Soil spectral characteristics and their predictive value in relation to spatial and temporal variability in wheat yield and soil quality within a long-term field trial

By

Nompumelelo Ngejane

Thesis presented in fulfilment of the requirements for the degree of

Master of Agricultural Sciences

at

Stellenbosch University

Department of Soil Science, Faculty of AgriSciences

Supervisor: Dr. Andrei B. Rozanov

Co-supervisor: Dr. Catherine E. Clarke

December 2019

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated). The reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2019

Summary

Soil is a heterogeneous growing medium, with complex processes and mechanisms that are not easy to be fully understood. Soil spatial variations may be encountered within short distances, and these variations, directly and indirectly, affect plant production and crop yields. The field of agriculture is facing an escalating demand of databases from a regional to a worldwide scale that will help agriculturists understand and be able to mitigate the impact of spatial variations in the field (soils and crops). However, to make such data available is expensive and involves tedious and labor-intensive methods. Rapid and cost-effective tools to measure variations in soil properties and crop yields for large areas are required. Soil spectroscopy appears to be a fast, nondestructive, cost-effective, environmental-friendly, reproducible, and repeatable analytical technique.

The study aims at evaluating the use of soil spectroscopy in predicting common selected soil properties and wheat yield, as well as exploring its potential in explaining the spatial variations in the field, both in soil properties and in wheat yield. The experiment was conducted as a long-term ongoing trial at the Langgewens research farm, Western Cape Department of Agriculture. The trial was laid out in an incomplete block design structure, across a 12 ha area made up of three cropping systems with varying degrees of crop diversity and four replicates allocated in 120 plots. Archived soil samples (for the year 2015) from all the 120 plots were used for the analysis of selected common soil properties and scanned to acquire the near-infrared (NIR) spectral signatures using a spectrophotometer (Bruker Multi-Purpose Analyzer). The NIR spectral signatures were pre-processed following two procedures that are de-noising (removal of the fringe bands with large noise) and data transformation (first derivative and straight-line subtraction) before performing the multivariate data analyses.

The partial least squares regression (PLSR) method was used to develop chemometric models to establish the relationship that the NIR spectral signatures have with wheat yield and soil properties. The prediction results of the PLSR models were fairly accurate and falling within the acceptable ranges. For the selected models, most correlation coefficients (R^2) ranged between 0.80 and 0.60 with the ratios of performance to deviation (RPD) ranging between 2.38 to 1.6 for wheat yield and selected soil properties (CEC, SOC, pH, Ca).

In 2019, the soil core samples at 0-5 cm depth (120 in total) were analyzed for some key soil parameters and were scanned to acquire the NIR spectral signatures. This was done to assess the temporal variations in wheat yield and changes in 5 cm soil spectral characteristics in the field trial area after four years (2015 to 2019). An overall significant difference was obtained between the averaged spectral absorbance for the years 2015 and 2019 ($p<0.05$),

an increase in absorbance for the year 2019 was observed. When assessing the changes that have occurred in some selected soil properties, bulk density was observed to have significantly decreased across the field ($p<0.05$). A decrease in soil organic carbon ($p<0.05$) was observed as well as in soil organic carbon stocks ($p<0.05$) within a fixed depth. However, no significant change was observed in soil carbon stocks when the depth was adjusted. Results obtained in this study show that to a certain extent, the spectral characteristics in the NIR region might be a good indicator of not only soil properties but also plant responses to the changes in soil properties across the field.

Opsomming

Die grond is 'n heterogene groeimedium, met komplekse prosesse en meganismes wat nie maklik verstaanbaar is nie. Ruimtelike variasies in die grond kan binne kort afstande voorkom, en hierdie variasies beïnvloed plantproduksie en oesopbrengste direk en indirek. Die landbouveld het 'n vinnig groeiende vraag na databasissevan 'n plaaslike tot 'n globale skaal, wat landboukundiges sal help om die impak van die ruimtelike variasies in die veld (grond en gewasse) te verstaan en te verminder. Om sulke data beskikbaar te stel, is egter duur en behels tydsame, arbeidsintensiewe metodes. Vinnige en kostedoeltreffende instrumente om variasies in grondeienskappe en oesopbrengste vir groot gebiede te meet, is nodig. Grondspektroskopie bied 'n vinnige, nie-vernietigende, koste-effektiewe, omgewingsvriendelike, en herhaalbare analitiese metode.

Die studie het ten doel om die gebruik van grondspektroskopie te evalueer om algemene geselekteerde grondeienskappe en koringopbrengste te voorspel, asook om die potensiaal daarvan te ondersoek om die ruimtelike variasies in die veld, sowel in grondeienskappe as koringopbrengs, te verklaar. Die eksperiment is uitgevoer as 'n langtermyn proef op die Langgewens-navorsingsplaas, van die Wes-Kaapse Departement van Landbou. Die proef is uitgevoer in 'n onvolledige blok ontwerpstruktuur , oor 'n oppervlakte van 12 ha wat bestaan uit drie teeltstelsels met wisselende gewasverskeidenheid en vier replikasies toegeken in 120 erwe. Grondmonsters vanuit die argief (die jaar 2015) van al 120 plotte is gebruik vir ontleding van geselekteerde gemeenskaplike grondeienskappe en geskandeer om die naby-infrarooi (NIR) spektrale handtekeninge te verkry, met behulp van 'n spektrofotometer (Bruker Multi-Purpose Analyzer). Die NIR-spektrale handtekeninge is vooraf verwerk volgens twee prosedures wat agtergrond geraas verwyder (die prosses waarby onnodige data bande verwyder word) en datatransformasie (eerste afgeleide en reguitlyn aftrek) voordat die multiveranderlike data-ontleding uitgevoer is.

Die metode van gedeeltelike kleinste kwadraatregressie (GKKR) is gebruik om chemometriese modelle te ontwikkel om die verwantskap wat die NIR-spektrale handtekeninge het met koringopbrengs en grondeienskappe te bepaal. Die voorspellingsresultate van die GKKR-modelle was redelik akkuraat en het binne aanvaarbare reekse geval. Vir die geselekteerde modelle het die meeste korrelasiekoëffisiënte (R^2) tussen 0,80 en 0,60 gewissel, met die verhoudings van prestasie tot afwyking (RPD) tussen 2,38 en 1,6 vir koringopbrengs en geselekteerde grondeienskappe (katioonuitruilkapasiteit, organiese koolstof, pH, uitruilbare kalsium).

In 2019 is die grondkernmonsters op 0-5 cm diepte (120 in totaal) geanaliseer vir enkele belangrike grondparameters en is geskandeer om die NIR-spektrale handtekeninge te bekom. Dit is gedoen om die tydelike variasies in koringopbrengs en veranderinge in 5 cm grondspektrale eienskappe in die veldproefgebied na vier jaar (2015 tot 2019) te bepaal. 'N beduidende verskil is verkry tussen die gemiddelde spektrale absorbansie vir die jare 2015 en 2019 ($p < 0,05$), 'n toename in absorbansie vir die jaar 2019 is waargeneem. By die ondersoek na die veranderinge wat plaasgevind het in sekere geselekteerde grondeienskappe, is daar waargeneem dat die massa digtheid oor die erwe beduidend afgeneem het ($p < 0.05$). 'N afname in organiese koolstof ($p<0.05$) sowel as in organiese koolstofvoorrade in die grond ($p<0.05$) is binne 'n vaste diepte waargeneem. Daar is egter geen noemenswaardige verandering in die grondstofkoolstofvoorraad waargeneem toe die diepte aangepas is nie. Resultate wat in hierdie studie verkry is, toon dat die spektrale eienskappe in die NIR-band tot 'n sekere mate 'n goeie aanduiding kan wees van nie net grondeienskappe nie, maar ook plantresponse op die veranderinge in grondeienskappe oor die hele veld.

Biographical sketch

Nompumelelo Ngejane was born in Mount Frere, Eastern Cape on February 12, 1995. She attended her primary and junior grades at Jersey Farm Junior Secondary School in Mthatha. In 2010, she started her high school grades at St John's College in Mthatha, where she obtained her matric. In 2013, she enrolled for a BSc degree in Agriculture at Fort Hare University, and completed it in 2017. Thereafter, in 2018, she enrolled for a Master's degree in Agriculture (Soil Science) at Stellenbosch University.

Acknowledgements

I wish to express my special and most sincere gratitude to the following:

- My supervisor, Dr. Andrei Rozanov, for the opportunity he has granted me. Thank you for your guidance and patience throughout the journey to make this project a success, I will forever be grateful.
- My co-supervisors, Dr. Cathy Clarke, for all her advice, motivation and moral support and compassion at times of need.
- Dr. Johann Strauss (Western Cape Department of Agriculture), for making sure we have access to the Langgewens research farm facilities and always providing us with all the information needed to successfully conduct the study.
- Devin Osborne, for always being there to answer countless clarity seeking questions whenever he was consulted.
- The lecturers and support staff at the Department of Soil Science, for their additional inputs and insights into the study.
- My postgraduate colleagues at the Soil Science Department, for creating a friendly working environment and for being there whenever I needed help.
- My mother (Sylvia Ngejane) and siblings (Zoleka, Siphumelele, Luleka and Noziphiwo) for their prayers and encouragement. I appreciate the genuine, unchanging and exceptional support you have given me throughout my studies.
- Abbey Maheso, for all the assistance, support, prayers and motivation throughout my journey. Thank you Ndhove for always encouraging me to perform at my level best.
- Citrus Academy, for supporting me financially.
- God Almighty, for giving me strength, ability and making it possible for me to start and finish this work.

Preface

This thesis is divided into five chapters.

Chapter 1 Contains the study background, problem statement, and also the aims and objectives.

Chapter 2 Presents the literature review on published work based on the use of near-infrared (NIR) spectroscopy in agriculture. The review focuses on evaluating whether the NIR spectral signatures can be used to estimate and predict the soil properties and wheat yield.

Chapter 3 Predominantly deals with developing prediction models for common selected soil properties and wheat yield using the NIR spectral signatures of the soil surface; the first objective will be fulfilled in this chapter.

Chapter 4 Assesses the temporal changes that have occurred over the period of four years of the field trial in the soil spectra, soil properties and in wheat yield; objective two and three will be fulfilled in this chapter.

Chapter 5 Presents general conclusions and recommendations.

Table of Contents

Figure Index

Table Index

Equations

Abbreviations

C	Chroma
Ca	Calcium
CEC	Cation exchange capacity
CIE	International Commission on Illumination -Commission Internationale de l'Eclairage
CO_2	Carbon dioxide
CO_3	Carbonates
CV	Cross-validation
DEM	Digital elevation model
DSM	Digital soil mapping
EC	Electric conductivity
EM	Electromagnetic
Fe	Iron
H	Hue
H_2O	Water/ Hydrogen oxide
ISFETs	Like ion-sensitive field effect transistors
K	Potassium
KCl	Potassium chloride
Kg	Kilogram
$Kg \cdot ha^{-1}$	Kilogram per hectare
Km	Kilometer
m	Meter
m^2	Square meter
Mg	Magnesium
$mg.g^{-1}$	Milligram per grams
mm	Millimeter
Mn	Manganese
MPA	Bruker Multi-Purpose Analyzer
MSLR	Multiple stepwise linear regression
NDVI	Normalized difference vegetation index
NIR	Near-Infrared

nm	Nanometer
OH	Hydroxide
OT	Overtone
PCA	Principal Component analysis
PLS	Partial least squares
PLSR	Partial least squares regression
PSS	Proximal soil sensing
QUANT	Quantitative
RF	Random forests
RGB	Red– Green–Blue
RMSEC	Root mean square error of calibration
RMSEE	Root mean square error of estimation
RMSECV	Root mean square error of cross validation
RMSEP	Root mean square error percentage
RPD	Ratio of performance to deviation
RT	Regression trees
SG	Savitzky–Golay
SNV	Standard normal variate
SO_4	Sulphate
SOC	Soil organic carbon
SOMC	Soil organic matter content
SSE	Sum of squared errors
$t.ha^{-1}$	Ton per hectare
TN	total nitrogen
TOC	Total organic carbon
V	Value
Vis	Visible
Vis-NIR	Visible-near-infrared
Zn	Zinc
Γ	Gamma
Λ	Wavelength
%R	Percentage reflectance

ρb	Bulk densit

CHAPTER 1: Introduction

This study focusses on assessing the soil spectral characteristics and their spatial variability in relation to wheat yield and soil properties. The purpose of this study is to evaluate whether the near-infrared (NIR) absorbance detection using laboratory analysis of samples collected in the field may be a good estimate of soil properties and wheat yield. It is also based on evaluating whether the spatial variation in NIR absorbance reflects the variation in some soil properties and wheat yield. This was done through developing models for predicting wheat yield and common selected soil properties using the NIR spectral signatures of the soil surface. Also, by inspecting the changes that have occurred in the field over a period of four years in the NIR spectral absorbance, wheat yield and in selected soil properties.

1.1 Background

Due to soil spatial variability within short distances in the field that have been observed to negatively influence the experimental results, it is vital to understand the spatial structure of any piece of land before it can be utilized (Van Es et al., 2007; Osborne, 2017). These variations directly and indirectly contribute a lot to variations in crop performances and production. The major reason behind the need to understand how soils vary spatially and how do they relate to variation in soil properties and crop yield is to be able to plan agricultural field trials in such a way that the variations will be catered for and mitigate their impact. Achieving this goal involves various actions and techniques that can be tedious, labor intensive and expensive. Therefore, the main aim is to develop and promote the use of rapid and cost-effective analytical methods. The conventional and/ or traditional methods, which are normally used to assess the spatial variations within the field, require complex analytical procedures with several parameters to be analyzed. Soil analysis is expensive and dense sampling is required to adequately characterize the spatial variability of an area, making broad-scale quantitative evaluation difficult (Nocita et al., 2015). This also, includes the means of quantifying and or predicting yield and doing production analysis.

Therefore, there is a need for cost-effective, environmentally friendly and rapid methods to be used for soil analysis and quantification of field spatial variation. Consequently, near-infrared spectroscopy appears to be a reliable and efficient strategy to determine if there is a significant variation within the field and in yield production (Malley et al., 1999; Ferrio et al., 2004; Brown et al., 2005; Feyziyev et al., 2016). Various authors have found the use of soil spectral signatures as a technique with a potential to successfully describe the spatial variations in

relation to soil parameters and crop yields (Weidong et al., 2002; Stenberg et al., 2010; Grunwald et al., 2015; Mohamed et al., 2018). It has been observed to be correlated to many parameters that are commonly observed including SOC, Ca, Mg, pH, moisture content, CEC, texture and many other parameters that are not normally analyzed (Luis Galvez-Sola et al., 2015; Šestak et al., 2018).The soil spectra is also being perceived as a general soil characteristic which can be related directly to production. So with the aid of soil spectroscopy, instead of measuring all of the soil parameters individually using the conventional or traditional analytical methods, the soil spectra can be inspected and analyzed to obtain its correlation with soil properties and find out how it relates to crop production (Islam et al., 2003; Du & Zhou, 2009; Peng et al., 2014)

1.2 Problem Statement

Observations on variability in topsoil properties and structure of soil cover were made by Osborne (2017) at Langgewens research farm, Western Cape. These spatial variations were shown to have a huge impact on yield and crop production. It was shown that no single soil characteristic may be seen as a driver of crop yields within the experimental setup, but the combination of all the parameters has a strong correlation with wheat yield. Therefore, user-friendly methods to study and quantify the spatial variations in the field are needed. The near-infrared (NIR) characterization of soil is seen as an integrating soil characteristic related to many soil properties and wheat yield. This study explores the co-variance of soil NIR spectroscopy with selected soil properties and wheat yield.

1.3 Aims and Objectives

This study aims to develop wheat yield and soil properties prediction models using the near-infrared spectroscopy. This would make it possible to evaluate whether does the variation observed in NIR absorbance correlate to the changes observed in the soil surface and wheat yield. The NIR spectral characterization can be used as a unique integrative soil characteristic of its own rather than a proxy for known and commonly used soil attributes in describing soil spatial variation better than the reference soil variables. The following specific objectives provide guidance to achieve the goal of this study:

i. To quantify the correlation between NIR spectral signatures with wheat yield and some common key soil properties.
ii. To detect temporal changes in soil organic matter (SOM) and soil carbon stocks in the first 5 cm of soil using proximal sensed data and routine soil analysis after four years of field trials.
iii. To evaluate whether the NIR spectral signatures reflect the changes in selected soil properties and wheat yield after four years of field trial.

CHAPTER 2: Literature review

2.1 Introduction

Soil properties may vary within short distances and they are rarely homogeneous at any spatial scale (Odlare et al., 2005). These variations are a result of many factors including parent material, topography (elevation, drainage), climatic conditions and soil management practice (Frogbrook et al., 2002). These factors act concurrently and may impact crops yield variations in space. Knowledge of variations in a field before laying out an experimental trial is important because, in agricultural experiments, variations are decisive for the results of the experiment. These variations have a huge impact on the treatment effect and can make it difficult to interpret the results fairly and accurately (Bilgili et al., 2011). To investigate soil spatial variation, various methods can be used, such as; soil maps and thematic maps based on several soil analysis (chemical, biological and physical) (Vågen et al., 2016). Yield maps and various remote sensing indices often correlate with mapped soil properties. Modern technology makes it possible to measure and map crop yields and enables the quantification of the within-field spatial variations (Stenberg et al., 2010).

In South Africa, the responsibility to protect biodiversity and the environment constantly increases while the cereal production gross margins are reported to be declining (DAFF, 2017). Therefore, there is a greater need for spatial field variation assessment and management using more convenient and environment-friendly methods. Visible and near-infrared spectroscopy (VNIRS) has been widely used to estimate various soil attributes for soil surveys, land use planning, and soil management purposes (Malley et al., 1999; Reeves, 1999, Viscarra Rossel et al., 2006a; Viscarra Rossel et al., 2006b; Stenberg et al., 2010; Reeves, 2010; Vasques et al., 2014; Soriano-Disla et al., 2014; Peng et al., 2014; Vohland et al., 2014; Bushong et al., 2015; Cambou et al., 2016; Mohamed et al., 2018). At present, there is no published work that directly relates soil VNIRS to crop yields. Meanwhile, the literature shows that VNIRS is widely used for various purposes in the production industry. This includes its use for routine, non-destructive assessment and estimation of grain quality in cereals, forage quality in soybean and nutritional elements estimation on citrus leaves (Ferrio et al., 2000; Luis et al., 2015; Asekova et al., 2016). Therefore, this study was conducted to further understand and outline the potential and benefits of using absorbance spectra for the estimation of crop (wheat) yield variation within the soil context of an agricultural trial. This literature review focuses on the use of proximal and remote sensing techniques for quantifying

soil spatial variability. It explores the use of near-infrared spectral signatures in assessing the spatial variability in the field (soil properties) and crop yields.

2.2 Proximal and remote soil sensing

The qualities and functions of soils have been inspected through conducting conventional methods including physical, chemical, mineralogical, and biological laboratory analysis (Cheng-Wen Chang et al., 2001). These conventional methods continue to serve the world of soil analysis well, but they can be time-consuming, expensive and complex. There is a growing demand for good quality, inexpensive information on how to use precision agriculture for improving the efficiency and sustainability of food production (Zhou and Shen, 2019). These applications have encouraged the development of sensors to measure soil properties and complement or replace the more conventional laboratory techniques. Sensors provide quantitative results and can be more time and cost-effective than conventional laboratory analyses (Rossel et al., 2011). Many devices can be used for proximal soil sensing (PSS), like ion-sensitive field-effect transistors (ISFETs) to measure soil pH and soil nutrients, or portable visible-near-infrared (vis-NIR) spectrometers to measure soil properties like organic carbon content and mineral composition. Proximal soil sensing refers to field-based techniques that can be used to measure soil properties from a distance of approximately less than 2 m above the soil surface. Field-based sensors are used to obtain signals from the soil when the sensor's detector is in contact with or close to (within 2 m) the soil (Viscarra Rossel et al., 2009).

On the other hand, remote sensing (RS) allows the mapping of the Earth's surface from satellite or airborne systems. Remote sensing is a method of collecting data that records the amount of electromagnetic radiation (EMR) emitted or reflected from objects on the Earth at different wavelength or wavenumber (Jensen, 2005). Different materials have different reflectance characteristics, and that includes soils, water, rocks, vegetation, and elements of the built environment (Brown, 2007). That makes it possible for RS to be able to play a role in the identification and mapping of soil attributes, whereby the impact of soil grain size, water content, and organic matter on soil spectral reflectance are identified (Jensen, 2005). Soil properties can be captured directly using RS (e.g., by images of bare soil) or inferred indirectly (e.g., by sensing biotic properties that are then used in a functional model to estimate them) (Grunwald et al., 2015). It gathers and can provide soil information over large geographical areas but its disadvantage is that it only senses the soil surface and in most cases, the topsoil (few millimeters of the profile). Meanwhile, some proximal soil sensors (e.g. EM38, GPR) can be used to measure both the top soil and sub soil or subsurface soil properties (Grunwald et al., 2015). The integration of vis–NIR spectroscopy, proximal and/ remote sensing and digital

soil mapping (DSM) is making it possible to study and gather information on soils over large regions of the world (Teng et al., 2018).

Amongst the benefits of using PSS, measurements are made at field conditions from the soil surface or within the soil profile and results are produced in a timely manner that is almost immediately (Grunwald et al., 2015). It provides analysts with an alternative and effective approach to learn and gather more information about soils. These sensors permit rapid and inexpensive collection of high-resolution, quantitative, and precise data, that can be used to better understand soil spatial variability (Rossel et al., 2011). There are various kinds of proximal soil sensors and may be identified by how they measure. Some are invasive, whereby measurements are carried in situ or ex-situ and noninvasive ones. They are also distinguished by the source of energy that they use (active or passive), how they operate (mobile or stationary) and whether they use direct or indirect inference when measuring the soil properties of interest (Fig. 2.1).

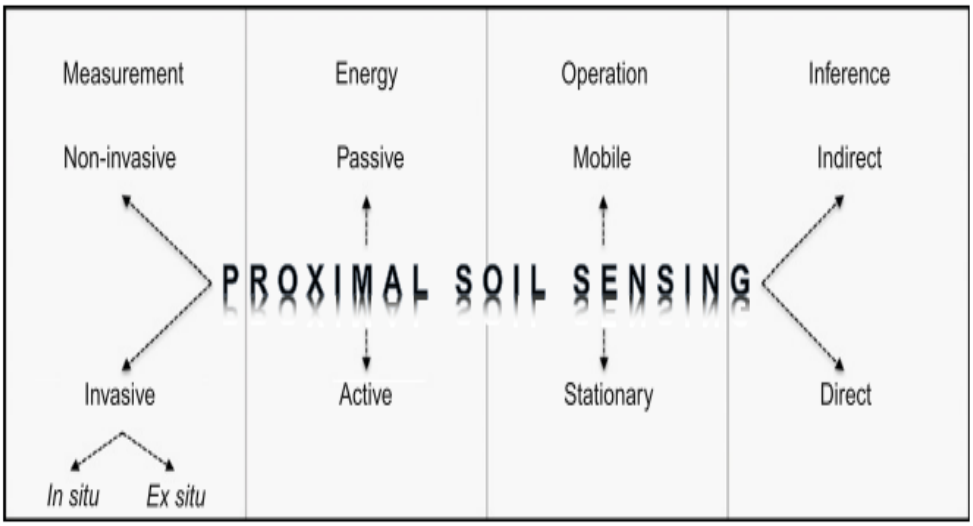


Figure 2.1: Illustration of different techniques for proximal soil sensing (source: https://ars.els-cdn.com/content/image/1-s2.0-B9780123864734000051-f05-02-9780123864734.jpg).

The proximal soil sensor is invasive if there is a contact between the sensor and the soil when doing measurements. Measurements with the invasive ones may be made in situ (measurements are undertaken within the soil) or ex-situ (measurements are done on collected soils, like measurements on soil cores). It is said to be noninvasive if there is no sensor-to-soil contact between the two (sensor and the soil). Active PSS produce their energy from an artificial source (i.e. halogen bulb) and the passive ones use naturally occurring radiation from the earth or sun. Typically, these sensors are used for fine-resolution soil mapping (Rossel et al., 2011). On the other hand, measurements when using a stationary sensor are done in a fixed manner.

The sensor is said to be direct when the measurement of the aimed soil property is based on a physical process. However, it is said to be indirect when the measurement is of a representative and inference is with a pedotransfer function. Figure 2.1 gives a description of the different techniques for SS and how do they function. Although proximal soil sensors results may not be as accurate per individual measurement like those of conventional laboratory analysis, they may be imprecise and/or biased. The logic for their use lies in the fact that they facilitate the collection of larger amounts of spatial data using simpler, cheaper, and less laborious techniques which are also more informative (Teng et al., 2018). These sensing techniques are of great importance and advantageous in precision agriculture. They offer conclusive knowledge on soil properties, crops yield and methods to increase soil productivity and decrease environmental risks with infrared spectroscopy being one of the techniques commonly widely used successfully (Grunwald et al., 2015; Ramirez-lopez et al., 2018; Zhou and Shen, 2019)

2.3 Soil spectral signatures

2.3.1 The Visible range

In the electromagnetic spectrum, visible light covers the wavelength range between 400 and 700 mm as depicted in Figure 2.2 below. When the visible radiation interacts with the soil, energy transition in atoms is produced and this is as a result of electron processes like crystal field effect and charge transfer (Vågen et al., 2016). In soils, colour is controlled by the broad absorption bands that are as a result of these electron processes in the visible wavelength. The scattering effects alter the albedo (the incident light proportion or reflected radiation by a surface) sequence of the spectrum base line. Even though the spectral response in the visible region is not very big, it is possible to obtain quantitative information from the spectral information adding it to the qualitative information observed by the naked eyes (Nocita et al.,2015).

2.3.2 The near infra-red spectral signatures

The interaction between a sample (soil) and infrared light that has been dispersed into individual wavelengths, usually by a prism defines near-infrared spectroscopy (Shepherd & Walsh, 2002). The near-infrared covers the region between 750-2500 nm (13333.33 - 4000.00 cm^{-1}) in the electromagnetic spectrum as shown below in Figure 2.2. Soil samples are scanned over the entire near-infrared region by the use of a monochromator. In the NIR region, the components of the complex organic mixture have different absorption properties as a result of the stretching and bending vibrations in molecular bonds (Odlare et al., 2005).

The NIR region is dominated by the weak overtones and combinations of these fundamental vibrations due to the stretching and bending of NH, OH and CH groups. Infrared spectroscopic techniques are highly sensitive to both organic and inorganic phases of the soil, making their use in the agricultural and environmental sciences particularly relevant (Viscarra et al., 2006). They also incorporate a great benefit because a single spectrum may contain wide-ranging information on numerous soil attributes, and can be used to predict them simultaneously (Vasques et al., 2014; Nocita et al.,2015).

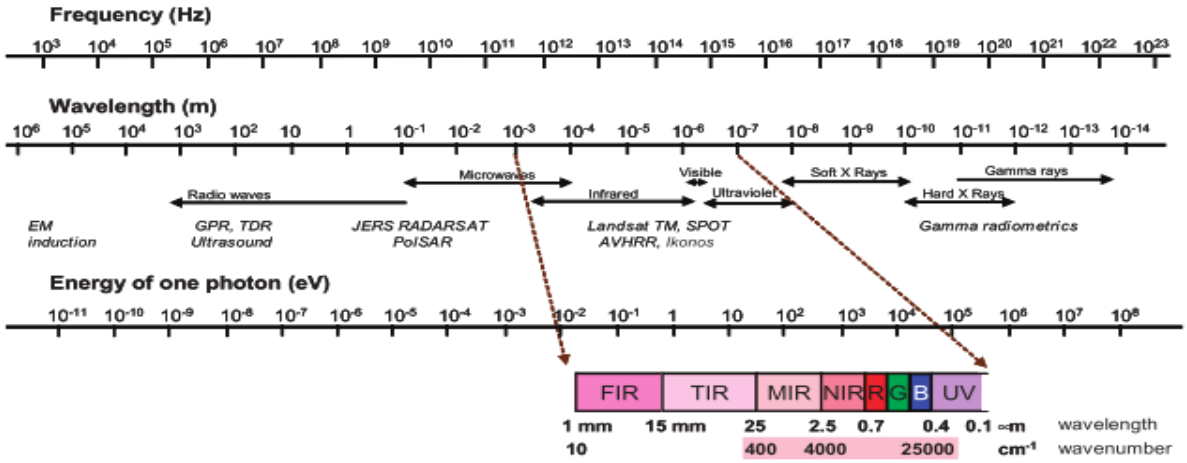


Figure 2.2: The electromagnetic spectrum (McBratney et al., 2003)

2.3.3 Characteristics of soil visible and near-infrared reflectance spectroscopy

A soil spectrum is produced by focusing on a soil sample radiation that has all the significant frequencies in a certain range of preference (Reeves, 2010. The radiation will make individual molecular bonds vibrate depending upon the constituents present in the soil, either by bending or extending, and they will absorb light to different degrees, with a particular energy quantum matching to the difference between two energy levels (McBratney et al., 2003). As the energy quantum is specifically linked to the frequency, the subsequent absorption spectrum creates a characteristic shape that can be utilized for explanatory purposes. The frequencies at which light is absorbed show up as a reduced sign of reflected radiation and are shown in percentage reflectance (%R), which can then be converted to apparent absorbance: A = log (1/R) as depicted in the vertical axis of Figure 2.3 (Šestak et al., 2018).

The wavelength at which the absorption occurs depends likewise on the chemical matrix and environmental factors. Those factors include neighboring functional groups and temperature, taking into account the identification of molecules, which may contain a similar type of bonds (Stenberg et al., 2010). The wide superimposed and weak vibrational modes characterize the

NIR region, giving soil NIR spectra few, broad absorption features. When radiation energy is high, electronic excitations become the main process in the visible region. Vis–NIR spectra contain fewer absorptions than the mid-IR; this is because of the broad and overlapping bands. Hence, leading to Vis-NIR spectra being more difficult to interpret. Moreover, useful information on organic and inorganic materials in the soil is also contained in this region (Luis Galvez-Solaet al., 2015).

Absorptions in the visible region are predominantly associated with minerals that contain iron like haematite and goethite (Mortimore et al., 2004). In the visible regions that are dominated by chromophores, soil organic matter (SOM) can also have broad absorption bands. In the NIR region absorptions result from the overtones of OH, SO_4, and CO_3 groups, and also from the combinations of fundamental features of H_2O and CO_2. Viscarra et al., (2006) also added that, “clay minerals can show absorption in the vis–NIR region due to metal-OH bend plus O–H stretch combinations”. With water having a strong influence on vis–NIR spectra of soils and the dominant absorption bands of water around 1400–1900 nm are characteristic of soil spectra, but there are weaker bands in other parts of the vis–NIR range (Weidong et al., 2002).

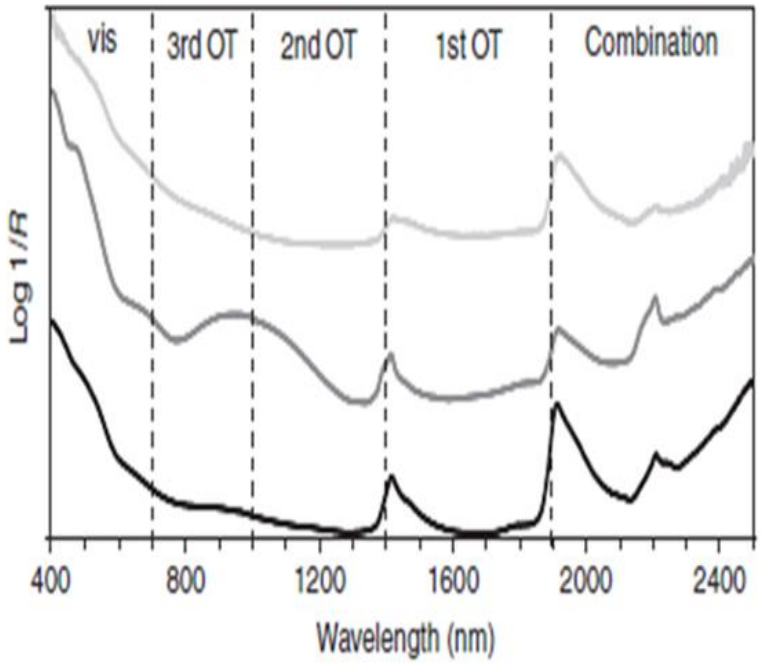


Figure 2.3: Soil vis-NIR 400-2500 nm spectra showing approximately where the combination, first, second and third overtone (OT) vibrations occur, as well as the visible (Vis) range (Stenberg et al., 2010)

2.3.4 Spectra transformation, wavelength optimization, and region selection

The measured spectra are easily influenced by individual differences (sample particle size, light intensity, measurement conditions), baseline variations, and substantial noise. Therefore, the pretreatment methods should be applied to minimize irrelevant and useless information of the spectra and increase the correlation between the spectra and the measured values. The frequently adopted pretreatment methods include normalization, Savitzky–Golay (SG), the first and second-order derivatives, multiplicative scatter correlation (MSC), standard normal variate (SNV), and detrending or any combination thereof (Lin et al., 2017). The pretreatment methods selected for our study include; Savitzky–Golay filter for smoothing, and a combination of the first-order and straight-line subtraction. The straight-line subtraction fits a straight line to the spectrum and subtracts it. This accounts for a tilt in the recorded spectrum (Luis Galvez-Sola et al., 2015). The first derivative calculates the first derivative of the spectrum. This method emphasizes steep edges of a peak. It is used to emphasize pronounced, but small features over a broad background (Zhou and Shen, 2019). Different spectral pretreatment methods are applied for minimizing the irrelevant and noisy parts of the spectra and increasing the spectra correlation with the measured values. Table 2.1 contains results from a study that developed prediction models for correlating the organic matter to Vis-NIR spectra using different pretreatment methods (Lin et al., 2017).

Table 2.1: Prediction Results of PCR Models with Different Pretreatments (Lin et al., 2017).

Pretreatment	PCS	R_{CC}	RMSEP	RDP
Smooth	15	0.5307	0.4123	1.189
1st derivative	7	0.6296	0.3547	1.3013
2^{nd} derivative	2	0.6808	0.3292	1.3116
MSC	9	0.6324	0.4053	1.3005
SNV	9	0.6206	0.4093	1.2879
MSC + 2^{nd} derivative	7	0.6261	0.3837	1.2704
MSC + 1^{st} derivative	6	0.5847	0.3927	1.2559
MSC + smooth	20	0.5779	0.4076	1.2079
SNV + smooth	18	0.7086	0.3434	1.4337
SNV + 1^{ST} derivative	6	0.5854	0.3928	1.2556
SNV + 2nd derivative	7	0.6415	0.6415	1.3058

It can be observed that different methods have different influences on the results, and the combination of the SG filter for smoothing, MSC and the derivatives exhibit the most favorable results. The preprocessing methods that yielded better results in the above study were also selected for the current.

Wavelength optimization on the full spectrum to enhance accuracy is still a challenging task, especially when the collected spectra display strong overlapping and imperceptible distinctive features (Han et al., 2014). The Vis-NIR range spectra are mainly composed of the overtones and combination bands of hydrogen groups, and the absorption peaks are of weak intensity and relatively low sensitivity. They also have wide absorption bandwidth, serious overlaps, and multiple correlations in spectral information. This also indicates one of the limiting factors, if the full spectra were involved in the model; it would not only increase the complexity of the model and calculation load but also reduce the prediction accuracy of the model owing to the irrelevant variables and collinearity between the variables (Stenberg et al., 2010).

2.4 Quantification of correlation between soil spectral signatures with soil properties and yield.

2.4.1 Near-infrared absorbance in relation to soil properties and its advantages.

Diffuse reflectance spectroscopy in the visible and near-infrared range (400–2500 nm) is one of the techniques that have the potential to improve the efficiency of soil survey and soil analysis at large (Rossel et al., 2016). Vis-NIR spectroscopy makes it possible to extract soil information on various soil properties including the amount of water present and it's particle-size distribution (Teng et al., 2018). The use of these spectral libraries as a tool for soil properties and yield predictions appear to be the most rapid method which can also be used to quantify soil properties, yield and soil spatial variability (Adeline et al., 2017). These libraries have a potential for the development of risk-based systems of soil interpretations. Whereby these interpretations are designed to quantify prediction uncertainty, so that users may be able to use such information in decision-making (Soriano-Disla et al., 2014). Table 2.2 depicts the results of a study by Ramirez-lopez et al. (2018) showing a correlation matrix of observed soil properties and mean vis-NIR absorbance. Across the correlation coefficients between soil properties and absorbance, Vis-NIR absorbance appears to have high correlations with soil properties.

Table 2.2: Correlation matrix of observed soil properties and mean vis-NIR absorbance of the Vis-NIR variables (Ramirez-lopez et al., 2018)

	Sand content	Silt content	Clay content	Ca^{++}	log (silt/ sand)	log (clay/sand)	Mean absorbance
Sand content	1.00	-0.82	-0.99	-0.70	-0.82	-0.85	-0.80
Silt content	-0.82	1.00	0.72	0.71	0.84	0.68	0.66
Clay content	-0.99	0.72	1.00	0.66	0.77	0.85	0.79
Ca^{++}	-0.70	0.71	0.66	1.00	0.61	0.56	0.62
log(silt/ sand)	-0.82	0.84	0.77	0.61	1.00	0.92	0.63
log(clay/ sand)	-0.85	0.68	0.85	0.56	0.92	1.00	0.65
Mean absorbance	-0.80	0.66	0.79	0.62	0.63	0.65	1.00

The use of the spectroscopic methods in soil analysis comes with many advantages, including its nature of being non-destructive, cost-effective and environmentally friendly (Soriano-Disla et al., 2014). Soil disturbance is minimized, as the already stored soil archives can be used for scanning (Viscarra et al., 2006). Nocita at al. (2015) compared the analysis cost of traditionally laboratory method (Walkley-Black) for analyzing total organic carbon with using Vis-NIR spectroscopy. Results obtained showed that Vis-NIR spectroscopy method is ten times cheaper and faster than the laboratory method.

The near-infrared reflectance spectroscopy method of analysis requires minimal sample preparation, coupled with the benefits of being environmentally friendly as they don't require the use of laboratory extracting chemicals (Bushong et al., 2015). Spectral reflectance or absorbance methods give spectra that are greatly characteristic of the soil composition, type and thus making it possible for a several soil properties (chemically, biologically and physically) to be analyzed (Viscarra et al., 2006). The near-infrared spectroscopy is currently being promoted to describe and develop more efficiently and economically precise information on the extent and variability of soil attributes, which affect crop growth and yield (Mohamed et al., 2018).

2.4.2 The capacity of NIR spectroscopy to predict soil properties and yield

Using wavelength ranges in the near-infrared (NIR, 750-2500 nm) researchers have successfully predicted several soil fertility parameters including organic carbon (SOC), inorganic carbon, total nitrogen (TN), cation exchange capacity (CEC), pH, potassium (K),

magnesium (Mg), calcium (Ca), zinc (Zn), iron (Fe), and manganese (Mn) with various levels of prediction accuracy (Viscarra Rossel et al., 2006; Awiti et al., 2008; Soriano-Disla, 2014).

It has been reported by several authors that, quantitative prediction of soil properties (e.g. CEC, organic carbon, pH, and heavy metals) can be achieved using near-infrared spectroscopy. However, Mohamed et al. (2018) discovered one of the main limiting factors in the assessment of the soil properties using this method, which is finding certain data pretreatment and calibration procedures, where the correlations between soil reflectance data and values of each soil properties could be achieved. Therefore, with the aid of various calibration models which are developed depending on the measured soil analysis and soil reflectance spectra, valid and accurate results can be obtained (Adeline et al., 2017; Lin et al., 2017; Mohamed et al., 2018). Table 2.3 depicts such an example with validation results for soil organic carbon (SOC) or total C with background data, sample and data pre-treatments (Stenberg et al., 2010). In our study, the OPUS spectroscopy software together with a built-in module called Quant2 have been used to calibrate and find relevant data pretreatment procedures to evaluate the covariance of NIR absorbance and soil properties with wheat yield.

When it comes to crop yield, near-infrared spectroscopy has been used for different purposes including; the analysis of the properties and chemical constituents related to baking quality and nutritive value of wheat (Osborne, 1992; Garnsworthy et al., 2000). Garnsworthy et al. (2000) also managed to predict some agronomic parameters including grain weight using NIRS. This NIRS method has been proposed as an alternative method for analysis, estimation and, prediction of different elements in plants for a variety of different purposes. For example, prediction of grain quality in cereals as well as in the estimation of nutritional elements on citrus leaves. (Ferrio et al., 2000; Luis Galvez-Sola et al., 2015). These findings are also supported by Asekova et al. (2016) who also concluded that near-infrared spectroscopy has the potential to be a useful tool in the quick analysis of many samples collected from various long term experiments.

Table 2.3: Validation results for soil organic carbon (SOC) or total C ($mg.g^{-1}$) with background data, sample and data pre-treatments. Farm-scale data sets and below are not included, extracted from Stenberg et al. (2010).

Sample origin	Soil type	Reference method for carbon	Range or S.D. of carbon	Drying/ grinding	λ-range	Mean absorbance	Cal.samples / val. samples data	R^2_{val}	RMSEP	Reference
NSW Australia	Top and subsoil	SOC Walkley & Black	0.6-49.5/ 8.2	Air dry/ <2 mm	400-2500 nm	Non/ PCR	121/40 selected from PCA scores	0.81	3.5	(Islam et al., 2003)
Southern Africa	Top soil	SOC by Dichromate oxidation	2.3-55.8/ NA	Air dry/ <2 mm	380-2460 nm	Non/ MARS	674/337 random	0.80	3.1	(Shepherd and Walsh, 2002)
Belgium	Top soil	SOC Walkley & Black	0.77-60 /9.5	Field Moist / < 2 mm	300 - 1700 nm	1st derivative after MSC / PLS	306/50	0.74	4.8	(Mouazen et al., 2007)
Sweden	Top soil	SOC by Carbonate Corrected dry combustion	4-207 / 17.421.6	Air dry / <2mm	1100-2500 nm	1st derivative / PLS	2060/680	0.46	7.2	(Stenberg et al., 2002)

However, analysis are not directly undertaken when using NIRS, there must be methods (algorithms) of extracting informational data form the spectral signatures. Therefore, empirical calibration techniques for NIRS computing must be used to establish the chemical basis of the connection between what is being analyzed like grain yield or other properties and the near-infrared spectrum. For empirical calibration, the multivariate methods, like partial least squares (PLSR), are the most recommended ones (Ferrio et al., 2004; Mohamed et al., 2018).

2.4.2.1 Multivariate data analysis techniques (Machine learning)

The soil spectral data is obtained in a form of spectral signatures containing various properties (Luis Galvez-Sola et al., 2015). To simplify the complexity of soil spectral data, multivariate data analysis techniques are used to establish the relationships between soil properties and reflectance spectra. There are various methods that are commonly used which include the multiple stepwise linear regression (MSLR), partial least squares regression (PLSR), regression trees (RT) and random forests (RF) (Ji et al., 2012). Partial least squares regression is widely used in the literature, we have also chosen it for our study. PLSR is the multivariate technique used to calibrate the predictive models to access the potential of NIR spectroscopy in predicting wheat yield and soil properties using the spectral signatures. Partial least squares regression is a nonparametric regression method based on factor analysis, which is the most standard and commonly used method in spectral analysis. Analysts prefer PLSR because it relates the response and predictor variables so that the model defines more of the variance in the response with less components. It has also been reported to be more interpretable and the algorithm is computationally quicker (Stenberg et al., 2010).

Partial least squares regression has been selected for this study in order to also get comparable results with other studies, as it has been widely used in the field of soil spectroscopy (Selige et al., 2006; Viscarra Rossel et al., 2006b; Madari et al., 2006; Vohland et al., 2014; Zhou, et al., 2014; Mohamed et al., 2018). This method incorporates the advantages of principal component analysis (PCA), canonical correlation analysis and linear regression analysis. The method is more suitable for spectrum analysis, which has sufficient independent variables. Leave-one-out cross-validation is used to test the PLSR model. The accuracy of prediction is evaluated using the determination coefficient (R^2), the root-mean-square error (RMSE) and the ratio of performance to deviation (RPD).

2.5 Soil colour measurement and its Importance

2.5.1 The Munsell color notation

The measurement of soil color is mostly done by the means of comparison with color charts using the Munsell soil colour chart (Munsell Colour Company, 1975), an extract from the complete Munsell Book of Colour (Munsell Colour Company, 1980). The soil's colour can be described in terms of a hue (H), value (V) and chroma (C) units, which are collectively referred to as Munsell notation, and this can be done through matching the colour of a soil with a particular colour chip. Hue refers to a definite colour in the visible light area of the electromagnetic spectrum. It differs with the predominant wavelength reflected from the particular surface. The degree of lightness or darkness of the soil colour is indicated by the value and the chroma denotes the strength or purity of the dominant colour (Marqués-mateu et al., 2018). A high chroma indicates a strong colour and if the chroma is low, the colour will in case of a low value tend to black, while it will tend to white if the value is high (Figure 2.4).

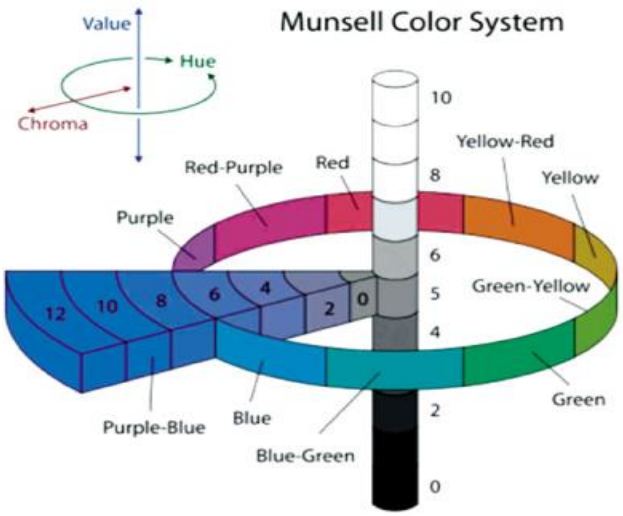


Figure 2.4: Diagrammatic representation of a Munsell soil colour system (source: http://commons.wikimedia.org /wiki/File:Munsell-system.svg)

Measuring colour using the Munsell system is relatively simple and affordable and that has resulted to various colour observers and different industries, including soil science to adopt it as the standard method for measuring and interpreting colour (Torrent & Barron, 1993). However, a range of disadvantages for Munsell soil colour description had been identified by a number of authors (Delgado & Huertas, 2014; Marqués-mateu et al., 2018). Although the Munsell system makes it possible to measure the soil colour easily and effectively, but it can be subjective. The accuracy with which it can be measured can be influenced by a number of

various factors as Munsell colour measurements are based on visual perceptions (Edwards, 1975). These factors include colour constancy, contrast and spreading along with temporal factors like light effects and colour blindness. Soil colour is of no direct agricultural significance but its significance in soils is based purely on its strong correlation with certain soil properties and processes that are in most cases relevant to land-use (van Huyssteen & Ellis, 1997).

However, Barrett (2002) observed poor correlation between soil color and many soil properties as a result caused by the restrictions of visual measurement techniques. It is evident that colour measurements in the field under variable daylight conditions are rarely accurate and that they might lead to numerous errors and invalid results (Sánchez-Marañón et al., 2011). Recommendations to ensure accurate soil colour measurements using the Munsell colour charts have been made by Melville & Atkinson (1985). These include but are not limited to making multiple observations by different observers, standardizing the preparations of all comparative samples, application of illuminating source preferably approximating illuminant D65 and adequately reporting the details of the colour measurement procedures followed.

One of the most important factors mentioned on the recommendations above, and probably one of the main focusses of the Melville and Atkinson paper under which colours by means of visual comparison with colour charts are measured is that of the lighting conditions. Melville & Atkinson (1985), further stated that because the different light sources will differ in terms of the luminosity radiated at each wavelength within the visible spectrum and therefore, the spectral power distribution for light sources used in measuring colour has been described and specified by an organization known as the CIE (International Commission on Illumination - Commission Internationale de l'Eclairage). Illuminant D65 is described as the standard by the CIE and therefore it should be used for Munsell soil colour analyses seeing that it is a close approximation of standard daylight conditions (CIE, 1971). The illuminating source is so important because different colored objects will have different spectral reflectance characteristics. The colour measurement under D65 lighting conditions require samples to be taken back to the laboratory.

The quantification of the different wavelengths of light that are reflected or absorbed from the surface of an object and therefore using its reflectance properties is a common practice. Spectrophotometric curves which compute spectral reflectance (%) for each wavelength of light can describe the reflectance properties of a soil. This represents a more complex level of colour determination than simple using vision and this is primarily measured using spectrophotometers (Rizzo et al., 2016). Various authors have reported the use of these instruments for soil colour determination to be both accurate and precise in this regard (Torrent et al. 1983; Post et al. 1993; Rizzo et al., 2016; Fan et al., 2017). This is supported by Viscarra

Rossel et al. (2009) who found the use of a portable near-visible infrared (Vis-NIS) spectrophotometer to generate estimates of soil colour that fairly agreed with what was visually measured. The advantages and benefits of using these spectrophotometers is that they are standardized instruments unlike human interpretation that consistently provide unbiased measurements of soil colour. However, these spectrophotometers are expensive and that makes in-field soil colour interpretation with handheld spectrophotometers by the average soil scientist an unlikely ideal to strive for in any classification system.

2.5.2 CIE-LAB Colour Space

A CIELAB colour space (with L*, a*, and b* as coordinates) is a color-opponent space with dimension L for lightness and a and b for the color-opponent dimensions which are the chromacity coordinates representing opposing red–green (+a reds, −a greens) and blue–yellow (+b yellows, −b blues) scales (Figure 2.5) (Delgado & Huertas, 2014). To enhance the description of colour in a uniform colour space and with suitable representation of perceived colour differences, the CIE developed the CIELu*v* and CIELa*b* systems (CIE, 1978). CIELUV uses the transformation of the x and y chromaticity coordinates to a uniform scale. The generation of the CIELAB values is done through non-linear transformations of XYZ. In both systems, L represents brightness or luminance as it was mentioned earlier and ranges from black (0) to white (100) and again; a* and u* represents a red (+)/green (-) scale; and b* and v* represents a yellow (+)/blue (-) scale. The representable model is based on nonlinearly compressed CIE XYZ color space coordinates; which consists of a central *y* axis (Y) and two horizontal *x* and *z* axes (+a*/u* to –a*/u* & +b*/v* to –b*/v*) that are perpendicular to each other (Viscarra Rossel et al., 2009).

Individual soil colors can be visually represented in a defined space and transformed between the different units used in each specified equations in these colour space models (Viscarra Rossel et al. 2006). Therefore, if the various used instruments give colour readings in different units of measurement, these units can be transformed into other desirable colour spaces using equations. This is possibly and most applicable where the applied instrument generates tristimulus values and the resulting values can be transposed to Munsell notation (HVC) (e.g. Torrent & Barron, (1993) and Post et al. (1993) used mostly systems of soil classification).

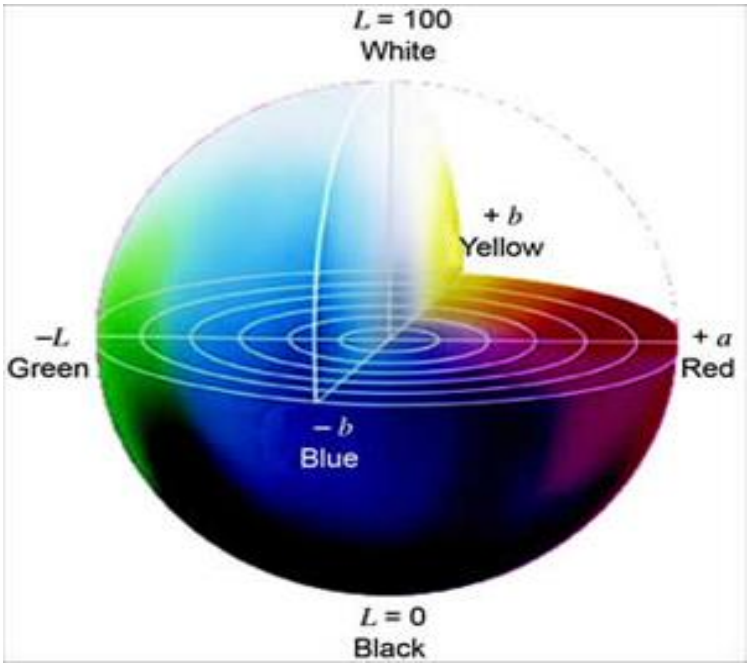


Figure 2.5: Graphical representation of the CIE Lab colour space. Source: https://www.colorcodehex.com/color-model.html

In the case of our study, this was done by the use of SpectraMagic NX software colour data software. Melville & Atkinson, (1985) alluded that through point representation in a visually defined space the mentioned above models are mostly suitable for comparisons between the individual colours themselves. For the purpose of soil colour analysis where the colour of a soil in its own sense has slight significance, the main interest would be to correlate soil colour to other colour-related variables indicative of a specific soil property or process or production. Our objective is also to correlate the soil colour to wheat yield and other soil properties.

2.6 Conclusions

Since the material composition largely determines its spectral properties, the spectral characteristics may be used to assess the spatial variation of multiple soil properties. Soil spectral characteristics in both visible and NIR ranges may be a good indicator of not only soil properties, but also plant responses to the changes in soil properties across the field.

Selection of specific spectral attributes through pre-processing of raw spectra can identify the parts of the spectrum and derivatives with highest co-variance between the transformed spectral characteristics and properties measured using a standard analytical method. PLSR may be a sufficiently good method to relate the spectral characteristic of material with its other properties or even the effects such a material may have on interaction with other materials.

In this case, we are particularly interested in correlating the spatial co-variance between soil spectra and wheat yield within a long-term experimental trial setup. Finding this co-variance and establishing a regression relationship between this two crop production factors requires extensive experimental work presented in the following chapters.

CHAPTER 3: Covariance of near-infrared absorbance and soil colour with wheat yield and selected soil properties

3.1 Introduction

The soil has numerous functions which include being a medium for plant growth (Bouma, 2014). It is of great importance to have a clear understanding of the spatial variations in the field. However, in order to understand the state of the soil, complex analytical procedures are required (FAO, 2015; Keesstra et al., 2016). Consistent monitoring of soil health depends on the development and implementation of fast and low-cost analytical methods (Keesstra et al., 2016). The results obtained by the standard analytical methods can be stored in databases for reinterpretation, record keeping and tracking temporal changes (Moebius-Clune et al., 2017). Over the past 30 years, soil reflectance spectroscopy (Vis-NIR-MIR) has been proved to be a fast, environmental-friendly cost-effective, nondestructive, repeatable and reproducible analytical technique (Viscarra Rossel et al., 2006; Soriano-Disla et al., 2014). Viscarra Rossel et al. (2009) also alluded that the technique is mainly used in the laboratory, but it is applicable in situ as new field sensors are developed.

As there is vast evidence on the successful use of soil spectroscopy to estimate soil properties, it is also important to evaluate whether this technique is good enough for routine soil analysis. That involves evaluating a lot of aspects around it such as ensuring that the errors associated with spectroscopic predictions are reduced to acceptable levels and best models are selected (Fernandez-ahumada et al., 2010; Rossel et al.,2016). The covariance of soil conditions and crop yield is the basis of precision agriculture (Shah & Wu, 2019). However, the covariance between soil spectral characteristics and crop yield has not been studied yet. The main reason for that might be because; only knowing the soil absorbance/ reflectance does not give the precise practical management tools. (Awiti et al., 2008).

The main aim of this chapter is to predict relative wheat yield and near-surface soil properties (averaged by core sampling to the depth of 5 cm) within the field trial area using near-infrared observations of the soil surface conditions and using experimental yield data. This is done using the multivariate calibration system of method development. The purpose of the calibration techniques is to evaluate the covariance between the quantities of the infrared radiation absorption with the properties of the system (wheat yield and common soil characteristics). The multivariate calibration make use of not only a single spectral point but take into account spectral features over a wide range (Madari et al., 2006). Therefore, the

information contained in the spectra of calibration samples will be compared to the information of the concentration values using a partial least square regression (PLSR).

The focus on 5 cm depth soil characteristics in this work is determined by the nature of the field trial, where various crop rotations are tested in a no-till system within fields recently converted from full tillage. The no-till system, on one hand limits the possible interventions to surface application of soil amendments, while the results of the trial are expected to be most pronounced close to the surface. Wheat, which was the main crop of the tested rotation systems was selected as the reference crop for our assessment of soil-yield relationship.

3.2 Materials and methods

3.2.1. Study area

The experiment was conducted at Langgewens Research Farm under the supervision of the Western Cape Department of Agriculture. The farm is located 18km North of Malmesbury in the Swartland region of the Western Cape of South Africa (33°16'34. 41" S, 18°45'51. 28" E). The climate in this area is described as a Mediterranean climate characterized by hot dry summers and cold wet winters. The Swartland is a sub-region of the Western Cape, which is a well-pronounced small grain producing area. A map of the Langgewens research farm and its location within the Western Cape Province can be found in Figure 3.1.

3.2.2 Trial design and layout

The trial was laid out as an ongoing component trial within a long-term crop and soil sustainability set up in an incomplete block design structure, across a 12 ha area. At the establishment of the trial, the design could not be set up as a complete randomized block design because the sizes of the camps were different.

The trial tests 3 cropping systems with varying degrees of crop diversity, which are:

- System 1 - low diversity - wheat/ canola/ wheat/ legume cover/ wheat/ wheat/ canola/ grass cover/ wheat/ canola.
- System 2 – more diversity - wheat/ faba beans/ legume cover/ wheat/ wheat/ canola/ grass cover/ vetch/ wheat
- System 3 – most diversity - oats/ linseed/ wheat/ legume cover/ wheat/ barley/ canola/ grass cover/ faba beans/ wheat

One hundred and twenty (120) plots were allocated to accommodate the 4 replicates of the trial as depicted in Figure 3.2. The three replicates were randomly assigned to camps 20 and 19C, while replicate 4 for all sequences was allocated to camp 19B. In this manner, the results of the experiment may be analyzed with either 3 or 4 replications.

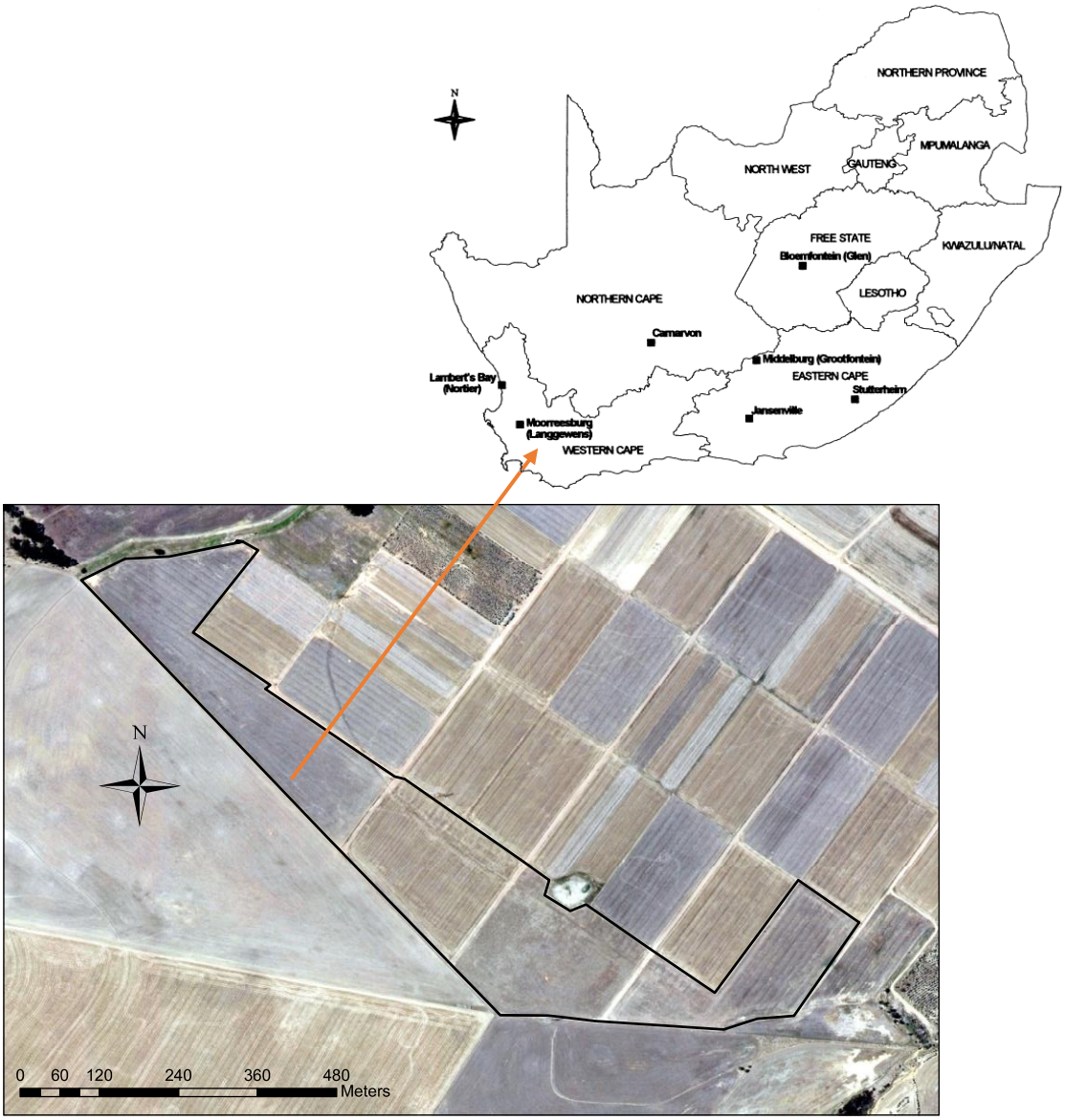


Figure 3.1: Map of the Langgewens research farm and its location within the Western Cape Province, South Africa

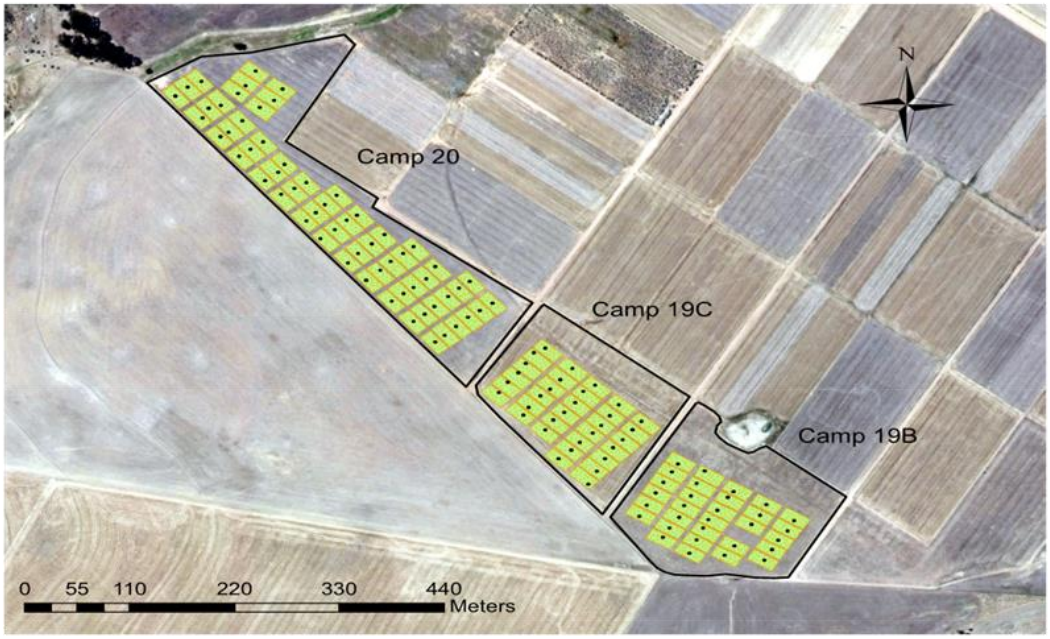


Figure 3.2: Map of experimental plot layout showing spatial near-surface (0-5 cm) soil sampling points per block (experimental unit).

3.2.3. Soil sampling and analysis

Some of the soil samples used in this chapter were archived samples housed in the Department of Soil Science, Stellenbosch University that were collected in the year 2015. For the soil properties, all the analyses took place at the Western Cape Department of Agriculture, Elsenburg, except for pH (KCl), EC, and coarse fragments determination, which were determined at the department of Soil Science, Stellenbosch University. All of these methods were adopted from the Handbook of standard soil testing methods for the advisory purpose, published by the Soil Analysis Work Committee (1990). In 2019 another sampling was done, soil colour was determined using a spectrophotometer at the department of Soil Science. Prior to analyses, samples were dried and then ground to pass through a 2-mm sieve. More information regarding the 2019 soil sampling is explained in chapter 4 materials and methods (section 4.2.1).

3.2.4. Wheat yield measurements.

Wheat yield in 2015 was measured using a New Holland T54 combine harvester equipped with a yield monitor. This system ran in combination with a global positioning system (GPS) in order to locate and visualize the spatial pattern in wheat yield over the trial area (Figure 3.3).

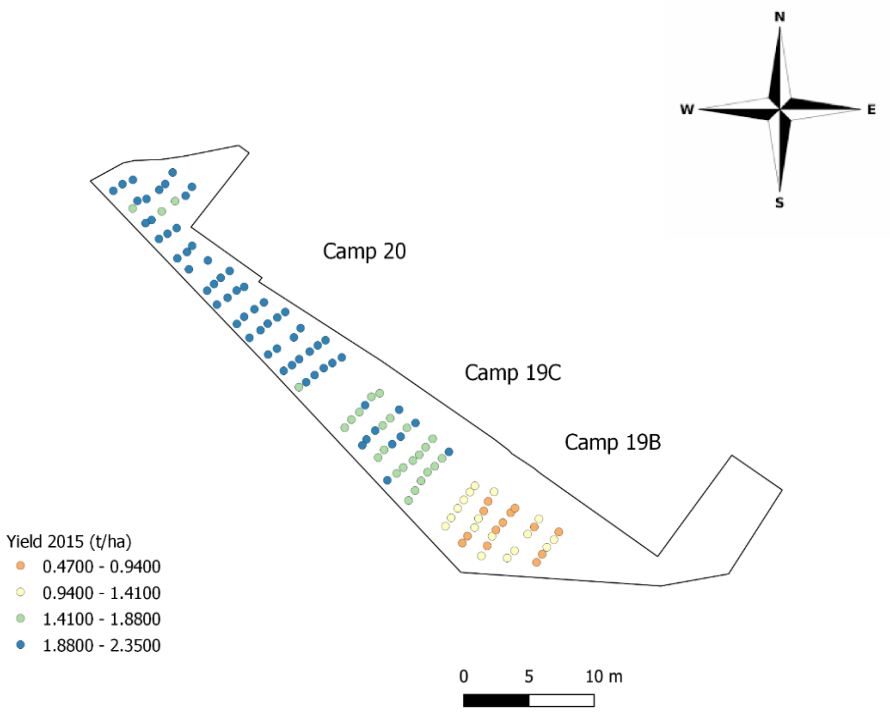


Figure 3.3: Yield map for the year 2015 showing wheat yield per plot in each sampling point per camp in the field trial area

Yield data was then normalized (rescaled to 0-1 range, where 1 is the highest and 0 is the lowest observed yield) and recorded as relative yield using the following equation:

$$z_i = \frac{xi - \min(x)}{\max(x) - \min(x)} \qquad \text{Equation 3.1}$$

where: x = (x_1,..., x_n) is the scaled wheat yield data and z_i is the i^{th} normalized data at the observation point (or plot).

The yield data obtained with the yield monitor was refined to remove edge effects (where incomplete portion of the field was harvested) and also averaged per block (as per Figure 3.2) to produce two types of data – point and block average.

3.2.5 LAB and Munsell (HVC) colour determination method

Colour measurements were made on all of the 120 soil samples using a Konica Minolta cm-600d spectrophotometer (Minolta, 2013). The spectrophotometer had to be first set up for this purpose. After the instrumental setup, measurements were standardized by calibration against a manufacturer-provided white plate of known reflectance, also called white calibration. The white plate is located on the inside of the white calibration cap with its calibration data stored in the internal memory of the instrument. Before a white calibration was performed, a zero calibration was also performed using the optional Zero Calibration Box CM-A182. The dry fine soils were then spread out on a flat surface of a glass lens individually. The spectrophotometer was then aligned over the surface of the glass lens and a reading was taken. The glass lens was removed and the following measurements in the dry state were taken. Following each measurement, the glass lens was cleaned with a fine-fibred cloth to prevent scratching or any form of damage that could influence the colour reading. The specified file was loaded and reflected on the input data. The name of the file was automatically displayed in the Tag field and the spectral data was displayed as a graph. The viewing conditions were set using SpectraMagic NX software, which is a colour data software that is designed to enable spectrophotometer measurements and graphical display of sample data. The instrument measured reflectance over an 8-mm-diameter circular area and is used with specular (gloss) component included (SCI). Colours measured by the spectrophotometer were reported both in CIE L*a*b* notation and in Munsell (HVC) notation.

3.2.6 Near-infrared spectral measurement method

Near-infrared spectral measurements were done using an NIR spectrophotometer (Bruker Multi-Purpose Analyzer), with a quartz beam splitter and RT-PbS detector (Infrared-photodetector). Each sample was placed in a 10 cm diameter, 1.5 cm deep sample holder (similar to a stainless steel petri dish) in a dark light-controlled room. It was ensured that the surface of the sample holder was fully covered with the soil sample, at least half a cm thick which is 20–25 g of the sample. They were scanned in the wavelength range between 750-2500 nm (wave numbers: 13300 – 4000 cm^{-1}) at 8 cm^{-1} using a rotating macro sample sphere at 64 scans per sample. The rotation assures a high reproducibility for heterogeneous samples as the rotation makes it possible to collect and further average the signal data from different points of the sample. In our case, 120 individual measurements were done. The viewing conditions were set using Opus software (©Bruker Optik) and absorbance was recorded.

Throughout this text, the wave number is used rather than wave length, where wave number n [nm] is related to wave length λ [cm^{-1}] as $n=10^6/\lambda$.

3.3. Statistical analysis

3.3.1. Spectral calibration and validation for prediction of soil properties and wheat yield

Near-infrared spectroscopy is an indirect analytical method that is based on the development of empirical models in which the concentration of a constituent is predicted from multivariate spectral data (Gobrecht et al., 2014). Partial least-squares regression (PLSR) was used to establish the relationships between the NIR spectra and wheat yield and soil properties, the software used to do this is OPUS 7.2.139.1294, including the statistical Quant2 module, which is supplied with the Bruker MPA (Vohral, 2011). Partial Least Squares regression is based on linear transformations from a large number of original descriptors to a new variable space based on a small number of orthogonal factors (latent variables) (Bayer et al., 2012). Latent variables are chosen in such a way as to provide maximum correlation with the dependent variable(s); thus, the PLS model, therefore, contains the smallest necessary number of factors (Miller & Miller, 2010).

One set of samples (120) representative for the multicomponent system was used to calibrate and validate the system. Before starting the calibration, one sample was excluded from the entity of samples for cross-validation. A calibration chemometric model using the PLSR was developed on 119 soil samples. The infrared spectra of these samples were used by Quant2 to calculate a calibration function, which essentially is the model used for the analysis of unknown samples later. Then, the calibration model was tested against the sample set aside for validation, and the cycle was repeated by separating a different sample until all samples were individually used for validation. The reproducibility of the measurements was checked for short and long time intervals, using a few test samples first.

The predictive ability of the PLSR model was assessed using statistical parameters typically used in chemometrics. That includes the coefficient of correlation (R^2), ratio of performance to deviation (RPD), the root mean square error of cross-validation (RMSECV) and also the rank (number of factors). These were taken as a criterion to judge the quality of the model. The coefficient of determination measures the proportion of total variation accounted for by the model, the remaining variation attributed to random error. The coefficient of determination approaches 100% as the fitted concentration values approach the true values (Awiti et al., 2008). The RMSECV was calculated from:

$$\text{RMSECV} = \sqrt{\frac{1}{N}\sum(x_i - y_i)^2} \qquad \text{Equation 3.2}$$

where: $(x_i - y_i)$ is the difference between the measured value by the laboratory methods and the predicted value by Quant2 (PLS method) using the spectral signatures and,
N is the total number of samples in the test.

The R^2 was calculated from:

$$R^2 = \left(1 - \frac{SSE}{\sum(x_i - y_i)^2}\right) \times 100 \qquad \text{Equation 3.3}$$

where: SSE (sum of squared errors) is the quadratic summation of residual values and residual (Res) is the difference between the true and the fitted value $(x_i - y_i)$.

3.3.2. The number of factors or eigenvectors (rank) selection

The quality of the chemometrics model also depends on the choice of the correct number of factors included; this is also called the rank of the model. The number of PLS vectors used is defined in the QUANT program by the size of the “rank”. Choosing a too small rank results in under-fitting so that not all features are explained by the model. On the other hand, including too many factors (rank too high) leads to overfitting and only adds noise, and degrades the model (Fernandez-ahumada et al., 2010). The root mean square error of cross validation (RMSECV) values versus the rank were plotted to determine the optimum rank in Figure 3.4 and the rank versus R^2 in Figure 3.5. If the RMSEP/CV is plotted against the rank used in each model, a minimum can be observed in this graph, indicating the optimum rank. Apparently, the model improves drastically up to rank 4, with rank 5 and 6 still giving slightly better predictions. However, ranks higher than 6 barely improve the model and basically represent the addition of fluctuations (noise, temperature differences of the samples etc.) which eventually leads to degradation of the result. Restricting the calculation to lower ranks saves processing time as the calibration set contains more samples.

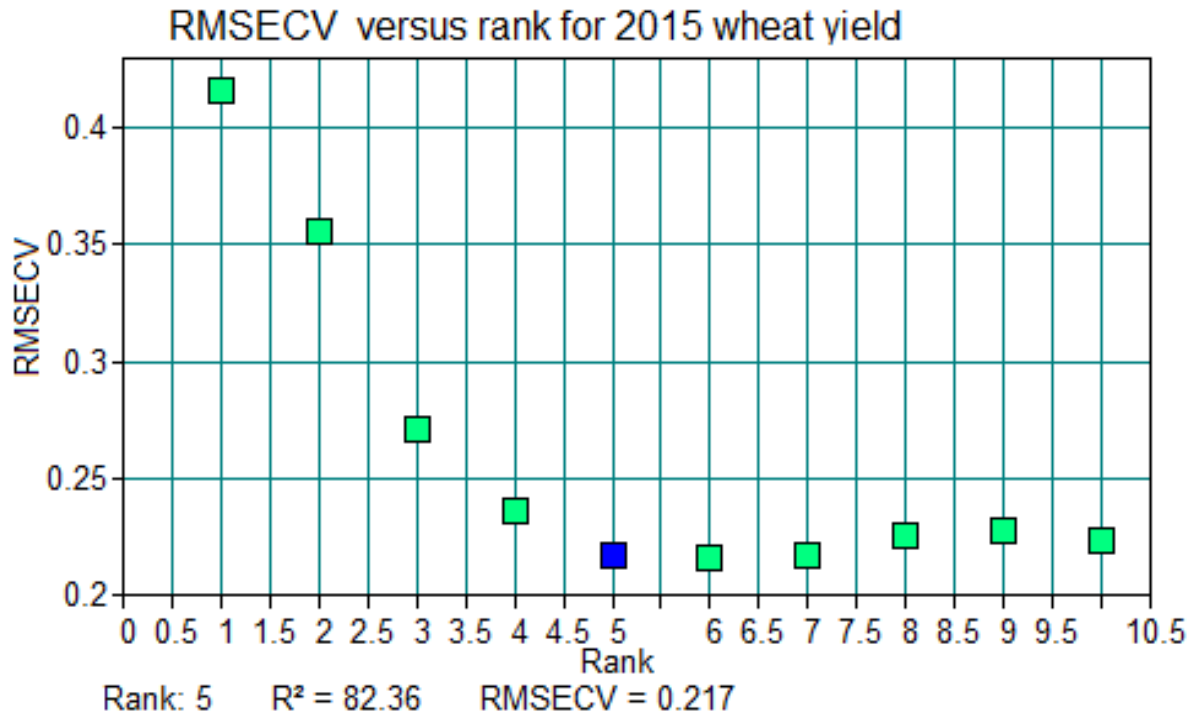


Figure 3.4: Plot of root mean square error of cross validation versus the rank for 2015 yield (block averages).

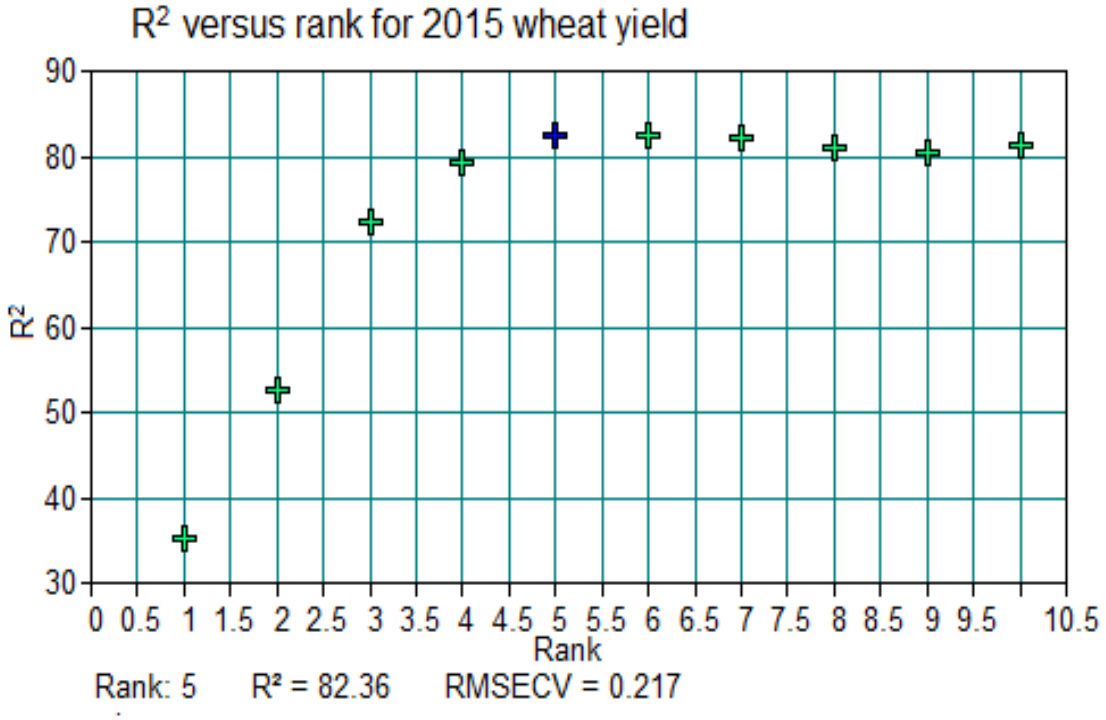


Figure 3.5: Plot of the coefficient of determination versus the rank for 2015 yield (block averages)

3.3.3 Spectral preprocessing methods

The spectral pre-treatment was done to eliminate spectral noise and ensure good quality results. Preprocessing included two procedures, which are de-noising, and data transformation. De-noising, the fringe bands with large noise in the spectra of each soil sample were removed. The removed wave ranges were those below 1111 nm and those above 2500 nm (below 9000 cm^{-1} and above 4000 cm^{-1}), as shown in Figure 3.6 grey shaded. Leaving the remaining part of the overall wavelength (1111–2500 nm), depicted as white are on Figure 3.6 to actually perform the calculations. In other words, the region have been removed do not contain any useful spectral information as noise prevails.

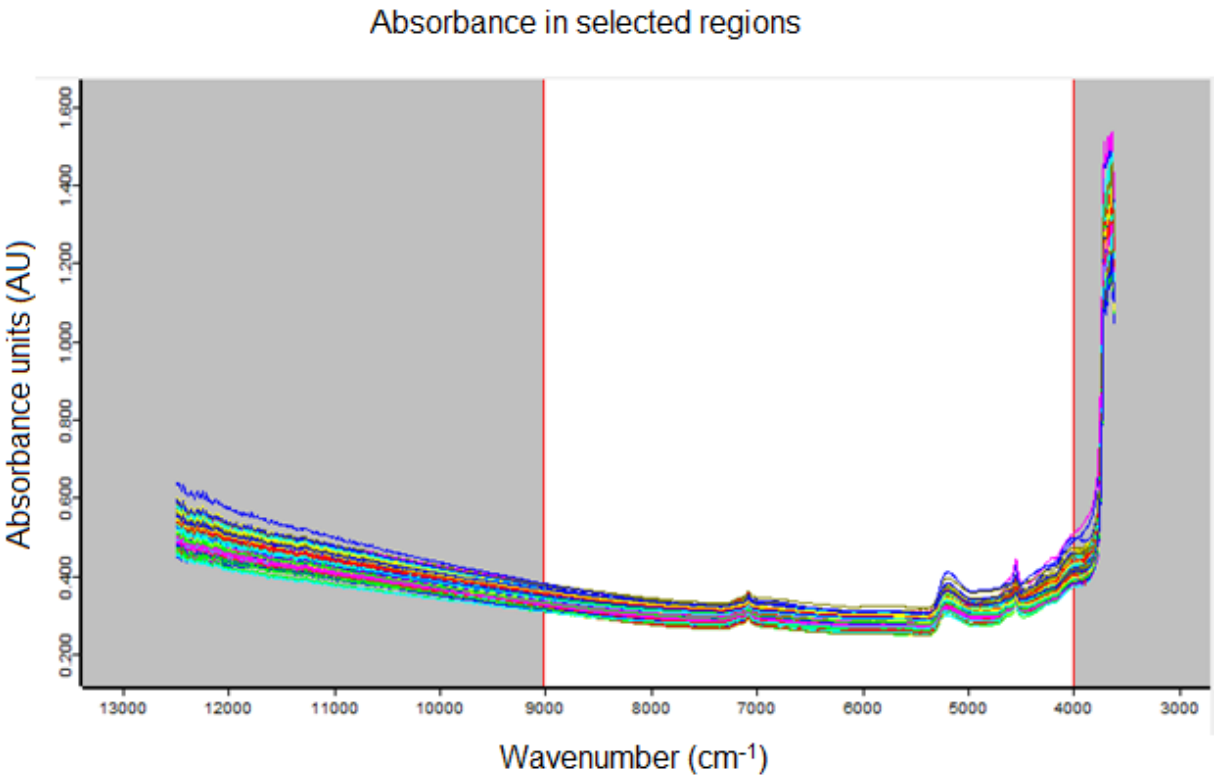


Figure 3.6: Absorbance in selected regions versus wavenumbers.

For data transformation, all the soil spectra were smoothed using the Savitzky-Golay method with first derivative and straight-line subtraction before performing multivariate data analyses. An example of transformed spectral signatures using these methods is illustrated in Figure 3.7. This was done to diminish the influence of differences in the optical environment and sample grinding and sieving processes (Zhou et al., 2014). These spectral pretreatment

methods used in our study have been found to be an optimal spectral pre-treatment in similar studies (Awiti et al., 2008; Lin et al., 2017).

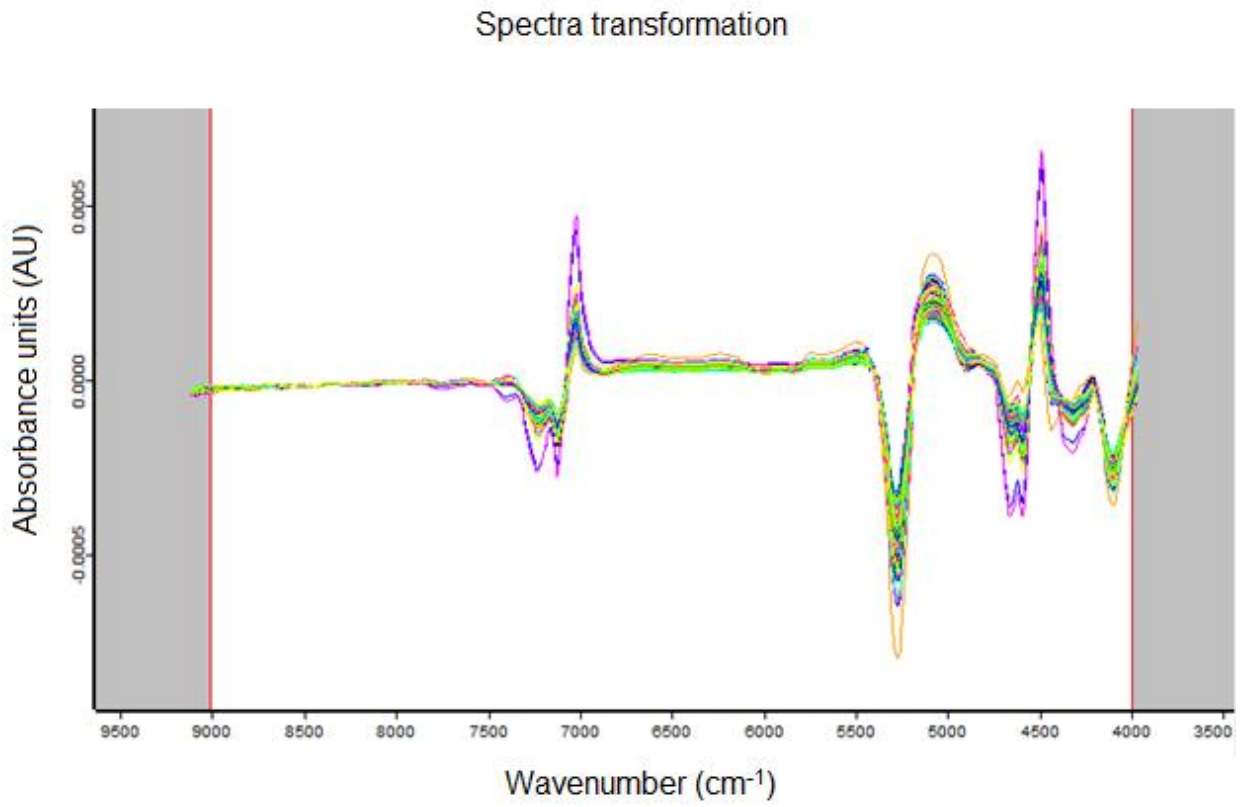


Figure 3.7: Spectral signatures that have been transformed using the first derivative and straight (base) line subtraction.

The selection of an appropriate pretreatment method is very important for the establishment of the NIR prediction models. As mentioned above, the NIR spectra are affected by many factors, including collinearity, physical properties, light scattering, and machine noise (Lin et al., 2017). They are easily influenced by differences (sample particle size, light intensity, measurement conditions, etc.), baseline variations, and substantial noises individually. Therefore, the pretreatment should be applied to minimize irrelevant and useless information of the spectra and increase the correlation between the spectra and the measured values.

3.4 Results and discussion

Calibration models to predict selected soil properties and wheat yield were developed using the partial least squares regression method. The number of components in PLS analyses was

determined using the Quant2 statistical software following the criteria explained in section 3.3.2.

3.4.1 Constructing the model

The data from the standard laboratory analysis was entered into the NIRS database (OPUS-QUANT 2) for the soil properties and wheat yield to derive a relationship with the spectral absorbance. Calibrations were carried out using the spectral signatures. In order to test whether adequate prediction models may be developed using the NIR spectral signatures, a calibration curve between the measured and predicted values was constructed (Figure 3.8). The calibration curve in Figure 3.8 (a) has a correlation coefficient (R^2 = 0.90), root mean square error of estimation (RMSEE = 0.0969) and ratio of performance to deviation (RPD = 3.2) for the normalized data (a). The results for Figure 3.8 b which is also a calibration curve (block averaged wheat yield) are; R^2 = 0.85, RPD =2.62, and RMSEE = 2.03. Results from these calibration curves were considered sufficient; the methods could be used to make predictions for the rest of the properties and wheat yield.

In addition, the black and red points in Figure 3.8 indicate values that may be either horizontal or vertical outliers. Which means their removal will tilt the regression model. The Quant2 program gives an option to remove these outliers but according to Miller & Miller (2010), outliers must only be removed with valid and proven reasons. It is not ideal to remove them without doing any background check from the reference data and a decision was taken, not to remove them in order to model reality as much as possible. The results from the wheat yield graphs appear in a form of two clusters, indicating the higher and the lower yielding areas within the same field.

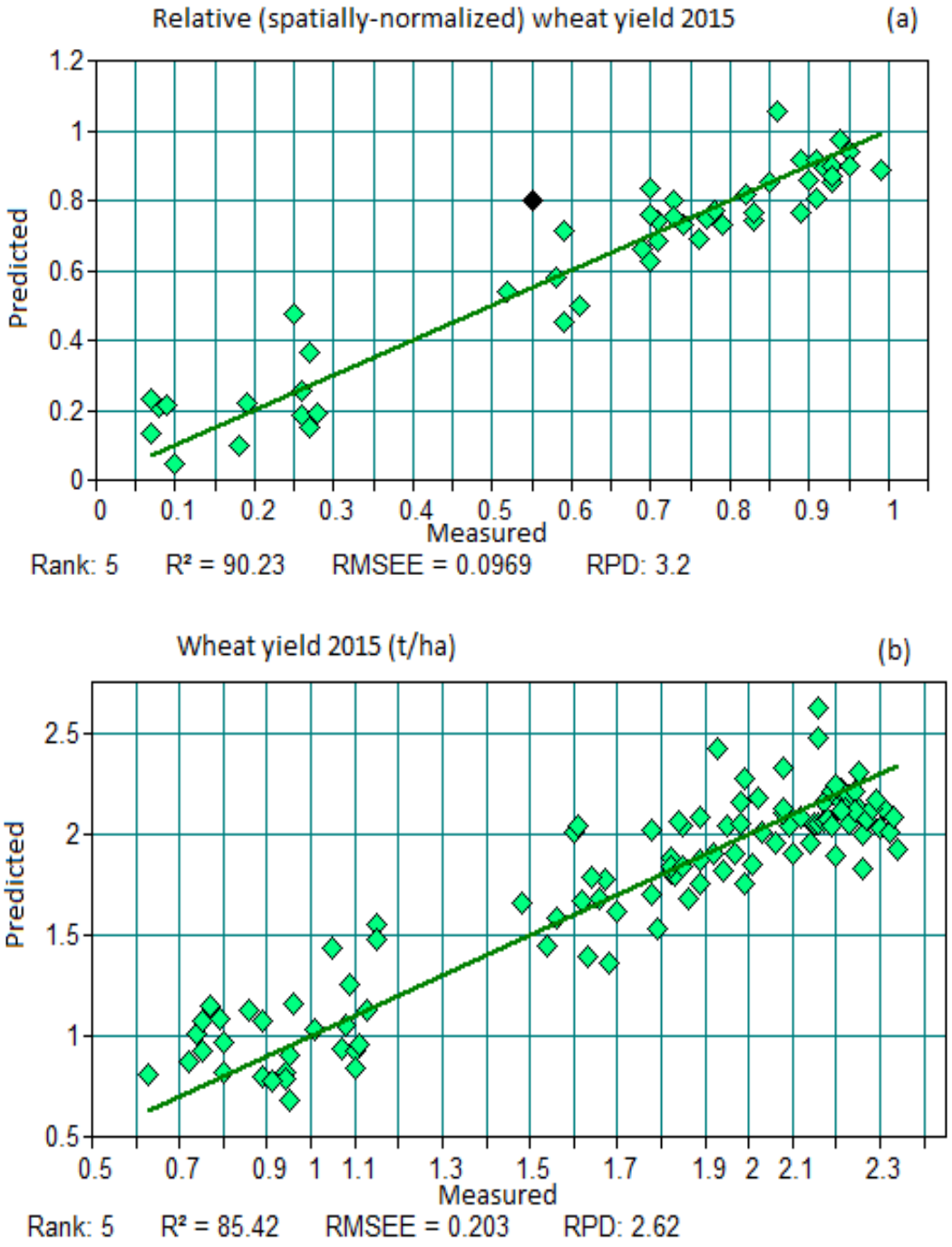


Figure 3.8: Calibration curve between measured and predicted wheat yields using the near- infrared spectral signatures; (a) relative wheat yield and (b) block averaged wheat yield.

3.4.2 Model cross-validation and prediction accuracy

Statistical tests were performed to cross-validate the performance of the model as mentioned in section 3.3. Prediction accuracy was assessed using the determination coefficient (R^2), root mean square error of cross-validation (RMSECV) or the root mean square error of estimation (RMSEE) for calibration and the ratio of performance to deviation (RPD). The rank was used to select an optimum number of factors for every system to avoid over or under fitting and collinearity. High values of R^2 and RPD, with low RMSECV values, represent good predictions. Generally, if 1.5<RPD<2, it indicates that the model only roughly estimates high and low values of predictions. If 2.0<RPD<2.5, it indicates that the model has a better predictive ability, while if 2.5<RPD<3.0, the model has very good predictive ability and if RPD>3.0, the model has an excellent predictive ability (Zhou et al., 2014).

The differences between the measured and predicted values were also calculated as depicted as an example on Figure 3.9 below. This was done to check how close the predicted values are to the reference measured values and values close to zero indicate high accuracy.

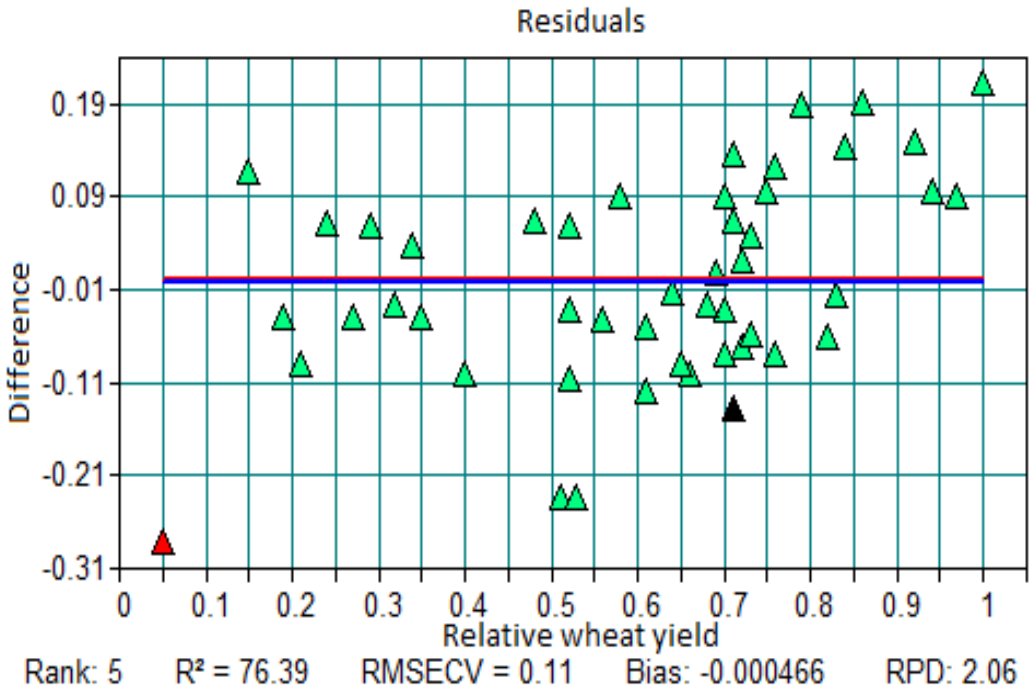


Figure 3.9: Differences between observed and predicted values of the relative wheat yield.

3.4.3 Covariance of the near- surface soil NIR spectra with wheat yield.

Shown below in Figure 3.10 is the illustration of the measured wheat yield (points) for the year 2015, Figure 3.10 (a) shows the prediction results without any spectral pre-treatment measures and no optimization or region selection occurred. Meanwhile, Figure 3.10 (b) shows the prediction results after limiting the frequency region and applying preprocessing procedures before doing the predictions. There are visual differences between the two Figures; Figure (a) has a rank that is 7, R^2 of 0.77, RMSECV of 0.235 and RPD of 2.07. However, after optimizing and preprocessing the same data, there has been an improvement on the R^2 value (0.80) and RPD (2.22) with a decrease in number of factors (Rank = 5) and a slight decrease in error (RMSECV = 0.22).

Therefore, the preprocessing methods appeared to have a vital role in improving the model and the rest of the results have been reported after there has been optimization and region-selection (preprocessing) of the spectral data where necessary. The prediction results for the wheat averages per block for 2015 are illustrated in Figure 3.11 below. More accurate results with high predictive ability were obtained for the year 2015 with a coefficient of determination of 0.82, RPD of 2.38 and with an error of 0.217 t/ha.

The spatially normalized wheat yield results are shown in Figure 3.12 below with R^2 of 0.81, RMSECV of 0.128 and RPD of 2.27 for the 2015 yield data. Data normalization allows for the data to be rescaled to 0-1 range (0 - minimum, 1 - maximum values). This was done in order to make it possible to compare the 2015 yields with those obtained later in the 10-year trial and be able to make a spatial assessment of the field performance. The rescaling of the data does not affect the regression model characteristics (R^2, and RPD), but affects the RMSE, which is not expressed in relative terms of the [0, 1] scale.

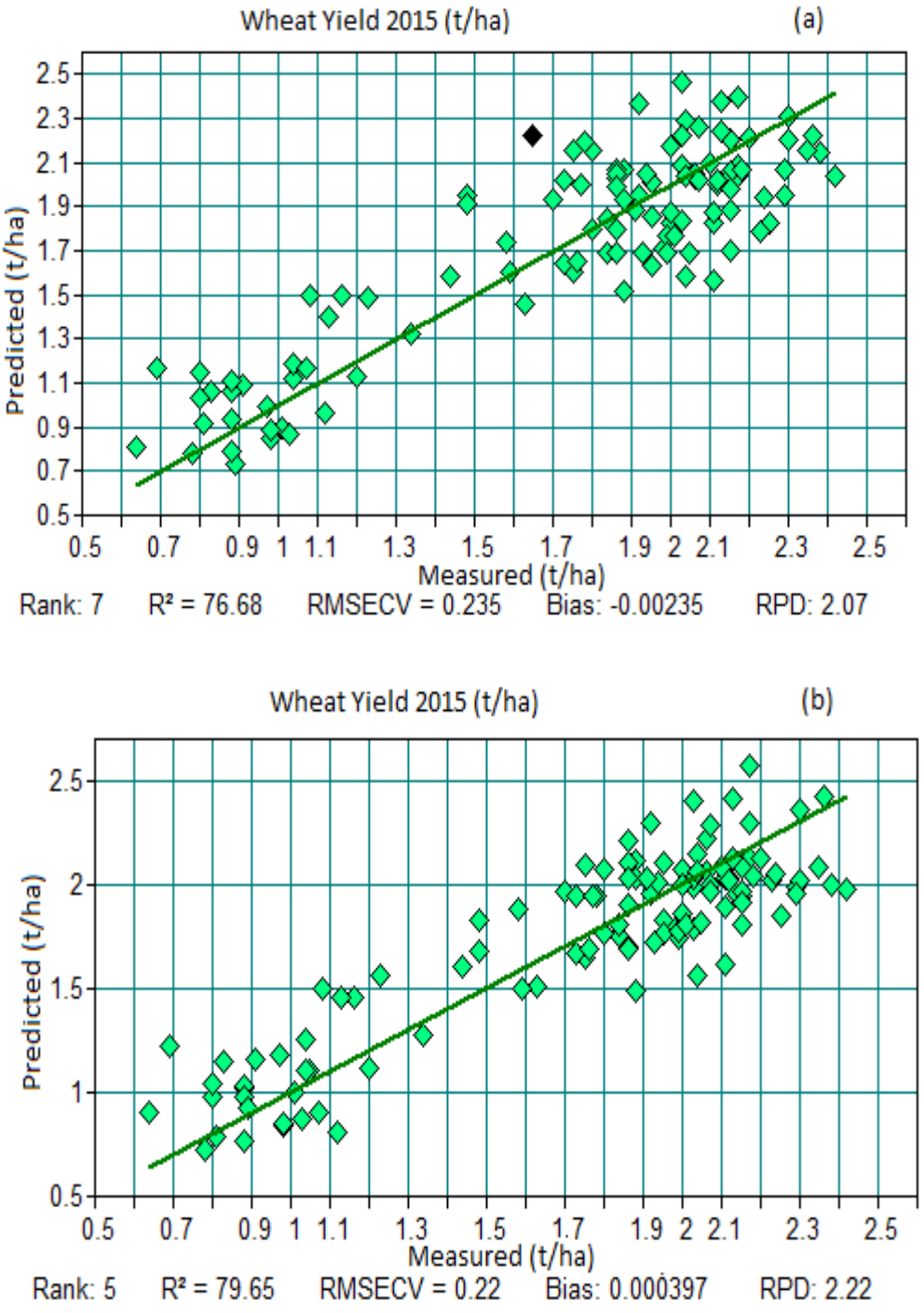


Figure 3. 10: Relationship between measured and predicted wheat yield (points) for 2015 to validate the near-infrared absorbance spectroscopy prediction models developed using the OPUS (Quant 2) software, (a) no spectral pre-processing (b) spectral pre-processed

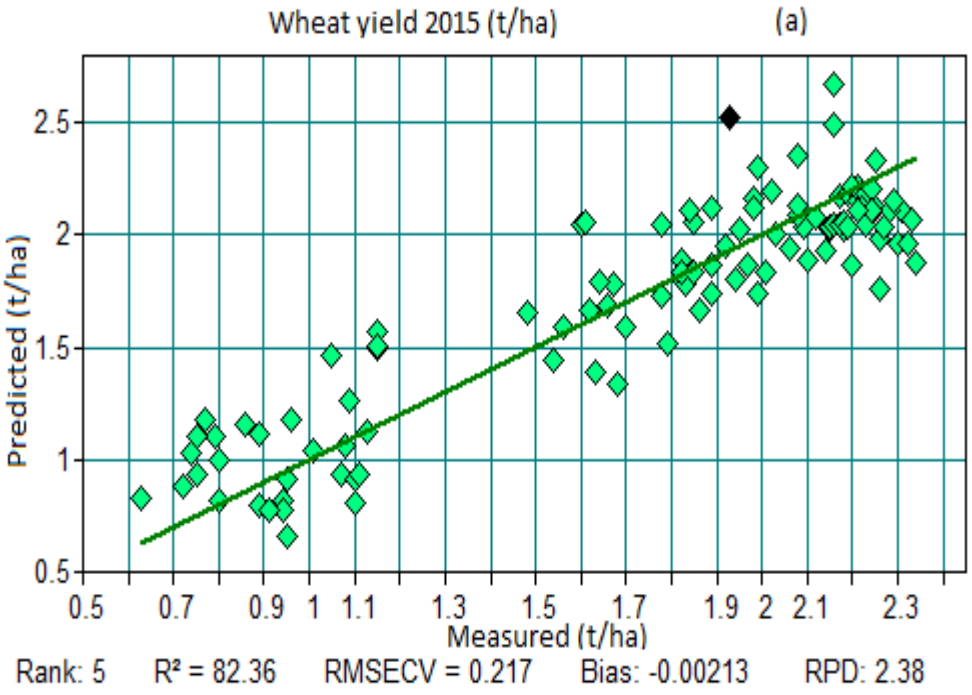


Figure 3.11: Relationship between measured and predicted wheat yield (block averages) for 2015 to validate the near-infrared absorbance prediction models developed using the OPUS (Quant 2) software.

The results for the models developed using the near-infrared spectral signatures to predict relative wheat yield, show promising results with reasonable correlation coefficients. Results for 2015 wheat yield predictions for point observations have an R^2 of 0.80, for average wheat yield per block for 2015 is 0.82, for relative wheat yield for 2015 is 0.81. These results are more or less the same as those reported by Mashaba (2016) who obtained R^2 of 0.73 for the regression models that predicted absolute winter wheat yield from normalized difference vegetation index (NDVI) and weather data. Also, almost all of their RPD's are greater than 2.0 which is an indication that the models have a acceptable predictive capability.

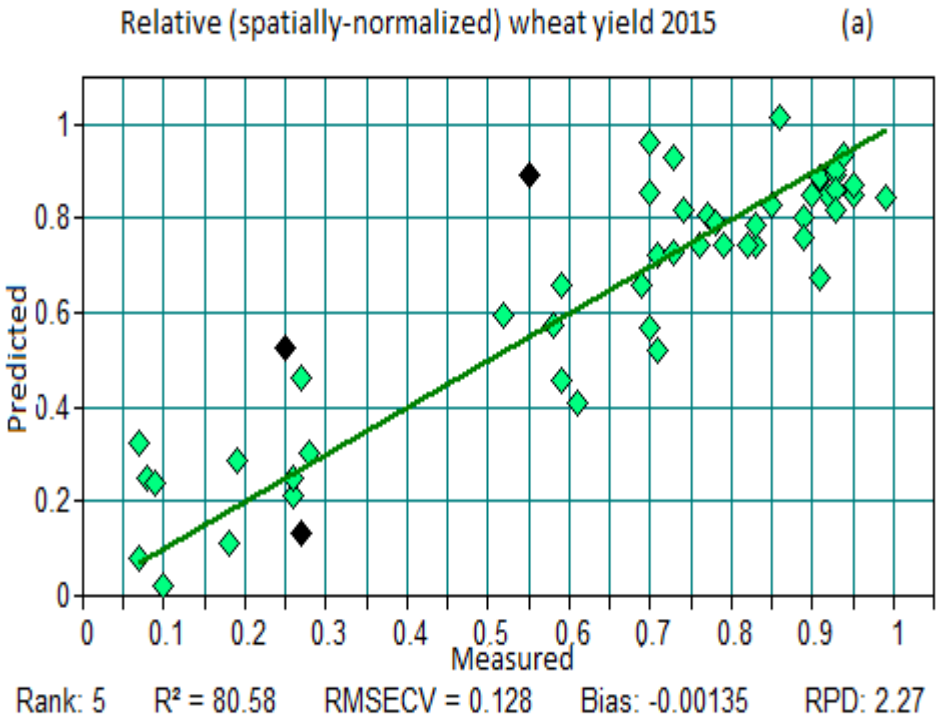


Figure 3.12: Relationship between measured and predicted relative wheat yield for 2015 to validate the near-infrared absorbance prediction model developed using the OPUS (Quant 2) software.

Here, prediction uncertainty was characterized on the set level (through the optimization of the calibration model) to make sure errors are eliminated and good prediction models are developed. Predictions were expected to be perfect, however there appears to be additional contributions to the RMSECV which would originate from a number of factors. As it was stated that on the graphs there are visual outliers (red and black points). Such outliers and irrelevant variables deteriorate any sound statistical evaluation of the model and might lead to misleading statistics (Klaas et al., 2005). Amongst the contributing factors to the uncertainty might be the optimum number of factors included (the rank), as some of the results with big RMSECV values have ranks which are greater than 6 or less than 4. Having a too small rank results in under-fitting and not all features get to be explained by the model. On the other hand,

having too many factors (rank too high) leads to overfitting and only adds noise, and degrades the model.

The outlying behavior can also come from the reference values because of: irregular behavior of the instrument, incorrect sample preparation, pollution in the sample, extremely high or low concentration of the analyte (Klaas et al., 2005). These outliers were not removed in this study because the samples were prepared in several dilution steps. Consequently, there are several uncertainties to be accounted for upon estimating the final uncertainty of the reference values and statistically for such reasons they cannot be removed. In this extreme case, RMSECV would mainly estimate the standard deviation (square root of the variance) of the measurement error of the reference measured values and it would not relate to the true prediction uncertainty at all (Klaas et al., 2005). Therefore, generally the prediction models developed in this study remain to be good ones as they have higher RPD's and reasonable correlation coefficients with the most optimum numbers of eigenvectors (ranks).

3.4.4 Covariance of the NIR spectra of the soil samples with some commonly determined soil properties

Figure 3.13 illustrates scatter plots of measured and predicted values obtained with the most accurate models and conventionally determined values, for the validation set for each soil property. They are all summarized in the following Table (3.1). The results obtained show a quite good potential for NIR to be used successfully for predicting the selected soil properties. The prediction model with an RMSECV of 1.5 explained about 70% of the variation on predicted values of cation exchange capacity (CEC) with an RPD of 1.88. With the rest of the observed soil properties results on the table also falling on the acceptable standard ranges.

Table 3.1: Soil properties' prediction results

Soil properties	R^2	RMSECV	RPD	Rank
CEC	71.57	1.5	1.88	6
S O C	64.02	0.33	1.67	6
Exchangeable calcium	69.48	1.36	1.81	7
pH	62.51	0.259	1.63	7

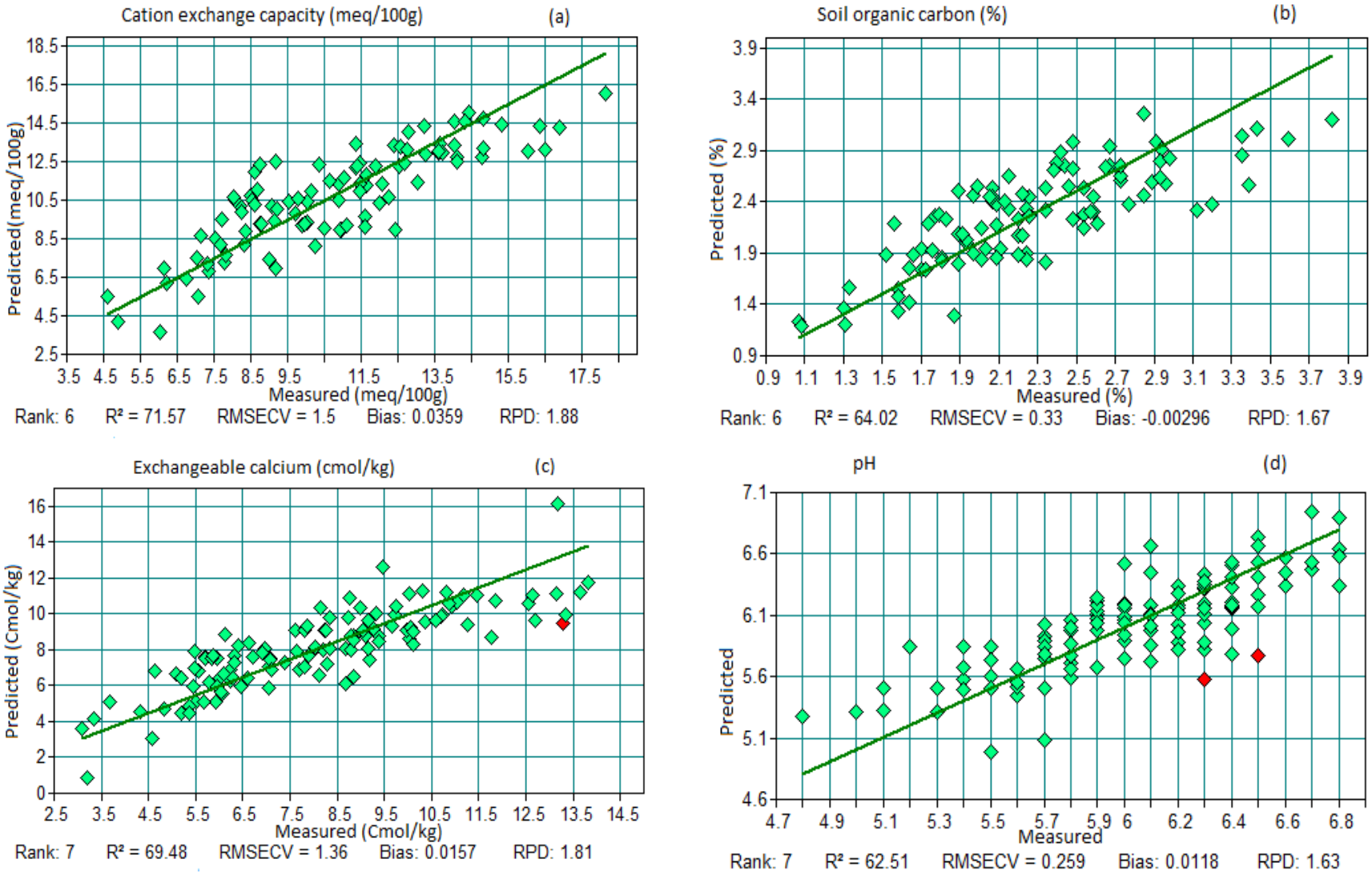


Figure 3.13: Relationship between measured and predicted various soil properties to validate the near- infrared absorbance prediction models developed using the OPUS (Quant 2) software, (a) CEC (b) SOC, (c) exchangeable calcium and (d) pH.

According to the results obtained in our study, the use of spectral libraries as a tool for soil properties' and crop yield prediction appears to yield promising results. A number of findings from various authors support our results such as those of Lin et al. (2017) who evaluated the ability of near-infrared spectroscopy to predict soil organic matter content (SOMC) with principal component regression (PCR) and obtained an R^2 of 0.71. With a determination coefficient up to 0.87 and RPD of 2.79, Wang et al. (2015) successfully analyzed the potential of Vis-NIRS to predict SOMC using two spectrometers, and the results showed that both spectrometers could achieve favorable results. These results are also in agreement with those of (Luis Galvez-Sola et al., 2015) who obtained a coefficient of determination for cross-validation (R_v) of 0.88 and the RPD of 2.84 in a study on rapid estimation of nutritional elements (Zinc (Zn)) on citrus leaves by near infrared reflectance spectroscopy and were considered acceptable. Increased accuracy on the selected models to make predictions was observed in a study by Asekova et al. (2016) on a study that estimated forage quality of soybean by near-infrared reflectance spectroscopy. On the study, two NIR equations developed for crude protein and crude fiber were observed to be the best prediction equations for estimating these parameters. They both had very strong correlation coefficients: 0.93 and 0.91 respectively. All of these comparable literature results prove the feasibility of the NIRs model in the prediction of soil properties and elements in plants and crop yields.

However, authors further complained that the noise and some irrelevant or collinear information included in Vis-NIRS can affect the accuracy of the method (Mohamed et al., 2018) . Therefore, the influence should be eliminated with applicable measures to ensure stable conditions before using the model. This is amongst the reasons why in this study (current-MSc work) it was made sure that the pre-treatment and region selection is done during the method development and all the models are developed on pre-processed spectra. The selection of appropriate pretreatment method is very important for the establishment of the NIR prediction model. As stated above, the NIR spectra are affected by many factors, such as collinearity, light scattering, physical properties, machine noise, and so on (Lin et al., 2017). The pretreatment could mine the weak signals and unnoticeable information through some transformations on the original spectral data. The combination and the sequence of different pretreatment methods need to be enhanced in the practical applications. In this study, the PLS-based prediction models for the wheat yield and soil properties after different pretreatment and optimization methods were developed using the training set, and the prediction set; results are listed in Table 3.1 (soil properties) and for the wheat yield in Figure 3.10-3.12. It can be seen that different methods have different influences as also mention by Lin et al. (2017) who obtained different results when using different methods or

procedures. The combination of the Savitzky–Golay (SG) filter for smoothing and a combination of the first-order and straight-line subtraction exhibit the most promising results.

3.4.5 Covariance of the soil NIR spectra with soil colour

Illustrated below in Table 3.2 are the prediction statistical results for soil colour. The colour has been predicted using the NIR spectral signatures both in the Munsell (HVC) and LAB notations. The spectral signatures appear to have a positive predictive ability for soil colour as there exist positive correlations across all the colour notations although they are a bit weak for the HVC notation. Acceptable RPD's for the L*a*b* notation were obtained, but those of HVC were not so good. Their optimal ranks were also mostly ridiculously big (greater than 6) which can serve as an indication of higher possibilities of overfitting, noise addition and model degradation.

Table 3.2: Soil colour prediction results using the NIR spectral signatures

Soil color notation	R^2	RMSECV	RPD	Rank
L*(D65)	82.97	0.498	2.42	8
a*(D65	56.85	0.669	1.52	9
b*(D65)	75.92	0.45	2.04	10
H	14.96	1.13	1.08	6
C	20.04	0.304	1.12	3
V	39.42	0.142	1.29	10

For the purpose of soil colour analysis where the colour of the soil in its own sense has slight significance, not forgetting that it is a major feature used in the classification and identification of soils (Viscarra Rossel et al., 2009). The main interest would be to correlate soil colour to other colour-related variables indicative of a specific soil property or process or production (Mattikalli, 1993). In our case, it can be noted that color has a strong relation with soil properties as it could be successfully predicted using the NIR spectral signatures of the soil surface, although it is not very strong for the HVC notation. These results (more especially L*(D65) and b*(D65) notations) are in agreement with those of Mattikalli (1993) who obtained an improved correlation coefficient of 0.8 after applying optimal rotational transformation of data with multiple linear regressions on spectral reflectance models to predict soil colour.

3.4.6 Correlation of the measured soil properties and wheat yield with NIR absorbance.

Table 3.3 gives the Pearson correlation coefficients between wheat yield, soil properties and the NIR absorbance. Due to the nature of the Munsell hue variable being a combination of both numbers and letters, a linear scale with substituted numerical values were used. The substituted values were 2.5YR = 2.5, 5YR = 5, 7.5YR = 7.5 and 10YR = 10.

The mean NIR absorbance has quite good correlations with wheat yield and with most of the measured soil properties as depicted in Table 3.3. With most of them being significantly correlated ($p<0.05$). The positive correlations of soil properties and wheat yield with NIR absorbance indicate that NIR is indeed an integrative soil characteristic that is related to multiple properties that were identified in our study and other studies. These results are in agreement with those of Ramirez-lopez et al. (2018) who worked on optimizing the use of Vis–NIR spectroscopy at a farm scale to predict soil properties robustly and found that there is a strong absolute correlation between soil properties and soil Vis–NIR absorbance by obtaining correlations larger than 0.6. However there are also some weak correlations and even negative correlations between the parameters, that might indicate that there are few parameters that it is slightly related to and that it cannot directly predict. These results are also indicating that no single soil characteristic appears to individually influence wheat yield performance but the combination of multiple properties has a huge impact on wheat performance, as the relationship between yield and soil parameters in most cases is significant.

Table 3.3: Pearson correlation coefficients between soil properties and wheat yield with NIR total absorbance

	Yield_15	SOC	CEC	Ca	pH	L*(D65)	a*(D65)	b*(D65)	H	V	C	NIR
Yield_15												
SOC	0.619*											
CEC	0.724*	0.800*										
Ca	0.688*	0.807*	0.962*									
pH	0.298*	0.433*	0.591*	0.697*								
L*(D65)	-0.412*	-0.543*	-0.639*	-0.684*	-0.565*							
a*(D65)	0.451*	0.0403	0.184	0.1109	-0.1217	-0.1646						
b*(D65)	-0.1451	-0.478*	-0.506*	-0.587*	-0.553*	0.736*	0.425*					
H	-0.0789	-0.1627	-0.1629	-0.1966	-0.330*	0.1810	0.1397	0.2208				
V	-0.277*	-0.398*	-0.502*	-0.554*	-0.482*	0.804*	-0.0993	0.644*	0.390*			
C	0.1778	-0.0489	-0.0597	-0.1366	-0.301*	0.284*	0.570*	0.690*	0.318*	0.388*		
NIR	0.475*	0.575*	0.631*	0.626*	0.474*	-0.718*	0.1169	-0.473*	-0.156	-0.503*	0.098	

3.4.7 Developing prediction models using colour and NIR spectral signatures

A colour-NIR model was developed to do predictions using the partial least squares regression in R statistical software and the same procedure as the one that was done in OPUS was followed. Guidelines followed to develop codes in R software were taken from Mevik and Wehrens (2018), the coding script is attached in appendix C. The developed model consists of the hue, chroma and value (HVC) colour notation and NIR sums as the predictor variables and in the example they were used to predict soil organic carbon (SOC). In overall, many of these variables are not significant; an R^2 of 0.25 for all the regressors was obtained. Also the only variable that showed significance is NIR (p-value<0.05), the rest of the variables had p-values greater than 0.05. This indicates a weak correlation between the regressors (HCV and NIR) and the response variable (SOC). The PLSR model was then fitted with four components. The validation results showed the root mean squared error of prediction (RMSEP) and the percentage of variance explained by the model. In order to check which components represent the best regression model, RMSEP versus components was plotted. This helps to determine the components with the lowest error (RMSEP) which would indicate the best component and analysis of the model be based on it (Mevik & Wehrens, 2018). Figure 3.14 shows the plot showing the criterion followed to decide how many components (optimal number of the components) to be involved in the model.

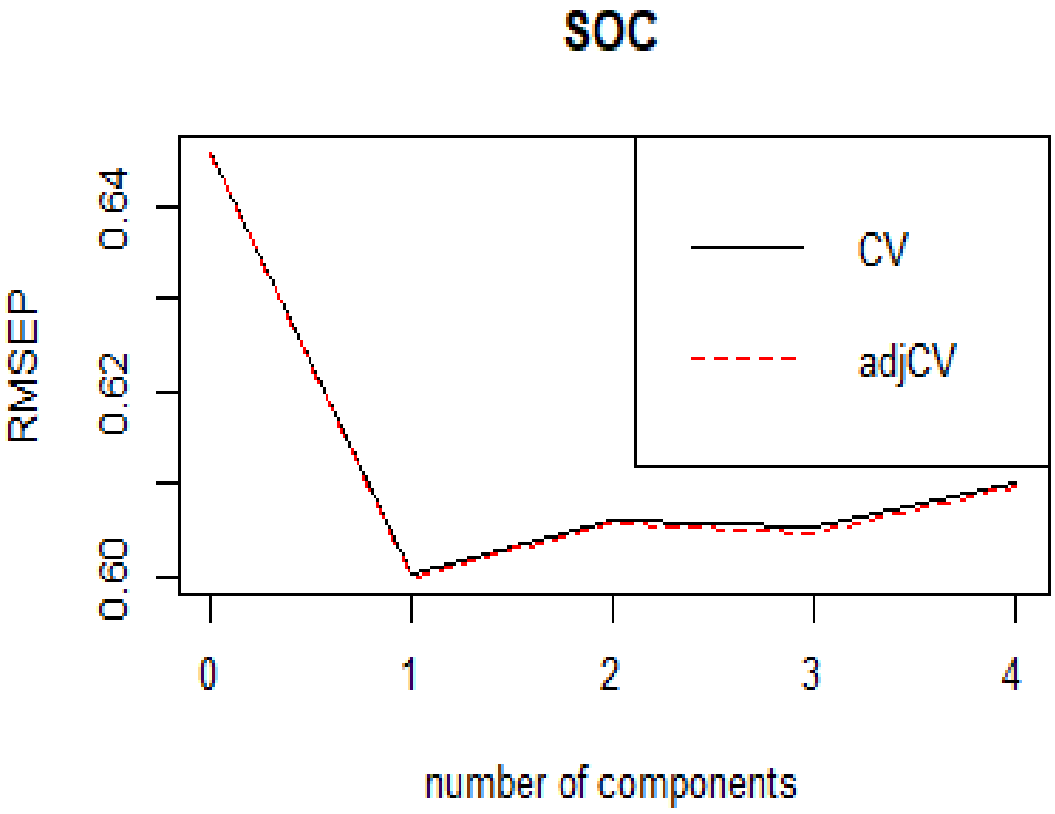


Figure 3.14: Number of components plotted against the RMSEP for the SOC prediction model developed.

From Figure 3.14 it can be deduced that, one component seems to be enough as it give the smallest error (RMSEP of 0.6). The model that seems to be the best has an R^2 of 0.2. Only about 20% of variation on Y (SOC) is explained by the model. After the number of components were chosen, the cross-validated predictions with one component versus measured values were carried out as illustrated in Figure 3.15 below. Some of the points in the Figure (3.15) follow the target line fairly, however there are other points that do not. This indicate a lot of noise as the points are also very scattered, and there is an indication of anomalies (Mevik & Wehrens, 2018).

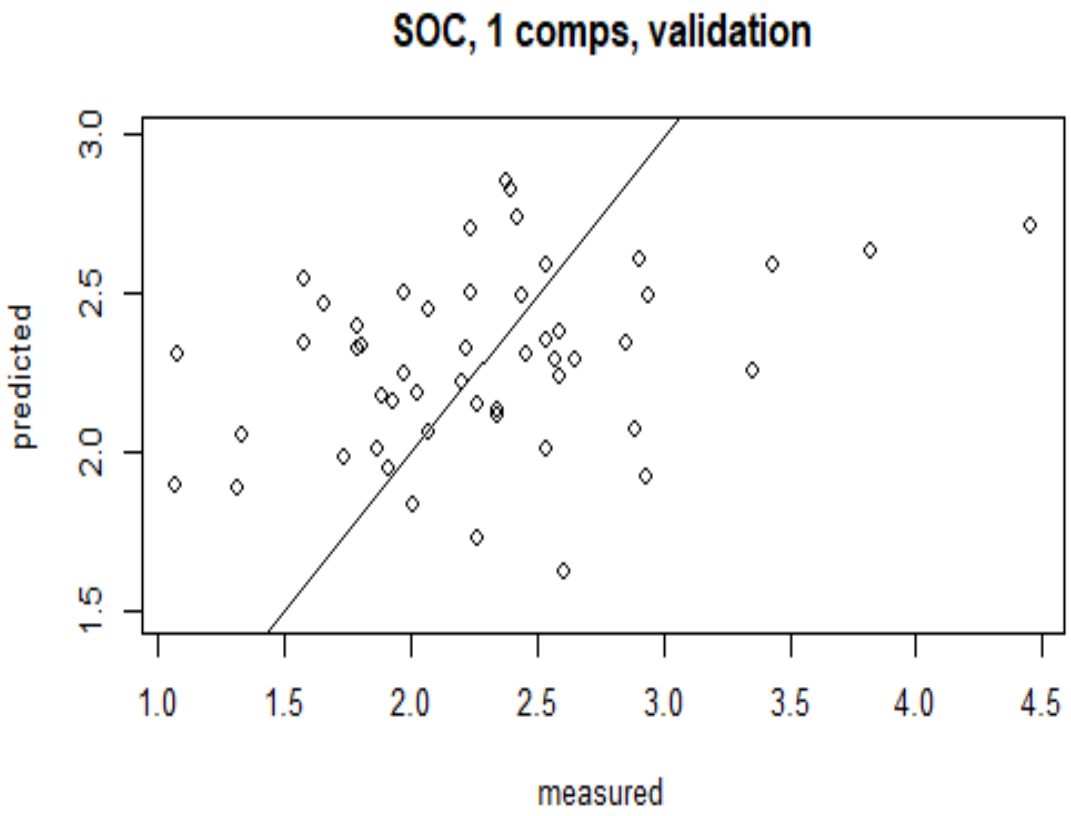


Figure 3.15: Cross-validation prediction results using one component.

The model developed using colour and NIR spectral signatures to predict soil variables (i.e. SOC) appear to be less accurate compared to the model that was developed using the whole NIR spectrum, which gave better soil organic carbon prediction results with an R^2 of 0.64 and RMSECV of 0.33. Therefore, only using the full NIR spectra to make predictions appear to yield results containing better predictive ability models.

3.5 Conclusions

The use of near-infrared soil spectroscopy may be an effective tool in predicting field-scale variation in crop productivity. It may be useful in optimized (as opposed to randomized) field trial design to achieve best results in experiments laid out in fields with known productivity bias due to variation in soil conditions.

The predictions of individual soil properties showed that the cross-validation results are comparable to and in agreement with other authors using PLS regression. To establish accurate and robust prediction models for wheat yield and soil properties, pre-treatment methods, sample selection methods, and wavelength optimization methods were applied. These preprocessing procedures appear to play an important role in the construction of NIRS prediction models. Results from this study show that the combination of the SG filter for smooth, first-order and straight-line subtraction can effectively eliminate the effects of noise and baseline drift and thus yield better results.

CHAPTER 4: Assessing temporal variations and changes that have occurred in the field after four years of crop rotation trial

4.1 Introduction

Variations in field conditions and in crop performance may be caused by various internal and external factors including soil quality and soil management practices. These variations are mostly linked either indirectly or directly to soil variability (Breysse et al., 2007). The fundamental key to managing and understanding crop variability needs a balanced approach that involves managing and understanding soil variability as well as its causes and effects (Bilgili et al., 2011; Sadegh et al., 2015). Various soil properties including the physical and chemical correlate differently with different landscape parameters. For example; clay content, sand content, and pH are well correlated with relative landscape position. Meanwhile, the water holding capacity and organic matter have also been observed to vary significantly with changes in slope position (Frogbrook et al., 2002, McBratney et al., 2003). It is therefore vital to characterize the variability of these soil properties to be aware of the significant effects they have on crop yield and performance within field trial experiments. This chapter focuses on evaluating the temporal changes that have occurred over time in the field, from the year 2015 to 2019. To observe changes for each field, the two spectra for 2015 and 2019 will be inspected to see whether are there any changes and identify the possible causes of them. Also, some selected soil properties including bulk density, soil organic carbon, and carbon stocks will be evaluated to see changes over the years.

4.2 Materials and Methods

4.2.1 Soil sampling, preparation and analysis

In 2019, surface soil samples were collected from Langgewens Research Farm before the field was planted. Samples were collected from the one hundred and twenty (120) plots of the field trial as illustrated in Chapter 3 (Figure 3.2). They were then taken to the laboratory, oven dried at 90°C for twenty-four hours and then weighed. Bulk density (ρ_b) was determined as the mass of oven-dried soil per unit bulk volume ($mg \cdot cm^{-3}$) (Al-shammary et al, 2018). Soil samples were then ground to pass through a 2 mm sieve to gravimetrically determine the stone content. The < 2 mm fraction of these samples collected in the study area were scanned using a Bruker Multi-Purpose Analyser spectrophotometer to obtain the NIR spectral signatures. This was done exactly following the NIR scanning methods explained in Chapter

3 (3.2.6). Total NIR absorbance was calculated as the sum of absorbance across all the measured wavelengths (wavenumbers: 4000-9000 cm^{-1}). The difference in absorbance (ΔA) between 2019 and 2015 observations was calculated per wavenumber for each sample. The soil organic carbon content was determined in the same laboratory using the same method described in Chapter 3.

4.2.2 Crop sequences of the trial area

The trial consisted of 3 cropping systems with varying degrees of crop diversity as detailed in Chapter 3. This entailed the allocation of 120 plots to accommodate the 4 replicates of the trial spanning over 10 years of crop rotations that started in 2016. In order to assess the variability in production, the whole trial area was planted to wheat as a uniformity trial in the year 2015 and for the other years, crops were rotated as shown in Table 4.1. Figure 4.1 shows the spatial allocation of all the rotation sequences starting in year 1 at a different point of 10 rotations.

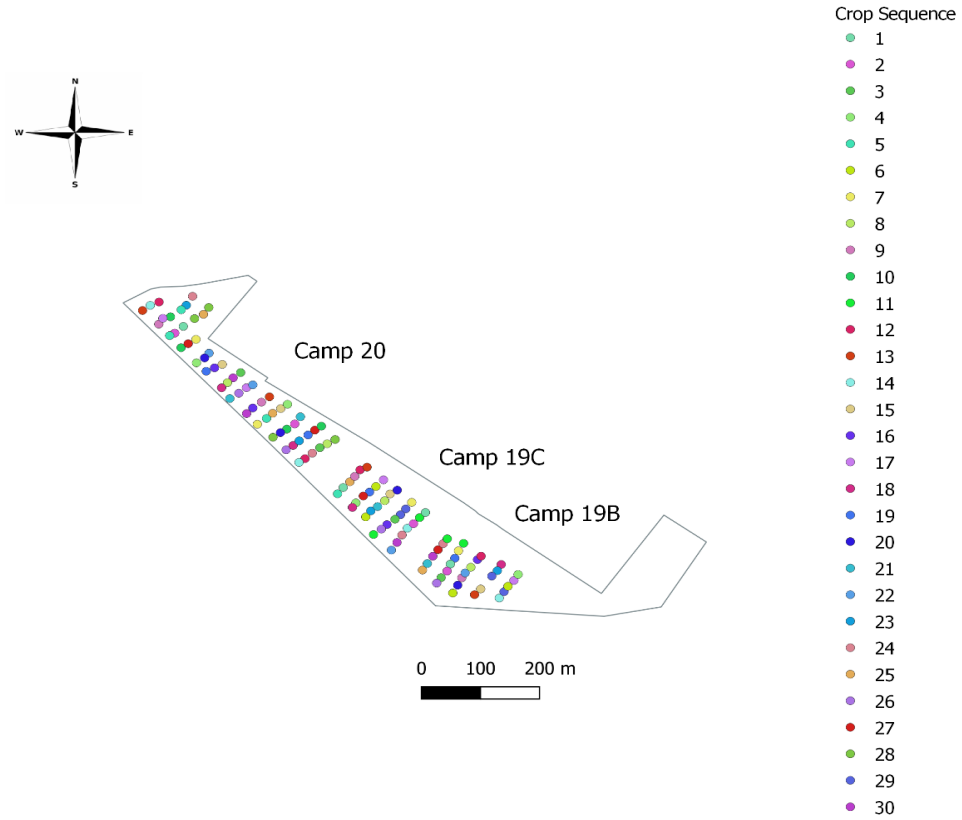


Figure 4.1: Illustration of the crop sequences detailed in Table 4.1

Table 4.1: Crop sequences over the years of the field trial at Langgewens research farm.

Crop Sequence	2015	2016	2017	2018	2019
1.	Wheat	Wheat	Legume cover	Wheat	Barley
2.	Wheat	Legume cover	Wheat	Barley	Canola
3.	Wheat	Wheat	Canola	Wheat	Canola
4.	Wheat	Canola	Grass cover	Wheat	Canola
5.	Wheat	Legume cover	Wheat	Wheat	Canola
6.	Wheat	Legume Cover	Wheat	Wheat	Canola
7.	Wheat	Barley	Canola	Grass cover	Faba beans
8.	Wheat	Vetch	Wheat	Wheat	Faba beans
9.	Wheat	Wheat	Barley	Canola	Grass cover
10.	Wheat	Wheat	Wheat	Canola	Grass cover
11.	Wheat	Wheat	Wheat	Canola	Grass cover
12.	Wheat	Wheat	Canola	Wheat	Legume cover
13.	Wheat	Wheat	Faba beans	Wheat	Legume cover
14.	Wheat	Oats	Linseed	Wheat	Legume cover
15.	Wheat	Faba beans	Wheat	Oats	Linseed
16.	Wheat	Grass Cover	Faba beans	Wheat	Oats
17.	Wheat	Wheat	Canola	Grass cover	Vetch
18.	Wheat	Canola	Wheat	Canola	Wheat
19.	Wheat	Grass cover	Wheat	Canola	Wheat
20.	Wheat	Canola	Grass cover	Faba beans	Wheat
21.	Wheat	Wheat	Wheat	Faba beans	Wheat
22.	Wheat	Wheat	Canola	Grass cover	Wheat
23.	Wheat	Canola	Wheat	Legume cover	Wheat
24.	Wheat	Faba beans	Wheat	Legume cover	Wheat
25.	Wheat	Linseed	Wheat	Legume cover	Wheat
26.	Wheat	Wheat	Oats	Linseed	Wheat
27.	Wheat	Canola	Grass Cover	Vetch	Wheat
28.	Wheat	Wheat	Legume cover	Wheat	Wheat
29.	Wheat	Wheat	Legume cover	Wheat	Wheat
30.	Wheat	Grass cover	Vetch	Wheat	Wheat

4.3 Results

4.3.1 Inspecting total NIR absorbance

To observe changes that have occurred in the field, the measured spectra (NIR absorbance) for the years (2015 and 2019) were plotted on the same axis as shown in Figure 4.2. From these results in Figure 4.2, it can be observed that there are substantial changes in the spectra. Total NIR absorbance in 2019 is higher than in 2015 for most of the plots, as the trend line for 2019 is above the one for 2015 (Figure 4.2). In addition, a large spread of the data points was observed as the standard deviation of 11.1 was obtained and an average of 7.7. To check whether the differences are significant across the field, a t-test was performed on the spectral absorbance between the year 2015 and 2019. The obtained p-value ($p=5.76781x10^{-9}$) indicates high significance in differences across the trial area.

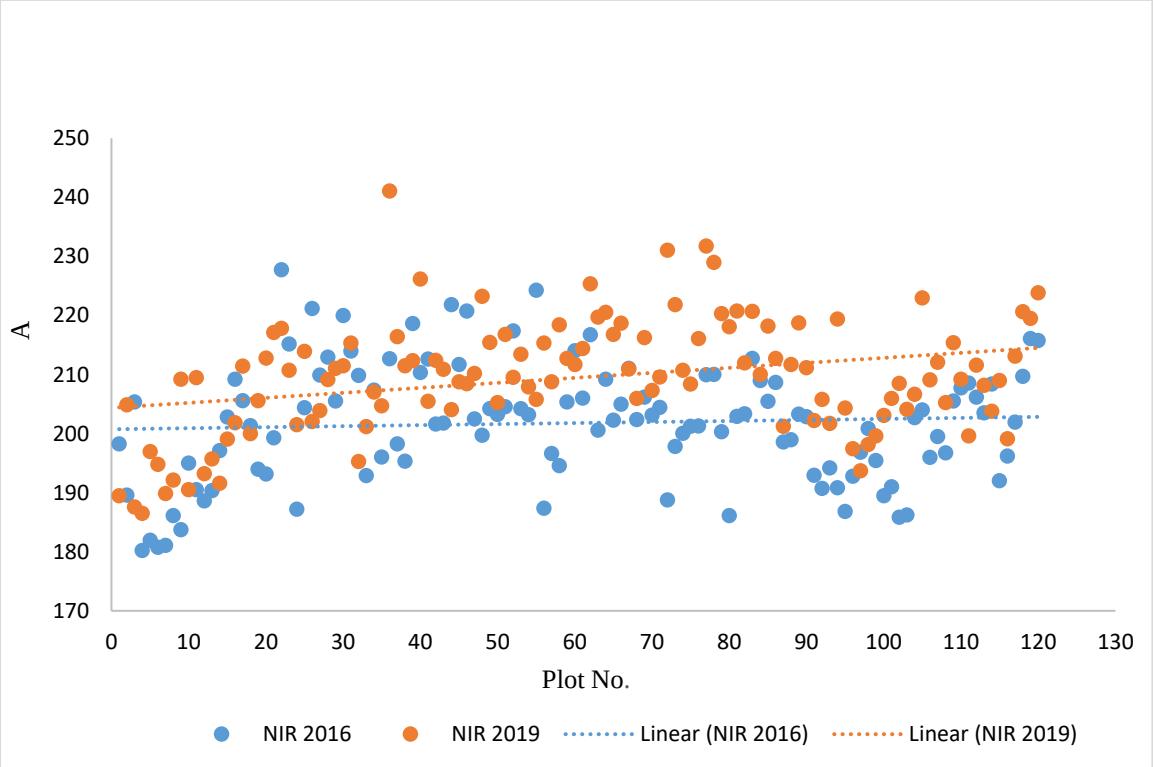


Figure 4.2: Total NIR absorbance (A) differences between the years 2015 and 2019 for all the experimental plots.

4.3.2 Inspecting the differences in the spectra for different years

To inspect variations in the spectra of the year 2015 and 2019, differences (Δ) across the absorbance for the plots per wave band were calculated. Results plotted in Figure 4.3 illustrate the absolute increase in absorbance from 2015 to 2019. The error-bars included in the graph demonstrate the standard error within each wave band. From the Figure, it can be observed that on average, there are visual changes that have occurred; these changes are definitely the overall shift of the absorbance towards higher ranges. As the wavenumber increases, also the absorbance increases at a consistent rate from above 5000 cm^{-1} wave band. It is actually around the wavenumbers of 5500-7000 cm^{-1} in the spectrum where differences appear to be consistent and that gives the much clearer understanding of the change. In these regions, as time passed on with many changes in the field, the soils started to absorb more light.

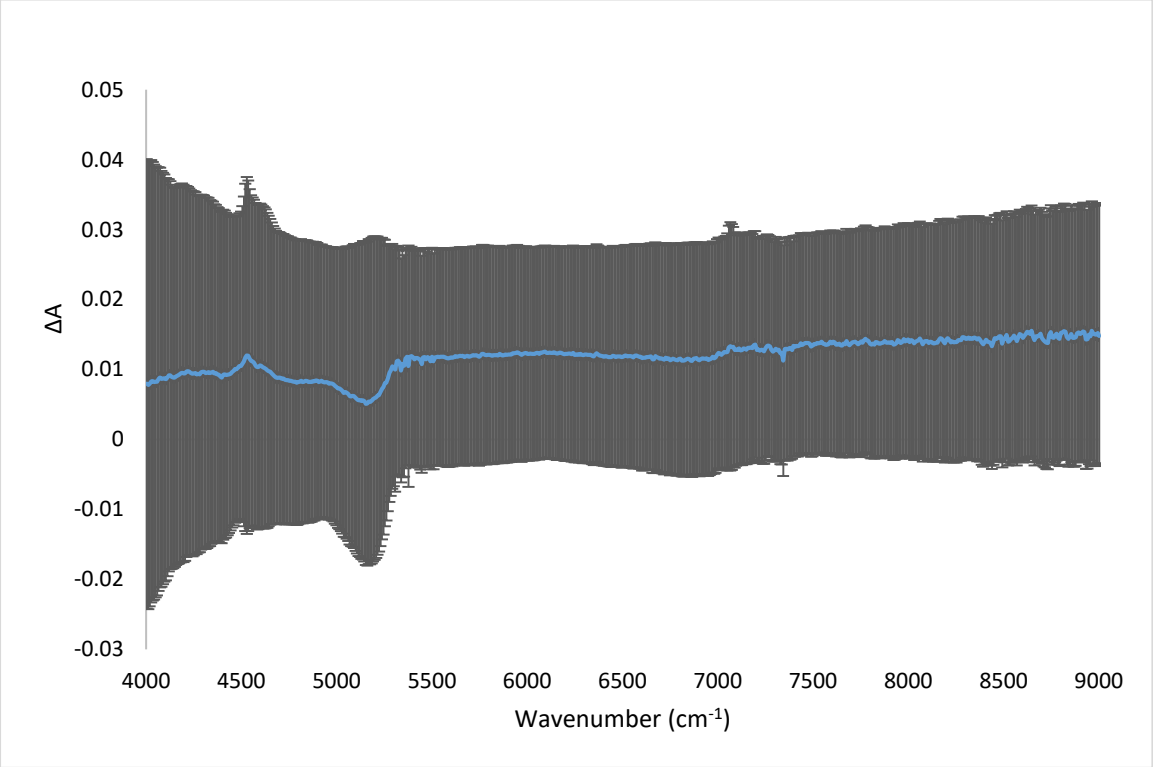


Figure 4.3: Mean ΔA vs wavenumber for all soil samples (2019-2015). The error-bars indicate the standard error for each wavenumber.

4.3.3 Analysis of NIR absorbance per bandwidth and per treatment

This section further evaluates whether the changes in the spectra illustrated in Figure 4.2 are reflected in wheat yield. Also, it identifies where the changes are and checks whether the magnitude of the changes may somehow be related to the treatments that were applied in the trial. The changes might be due to a lot of internal and external factors including the changes

in the crop rotation sequences over the period of the experimental years. Statistical analysis (t-tests) per bandwidth was performed, very small p-values were obtained. The differences between 2019 and 2015 are highly significant in all of the 649 bands that were used for analysis (p-values<0.05). This shows that there are significant differences across the plots at different bandwidths. After noticing these differences, the plots were grouped according to treatments. Average absorbance per treatment was calculated and plotted against the wavenumbers (Figure 4.4). This was done in order to be able to examine whether are there any differences within the treatments and how do they differ from the mean.

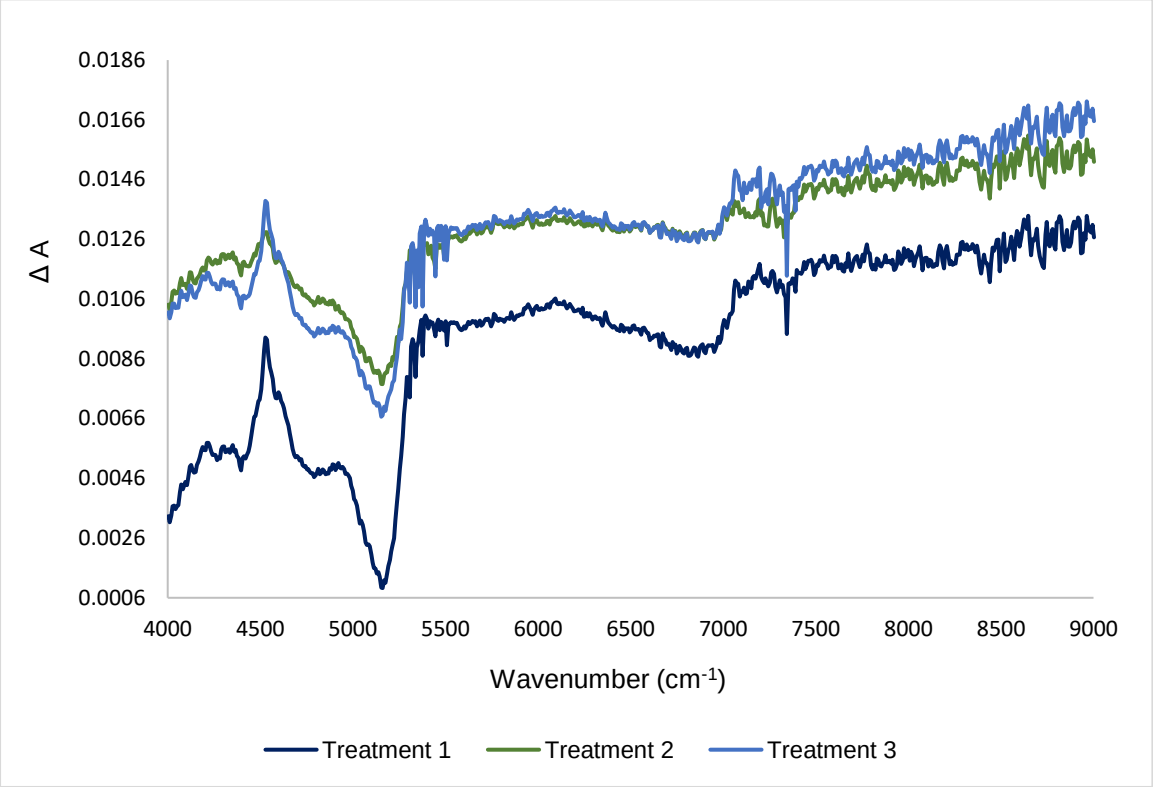


Figure 4.4: Average ΔA per treatment (averaged per wave number from 40 replications) and plotted against the wavenumbers, to explore differences per treatment.

From Figure 4.4, it can be seen that treatment two and three have higher absorbance and they follow a similar trend in terms of absorbance per wavenumber. However, treatment 1 differs from both of them (2 and 3), it has the smallest absorbance.

Since there are noticeable changes, further analysis were carried out to check if there are any statistical significant differences within and between the treatments. The calculated p-values

between the treatments are presented in Table 4.2. From the results it can be seen that indeed treatment 1 is significantly different from the two ($p<0.05$), and treatment 2 and 3 are not statistically significant ($p>0.05$). Then the bands that showed a lot of variation were further inspected (wavenumber range from 5160-4004 cm^{-1}). The results are still not very different from the whole NIR spectral region, they are all significant.

Table 4.2: Treatment t-test results

Treatment	p-value	selected bands p-value
1 vs 2	**$7.2x10^{-95}$**	**$4.2x10^{-47}$**
1 vs 3	**$3.8x10^{-114}$**	**$1.9x10^{-38}$**
2 vs 3	0.15	0.009

4.3.4 Assessment of variations in certain soil properties and yield

4.3.4.1 Bulk density analysis

Changes that have occurred over the period of 4 years in soil bulk density (ρ_b/BD) were assessed. Figure 4.5 gives a graphical illustration of the bulk density observations for each of the 120 plots for the year 2015 and 2019. From the results it can be seen that there are differences in bulk density of the two years described by West-East trend lines (from plot 1 to plot 120, which run in parallel for the two years). A decrease in bulk density is observed throughout the area, as the 2019 trend line is lower than that of 2015 bulk density. The year 2015 has an average bulk density of 1.36 $mg \cdot m^{-3}$ meanwhile that of 2019 is 1.23 $mg \cdot m^{-3}$. With a decrease in bulk density in 2019, generally, a decrease in bulk density means an improvement in soil health (through better aeration and water infiltration). According to Al-shammary et al. (2018), as a rule of thumb, rocks mostly have a ρ_s of 2.65 $mg \cdot m^{-3}$ used further in the carbon stocks assessment to correct for the volume of stones (Equation 4.2).

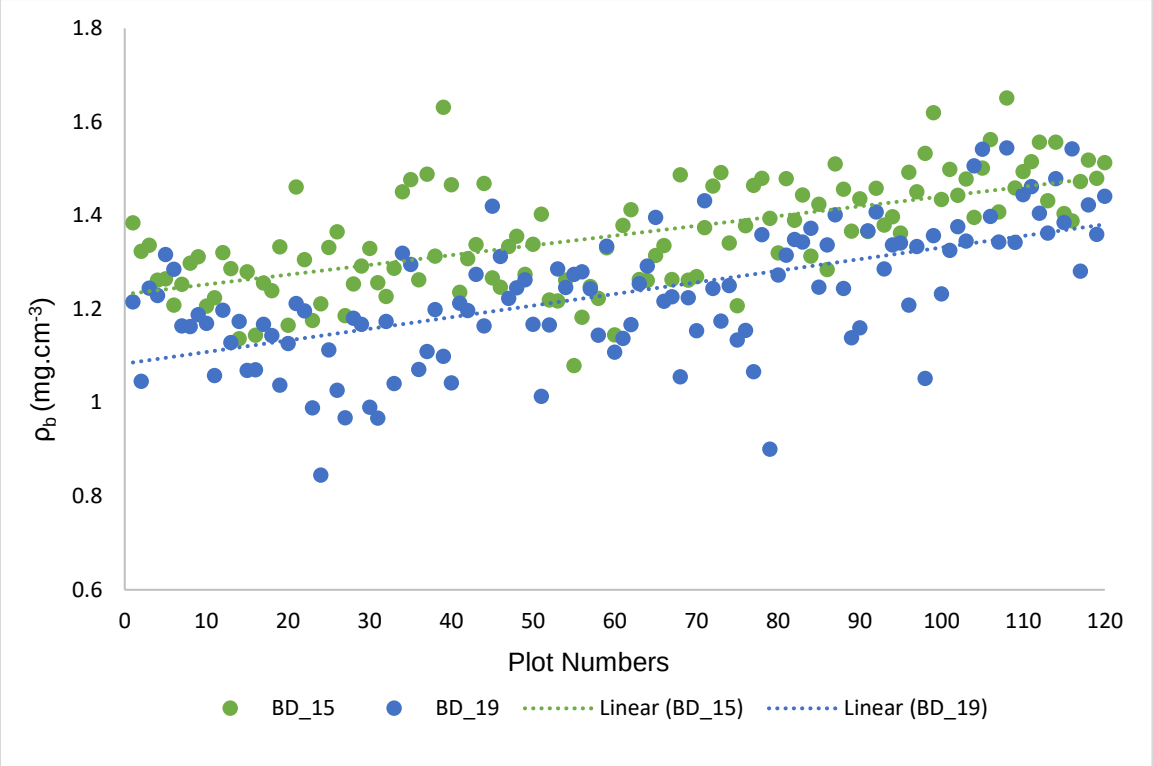


Figure 4.5: Soil bulk density for the year 2015 and 2019

Figure 4.6 represents an illustration of ρ_b changes per system. In average the ρ_b for 2019 is lower than that of 2015, with standard deviations in overall as well being smaller. Additionally, t-tests were performed to check whether the differences between the two years are significant or not, both per treatment and per plot. The t-test results for the treatments are presented in Table 4.3 and those showing count from different alphas (α=0.05 and α=0.01) per plot can be found in Appendix C. From the results, count for α=0.01 has 29 significant values out of 120 data points and count for α=0.05 has only 13 significant values. The p-values obtained for the treatments are significant with treatment 3 being highly significant followed by 2 then 1 (p-values<0.05).

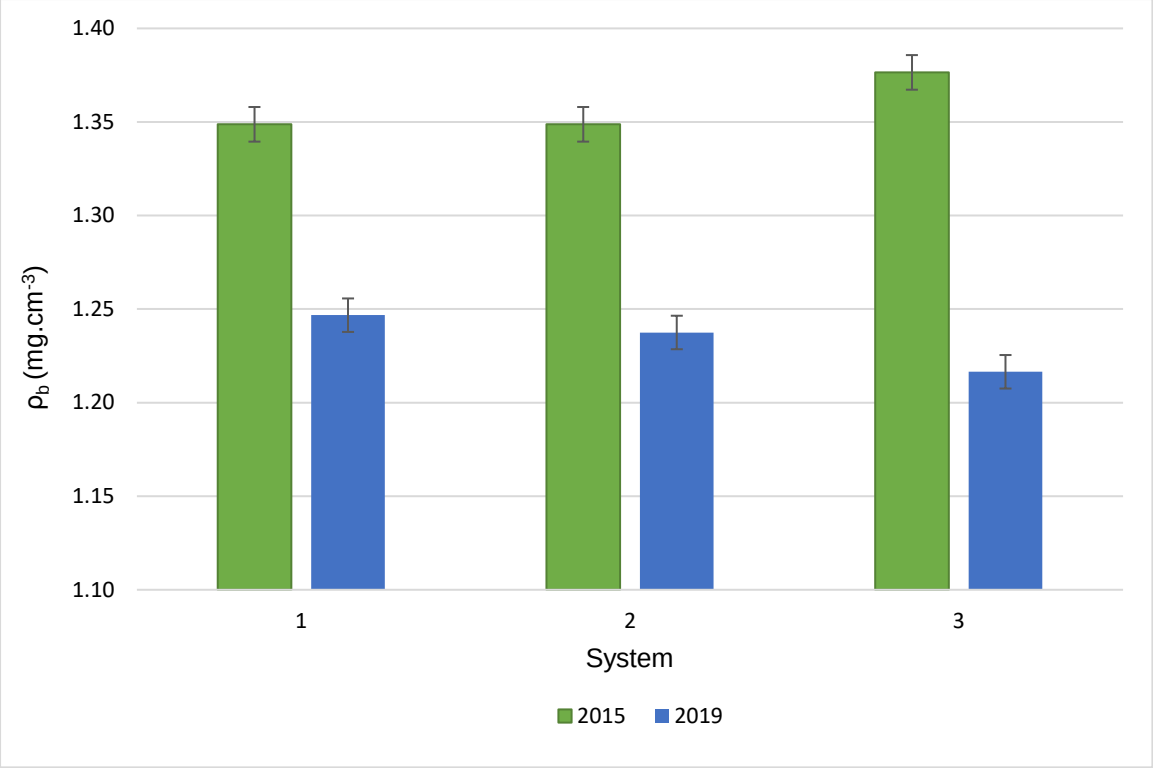


Figure 4.6: Soil bulk density for the year 2015 and 2019 averages grouped per system (treatment), with standard error bars.

Table 4.3: Bulk density descriptive statistical results including the t-test values to evaluate the significance within the applied treatments over the years.

System	Average		σ		St. Error		p-value
Year	**2015**	**2019**	**2015**	**2019**	**2015**	**2019**	
System 1	1.35	1.25	0.11	0.13	0.02	0.02	$9.7x10^{-6}$
System 2	1.35	1.24	0.11	0.13	0.02	0.02	$2.2x10^{-6}$
System 3	1.38	1.22	0.13	0.16	0.02	0.02	$1.4x10^{-7}$

The following Table (4.4) shows descriptive results of the crop sequences over the period of four years. Half (15/30) of the observed changes in crop sequences in terms of bulk density are also significant (highlighted in bold), and half of them are not significant as shown in Table 4.4 below.

Table 4.4: Descriptive results of the crop sequences over a period of four years for bulk density

Crop Sequence				$\Delta\rho_b$	σ	P-value
1. Wheat	Legume cover	Wheat	Barley	-0.17	0.11	**0.049**
2. Legume cover	Wheat	Barley	Canola	-0.15	0.07	**0.021**
3. Wheat	Canola	Wheat	Canola	-0.04	0.17	0.707
4. Canola	Grass cover	Wheat	Canola	-0.1	0.11	0.155
5. Legume cover	Wheat	Wheat	Canola	-0.11	0.08	**0.064**
6. Legume Cover	Wheat	Wheat	Canola	-0.02	0.12	0.842
7. Barley	Canola	Grass cover	Faba beans	-0.34	0.17	**0.027**
8. Vetch	Wheat	Wheat	Faba beans	-0.11	0.09	**0.089**
9. Wheat	Barley	Canola	Grass cover	-0.13	0.12	0.115
10. Wheat	Wheat	Canola	Grass cover	-0.12	0.13	0.166
11. Wheat	Wheat	Canola	Grass cover	-0.19	0.09	**0.025**
12. Wheat	Canola	Wheat	Legume cover	-0.05	0.05	0.141
13. Wheat	Faba beans	Wheat	Legume cover	-0.1	0.14	0.232
14. Oats	Linseed	Wheat	Legume cover	-0.08	0.18	0.437
15. Faba beans	Wheat	Oats	Linseed	-0.24	0.19	**0.082**
16. Grass Cover	Faba beans	Wheat	Oats	-0.17	0.04	**0.002**
17. Wheat	Canola	Grass cover	Vetch	-0.14	0.17	0.205
18. Canola	Wheat	Canola	Wheat	-0.1	0.03	**0.009**
19. Grass cover	Wheat	Canola	Wheat	-0.15	0.1	**0.061**
20. Canola	Grass cover	Faba beans	Wheat	-0.02	0.17	0.819
21. Wheat	Wheat	Faba beans	Wheat	-0.15	0.12	**0.084**
22. Wheat	Canola	Grass cover	Wheat	-0.09	0.15	0.305
23. Canola	Wheat	Legume cover	Wheat	-0.12	0.1	**0.077**
24. Faba beans	Wheat	Legume cover	Wheat	0.03	0.07	0.447
25. Linseed	Wheat	Legume cover	Wheat	-0.18	0.28	0.298
26. Wheat	Oats	Linseed	Wheat	-0.16	0.13	**0.096**
27. Canola	Grass Cover	Vetch	Wheat	-0.25	0.2	**0.089**
28. Wheat	Legume cover	Wheat	Wheat	0.02	0.1	0.729
29. Wheat	Legume cover	Wheat	Wheat	-0.22	0.19	0.104
30. Grass cover	Vetch	Wheat	Wheat	-0.2	0.11	**0.033**

4.3.4.2 Soil organic carbon

Soil organic carbon (SOC) does not only play an important role in sustaining, improving food production and enhancing soil quality but it also alleviates the effects of greenhouse gases (Sakİn, 2012). Changes in SOC over the trial period were inspected and comparisons between the 2015 and 2019 measured carbon are represented in Figure 4.7. The results obtained show that in overall, there is a significant difference between the 2015 and 2019 measured carbon data ($p<0.05$). An average decrease of -0.16 and a quite high standard deviation of 0.69 were obtained. The differences (residuals and deltas) are available in Appendix A.

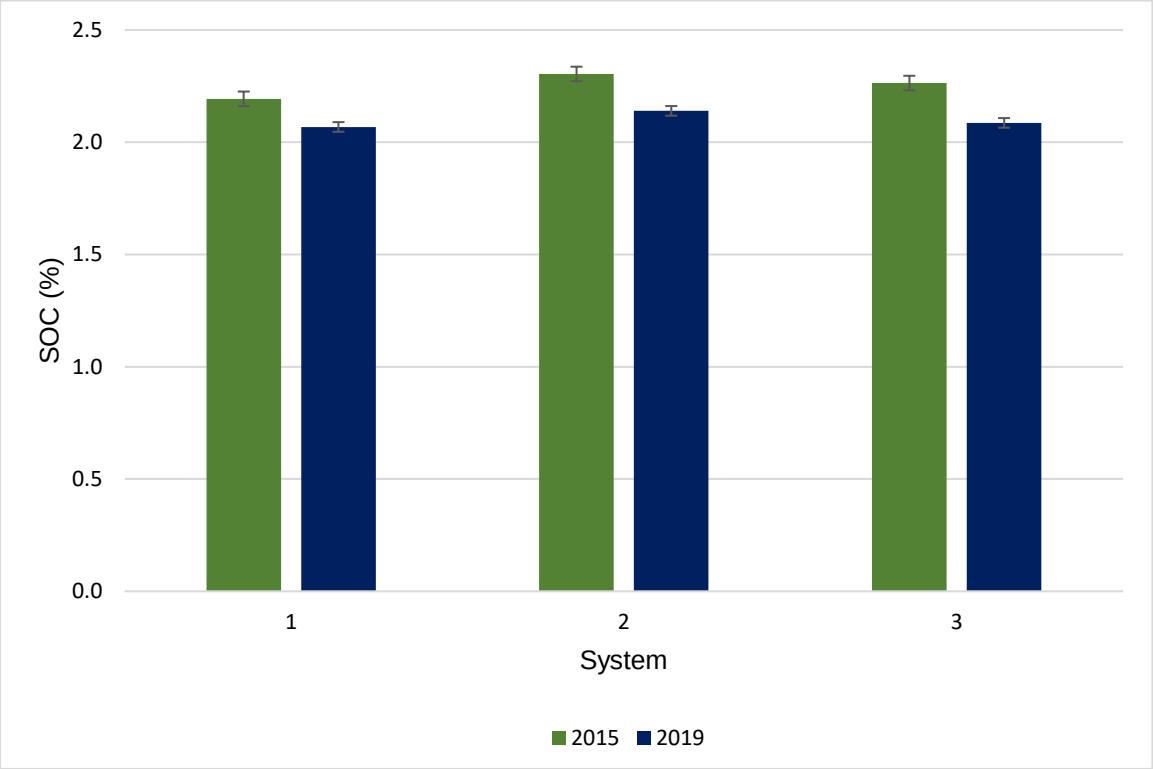


Figure 4.7: Soil organic carbon (SOC) for the year 2015 and 2019, averages grouped per system (treatment) with standard error bars.

Figure 4.8 presents a summary of the descriptive statistics for the differences per system and per replication over the years. In system one there are no significant changes ($p>0.05$) for both replicate 3 and 4. In system 2, the changes are only significant at alpha = 0.1 level of significance for 4 replicates. For system 3, the differences are significant for both replications.

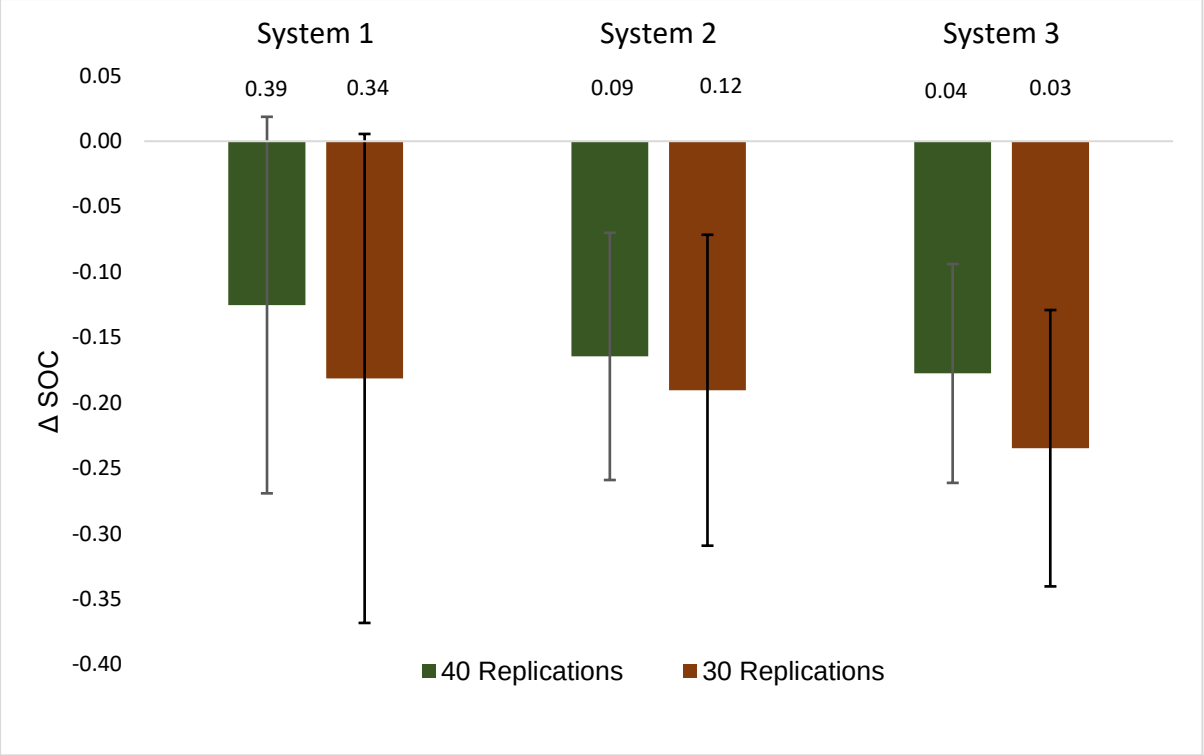


Figure 4.8: Plot showing the SOC content reduction per system and per number of replications between the two years (2015 and 2019), p-values are indicated under each replication of the system and their standard errors as error bars.

While it was expected that the part of the field allocated to three randomly distributed replications (excluding section 19B hosting replication 4) will produce results that are more reliable, this did not happen in case of SOC observations and a statistically significant decline was observed only in system 3, using either 3 or 4 replications of the crop sequences. Table 4.5 shows descriptive results of the crop sequences over four years. Across the observations per rotation, a decrease in soil carbon is observed, except for crop sequence 11, 19, 29 and 30 where an increase is observed. However, there is only one crop sequence (number 18) where by the changes are significant with a p-value of 0.018.

Table 4.5: Descriptive results of the crop sequences over a period of four years for soil organic carbon

Crop Sequence					ΔC_{org}	σ	p-value
1.	Wheat	Legume cover	Wheat	Barley	-0.64	0.93	0.264
2.	Legume cover	Wheat	Barley	Canola	-0.09	0.2	0.455
3.	Wheat	Canola	Wheat	Canola	-0.46	1.52	0.589
4.	Canola	Grass cover	Wheat	Canola	-0.09	2.01	0.937
5.	Legume cover	Wheat	Wheat	Canola	-0.44	0.97	0.438
6.	Legume cover	Wheat	Wheat	Canola	-0.17	0.48	0.524
7.	Barley	Canola	Grass cover	Faba beans	0.21	0.27	0.217
8.	Vetch	Wheat	Wheat	Faba beans	-0.03	0.54	0.945
9.	Wheat	Barley	Canola	Grass cover	-0.31	0.5	0.302
10.	Wheat	Wheat	Canola	Grass cover	-0.29	0.84	0.549
11.	Wheat	Wheat	Canola	Grass cover	0.2	0.57	0.544
12.	Wheat	Canola	Wheat	Legume cover	-0.11	0.29	0.507
13.	Wheat	Faba beans	Wheat	Legume cover	-0.18	0.42	0.46
14.	Oats	Linseed	Wheat	Legume cover	-0.14	0.61	0.686
15.	Faba beans	Wheat	Oats	Linseed	-0.2	0.45	0.437
16.	Grass cover	Faba beans	Wheat	Oats	-0.18	0.53	0.551
17.	Wheat	Canola	Grass cover	Vetch	-0.26	0.58	0.434
18.	Canola	Wheat	Canola	Wheat	-0.63	0.27	**0.018**
19.	Grass cover	Wheat	Canola	Wheat	0.28	0.32	0.173
20.	Canola	Grass cover	Faba beans	Wheat	-0.42	0.43	0.146
21.	Wheat	Wheat	Faba beans	Wheat	-0.35	0.9	0.5
22.	Wheat	Canola	Grass cover	Wheat	-0.15	0.9	0.769
23.	Canola	Wheat	Legume cover	Wheat	-0.15	0.97	0.778
24.	Faba beans	Wheat	Legume cover	Wheat	-0.17	0.44	0.505
25.	Linseed	Wheat	Legume cover	Wheat	-0.15	0.72	0.706
26.	Wheat	Oats	Linseed	Wheat	0.12	0.47	0.665
27.	Canola	Grass cover	Vetch	Wheat	0.14	0.76	0.735
28.	Wheat	Legume cover	Wheat	Wheat	-0.55	0.47	0.1
29.	Wheat	Legume cover	Wheat	Wheat	0.37	0.61	0.307
30.	Grass Cover	Vetch	Wheat	Wheat	0.08	0.45	0.746

4.3.4.3 Soil carbon stocks

Carbon stocks per sample were calculated for both 2015 and 2019 data using the formula extracted from Wiese et al. (2016):

$C_{stock} = Cv \cdot \Delta z$ Equation 4.1

where, Cv is the volumetric SOC content and Δz is the depth increment (0.05 m).

C_V accounting for the volume of stones calculated as:

$Cv = 10.SOC \cdot \rho_b \cdot (1 - S_m \cdot \rho_b / \rho_s)$ Equation 4.2

where, 10 is a unit conversion factor, $Cv_{(2mm)}$ is the volumetric carbon [$mg \cdot m^{-3}$] content in the <2mm fraction, SOC is soil organic carbon content [%wt], ρb is soil bulk density [$mg \cdot cm^{-3}$], S_m is the mass fraction of stones of the bulk sample determined gravimetrically, and ρ_s is the particle density in this case estimated as 2.65 $mg \cdot cm^{-3}$. The results of the calculated carbon stocks per system are shown in Figure 4.9 below.

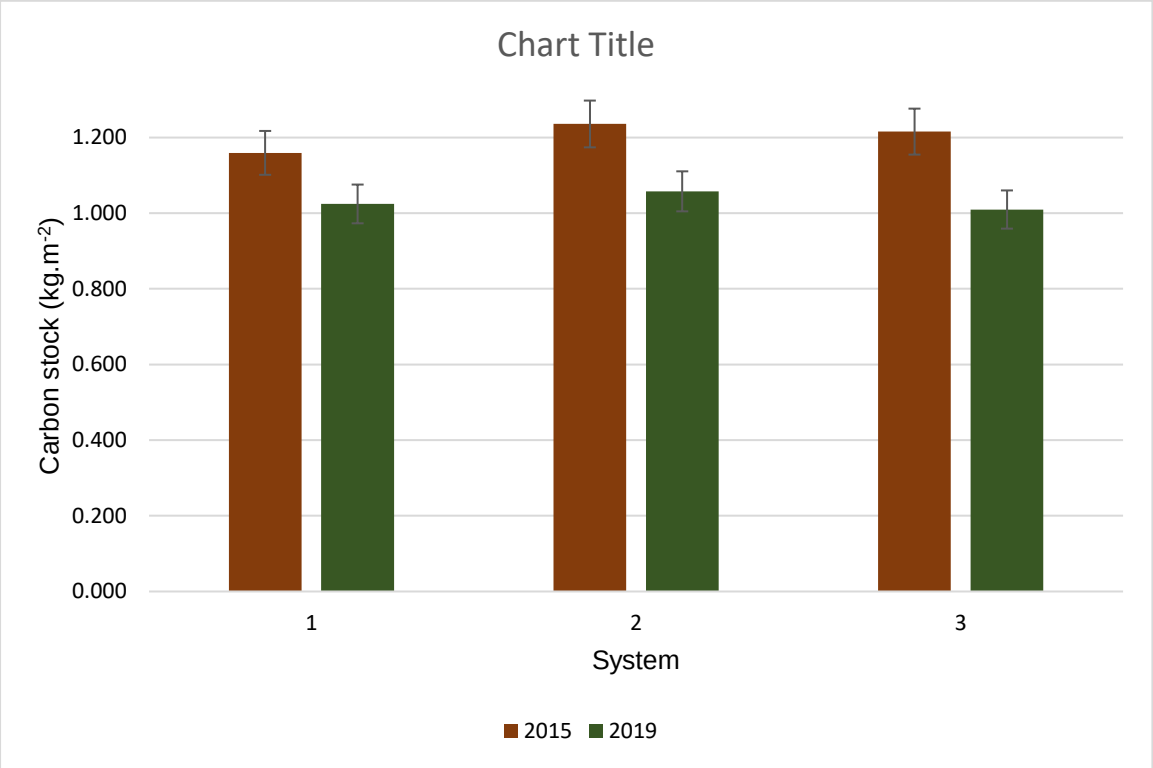


Figure 4.9: Averaged carbon stocks for the years 2015 and 2019 grouped per system (treatment), with standard error bars.

However, many authors (Wendt & Hauser, 2013) pointed out the inadequacy of comparing changes in soil carbon stocks within fixed depth intervals following the work on Equivalent Soil Mass (Ellertl & Bettany, 1995). Here a a simple depth adjustment is implemented for the calculation of carbon stocks (Equation 4.3), which accounts for compaction / de-compaction within the layer of interest, assuming that the change in volume is unidirectional. In case of de-compaction, is fully accounted by up swelling of the soil material. While criticized for deep soil carbon change assessments (Lee et al., 2009), this method may be sufficiently accurate to represent the effect of bulk density change near the soil surface.

The depth was then adjusted using the following equation:

$$Z_e = Z \cdot \frac{\rho_i}{\rho_f} \qquad \text{Equation 4.3}$$

From Table 4.6 it can be observed that all the differences in carbon stocks per system are highly significant (p-values<0.05). A decrease in soil organic carbon stocks (P<0.05) was observed within the 5 cm fixed depth, while no change in SOC stocks was observed on equivalent soil mass basis.

Table 4.6: Summary descriptive statistics for the carbon stocks results

	System	Average		ΔC_{stock}	σ		St.Error		p-value
	Year	2019	2015		2019	2015	2019	2015	
Fixed depth (5 cm)	System 1	1.025	1.160	-0.135	0.247	0.302	0.04	0.05	0.04
	System 2	1.058	1.236	-0.178	0.223	0.250	0.04	0.04	2×10^{-4}
	System 3	1.010	1.216	-0.206	0.199	0.259	0.03	0.04	5×10^{-6}
Equivalent soil mass									
	System 1	1.125	1.160	-0.034	0.339	0.302	0.054	0.048	0.663
	System 2	1.168	1.236	-0.068	0.293	0.250	0.046	0.040	0.214
	System 3	1.159	1.216	-0.057	0.281	0.259	0.044	0.041	0.250

Table 4.7 shows descriptive results of the crop sequences over the four years period. Across the observations per rotation, a decrease in carbon stocks is observed, except for crop sequence 11, and 29 where an increase is observed. On the p-values column, significant observations are made bold and about five crop sequences are significant (p<0.05), while the rest are not (p>0.05).

Table 4.7: Statistical results of the crop sequences over a period of four years for soil carbon stocks

Crop sequence					Fixed depth (5 cm)			Equivalent soil mass		
					ΔC_{stock}	σ	p-value	ΔC_{stock}	σ	p-value
1.	Wheat	Legume cover	Wheat	Barley	-0.37	0.34	0.124	0.02	0.33	0.904
2.	Legume cover	Wheat	Barley	Canola	-0.16	0.09	**0.042**	0.13	0.24	0.363
3.	Wheat	Canola	Wheat	Canola	-0.23	0.58	0.488	-0.23	0.58	0.487
4.	Canola	Grass cover	Wheat	Canola	-0.11	0.91	0.838	-0.10	0.91	0.838
5.	Legume cover	Wheat	Wheat	Canola	-0.35	0.44	0.212	-0.35	0.44	0.212
6.	Legume cover	Wheat	Wheat	Canola	-0.15	0.23	0.291	-0.14	0.23	0.290
7.	Barley	Canola	Grass cover	Faba beans	-0.09	0.18	0.389	-0.09	0.17	0.388
8.	Vetch	Wheat	Wheat	Faba beans	-0.09	0.25	0.526	-0.09	0.24	0.526
9.	Wheat	Barley	Canola	Grass cover	-0.22	0.33	0.279	-0.21	0.33	0.279
10.	Wheat	Wheat	Canola	Grass cover	-0.2	0.43	0.421	-0.20	0.42	0.421
11.	Wheat	Wheat	Canola	Grass cover	0.09	0.23	0.491	0.09	0.23	0.490
12.	Wheat	Canola	Wheat	Legume cover	-0.13	0.23	0.343	-0.13	0.22	0.342
13.	Wheat	Faba beans	Wheat	Legume cover	-0.24	0.13	**0.034**	-0.23	0.13	**0.033**
14.	Oats	Linseed	Wheat	Legume cover	-0.14	0.39	0.53	-0.14	0.39	0.530
15.	Faba beans	Wheat	Oats	Linseed	-0.29	0.18	**0.045**	-0.29	0.17	**0.045**
16.	Grass Cover	Faba beans	Wheat	Oats	-0.25	0.32	0.211	-0.25	0.31	0.210
17.	Wheat	Canola	Grass Cover	Vetch	-0.22	0.31	0.259	-0.22	0.31	0.258
18.	Canola	Wheat	Canola	Wheat	-0.4	0.16	**0.015**	-0.40	0.15	**0.014**
19.	Grass cover	Wheat	Canola	Wheat	0.05	0.15	0.54	0.05	0.14	0.540
20.	Canola	Grass cover	Faba beans	Wheat	-0.28	0.21	0.075	-0.28	0.21	0.074
21.	Wheat	Wheat	Faba beans	Wheat	-0.36	0.38	0.156	-0.35	0.37	0.156
22.	Wheat	Canola	Grass cover	Wheat	-0.12	0.38	0.592	-0.11	0.38	0.591
23.	Canola	Wheat	Legume cover	Wheat	-0.15	0.56	0.639	-0.14	0.55	0.638
24.	Faba beans	Wheat	Legume cover	Wheat	-0.12	0.25	0.418	-0.12	0.25	0.418
25.	Linseed	Wheat	Legume cover	Wheat	-0.18	0.3	0.306	-0.18	0.29	0.306
26.	Wheat	Oats	Linseed	Wheat	-0.12	0.18	0.277	-0.12	0.17	0.276
27.	Canola	Grass cover	Vetch	Wheat	-0.14	0.25	0.368	-0.13	0.25	0.367
28.	Wheat	Legume cover	Wheat	Wheat	-0.33	0.18	**0.034**	-0.33	0.18	**0.034**
29.	Wheat	Legume cover	Wheat	Wheat	0.04	0.14	0.664	0.03	0.13	0.664
30.	Grass cover	Vetch	Wheat	Wheat	-0.05	0.2	0.665	-0.05	0.19	0.665

4.4 Validation of true predictions

Amongst the calibration models that were developed in section 3.4.4, one of the models was developed to check the ability of NIR to predict soil organic carbon, the results of the developed model showed that the NIR spectral signatures have the ability to fairly predict SOC; with an R^2 = 0.64, RPD = 1.67 and RMSECV = 0.33. New carbon values were then predicted using this model and the 2019 soil spectral signatures in OPUS Qunt2 software. Results obtained from the predictions show that in overall there is a significant difference (p=0.0045) between the predicted and measured carbon values. An overall decrease of -0.21 (average of the difference between measured and predicted values) was also obtained and this decrease is accompanied by a relatively high standard deviation (0.68).

Table 4.8 shows the summary of the predictions per system between the predicted carbon using NIR and reference measured values using the laboratory methods as suggested by Schumacher (2002). In system one there are no significant changes (p>0.05) for both replicates either 3 (in more uniform area) or 4 (in more variable area) times as well as in system 3. It is only in system 2, where there are significant changes observed using all 4 replications (p=0.04). All of these changes are a decline with high standard deviations. These results show that the differences between the predicted and the measured values are relatively small, because significant differences are only observed in one system (2), although in overall the differences are significant.

Table 4.8: Summary descriptive statistics for the assessment of SOC content change - ΔC (NIR) - predicted using NIR

	System1			System2			System3		
ΔC (NIR)	Average	σ	P	Average	σ	P	Average	σ	p
4 Reps	-0.19	0.76	0.12	-0.17	0.51	0.04	-0.08	0.48	0.31
3 Reps	-0.22	0.82	0.17	-0.16	0.55	0.13	-0.10	0.53	0.34

Furthermore, the results were plotted and the regression is shown in Figure 4.10 below. With the statistical results on the figure; R^2=0.3182, and RMSE = 0.55, it can be seen that the model can predict new values, but it is not very accurate as the coefficient of determination obtained is not very strong with a high standard error as well. The changes in SOC seem not to be easily predicted using this model as only about 30% of the variation in carbon between the years is explained by the model. While the model quite accurately predicts the mean for all the 120 samples (C_{org} = 2.10%) with standard deviation (σ = 0.46) as opposed to σ = 0.54 for Walkley and Black analysis 9reference laboratory measured), the paired t-test produces

unsatisfactory results (p=0.015), showing statistically significant differences between the observed and predicted values.

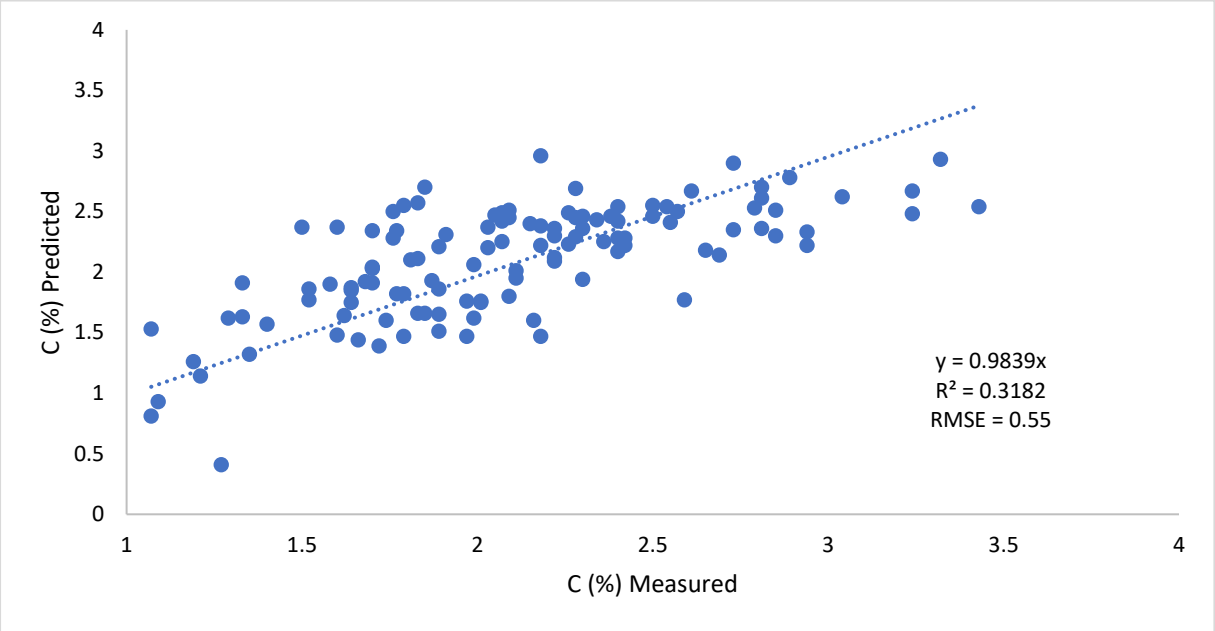


Figure 4.10: Relationship between measured and predicted soil organic carbon values.

The following Figure 4.11 was constructed to inspect whether it can be possible to pick up the differences between 2015 and 2019 from both the residues of the Walkley-Black method and those of the NIR spectral signatures. To a certain degree, the changes can be observed in both ways, with R^2 of 0.52, which means at least 50% of the variations is explained. Therefore, both NIR and the laboratory measurements give quite similar predictions; they both show a decrease in organic carbon all together after these four years.

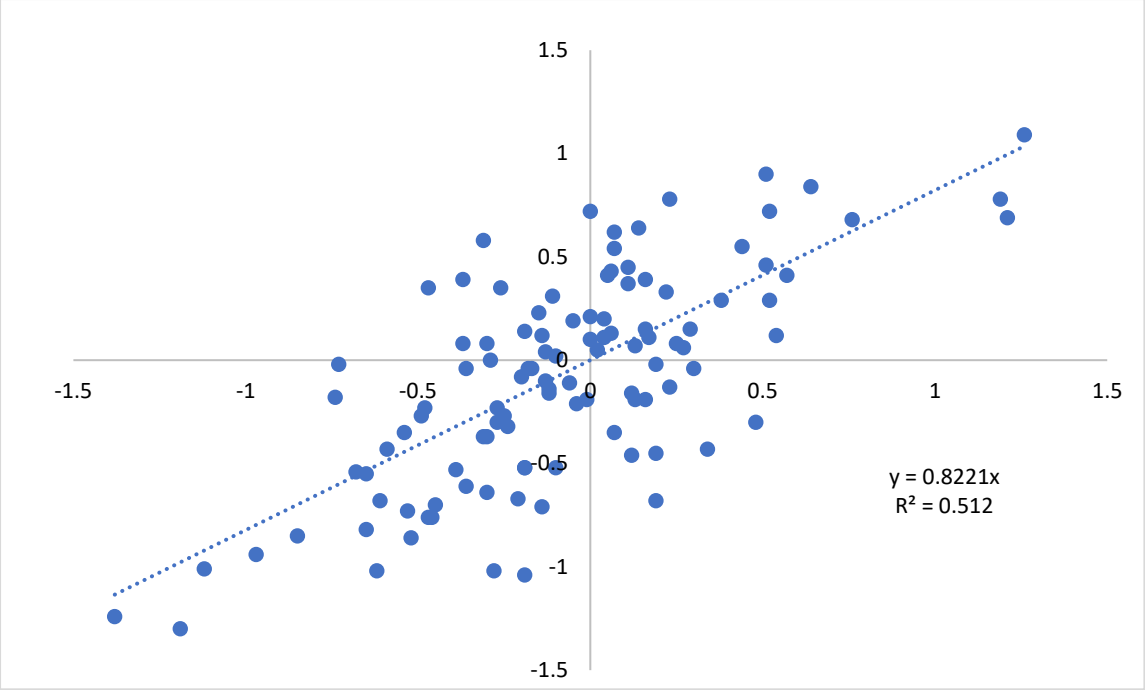


Figure 4.11: Residues of the Walkley-Black method and those of the NIR spectral signatures from 2015 and 2019.

4.5 Discussion

4.5.1 The observed changes in NIR absorbance

An upward shift in the measured spectra was observed and the changes were significant ($p<0.05$). The observed increase in absorbance from the year 2015 to the year 2019 may be an indication that the soils are becoming darker as it was stated by Zhou et al. (2014) that dark soils tend to have higher absorbance in the NIR region. Additionally, the major factors that affect soil absorbance and reflectance include moisture content, soil texture (proportion of sand, silt and clay), bulk density, presence of iron oxide, and organic matter content (Cheng-Wen Chang et al., 2001; Madari et al., 2006; Bushong et al., 2015). In Figure 4.3 where the variations between the two years were assessed per wave band, the influence of water content was observed. Around the wavenumbers of 4000-5000cm^{-1} (1400–1900 nm) there appears to be a presence of some fluctuation, which might be as a results of water having a strong influence on NIR spectra of soils as this is a dominant water absorption region (Weidong et al., 2002; Viscarra Rossel et al., 2009).

Moreover, according to Nocita et all. (2015), NIR spectroscopy also provides information about the relative proportions of bonds such as C-H, N-H, S-H and O-H, present in the organic compounds with the O-H also included in the water molecule. Even though the samples were

oven dried prior to scanning there will still be water adsorbed on the surface areas of clay minerals and organic matter, and that is why changes in the water NIR dominant regions can be detected (Vasqueset al., 2014). When the spectral absorbance results were grouped according to treatments, treatment 2 and 3 appeared to have the highest absorbance with treatment 1 having the lowest. This is attributed to the different levels of diversity as treatment 1 had low diversity, meanwhile 2 and 3 contain more and most diversity respectively. These results are comparable and in agreement with those of Sadegh et al. (2015) who obtained high NIR absorbance in soils with permanent ground cover and low absorbance in bare soils.

4.5.2 Variations in the observed soil properties and the impact of the crop sequences

Bulk density (ρb) is one of the parameters that were measured both in 2015 and in 2019. A decrease in ρb for the year 2019 was observed, and the results were significant across the treatments (systems) with treatment 3 ($p=1.4x10^{-7}$) being highly significant followed by 2 ($p=2.2x10^{-6}$) and then 1 ($p=9.7x10^{-6}$). Bulk density affects plant growth, rooting depth, infiltration, plant nutrient availability, available water capacity, soil porosity, and soil microbial activities, which influence key soil processes and productivity (FAO, 2018). Generally, a decrease in bulk density means an improvement in soil aeration (Minasny et al., 1999). According to Al-shammary et al. (2018), as a rule of thumb, a silt loam soil has about 50% pore space and a bulk density of about 1.33 $mg \cdot m^{-3}$. Our bulk density for the year 2019 has an average density of 1.255 $mg \cdot m^{-3}$ ≈1.3 $mg \cdot m^{-3}$ that may serve as an indicator that the soils are porous and well-aggregated. This might also explain the upward shift (increase) in NIR spectral absorbance because according to Soriano-Disla et al. (2014), silt loam soils with good water holding capacity tend to absorb more radiation in the NIR region (800-2800 nm).

Bulk density analysis were also done per crop sequences as illustrated in Table 4.4. Across the observations per rotations, a decrease in soil bulk density was observed, except for crop sequence number 24 where an increase is observed, though it is not significant. About 50% of the crop sequences are found to be significantly different. The decrease in soil bulk density may be an indication that diversity in soil cover has a huge impact in determining soil compaction. High diversity seem to improve bulk density as system 3 (more diversity) is highly significant. These findings are similar to those of (Pattanayak, 2019) who found crop diversification to improve a lot of chemical and biophysical properties including soil aggregate stability, porosity, root penetration, nutrient availability, water holding capacity, and infiltration rate.

Changes in soil organic carbon were measured, and on average a significant decrease was observed in SOC between the year 2015 and 2019 ($p=0.015$). Analysis per system showed that in system one there are no significant changes ($p>0.05$) for both replicate 3 and 4. In system 2, the changes were only significant at 0.1 level of significance for 4 replicates, and for system 3 the differences are significant for both replications. Changes were also inspected in terms of crop sequence (Table 4.5) results obtained showed that the decline observed was only significant in one crop sequence. The rest of the changes were not significant.

A similar trend was also observed in carbon stocks. The results obtained showed a significant overall decrease across the 120 plots ($p=1.6x10^{-7}$) within the 5 cm soil depth. No decrease in stocks, occurred ($p>0.05$) when accounting on equivalent mass basis, while factoring in the soil volume change due to de-compaction. The actual size of SOC stocks depends on the rate of C gain or loss over time as well as on bulk density (FAO, 2018). Conant et al. (2011) further stated that the major problem when measuring SOC stocks is how to deal with the spatial variability in SOC content and in chemical properties associated with different factors such as soil and vegetation types, climate, land use and management. Additionally, Pattanayak (2019) conducted a study on evaluating the effect of climate in soil organic carbon, and concluded that global soil organic carbon stocks have decreased due to changes in climate and land cove. Pattanayak (2019) also stated that soils in warmer climates tend to contain less organic matter than those in cooler climates because organic matter tend to decay more quickly at higher temperatures. This might also have an impact on the observed SCO decrease.

The “darkening” of samples in the NIR band contradicts the observed loss in organic matter. This may be to some extent explained by the rearrangement of organic matter in the topsoil towards the surface of micro-aggregates. The loss of organic matter accompanies the observed soil de-compaction associated with better aeration and oxidation of organic material. At the same time micro-aggregation (not measured) is quite likely to have improved with development of organic coatings on micro-aggregates resulting in overall increase of absorbance (Erktan et al., 2016; Waruru et al., 2016).

4.6 Conclusions

The study was conducted to explore the changes in spectral absorbance over four years and to explore whether those changes are reflected in soil properties and wheat yield. On average, the measured spectral signatures for the two years (2015 and 2019) showed an increase in absorbance. In terms of absorbance per system (treatment), system two and three had higher absorbance, with treatment one having the lowest. This may be attributed to the impact of the different levels of diversity as treatment 1 had low diversity, meanwhile 2 and 3 contained more and most diversity respectively.

A slight decrease in organic carbon was observed as well and a significant decrease in carbon stocks ($p<0.05$) in 5 cm depth. No decrease in stocks, however occurred ($p>0.05$) when accounting on equivalent mass basis, while factoring in the soil volume change due to de-compaction.

The variations in the field observed in the near infrared spectral signatures were also observed in some soil properties. This indicates that the use of the near infrared spectroscopy has a great potential in showing the changes that are occurring in the field.

With soil carbon being of particular interest to this trial, it should be noted that the true validation by soil resampling and analysis of the same field blocks was disappointing. The best model produced by cross-validation resulted in unsatisfactory predictions in true validation. The validation of true predicted values against laboratory analysis showed a slope of the correlation (0.98) while close to 1:1 line with an R^2 of 0.3 and an RMSE of 0.55% of carbon content.

CHAPTER 5: Conclusions and recommendations

5.1 Conclusions

Spatial variations within short distances in the soil continue to be a limiting factor in obtaining reliable and accurate results in experimental trials. There is an increasing demand for techniques and methods that will help understand these variations better. Ways of doing soil analysis (techniques) are not only expensive but also involve tedious and labor-intensive methods to be followed. Rapid and cost-effective tools to study the spatial variations in soil properties and crop yields for large areas are required. Soil spectroscopy appears to be a fast, nondestructive, cost-effective, environmental-friendly, reproducible, and repeatable analytical technique. Visible and near-infrared reflectance spectroscopy have been widely used to estimate various soil attributes for soil surveys, land use planning, and soil management purposes.

This study aimed at evaluating whether the near-infrared (NIR) absorbance detection using laboratory analysis of samples collected in the field may be a good estimate of soil properties and wheat yield. It also evaluated whether does the temporal-spatial variation in NIR absorbance reflects the variation in some selected common soil properties and wheat yield. The following conclusions were reached from this study:

- Soil spectral characteristics in the NIR ranges (750-2500 nm) may be a good indicator of not only soil properties but also plant responses to the changes in soil properties across the field. The results obtained on the partial least squares regression models that were developed to relate the soil properties and wheat yield to NIR spectral signatures showed that in most cases 80% to 60% of the variation in the predicted variable is accounted for by the regression model or by the regressors.
- Partial least squares regression method was found to be a sufficiently good chemometric method used to relate the spectral characteristic of material with its other properties.
- The selection of specific spectral attributes through pre-processing (de-noising and spectra transformation) of the raw spectra can identify the parts of the spectrum and derivatives with highest co-variance between the transformed spectral characteristics and properties measured using a standard analytical method.

Observations on soil change after four years of crop rotation with different levels of crop diversity have shown that near-surface soil spectral characteristics and some key soil parameters have changed significantly.

- The change in soil spectral characteristics was significant across the full range of the spectrum, with the most prominent changes occurring in systems with higher level of diversity.
- Soil de-compaction was observed throughout the trial area with largest reduction in bulk density occurring in system 3 characterized by the highest level of crop diversity.
- A minor decrease in total SOC content was observed throughout the trial area.
- A statistically significant decrease in soil carbon stocks was reported within the 5 cm fixed depth interval, which on the contrary has shown no statistically significant change in soil carbon stocks on equivalent mass basis.
- With soil carbon being of particular interest to this trial, its results for true validation by soil resampling and analysis of the same field blocks were a bit disappointing as only about 30% of the variations on the predicted values could be explained by the measured values with RMSE of 0.55%.
- Due to the fact that soil spectral characteristics have changed without significant changes in soil carbon content or soil carbon stocks on equivalent mass basis, it has been assume that such changes in spectral characteristics may have occurred due to other parameters (e.g. increased micro-aggregation) not measured in this study.

5.2 Recommendations

This research has presented the importance of finding ways to better understand the spatial variations in the structure of the field, both in terms of soil properties and wheat yield. The developed PLSR chemometric models can be used to predict soil properties and yield. It is therefore suggested that further studies be aimed at understanding soil variability using the NIR spectroscopy and using the invasive proximal soil sensing spectroscopy tools. So as to explore the differences between analyzing sampled soils and scanning the soils directly in situ. To avoid dense core sampling and the procedures (drying, grinding, sieving) that have to be followed before scanning the samples as these procedures are also prone to introduce errors to a certain extent. Further studies can also be aimed at exploring how proximal colour and NIR sensing is related to remotely sensed imagery of the bare soil. This can be achieved by using a remotely sensed data captured using drone flights at RGB and NIR, and then be integrated into the NIR spectra to improve the quality of comparable data.

References

Adeline, K.R.M., Gomez, C., Gorretta, N. & Roger, J.M. 2017. Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data. *Geoderma*. 288:143–153.

Andales, A. A., Green, T. R., Ahuja, L. R., Erskine, R. H., Peterson, G. A. 2007. Temporally stable patterns in grain yield and soil water on a dryland catena. Agricultural Systems, 94, pp.119-127.

Al-shammary, A.A.G., Kouzani, A.Z., Kaynak, A., Khoo, S.Y., Norton, M. & Gates, W. 2018. Soil Bulk Density Estimation Methods : A Review Soil Bulk Density Estimation Methods. *Pedosphere: An International Journal*. 28(4):581–596.

Asekova, S., Han, S., Choi, H., Park, S. & Shin, D. 2016. Determination of forage quality by near-infrared reflectance spectroscopy in soybean Determination of forage quality by near-infrared reflectance spectroscopy in soybean. (January).

Awiti, A.O., Walsh, M.G., Shepherd, K.D. & Kinyamario, J. 2008. Soil condition classification using infrared spectroscopy : A proposition for assessment of soil condition along a tropical forest-cropland chronosequence. 143:73–84.

Barrett, L.R., 2002. Spectrophotometric color measurement in situ in well drained sandy soils. Geoderma, 108, 49–77.

Bayer, A., Bachmann, M., Andreas, M. & Kaufmann, H. 2012. A Comparison of Feature-Based MLR and PLS Regression Techniques for the Prediction of Three Soil Constituents in a Degraded South African Ecosystem. 2012(1).

Bilgili, A.V., Akbas, F. & Es, H.M. Van. 2011. Combined use of hyperspectral VNIR reflectance spectroscopy and kriging to predict soil variables spatially Combined use of hyperspectral VNIR reflectance spectroscopy and kriging to predict soil variables. (June).

Bouma, J. 2014. Review article Soil science contributions towards Sustainable Development Goals and their implementation : linking soil functions with ecosystem services §. 111–120.

Breysse, D., Borderie, C. La, Elachachi, S.M., Niandou, H., Borderie, C. La, Elachachi, S.M. & Spatial, H.N. 2007. Civil Engineering and Environmental Systems Spatial variations in soil properties and their influence on structural reliability. 6608.

Brown, D.J. 2007. Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. *Geoderma*. 140(4):444–453.

Brown, D.J., Bricklemyer, R.S. & Miller, P.R. 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. *Geoderma*. 129(3–4):251–267.

Bushong, J.T., Norman, R.J. & Slaton, N.A. 2015. Near-Infrared Reflectance Spectroscopy as a Method for Determining Organic Carbon Concentrations in Soil. *Communications in Soil Science and Plant Analysis*. 46(14):1791–1801.

Cambou, A., Cardinael, R., Kouakoua, E., Villeneuve, M., Durand, C. & Barthès, B.G. 2016. Geoderma Prediction of soil organic carbon stock using visible and near infrared re fl ectance spectroscopy (VNIRS) in the field. *Geoderma*. 261:151–159.

CIE, 1971. Colorimetry: official recommendations of the International Commission on Illumination, Paris.

CIE, 1978. Recommendations on uniform color spaces, color difference equations and psychometric color terms, Paris.

Cheng-Wen Chang, David A. Laird, Maurice J. Mausbach, and Charles R. Hurburgh, J. 2001. Near-Infrared reflectance Spectroscopy-Principal Component Regression Analyses of Soil Properties. 480–490.

DAFF. 2017. Trends in the Agricultural Sector.

Delgado, R. & Huertas, R. 2014. Four soil color charts compared in CIELAB color space Colour variation in standard soil-colour charts. (May 2014).

Du, C. & Zhou, Æ.J. 2009. Evaluation of soil fertility using infrared spectroscopy : a review. 97–113.

Edwards, S.J., 1975. The science of colour. Physics Education, June 1975, 316–321.

Ellertl, B.H. & Bettany, J.R. 1995. Calculation of organic matter and nutrients stored in soils under contrasting management regimes.

Erktan, A., Legout, C., Danieli, S. De, Daumergue, N. & Cécillon, L. 2016. Geoderma Comparison of infrared spectroscopy and laser granulometry as alternative methods to estimate soil aggregate stability in Mediterranean badlands. *Geoderma*. 271:225–233.

Fan, Z., Herrick, J.E., Saltzman, R. & Zee, J. Van. 2017. Measurement of Soil Color: A Comparison Between Smartphone Camera and the Munsell Color Charts. (May):1139–1146.

FAO. 2015. Soils Challenge Badge.

FAO. 2018. Measuring and modelling soil carbon stocks and stock changes in livestock production systems.

Fernandez-ahumada, E., Palagos, B., Roger, J. & Mcbratney, A. 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. Trends in Analytical Chemistry. 29(9):1073–1081.

Ferrio, Juan Pedro, E. Bertran, M.M. Nachit, J. Catal, J.L.A. 2004. Estimation of grain yield by near-infrared reflectance spectroscopy in durum wheat Estimation of grain yield by near-infrared reflectance spectroscopy in durum wheat. (September 2015).

Ferrio, J.P., Araus, J.L., Bertran, E., Nachit, M. & Royo, C. 2000. Near infrared reflectance

spectroscopy as a new screening tool to increase durum wheat yield. 144:141–144.

Feyziyev, F., Babayev, M., Priori, S. & Abate, G.L. 2016. Using Visible-Near Infrared Spectroscopy to Predict Soil Properties of Mugan. (March):52–58.

Frogbrook, Z.L., Oliver, M.A., Salahi, M. & Ellis, R.H. 2002. Exploring the spatial relations between cereal yield and soil chemical properties and the implications for sampling.

Garten, C.T. & Hanson, P.J. 2006. Measured forest soil C stocks and estimated turnover times along an elevation gradient. *Geoderma*. 136(1–2):342–352.

Gobrecht, A., Roger, J.M. & Bellon-Maurel, V. 2014. Major Issues of Diffuse Reflectance NIR Spectroscopy in the Specific Context of Soil Carbon Content Estimation. A Review. Vol. 123.

Garnsworthy, P.C., J. Wiseman & K. Fegeros, 2000. Prediction of chemical, nutritive and agronomic characteristics of wheat by near infrared spectroscopy. J Agric Sci 135: 409–417.

Grunwald, S., Vasques, G.M. & Rivero, R.G. 2015. *Fusion of Soil and Remote Sensing Data to Model Soil Properties*. Vol. 131. Elsevier Ltd.

Han, Y., Zhao, Y. & Wang, Y. 2014. Study on Polarized Spectral Characteristics of Soil with Different Water Content. *Journal of the Indian Society of Remote Sensing*. 42(4):727–732.

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.

Jensen, J.R. 2005. 2005. Introductory Digital Image Processing.

Ji W J, Li X, Li C X, et al. 2012. Using different data mining algorithms to predict soil organic matter based on visible-near infrared spectroscopy. Spectrosc Spect Anal, 32: 2393–2398

Keesstra, S.D., Bouma, J., Wallinga, J., Tittonell, P., Smith, P., Cerdà, A., Montanarella, L., Quinton, J.N., et al. 2016. The significance of soils and soil science towards realization of the United Nations Sustainable Development Goals. (i):111–128.

Klaas, N., Faber, M. & Bro, R. 2005. Standard error of prediction for multilinear PLS 2 . Practical implementation in fluorescence spectroscopy. 75:69–76.

Lee, J., Hopmans, J.W., Rolston, D.E., Baer, S.G. & Six, J. 2009. Agriculture , Ecosystems and Environment Determining soil carbon stock changes : Simple bulk density corrections fail. 134:251–256.

Lin, Z.D., Wang, Y.B., Wang, R.J., Wang, L.S., Lu, C.P., Zhang, Z.Y., Song, L.T. & Liu, Y. 2017. Improvement of the Vis-NIRs model in the prediction of soil organic matter content using spectral pretreatments, sample selection and wavelenght optimization. 84(3):529–534.

Luis Galvez-Sola, Francisco García-Sánchez, J.G.P.-P., Vicente Gimeno, Josefa M. Navarro , Raul Moral, J.J.M.-N. and & Nieves1, M. 2015. Rapid estimation of nutritional elements on citrus leaves by near infrared reflectance spectroscopy. 6(July):1–8.

Madari, B.E., Reeves, J.B., Machado, P.L.O.A., Guimarães, C.M., Torres, E. & McCarty, G.W. 2006. Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. *Geoderma*. 136(1–2):245–259.

Malley, D.F., 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(7–8):999–1012.

Marqués-mateu, Á., Moreno-ramón, H., Balasch, S. & Ibáñez-asensio, S. 2018. Catena Quantifying the uncertainty of soil colour measurements with Munsell charts using a modi fi ed attribute agreement analysis. *Catena*. 171(June):44–53.

Mashaba, Z.O. 2016. Modelling dryland winter wheat yield using remotely sensed imagery and agrometeorological parameters.

Mattikalli, M. 1993. Soil Color Modeling for the Visible and Near- Infrared Bands of Landsat Sensors Using Laboratory Spectral Measurements. 4257(96).

McBratney, A.B., Mendonça Santos, M.L. & Minasny, B. 2003. On digital soil mapping. Vol. 117.

Melville, M.D. & Atkinson, G., 1985. Soil colour: its measurement and its designation in models of uniform colour space. Journal of Soil Science, 36, 495–512.

Mevik, B. & Wehrens, R. 2018. Introduction to the pls Package. (Section 7).

Miller, J.N. & Miller, J.C. 2010. *S*tatistics and Chemometrics for Analytical Chemistry. Sixth edit ed.

Minasny, B., McBratney, A. & Bristow, K. 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. *Geoderma*. 93:225–253. [Online], Available: http://www.sciencedirect.com/science/article/pii/S0016706199000610.

Minolta. 2013. Spectra MagicNX Colour data software CM 600d spectrophotometer.

Moebius-Clune B.N, Moebius-Clune, D.J. Gugino B.K. , Idowu O.J., Schindelbeck R.R., Ristow A.J., van Es H.M., Thies J.E., Shayler H.A., McBride M.B., Kurtz K.S.M., Wolfe D.W., and A.G.S. 2017. Comprehensive Assessment of Soil Health.

Mohamed, E.S., Saleh, A.M., Belal, A.B. & Gad, A.A. 2018. Application of near-infrared reflectance for quantitative assessment of soil properties. *Egyptian Journal of Remote Sensing and Space Science*. 21(1):1–14.

Mortimore, J. L., Marshall, L. J. R., Almond, M. J., Hollins, P., and Matthews, W. (2004). Analysis of red and yellow ochre samples from Clearwell Caves and Catalhoyuk by vibrational spectroscopy and other techniques. Spectrochim. Acta A Mol. Biomol.

Spectrosc. 60, 1179–1188.

Munsell Color Company, 1975. Munsell soil color charts, Maryland, U.S.A.: Macbeth Division of Kollmorgen Corporation.

Munsell Color Company., 1980. Munsell soil color charts, 1980 ed. Munsell Color Co. Inc., Baltimore, MD.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Dor, E. Ben, Brown, D.J., et al. 2015. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. Advances in Agronomy. 132:139–159.

Odlare, M., Svensson, K. & Pell, M. 2005. Near infrared reflectance spectroscopy for assessment of spatial soil variation in an agricultural field. 126:193–202.

Osborne, D.G. 2017. Spatial variation , bias and experimental design in agronomic field trials : a case study of a farming systems trial in the Western Cape province of South Africa By. (December).

Pattanayak, U. 2019. Crop Diversification and Sustainable Agriculture in India Indialics network for economics of learning , Innovation , and competence building system Crop Diversification and Sustainable Agriculture in India Minati Mallick Urmi Pattanayak Working Paper. (January 2017).

Peng, X., Shi, T., Song, A., Chen, Y. & Gao, W. 2014. Estimating Soil Organic Carbon Using Vis/NIR Spectroscopy with SVMR and SPA Methods. 2699–2717.

Post, D.F., Bryant, R.B., Batchily, A.K., Huete, A.R., Levine, S.J., Mays, M.D. & Escadafal, R., 1993. Correlations between field and laboratory measurements of soil colour. In J. M. Bigham & E. J. Ciolkosz, eds. Soil color. Madison, Wisconsin: Soil Science Society of America and American Society of Agronomy, 35–49.

Ramirez-lopez, L., Wadoux, A. & Franceschini, M.H.D. 2018. Robust soil mapping at the farm scale with vis – NIR. (November).

Reeves, J.B. 1999. Near infrared reflectance spectroscopy for the analysis of agricultural soils. 193:179–193.

Reeves, J.B. 2010. Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? *Geoderma*. 158(1–2):3–14.

Rizzo, R., Demattê, J.A.M., Lepsch, I.F., Gallo, B.C. & Fongaro, C.T. 2016. Geoderma Digital soil mapping at local scale using a multi-depth Vis – NIR spectral library and terrain attributes. 274:18–27.

Roger M. & Michael L. 2003. Soil carbon stocks and bulk density : spatial or cumulative mass coordinates as a basis of expression ? 1507–1514.

Rossel, R.A.V., Adamchuk, Ã.V.I., Sudduth, K.A., Mckenzie, N.J. & Lobsey, C.Ã. 2011. Proximal Soil Sensing : An Effective Approach for Soil Measurements in Space and

Time. Vol. 113. Elsevier Inc.

Rossel, R.A.V., Behrens, T., Ben-dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., et al. 2016. Earth-Science Reviews A global spectral library to characterize the world ' s soil. 155(January):198–230.

Sadegh, M., Rourke, S.M.O. & Holden, N.M. 2015. Geoderma Evaluation of soil quality for agricultural production using visible – near-infrared spectroscopy. *Geoderma*. 243–244:80–91.

Sakİn, E. 2012. Organic carbon organic matter and bulk density relationships in arid-semi arid soils in Southeast Anatolia region. 11(6):1373–1377.

Sánchez-Marañón, M., García, P. A., Huertas, R., Hernández-Andrés, J. & Melgosa, M., 2011. Influence of Natural Daylight on Soil Color Description: Assessment Using a Color- Appearance Model. Soil Science Society of America Journal, 75(3): 984-993.

Schumacher, B.A. 2002. Methods for the determination of total organic carbon (toc) in soils and sediments Brian. (April).

Selige, T., Böhner, J. & Schmidhalter, U. 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma*. 136(1–2):235–244.

Šestak, I., Mesić, M., Zgorelec, Ž., Perčin, A. & Stupnišek, I. 2018. Visible and near infrared reflectance spectroscopy for field-scale assessment of Stagnosols properties. 64(6):276–282.

Shah, F. & Wu, W. 2019. Soil and Crop Management Strategies to Ensure Higher Crop Productivity within Sustainable Environments. 1–19.

Shepherd, K.D. & Walsh, M.G. 2002. Development of Reflectance Spectral Libraries for Characterization of Soil Properties. 988–998.

Soil Analysis Work Committee. 1990. Handbook of Standard Soil Testing Methods for Advisory Purposes. Soil Science Society of South Africa, Republic of South Africa.

Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., MacDonald, L.M. & McLaughlin, M.J. 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Applied Spectroscopy Reviews*. 49(2):139–186.

Stenberg, B., Rossel, R.A.V. & Mouazen, A.M. 2010. Visible and near infrared spectroscopy in soil science. 107(10):1–44.

Teng, H., Viscarra, R.A., Shi, Z. & Behrens, T. 2018. Catena Updating a national soil classification with spectroscopic predictions and digital soil mapping. *Catena*. 164(February):125–134.

Torrent, J., Scwertmann, U., Fechter, H. & Alferez, F., 1983. Quantitative relationships between soil color and Hematite content. Soil Science, 136(6): 354-358.

Torrent, J. & Barron, V., 1993. Laboratory measurements of Soil Color: theory and practice. In J. M. Bigham & E. J. Ciolkosz, eds. Soil color. Madison: Soil Science Society of America, 21–33.

Vågen, T., Winowiecki, L.A., Tondoh, J.E., Desta, L.T. & Gumbricht, T. 2016. Geoderma Mapping of soil properties and land degradation risk in Africa using MODIS re fl ectance. *Geoderma*. 263:216–225.

Van Es, H.M., Gomes, C.P., Sellmann, M. & van Es, C.L. 2007. Spatially-Balanced Complete Block designs for field experiments. 140:346–352.

Van Huyssteen, C. W., Ellis, F. & Lambrechts, J. J. N., 1997. The relationship between subsoil colour and degree of wetness in a suite of soils in the Grabouw district, Western Cape II. Predicting duration of water saturation and evaluation of colour definitions for colour-defined horizons South African Journal of Plant and Soil, 14(4): 154-157.

Vasques, G.M., Demattê, J.A.M., Viscarra, R.A., Ramírez-lópez, L. & Terra, F.S. 2014. Geoderma Soil classi fi cation using visible / near-infrared diffuse re fl ectance spectra from multiple depths. *Geoderma*. 223–225:73–78.

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

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

Viscarra Rossel, R.A., Minasny, B., Roudier, P. & McBratney, A.B., 2006. Colour space models for soil science. Geoderma, 133, 320–337.

Viscarra Rossel, R.A., Cattle, S.R., Ortega, A. & Fouad, Y. 2009. In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. *Geoderma*. 150(3–4):253–266.

Vohland, M., Ludwig, M., Thiele-Bruhn, S. & Ludwig, B. 2014. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. *Geoderma*. 223–225(1):88–96.

Vohral, M. 2011. Bruker MPA User Manual. 1–42.

Waruru, B.K., Shepherd, K.D., Ndegwa, G.M. & Sila, A.M. 2016. ScienceDirect Special Issue : Proximal Soil Sensing Estimation of wet aggregation indices using soil properties and diffuse reflectance near infrared spectroscopy : An application of classification and regression tree analysis. Biosystems Engineering. 152:148–164.

Weidong, L., Baret, F., Xingfa, G., Qingxi, T., Lanfen, Z. & Bing, Z. 2002. Relating soil surface moisture to reflectance. 81:238–246.

Wendt J.W. & Hauser S. 2013. An equivalent soil mass procedure for monitoring soil organic

carbon in multiple soil layers. (February):58–65.

Wiese, L., Ros, I., Rozanov, A., Boshoff, A., Clercq, W. De & Seifert, T. 2016. Geoderma An approach to soil carbon accounting and mapping using vertical distribution functions for known soil types. *Geoderma*. 263:264–273.

Zhou, S.H.I., Qianlong, W., Jie, P., Wenjun, J.I., Huanjun, L.I.U., Xi, L.I. & Rossel, R.A.V. 2014. Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations. 57(7):1671–1680.

Zhou J. M. and Shen Y. Z. 2019. Investigation of soil properties using different techniques of mid-infrared spectroscopy. (January):96–106.

Appendix A: Raw analytical data

Yield (t/ha) and crop sequence for the period of four years (2015-2018)

Plot no.	Crop_15	Yield_15	Crop_16	Yield_16	Crop_17	Yield_17	Crop_18	Yield_18
1	Wheat	1.98	Wheat	3.41	Faba beans		Wheat	3.03
2	Wheat	1.99	Oats		Linseed		Wheat	2.31
3	Wheat	1.92	Wheat	3.24	Canola	1.35	Wheat	2.87
4	Wheat	2.01	Faba beans		Wheat	2.65	Legume cover	
5	Wheat	2.09	Wheat	3.92	Legume Cover		Wheat	2.90
6	Wheat	2.15	Linseed		Wheat	2.74	Legume cover	
7	Wheat	2.10	Canola		Wheat	2.02	Legume cover	
8	Wheat	1.89	Legume cover		Wheat	2.39	Wheat	2.30
9	Wheat	1.48	Wheat	3.93	Legume Cover		Wheat	2.11
10	Wheat	2.08	Wheat	3.51	Wheat	2.34	Canola	
11	Wheat	1.86	Wheat	3.94	Legume Cover		Wheat	2.51
12	Wheat	1.95	Legume cover		Wheat	1.37	Barley	
13	Wheat	2.17	Legume cover		Wheat	1.98	Wheat	2.04
14	Wheat	1.83	Wheat	4.11	Barley		Canola	
15	Wheat	2.13	Wheat	4.42	Canola		Grass cover	
16	Wheat	2.21	Wheat	3.48	Wheat	2.03	Canola	
17	Wheat	1.98	Canola		Grass cover		Vetch	
18	Wheat	2.03	Barley		Canola		Grass cover	
19	Wheat	2.24	Wheat	3.82	Canola		Grass cover	

20	Wheat	2.17	Canola		Grass cover		Faba beans	
21	Wheat	2.09	Canola		Grass cover		Wheat	2.30
22	Wheat	2.31	Grass cover		Wheat		Canola	
23	Wheat		Grass cover		Faba beans		Wheat	2.09
24	Wheat	2.15	Faba beans		Wheat	2.79	Oats	
25	Wheat	2.16	Wheat	4.19	Canola		Wheat	2.17
26	Wheat	2.17	Grass cover		Vetch		Wheat	1.99
27	Wheat	2.18	Vetch		Wheat	2.61	Wheat	2.47
28	Wheat	2.19	Canola		Wheat	2.32	Canola	
29	Wheat	2.2	Wheat	3.97	Wheat	2.44	Faba beans	0.52
30	Wheat	2.21	Wheat	4.06	Oats		Linseed	
31	Wheat	2.22	Wheat	4.60	Canola		Grass cover	
32	Wheat	2.23	Wheat	4.24	Canola		Grass cover	
33	Wheat	2.24	Wheat	3.72	Faba beans		Wheat	2.47
34	Wheat	2.26	Wheat	4.02	Barley		Canola	
35	Wheat	2.27	Grass cover		Faba beans		Wheat	2.35
36	Wheat	2.28	Grass cover		Vetch		Wheat	2.26
37	Wheat	2.29	Barley		Canola		Grass cover	
38	Wheat	2.30	Legume cover		Wheat	2.8	Wheat	2.50
39	Wheat	2.31	Linseed		Wheat	3.08	Legume cover	
40	Wheat	2.32	Faba beans		Wheat	3.07	Oats	
41	Wheat	2.33	Canola		Grass cover		Wheat	2.09
42	Wheat	2.34	Wheat	3.77	Wheat	2.47	Faba beans	
43	Wheat	2.35	Legume cover		Wheat	2.96	Barley	

44	Wheat		Wheat		Wheat	2.89	Canola	
45	Wheat	2.15	Canola		Grass cover		Faba beans	
46	Wheat	2.16	Wheat	3.87	Legume Cover		Wheat	2.53
47	Wheat	2.17	Wheat	3.53	Oats	1.37	Linseed	
48	Wheat	2.18	Canola		Wheat	2.85	Canola	
49	Wheat	2.19	Canola		Wheat	2.64	Legume cover	
50	Wheat	2.20	Grass cover		Wheat	2.61	Canola	
51	Wheat	2.21	Canola		Grass cover		Vetch	
52	Wheat	2.22	Wheat	3.80	Wheat		Canola	
53	Wheat	2.23	Wheat	3.90	Legume Cover		Wheat	2.45
54	Wheat	2.24	Vetch		Wheat	2.83	Wheat	2.54
55	Wheat	2.26	Wheat	3.98	Canola		Wheat	2.62
56	Wheat	2.27	Faba beans		Wheat	3.2	Legume cover	
57	Wheat	1.98	Wheat	3.92	Canola		Wheat	2.25
58	Wheat	1.71	Oats		Linseed		Wheat	2.08
59	Wheat	1.79	Legume cover		Wheat	2.63	Wheat	2.63
60	Wheat	1.84	Wheat	3.93	Legume Cover		Wheat	2.38
61	Wheat	1.60	Linseed		Wheat	2.86	Legume cover	
62	Wheat	1.89	Wheat	4.21	Barley		Canola	
63	Wheat	1.58	Wheat	4.25	Canola		Wheat	1.97
64	Wheat	1.83	Wheat	4.53	Faba beans		Wheat	2.15
65	Wheat	2.00	Wheat	3.71	Canola		Grass cover	
66	Wheat	1.83	Legume cover		Wheat	2.69	Wheat	2.39
67	Wheat	1.62	Grass cover		Wheat	2.65	Canola	

68	Wheat	2.21	Canola		Grass cover		Vetch	
69	Wheat	1.93	Canola		Grass cover		Wheat	1.62
70	Wheat	2.09	Canola		Wheat	2.83	Canola	1.23
71	Wheat	1.83	Legume cover		Wheat	2.44	Wheat	2.31
72	Wheat	1.68	Canola		Wheat	2.51	Legume cover	
73	Wheat	1.94	Wheat	3.98	Wheat	2.64	Faba beans	1.08
74	Wheat	2.06	Vetch		Wheat	2.89	Wheat	2.15
75	Wheat	1.79	Faba beans		Wheat	2.77	Oats	
76	Wheat	2.03	Canola		Grass cover		Faba beans	
77	Wheat	1.83	Barley		Canola		Grass cover	
78	Wheat	1.64	Wheat	3.51	Legume Cover		Wheat	2.14
79	Wheat	1.69	Wheat	3.64	Legume Cover		Wheat	2.19
80	Wheat	1.63	Wheat	4.05	Canola		Wheat	2.19
81	Wheat	1.86	Grass cover		Faba beans		Wheat	2.11
82	Wheat	1.82	Wheat	4.28	Oats		Linseed	
83	Wheat	2.07	Wheat	4.47	Wheat	2.80	Canola	
84	Wheat	1.86	Wheat	3.52	Canola		Grass cover	
85	Wheat	1.55	Grass cover		Vetch		Wheat	2.14
86	Wheat	1.56	Faba beans		Wheat	2.54	Legume cover	
87	Wheat	1.79	Oats		Linseed		Wheat	2.13
88	Wheat	1.66	Legume cover		Wheat	2.62	Barley	
89	Wheat	1.64	Wheat	3.93	Wheat	2.75	Canola	
90	Wheat	1.89	Wheat	3.97	Legume Cover		Wheat	1.89
91	Wheat	1.13	Linseed		Wheat	2.04	Legume cover	

92	Wheat	1.07	Wheat	2.88	Wheat	2.29	Faba beans	
93	Wheat	1.16	Grass cover		Vetch		Wheat	1.94
94	Wheat	1.16	Canola		Grass cover		Vetch	
95	Wheat	1.16	Faba beans		Wheat	2.49	Legume cover	
96	Wheat	1.09	Wheat	3.60	Wheat	2.68	Canola	
97	Wheat	1.02	Wheat	2.30	Wheat	1.94	Canola	
98	Wheat	0.74	Barley		Canola		Grass cover	
99	Wheat	0.73	Grass cover		Wheat	1.88	Canola	1.01
100	Wheat	1.06	Wheat	2.33	Legume Cover		Wheat	2.10
101	Wheat	0.97	Legume cover		Wheat	2.18	Barley	
102	Wheat	0.77	Wheat	2.47	Canola		Wheat	2.52
103	Wheat	0.8	Wheat	3.12	Oats		Linseed	
104	Wheat	0.95	Legume cover		Wheat	1.90	Wheat	2.09
105	Wheat	0.89	Canola		Grass cover		Faba beans	
106	Wheat	1.08	Wheat	3.06	Barley		Canola	
107	Wheat	0.76	Wheat	2.45	Canola		Grass cover	
108	Wheat	0.81	Vetch		Wheat	1.89	Wheat	2.26
109	Wheat	0.89	Grass cover		Faba beans		Wheat	2.08
110	Wheat	0.75	Wheat	2.68	Canola		Wheat	2.33
111	Wheat	1.10	Canola		Wheat	1.80	Canola	
112	Wheat	0.87	Canola		Wheat	1.82	Legume cover	
113	Wheat	0.94	Wheat	3.00	Legume Cover		Wheat	2.11
114	Wheat	1.11	Faba beans		Wheat	1.92	Oats	
115	Wheat	1.12	Wheat	2.94	Faba beans		Wheat	2.26

116	Wheat	0.64	Oats		Linseed		Wheat	2.54
117	Wheat	0.80	Wheat	2.80	Legume Cover		Wheat	2.60
118	Wheat	0.95	Legume cover		Wheat	2.00	Wheat	2.35
119	Wheat	0.96	Wheat	3.10	Canola		Grass cover	
120	Wheat	0.91	Canola		Grass cover		Wheat	2.05

2015 soil properties data

Plot no.	BD (g/m3)	C(%)	Cstock (kg/m2)	CEC (Meq/100g)	Ca (Cmol/Kg)	pH	Total NIR (Abs)_2016
1	1.39	2.03	1.32	9.29	5.8	5.5	198.26
2	1.33	2.15	1.24	9.16	5.57	5.6	189.59
3	1.34	1.79	1.04	8.07	4.57	5.5	205.34
4	1.27	2.96	1.47	9.54	6.97	5.5	180.2
5	1.27	2.61	1.43	10.25	6.6	5.3	181.93
6	1.21	3.39	1.75	11.99	8.19	5.8	180.73
7	1.26	3.12	1.72	10.49	6.24	5.1	181.07
8	1.3	2.54	1.42	8.22	5.71	5.7	186.11
9	1.32	2.26	1.21	8.35	5.08	5	183.7
10	1.21	2.89	1.4	11.64	8.34	6	194.98
11	1.23	2.93	1.52	10.89	8.23	6	190.52
12	1.33	2.13	1.27	16.17	10.92	6.3	188.59
13	1.29	2.07	1.17	13.34	10.04	6.3	190.34
14	1.14	2.26	1.12	12.26	9.38	6.2	197.09
15	1.28	2.46	1.33	16.49	12.57	6.4	202.81
16	1.15	2.44	1.15	16.03	12.7	6.4	209.19
17	1.26	2.67	1.41	13.63	9.72	6.3	205.57
18	1.24	2.22	1.22	18.22	12.64	6.8	201.35
19	1.34	2.07	1.06	10.37	7.62	6.1	193.99
20	1.17	3.35	1.59	13.64	10.36	5.8	193.18
21	1.47	0.53	0.32	11.09	7.69	5.8	199.27
22	1.31	2.48	1.31	13.2	9.74	6.1	227.71
23	1.18	2.73	1.35	11.37	8.33	5.7	215.17
24	1.22	2.85	1.45	15.37	11.76	6.4	187.21
25	1.34	2.54	1.39	19.44	13.66	6.7	204.39
26	1.37	2.34	1.24	19.59	13.14	6.8	221.17
27	1.19	2.85	1.29	12.52	9.11	5.9	209.88
28	1.26	2.22	1.17	11.94	8.69	5.7	212.93
29	1.3	2.59	1.55	15.4	11.04	5.9	205.49
30	1.33	2.4	1.24	13.62	9.28	6.2	219.98
31	1.26	2.91	1.45	14.42	10.32	5.9	213.96
32	1.23	2.94	1.45	16.37	11.85	6.3	209.84
33	1.29	2.54	1.35	14.03	10.88	6.4	192.86
34	1.46	2.07	1.15	14.12	9.97	6.4	207.37
35	1.48	2.48	1.5	10.89	7.96	5.7	196.04
36	1.27	2.93	1.48	13.24	9.32	5.8	212.64
37	1.49	1.89	1.06	10.64	7.8	6	198.24
38	1.32	2.09	1.19	12.41	8.76	6.3	195.32
39	1.64	1.5	0.9	9.19	6.02	5.7	218.61
40	1.47	2.18	1.18	17.91	13.3	6.8	210.31
41	1.24	3.59	1.82	16.9	13.34	6.4	212.59
42	1.31	3.35	1.81	15.32	11.45	6.3	201.63

43	1.34	2.48	1.35	12.58	9.48	6.1	201.79
44	1.47	2.34	1.29	10.05	7.04	5.5	221.85
45	1.27	1.77	0.91	8.51	6.31	6.1	211.71
46	1.25	2.38	1.22	11.46	8.27	5.9	220.73
47	1.34	2.65	1.36	13.7	10.74	6.1	202.45
48	1.36	2.15	1.18	8.52	5.49	5.4	199.73
49	1.28	2.73	1.37	11.91	9.65	5.9	204.22
50	1.34	1.7	0.97	11.59	9.18	6.2	203.24
51	1.41	1.64	0.97	7.16	4.64	5.2	204.51
52	1.22	2.42	1.16	12.56	9.15	6	217.4
53	1.22	2.85	1.46	13.6	10.71	6.3	204.23
54	1.27	2.85	1.51	12.67	10.05	6.1	203.16
55	1.08	4.45	1.98	17.89	13.17	6.7	224.27
56	1.19	2.22	1.19	11.03	7.87	5.9	187.33
57	1.25	2.34	1.26	14.03	10.02	6.6	196.6
58	1.23	2.07	1.07	9.18	6.24	5.3	194.56
59	1.34	3.2	1.73	11.41	9.04	6	205.37
60	1.15	3.43	1.49	14.83	10.82	6.2	214.04
61	1.38	2.59	1.43	14.32	11.17	6.5	206.01
62	1.42	3.82	2.06	18.15	13.82	6.5	216.76
63	1.27	2.59	1.46	14.79	10.6	6.3	200.57
64	1.27	2.24	1.21	13.03	8.84	6.3	209.22
65	1.32	2.57	1.39	14.09	11.25	6.5	202.27
66	1.34	2.2	1.18	9.77	7.9	6	204.97
67	1.27	2.73	1.37	12.77	8.98	6.1	211.05
68	1.49	2.01	1.18	8.62	6.64	6.1	202.37
69	1.27	2.98	1.46	14.83	10.1	6.5	206.18
70	1.27	2.36	1.26	13.88	10.11	6.4	203.12
71	1.38	2.13	1.19	8.2	5.49	5.3	204.44
72	1.47	1.58	0.93	8.32	5.51	5.4	188.76
73	1.5	1.89	1.1	8.25	6.09	5.5	197.82
74	1.35	1.99	1.14	8.25	5.71	5.7	200.03
75	1.21	2.24	1.19	12.42	8.79	6.1	201.23
76	1.38	2.26	1.31	12.75	8.73	6.2	201.28
77	1.47	2.67	1.44	9.17	6.32	5.6	209.96
78	1.48	1.58	0.97	6.41	5.08	6.1	209.98
79	1.4	1.97	1.12	9.07	6.9	6.3	200.3
80	1.32	2.01	1.13	9.22	7.08	6.2	186.11
81	1.48	2.11	1.25	8.08	5.95	5.7	202.91
82	1.4	1.81	1.07	10.96	7.81	6	203.35
83	1.45	2.54	1.37	13.02	9.16	6.3	212.74
84	1.32	1.97	1.01	10.13	7.08	6.2	208.96
85	1.43	1.76	1.02	6.85	5.21	6	205.47
86	1.29	2.05	1.06	8.03	5.85	5.7	208.65
87	1.52	2.09	1.17	8.68	6.4	6	198.59
88	1.46	1.99	1.15	11.37	8.16	6.4	198.94

89	1.37	2.22	1.16	11.47	8.02	6.6	203.26
90	1.44	1.79	1.01	8.75	6.11	5.9	202.84
91	1.37	2.09	1.07	8.1	5.96	6	192.93
92	1.46	1.74	0.94	7.73	5.88	6	190.73
93	1.38	1.7	0.89	7.53	5.45	5.8	194.17
94	1.4	2.77	1.53	11.63	8.8	6.6	190.82
95	1.37	1.93	1.11	11.6	8.68	6.3	186.82
96	1.5	1.33	0.78	7.8	5.37	6.1	192.8
97	1.46	1.93	1.01	8.75	6.46	6.2	196.83
98	1.54	1.56	0.83	8.78	6.29	6.2	200.84
99	1.62	1.76	0.99	9.91	7.11	6.8	195.43
100	1.44	1.91	1.02	8.62	6.71	6.5	189.47
101	1.5	1.72	0.94	10.94	8.28	6.8	191.01
102	1.45	1.07	0.63	4.9	3.08	5.4	185.85
103	1.48	1.31	0.8	4.61	3.35	5.6	186.19
104	1.4	1.81	1.01	7.05	5.2	6.4	202.68
105	1.51	2.03	1.15	10.04	8.11	6.4	204.02
106	1.57	2.34	1.24	7.69	6.07	5.8	195.96
107	1.41	2.2	1.16	9.98	7.64	6.3	199.47
108	1.66	1.58	0.91	7.36	5.94	6.4	196.71
109	1.46	2.09	1.22	6.2	3.67	5.1	205.5
110	1.5	1.66	0.89	6.15	4.33	5.4	207.71
111	1.52	1.52	0.83	11.09	8.86	6.5	208.53
112	1.56	1.83	0.98	11.47	9.38	6.5	206.16
113	1.44	1.58	0.78	7.31	5.67	5.8	203.51
114	1.56	1.89	1.02	8.99	7.04	6.3	208.37
115	1.41	1.87	1	6.03	3.2	4.8	192.04
116	1.39	1.3	0.72	6.74	4.83	5.8	196.19
117	1.48	1.08	0.65	7.07	5.36	5.6	201.91
118	1.52	1.64	0.86	7.82	6.04	5.9	209.7
119	1.48	2.24	1.15	12.06	8.86	6.7	216.07
120	1.52	2.2	1.15	9.73	7.39	5.9	215.73

2019 soil properties data

				Colour (2015 samples)						Total
Plot no.	C (%)	ρ (g/cm3)	C_{stock} (kg/m2)	L*(D65)	a*(D65)	b*(D65)	Hue	Value	Chroma	NIR (Abs)
1	2.36	1.22	1.19	57.36	9.98	20.08	7.5YR	6	4	189.48
2	1.7	1.25	0.77	56.52	8.88	17.95	7.5YR	6	3	204.86
3	1.33	1.25	0.65	57.94	10.08	20.6	7.5YR	6	4	187.58
4	2.42	1.23	1.15	55.96	7.27	16.32	7.5YR	6	3	186.51
5	2.5	1.32	1.24	56.59	7.64	16.73	7.5YR	6	3	196.94
6	2.54	1.29	1.34	55.49	8.31	16.66	7.5YR	5	3	194.78
7	1.64	1.17	0.78	55.23	8.7	16.99	7.5YR	5	3	189.84
8	1.07	1.17	0.57	56.17	7.77	16.89	7.5YR	6	3	192.12
9	2.07	1.19	1.05	56.04	7.34	16.95	7.5YR	6	3	209.21
10	1.85	1.17	0.9	53.7	8.36	16.01	7.5YR	5	3	190.51
11	1.91	1.26	0.92	54.41	7.11	15.11	7.5YR	5	3	209.48
12	2.28	1.2	1.17	53.85	9.06	16.56	7.5YR	5	3	193.20
13	2.03	1.13	0.98	54.37	7.72	15.51	7.5YR	5	3	195.72
14	2.07	1.18	1.06	54.37	7.84	15.36	7.5YR	5	3	191.57
15	2.57	1.07	1.19	54.19	7.64	15.42	7.5YR	5	3	199.08
16	2.09	1.18	0.95	53.19	7.61	14.97	7.5YR	5	3	201.81
17	2.3	1.17	1.12	53.87	7.58	14.97	7.5YR	5	3	211.41
18	2.09	1.15	0.97	54.49	8.66	15.9	7.5YR	5	3	199.98
19	2.85	1.24	1.31	55.25	9.6	17.03	5YR	5	3	205.56
20	2.34	1.13	1.04	54.78	8.51	15.96	7.5YR	5	3	212.78

21	3.24	1.22	1.47	56.14	9.03	17.14	7.5YR	6	3	217.12
22	2.4	1.2	1.16	53.77	7.39	14.71	7.5YR	5	3	217.84
23	2.5	0.99	1.08	54.11	8.23	15.5	7.5YR	5	3	210.74
24	3.24	0.85	1.26	55.59	9.11	16.72	7.5YR	5	3	201.47
25	2.61	1.12	1.2	54.25	8.07	15.49	7.5YR	5	3	213.92
26	2.89	1.23	1.31	53.59	7.84	15.3	7.5YR	5	3	202.05
27	3.43	1.17	1.5	53.46	8.18	15.4	7.5YR	5	3	203.90
28	1.4	1.19	0.67	53.66	7.7	14.84	7.5YR	5	3	209.16
29	2.94	1.17	1.38	53.97	7.83	15.28	7.5YR	5	3	211.03
30	2.85	1	1.16	53.95	7.93	15.49	7.5YR	5	3	211.47
31	2.18	1.1	0.93	53.11	7.2	14.56	7.5YR	5	3	215.33
32	1.64	1.18	0.86	53.03	7.23	14.48	7.5YR	5	3	195.27
33	2.4	1.05	1.05	54.37	8.22	15.97	7.5YR	5	3	201.16
34	2.22	1.32	1.12	53.84	8.18	16.05	7.5YR	5	3	207.01
35	1.77	1.3	0.89	53.82	7.88	15.44	7.5YR	5	3	204.70
36	2.4	1.28	1.15	53.49	7.78	15.41	7.5YR	5	3	241.05
37	2.3	1.11	1.05	55	7.81	15.92	7.5YR	5	3	216.39
38	2.81	1.2	1.35	52.92	8.05	15.06	7.5YR	5	3	211.46
39	2.28	1.1	1.02	54.27	8.27	16.4	7.5YR	5	3	212.36
40	1.5	1.25	0.65	53.85	7.45	15.53	7.5YR	5	3	226.18
41	1.79	1.22	0.91	53.45	7.59	15.31	7.5YR	5	3	205.45
42	1.89	1.2	0.95	54.3	7.32	15.66	7.5YR	5	3	212.40
43	2.18	1.28	1.1	51.93	7.57	14.32	7.5YR	5	3	210.87
44	1.7	1.17	0.81	54.1	9.58	17.2	7.5YR	5	3	204.06

45	1.58	1.42	0.85	55.57	8.74	17.12	7.5YR	5	3	208.79
46	1.52	1.32	0.78	53.78	7.32	15.31	7.5YR	5	3	208.43
47	2.73	1.23	1.3	54.11	7.43	15.79	7.5YR	5	3	210.19
48	1.27	1.25	0.63	54.8	7.48	15.75	7.5YR	5	3	223.24
49	2.81	1.27	1.46	53.76	8.11	15.63	7.5YR	5	3	215.45
50	2.38	1.17	1.12	53.6	8.14	15.82	7.5YR	5	3	205.23
51	2.73	1.22	1.18	53.42	7.85	15.33	7.5YR	5	3	216.80
52	3.32	1.17	1.56	53.19	7.23	14.99	7.5YR	5	3	209.53
53	1.83	1.29	0.96	53.89	7.83	15.82	7.5YR	5	3	213.44
54	2.15	1.25	1.14	53.46	7.62	15.86	7.5YR	5	3	207.91
55	1.76	1.28	0.94	52.33	6.9	14.44	7.5YR	5	3	205.80
56	1.7	1.29	0.87	55.06	7.85	17.17	10YR	5	3	215.30
57	2.18	1.25	1.15	54.65	6.86	16.13	7.5YR	5	3	208.74
58	2.69	1.15	1.39	55.25	7.18	17.01	10YR	5	3	218.43
59	2.26	1.34	1.22	53.08	7.54	15.56	7.5YR	5	3	212.73
60	1.68	1.11	0.78	52.89	7.25	15.11	7.5YR	5	3	211.69
61	2.07	1.14	0.98	53.42	7.07	14.62	7.5YR	5	3	214.39
62	2.81	1.17	1.36	53.82	7.03	14.79	7.5YR	5	3	225.32
63	2.55	1.26	1.3	54.45	8.58	16.17	7.5YR	5	3	219.70
64	2.03	1.3	1.07	54.17	7.82	15.19	7.5YR	5	3	220.55
65	1.81	1.4	1	54.1	7.71	15.75	7.5YR	5	3	216.77
66	2.28	1.22	1.09	53.89	7.39	14.9	7.5YR	5	3	218.69
67	3.04	1.23	1.51	54.31	7.55	15.23	7.5YR	5	3	210.95
68	2.4	1.26	1.07	55.29	8.19	16.27	7.5YR	5	3	205.93

69	1.74	1.23	0.87	54.5	8.4	16.15	7.5YR	5	3	216.23
70	1.99	1.16	0.97	55.15	8.13	16.19	7.5YR	5	3	207.29
71	1.52	1.44	0.87	55.68	6.81	15.73	7.5YR	5	3	209.58
72	2.42	1.25	1.25	55.69	8.63	16.59	7.5YR	5	3	231.06
73	2.3	1.18	1.11	54.89	8.33	16.27	7.5YR	5	3	221.84
74	1.89	1.26	0.98	54.42	8.99	16.1	7.5YR	5	3	210.70
75	2.01	1.14	0.93	53.28	9.24	16.09	7.5YR	5	3	208.39
76	2.22	1.16	1.11	53.76	8.14	15.43	7.5YR	5	3	216.09
77	2.79	1.07	1.24	54.59	5.99	15.55	7.5YR	5	3	231.73
78	1.64	1.36	0.84	55.84	6.69	15.97	7.5YR	6	3	228.98
79	3.24	0.91	1.32	54.44	7.65	15.97	7.5YR	5	3	220.31
80	2.11	1.28	1.14	53.49	8.46	15.47	7.5YR	5	3	218.08
81	2.65	1.32	1.4	52.67	8.32	15.45	7.5YR	5	3	220.72
82	1.27	1.35	0.71	52.77	9.44	16.66	7.5YR	5	3	211.98
83	2.22	1.35	1.21	52.42	8.1	15.37	7.5YR	5	3	220.67
84	2.26	1.38	1.14	54.11	5.94	15.93	7.5YR	5	3	210.02
85	1.87	1.25	1.01	54.38	6.14	15.76	7.5YR	5	3	218.20
86	2.34	1.34	1.24	54.19	6.94	16.26	7.5YR	5	3	212.73
87	1.33	1.41	0.73	53.95	7.18	16.05	7.5YR	5	3	201.23
88	1.83	1.25	0.94	53.54	8.41	16.26	7.5YR	5	3	211.68
89	2.94	1.14	1.45	53.15	8.05	15.45	7.5YR	5	3	218.73
90	1.83	1.17	0.92	54.32	8.64	16.44	7.5YR	5	3	211.14
91	2.09	1.37	1.09	56.01	5.09	15.67	10YR	6	3	202.21
92	1.07	1.41	0.56	56.36	5.54	16.47	10YR	6	3	205.80

93	1.89	1.29	0.98	56.87	5.36	16.66	10YR	6	3	201.73
94	2.22	1.34	1.19	55.38	5.18	15.37	2.5Y	5	2	219.38
95	2.05	1.35	1.1	56.84	6.01	17.22	10YR	6	3	204.31
96	1.97	1.21	1.04	56.49	5.99	16.79	10YR	6	3	197.41
97	1.66	1.34	0.98	55.63	5.14	15.81	2.5Y	5	3	193.70
98	1.99	1.06	0.95	55.06	5.51	15.66	7.5YR	5	3	198.14
99	1.97	1.36	1.06	55.73	5.21	15.91	2.5Y	5	3	199.61
100	2.11	1.24	0.98	55.93	4.96	15.24	2.5Y	5	2	203.07
101	1.7	1.33	0.88	55.29	6.5	15.74	7.5YR	5	3	205.92
102	1.76	1.38	0.94	57.77	6.32	16.87	10YR	6	3	208.48
103	1.77	1.35	0.84	58.05	6.04	17.09	10YR	6	3	204.08
104	1.29	1.51	0.7	54.52	6.05	15.7	7.5YR	5	3	206.63
105	1.6	1.55	0.85	53.79	5.64	15.08	7.5YR	5	3	222.96
106	2.16	1.4	1.17	55.19	6.23	15.64	7.5YR	5	3	209.09
107	1.85	1.35	0.92	54.01	6.37	15.09	7.5YR	5	3	212.05
108	1.72	1.55	0.9	55.62	6.85	16.26	7.5YR	5	3	205.24
109	1.79	1.35	0.96	55.52	6.31	15.8	7.5YR	5	3	215.39
110	1.89	1.45	1.04	55.13	6.92	16.07	7.5YR	5	3	209.17
111	1.09	1.47	0.58	53.78	5.9	15	7.5YR	5	3	199.59
112	1.79	1.41	0.94	53.6	6.67	15.04	7.5YR	5	3	211.56
113	1.6	1.37	0.83	53.96	7.4	15.73	7.5YR	5	3	208.16
114	1.62	1.48	0.86	53.81	6.66	15.11	7.5YR	5	3	203.79
115	1.19	1.39	0.62	55.43	7.71	17.17	10YR	5	3	208.99
116	1.35	1.55	0.76	54.71	8.04	16.12	7.5YR	5	3	199.09

117	1.21	1.29	0.66	53.76	8.37	16.17	7.5YR	5	3	213.09
118	2.01	1.43	1.01	52.92	7.67	15.07	7.5YR	5	3	220.62
119	2.59	1.36	1.33	52.86	8.22	15.3	7.5YR	5	3	219.52
120	2.18	1.45	1.11	52.93	7.58	15.09	7.5YR	5	3	223.82

Soil organic carbon prediction results and Bulk density t.tests (p-values) per plot for 2015 vs 2019

Plot no.	Prediction	Residuals	Delta NIR	Delta Cwb	ρ P-values
1	2.25	0.11	0.22	0.33	0.041
2	2.34	-0.64	0.19	-0.45	0.049
3	1.91	-0.58	0.12	-0.46	0.143
4	2.28	0.14	-0.68	-0.54	0.78
5	2.55	-0.05	-0.06	-0.11	0.623
6	2.54	0	-0.85	-0.85	0.376
7	1.87	-0.23	-1.25	-1.48	0.549
8	0.81	0.26	-1.73	-1.47	0.365
9	2.25	-0.18	-0.01	-0.19	0.067
10	2.7	-0.85	-0.19	-1.04	0.76
11	2.31	-0.4	-0.62	-1.02	0.058
12	2.29	-0.01	0.16	0.15	0.254
13	2.37	-0.34	0.3	-0.04	0.335
14	2.42	-0.35	0.16	-0.19	0.736
15	2.5	0.07	0.04	0.11	0.101
16	2.51	-0.42	0.07	-0.35	0.323
17	2.36	-0.06	-0.31	-0.37	0.497
18	2.45	-0.36	0.23	-0.13	0.38
19	2.3	0.55	0.23	0.78	0.128
20	0.31	2.03	-3.04	-1.01	0.872
21	2.67	0.57	2.14	2.71	0.115
22	2.28	0.12	-0.2	-0.08	0.344
23	2.46	0.04	-0.27	-0.23	0.294
24	2.48	0.76	-0.37	0.39	0.297
25	2.67	-0.06	0.13	0.07	0.07
26	2.78	0.11	0.44	0.55	0.083
27	2.54	0.89	-0.31	0.58	0.155
28	1.57	-0.17	-0.65	-0.82	0.744
29	2.33	0.61	-0.26	0.35	0.513
30	2.51	0.34	0.11	0.45	0.064
31	2.38	-0.2	-0.53	-0.73	0.032
32	1.75	-0.11	-1.19	-1.3	0.51
33	2.42	-0.02	-0.12	-0.14	0.348

34	2.36	-0.14	0.29	0.15	0.079
35	2.34	-0.57	-0.14	-0.71	0.359
36	2.54	-0.14	-0.39	-0.53	0.133
37	2.46	-0.16	0.57	0.41	0.009
38	2.61	0.2	0.52	0.72	0.157
39	2.69	-0.41	1.19	0.78	0.018
40	2.37	-0.87	0.19	-0.68	0.032
41	2.55	-0.76	-1.04	-1.8	0.887
42	2.21	-0.32	-1.14	-1.46	0.216
43	2.96	-0.78	0.48	-0.3	0.591
44	2.04	-0.34	-0.3	-0.64	0.054
45	1.9	-0.32	0.13	-0.19	0.123
46	1.86	-0.34	-0.52	-0.86	0.522
47	2.35	0.38	-0.3	0.08	0.386
48	0.41	0.86	-1.74	-0.88	0.525
49	2.36	0.45	-0.37	0.08	0.952
50	2.46	-0.08	0.76	0.68	0.055
51	2.9	-0.17	1.26	1.09	0.065
52	2.93	0.39	0.51	0.9	0.778
53	2.57	-0.74	-0.28	-1.02	0.607
54	2.4	-0.25	-0.45	-0.7	0.873
55	2.5	-0.74	-1.95	-2.69	0.244
56	2.03	-0.33	-0.19	-0.52	0.479
57	2.22	-0.04	-0.12	-0.16	0.967
58	2.14	0.55	0.07	0.62	0.471
59	2.23	0.03	-0.97	-0.94	0.98
60	1.92	-0.24	-1.51	-1.75	0.787
61	2.49	-0.42	-0.1	-0.52	0.211
62	2.7	0.11	-1.12	-1.01	0.186
63	2.41	0.14	-0.18	-0.04	0.903
64	2.2	-0.17	-0.04	-0.21	0.761
65	2.1	-0.29	-0.47	-0.76	0.274
66	2.45	-0.17	0.25	0.08	0.2
67	2.62	0.42	-0.11	0.31	0.754
68	2.17	0.23	0.16	0.39	0.029
69	1.6	0.14	-1.38	-1.24	0.416
70	2.06	-0.07	-0.3	-0.37	0.044
71	1.77	-0.25	-0.36	-0.61	0.579

72	2.22	0.2	0.64	0.84	0.169
73	1.94	0.36	0.05	0.41	0.228
74	1.86	0.03	-0.13	-0.1	0.464
75	1.76	0.25	-0.48	-0.23	0.347
76	2.09	0.13	-0.17	-0.04	0.062
77	2.53	0.26	-0.14	0.12	0.042
78	1.85	-0.21	0.27	0.06	0.429
79	-0.11	3.35	-2.08	1.27	0.016
80	2.01	0.1	0	0.1	0.7
81	2.18	0.47	0.07	0.54	0.355
82	-0.56	1.83	-2.37	-0.54	0.727
83	2.3	-0.08	-0.24	-0.32	0.668
84	2.49	-0.23	0.52	0.29	0.571
85	1.93	-0.06	0.17	0.11	0.12
86	2.43	-0.09	0.38	0.29	0.697
87	1.63	-0.3	-0.46	-0.76	0.428
88	2.11	-0.28	0.12	-0.16	0.276
89	2.22	0.72	0	0.72	0.065
90	1.66	0.17	-0.13	0.04	0.01
91	1.8	0.29	-0.29	0	0.991
92	1.53	-0.46	-0.21	-0.67	0.466
93	1.65	0.24	-0.05	0.19	0.405
94	2.12	0.1	-0.65	-0.55	0.439
95	2.47	-0.42	0.54	0.12	0.894
96	1.47	0.5	0.14	0.64	0.216
97	1.44	0.22	-0.49	-0.27	0.259
98	1.62	0.37	0.06	0.43	0.013
99	1.76	0.21	0	0.21	0.014
100	1.95	0.16	0.04	0.2	0.114
101	1.91	-0.21	0.19	-0.02	0.09
102	2.28	-0.52	1.21	0.69	0.412
103	1.82	-0.05	0.51	0.46	0.249
104	1.62	-0.33	-0.19	-0.52	0.254
105	2.37	-0.77	0.34	-0.43	0.456
106	1.6	0.56	-0.74	-0.18	0.086
107	1.66	0.19	-0.54	-0.35	0.568
108	1.39	0.33	-0.19	0.14	0.293
109	1.82	-0.03	-0.27	-0.3	0.27

110	1.51	0.38	-0.15	0.23	0.824
111	0.93	0.16	-0.59	-0.43	0.769
112	1.47	0.32	-0.36	-0.04	0.076
113	1.48	0.12	-0.1	0.02	0.363
114	1.64	-0.02	-0.25	-0.27	0.215
115	1.26	-0.07	-0.61	-0.68	0.641
116	1.32	0.03	0.02	0.05	0.15
117	1.14	0.07	0.06	0.13	0.065
118	1.75	0.26	0.11	0.37	0.437
119	1.77	0.82	-0.47	0.35	0.43
120	1.47	0.71	-0.73	-0.02	0.624

Appendix B: Statistical model development report (from the OPUS software)

An example of how the statistical analysis was performed in OPUS (Quant 2) software for model development

Validation Report

General Information

Method File:	Yield_2016_blocks_N.q2
Standards (total):	51
Calibration Spectra:	48
Test Spectra:	0
Data Block:	AB
Components (total):	2
Frequency Regions	1
Selected Datapoints:	1154
Mean Centering:	Yes
Preprocessing:	First derivative + Straight line subtraction
Smoothing Points:	17

Frequency Regions

from	to
3594.8	12489.4

Mean prediction error

	Rank	R^2	RMSECV	Bias	RPD	Offset	Slope
1	1	62.72	0.138	-0.00537	1.64	0.245	0.602
2	2	63.54	0.137	-3.68E-006	1.66	0.199	0.670
3	3	62.28	0.139	-0.00606	1.63	0.190	0.695
4	4	64.99	0.134	-0.0012	1.69	0.161	0.735
5	5	76.39	0.11	-0.000466	2.06	0.122	0.799
6	6	74.76	0.114	-0.00195	1.99	0.098	0.841
7	7	69.07	0.126	5.4E-005	1.8	0.102	0.831
8	8	62.28	0.139	0.000385	1.63	0.083	0.861
9	9	59.74	0.144	-0.000813	1.58	0.088	0.855
10	10	58.4	0.146	-0.00123	1.55	0.089	0.854

Component values (reference measured, predicted values and the difference)

	File Name	True	Prediction	Difference	Component Value Density
1	(1).0	0.48	0.4147	0.0653	18.24
2	(3).0	0.4	0.4993	-0.0993	56.03
3	(5).0	0.7	0.7323	-0.0323	137.93
4	(9).0	0.71	0.6457	0.0643	68.21
5	(10).0	0.52	0.4604	0.0596	21.24
6	(11).0	0.71	0.5726	0.137	42.23
7	(14).0	0.79	0.6011	0.189	51.38
8	(15).0	0.92	0.7715	0.149	63.58
9	(16).0	0.51	0.7435	-0.233	80.00
10	(19).0	0.66	0.7599	-0.0999	100.00
11	(25).0	0.82	0.8796	-0.0596	26.82
12	(29).0	0.72	0.7912	-0.0712	43.30
13	(30).0	0.76	0.8388	-0.0788	80.00
14	(31).0	1	0.7855	0.215	44.44
15	(32).0	0.84	0.6968	0.143	244.24
16	(33).0	0.61	0.7291	-0.119	200.00
17	(34).0	0.75	0.6536	0.0964	80.00
18	(42).0	0.64	0.6531	-0.0131	80.00
19	(46).0	0.68	0.7059	-0.0259	218.67
20	(47).0	0.53	0.7621	-0.232	90.48
21	(52).0	0.65	0.7383	-0.0883	85.98
22	(53).0	0.69	0.6803	0.00965	81.12
23	(55).0	0.73	0.7886	-0.0586	44.44
24	(57).0	0.7	0.7792	-0.0792	51.05
25	(60).0	0.71	0.8482	-0.138	60.19
26	(62).0	0.83	0.8445	-0.0145	67.74
28	(64).0	0.97	0.8801	0.0899	26.64
29	(65).0	0.61	0.6602	-0.0502	79.45
30	(73).0	0.73	0.6823	0.0477	88.12

	File Name	True	Prediction	Difference	Component Value Density
32	(79).0	0.58	0.4883	0.0917	42.80
33	(80).0	0.76	0.6382	0.122	62.88
34	(82).0	0.86	0.6673	0.193	66.67
35	(83).0	0.94	0.8438	0.0962	69.53
36	(84).0	0.52	0.6249	-0.105	57.14
37	(89).0	0.7	0.6084	0.0916	63.39
38	(90).0	0.72	0.6966	0.0234	237.25
39	(92).0	0.24	0.1781	0.0619	29.90
40	(96).0	0.56	0.6012	-0.0412	51.61
42	(102).0	0.35	0.3894	-0.0394	23.69
43	(103).0	0.32	0.3451	-0.0251	44.44
44	(106).0	0.05	0.3304	-0.28	50.00
45	(107).0	0.15	0.03111	0.119	8.41
46	(110).0	0.29	0.23	0.06	36.36
47	(113).0	0.19	0.2289	-0.0389	36.36
49	(115).0	0.27	0.3092	-0.0392	40.00
50	(117).0	0.21	0.2993	-0.0893	40.00
51	(119).0	0.34	0.302	0.038	40.00

Appendix C: Statistical analysis script from R software

```
###Multiple correlations between soil properties, wheat yield and the mean NIR absorbance
###Importing the data
Multiple_Correlations_data1 <- read.csv("Research/Data/Multiple Correlations_data1.csv")

dim(Multiple_Correlations_data1)
str(Multiple_Correlations_data1)

cor(Multiple_Correlations_data1)
round(cor(Multiple_Correlations_data1),digits = 3)

cor(Multiple_Correlations_data1,method="peason")
##############################################################################################
#### Changes for each field observations

##Importing the data
Field_Changes_Data <- read.csv("Research/Combining the soil spectra/Field Changes Data.csv")
str(Field_Changes_Data)
library(psych)
library(modeest)
library(RcmdrMisc)
library(ggplot2)

##paired t-test
t.test(Field_Changes_Data$NIR.Abs.2015,Field_Changes_Data$ NIR.Abs.2019,
      alternative='greater',conf.level=.95, paired=T)

##Calculating the differences
Field_Changes_Data$Dif <- Field_Changes_Data$NIR.Abs.2015 - Field_Changes_Data$NIR.Abs.2019

#Checking the normality of the differences
shapiro.test(Field_Changes_Data$Dif)

#one-sample t-test for the calculated differences in the paired value
t.test(Field_Changes_Data$Dif,alternative='greater',conf.level=.95, paired=F)

Developing prediction models using Partial least squares regression
```

```
ModelSOC <- lm(SOC ~ H + V + C + NIR, data = Data_PLSR)
summary(ModelSOC)
## From the summary output, many of these variables are not significantly different, overall R2 for all regressors=0.2491
library(pls)
library(plsr)
DataTrain <- Data_PLSR[1:41,]
DataTest <- Data_PLSR[42:51,]
#Ftting a PLSR model
ModelSOC <- plsr(SOC ~ H + V + C + NIR, data = Data_PLSR, validation = "LOO")

## an overview of the fit and validation results with the summary function:
summary(ModelSOC)
# judging the RMSEPs by plotting them:
plot(RMSEP(ModelSOC), legendpos = "topright")

##) one component seem to be enough. This gives an RMSEP of 0.6

## Prediction plot
plot(ModelSOC, ncomp = 1, asp = 1, line = TRUE)

## Extracting the explained variances explicitly with:
explvar(ModelSOC)
plot(ModelSOC, plottype = "scores", comps = 1:3)
plot(ModelSOC, "loadings", comps = 1:2, legendpos = "topright")
predict(ModelSOC, ncomp = 1, newdata = DataTest)
```

Appendix D: The sequence of yield maps for the 4 years of the experimental trial

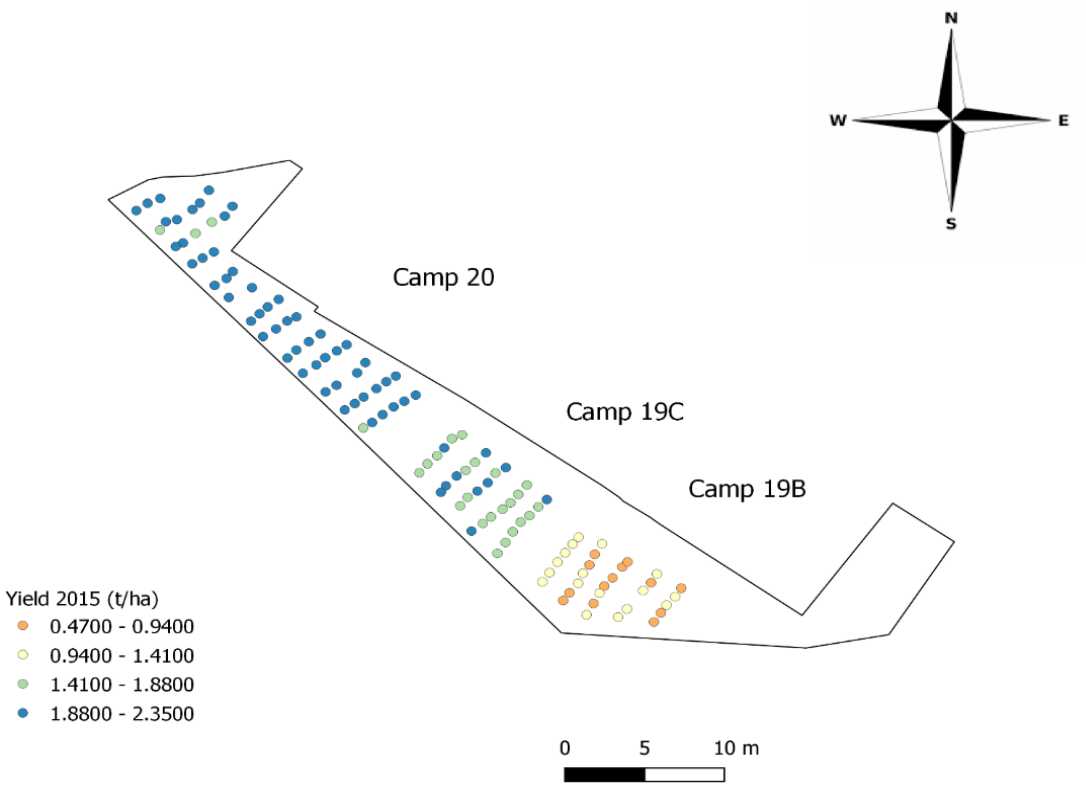


Figure D.1: Yield map for the year 2015 showing wheat yield per plot in each sampling point per camp of the field trial area.

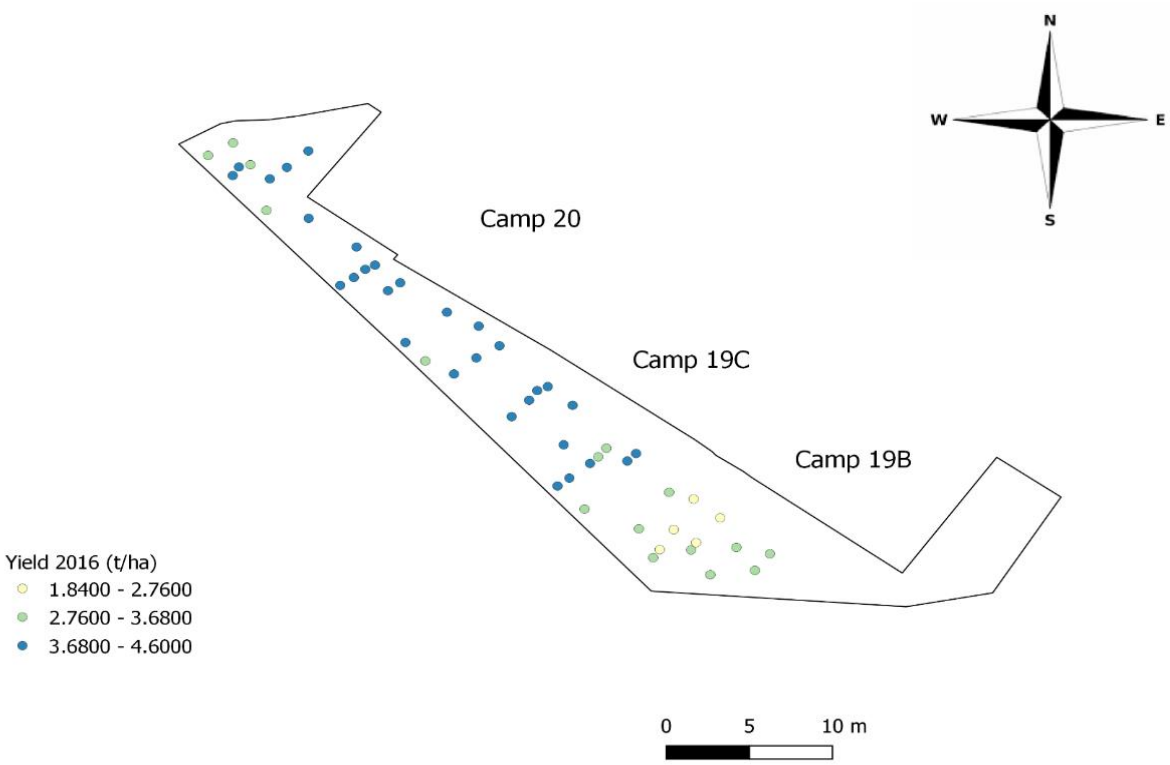
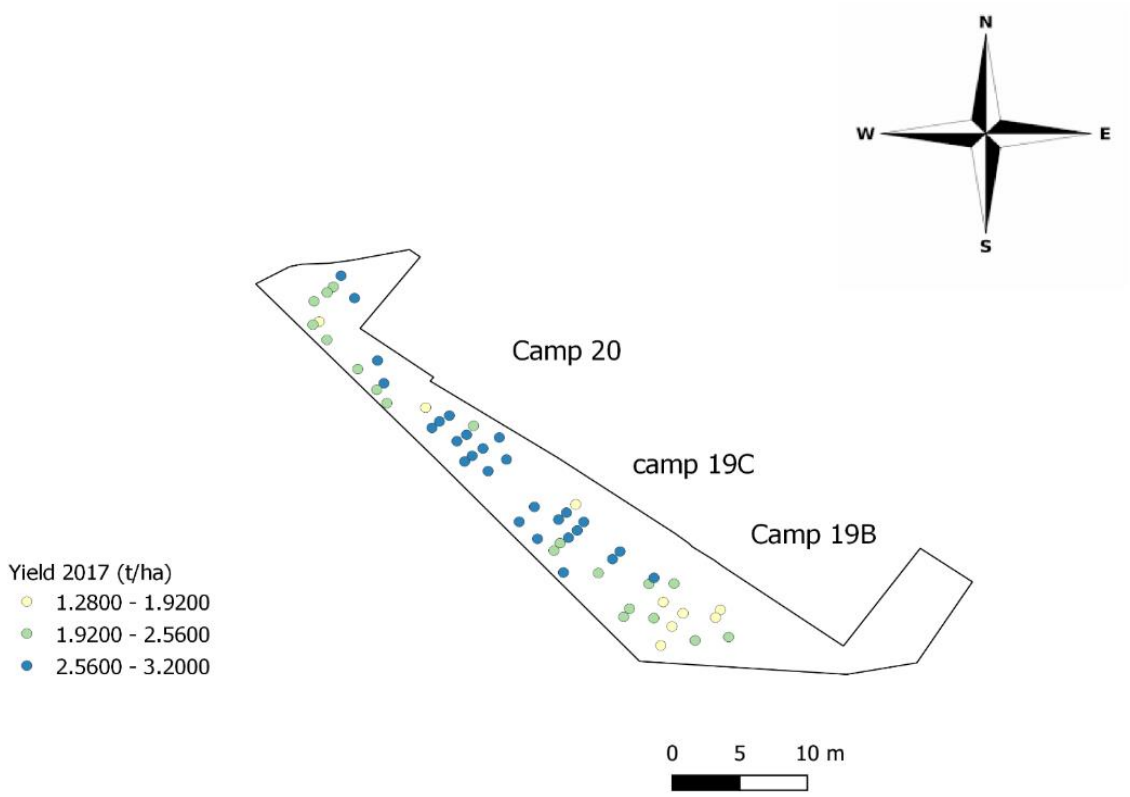


Figure D.2: Yield map for the year 2016 showing wheat yield per plot in each sampling point per camp of the field trial area.



FigureD.3: Yield map for the year 2017 showing wheat yield per plot in each sampling point per camp of the field trial area.

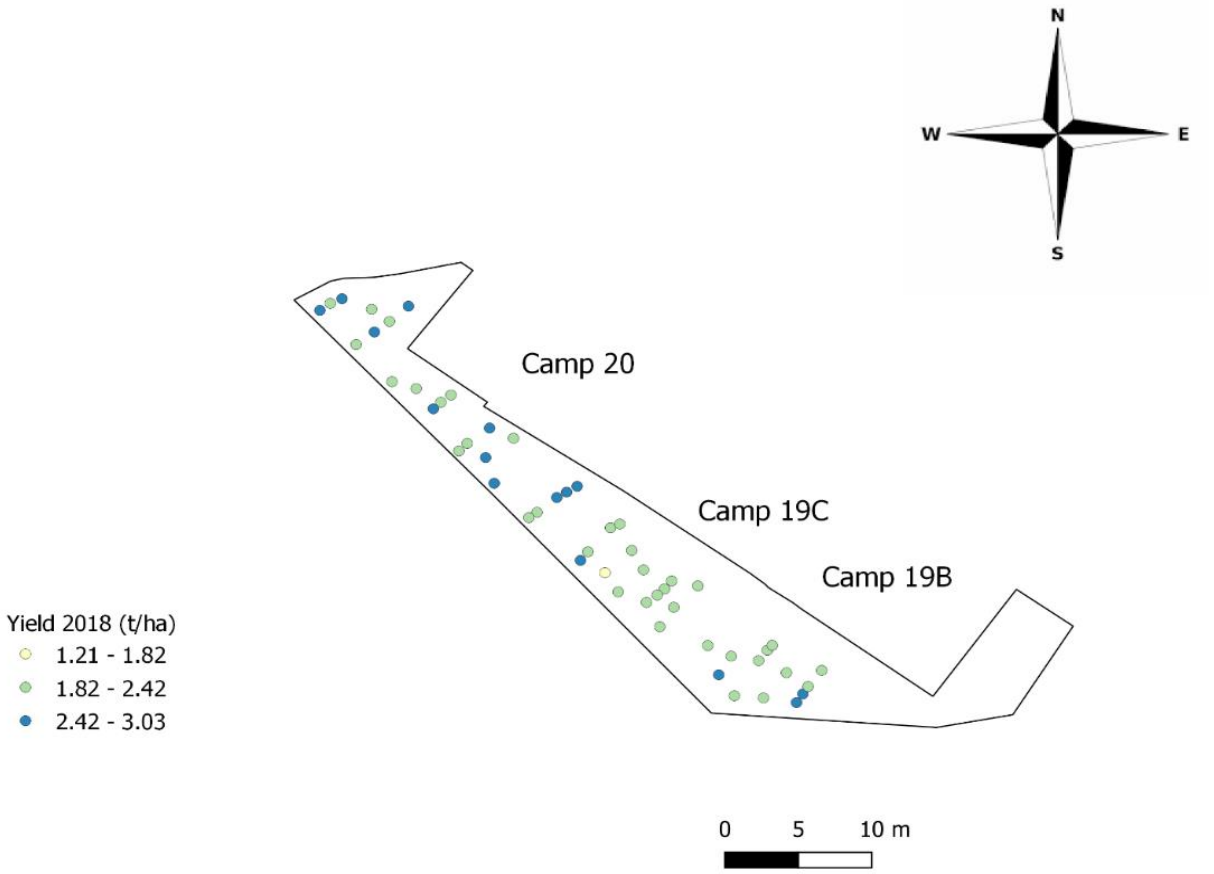


Figure D.4: Yield map for the year 2018 showing wheat yield per plot in each sampling point per camp of the field trial area.