil C and N with moderately useful and with moderately successful levels of accuracy (Malley *et al.* 2004), despite the lack of sample preparation and the variably moist soil conditions (11-82%) in our field determination of spectral reflectance. Soil texture of these Pumice, Allophanic and Tephric Recent Soils which are mostly coarse texture may be the reason for this moderate accuracy. Dalal and Henry (1986) found higher error of organic C and total N prediction in coarsely ground (< 2 mm) than in finely ground (< 0.25 mm) soil samples. Barthes *et al*. (2006) found less accurate prediction of total C and N of oven-dried < 2 mm than oven-dried < 0.2 mm particle size samples. In addition, high water content of this soil samples (range WC 11-82%; 46% samples have > 50% WC) may cause this lack of accuracy because it can reduce the strength of important absorption feature of C and N at a 2200 nm band by increasing water content (Lobell and Asner 2002). Malley *et al*. (2002) reported less accurate prediction of e.g. organic matter and NH_4-N with field moist soils than dry soils (dried at 40^{o}C). Kooistra *et al*. (2003)

noticed negative impact of water content on organic matter and clay content prediction. Some improvement in the prediction of soil carbon content is desired but the accuracy of soil C estimation in this study will be quite useful in field surveys to establish soil zones that require separate treatment or management. We are pleased with the accuracy with which we are able to predict soil N content (Table 6.6) and our aim for future studies is to develop techniques to improve the prediction of soil C content from a level of moderately useful estimation to successful or excellent estimation.

6.4. Conclusions

We have developed a purpose-built contact probe for acquiring soil reflectance with a field spectroradiometer. PLSR calibration models based on the first derivative of the acquired field soil reflectance spectra show potential to be used for *in situ* estimation of total soil C and N in pastoral soils. PLSR models have a greater chance of being more accurate if the data sets used for calibration have a similar range of spectral and reference chemical properties to the target unknown data set. This current study shows that when reference chemical data of the target set are unknown, similarity can be based on selecting calibration data that have spatial and management histories similar to the target unknowns, or selection can be based on similarities in the spectral data alone. Our sample processing techniques based on the variance observed in the spectral data allow selection of the calibration data set to form a PLSR model that accommodates the spectral properties of the target unknown soil core sample. This has the potential to improve the accuracy of the prediction model. This study indicates that *in situ* assessment of soil C and N by field spectroscopy has considerable potential for rapid mapping of soil C and N across the landscape.

CHAPTER 7

Improving the Accuracy of Predicting Soil Carbon and Nitrogen Concentrations from Proximally Sensed Soil Spectral Reflectance

From experiences gained in earlier Chapters (4, 5 and 6) it became apparent that the method used to acquire spectral reflectance from flat horizontal cross-sections of soil cores caused an intrinsic calibration error because the spectra were acquired from "a field of view" that could not be replicated by the sample taken for the laboratory determination of the reference measurement of soil C and N concentration. The aim of this chapter is to compare the original method with a method that acquired the spectra from the curved vertical wall of a cylindrical soil core.

A paper summarizing some results from this study was reviewed for oral presentation and the proceeding of the First Global Workshop on High Resolution Digital Soil Sensing & Mapping, 5-8 February 2008, Sydney, Australia (the Australian Centre for Precision Agriculture, University of Sydney and CSIRO Land and Water). An extended paper has been submitted and reviewed for a book chapter and will be published along with other papers from this conference. A research article summarizing the complete results of this study will be submitted to *Australian Journal of Soil Research*, in 2009.

7.1. Introduction

Mapping soil properties using reflectance spectroscopy is an emerging technology (Gehl and Rice 2007), as portable spectroradiometers and geographic positional system (GPS) become more widely available (Adamchuk *et al.* 2004). Near infrared reflectance spectroscopy (NIRS), which is a rapid and non-destructive technique, has been used to predict soil carbon (C) from "as is" field soil samples using a mobile on-the-go technique (e.g. Mouazen *et al.* 2007). A recent development (Kweon *et al.* 2008) with a sapphire windowed soil probe has the potential to allow a map of soil C to be produced both in horizontal and vertical dimensions. A manual coring approach, followed by acquiring reflectance spectra with a backpack field spectroradiometer (Analytical Spectral Devices FieldSpecPro, Boulder, CO, USA), has been developed by Kusumo *et*

al. (2008a) to estimate soil C and nitrogen (N) to depth and plant root density (Kusumo *et al.* 2009b; Kusumo *et al.* 2009a). Such capabilities have obvious importance for signatories to the Kyoto Protocol that wish to adopt Article 3.4 and be allowed to offset greenhouse gas emissions with audited proof of increased soil carbon to 30 cm soil depth.

The coring technique of Kusumo *et al*. (2008a) involved cross-sectioning intact cores and acquiring the spectra from the freshly cut horizontal surface. Partial least squares regression (PLSR) was used to develop a calibration model of the first derivative of the reflectance spectra against laboratory measured reference soil C and N concentrations. Under the conditions of the study, the calibration model produced by this technique was able to deliver only "moderate accuracy" in the prediction of the laboratory measured reference soil C and N concentrations (Malley *et al.* 2004). Improvement of this technique is required to predict soil C and N more accurately. Soil C and N concentrations have considerable variance with depth, therefore one possible source of error with the technique may arise because the C and N concentrations of the few millimetres of soil surface from which the diffusely reflected spectra were acquired is different from the soil C and N concentrations in the 1.5 cm of soil section below that surface, which was taken for laboratory measurements. Another source of error may arise because diffusely reflected NIR radiation from a flat soil surface contains significant specular reflectance that is "noise" containing little information from the surface material.

The objective of this paper is to report differences in soil reflectance spectra caused by changing the field of view of the spectroradiometer from a horizontal cross-section of a soil core to the cylindrical vertical side of a soil core and the effect of these different acquisition methods on the accuracy of predicting soil C and N in pastoral soils.

7.2. Materials and Methods

7.2.1. Contact Probe Modification and Measurement Techniques

A modified soil probe was developed based on the plant contact probe (supplied by ASD, Boulder, CO,USA) (Figure 7.1); its detail was explained in Kusumo *et al*.

(2008a). This soil probe was used to acquire soil reflectance spectra using two methods; first, method **H**, involved measurements made from a flat sectioned horizontal soil surface of a soil core (Figure 7.1a), and second, method **V,** from a curved vertical side of a cylindrical soil core (Figure 7.1b). The soil core was rotated 360^{o}, which produced a flat field of view of 561 mm^{2} for the **H** method and a 4707.4 mm^{2} cylinder field of view for the **V** method. Both methods have the same distance (30.5 mm, the shortest distance to curved surface in **V** method) between the object (soil surface) and the probe window.

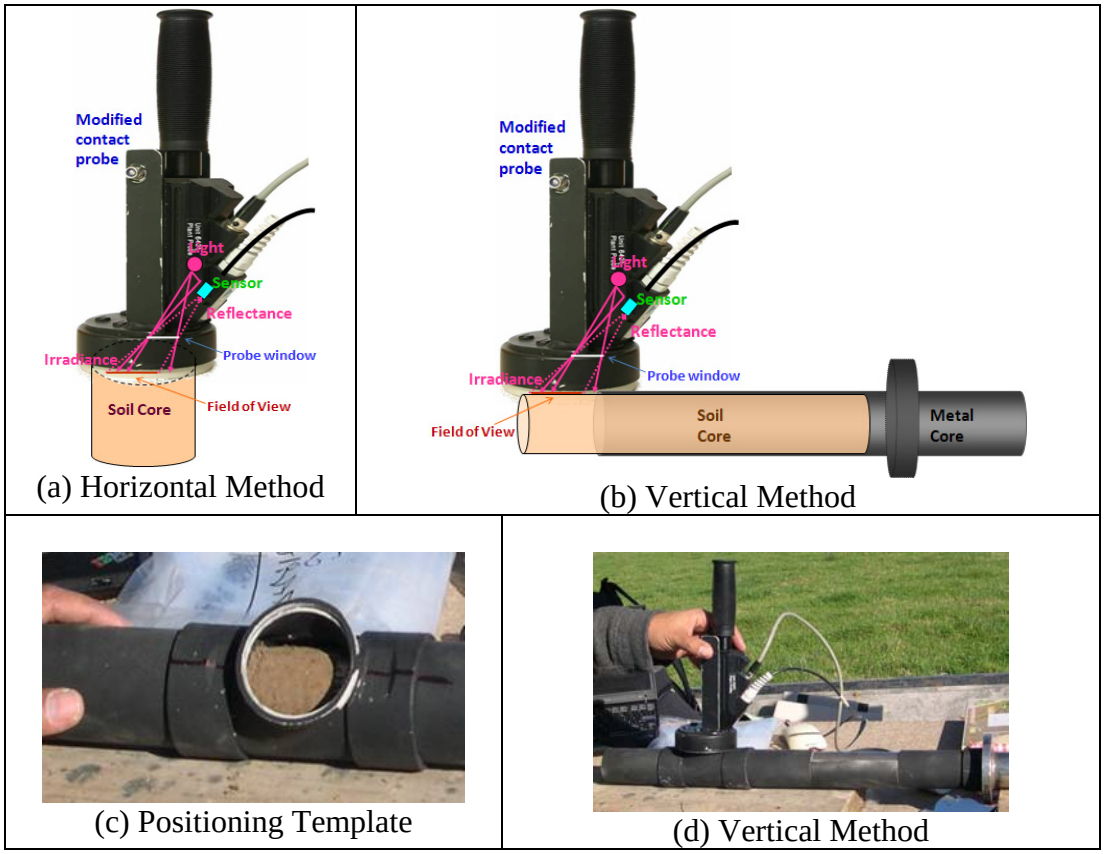


Figure 7.1 A prototype soil probe for acquiring soil reflectance spectra using the (a) horizontal and (b) vertical measurements and (c) the positioning template for (d) vertical method.

7.2.2. Site Locations and Sample Collection

The **H** and **V** methods were compared for soil C and N assessment under permanent pasture on Manawatu fine sandy loam [Fluvial Recent soil (Hewitt 1998); Dystric Fluventic Eutrudept (USDA Soil Survey Staff 1999) on greywacke alluvium] in the Manawatu region, New Zealand (Hewitt 1998). The permanent ryegrass (*Lolium*

perenne) white clover (*Trifolium repens*) dominant pastures at both sites had been present for more than 20 years. Soil cores were collected (November 2007) from 3 transects that were 20 m apart, with 6 cores in each transect at 15 m intervals. A total of 18 soil cores were sampled at 6 depths at 5 cm intervals from 1.5 to 31.5 cm depth, totalling 108 samples. Field soil reflectance spectra were acquired from the horizontal (Figure 7.1a) and vertical (Figure 7.1b) side of a 5 cm cylindrical soil core at each depth. Each soil slice was sectioned using a knife and collected in a plastic bag for laboratory analysis.

7.2.3. Measurement of Soil Properties

Samples sliced from soil cores were weighed field moist, crumbled and allowed to air dry before reweighing. Soil moisture content was expressed as a fraction of air dry weight. Air dry soils were then ground to < 500 µm particle size for total C and N analysis using a LECO FP-2000 CNS Analyser (Blakemore *et al.* 1987).

7.2.4. Spectral Pre-processing and Development of Calibration Model

Each replicate spectra acquired by the modified soil probe (attached by fibre optic cable to the spectroradiometer (ASD FieldSpecPro, Boulder, CO.) was the time average of 10 acquisitions. By rotating the probe 360º, ten replicates of a time-averaged reflectance spectrum were acquired from each soil sample. The spectroradiometer records spectra from 350-2500 nm and samples at intervals of 1.4 nm for the region 350-1000 nm and 2 nm for the region 1000-2500 nm. The data processing software supplied with the spectroradiometer interpolates the data into 1-nm spaced data. Spectral data (1-nm spaced) were pre-processed (Hueni and Tuohy 2006) using a Savitzky-Golay smoothing filter with window size 31 nm and a 4^{th} order polynomial algorithm (Savitzky and Golay 1964). The smoothed data were reduced by taking every fifth waveband, then transformed into the first derivative, and finally the 10 spectra acquired per sample were averaged. The first derivative data were imported to Minitab 14 (MINITAB Inc. 2003) for principal component analysis (PCA) and partial least squares regression (PLSR). A score plot of PCA components PC1 and PC2 which accounted for the greatest variance in the pre-processed spectral data was used to observe the pattern of sample scattering. PLSR was used to develop calibration models from the first derivative spectral data and the reference analytical data (C and N).

During PLSR processing, samples which had a standardized residual > 2.0 were removed as outliers from the calibration and validation data sets. The accuracy of the models were tested internally using the leave-one-out cross-validation method (Kusumo *et al.* 2009a) and externally using a separate validation data set (Kusumo *et al.* 2008a).

7.2.5. Spectral Noise

Spectral noise was estimated by calculating the variance between raw and smooth spectra. The root mean square (*rms*) noise was calculated using the following equation (Adam 2004):

$$rms\ noise = \sqrt{\frac{\sum (x_{is} - x_i)^2}{n-1}}$$

where x_{is} is the smooth reflectance value at *i* wavelength, x_i is the raw reflectance value at the *i* wavelength and *n* is the number of variables (wavelengths).

7.2.6. Regression Model Accuracy

The ability of the PLSR model to predict soil properties was assessed using the statistics: RMSE (root mean square error) of measured and predicted soil property values, coefficient determination (r^2), RPD (ratio of prediction to deviation) which is the ratio of the standard deviation of measured value of soil properties to the RMSE, and RER (ratio error range) which is the ratio of the range of measured values of soil properties to the RMSE. The explanation of these statistical values is presented in Kusumo *et al*. (2008a). The best prediction model is shown by the largest RPD, r^2, RER and the smallest RMSE (in cross-validation or validation procedure).

7.3. Results and Discussion

7.3.1. Summary of Soil Properties

A summary of soil total C, total N and water content is presented in Table 7.1. The data show a wide range of soil C and N concentrations, which decrease with soil depth (Figure 7.2). The high C and N concentrations (mean C 4.36%, range C 3.47-5.20%; mean N 0.42%, range N 0.33-0.50%) in the top soil (1.5-6.5 cm depth) are consistent with this Fluvial Recent soil being under permanent pasture and not having been cultivated for more than 20 years (Shepherd *et al.* 2001). Such soils accumulate organic matter in topsoils through excreta return and fine root turnover and litter decomposition (Guo *et al.* 2005).

Table 7.1 Soil properties of the 108 soil samples.

Soil Properties	Range		Median	Mean	Variance
	Minimum	Maximum			
Total Carbon (%)	0.27	5.20	1.41	1.85	1.681
Total Nitrogen (%)	0.02	0.50	0.14	0.18	0.015
Water Content (%)	20.7	53.6	28.6	30.7	39.7

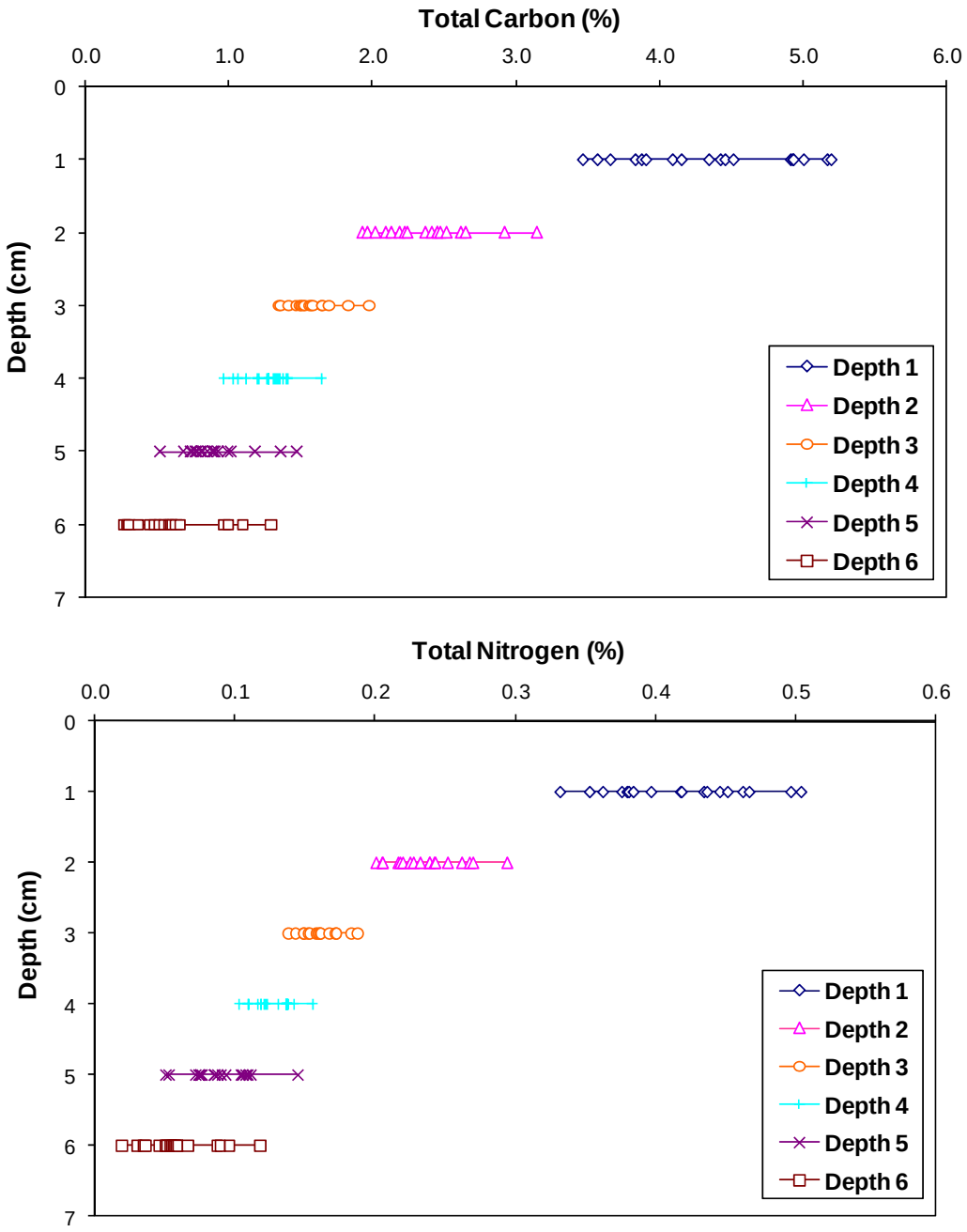


Figure 7.2 (a) The total carbon and (b) nitrogen concentrations at each sampling depth (depth; 1 (1.5-6.5 cm), 2 (6.5-11.5 cm), 3 (11.5-16.5 cm), 4 (16.5-21.5 cm), 5 (21.5-26.5 cm), and 6 (26.5-31.5 cm).

7.3.2. Spectral Shapes Recorded by H and V Method

The reflectance spectra acquired from six different depths of the same soil core recorded by both **H** and **V** methods are presented in Figure 7.3. A slightly concave shape of spectrum at about 900 nm, which is perhaps due to absorbance by iron oxides (Demattê and Garcia 1999), occurs in subsoil samples (11.5 to 31.5 cm depth). Disappearance of the concavity at this 900 nm band in samples from shallower depths (1.5 to 11.5 cm) probably results from higher organic matter content masking iron oxides (Demattê *et al.* 2004a). Strong water absorption bands at around 1455 nm results from the first overtones of O-H stretching bond, and bands around 1945 nm are due to the combinations of H-O-H bend with O-H stretching (Clark 1999).

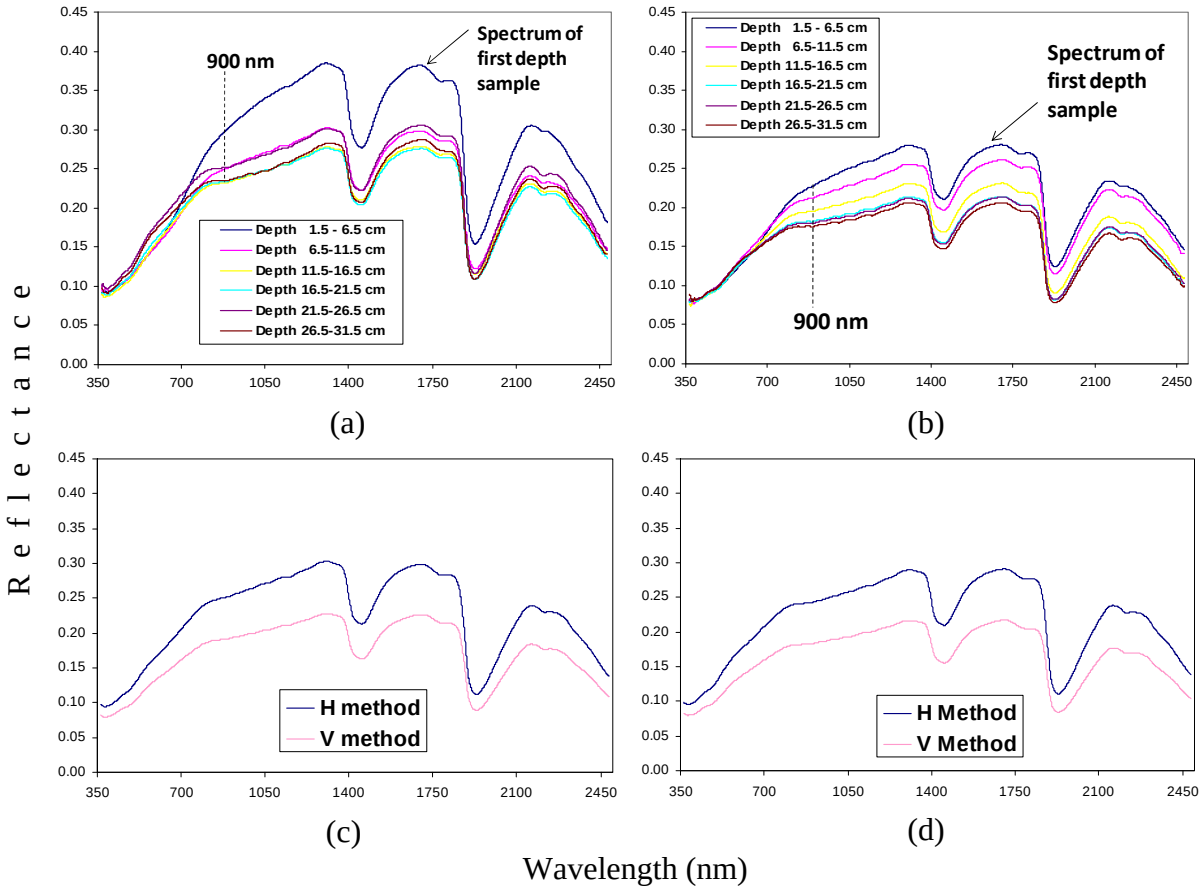
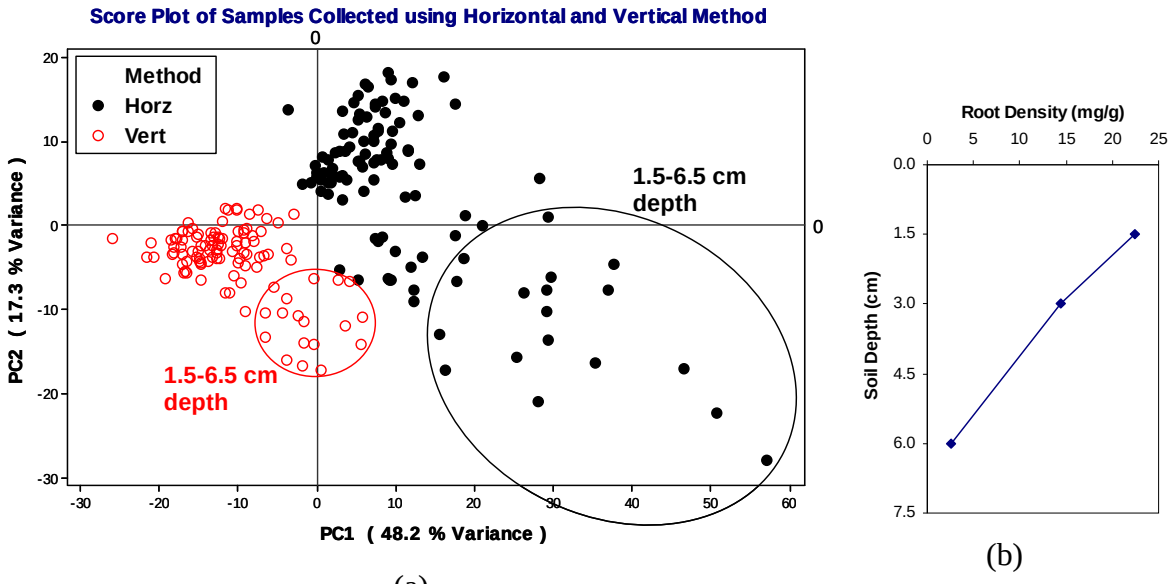


Figure 7.3 Reflectance spectra acquired from six depths of the same soil core recorded using (a) the horizontal and (b) vertical method, and the average of (c) all spectra and (d) the 2nd-6th depth spectra.

The spectrum from the shallowest soil depth (1.5-6.5 cm) recorded using the **H** method exhibits much higher reflectance especially in the NIR region (780-2500 nm) compared to the deeper samples (6.5-31.5 cm depth) (Figure 7.3a). The higher root

density and organic residues in the top-soil accounts for this (Figure 7.4b, root data acquired from the same site). Increasing pasture root density (Kusumo *et al.* 2009b; Kusumo *et al.* 2009a) or crop residues (Demattê *et al.* 2004b) tend to increase reflectance especially in the NIR region. Higher reflectance in the NIR region is also related to a large number of air voids present in roots and crop residues; more air-cell interfaces cause higher reflection (Baumgardner *et al.* 1985). In addition, the first soil depth spectrum resembles the spectrum of minimally decomposed soil material (fibric) reported by Stoner and Baumgarner (1981).

Higher reflectances are obtained with the **H** method than the **V** method at all soil depths (Figure 7.3a c.f. 7.3b). This is most obvious with the first organically rich, high root content soil depth (Figure 7.3a c.f. 7.3b), and when this depth is included in the average spectrum (Figure 7.3c). However, this phenomenon remains significant even when the first depth is absent from averaged spectra (Figure 7.3d). The higher reflectance is more a function of the method of acquisition rather than soil depth, and results from the differences in shape of the soil surface (curved in **V** method and flat in **H** method). Less light is reflected from the curved surface of the **V** method. The variance of the reflectances acquired by the **V** method was smaller, which is shown by less scattering of observations (smaller convex hull, Figure 7.4a); especially with observations from the topsoil samples (1.5-6.5 cm, enclosed in circles in Figure 7.4a).



(a) (b)

Figure 7.4 (a) A score plot of the first two PCs explaining 65% of the variance in the acquired soil reflectance spectra and (b) the distribution of root density with soil depth (root data from Kusumo *et al.* 2009b).

7.3.3. Comparison of Accuracy of C and N Predicted by the H and V Method

Compared to the **H** method, the **V** method gave more accurate predictions of soil C and N concentrations, both with and without removing outliers. This is shown by larger RPD, r^2 and RER, and by smaller error (RMSECV) (Table 7.2). Two main reasons can be suggested for the improved accuracy of the **V** method. These are (a) the larger, vertical field-of-view (Figure 7.1) capturing the reflectance in the **V** method and (b) the spectral sample from the cylindrical surface area more closely represents the soil sample taken for laboratory analysis. Both account for the large and non-uniform variation in soil organic matter and root density with soil depth. The field of view for the **H** method is smaller and reflectance spectra acquired from a few mm thickness of soil surface do not account for the variation in the C and N concentrations determined in the 5 cm of section of soil taken for laboratory measurements. Indeed, the difference in the sample used for wet chemistry and spectroscopy is one of the major sources of calibration error (Mark and Workman 2003). Reflectance recorded from the curved surface produces less noise and possibly less specular reflectance. Specular reflectance contains little information from the surface materials. Evidence for reduced noise is slight. For example the root mean square noise (between raw and smooth spectrum of the first soil core) for the **H** and **V** methods were 0.001150 and 0.001104, respectively.

Table 7.2 Prediction values of soil C and N with and without removing outliers.

Method	Prediction values for C							
	n	Comp	r^2	RMSECV	RPD	RER	Bias	Slope
Horizontal [#)]	108	5	0.88	0.45 %	2.90	11.05	0.0030	0.92
Vertical [#)]	108	6	0.95	0.29 %	4.52	17.22	0.0042	0.96
Horizontal [##)]	103	5	0.95	0.27 %	4.45	18.47	-0.0001	0.95
Vertical [##)]	104	6	0.97	0.21 %	5.80	23.14	0.00066	0.97
	Prediction values for N							
Horizontal [#)]	108	3	0.90	0.04 %	3.11	12.27	0.00003	0.91
Vertical [#)]	108	5	0.95	0.03 %	4.52	17.84	0.00028	0.96
Horizontal [##)]	103	5	0.94	0.03 %	4.25	18.49	0.00022	0.96
Vertical [##)]	104	6	0.96	0.02 %	5.17	21.17	0.00040	0.97

Note: n = number of samples; Comp = component (factor or latent variable); [#)] without removing outlier; [##)] with removing outliers; Bias and Slope (see previous Chapters)

7.3.4. PLSR Coefficients Explaining Important Wavelengths

The PLSR model, using the first derivative of reflectance, for predicting soil C and N concentration by the **H** and **V** method is represented graphically in Figure 7.5. The size of the coefficient (positive or negative) represents the importance of the reflectance band in terms of accounting for the variance in measured soil C and N. In general, the graphical pattern of coefficients produced from the **H** and **V** methods of reflectance capture are similar. For example, for both methods the PLSR technique selects important bands from the visible region at 465 nm and 585 nm. This indicates that the degree of reflectance from these bands in the current sample population is influenced by soil C variation. Other important bands, appear in the NIR region (780-2440 nm), and are presented in Figure 7.5.

It is interesting to note that regression coefficients of the **V** method at bands 730-990 nm are higher than that of the **H** method, in predicting both soil C and N concentrations (Figure 5a and 5b). This indicates that the **V** method allows reflectance in this band to become more significant in building the PLSR calibration model for soil C and N prediction than with the **H** method. An important slope change (change in reflectance at 800 nm) in this band region for the **V** method is consistent with changes in soil organic matter content (Krishnan *et al.* 1980). In addition, Ben-Dor *et al.* (1997) show that spectral reflectance change between 680-800 nm can occur as organic matter ages during decomposition. Kusumo *et al*. (2009b; 2009a) report that changes in both root density and soil organic matter can affect reflectance between 830-940 nm. A further PCA was conducted on the first derivative data selected only from the 730-990 nm spectral region. PC1 accounted for 93.5% of the spectral variance in this range (data not shown) and correlated very strongly with changes in soil C concentration for both the **V** method (r = -0.93) and **H** method (r = -0.90). This is further evidence that the V method may allow the capture of spectra that contain improved information on changes in soil C and N concentration.

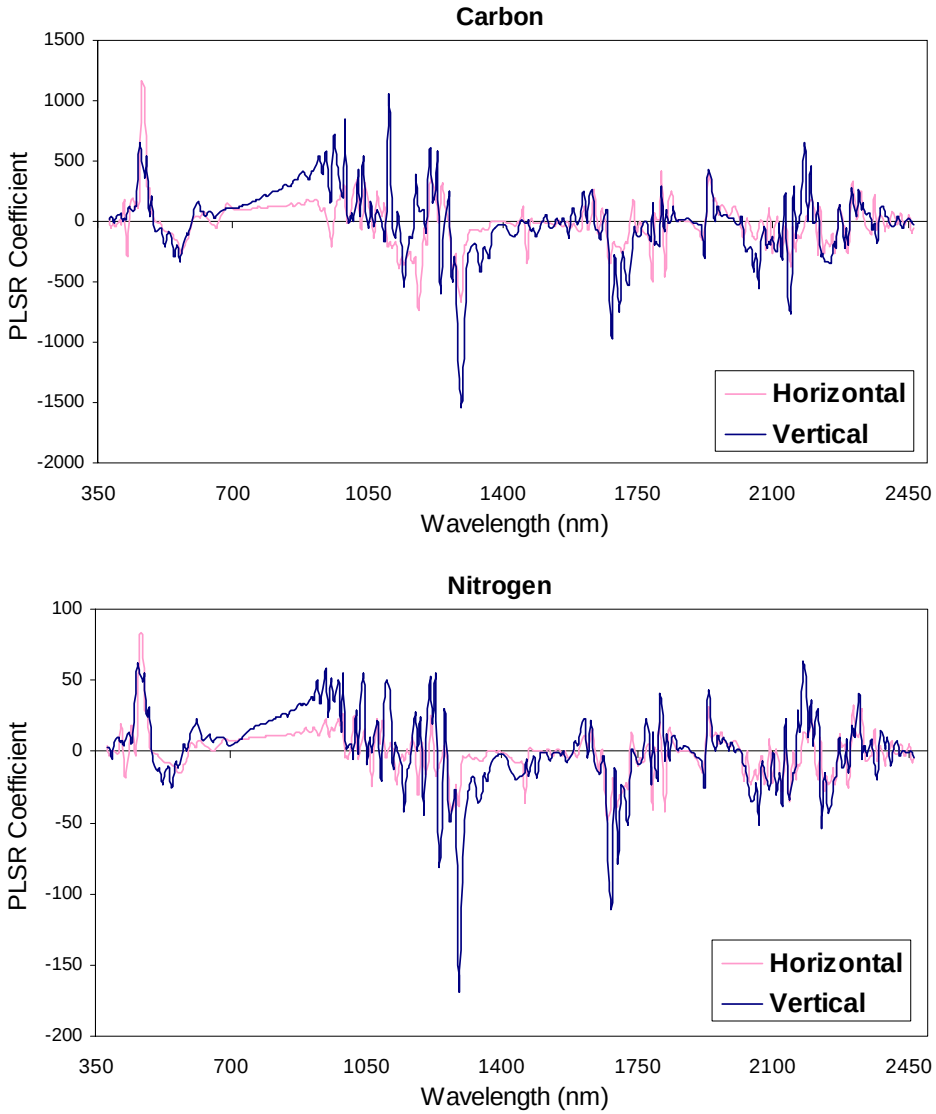


Figure 7.5 PLSR coefficients of the calibration model of (a) total soil carbon and (b) total soil nitrogen using reflectance spectra acquired by the horizontal and vertical methods.

7.3.5. Validation of the V Method Using a Ratio of 1:1 and 3:1 of the Calibration and Validation Sets

The **V** method gave very accurate predictions of soil C and N concentrations when the calibration model was tested internally using the leave-one-out cross validation procedure. To further test the robustness of the **V** method, the data was divided into separate calibration and validation sample sets at ratios of 1:1 and 3:1. Allocating samples into calibration and validation sets was achieved by arranging the sample data array from the smallest to the largest C-LECO (or N-LECO) values. Then either alternate rows (1:1) or 3 out of 4 rows (3:1) were allocated to the calibration set and the

remaining values to the validation set. The prediction accuracy of soil C and N concentrations using this external test (Figure 7.6) is still excellent (Malley *et al.* 2004). The larger RPD is obtained when the ratio of calibration and validation set is increased from 1:1 to 3:1. The larger coverage of spectral variance in the calibration set at the 3:1 ratio than that of the 1:1 ratio probably account for this.

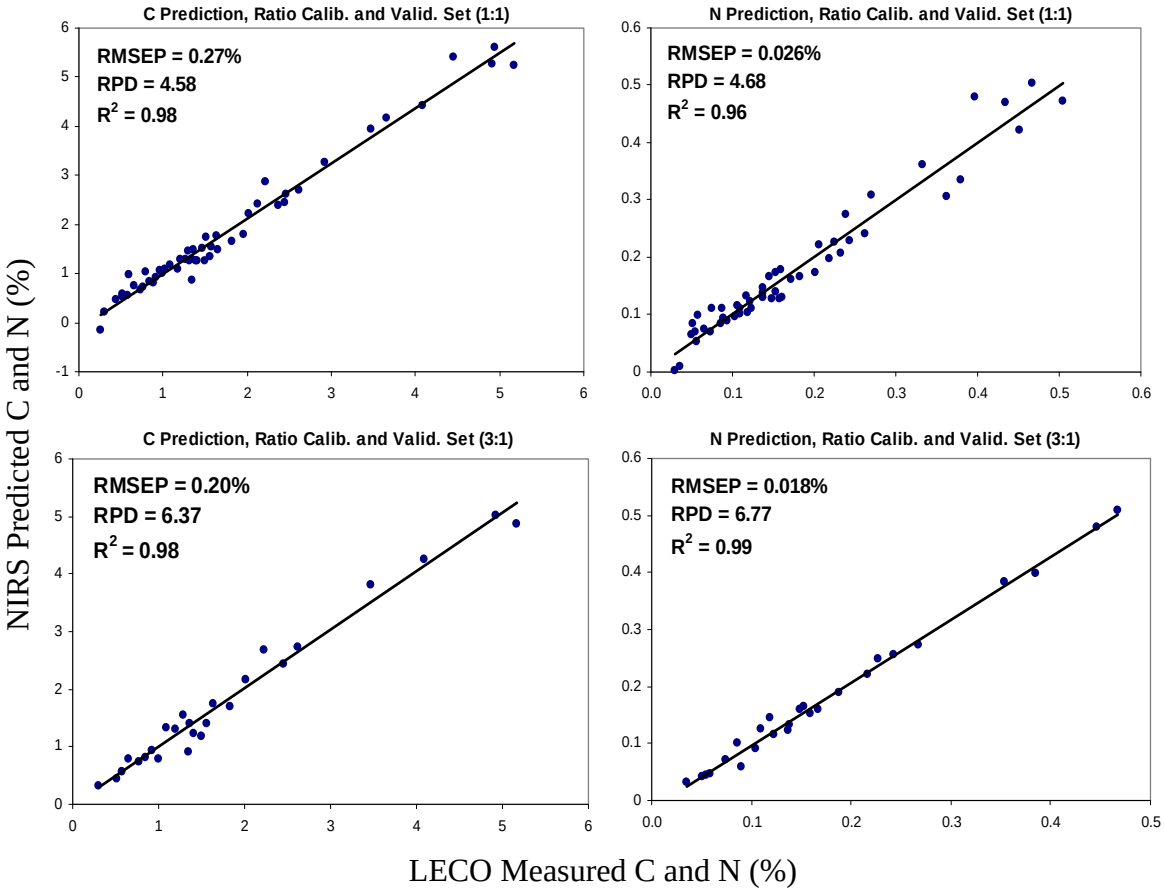


Figure 7.6 Prediction accuracy using different ratios of data in the PLSR model calibration and validation sets.

Figure 7.7 illustrates the average of laboratory measured soil C and N concentrations plotted in relation to soil depth and those predicted by the PLSR calibration model obtained using the **V** method. Accurate prediction of soil C and N concentration from spectral reflectance is shown at each depth, from the large C and N concentrations in the topsoil, to the small concentrations in the subsoil. This indicates that the wide range of soil chromophores that relate to changing soil C and N concentrations are "sensed" in the reflectance spectra and can be extracted by the PLSR technique and usefully used to predicted soil C and N concentrations with excellent accuracy (RPD > 4.0 , r^2 > 0.95 and RER > 20, Malley *et al*. (2004)).

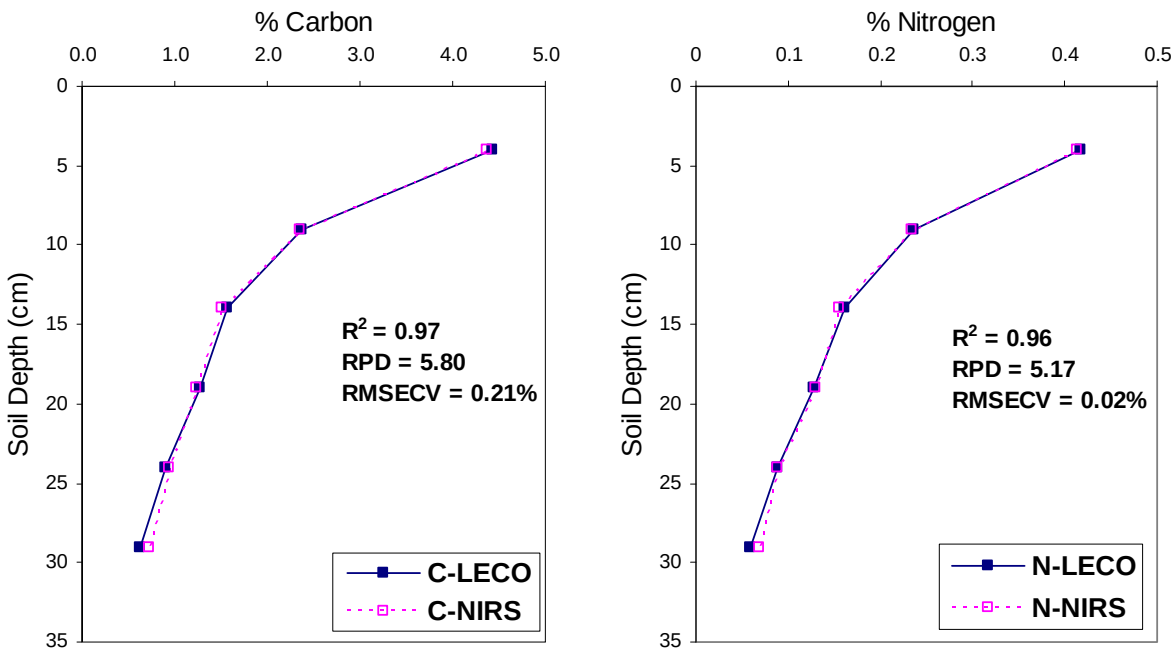


Figure 7.7 The average laboratory measured (LECO) total C and N concentrations at different soil depths and those predicted from soil spectral reflectance (NIRS) using the V method. Values are predicted using leave-one-out cross-validation procedure.

7.4. Conclusion

We conclude that rapid *in situ* assessment of soil C and N concentrations can be achieved using information in the reflectance spectra acquired from soil cores using a backpack spectroradiometer. Soil C and N concentrations can be predicted with an excellent level of accuracy (RPD > 5.0, r^2 > 0.95 and RER > 20) when the spectra are acquired from the curved vertical wall of a cylindrical section of soil core (**V** method), which represents the same sample position and dimensions that are taken to provide the reference set of laboratory determined soil C and N concentrations. In future studies, the **V** method is recommended over the **H** method because the spectroradiometer's field of view can be more easily matched to the dimensions (and therefore account for variability in soil properties with depth) of the soil core sample taken to provide the reference laboratory determined soil C and N analyses. In addition the V method may allow the capture of spectra that contain improved information on changes in soil C and N concentration but this attribute requires further testing on more soil types. This study provides further evidence that careful acquisition of soil reflectance spectra in the field can be used for rapid highly accurate mapping of the horizontal and vertical distribution of soil C and N.

CHAPTER 8

Temporal Robustness of Calibration Models in Predicting Soil Carbon and Nitrogen Concentrations from Proximally Sensed Soil Spectral Reflectance

Improved prediction accuracy for soil C and N concentrations was obtained (in Chapter 7) using "Vis-NIRS-soil core" technique that acquired spectral reflectance from curved vertical wall of cylindrical soil core (**V** method) compared to the flat horizontal cross-section of soil core (**H** method). In this chapter the robustness of calibration models developed from spectral data collected in different seasons and soil moisture conditions is investigated.

Some data from this study have been reviewed and presented in the Proceedings of the First Global Workshop on High Resolution Digital Soil Sensing & Mapping, 5-8 February 2008, Sydney, Australia (the Australian Centre for Precision Agriculture, University of Sydney and CSIRO Land and Water). An extended paper has been submitted and reviewed for a book chapter and will be published along with other papers from this conference. A research article summarizing the full results of this study will be submitted to *Australian Journal of Soil Research,* in 2009.

8.1. Introduction

Rapid methods for monitoring and auditing soil carbon change may be required by signatories to the Kyoto Protocol that adopt Article 3.4. This Article allows the offset of greenhouse gas emissions with verified increases in soil carbon. Increases that may result from soil management and/or land use change. The IPCC (Intergovernmental Panel on Climate Change) recommends soil carbon changes are monitored to 30 cm soil depth. Research required to identify agricultural and forestry management systems that lead to soil C changes will also require rapid and inexpensive field techniques for measuring SOM content changes (Kimble *et al.* 2001; Lal 2008; Paustian *et al.* 2000; Smith 2008).

Recent research (e.g. Chang and Laird 2002; Moron and Cozzolino 2004) has developed methods for using near infrared reflectance spectroscopy (NIRS) to predict

soil C and N concentrations. Kusumo *et al*. (2008a) have extended that methodology to *in situ* rapid measurement of soil C and N in soil profiles under permanent pasture and maize (Chapter 4 and 5). Improved accuracy in predicting soil C and N concentrations (Chapter 7) has been gained by acquiring the reflectance spectra from the curved vertical wall of soil cores rather than flat horizontal surfaces cut across the soil core; a technique used in the earlier studies (Chapter 6; Kusumo et al 2008a). The more accurate calibration model derived using the vertical method (**V** method) results from the spectra being acquired from the circumference of the same soil core that is taken for laboratory measurement of C and N concentrations (i.e. the spectral data and the laboratory reference measurement are made on the same soil sample).

Although predicting soil carbon and nitrogen concentrations from Vis-NIR reflectance spectra can be rapid and inexpensive, field spectroradiometers are expensive and the calibration of spectral data against reference data can also be expensive. The smaller the number of data sets required for calibration, the more cost effective field spectrometry will be. Calibration models for soil C and N may vary because the soil reflectance spectra may alter slightly depending upon small changes in soil type, sample location, seasonal soil moisture differences and sample preparation techniques. The cost of developing a calibration model can be expensive if each time this technique is used it needs a new calibration model. Some researchers (e.g. Bogrekci and Lee 2004; Kooistra *et al.* 2003; Malley *et al.* 2002) have shown less accuracy of calibration models due to changes in soil moisture, although others have been able to accurately predict soil C and N using moist soil samples (Chang *et al.* 2005; Fystro 2002; van Vuuren *et al.* 2007).

This chapter examines how seasonal changes, mostly resulting from soil moisture differences but also small changes in location of sampling site, influence the accuracy of calibration models used to predict soil C and N concentrations to 30 cm soil depth.

8.2. Materials and Methods

8.2.1. Contact probe and spectral acquisition

The details of acquiring soil reflectance spectra using the **v**ertical method (Figure 8.1) were described in Chapter 7, Section 7.2.1. In brief, reflectance spectra were

acquired from the curved vertical wall of a cylindrical soil core. The soil core was rotated 360^{o} to get the field of view 4707.4 mm^{2} cylinder (Figure 8.1).

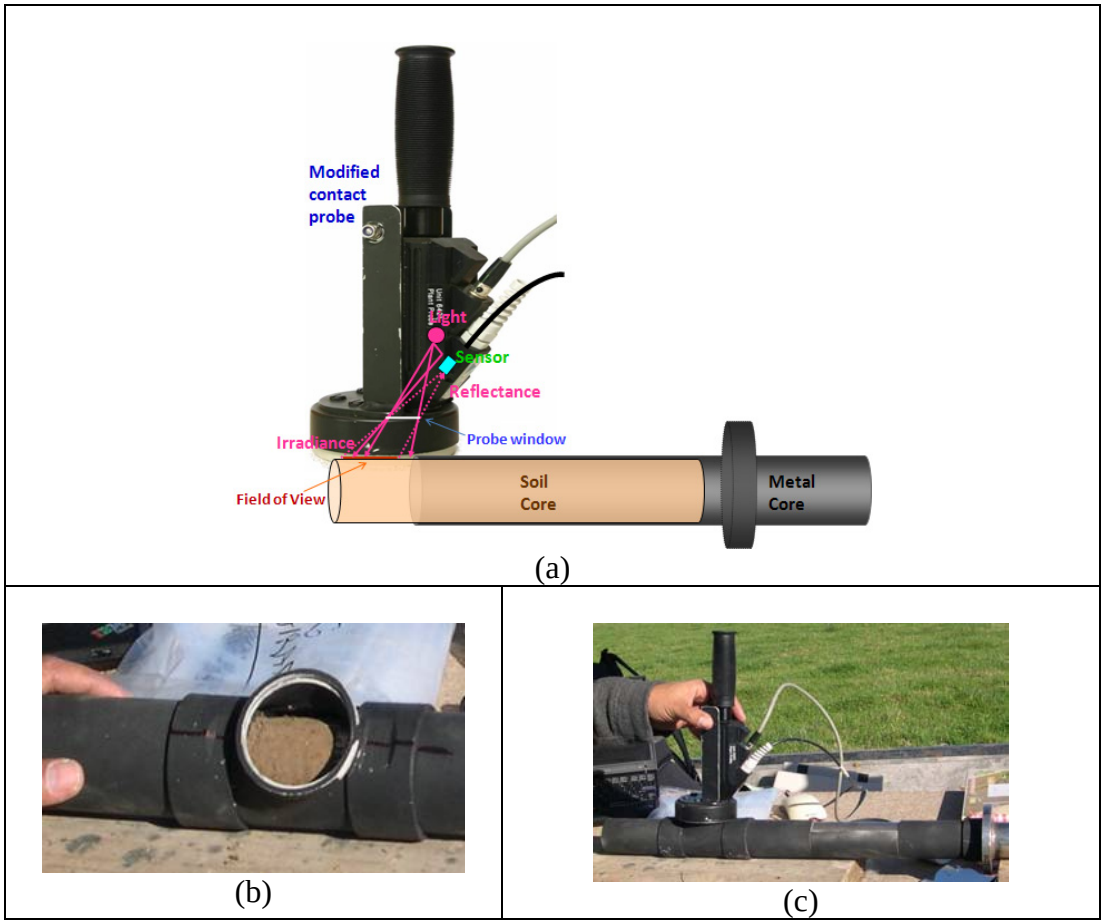


Figure 8.1 (a) The soil probe and template used to collect reflectance spectra from the curved vertical wall of a soil core, and (b) positioning template for (c) vertical acquisition of spectral reflectance.

8.2.2. Site Locations and Sample Collection

The soil cores used for measurement of reflectance and laboratory determination of soil C and N concentrations were taken from a permanent pasture on Fluvial Recent soil, in the Manawatu region, New Zealand, in May and November 2007. In May, a total of 17 soil cores were sampled at 6 depths at 5 cm intervals from 1.5 to 31.5 cm depth, totalling 102 samples. In November, a total of 18 soil cores were sampled at the same depths and intervals, totalling 108 samples. The soil cores were collected from 3 transects that were 20 m apart, with 6 cores in each transect at 15 m intervals. Field soil reflectance spectra were acquired from the curved vertical side of a 5 cm cylindrical soil core, which was subsequently analysed in a laboratory for reference data.

8.2.3. Measurement of Soil Properties

Soil cores were weighed field moist, crumbled and allowed to air dry before reweighing. Soil moisture content was expressed as a fraction of air dry weight. Air dry soils were then ground to < 500 µm particle size for total C and N analysis using the dry combustion method of a LECO FP-2000 CNS Analyser (LECO Corp., St Joseph, MI, USA) (Blakemore *et al.* 1987).

8.2.4. Spectral Pre-processing and Development of Calibration Model

Ten replicates of the reflectance spectra were acquired from each soil sample and the spectral data were pre-processed as described in Chapter 7 Section 7.2.4. A PCA score plot was used to observe the pattern of sample scattering. PLSR was used to develop calibration models from the pre-processed spectral data against the reference analytical data (LECO-measured C and N). During PLSR processing, samples which had a standardized residual > 2.0 were removed as outliers (MINITAB Inc. 2003) from the calibration and validation set. The accuracy of the models was tested internally using leave-one-out cross-validation method and externally using a separate validation set.

8.2.5. Regression Model Accuracy

The ability of the PLSR model to predict soil properties was assessed using the following statistics. RMSE (root mean square error) is the standard deviation of the difference between the measured and the predicted soil property values. RMSE which is calculated from cross-validation is called RMSECV, and from validation data is RMSEP. RPD (ratio of prediction to deviation) is the ratio of the standard deviation of measured value of soil properties to the RMSE. RER (ratio error range) is the ratio of the range of measured values of soil properties to the RMSE. The best prediction model is shown by the largest RPD, RER, r^2 and the smallest RMSECV or RMSEP; for more detailed explanations of these statistics see Chapter 3 and 6, Section 3.2.7 and 6.2.6 or Kusumo et al. (2008a).

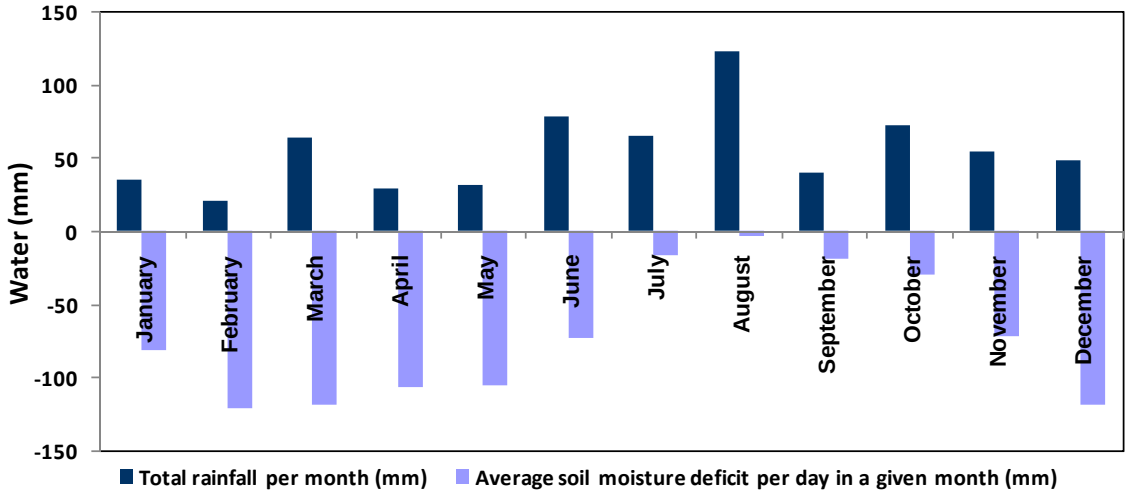
8.3. Results and Discussion

8.3.1. Summary of Soil Properties

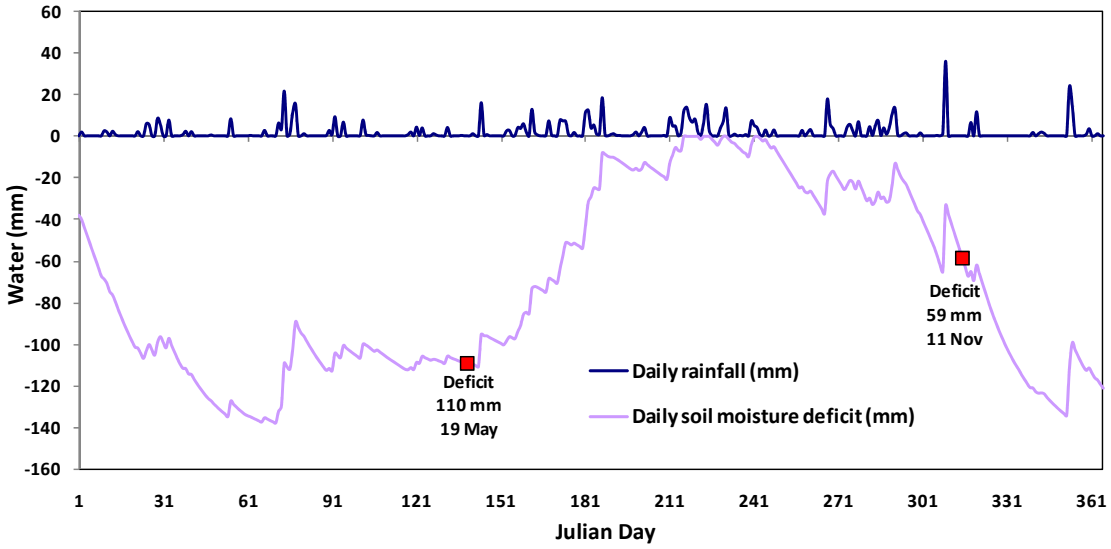
A summary of the laboratory analysis of soil C, N and water contents is presented in Table 8.1. May and November data show relatively similar ranges for mean and variance of total C and N concentrations. During the interval between May and November the pasture was grazed by rising 1 and 2 year old Friesian, dairy heifers. There was normal dung and urine return and a small amount of hoof treading damage but no cultivation during this period that would have disturbed the soil profile. In contrast to the small seasonal differences in soil C and N, the range and mean of water content from the November samples (end of spring) are larger than the May samples (autumn). Climatic data (Figure 8.2) show that November was a wetter month than May (total rainfall 55.4 mm cf. 31.6 mm) (Figure 8.2a) with a smaller soil moisture deficit (59 mm cf. 110 mm) on the day of sampling (Figure 8.2b). Between the two samplings (May and November) the soils have experienced water saturation in August (221 – 230 Julian days) when the hoof treading damage occurred.

Table 8.1 Soil properties of all 102 (May) and 108 (November) soil samples.

Soil Properties	Range		Median	Mean	Variance
	Minimum	Maximum			
	Data collected on May				
Total Carbon (%)	0.31	5.43	1.41	1.75	1.508
Total Nitrogen (%)	0.02	0.49	0.12	0.16	0.014
Water Content (%)	11.5	39.7	20.7	21.6	27.1
	Data collected on November				
Total Carbon (%)	0.27	5.20	1.41	1.85	1.681
Total Nitrogen (%)	0.02	0.50	0.14	0.18	0.015
Water Content (%)	20.7	53.6	28.6	30.7	39.7



(a)



(b)

Figure 8.2 Total rainfall (mm) and soil moisture deficit (mm) (a) monthly and (b) daily during 2007 collected from Palmerston North Airport station, the station closest to the experimental site (source: http://cliflo.niwa.co.nz). The square markers (Figure 8.2b) are the soil moisture deficit when the soil samples collected.

8.3.2. Spectral Characteristics

The soil spectrum representing the average spectral data collected in May or November is shown in Figure 8.3. November data, which have higher water content than May (Table 8.1), show lower reflectance intensity (Figure 8.3). Decreased reflectance due to higher soil water content has been reported by previous workers (Baumgardner *et al.* 1985; Bowers and Hanks 1965; Demattê *et al.* 2006; Mouazen *et al.* 2005). The OH^- bond, abundant in water, is known as the strongest NIR absorber (Malley *et al.* 2004); strongest absorptions occur near 1455 and 1930 nm. Slight

differences in C and N content between the May and November data (Table 8.1) are unlikely to cause the significant difference in reflectance intensity shown in Figure 8.3. Larger differences in organic matter quality or amount can reduce spectral reflectance intensity throughout the Vis-NIR bands (Bowers and Hanks 1965; Demattê *et al.* 2003) or parts of the bands (500-1200 nm) (Mathews *et al.* 1973b) .

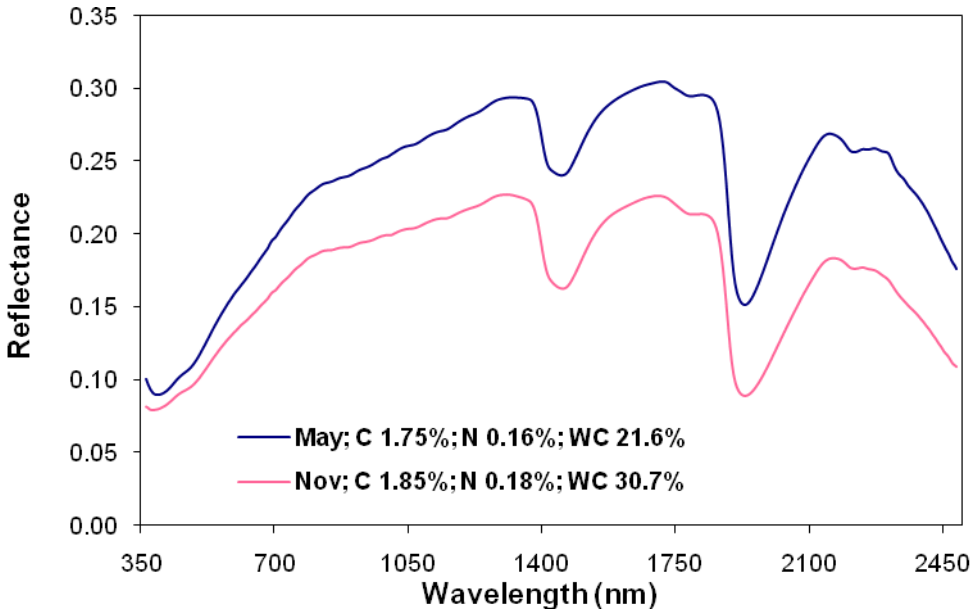


Figure 8.3 The average of the smoothed reflectance spectra for May and November sampling. C, N and water content (WC) in the figure legend are averaged from all samples.

The variances in the reflectance spectra sampled in May and November are represented in the score plot of the first two principal components (PC1 and PC2, Figure 8.4). Samples from 1.5-6.5 cm depth are clearly differentiated from the deeper samples. This is probably related to the higher organic matter content at the shallow depth, which is derived from pasture litter, roots and dung return. The scores of some samples from the 11.5-31.5 cm depths, and particularly the 11.5 – 21.5 cm depths, are tightly clustered indicating similar spectral characteristics for these samples with depth and across the paddock. It is expected that the chromophores such as water, soil organic matter, clay and non-clay minerals, iron oxides, and parent material (Baumgardner *et al.* 1985; Ben-Dor 2002) would be similar for these tightly clustered samples.

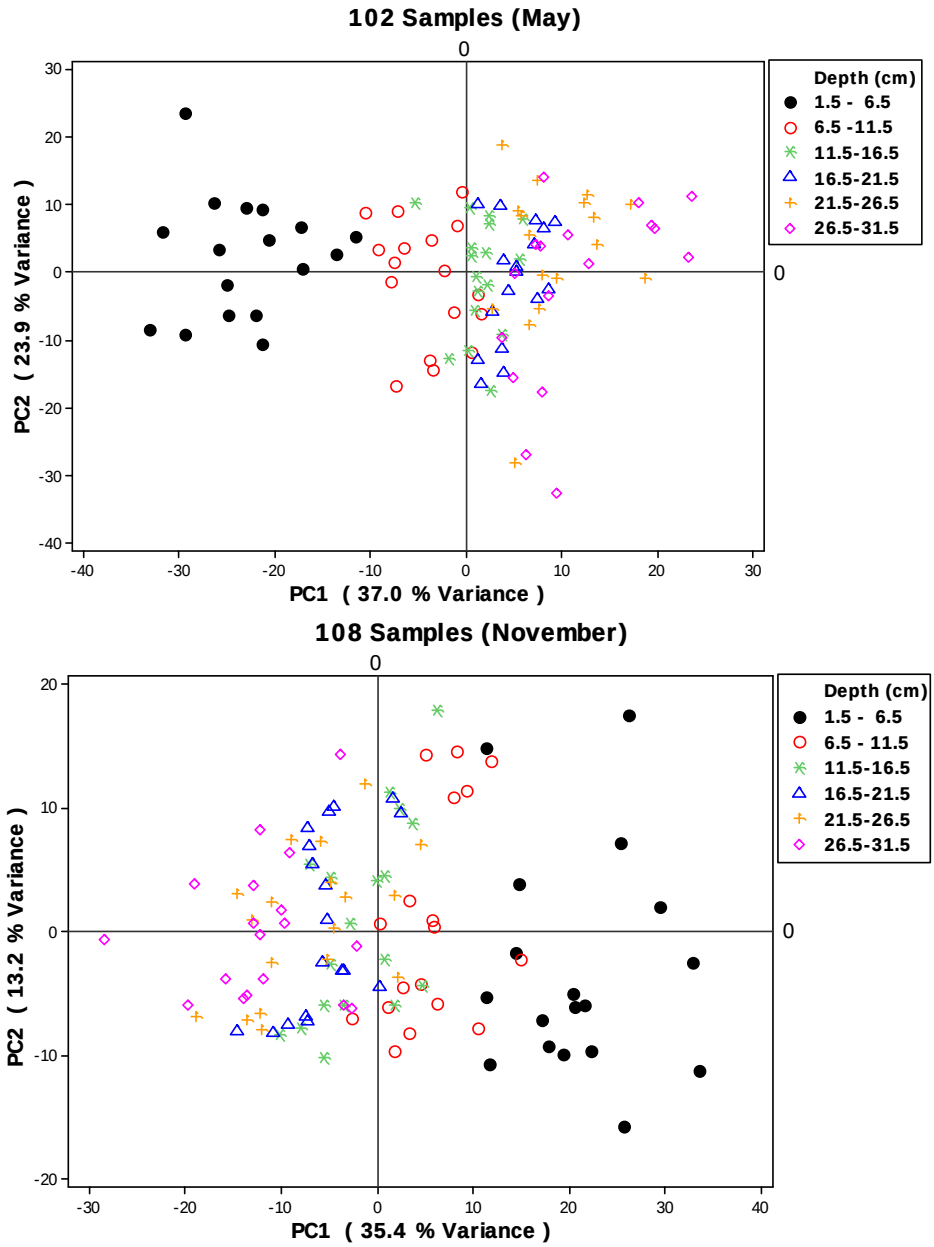
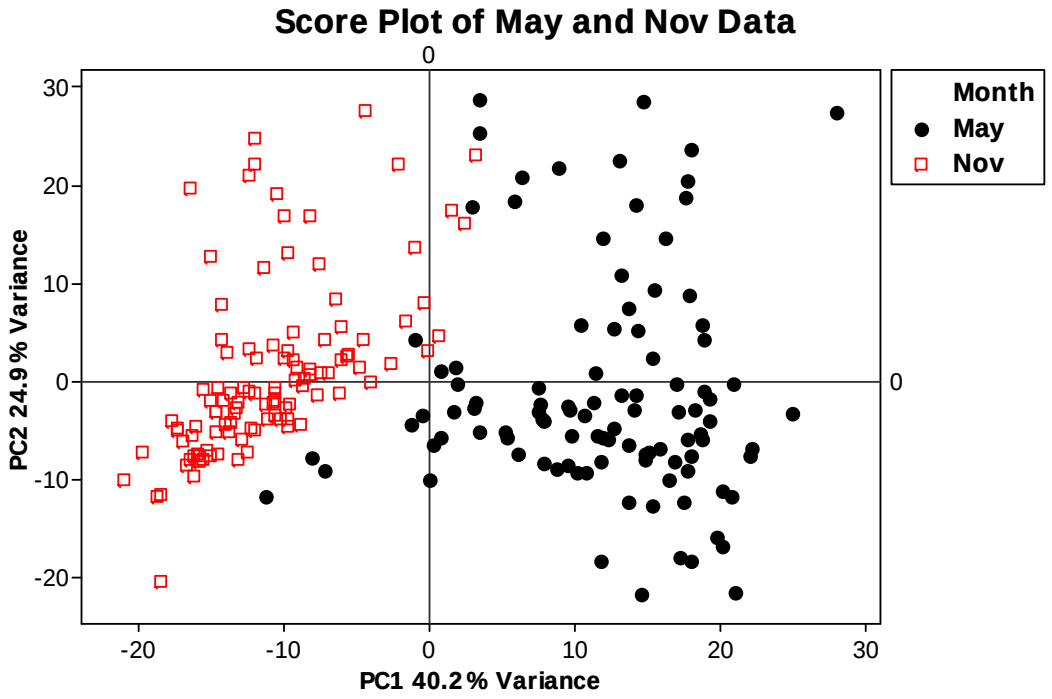


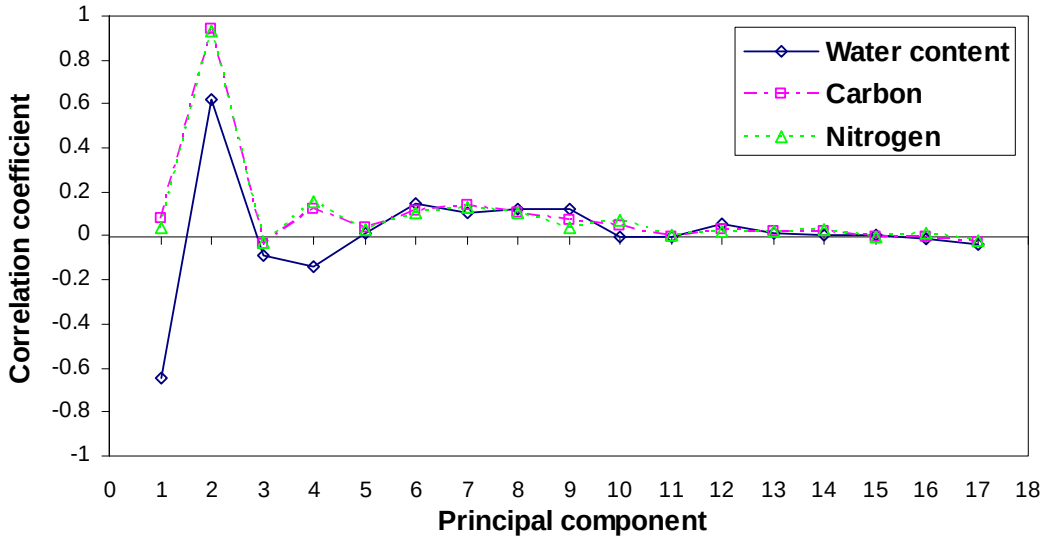
Figure 8.4 The score plot of first and second component from the PCA of reflectance spectra acquired from soil samples taken in May (above) and November (below) based on depth.

Obvious separation between May and November spectral data on the PCA score plot can be seen in Figure 8.5a. PC1 which accounts for the greatest spectral variance (40.2%) clearly separates May and November samples. There is a strong correlation (Figure 8.5b) between water content and PC1 (r = -0.65) and PC2 (r = 0.62) indicating that soil moisture content differences discriminates the two groups of data. Variation in soil organic matter content does not strongly influence PC1 (r = 0.08 with soil C; r =

0.03 with soil N; Figure 8.5b) but does strongly influence PC2 (r = 0.94 with soil C; r = 0.94 with soil N), which accounts for 24.9% of the spectral variance.



(a)



(b)

Figure 8.5 (a) Distribution of May and November samples on the PCA score plot and (b) correlation coefficient of each principal component with soil water content, soil C and N concentration.

8.3.3. C and N Prediction Using May and November Data Separately

The prediction of soil C and N concentrations using PLSR calibration models developed from each data set (May and November) separately are presented in Figure 8.6. Two calibration and validation techniques are used. Firstly all data in a sample set were used to build a calibration model and its validity was tested using leave-one-out cross-validation. Secondly, half the sample set was used to build a calibration model and its validity was tested by predicting the C and N content of the remaining data (external testing using a ratio of calibration to validation set of 1:1). The calibration models developed from spectral and laboratory reference C and N concentration data for the same sampling time (either May or November) allow accurate prediction of soil C and N concentrations (Figure 8.6). The accuracy is expressed by large RPD and r^2 and small error (RMSECV or RMSEP) (see Figure 8.6). Prediction accuracy with an RPD > 4 (with $r^2 > 0.95$ and RER > 20) can be categorised as excellent (Malley *et al.* 2004).

The prediction accuracy of the calibration conducted with data collected in May (RPD 5.95-7.31) is slightly more accurate than the November calibration (RPD 4.58-5.80). Drier conditions in May soil samples may account for this (see Table 8.1). Water may mask absorption features of overtones and combinations of bands associated with soil organic matter, which occur near the water bands (950, 1200, 1400, 1900, 2200 nm). This may then reduce the ability of NIRS to accurately predict other soil properties. Malley *et al.* (2002) reported less accurate prediction of organic matter and NH_4-N concentrations with field moist soils than with dry soils (dried at 40^oC). Kooistra *et al.* (2003) reported negative impact of water content on soil organic matter and clay content prediction. In addition, lower phosphorus prediction accuracy was reported by Bogrekci and Lee (2004) as the water content increased.

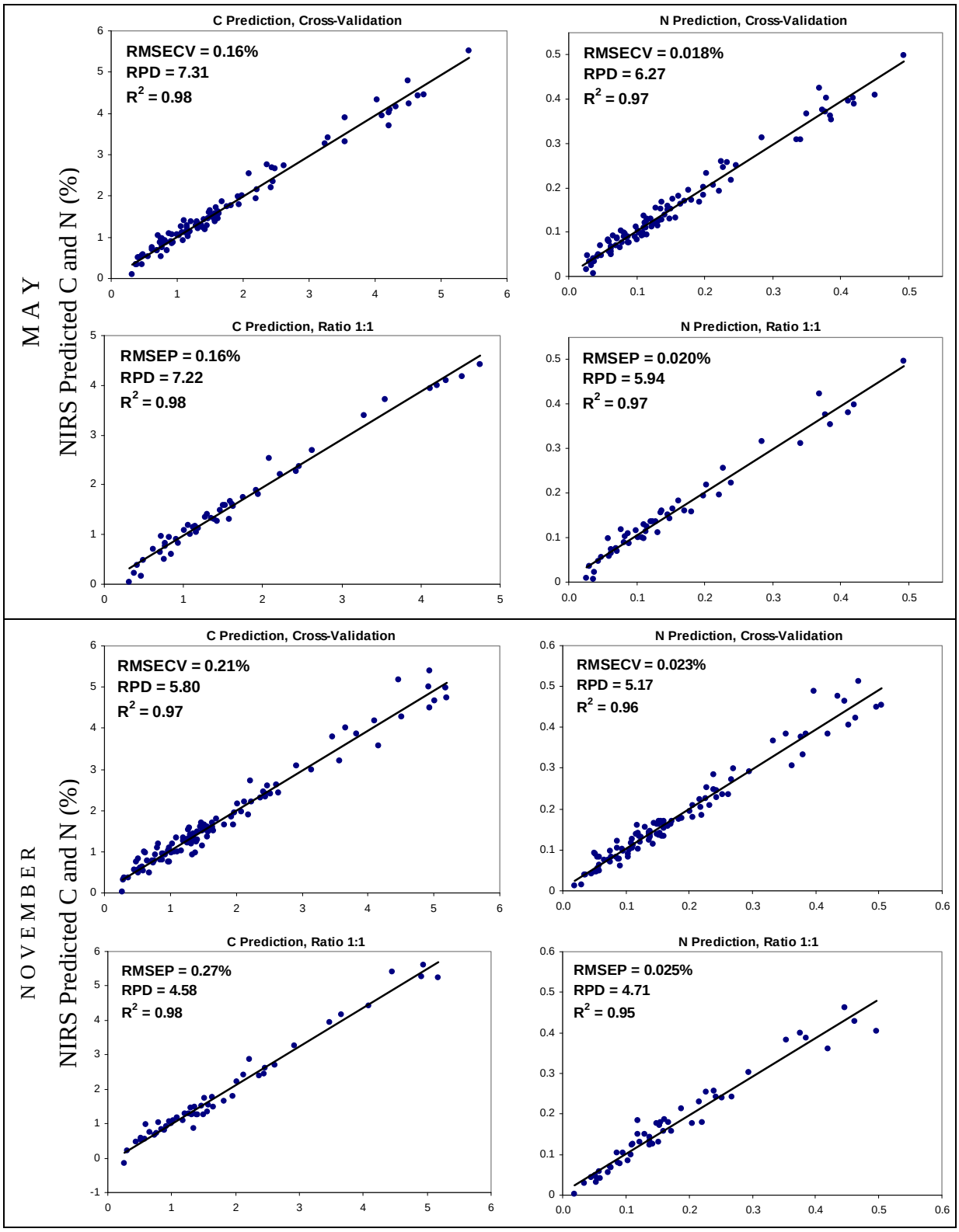


Figure 8.6 Regression of predicted and measured soil C and N contents of each sample population based on the time collection.

8.3.4. Is the Calibration Model Influenced by Temporal Variations in the Soil?

If the calibration of spectral data with soil C and N concentrations is stable over time then less effort will be required in re-calibrating the prediction model, which avoids cost and labour. It was considered that changes in soil colour (mottling), a feature of seasonal water table fluctuations, may significantly influence the robustness of the calibration model.

A PLSR calibration model built using spectral and soil C and N data collected in May was used to predict soil C and N concentrations using spectral data collected in November (Table 8.2 and Figure 8.7) and vice versa. Even though the core samples taken in the wet conditions in November were more mottled than those taken in dry conditions in May, accurate predictions of soil C and N concentrations were achieved. The accuracy of prediction is shown by large RPD, RER and r^2 (Table 8.2). Malley *et al*. (2004) categorise this prediction accuracy into "successful", when the RPD is between 3-4 (with r^2 0.90-0.95 and RER 15-20). Under the conditions and soil types found in this experiment it would appear that acquiring spectra using the **V** method provides "one season" (either May or November) calibration models for *in situ* prediction of soil C and N that are reasonably temporally robust. There is some loss of accuracy, however, when the "one season" calibrations are used to predict soil C and N concentrations measured in a different season (November or May) (cf. Figure 8.6 and Table 8.2).

Table 8.2 Statistical descriptors of soil C and N prediction when calibration models built for data collected in May are used to predict November data, and vice versa.

Soil properties	Calibration set		Validation set		----------Prediction values----------						
	Month	n	Month	n	Comp	RMSEP	r^2	RPD	RER	Bias	Slope
Total C (%)	May	100	Nov	104	3	0.36%	0.97	3.43	13.72	0.094	0.75
	Nov	104	May	100	6	0.36%	0.97	3.35	14.03	-0.100	1.20
Total N (%)	May	100	Nov	104	5	0.034%	0.95	3.44	14.13	0.007	0.78
	Nov	104	May	100	6	0.034%	0.96	3.42	13.97	0.009	1.16

Note: n = number of samples; comp = component (factor or latent variable); Nov = November; Bias = mean difference between measured and predicted soil C and N; Slope = regression coefficient between measured and predicted soil C and N

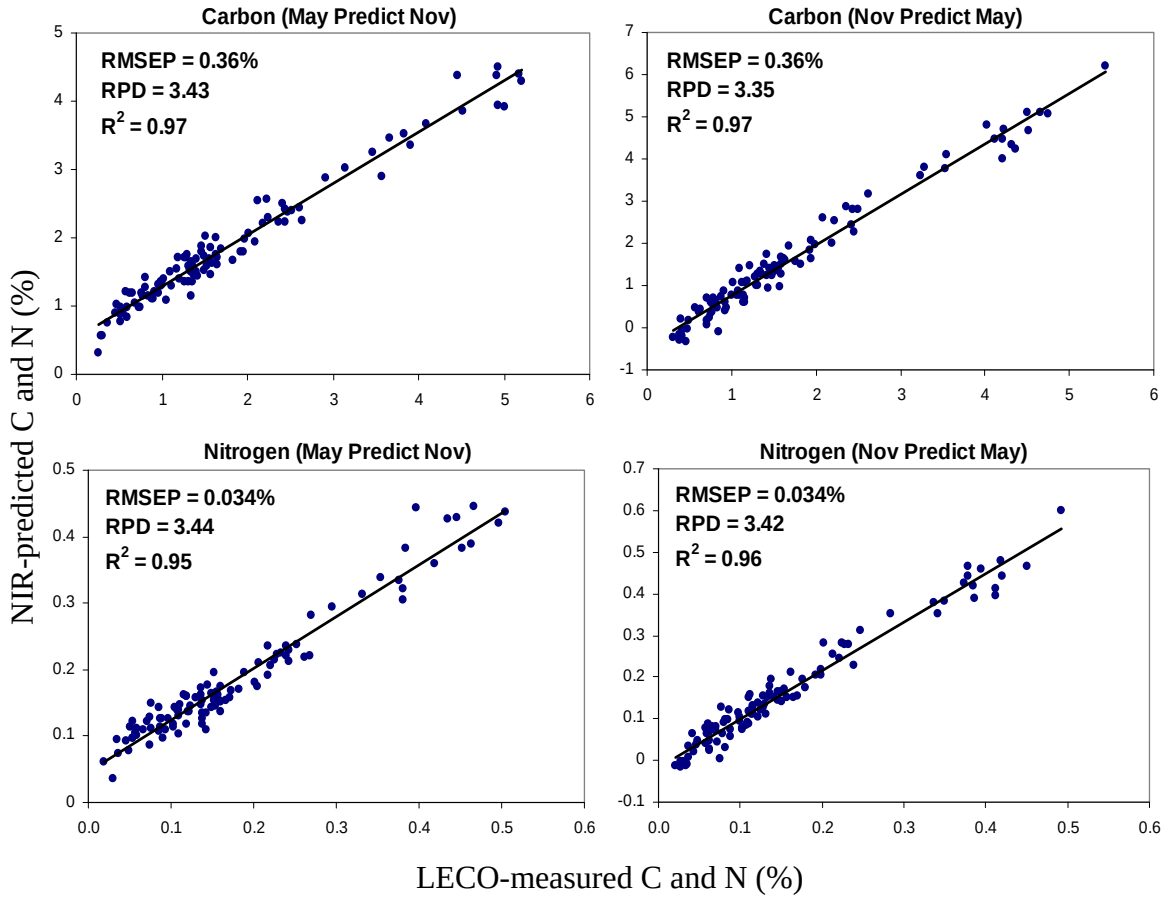


Figure 8.7 The May PLSR calibration model is shown predicting soil C and N concentrations from spectra acquired in November (and vice versa).

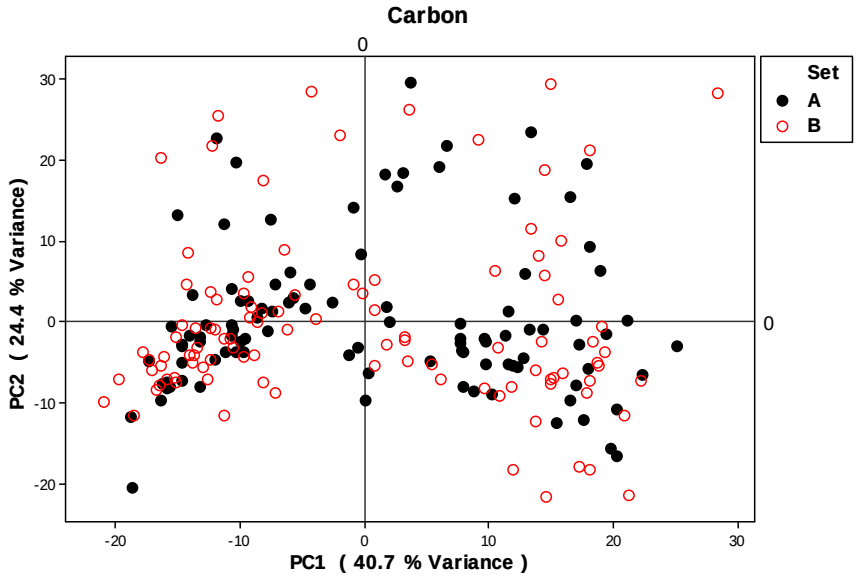
8.3.5. Should Unknown Samples Lie within the Spatial Distribution of the PCA Score Plot of the Calibration Set?

Despite the separate spatial distribution shown by the May and November samples on the PCA score plot, which was caused by differences in soil moisture content (Figure 8.5a), it was possible to produce a successful calibration model developed from the May data to predict the November data, and vice versa (RPD 3.35-3.44; r^2 0.95-0.97).

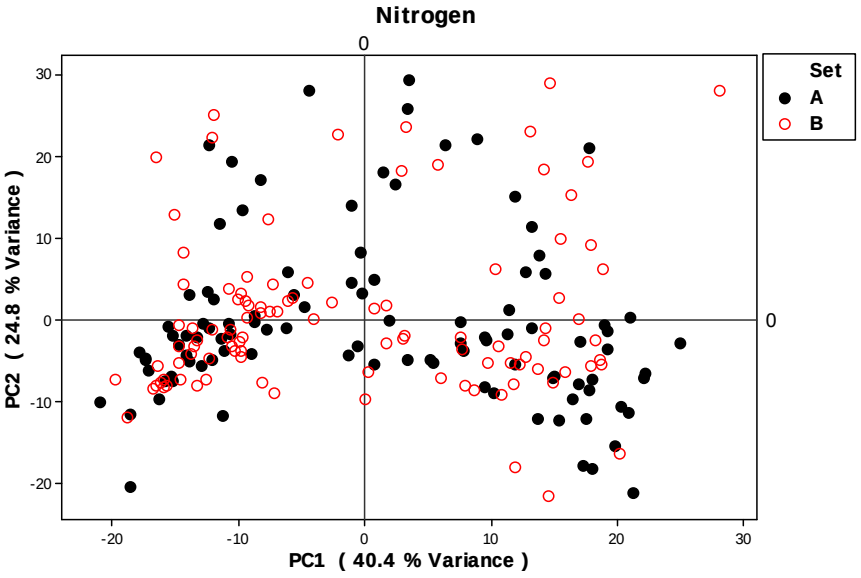
As previously discussed in Chapter 6 (the example with soil C and N prediction in Pumice, Allophanic and Tephric Recent soils), the question arises – can an improved calibration model be built when the calibration data set is carefully chosen to reflect the full spread (in PCA score plot) of the combined May and November data sets? To examine this, May and November samples were amalgamated and the samples were separated into two sets (A and B) by ranking them from the lowest to the highest

measured soil C (or N) values and then allocating samples with odd ranking numbers to set A and even numbers to set B. This resulted in a ratio of 1:1 data split between calibration and validation sets. The distribution of samples on the PCA score plot due to this process is presented in Figure 8.8.

When the distribution pattern of data in the calibration and validation set was similar (Figure 8.8) as suggested by Esbensen *et al*. (2006), better prediction of C and N was obtained (RPD 4.53-5.56; RMSEP for C 0.22-0.23%; RMSEP for N 0.026%; Figure 8.9), compared to when the calibration data was drawn from the May data and used to predict the November data set, and vice versa (RPD 3.35-3.44; RMSEP for C 0.36%; RMSEP for N 0.034%; Figure 8.7). This indicates that the spatial distribution of validation samples set on the PCA score plot should ideally lie within the calibration set distribution. In addition, the samples sets A and B shown in Figure 8.8 depict sample clusters (or neighbouring data values) that have similar spectral data dimensions. As shown in Chapter 6, improved prediction accuracy can be obtained if the calibration and validation set embrace the same variance dimensions (Williams 2001).



(a)



(b)

Figure 8.8 Sample distribution of set A and B on the PCA score plot.

An appropriate test for assessing whether a calibration model will give an accurate prediction of the attributes of an unknown sample is whether the spectral attributes of the unknown sample fall within the distribution of the calibration set on a PCA score plot (Kusumo *et al.* 2008a). If the unknown sample falls outside this range, then the reference attributes (e.g. in this case soil C and N determined) of the sample should be analysed in the laboratory. The laboratory analysis may identify the sample as a spectral

outlier (not a reference value outlier) that requires re-sampling. If the spectral and reference values are consistent outliers then this sample could be added to the calibration set when unknown samples with similar spectral attributes to it appear in the unknown sample set. Likewise, samples in the calibration set with spectral properties out of range of the unknown could be removed as new samples enter the calibration set. In this way calibration sets can be flexible and chosen to reflect the spectral properties of the unknown samples.

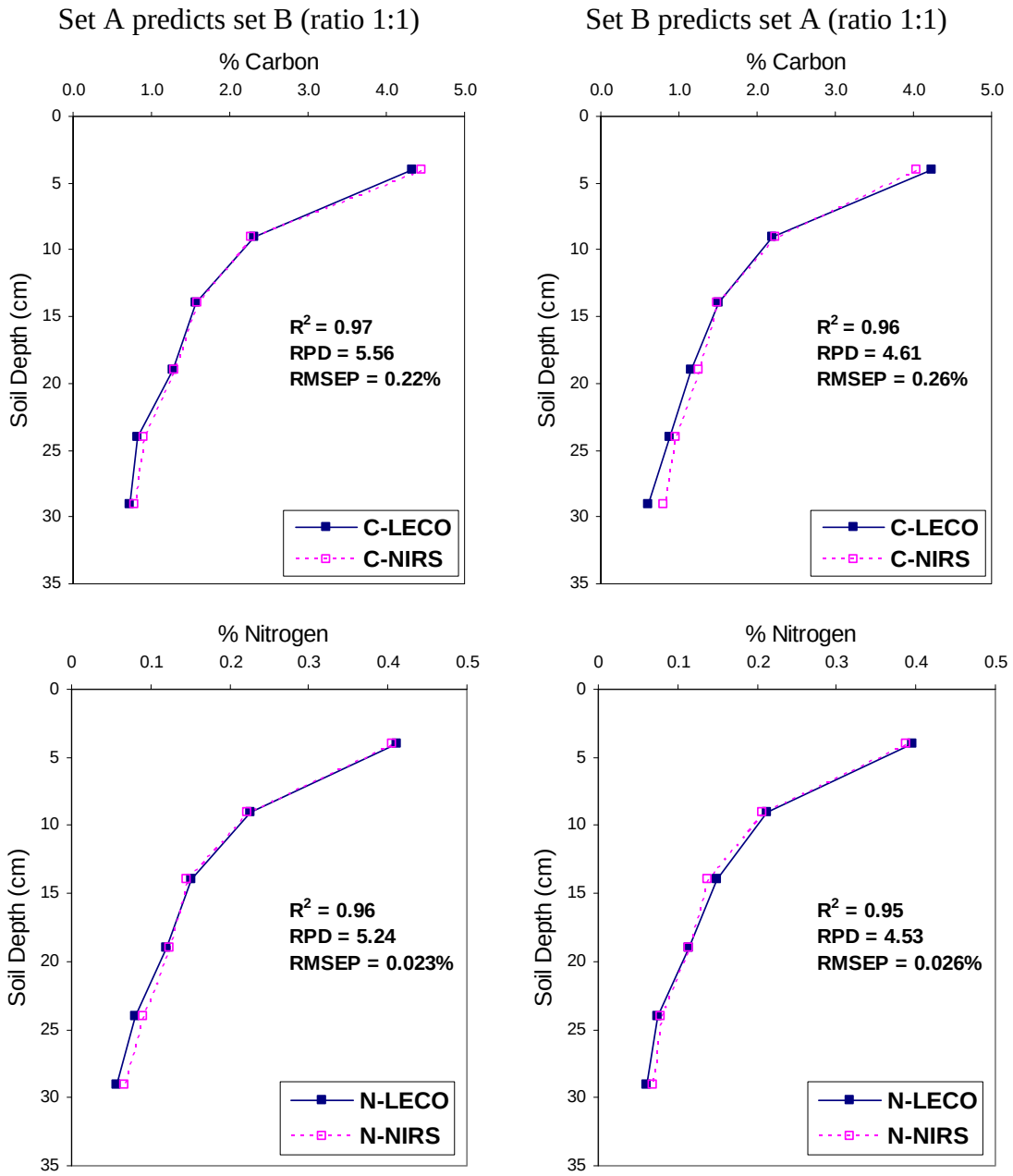


Figure 8.9 Mean of C and N in each soil depth measured by LECO and predicted by NIRS.

8.3.6. Soil Water Content Ranges in Calibration Set Data

The development of the combined data sets (set A and B, Figure 8.8) overcame the problem that calibration data sets derived separately from either the May or November data had contrasting soil water contents. The new calibration and validation sets (set A and B) drawn from the combined May and November data have descriptive statistics for water contents that are relatively similar [(WC set A; range 11.9-53.6%; mean 25.7%; median 26.2%) (WC set B; range 11.5-51.3%; mean 26.4%; median 26.8%)]. Data with similar water contents in the calibration and validation sets produce more accurate C and N prediction (Figure 8.9, RPD 4.53-5.56; RMSEP for C 0.22-0.23%; RMSEP for N 0.026%), compared to when drier May data (with WC range 11.5-39.7%; mean 21.6%; median 20.7%) were used to predict the wetter November data (with WC range 20.7-53%; mean 30.7%; median 28.6%) and vice versa (Figure 8.7, RPD 3.35-3.44; RMSEP for C 0.36%; RMSEP for N 0.034%). Thus for practical purposes, reflectance spectra to be used for *in situ* prediction of soil C and N concentrations should be acquired when field soil water contents are similar to (within the same range) the water contents when the calibration set data were acquired. The results presented in this paper indicate that soil C and N concentrations can be predicted successfully even when the range of water contents is relatively large (11-53%). However, this range is unlikely to extend to wetter conditions and further research is required to set maximum limits.

8.4. Conclusions

Calibration models developed from reference and spectral data acquired from curved surfaces of soil cores collected from drier soils in May produced successful predictions of soil C and N concentrations in wetter soils in November, and vice versa. This indicates that the spectral acquisition technique was a temporally robust method for *in situ* measurement of soil C and N for the Fluvial Recent soil in this study. Soil water content differences between the May and November samplings only slightly decreased the prediction accuracy of soil C and N concentration. Improved accuracy was achievable, however, when the data was manipulated so that both the calibration and validation data sets embraced the same spectral variance dimensions and similar water content ranges. To obtain the most accurate *in situ* field predictions of soil C and N concentration it is recommended that efforts are taken to ensure unknown samplings are

conducted at soil moisture contents that embrace the same range of soil moisture contents of the calibration data. The results, in conjunction with the findings of previous studies, continue to show that using portable near infrared reflectance spectroscopy (NIRS) to predict soil C and N concentrations has considerable potential for mapping and assessing the impacts of land-use change on soil organic matter stocks.

CHAPTER 9

Overall Summary and Future Works

This chapter summarizes the major research findings reported in this thesis related to the development of procedures for predicting root density and soil C and soil N concentrations from Vis-NIR reflectance spectra acquired from soils under field conditions. It concludes with some recommendations for future research.

9.1. Overall Summary

9.1.1. NIRS for Field Soil Property Measurement

The literature review identified that considerable progress has been made by other researchers on developing techniques for predicting a range of soil properties from soil spectral reflectance. Most researchers had concentrated their studies on and developed successful methods for the laboratory analysis of dried and sieved soils using near-infrared spectroscopy (NIRS). Only a small number of researchers had attempted to develop field methodologies. The most advanced of these, a truck or tractor-mounted, automated, soil probe for acquiring soil reflectance spectra (P4000 VIS-NIR-EC-Force Probe, Veris Technologies Vis-NIR soil probe and corer; http://www.veristech.com/products/probe.aspx; http://www.veristech.com/images/VerisP4000.pdf) is a very recent development. No research literature using this probe had been published in peer reviewed literature.

Research gaps existed in this area: No NIR spectroscopy research had been conducted on field soils in New Zealand. Globally, rapid and inexpensive field techniques for measuring SOM content are required to identify agricultural and forestry management systems that lead to soil C changes. These methods may be required to comply with IPCC protocol that requires soil carbon accounting to 30 cm soil depth. Such methods are required not only for easily accessible terrain but also for terrain only accessible on foot.

NIR soil reflectance is a result of interaction between NIR radiation and covalent bonds of small atoms (particularly O, H, C and N) of soil components. This makes NIR reflectance spectrometry sensitive to soil organic matter, soil water and hydrous oxide

groups of amorphous and crystalline clay minerals. These have been termed “soil chromophores”. NIRS has been used to successfully predict soil properties such as total C, organic C, total N, CEC and soil moisture content (Malley and Martin 2003). The most common technique for extracting relevant information from the complex spectral data (recorded at 1-2 nm band widths) is to smooth the spectra, calculate the first derivative over an expanded band width (5 to 10 nm) and use partial least squares regression to calibrate the first derivative spectral data to the reference soil property data. The calibration can then be used to predict the soil property of “unknown” soil samples that only have spectral data.

According to published literature the successful development of field methodologies required the solutions to the following problems that cause variation and interferences in the spectral data: excessive and varying soil moisture content, varying shapes of soil aggregates, varying soil textures, temporal change in soil chromophores such as hydrous oxides and root densities, and non-uniform soil surface to sensor distances.

Problems relating to spectral acquisition were addressed in Chapter 2 section 2.6. Inaccuracies likely to be caused by variable soil-to-sensor distances were overcome in this study by designing a soil probe which had fixed soil-to-sensor distance. The soil probe also has its own light source and is designed to eliminate the effect of ambient light. The probe, connected by a fibre optic cable to an ASD FieldSpecPro portable spectroradiometer, was designed to acquire soil spectral reflectance from flat cross-sectioned surfaces of soil cores (**H** method). Further improvements involved spectral acquisition from a larger field of view for better representation of field soil variation. This was achieved (Chapters 7 and 8) by building a template to acquire soil spectral reflectance from the curved vertical surface of soil cores (**V** method). A larger field of view was achieved by rotating the soil core. This system of spectral acquisition allows the spectral reflectance to be obtained from the same soil sample dimensions as that taken for reference soil property measurement. Results showed that improved calibrations could be achieved to predict soil C and N concentrations using this technique.

9.1.2. Root Density Predictions

Despite the importance of root dynamics in soil organic matter formation, no studies could be found in the literature where Vis-NIRS had been used to predict plant root densities or follow root dynamics. Early experimental work (Chapter 3), demonstrated that Vis-NIR reflectance was sensitive to root density change. Vis-NIRS plus PLSR produce calibration models capable of predicting ryegrass root densities in a pot trial involving Allophanic and Fluvial Recent soils. This successful result was repeated in field studies (Chapter 4) that showed that root densities could be predicted in permanent pastoral soils (Ramiha silt loam (Allophanic) and Manawatu fine sandy loam (Fluvial Recent). More accurate root density prediction was achieved when calibration sets were developed separately for each of the two contrasting soils and their contrasting root densities. Better prediction accuracy was obtained in the Allophanic soil (r^2 0.83, RPD 2.42, RMSECV 1.96 mg g^{-1}) than that in Fluvial Recent soil (r^2 0.75, RPD 1.98, RMSECV 5.11 mg g^{-1}). Interestingly, root density can be predicted in the field, despite the presence of various interfering soil chromophores. For example, interferences between root content and iron oxides were seen in reflectance bands near 900 nm. Unfortunately, the calibration model developed to predict root densities in the glasshouse data could not be used to predict root density from field spectral data. A clear difference was seen in the spectral attributes of the glasshouse and field soils. This was partly because of the widely different physical nature of the field and laboratory soils, and partly because the range of the glasshouse root densities did not cover the field observations; and also spectral properties of the glasshouse calibration data do not embrace the field spectral variance. This indicated that the development of a calibration model should be part of the field sampling programme.

Vis-NIRS was also successful in predicting maize root density in arable soil (Chapter 5). The assessment was carried out on Gley Soils (Kairanga silt loam and fine sandy loam) under 90-day old maize. The Vis-NIRS technique was able to describe root distribution within the soil profile and was sensitive to the differences in root density between the two texturally differing soil types. Interestingly, because measured maize root density was poorly correlated with the distribution of measured soil C and N content, it was possible to show that different spectral attributes were used in the PLSR

calibration model to predict root density compared with those used to predict soil C and N concentrations.

The ability of Vis-NIRS technique developed in this thesis to predict root density, soil C and soil N from deep subsoil will allow future researchers to explore deep rooting plant species that may add soil C deep in the soil, where it is known to have longer residence time. This ability will be important in designing land management strategies to sequester soil C, a partial solution to global warming.

9.1.3. Improved Procedures for Selection of Calibration and Validation Data Sets

Results in Chapter 6 showed that soil C and N concentrations in selected soils (Allophanic, Pumice and Tephric Recent) could be predicted using Vis-NIRS (**H** method), in which soil reflectance spectra were acquired from the flat horizontal cross-section of soil cores. Moderate accuracy of C and N prediction was obtained in this study, which is probably because of the coarse soil textures and the high soil water contents of most of the samples.

Various selection methods were used to compile the calibration and validation data sets. The selection methods presented were based on the near neighbor sample pairs (soil type, location, transect, site and depth), near neighbour reference analysis values (based on C and N analyses) and near neighbour attributes in spectral variance using distribution on PCA score plot. The latter technique was used in Chapters 6 and involved firstly carrying out PCA analysis on spectral data of both “known” and “unknown” samples. The known samples have associated reference data (e.g. laboratory measured soil C) and are the spectral data used to develop the calibration model. Then, the spectral data of the “unknown” samples are plotted together with the “known” samples on the score plot. If the spectral data of “unknown” samples are spatially distributed within the “known” samples, good prediction of soil C, N and root density may be obtained.

The exercise proved that, when there is no prior knowledge of sample reference measurements for the “unknown” sample set, improved prediction accuracy is likely to be achieved when the calibration and unknown data sets embrace the same spectral

variance dimensions in the PCA score plot. A similar conclusion was drawn in Chapter 8 when choosing samples to predict soil C and N concentrations in Fluvial Recent soils under permanent pasture. The major conclusion was that, improved prediction of soil C and N concentrations was achieved when both the calibration and validation data sets embraced the same spectral variance dimensions and similar water content ranges. To obtain the most accurate *in situ* field predictions of soil C and N concentrations it is recommended that efforts are taken to ensure "unknown" samplings are conducted at soil moisture contents that fall within the same range of soil moisture contents at which the calibration data was acquired.

9.2. Implications for Future Works

The experimental work conducted in this thesis was constrained by the recommended period of study for a PhD thesis. There are a number of areas of research into the use of Vis-NIRS for predicting soil properties that require further investigation.

- Although Vis-NIRS has been successfully tested to measure root density in the glasshouse study and in the field, the accuracy of this technique for field routine measurement should be further investigated. So far, root density measurement has been done in three soils of different orders (Allophanic, Fluvial Recent and Gley Soils). Our knowledge of the ability of Vis-NIRS to measure root content will be more complete if the root density assessment is conducted in a wider range of soil types. This technique should be tested also for soils under other crops (in addition to ryegrass grass and maize). The number of samples collected should be large enough to provide samples for a calibration and validation set. The robustness of root prediction model should be also tested in different seasons.

- All field root content predictions in this thesis were carried out using the **H** method (acquiring the spectral reflectance from the flat cross section of a cylindrical soil core). Even though we recommend the use of the horizontal technique to measure root density in point sampling thin sections of soil, the accuracy of the vertical method for field root measurement should be investigated. The larger sample size involved in this technique may be more appropriate when greater depths of the soil profile are considered.

- Vis-NIRS prediction of soil C and N has been conducted in a range of New Zealand soil types, classified as Pumice, Allophanic, Tephric Recent and Fluvial Recent soils. Its value needs to be tested in other dominant soil types in New Zealand. Further work should use the vertical technique (**V** method) and should involve an investigation of the appropriate selection of data to build calibration models appropriate for predicting soil properties in a soil type, or a soil order, or groups of soil orders. In addition, predicted soil properties may be extended to the measurement of soil TP, NO_3^-, pH, and CEC. Although these are soil properties that are only indirectly related to the chromophores active in the visible and near infrared wavelengths, previous researchers have reported various successes of prediction (poor to good) for these properties by NIRS.

- Because soil water content influences the accuracy of NIRS measurement, further investigations should focus on the maximum range of field soil water contents that allow accurate prediction of soil properties by NIRS.

- Finally work needs to begin on turning this essentially manual coring technique (**V** method) into a user friendly, automated process. To build accurate calibration models there is a need to ensure that the spectra used for calibration are acquired from the same sample of soil as the reference measure of the soil property. An issue that still has to be addressed by other researchers.

References

Aber J, Wessman CA, Peterson DL, Mellilo JM, Fownes JH (1990) Remote sensing of litter and soil organic matter decomposition in forest ecosystems. In 'Remote sensing of biosphere functioning'. (Eds RJ Hobbs, HA Mooney) pp. 87-101. (Springer-Verlag: New York Berlin Heidelberg London Paris Tokyo Hong Kong).

Adam MJ (2004) 'Chemometrics in analytical spectroscopy.' (The Royal Society of Chemistry: Cambridge, UK).

Adamchuk VI, Hummel JW, Morgan MT, Upadhyaya SK (2004) On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture* **44**, 71-91.

Adnan N, Ahmad MH, Adnan R (2006) A Comparative study on some methods for handling multicollinearity problems. *Matematika* **22**, 109-119.

Analytical Spectral Devices (1999) 'Analytical spectral devises, Inc. (ADS) technical guide.'

Anderson RJ, Bendell DJ, Groundwater PW (2004) 'Organic spectroscopic analysis.' (The Royal Society of Chemistry: Cambridge, UK).

Baldock JA, Nelson PN (2000) Soil organic matter. In 'Handbook of Soil Science'. (Ed. ME Sumner) pp. B25-B84. (CRC Press: Boca Raton, FL).

Barnes RJ, Dhanoa MS, Lister SJ (1989) Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy* **43**, 772-777.

Barthes BG, Brunet D, Ferrer H, Chotte JL, Feller C (2006) Determination of total carbon and nitrogen content in a range of tropical soils using near infrared spectroscopy: influence of replication and sample grinding and drying. *Journal of Near Infrared Spectroscopy* **14**, 341-348.

Baumgardner MF, Kristof SJ, Johannsen CJ, Zachary AL (1970) Effect of organic matter on multispectral properties of soils. In 'Proceeding of the Indian Academy of Science' pp. 413-422.

Baumgardner MF, Silva LRF, Biehl LL, Stoner ER (1985) Reflectance properties of soils. *Advances in Agronomy* **38**, 1-44.

Beck RH, Robinson BF, McFee WH, Peterson JB (1976) Information note 081176. Laboratory Application of Remote Sensing, Purdue University, West Lafayette, IN.

Ben-Dor E (2002) Quantitative remote sensing of soil properties. *Advances in Agronomy* **75**, 173-243.

Ben-Dor E, Banin A (1990a) Diffuse reflectance spectra of smectite minerals in the near infrared and their relation to chemical composition. *Sciences Geologigues Bulletin* **43**, 117-128.

Ben-Dor E, Banin A (1990b) Near infrared reflectance analysis of carbonate concentration in soils. *Applied Spectroscopy* **44**, 1064-1069.

Ben-Dor E, Banin A (1995) Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Science Society of America Journal* **59**, 364-372.

Ben-Dor E, Inbar Y, 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 Sensing of Environment* **61**, 1-15.

Ben-Dor E, Irons JR, Epema G (1999a) Soil reflectance. In 'Remote sensing for the earth science: Manual of remote sensing'. (Ed. AN Rench). (John Wiley & Sons: New York).

Ben-Dor E, Patkin K, Banin A (1999b) Detection of soil properties using the DAIS-7915 hyperspectral scanner data over a saline area. In 'Proceeding of the 13th International Conference on Applied Geology Remote Sensing'. Vancouver, Canada pp. 316-323.

Bingham IJ, Baddeley JA, Watson CA (2005) Evaluation of air-knife excavation and digital photography for rapid assessment of cereal roots on profile walls. In 'Conference of Root and Soil Environment II'. Sutton Bonington, UK. (Eds J Foulkes, G Russel, M Hawkesford, M Gooding, D Sparkes, E Stockdale) pp. 93-98.

Blakemore LC, Searle PL, Daly BK (1987) 'Methods for chemical analysis of soils.' (NZ Soil Bureau Scientific Report : 80, Department of Scientific and Industrial Research: Lower Hutt, New Zealand).

Boddey RM, Urquiaga S, Alves BJR, Fisher MJ (2002) Potential for C accumulation under *Brachiaria* pastures. In 'Agricultural Practices and Policies for C Sequestration in Soil'. (Eds JM Kimble, R Lal, RF Follet) pp. 395-408. (Lewis Publisher, CRC Press Company: Boca Raton).

Bogrekci I, Lee WS (2004) The effects of soil moisture content on reflectance spectra of soils using UV-VIS-NIR spectroscopy. In 'Proceedings of the 7th International Conference on Precision Agriculture and Other Precision Resources Management 25-28 July, 2004'. Hyatt Regency, Minneapolis, MN, USA. (Ed. DJ Mulla) pp. 1307-1317.

Bogrekci I, Lee WS (2005) Improving phosphorus sensing by eliminating soil particle size effect in spectral measurement. *Transactions of the ASAE* **48,** 1971-1978.

Böhm W (1979) 'Methods of studying root systems.' (Springer: New York).

Bolster KL, Martin ME, Aber JD (1996) Determination of carbon fraction and nitrogen concentration in tree foliage by near infrared reflectance: a comparison of statistical methods. *Canadian Journal of Forest Research* **26**, 590-600.

Börjesson T, Sternberg B, Linden B, Jonsson A (1999) NIR spectroscopy, mineral nitrogen analysis and soil incubations for the prediction of crop uptake of nitrogen during the growing season. *Plant and Soil* **214**, 75-83.

Boudot JP, Bel Hadi Brahim A, Chone T (1988) Dependence of carbon and nitrogen mineralization rates upon amorphous metallic constituents and allophanes in highland soils. *Geoderma* **42**, 245-260.

Boulesteix AN, Strimmer K (2007) Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics* **8**, 32-44.

Bowers S, Hanks RJ (1965) Reflectance of radiant energy from soils. *Soil Science* **100**, 130-138.

Bowers SA, Smith SJ (1972) Spectrophotometric determination of soil water content. *Soil Science Society of America Proceeding* **36**, 978-980.

Bowman DC, Devitt DA, Engelke MC, Rufty Jr TW (1998) Root architecture affects nitrate leaching from bentgrass turf. *Crop Science* **38**, 1633-1639.

Brereton RG (2003) 'Chemometrics: Data analysis for the laboratory and chemical plant.' (John Wiley & Sons: Chichester, England).

Brimmer PJ, DeThomas FA, Hall JW (2001) Method development and implementation of near-infrared spectroscopy in industrial manufacturing support laboratories. In 'Near-infrared technology in the agriculture and food industries'. (Eds P Williams, K Norris) pp. 199-214. (American Association of Cereal Chemist, Inc.: Minnesota, USA).

Brown DJ, Bricklemyer RS, Miller PR (2005) Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. *Geoderma* **129**, 251-261.

Burns D, Cziurczak E (Eds) (1992) 'handbook of near-infrared spectroscopy.' (Marcel and Dekker: New York).

Cambardella CA, Gadja AM, Doran JW, Wienhold BJ, Kettler TA (2001) Estimation of particulate and total organic matter by weight loss-on-ignition. In 'Assessment methods for soil carbon'. (Ed. Lal R. et.al.) pp. 349-359. (Lewis Publ.: Boca Raton, FL).

Campbell JB (1996) 'Introduction to remote sensing.' (The Guildford Press: New York).

Cariati F, Erre L, Micera G, Piu P, Gessa C (1983) Polarization of water molecules in phyllosylicates in relation to exchange cations as studied by near infrared spectroscopy. *Clays and Clay Minerals* **31**, 155-157.

Chandra AM, Gosh SK (2006) 'Remote sensing and geographical information.' (Alpha Science International Ltd.: Oxford, UK).

Chang CW, Laird DA (2002) Near-infrared reflectance spectroscopic analysis of soil C and N. *Soil Science* **167**, 110-116.

Chang CW, Laird DA, Hurburg Jr CR (2005) Influence of soil moisture on near-infrared reflectance spectroscopic measurement of soil properties. *Soil Science* **170**, 244-255.

Chang CW, Laird DA, Mausbach MJ, Hurburg Jr CR (2001) Near-infrared reflectance spectroscopy - Principal component regression analysis of soil properties. *Soil Science Society of America Journal* **65**, 480-490.

Chen K, Kissel DE, West LT, Adkins W (2000) Field-scale mapping of surface soil organic carbon using remotely sensed imagery. *Soil Science Society America of Journal* **64**, 746-753.

Cheng B, Wu X (2006) A modified PLSR method in prediction. *Journal of Data Science* **4**, 257-274.

Christy AA, Kvalheim OM (2007) Latent-variable analysis of multivariate data in infrared spectrometry. In 'Near-infrared spectroscopy in food science and technology'. (Eds Y Ozaki, WF McClure, AA Christy). (Wiley-Interscience, John Wiley & Sons, Inc.: Hoboken, New Jersey).

Clark DH, Cary EE, Mayland HF (1989) Analysis of trace elements in forages by near infrared reflectance spectroscopy. *Agronomy Journal* **81**, 91-95.

Clark DH, Maryland HF, Lamb RC (1987) Mineral analysis of forages with near infrared reflectance spectroscopy. *Agronomy Journal* **79**, 485-490.

Clark RN (1981) The reflectance of water-mineral mixtures at low temperature. *Journal of Geophysical Research* **86**, 3074-3086.

Clark RN (1999) Spectroscopy of rocks and minerals, and principles of spectroscopy. In 'Remote sensing for the earth science: Manual of remote sensing'. (Ed. AN Rench). (John Wiley & Sons: New York).

Clark RN, King TVV, Klejwa M, Swayze G, Vergo N (1990) High spectral resolution reflectance spectroscopy of minerals. *Journal of Geophysical Research* **95**, 12653-12680.

Clark RN, Roush TL (1984) Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research* **89**, 6329-6340.

Collins WS, Chang SH, Raines G, Canney F, Ashley R (1983) Airborne biogeophysical mapping of hidden mineral deposits. *Economic Geology* **78**, 737-749.

Conant RT, Paustian K, Elliot ET (2001) Grassland management and conversion into grassland: Effect on soil carbon. *Ecological Applications* **11**, 343-355.

Condit HR (1970) The spectral reflectance of American soils. *Photogrammetric Engineering & Remote Sensing* **36**, 955-966.

Cowie JD (1978) 'Soil and agriculture of Kairanga Country, North Island, New Zealand. Soil Bureau Bulletin 33.' (New Zealand Scientific and Industrial Research, Government Printer Wellington: New Zealand).

Cozzolino D, Montossi F, San Julian R (2005) The use of visible (VIS) and near infrared (NIR) reflectance spectroscopy to predict fibre diameter in both clean and greasy wool samples. *Animal Science* **80**, 333-337.

Cozzolino D, Moron A (2006) Potential of near-infrared reflectance spectroscopy and chemometrics to predict soil carbon fractions. *Soil & Tillage Research* **85**, 78-85.

Cozzolino D, Murray I (2004) Analysis of animal by-products. In 'Near-infrared spectroscopy in agriculture, Agronomy Monograph no. 44'. (Eds KA Barbarick, CA Roberts, WA Dick, JJ Workman, JB Reeves III, L Al-Amoodi) pp. 647-662. (American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.: Madison, Wisconsin, USA).

Cremers DA, Ebinger MH, Breshears DD, Unkefer PJ, Kammerdiener SA, Ferris MJ, Katlett KM, Brown JR (2001) Measuring total soil carbon with laser-induced breakdown spectroscopy (LIBS). *Journal of Environmental Quality* **30**, 2202-2206.

Crush JR, Easton HS, Waller JE, Hume DE, Faville MJ (2007) Genotypic variation in patterns of root distribution, nitrate interception and response to moisture stress of a perennial ryegrass (*Lolium perenne* L.) mapping population. *Grass and Forage Science* **62**, 256-273.

Csillag F, Pasztor L, Biehl LL (1993) Spectral band selection for the characterization of salinity status of soils. *Remote Sensing of Environment* **43**, 231-242.

Cullen BR, Chapman DF, Quigley PE (2006) Comparative defoliation tolerance of temperate perennial grasses. *Grass and Forage Science* **61**, 405-412.

Curran PJ (1989) Remote sensing for foliar chemistry. *Remote Sensing of Environment* **30**, 271-278.

Curran PJ, Dungan JL, Macler BA, Plummer SE, Peterson DL (1992) Reflectance spectroscopy of fresh whole leaves for estimation of chemical concentration. *Remote Sensing of Environment* **39**, 153-166.

Dalal RC, Henry RJ (1986) Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance. *Soil Science Society of America Journal* **50**, 120-123.

Daly BK, Rijkse WC (1974) Chemistry: A typical sequence of soils derived from air-fall tephra. In 'Soil Groups of New Zealand: part 1 Yellow-Brown Pumice Soils'. (Ed. NE Read). (New Zealand Society of Soil Science: Wellington).

Daniel KW, Tripathi NK, Honda K (2003) Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lop Buri (Thailand). *Australian Journal of Soil Research* **41**, 47-59.

Davies AMC (1996) Accolades, accusations and anecdotes. *Journal of Near Infrared Spectroscopy* **4**, 3-21.

Delwiche SR (2004) Analysis of small grain crops. In 'Near-infrared spectroscopy in agriculture, Agronomy Monograph no. 44'. (Eds KA Barbarick, CA Roberts, WA Dick, JJ Workman, JB Reeves III, L Al-Amoodi) pp. 269-320. (American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.: Madison, Wisconsin, USA).

Demattê JAM, Campos RC, Alves MC, Fiorio PR, Nanni MR (2004a) Visible-NIR reflectance: a new approach on soil evaluation. *Geoderma* **121**, 95-112.

Demattê JAM, Gama MAP, Cooper M, Araújo JC, Nanni MR, Fiorio PR (2004b) Effect of fermentation residue on the spectral reflectance properties of soils. *Geoderma* **120**, 187-200.

Demattê JAM, Garcia PA (1999) Alteration of soil properties through a weathering sequence as evaluated by spectral reflectance. *Soil Science Society of America Journal* **63**, 327-343.

Demattê JAM, Pereira HS, Nanni MR, Cooper M, Fiorio PR (2003) Soil chemical alterations promoted by fertilizer application assessed by spectral reflectance. *Soil Science* **168**, 730-747.

Demattê JAM, Sousa AA, Alves MC, Nanni MR, Fiorio PR, Campos RC (2006) Determining water status and other soil characteristics by spectral proximal sensing. *Geoderma* **135**, 179-195.

Dillon WR, Goldstein M (1984) 'Multivariate analysis methods and applications.' (John Wiley & Sons: New York).

Drapper NR, Smith H (1998) 'Applied regression analysis.' (John Wiley and Sons Inc.: New York).

Duckworth J (2004) Mathematical data preprocessing. In 'Near-infrared spectroscopy in agriculture, agronomy monograph no. 44'. (Eds KA Barbarick, CA Roberts, WA Dick, JJ Workman, JB Reeves III, L Al-Amoodi) pp. 115-132. (American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.: Madison, Wisconsin, USA).

Dunbabin V, Diggle A, Rengel Z (2003) Is there an optimal root architecture for nitrate capture in leaching environment? *Plant, Cell and Environment* **26**, 835-844.

Dunn BW, Beecher HG, Batten GD, Ciavarella S (2002) The potential of near-infrared reflectance spectroscopy for soil analysis - A case study from the Riverine Plain of south-eastern Australia. *Australian Journal of Experimental Agriculture* **42**, 607-614.

Dyer DJ (2004) Analysis of oilseeds and coarse grains. In 'Near-infrared spectroscopy in agriculture, agronomy monograph no. 44'. (Eds KA Barbarick, CA Roberts, WA Dick, JJ Workman, JB Reeves III, L Al-Amoodi) pp. 321-344. (American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.: Madison, Wisconsin, USA).

Ebinger MH, Norfleet ML, Breshears DD, Cremers DA, Ferris MJ, Unkefer PJ, Lamb MS, Goddard KL, Meyer CW (2003) Extending the applicability of laser-induced breakdown spectroscopy for total carbon measurement. *Soil Science Society America of Journal* **67**, 1616-1619.

Elliot ET, Coleman DC (1988) Let the soil do the work for us. *Ecology Bulletin* **39**, 23-32.

Elvidge CD (1990) Visible and near infrared reflectance characteristics of dry plant materials. *International Journal of Remote Sensing* **11**, 1775-1795.

Esbensen KH, Guyot D, Westad F, Houmoller LP (2006) 'Multivariate data analysis - in practice: An introduction to multivariate data analysis and experimental design.' (Camo: Oslo, Norway).

Escamilla JA, Comerford NB, Neary DG (1991) Soil core-break method to estimate pine root distribution. *Soil Science Society of America Journal* **55**, 1722-1726.

Fernandez RN, Schulze DG, Goffin DJ, Van Scoyoc GE (1988) Color, organic matter, pesticide adsorption relationships in a soil landscape. *Soil Science Society America of Journal* **52**, 1023-1026.

Fisher MJ, Braz SP, Dos Santos RSM, Urquiaga S, Alves BJR, Roddey RM (2007) Another dimension to grazing systems: Soil carbon. *Tropical Grasslands* **41**, 65-83.

Fisher MJ, Rao IM, Ayarza MA, Lascano CE, Sanz JI, Thomas RJ, Vera RR (1994) Carbon storage by introduced deep-rooted grasses in the South American savannas. *Nature* **371**, 236-238.

Fliege NJ (1994) 'Multivariate digital signal processing.' (John Wiley & Sons: Chichester).

Follet RF, Leavitt SW, Kimble JM, Pruessner EG (2003) Paleoenvironmental inferences from $\delta^{13}C$ of soil organic C in ^{14}C-dated profiles in the U.S.Great Plains. In 'XVI INQUA Congress, Reno, NV, 23-30 July, 2003'. (Ed. DA McGilloway).

Fystro G (2002) The prediction of C and N content and their potential mineralisation in heterogeneous soil samples using Vis-NIR spectroscopy and comparative methods. *Plant and Soil* **246**, 139-149.

Gausman HW, Allen WA, Cardenas R, Bowen RL (1970) Color photos, cotton leaves and soil salinity. *Photogrammetric Engineering & Remote Sensing* **36**, 454-459.

Gausman HW, Gerbermann AH, Wiegend CL, Leamer RW, Rodriguez RR, Noriega JR (1975) Reflectance differences between crop residues and bare soils. *Soil Science Society of America Proceeding* **39**, 752-755.

Gehl RJ, Rice CW (2007) Emerging technologies for in situ measurement of soil carbon. *Climatic Change* **80**, 43-54.

Geladi P, MacDougall D, Martens H (1985) Linearisation and scatter correction for NIR reflectance spectra of meat. *Applied Spectroscopy* **39**, 491-500.

Goetz FH, Chabrillat S, Lu Z (2001) Field reflectance spectroscopy for detection of swelling clays at construction sites. *Field Analytical Chemistry and Technology* **5**, 143-155.

Goldchin A, Oades JM, Skjemstad JO, Clarke P (1994) Soil structure and carbon cycling. *Australian Journal of Soil Research* **32**, 1043-1068.

Gower JC (1982) Data analysis: Multivariate or univariate and other difficulties. In 'Food Research and Data Analysis, Proceeding from IUFoST Symposium September 20-23'. Oslo, Norway. (Eds H Martens, H Russwurm Jr) pp. 39-67. (Applied Science Publishers).

Grove CI, Hook SJ, Paylor ED (1992) 'Laboratory reflectance spectra of 160 minerals, 0.4 to 2.5 Micrometers.' (Jet Propulsion Laboratory Publication 92-2: Pasadena).

Guo LB, Halliday MJ, Siakimotu SJM, Gifford RM (2005) Fine root production and litter input: Its effects on soil carbon. *Plant and Soil* **272**, 1-10.

Halvin JL, Beaton JD, Tisdale SL, Nelson WL (1999) 'Soil Fertility and Fertilizers: An Introduction to Nutrient Management.' (Prentice Hall: NJ, USA).

Hardisky MA, Klemas V, Smart RM (1983) The influence of soil salinity, growth form and leaf moisture on spectral radiance of Spartina Alterniflora canopies. *Photogrammetric Engineering & Remote Sensing* **49**, 77-83.

Harris RD, Cremers DA, Ebinger MH, Bluhm BK (2004) Determination of nitrogen in sand using laser-induced breakdown spectroscopy. *Applied Spectrocopy* **58**, 770-775.

Hassink J (1992) Effect of soil texture and structure on carbon and nitrogen mineralization in grassland soils. *Biology and Fertility of Soils* **14**, 126-134.

He Y, Huang M, García A, Hernández A, Song H (2007) Prediction of macronutrients content using near-infrared spectroscopy. *Computers and Electronics in Agriculture* **58**, 144-153.

He Y, Song H, García PA, Hernández GA (2005) Measurement and analysis of soil nitrogen and organic matter content using near-infrared spectroscopy techniques. *Journal of Zhejiang University of SCIENCE* **6B**, 1081-1086.

Hedley CB, Kusumo BH, Sanches ID, Tuohy MP, Hedley MJ, Hawke M (2007) Proximal sensing of soil and pasture development under "forest to farm" land use change. In 'Designing Sustainable Farms : Critical Aspects of Soil and Water Management, 8th and 9th February 2007'. Massey university, Palmerston North. (Fertilizer & Lime Research Centre Massey University).

Hedley CB, Yule IJ (2008) A high resolution of soil water status mapping method for irrigation scheduling and two variable rate scenarios for pasture and maize irrigation. In 'Proceeding of 9th International Precision Agriculture Conference, July 20-23, 2008'. Denver, USA.

Henderson TL, Baumgardner MF, Franzmeier DP, Stott DE, Coster DC (1992) High dimensional reflectance analysis of soil organic matter. *Soil Science Society of America Journal* **56**, 865-872.

Herzberg G (1945) 'Molecular spectra and molecular structure, Vol. 2, Infrared and Raman spectra of polyatomic molecules.' (Van Nostrand, Reinhold: New York).

Hewitt AE (1998) 'New Zealand soil classification.' (Landcare Research Science Series No. 1. Manaki Whenua Press: Lincoln, Canterbury, New Zealand).

Hick RT, Russell WGR (1990) Some spectral considerations for remote sensing of soil salinity. *Australian Journal of Soil Research* **28**, 417-431.

Hirschfeld T (1985) Salinity determination using NIRA. *Applied Spectroscopy* **39**, 740-741.

Hoffer RM, Johannsen CJ (1969) Ecological potentials in spectral signature analysis. In 'Remote sensing in ecology'. (Ed. PL Johnson) pp. 1-16. (University of Georgia Press: Athens).

Hruschka WR (1987) Data analysis: Wavelength selection methods. In 'Near-Infrared Technology in the Agriculture and Food Industries'. (Eds P Williams, K Norris) pp. 35-55. (American Association of Cereal Chemist, Inc.: St Paul, MN).

Hruschka WR (2001) Data analysis: wavelength selection methods. In 'Near-Infrared Technology in Agriculture and Food Industries'. (Eds P Williams, K Norris) pp. 39-58. (American Society of Cereal Chemists, Inc.: St Paul, Minnesota, USA).

Hueni A (2006) Field spectroradiometer data: Acquisition, organisation, processing and analysis on the example of New Zealand native plants. Master Thesis, Massey University.

Hueni A, Tuohy M (2006) Spectroradiometer data structuring, pre-processing and analysis - An IT based approach. *Journal of Spatial Science* **51**, 93-102.

Hummel JW, Gaultney LD, Sudduth KA (1996) Soil property sensing for site-specific crop management. *Computers and Electronics in Agriculture* **14**, 121-136.

Hummel JW, Sudduth KA, Hollinger SE (2001) Soil moisture and organic matter prediction of surface and subsurface soils using an NIR soil sensor. *Computers and Electronics in Agriculture* **32**, 149-165.

Hunt GR (1979) Near-infrared (1.3 - 2.4 um) spectra of alteration minerals: potential for use in remote sensing. *Geophysics* **44**, 1974-1986.

Hunt GR, Salisbury JW (1970) Visible and near infrared spectra of minerals and rocks: I. Silicate minerals. *Modern Geology* **1**, 283-300.

Hunt GR, Salisbury JW (1971) Visible and near infrared spectra of minerals and rocks: II. Carbonates. *Modern Geology* **2**, 23-30.

Hunt GR, Salisbury JW, Lenhoff CJ (1971c) Visible and near-infrared spectra of minerals and rocks: Halides, phosphates, arsenates, vanadates and borates. *Modern Geology* **3**, 121-132.

Hunt GR, Salisbury JW, Lenhoff CJ (1973b) Visible and near infrared spectra of minerals and rocks: Intermediate igneous rocks. *Modern Geology* **4**, 237-244.

Hunt GR, Salisbury JW, Lenhoff CJ (1973c) Visible and near infrared spectra of minerals and rocks: Acidic igneous rocks. *Modern Geology* **4**, 217-224.

Hunt GR, Salisbury JW, Lenhoff CJ (1974) Visible and near infrared spectra of minerals and rocks: Basic & ultrabasic igneous rocks *Modern Geology* **5**, 15-22.

Islam K, Singh B, 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.

Jackman RH (1960) Organic matter stability and nutrient availability in Taupo pumice. *New Zealand Journal of Agricultural Research* **3**, 6-23.

Kano Y, McClure WF, Skaggs RW (1985) A near infrared reflectance soil moisture meter. *Transactions of the ASAE* **28**, 1852-1855.

Kimble JM, Lal R, Follet RF, Stewart BA (Eds) (2001) 'Assessment methods for soil carbon.' (Lewis Publishers: Boca Raton).

Kokko EG, Volkmar KM, Gowen BE, Entz T (1993) Determination of total root surface area in soil core samples by image analysis. *Soil & Tillage Research* **26**, 33-43.

Kooistra L, Wanders J, Epema GF, Leuven SREW, Wehrens R, Buydens LMC (2003) The potential of field spectroscopy for the assessment of sediment properties in river floodplains. *Analytica Chimica Acta* **484**, 189-200.

Kramer R, Workman JJ, Reeves III JB (2004) Qualitative analysis. In 'Near-infrared spectroscopy in agriculture'. (Eds KA Barbarick, CA Roberts, WA Dick, JJ Workman, JB Reeves III, L Al-Amoodi) pp. 175-206. (American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.: Madison, Wisconsin, USA).

Krishnan P, Alexander JD, Bulter BJ, Hummel JW (1980) Reflectance technique for predicting soil organic matter. *Soil Science Society of America Journal* **44**, 1282-1285.

Krishnan P, Bulter BJ, Hummel J (1981) Close-range sensing of soil organic matter. *Transactions of the ASAE* **24**, 306-311.

Kusumo BH, Hedley CB, Hedley MJ, Hueni A, Tuohy MP, Arnold GC (2008a) The use of diffuse reflectance spectroscopy for in situ carbon and nitrogen analysis of pastoral soils. *Australian Journal of Soil Research* **46**, 623-635.

Kusumo BH, Hedley MJ, Hedley CB, Arnold GC, Tuohy MP (2009b) Predicting pasture root density from soil spectral reflectance: field measurement. *European Journal of Soil Science*, (Paper submitted).

Kusumo BH, Hedley MJ, Hedley CB, Hueni A, Arnold GC, Tuohy MP (2009a) The use of Vis-NIR spectral reflectance for determining root density: evaluation of ryegrass roots in a glasshouse trial. *European Journal of Soil Science* **60**, 22-32.

Kusumo BH, Hedley MJ, Tuohy MP, Hedley CB, Arnold GC (2008b) Predicting soil carbon and nitrogen concentrations and pasture root densities from proximally sensed soil spectral reflectance. In '1st Global Workshop on High Resolution Digital Soil Sensing & Mapping'. 5-8 February 2008 Sydney, Australia. (The Australian Centre for Precision Agriculture, University of Sydney & CSIRO Land and Water).

Kweon G, Lund E, Maxton C, Drummond P, Jensen K (2008) In situ measurement of soil properties using a probe-based VIS- NIR spectrophotometer. In 'Abstracts of 2008 ASABE Annual International Meeting, June 29 - July 2'. Rhode Island, USA.

Lacelle B, Waltman S, Bliss N, Orozco-Chavez F (2001) Method used to create the North American Soil Organic Carbon Digital Database. In 'Assessment methods for soil carbon'. (Ed. L R.) pp. 485-494. (Lewis Publ: Boca Raton, FL).

Lal R (2003) Offsetting global CO_2 emissions by restoration of degraded soils and intensification of world agriculture and forestry. *Land Degradation and Development* **14**, 309-322.

Lal R (2007) Carbon management in agricultural soils. *Mitigation and Adaptation Strategies for Global Change* **12**, 303-322.

Lal R (2008) Soils and sustainable agriculture. A review. *Agronomy for Sustainable Development* **28**, 57-64.

Lal R, Follet RF, Stewart BA, Kimble JM (2007) Soil carbon sequestration to mitigate climate change and advance food security. *Soil Science* **172**, 943-956.

Lebowitz RJ (1988) Digital image analysis measurement of root length and diameter. *Environmental and Experimental Botany* **28**, 267-273.

Liao L, Jarecke D, Greinchauf D, Hedman T (2000) Performance characterization of the hyperion imaging spectrometer instrument. *Proceeding of SPIE* **4135**, 264-275.

Lillesand TM, Kiefer RW, Chipman JW (2004) 'Remote sensing and image interpretation.' (John Wiley & Sons Inc.: USA).

Lindberg JD, Snyder DG (1972) Diffuse reflectance spectra of several clay minerals. *The American Mineralogist* **57**, 485-493.

Lobell DB, Asner GP (2002) Moisture effects on soil reflectance. *Soil Science Society of America Journal* **66**, 722-727.

Locher F, Heuwinkel H, Gutser R, Schmidhalter U (2005) Development of near infrared reflectance spectroscopy calibrations to estimate legume content of multispecies legume-grass mixtures. *Agronomy Journal* **97**, 11-17.

Ludwig B, Khanna PK, Bauhus J, Hopmans P (2002) Near infrared spectroscopy of forest soils to determine chemical and biological properties related to soil sustainability. *Forest Ecology and Management* **171**, 121-132.

Lynch JP (2007) Roots of the second green revolution. *Australian Journal of Botany* **55**, 493-512.

Madden HM (1978) Comments on the Savitzky-Golay convolution method for least-square-fit smoothing and differentiation of digital data. *Analytical Chemistry* **50**, 1383-1386.

Malengreau N, Bedidi A, Muller JP, Herbillon AJ (1996) Spectroscopic control of iron oxide dissolution in two ferralitic soils. *European Journal of Soil Science* **47**, 13-20.

Malley DF, Martin PD (2003) The use of near-infrared spectroscopy for soil analysis. In 'Tools for nutrient and pollutant management: Application to agriculture and environmental quality'. Massey University, Palmerston North, New Zealand. (Eds LD Currie, JA Hanly) pp. 371-404. (Fertilizer and Lime Research Centre).

Malley DF, Martin PD, Ben-Dor E (2004) Application in analysis of soils. In 'Near-infrared spectroscopy in agriculture, Agronomy Monograph no. 44'. (Eds KA Barbarick, CA Roberts, WA Dick, JJ Workman, JB Reeves III, L Al-Amoodi) pp. 729-784. (American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.: Madison, Wisconsin, USA).

Malley DF, Martin PD, McClintock LM, Yesmin L, Eilers RG, Haluschak P (2000) Feasibility of analyzing archived Canadian prairie agricultural soils by near infrared reflectance spectroscopy. In 'Near Infrared Spectroscopy: Proceeding of the 9th International Conference'. Norwich, UK. (Eds AMC Davies, R Giangiacomo) pp. 579-585. (NIR Publications).

Malley DF, Yesmin L, Eilers RG (2002) Rapid analysis of hog manure and manure-amended soils using near-infrared spectroscopy. *Soil Science Society of America Journal* **66**, 1677-1686.

Malley DF, Yesmin L, Wray D, Edwards S (1999) Application of near-infrared spectroscopy in analysis of soil mineral nutrients. *Communications in Soil Science and Plant Analysis* **30**, 999-1012.

Mannetje L.‘t (2007) The role of grasslands and forests as carbon stores. *Tropical Grasslands* **41**, 50-54.

Mark H, Workman JJ (2003) 'Statistics in spectroscopy.' (Elsevier Academic Press: San Diago).

Martens H, Naes T (2001) Multivariate Calibration by Data Compression. In 'Near-Infrared Technology in Agriculture and Food Industries'. (Eds P Williams, K Norris) pp. 59-100. (American Society of Cereal Chemists, Inc.: St Paul, Minnesota, USA).

Martens H, Stark E (1991) Extended multiplicative signal correction and spectral interference subtraction: New processing methods for near infrared spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis* **9**, 625-635.

Martens M, Martens H (1986) Partial least square regression. In 'Statistical procedures in food research'. (Ed. JR Piggot). (Elsevier Applied Science: London).

Martin PD, Malley DF, Manning G, Fuller L (2002) Determination of soil organic carbon and nitrogen at the field level using near-infrared spectroscopy. *Canadian Journal of Soil Science* **82**, 413-422.

Massart DL, Vandeginste BGM, Deming SN, Michotte Y, Kaufman L (1988) 'Chemometrics: a textbook.' (Elsevier: Amsterdam).

Mathers AC, Lotspeich FB, Laase GR, Wilson GC (1966) Strength of compacted Amarillo fine sandy loan as influenced by moisture, clay content, and exchangeable cation. *Soil Science Society of America Proceeding* **30**, 788-791.

Mathews HL, Cunningham RL, Cipra JE, West TR (1973a) Application of multispectral remote sensing to soil survey research in Southeastern Pennsylvania. *Soil Science Society of America Proceeding* **37**, 88-93.

Mathews HL, Cunningham RL, Petersen GW (1973b) Spectral reflectance of selected Pennsylvania Soils. *Soil Science Society of America Proceeding* **37**, 421-424.

Matuo PK, Sheperd KD, Albrecht A, Cadisch G (2006) Prediction of carbon mineralization rates from different soil physical fractions using diffuse reflectance spectroscopy. *Soil Biology and Biochemistry* **38**, 1658-1664.

McBratney AB, Santos MML, Minasny B (2003) On digital soil mapping. *Geoderma* **117**, 3-52.

McCarty GW, Reeves III JB (2000) Development of rapid instrumental methods for measuring soil organic carbon. In 'Assessment method for soil carbon'. (Eds R Lal, et al.) pp. 371-380. (Lewis Publishers: Boca Raton, FL).

McCarty GW, Reeves III JB (2006) Comparison of near infrared and mid infrared diffuse reflectance spectroscopy for field-scale measurement of soil fertility parameters. *Soil Science* **171**, 94-102.

McCarty GW, Reeves III JB, Follet RF, Kimble JM (2002) Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Science Society of America Journal* **66**, 640-646.

McClure WF (1985) Status of NIR in the tobacco industry. In 'NIR-84'. (Eds D Miskelly, DP Law, T Clucas). (Cereal Chemistry Division, Royal Australian Chemical Institute: Melbourne).

McClure WF (2001) Near-Infrared Instrumentation. In 'Near-infrared technology in the agriculture and food industries'. (Eds P Williams, K Norris) pp. 109-127. (American Association of Cereal Chemist, Inc.: Minnesota, USA).

McLellan TM, Aber JD, Martin ME, Malillo JM, Nadelhoffer KJ (1991a) Determination of nitrogen and cellulose content of decomposing leaf material by near infrared reflectance spectroscopy. *Canadian Journal of Forest Research* **21**, 1684-1688.

McLellan TM, Aber JD, Martin ME, Malillo JM, Nadelhoffer KJ, Dewey B (1991b) Comparison of wet chemistry and near infrared reflectance measurements of carbon-fraction chemistry and nitrogen concentration of forest foliage. *Canadian Journal of Forest Research* **21**, 1689-1893.

McTiernan KB, Garnett MH, Mauquoy D, Ineson P, Coûteaux MM (1998) Use of near-infrared reflectance spectroscopy (NIRS) in palaeoecological studies of peat. *The Holocene* **8**, 729-740.

Miller CE (2001) Chemical principles of near-infrared technology. In 'Near-infrared technology in the agricultural and food industries'. (Eds P Williams, K Norris) pp. 19-37. (American Association of Cereal Chemist, Inc.: St. Paul, Minnesota, USA).

Miller JN, Miller JC (2005) 'Statistics and chemometrics for analytical chemistry.' (Pearson Prentice Hall: Harlow, England).

Milman AS (1999) 'Mathematical principal of remote sensing: making interferences from noisy data.' (Sleeping Bear Press: Chelsea).

Mimmo T, Reeves III JB, McCarty GW, Galletti G (2002) Determination of biological measures by mid-infrared diffuse reflectance spectroscopy in soils within a landscape. *Soil Science* **167**, 281-287.

MINITAB Inc. (2003) 'MINITAB Statistical Software.' (State College, Pennsylvania).

Montgomery OL (1976) An investigation of the relationship between spectral reflectance and the chemical, physical and genetic characteristics of soils. Ph.D. Thesis, Purdue University.

Moron A, Cozzolino D (2002) Application of near infrared reflectance spectroscopy for the analysis of organic C, total N and pH in soils of Uruguay. *Journal of Near Infrared Spectroscopy* **10**, 215-221.

Moron A, Cozzolino D (2003) Exploring the use of near infrared reflectance spectroscopy to study physical properties and microelements in soils. *Journal of Near Infrared Spectroscopy* **11**, 145-154.

Moron A, Cozzolino D (2004) Determination of potentially mineralizable nitrogen and nitrogen in particulate organic matter fractions in soil by visible and near-infrared reflectance spectroscopy. *Journal of Agricultural Science* **142**, 335-343.

Morra MJ, Hall MH, Freeborn LL (1991) Carbon and nitrogen analysis of soil fractions using near infrared reflectance spectroscopy. *Soil Science Society of America Journal* **55**, 288-291.

Mosier-Boss PA, Lieberman SH, Theriault GA (2002) Field demonstrations of a direct push FO-LIBS metal sensor. *Environmental Science and Technology* **36**, 3968-3976.

Mouazen AM, Baerdermaeker JD, Ramon H (2005) Towards development of on-line soil moisture content sensor using a fibre-type NIR spectrophotometer. *Soil & Tillage Research* **80**, 171-183.

Mouazen AM, De Baerdemaeker J, Ramon H (2006a) Effect of wavelength range on the measurement accuracy of some selected soil constituents using visual-near infrared spectroscopy. *Journal of Near Infrared Spectroscopy* **14**, 189-199

Mouazen AM, Karoui R, De Baerdemaeker J, Ramon H (2006b) Characterization of soil water content using measured visible and near infrared spectra. *Soil Science Society of America Journal* **70**, 1295-1302.

Mouazen AM, Maleki MR, De Baerdemaeker J, Ramon H (2007) On-line measurement of selected soil properties using a VIS-NIR sensor. *Soil & Tillage Research* **93**, 13-27.

Murray I, Cowe I (2004) Sample preparation. In 'Near-infrared spectroscopy in agriculture, Agronomy Monograph no. 44'. (Eds L Al-Amoodi, CA Roberts, JJ Workman, JB Reeves III, KA Barbarick, WA Dick) pp. 75-112. (American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc.: Madison, Wisconsin, USA).

Murray I, Williams PC (1987) Chemical principles of near-infrared technology. In 'Near-infrared technology in the agricultural and food industries'. (Eds PC Williams, K Norris). (American Association of Cereal Chemist Inc.: St Paul, Minnesota, USA).

Nelson DW, Sommer LE (1996) Total carbon, organic carbon, organic matter. In 'Methods of soil analysis. Part 3 - Chemical methods'. (Ed. Sparks D.L. et.al.). (SSSA Book Ser 5 SSSA and ASA: Madison, WI).

Newman EI (1966) A method of estimating the total length of root in a sample. *Journal of Applied Ecology* **3**, 139-145.

Nie ZN, Mackay AD, Valentine I, Barker DJ, Hodgson J (1997) Influence of pastoral fallow on plant root growth and soil physical and chemical characteristics in a hill pasture. *Plant and Soil* **197**, 201-208.

Norris KH (1996) History of NIR. *Journal of Near Infrared Spectroscopy* **4**, 31-37.

Norris KH, Barnes RF, Moore JE, Shenk JS (1976) Predicting forage quality by infrared reflectance spectroscopy. *Journal of Animal Science* **47**, 1558-1561.

Obukhov AI, Orlov DC (1964) Spectral reflectance of the major soil groups and the possibility of using diffuse reflection in soil investigations. *Soviet Soil Science* **2**, 174-184.

Odlare M, Svensson K, Pell M (2005) Near infrared reflectance spectroscopy for assessment of spatial soil variation in an agricultural field. *Geoderma* **126**, 193-202.

Ortiz-Ribbing LM, Eastburn DM (2003) Evaluation of digital image acquisition methods for determining soybean root characteristics. *Online. Crop Management*, doi:10.1094/CM-2003-0702-1001-RS.

Ozaki Y, Morita S, Du Y (2007) Spectral analysis. In 'Near-infrared spectroscopy in food science and technology'. (Eds Y Ozaki, WF McClure, AA Christy). (Wiley-Interscience, John Wiley & Sons, Inc.: Hoboken, New Jersey).

Palacios-Orueta A, Pinzón JE, Ustin SL, Roberts DA (1999) Remote sensing of soils in the Santa Monica Mountains: II. Hierarchical foreground and background analysis. *Remote Sensing of Environment* **68**, 138-151.

Pan WL, Bolton RP (1991) Root quantification by edge discrimination using a desktop scanner. *Agronomy Journal* **83**, 1047-1052.

Paustian K, Collins HP, Paul EA (1997) Management controls of soil carbon. In 'Soil organic matter in temperate agroecosystems. Long-term experiments in North America'. (Eds EA Paul, ET Elliot, K Paustian, CV Cole) pp. 15-49. (CRC Press: Boca Raton, FL).

Paustian K, Six J, Elliot E, Hunt HW (2000) Management options for reducing CO2 emissions for agricultural soils. *Biogeochemistry* **48**, 147-163.

Pearlman JS, Barry PS, Segal CS, Shepanski, Beiso D, Carman SL (2003) Hyperion, a space-based imaging spectrometer. *IEEE Transactions on Geoscience and Remote Sensing* **41**, 1160-1173.

Pirie A, Singh B, Islam K (2005) Ultra-violet, visible, near-infrared, and mid-infrared diffuse reflectance spectrsocopic techniques to predict several soil properties. *Australian Journal of Soil Research* **43**, 713-721.

Poirier V, Angers DA, Rochette P, Chantigny MH, Ziadi N, Tremblay G, Fortin J (2009) Interactive effects of tillage and mineral fertilization on soil carbon profiles. *Soil Science Society America of Journal* **73**, 255-261.

Poss JA, Russell WB, Grieve CM (2006) Estimating yields of salt- and water-stressed forages with remote sensing in the visible and near infrared. *Journal of Environmental Quality* **35**, 1060-1071

Post WM, Izaurralde RC, Mann LK, Bliss N (2001) Monitoring and verifying changes of organic carbon in soil. *Climate Change* **51**, 73-99.

Qing Z, Bin Z, YangZhu Z, RenChao W (2004) Influence of parent materials on paddy soil hyperspectral characteristics and SOM spectral-parameter-models. *Acta Pedologica Sinica* **41**, 905-911.

Rao BRM, Ravi Sankar T, Dwivedi RS, Thammappa SS, Venkataratnam L (1995) Spectral behavior of salt-affected soils. *International Journal of Remote Sensing* **16**, 2125-2136.

Rees RM, Bingham IJ, Baddeley JA, Watson CA (2005) The role of plants and land management in sequestering soil carbon in temperate arable and grassland ecosystems. *Geoderma* **128**, 130-154.

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

Reeves III JB, McCarty GW (2000) The potential of near infrared reflectance spectroscopy as a tool for spatial mapping of soil composition for use in precision agriculture. In 'Near Infrared Spectroscopy: Proceeding of the 9th International Conference'. Norwich, UK. (Eds AMC Davies, R Giangiacomo) pp. 587-591. (NIR Publications).

Reeves III JB, McCarty GW (2001) Quantitative analysis of agricultural soils using near infrared reflectance spectroscopy and a fibre optic probe. *Journal of Near Infrared Spectroscopy* **9**, 25-34.

Roberts CA, Stuth J, Flinn P (2004) Analysis of forages and feedstuffs. In 'Near-infrared spectroscopy in agriculture, Agronomy Monograph no. 44'. (Eds KA Barbarick, CA Roberts, WA Dick, JJ Workman, JB Reeves III, L Al-Amoodi) pp. 231-267. (American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.: Madison, Wisconsin, USA).

Rosell RA, Gasparoni JC, Galantini JA (2001) Soil organic matter evaluation. In 'Assessment methods for soil carbon'. (Ed. Lal R. et.al.) pp. 311-322. (Lewis Publ.: Boca Raton, FL).

Russel CA (2003) Sample preparation and prediction of soil organic matter by near infra-red reflectance spectroscopy. *Communications in Soil Science and Plant Analysis* **34**, 1557-1572.

Saeys W, Mouazen AM, Ramon H (2005) Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. *Biosystems Engineering* **91**, 393-402.

Saggar S, Mackay AD, Hedley CB (1999) Hill slope effects on the vertical fluxes of photosynthetically fixed ^{14}C in a grazed pasture. *Australian Journal of Soil Research* **37**, 655-666.

Sandorfy C, Buchet R, Lachenal G (2007) Principal of molecular vibrations for near-infrared spectroscopy. In 'Near-infrared spectroscopy in food science and technology'. (Eds Y Ozaki, WF McClure, AA Christy). (Wiley-Interscience, John Wiley & Sons, Inc.: Hoboken, New Jersey).

Savitzky A, Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedure. *Analytical Chemistry* **36**, 1627-1638.

Scheinost AC, Chavernas A, Barron V, Torrent J (1998) Use and limitations of second-derivative diffuse reflectance spectroscopy in the visible to near-infrared range

to identify and quantify Fe oxide minerals in soils. *Clays and Clay Minerals* **46**, 528-536.

Schmidt KS, Skidmore AK (2004) Smoothing vegetation spectra with wavelets. *International Journal of Remote Sensing* **25**, 1167-1184.

Shenk JS, Landa I, Hoover MR, Westerhaus MO (1981) Description and evaluation of near infrared reflectance spectro-computer for forage and grain analysis. *Crop Science* **21**, 355-358.

Shenk JS, Westerhaus MO (1991a) Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. *Crop Science* **31**, 469-474.

Shenk JS, Westerhaus MO (1991b) Population structuring of near infrared spectra and modified partial least squared regression. *Crop Science* **31**, 1548-1555.

Shenk JS, Westerhaus MO (1996) Calibration the ISI way. In 'Near infrared spectroscopy: the future waves'. (Eds AMC Davies, P Williams) pp. 198-202. (NIR Publications: Chichester, UK).

Shepherd KD, Palm CA, Gachengo CN, Vanlauwe B (2003) Rapid characterization of organic resource quality for soil and livestock management in tropical agroecosystem using near-infrared spectroscopy. *Agronomy Journal* **95**, 1314-1322.

Shepherd KD, Vanlauwe B, Gachengo CN, Palm CA (2005) Decomposition and mineralization of organic residues predicted using near infrared spectroscopy. *Plant and Soil* **277**, 315-333.

Shepherd KD, Walsh MG (2002) Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal* **66**, 988-998.

Shepherd TG, Saggar S, Newman RH, Ross CW, Dando JL (2001) Tillage-induced changes to soil structure and organic carbon fractions in New Zealand soils. *Australian Journal of Soil Research* **39**, 465-489.

Shibusawa S, Ehara K, Okayama T, Umeda H, Hirako S (2005) A real-time multi-spectral soil sensor: predictability of soil moisture and organic matter content in a small field. In 'Papers presented at the 5th European Conference on Precision Agriculture'. Uppsala, Sweden. (Ed. JV Stafford) pp. 495-502.

Shibusawa S, I Made Anom, Sato SW, Sasao A, Hirako S (2001) Soil mapping using the real-time soil spectrophotometer. In 'Third European Conference on Precision Agriculture (ECPA)'. (Eds G Granier, S Blackmore) pp. 497-508. (Agro Montpellier).

Shibusawa S, I. Made Anom SW, Hache C, Sasao A, Hirako S (2003) Site-specific crop response to temporal trend of soil variability determined by the real-time soil spectrofotometer. In 'Precision Agriculture. Proceeding of the Joint European Conference of ECPA-ECPLF'. Berlin, Germany. (Ed. JV Stafford) pp. 639-643. (Wageningen Academic Publishers).

Shonk JL, Gaultney LD, Schultze DG, Van Scoyoc GE (1991) Spectroscopic sensing of soil organic matter content. *Transactions of the ASAE* **34**, 1978-1984.

Six J, Paustian K, Elliot ET, Combrink C (2000) Soil structure and soil organic matter. I. Distribution of aggregate size classes and aggregate-associated carbon. *Soil Science Society of America Journal* **64**, 681-689.

Slaughter DC, Abbott JA (2004) Analysis of fruits and vegetables. In 'Near-infrared spectroscopy in agriculture, Agronomy Monograph no. 44'. (Eds KA Barbarick, CA Roberts, WA Dick, JJ Workman, JB Reeves III, L Al-Amoodi) pp. 377-398. (American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.: Madison, Wisconsin, USA).

Slaughter DC, Pelletier MG, Upadhyaya SK (2001) Sensing soil moisture using NIR spectroscopy. *Applied Engineering in Agriculture* **17**, 241-247.

Smith P (2004) Carbon sequestration in croplands: The potential in Europe and the global context. *European Journal of Agronomy* **20**, 229-236.

Smith P (2008) Land use change and soil organic carbon dynamics. *Nutrient Cycling in Agroecosystems* **81**, 169-178.

Solberg ED, Nyborg M, Izaurralde RC, Malhi SS, Janzen HH, Molina-Ayala M (1998) Carbon storage in soils under continuous cereal grain cropping: N fertiliser and straw. In 'Management of carbon sequestration in soil'. (Eds R Lal, K Kimble, R Follett, B Stewart) pp. 235-254. (CRC Press. 456 p: New York).

Solomonson VV, Smith PL, Park Jr AB, Webb WC, Lynch TJ (1980) An overview of progress in the design and implementation of *Landsat*-D systems. *IEEE Transactions on Geoscience and Remote Sensing* **18**, 137-146.

Sorensen LK, Dalsgaard S (2005) Determination of clay and other soil properties by near infrared spectroscopy. *Soil Science Society of America Journal* **69**, 159-167.

Soussana JF, Loiseau P, Vuichard N, Ceschia E, Balesdent J, Chevallier T, Arrouays D (2004) Carbon cycling and sequestration opportunities in temperate grasslands. *Soil Use and Management* **20**, 219-230.

Starr C, Morgan AG, Smith DB (1981) An evaluation of near-infrared reflectance analysis in some plant-breeding programmes. *Journal of Agricultural Science* **97**, 107-118.

Stenberg B, Nordkvist E, Salmonsson L (1995) The use of near infrared reflectance spectra of soils for objective selection of samples. *Soil Science* **159**, 109-114.

Stenberg B, Rogstrand G, Bölenius E, Arvidsson J (2007) On-line soil NIR spectroscopy: identification and treatment of spectra influenced by variable probe distance and residue contamination. In 'ECPA 2007'. (Ed. JV Stafford) pp. 125-131. (Wageningen Academic: Skiathos).

Stoner ER (1979) Physicochemical, site and bi-directional reflectance factor characteristics of uniformly moist soil. Ph.D. Thesis, Purdue University.

Stoner ER, Baumgardner MF (1981) Characteristic variation of reflectance of surface soils. *Soil Science Society of America Journal* **45**, 1161-1165.

Stuart B (2004) 'Infrared spectroscopy: fundamentals and application.' (John Wiley & Sons: Chichester, England).

Sudduth KA, Hummel JW (1991) Evaluation of reflectance methods for soil and organic matter sensing. *Transactions of the ASAE* **34**, 1900-1909.

Sudduth KA, Hummel JW (1993) Soil organic matter, CEC, and moisture sensing with a prototype NIR spectrometer. *Transactions of the ASAE* **36**, 1571-1582.

Sudduth KA, Hummel JW (1996) Geographic operating range evaluation of an NIR soil sensor. *Transactions of the ASAE* **39**, 1599-1604.

Sudduth KA, Hummel JW, Birrell SJ (1997) Sensors for site-specific management. In 'The state of site-specific management for agriculture'. (Eds FT Pierce, EJ Sadler) pp. 183-210. (ASA-CSSA-SSSA: Madison, Wisconsin).

Tan ZX, Lal R, Smeck NE, Calhoun FG (2004) Relationship between surface soil organic carbon pool and site variables. *Geoderma* **121**, 187-195.

Theng BKG, Churchman GJ, Newman RH (1986) The occurrence of interlayer clay-organic complexes in two New Zealand soils. *Soil Science* **142**, 262-266.

Thenkabail PS, Enclona EA, Ashton MS (2004) Accuracy assessment of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment* **90**, 354-376.

Thorup-Kristensen K (2006) Effect of deep and shallow root systems on the dynamics of soil inorganic N during 3-year crop rotations. *Plant and Soil* **288**, 233-248.

Tisdale JM, Oades JM (1982) Organic matter and water-stable aggregates in soils. *Journal of Soil Science* **33**, 141-163.

Tomasson JA, Sui R, Cox MS, Al-Rajehy A (2001) Soil reflectance sensing for determining soil properties in precision agriculture. *Transactions of the ASAE* **44**, 1445-1453.

Trujillo W, Fisher MJ, Lal R (2006) Root dynamics of native savanna and introduced pastures in the Eastern Plains of Colombia. *Soil & Tillage Research* **87**, 28-38.

Tsai F, Philphot W (1998) Derivative analysis of hyperspectral data. *Remote Sensing of Environment* **66**, 41-51.

Udelhoven T, Emmerling C, Jarmer T (2003) Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study. *Plant and Soil* **251**, 319-329.

Unger PW, Kaspar TC (1994) Soil compaction and root growth: a review. *Agronomy Journal* **86**, 759-766.

Urselmans TT, Mitchel K, Helfrich M, Flessa H, Ludwig B (2006) Near-infrared spectroscopy can predict the composition of organic matter in soil and litter. *Journal of Plant Nutrition and Soil Science* **169**, 168-174.

USDA Soil Survey Staff (1999) 'Soil Taxonomy a basic system of soil classification for making and interpreting soil surveys. USDA Agriculture Handbook No. 436.' (Published by US Government Printing Office: Washington, DC, USA).

Valdes C, Andres S, Giraldez FJ, Garcia R, Calleja A (2006) Potential use of visible and near infrared reflectance spectroscopy for the estimation of nitrogen fractions in forages harvested from permanent meadows. *Journal of the Science of Food and Agriculture* **86**, 308-314.

van Vuuren JAJ, Meyer JH, Claassens AS (2007) Potential use of near infrared reflectance monitoring in precision agriculture. *Communications in Soil Science and Plant Analysis* **37**, 2171-2184.

Vinogradov BV (1981) Remote sensing of humus content of soils. *Soviet Soil Science* **11**, 114-123.

Viscarra Rossel RA (2008) ParleS: Software for chemometric analysis of spectrocopic data. *Chemometrics and Intelligent Laboratory Systems* **90**, 72-83.

Viscarra Rossel RA, McBratney AB (1998) Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. *Geoderma* **85**, 19-39.

Viscarra Rossel RA, Walvoort DJJ, McBratney AB, Janik LJ, Skjemstad JO (2006) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **131**, 59-75.

Vucetich CG, Wells N (1978) 'Soils, agriculture, and forestry of Waiotapu Region, Central North Island, New Zealand.' (DSIR Soil Bureau Bulletin 31. Published by Landcare Research, New Zealand).

Waes CV, Mestdagh I, Looten P, Carlier L (2005) Possibilities of near infrared reflectance spectroscopy for the prediction of organic carbon concentrations in grassland soils. *Journal of Agricultural Science* **143**, 487-492.

Walkley A, Black IA (1934) An examination of the Degtjareff method for determining organic carbon in soils: Effect of variation in digestion conditions and of inorganic soil constituents. *Soil Science* **63**, 251-263.

Wang D, Shannon MC (2001) Detecting salinity effect on soybean growth using a multi-spectral radiometer. In 'Proceedings of the 5th International Conference on Precision Agriculture, 16-19 July, 2000'. Bloomington, Minnesota, USA. (Eds PC Robert, RH Rust, WE Larson).

Westerhaus M, Workman Jr J, Reeves III JB, Mark H (2004) Quantitative analysis. In 'Near-infrared spectroscopy in agriculture, Agronomy Monograph no. 44'. (Eds KA Barbarick, CA Roberts, WA Dick, JJ Workman, JB Reeves III, L Al-Amoodi) pp. 133-174. (American Society of Agronomy, Inc., Crop Science Society of America, Inc., Soil Science Society of America, Inc.: Madison, Wisconsin, USA).

Wielopolski L, Hendrey G, Johnsen KH, Mitra S, Prior SA, Rogers HH, Torbert HA (2008) Nondestructive system for analyzing carbon in soil. *Soil Science Society America of Journal* **72**, 1269-1277.

Wielopolski L, Mitra S, Hendrey G, Rogers H, Torbert A, Prior S (2003) Non-destructive in situ soil carbon analysis: principles and results. In 'Proceeding 2nd Nat. Conf. Carbon Sequestration: Developing and Validating the Technology Base to Reduce Carbon Intensity. 5-8 May, 2003'. Alexandria, VA.

Wielopolski L, Orion I, Hendrey G, Rogers H (2000) Soil carbon measurement using inelastic neutron scattering. *IEEE Transactions on Nuclear Science* **47**, 914-917.

Williams PC (1992) The Phil Williams' episode. *NIR News* **3**, 3-4.

Williams PC (1995) Near infrared technology in Canada. *NIR News* **6**, 12-13.

Williams PC (2001) Implementation of near-infrared technology. In 'Near-infrared technology in the agriculture and food industries'. (Eds P Williams, K Norris) pp. 145-169. (American Association of Cereal Chemist, Inc.: Minnesota, USA).

Williams PC, Norris K (2001) Variables affecting near-infrared spectroscopic analysis. In 'Near-infrared technology in the agriculture and food industries'. (Eds P Williams, K Norris) pp. 171-185. (American Association of Cereal Chemist, Inc.: Minnesota, USA).

Workman JJ, Shenk J (2004) Understanding and using the near-infrared spectrum as an analytical method. In 'Near-infrared spectroscopy in agriculture, Agronomy Monograph no. 44'. (Eds L Al-Amoodi, CA Roberts, JJ Workman, JB Reeves III, KA Barbarick, WA Dick) pp. 3-10. (American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc.: Madion, Wisconsin, USA).

Wu HX, Wang RC (1991) Soil spectral characteristics and its quantitative analysis in soil classification. *Acta Pedologica Sinica* **28**, 177-184.

www.wikipedia.org Endmember (mineralogy). In 'Wikipedia, The free of encyclopedia'.

Yamamoto KY, Cremers DA, Ferris MJ, Foster LE (1996) Detection of metals in the environment using a portable laser-induced breakdown spectroscopy instrument. *Applied Spectrocopy* **50**, 222-233.

YingFeng H, TengHui L (1995) Spectral characteristics of main-types of soil in southern China and soil classification. *Acta Pedologica Sinica* **31**, 58-68.

Zanoni V, Davis B, Ryan R, Gasser G, Blonski S (2002) Remote sensing requirements development: A simulation-based approach. (ISPRS.www.isprs.org/ commision1/proceedings/paper/VZanoni ISPRS2002. pdf.