Refining the methods for determining soil organic carbon stocks in South Africa

WH Cloete

orcid.org/ 0000-0002-7295-7181

Thesis accepted in fulfilment of the requirements for the degree *Doctor of Philosophy in Science with Environmental Sciences* at the North-West University

Promoter: Prof GM van Zijl

Co-promoter: Prof GC du Preez

Graduation: July 2025

DECLARATION

I confirm that this thesis, Refining the methods for determining soil organic carbon stocks in South Africa, presented for the *Doctor of Philosophy in Science with Environmental Sciences* degree is entirely my own, original work and therefore has not been submitted to any other academic institution.

Furthermore, I acknowledge and agree that the North-West University possesses exclusive publication rights to this thesis.

.....

W.H. Cloete

22 March 2025

ACKNOWLEDGEMENTS

The author would like to thank the following people for their contributions and support to this thesis.

Prof George van Zijl and Prof Gerhard du Preez for their guidance, determination and support towards the completion of this thesis.

Joaquim “Tiny” Geraldes, Charl Strydom and Edrich Louw for their friendship and assistance in fieldwork and in the soil science laboratory (North-West University, Potchefstroom Campus).

My colleagues, Anru Kock and Molebaleng Sehlapelo for their friendship and support during the completion of this thesis.

Cleaning ladies, Thabang and Baby for their friendship and laughter helping me thank God for my blessings.

My Bible study friends, Pieter Venter, Enette Venter, Laura Venter, Tyron de Bruyn, Mauritz Kotze, Niné Jansen van Vuuren, Richter Jansen van Vuuren, Shinji “muscles” Nakatani and JP Retief for their support and belief encouraged me to complete this thesis.

My mother (Magdeleen Cloete), father (Petrus Cloete), grandmother (Lenie Swanepoel) and sister (Cari Cloete) whose love, support and belief encouraged me to complete this thesis.

My best friends, Keegan Davel, Schalk Buys, Wihan de Beer, Job Vorster and Jay Moore for their unconditional friendship and support (even if they had no idea what I was doing other than looking at dirt).

My fiancée, Anri “Gogga” Venter for her care, providing a stress-free relationship and unconditional love and friendship.

Finally, my Lord and Saviour, Father God, His Son Jesus Christ and the Holy Spirit for getting the opportunity to start a PhD thesis, never mind completing the thesis.

God is goed. God is good. Modimo omogolo.

PREFACE

This thesis is presented as a compilation of 7 chapters, where Chapters 3 – 6 have been presented as separate journal publications and have all been submitted and are currently either published or under revision in peer-reviewed journals at date of submission. There is some repetition due to the thesis being presented as publications, however, each publication has been modified to reduce the repetition without losing relevant information.

The first chapter provides a short introduction regarding soil organic carbon, the gaps that this study tried to address and the aim and objectives of the research. Chapter 2 provides a more detailed description of (i) soil organic carbon content and its importance in agriculture, (ii) the agricultural management practices that lead to soil organic carbon sequestration, (iii) the analytical methods used to determine soil organic carbon (including near-infrared spectroscopy), (iv) the methods used to map soil organic carbon (including ordinary kriging and digital soil mapping) and finally, (v) the designs and methods used for sampling, calculating and modelling soil organic carbon stocks.

Chapter 3 focussed on determining which analytical method should be used for determining soil organic carbon content for the assessment of carbon credits in South Africa. This research has been published in *Geoderma Regional* (https://www.sciencedirect.com/science/article/pii/S235200942500032X) and was also presented at the Combined Congress in Pretoria in January 2023. Chapter 4 focussed on creating a near-infrared spectroscopy calibration algorithm of soil organic carbon content that could be used for South Africa. This research was also presented at the Combined Congress in Wilderness in January 2024 and has been published in *Soil Advances* (https://doi.org/10.1016/j.soilad.2025.100039). Chapter 5 aimed to determine whether conventional mapping using ordinary kriging or digital soil mapping with machine learning should be used for mapping soil organic carbon stocks at field scale in South Africa. This research is currently in press (accepted with minor revision) in *Soil Advances*. Chapter 6 focussed on comparing two soil organic carbon stocks quantification approaches, modelling and re-measure, at field scale in South Africa. This research is also currently at the editor in *South African Journal of Plant and Soil*. The final chapter gives a general conclusion of all the chapters and highlights future research that is needed.

ABSTRACT

The aim of this research was to refine the methods for determining soil organic carbon stocks (SOCS) in South Africa. There were a few gaps that this study tried to address that the Verified Carbon Standard methodology, and the Gold Standard methodology are not being clear on or specific about. These gaps have a direct impact on the South African industry and includes: (i) the uncertainty regarding which analytical method should be used for determining soil organic carbon (SOC) content in South Africa, (ii) the unavailability of a near-infrared (NIR) calibration algorithm for SOC content in South Africa, (iii) the uncertainty on how to map SOCS at field scale in South Africa, and (iv) the uncertainty of using either a modelling or re-measuring approach be for quantifying SOCS at field scale in South Africa.

Key objectives of this study included to: (i) evaluate the preferred method to be used for determining SOC content and the application for South Africa conditions, (ii) develop a NIR calibration algorithm for SOC content in South Africa on three different scales, (iii) determine how to map SOCS specifically at field scale in South Africa, and finally (iv) compare SOCS quantification approaches specifically at field scale in South Africa. The following paragraphs indicate the methodology for achieving the objectives, as well as the main findings and suggestions for the way forward.

Accurate quantification of SOC content is essential for the assessment of carbon credits. In South Africa, the standard methodologies for carbon credit assessment does not specify which analytical method should be used for determining SOC content. The study aimed to determine which analytical method should be used for determining SOC content for the assessment of carbon credits. Secondly, it determined whether pedotransfer functions could be used for transferring SOC content values between methods, especially for transferring historic determined SOC content to more recently determined values using more modern methods. Two-hundred-and-twenty topsoil (0–30 cm) samples were collected from five catchments in South Africa and analysed for SOC content with the three analytical methods: Walkley-Black wet-oxidation (WB), total dry combustion (TDC) and loss-on-ignition (LOI). The study found that the TDC method provided the most reliable SOC content values and should still be considered the preferred method for determining SOC content for the assessment of carbon credits in South Africa. The WB method should be avoided if a soil is expected to have a high SOC content, while the LOI method could still be used for determining SOM, however, this method should be avoided when determining SOC content. The study also reached the second aim by successfully creating pedotransfer functions between all three methods. However, only the WB and TDC methods had a very strong relationship ($R^2 = 0.91$) and showed that accuracy start to decrease significantly

above 2.5% SOC content. Therefore, the pedotransfer function (SOC_{WB} = -0.157 + 0.895 x SOC_{TDC} – 0.0149 x SOC_{TDC}^2 – 0.000606 x SOC_{TDC}^3) could be used for transferring SOC content values with SOC content up to 2.5%.

Near-infrared (NIR) spectroscopy has emerged as an easy, rapid and cost-effective alternative for SOC analysis and the accounting of carbon credits. South Africa currently lacks a calibration algorithm for predicting SOC content from NIR spectroscopy. This study aimed to develop a NIR spectroscopy calibration algorithm for SOC content, specific to South Africa. Soil samples were collected from 2 fields and 3 catchments across South Africa. These samples were analysed using the total dry combustion (TDC) method and scanned with a NIR spectrometer. Sixty NIR calibration algorithms were developed on a regional scale. The impact of methodological parameters, such as sample state, sampling design, processing and machine learning models, on the root mean square error (RMSE) of the validation statistics was also assessed. Although 60 regional-scale calibration algorithms were developed, none were suitable (RMSE = 0.39 % and RPIQ > 2) for SOC content prediction, which was attributed to the small sample size (n = 238). However, local calibration models for the Tsitsa catchment and Ottosdal fields presented great accuracy (RMSE < 0.1 and RPIQ > 1.5) that can be used for future SOC content prediction. The study found that the open spectral library global prediction model poorly predicted SOC content using local data (RMSE = 1.23 % and R^2 = - 0.83). This was attributed to South African samples being underrepresented in the global dataset. Sample state and sampling design were the most influential parameters affecting RMSE. To develop a national calibration algorithm, effort should be placed on developing accurate calibration algorithms for smaller areas that could be added to the national spectral library.

Reliable SOCS maps are important for accurately assessing carbon credits. To address this gap, the study aimed to determine whether conventional mapping using ordinary kriging (OK) or digital soil mapping (DSM) with machine learning (ML) should be used for mapping SOCS at field scale in South Africa. Fifty samples from three depths, 0-5, 5-15 and 15-30 cm were collected from two farms in South Africa and analysed for SOC content using the TDC method, while for dry bulk density (ρb) a pedo-transfer was created for predicting the ρb at unsampled locations. Maps for SOC content, ρb and SOCS were created using OK and DSM with ML and assessed with evaluation statistics including root mean square error (RMSE) and Lin's concordance correlation coefficient (ρc). A calculate-first approach and map-first approach were used for calculating the total SOCS for each study site. Findings of this study demonstrated that DSM with ML should be used rather than OK for mapping SOCS at field scale in South Africa. Utilising DSM with ML will lead to more accurate and reliable SOCS maps. However, the results also indicated that the SOC content, ρb and SOCS maps for DSM with ML need to be improved. To increase the accuracy of maps generated using DSM with ML, additional soil samples could be collected, and field scale

covariates should be incorporated. The results were inconclusive regarding whether the calculate-first approach or the map-first approach should be used when utilising OK or DSM with ML. Although both approaches have several advantages and disadvantages, more research is needed to determine which approach should be used when utilising OK or DSM with ML. However, the study concluded that the calculate-first approach should be endorsed, due to this approach being simpler and might lead to more reliable and accurate SOCS estimations.

There is an ongoing debate about the most appropriate approach for quantifying changes in soil organic carbon stocks SOCS. This study aimed to compare two SOCS quantification approaches, modelling and re-measure, on field scale in South Africa. The baseline assessment (2022) for two fields in South Africa was established beforehand and afterwards, fifty samples from three depth increments, 0-5, 5-15 and 15-30 cm were collected, and the SOCS calculated for two study sites (2024) using the re-measure approach and the modelling approach using the Rothamsted carbon (RothC) model. The sampling cost was also estimated against the value of sequestered SOCS (2022-2024) and compared between the two approaches. The main finding from the study indicated that it was inconclusive if a modelling or re-measure approach should be used for quantifying SOCS on field scale in South Africa. The spatial variation of SOC was the main factor influencing the total SOCS, sequestered SOCS and carbon credit values. The results indicated that the modelling approach using the RothC model slightly underestimated the total SOCS and sequestered SOCS when utilising all 50 observations for both study sites. The industry-represented, “One-observation”, did not have a consistent trend and showed overprediction and underprediction for both study sites using the different approaches. This had significant financial implications using both approaches. The study also found that the spatial variation of the SOC must be taken into account at the onset of measuring the SOC content in year one, independent of whether the SOCS would be modelled or re-measured. The carbon sequestration industry must take the model limitations (such as field scale dynamics and sensitivity to short-term changes) into account, and find ways, such as an improved sampling design to address these accuracies of RothC at field scale.

Key words: Soil organic carbon, carbon credits, near-infrared spectroscopy, RothC, digital soil mapping, satellite imagery, machine learning.

ABBREVIATIONS

ANNs: Artificial neural networks

BIO: Microbial biomass

cLHS: Conditioned Latin Hypercube sampling

DPM: Decomposable plant material

DSM: Digital soil mapping

HUM: Humified organic matter

IOM: Inert organic matter

ML: Machine learning

OSSL: Open Source Spectral Library

ρb: Dry bulk density

PTF: Pedo-transfer function

ρc: Lin’s concordance correlation coefficient

RA: Regenerative agriculture

RF: Random Forest

RMSE: Root means square error

RothC: Rothamsted Carbon

RPD: ratio of performance to deviation

RPIQ: ratio of performance to inter-quartile distance

RPM: Resistant plant material

RS: Random sampling

SIC: Soil inorganic carbon

SOC: Soil organic carbon

SOCS: Soil organic carbon stocks

SOM: Soil organic matter

SVM: Support Vector Machines

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

CHAPTER 1 INTRODUCTION

1.1 Background and Problem Statement

The increased atmospheric carbon dioxide and other greenhouse gas emissions are a pressing local and global issue, leading to climate change that threatens environmental and social health (Lal, 2009; Zhou et al., 2019). Carbon emissions, including carbon dioxide and methane, can be decreased by the sequestration of carbon, which refers to the long-term storage of carbon in oceans, soils, vegetation, and geological formations (Lal, 2007). The potential reduction of carbon emissions through the sequestration of carbon in the soil has shown to be a viable option in recent years (Nair, 2012; Rhodes, 2017; Baurov, 2021; Zhang et al., 2021).

Local land managers or farmers have the unique opportunity to sequester carbon and sell units of sequestered carbon (also called carbon credits) to the government or other businesses to generate an income (Sharma et al., 2021). The selling of these carbon credits is becoming increasingly sought after in South Africa. Accurate quantification of soil organic carbon (SOC) content is essential to account for these carbon credits. Currently, South Africa predominantly uses the Verified Carbon Standard methodology (Shoch & Swails, 2020) and the Gold Standard (Aynekulu et al., 2011; TREES Consulting, 2020) as frameworks for determining SOC content. These two frameworks are also used for mapping and modelling SOC content and soil organic carbon stocks (SOCS).

There are a few gaps that this study tried to address that the Verified Carbon Standard methodology, and the Gold Standard methodology are not being clear on or specific about. These gaps include:

- the measurement of SOC content,
- predicting SOC content with near-infrared spectroscopy,
- mapping of SOCS and finally,
- the re-measure and modelling approaches of SOCS.

Firstly, the Verified Carbon Standard does not specify which analytical method to determine SOC content, while the Gold Standard states that only the total dry combustion method may be used (Shoch & Swails, 2020; TREES Consulting, 2020). This uncertainty and variability in SOC content quantification introduces a level of risk, potentially affecting the credibility and reliability of carbon credit calculations. Therefore, to ensure credible carbon credits, SOC content must be measured using a single, dependable analytical method or a set of methods that can be consistently demonstrated to be interchangeable in their measurements.

Secondly, only two previous studies have focused on creating NIR spectroscopy calibration algorithms in South Africa (Nocita, 2009; Zahinda, 2020). Importantly, these algorithms have been made on small scales, are somewhat imprecise, and can only be used to predict SOC content for that specific area. While global studies have been conducted, the need for these studies to be specific to South Africa arises from the unique local conditions, such as soil types and climate, which can significantly affect the accuracy and reliability of NIR spectroscopic calibration algorithms. A global initiative has led to the development of the Open Soil Spectral Library (OSSL), which has an estimation service for utilising calibrated prediction models from open-source NIR spectral data (Safanelli et al., 2023). However, recent studies have found that models from the OSSL underperform because they do not capture the complete representation of South African soil spectral properties (Kock et al., 2024). Therefore, there isn't a readily available NIR spectroscopy calibration algorithm that can be used to predict SOC content for South Africa.

Thirdly, greater precision in SOC maps at field scale is essential for refining agricultural management practices, assessing greenhouse gas mitigation strategies for agriculture, and calculating carbon credits. Although previous studies have utilised ordinary kriging (OK) and digital soil mapping (DSM) with machine learning (ML) for mapping SOCS (Wiese et al., 2016; Flynn et al., 2019; Venter et al., 2021; Flynn et al., 2022), South Africa does not currently have a standard and/or reliable method for accurately mapping SOCS at field scale (Aynekulu et al., 2011; TREES Consulting, 2020; Shoch & Swails, 2020). The unavailability of a standard mapping method leads to inconsistent and incomparable SOCS maps, which could lead to poor land management decisions. By having a standard method, the accuracy and consistency of mapping SOCS at field scale could be improved by reducing potential errors or biases that could arise from using different or untested techniques.

Fourthly, the measure and re-measure approach is not commonly used, as it takes time and is expensive. Therefore, South African companies commonly employ the measure and modelling approach using, for example, the Rothamsted Carbon (RothC) Model. However, this approach has been debated and could influence the reliability and accuracy of SOCS assessment. Therefore, there is a need to evaluate the most reliable approach for assessing SOCS in cultivated fields.

1.2 Aims and Objectives

This study aimed to refine the methods for determining SOCS in South Africa. Objectives of this study included to:

- Evaluate which analytical method should be used for determining SOC content in South Africa.

- Create a NIR calibration algorithm for SOC content in South Africa.
- Determine how to map SOCS at field scale in South Africa.
- Compare SOCS quantification approaches at field scale in South Africa.

1.3 Hypotheses of the study

Hypotheses of the study included:

- The study hypothesised that different methods work on different principles and would thus provide different SOC content values of the same samples. Secondly, it is hypothesised that transfer functions could be used for transferring SOC content values from one method to another.
- The study hypothesised that NIR calibration algorithms would perform better at local scales compared to regional scales. The study also hypothesised that a NIR spectrometer could be used as a field instrument.
- The study hypothesised that DSM with ML would create more accurate SOCS maps compared to OK.
- The study hypothesized that the re-measure approach would provide a more accurate assessment of SOCS compared to the modelling approach.

CHAPTER 2 LITERATURE REVIEW

2.1 Soil organic carbon content

2.1.1 General Importance

Soil carbon consists of two fractions: SOC and soil inorganic carbon (SIC). The SOC fraction is the primary carbon component of soil organic matter (SOM), which is defined as "*a heterogeneous pool of C comprised of diverse materials including fine fragments of litter, roots and soil fauna, microbial biomass C, products of microbial decay and other biotic processes (i.e., such as particulate organic matter), and simple compounds such as sugar and polysaccharides*" (FAO, 2019). SOC's quantity (content) is arguably the most critical soil indicator because of its central role in various soil functions (Stockmann et al., 2015). SOC is known to improve soil fertility and productivity and contribute to worldwide food security. Regarding soil fertility, SOC enhances nutrient availability, soil structure and physical properties, and biological soil health and acts as a buffer against harmful substances (Chan, 2008).

2.1.2 Soil organic carbon sequestration

The increased atmospheric carbon dioxide and other greenhouse gas emissions are a pressing local and global issue, leading to climate change that threatens environmental and social health (Lal, 2009; Zhou et al., 2019). Carbon emissions, including carbon dioxide and methane, can be decreased by the sequestration of carbon, which refers to the long-term storage of carbon in oceans, soils, vegetation, and geological formations (Lal, 2009). The potential reduction of carbon emissions through the sequestration of carbon in the soil has shown to be a viable option in recent years (Nair, 2012; Rhodes, 2017; Baurov, 2021; Zhang et al., 2021).

SOC is the largest component of the terrestrial carbon pool (1 600 gigatons) and approximately twice the amount of carbon in the atmosphere (750 gigatons) and in vegetation (610 gigatons) (Chan, 2008). The amount of SOC stored in the soil profile can be considerable. For example, if the soil has a SOC content of 1% within 0-30 cm depth, the amount of SOC stored over 1 hectare of land can potentially weigh about 42 tons (Chan, 2008). Grasslands and rangelands cover almost 68% of the total agricultural area globally and have contributed to SOC sequestration (FAO, 2019). Therefore, SOC sequestration is a necessary process that can reduce carbon emissions and help alleviate the problems of global warming and climate change.

2.2 Agricultural Management Practices

Agricultural management practices may significantly influence the capacity of the soil to sequester SOC. To obtain the greatest potential soil carbon sequestration, the established carbon sinks must be protected from disturbances over time, such as maintaining native plant cover or, most

importantly, continued best management practices (Chan, 2008). Agriculture management practices that enhance carbon sequestration, as well as increase carbon credits, include: (i) crop and residue management, such as erosion and irrigation control; (ii) conservation tillage, including no- or reduced tillage; (iii) pasture management, such as fertiliser management and improved grass species and legumes, and (iv) organic amendments, such as animal manure and recycled organics (Chan, 2008).

2.2.1 Crop and residue management

Crop and residue management practices are critical factors in increasing carbon sequestration. The use of cover crops and intercropping practices can also contribute to increasing SOC content levels and improving soil health, which can lead to more productive and resilient agricultural systems (Gzubuyuk et al., 2015; Dakhalla et al., 2016; Merante et al., 2017). Therefore, crop and residue management practices, such as selecting specific crops, using appropriate crop rotations, and returning residues to the soil can significantly impact increasing carbon sequestration. Nitrogen leaching can also be significantly reduced by cultivating specific crop combinations, such as barley with Italian ryegrass rather than pea and grasslands (Merante et al., 2017). In addition, residue management directly affects SOM and, therefore, soil properties such as pH, CEC, nutrient cycling, particle aggregation, moisture content, and soil temperature (Merante et al., 2017). For instance, returning residues to the soil, rather than removing them, can quickly increase SOC content levels, leading to significant carbon sequestration.

2.2.2 Tillage practices

When it comes to decreased soil disturbance, reduced tillage or zero-tillage is an increasingly applied conservation or regenerative practice. It consists of cultivating crops while adopting agricultural practices that leave the soil undisturbed. The only soil disturbance occurring under reduced or zero-tillage is related to the movement of the agrarian machines generally used for weeding, seeding, and harvest (Merante et al., 2017). Reduced tillage reduces soil disturbance by the shallower depth of tillage and reduced frequency of passes over soil or by tilling only specific field portions (strip or ridge tillage). This reduced soil disturbance brings different benefits for the soil, especially over long periods, including (i) increased macro-aggregate stability and water infiltration improvement (Franzluebbers, 2002; Merante et al., 2017), (ii) soil erosion reduction (de Freitas & Landers, 2014; Merante et al., 2017), (iii) increased soil aeration (Batey, 2009; Merante et al., 2017) and (iv) improved soil moisture and compaction reduction (Jin et al., 2011; Merante et al., 2017). Reduced or zero-tillage can protect organic material from high temperatures and reduce mineralization (Alvarez et al., 2014; Merante et al., 2017). Reduced or zero-tillage has also improved biodiversity by creating a more favourable habitat for soil organisms, including earthworms and other beneficial microbes. These practices can also reduce greenhouse gas emissions by minimizing soil disturbance and decreasing fuel use associated

with tillage operations (Merante et al., 2017). Adopting reduced or zero-tillage practices can also lead to increased carbon sequestration, translating into additional carbon credits for farmers participating in the carbon markets.

2.2.3 Pasture management

Pasture management practices, such as fertiliser management and improved grass species and legumes, can increase carbon sequestration in soil and vegetation. Fertiliser management involves carefully using nitrogen and other nutrients to optimise plant growth and minimise environmental nutrient loss (Singh et al., 2019). This can increase the amount of carbon stored in plant biomass and SOM. Additionally, improved grass species and legumes can increase productivity and improve soil health, increasing carbon sequestration (Yu et al., 2020). Regenerative agricultural practices, such as crop rotation, cover cropping and the use of livestock to graze and fertilise the land can increase the amount of carbon stored in soil and vegetation, increase water retention and reduce erosion (Singh et al., 2019).

Grazing intensity and grazing frequency has been shown to be an important factor affecting rangeland health and ultimately SOC. Continuous grazing could lead to excessive defoliation of preferred forage and lead to localised soil degradation and erosion (Teutscherova et al., 2021). A recent study has also found that intensive, short-duration rotational grazing could improve soil quality and ultimately increase SOC content (Teutscherova et al., 2021). The phenomenon, bush encroachment has also shown to be cause of severe degradation in South Africa (Turpie et al., 2019). Bush encroachment refers to *"the process whereby the cover of indigenous woody plants (trees and shrubs) in a grassy ecosystem (savanna or grassland) increases substantially relative to the indigenous woody cover of some (historical) reference state."* (Turpie et al., 2019). Although bush encroachment can lead to an increase in SOCS, as woody plants store more carbon than grasses, this phenomenon has shown to severely reduce the herbaceous plant biomass and density, leading to a decrease in SOCS, especially in the topsoil.

2.2.4 Organic amendments

Practices that include organic amendments, such as "green manuring" have been increasingly promoted to elevate SOC content levels (Lessmann et al., 2021). Organic amendments can improve soil fertility by increasing the availability of nutrients for plants and promoting microbial activity. Organic amendments can also enhance soil structure and porosity, improving soil water-holding capacity and reducing erosion (Kumar et al., 2019). Different types of organic amendments can have varying effects on SOC content levels. For example, a study by Singh et al. (2018) found that applying animal and green manure resulted in higher SOC content levels than crop residues. In addition to improving soil health, the use of organic amendments can also have a positive environmental impact by reducing the need for synthetic fertilisers. The use of

organic amendments can also contribute to reducing greenhouse gas emissions. For example, using livestock manure can decrease methane and nitrous oxide emissions by replacing synthetic fertilisers (Chen et al., 2019). Moreover, the type and quantity of organic amendments used can affect the microbial community composition and diversity in the soil, impacting nutrient cycling, SOM decomposition, and overall soil health. Finally, it is essential to note that the effectiveness of organic amendments in increasing SOC content levels can depend on various factors, such as soil type, climate, and the specific amendment used. Additionally, the increased SOC levels resulting from organic amendment practices can lead to carbon sequestration in the soil, which can contribute to generating carbon credits.

2.3 Methods for determining soil organic carbon content

In South Africa, there are currently three different analytical methods used for measuring SOC content, including the Walkley-Black wet-oxidation (WB) (Walkley & Black, 1934), total dry combustion (TDC) (Nelson & Sommers, 1996; FAO, 2019) and loss-on-ignition (LOI) methods (Nelson & Sommers, 1996; FAO, 2019). NIR spectroscopy is also a relatively new method for determining soil properties in South Africa (FAO, 2019; Seboko et al., 2023; Kock et al., 2024). However, this method first requires determined SOC content data from one of the abovementioned methods.

2.3.1 Walkley-Black wet-oxidation

SOC content analysis using the WB method (Walkley & Black, 1934) has been the most widely used method since its development (O'Rourke & Holden, 2011). This is because it is rapid and requires minimum equipment compared to other methods (Nelson & Sommers, 1996; Chen et al., 2015). The WB method measurements of SOC involve using dichromate ($Cr_2O_7^{2-}$) to oxidize SOC (wet oxidation), and iron sulphate ($FeSO_4$) to reduce excess $Cr_2O_7^{2-}$ in solution (Roper et al., 2019).

The WB method is also based on several assumptions (De Vos et al., 2007), which include: (i) that carbon is the chief reducing agent present in soils and (ii) that carbon, oxidisable hydrogen, and reducible nitrogen are present in SOM in about the same proportions (Walkley & Black, 1934). Concern has also been raised regarding the interference of other soil constituents, such as chloride, ferrous iron, and higher oxides of manganese that have been shown to undergo redox reactions in chromic acid mixtures, leading to incorrect values of SOC content (Walkley, 1947; De Vos, 2007). Other limitations of this method include incomplete oxidation, which leads to the underestimation of SOC content. SOC content values have been adjusted with a correlation factor of 1.47 to counter this underestimation (Meersmans et al., 2009; O'Rourke & Holden, 2011). Another concern in developing countries like South Africa is the high cost of imported chemicals.

Due to the hazardous nature of the reagents, disposing thereof (after laboratory analysis) could harm the environment or have an additional cost component.

2.3.2 Total dry combustion

The TDC method is considered the most reliable for measuring SOC content (Roper et al., 2019). The TDC method involves using a CN elemental analyser for measuring SOC content based on carbon dioxide released from thermally oxidised soil at very high temperatures (between 950 and 1 800 °C) (Nelson & Sommers, 1996; Abella & Zimmer, 2007; Roper et al., 2019; FAO, 2019). After combustion, the product is quantified by gas chromatography using thermal conductivity or a flame ionisation detector (FAO, 2019). For SOC content, the end-product is carbon dioxide, while for nitrogen, it is nitrogen gas or oxides of nitrogen (Thompson, 2008). If other elements, such as chlorine, are present, they are converted to hydrogen chloride. During the quantification process, the end-products are carried by inert carrier gas, such as helium, and passed over heated high-purity copper. Copper removes oxygen that is not consumed during the initial combustion process (Thompson, 2008).

The main advantage of this method is that it ensures complete combustion of all SOC content present in the sample and allows a relatively large number of samples to be processed per unit time (FAO, 2019). The most significant disadvantage is the high initial economic investment associated with the purchase of specific instruments, specifically in developing countries such as South Africa (Wang et al., 2012; Roper et al., 2019). Another disadvantage is the possible measurement of SIC content fraction with the SOC content, which leads to an overestimation of SOC content (FAO, 2019). This is especially true for arid Southern Africa, where SIC can contribute considerably to the total soil carbon.

2.3.3 Loss-on-ignition

The LOI method involves the thermal decomposition of SOM using a muffle furnace (Roper et al., 2019). The mass lost during thermal decomposition indicates the mass of SOM lost due to oxidation and volatilisation. A known sample weight is placed in a ceramic crucible (or similar vessel) and then heated to between 350 and 550 ºC for a specific time, usually 3 to 8 hours (Nelson & Sommers, 1996). Although there is no definite standard procedure, the method is based on gravimetric mass loss (Hoogsteen et al., 2015). The sample is then cooled in a desiccator and weighed. SOM is calculated as the difference between the initial and final sample weights divided by the initial sample weight multiplied by 100 (Roper et al., 2019).

One primary concern about this method is that some clay minerals will lose structural water (i.e., water that is part of the matrix) or hydroxyl groups at the temperatures used to combust the samples. This structural water loss will increase the total sample weight loss, leading to

overestimating the SOM content (Schumacher, 2002). Several authors (including De Vos et al., 2005; Hoogsteen et al., 2015) have proposed clay correction factors to account for water losses. Another factor that affects the accuracy of SOC content estimations is the presence of carbonates (i.e., $CaCO_3$ or $CaMg(CO_3)_2$), which decompose at ignition temperatures above 600 ºC (Kasozi et al., 2009; Hoogsteen et al., 2015).

Typical LOI procedures assume the mass lost during thermal oxidation is entirely comprised of SOM, with estimates of SOC content derived using adjustment factors (Chatterjee et al., 2009; Roper et al., 2019). Although widely used, the validity of the conventional LOI-to-SOC conversion factor of 0.58 (or 1.74) remains dubious (Pribyl, 2010; Jensen et al., 2018). The basic assumption is that LOI is due to the combustion of SOM and that the SOC content in SOM is constant at 58% (Christensen & Malmros, 1982; Jensen et al., 2018). However, the traditional conversion factor 0.58 for the LOI-to-SOC is considered too high. Although the accuracy of the LOI method remains dubious, some laboratories are transitioning from the WB method to the LOI method due to the human health problems and environmental conditions that the WB method is responsible for (Barwari et al., 2017).

2.3.4 Near-infrared diffuse reflectance spectroscopy

Research using NIR diffuse reflectance spectroscopy in soil science has increased rapidly in recent years (Wetterlind et al., 2008; Stenberg et al., 2010). Usually, the focus is on soil organic matter (SOM), texture, structure, and microbial activity in the soil (Stenberg et al., 2010). Reasons for interest in NIR include: (i) simple sample preparation, (ii) sample unaffected by analysis in any way, (iii) non-hazardous chemicals are required, (iv) measurement only takes a few seconds, (v) several soil properties can be estimated from a single scan, and (vi) the technique can be used both in the laboratory and in situ (Viscarra Rossel et al., 2006; Stenberg et al., 2010; Orr et al., 2017).

2.3.4.1 Fundamentals of soil near-infrared diffuse reflectance spectroscopy

Electromagnetic radiation is defined as waves of electromagnetic fields travelling through space with electromagnetic (EM) energy, which includes radio waves, microwaves, infrared, visible radiation (light), X-rays and gamma-rays (Atkins & De Paula, 2010; Kock, 2021). The infrared section of the EM spectrum ranges from 10 to 14 790 cm^{-1} and is divided into three regions, namely near- (NIR) (14 290 − 4 000 cm^{-1}), mid- (MIR) (4 000 − 400 cm^{-1}) and far infrared (FIR) (400 – 10 cm^{-1}). NIR is conventionally specified in wavelength in nanometer (nm, 10^{-9}) or micrometer (µm). In contrast, MIR is conventionally specified in wavenumber (cm^{-1}), which means how many EM waves fit into a length of one centimeter (FAO, 2022).

To generate a soil spectrum, radiation containing all relevant frequencies in the range is directed to the soil sample (Figure 2-1, Stenberg et al., 2010). The radiation or EM energy can be absorbed, transmitted, or deflected (Figure 2-1, FAO, 2022). In soil spectroscopy, the reflected energy from the soil surface is of most interest. There are two types of reflection: (i) specular (mirror) reflection and (ii) diffuse reflection. Diffuse reflectance penetrates and sufficiently interacts with the soil matrix, therefore containing more information regarding the soil constituents (FAO, 2022). The energy quantum is directly related to frequency (inversely related to wavelength); thus, the absorption spectrum produces a characteristic shape, which can be used for analytical purposes (Stenberg et al., 2010). Soil spectra can be represented in either reflectance or absorbance, therefore, practitioners should be careful about whether they are working with reflectance or absorbance spectra (FAO, 2022). Ultimately, the frequencies at which light is absorbed appear as a reduced signal of reflected radiation and are displayed in % reflectance (R), which can then be transformed to apparent absorbance: $A = \log(1/R)$ (Stenberg et al., 2010). Chemical and environmental factors influence the wavelength at which absorption takes place, such as neighbouring functional groups and temperature, allowing for the detection of a range of molecules which may contain the same type of bonds.

Another essential part of soil spectroscopy is the modes of interaction between the EM energy and soil. Depending on the constituents present in the soil the EM energy or radiation will cause individual molecular bonds to vibrate, either by bending or stretching, and they will absorb light to various degrees, with a specific energy quantum corresponding to the difference between two energy levels (Stenberg et al., 2010; FAO, 2022). MIR energies cause molecular vibrations of chemical bonds commonly present in soils' organic and mineral compounds. The chemical bonds in organic compounds include O-H, N-H, C-H, C-C, C=C, C-N, C=O, etc. (FAO, 2022). Notably, the overtones and combinations of these vibrational absorption bangs appear in the NIR region. The NIR region is characterised by broad, superimposed, and weak vibrational modes, giving soil NIR spectra few, broad absorption features (Stenberg et al., 2010).

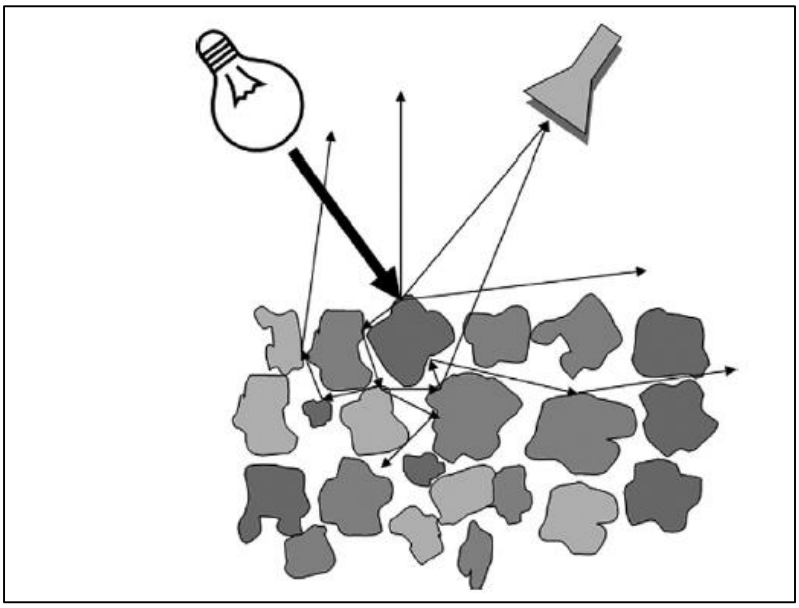


Figure 2-1: Illustration of diffuse reflectance spectroscopy, with incoming light from the left scattered, absorbed, and reflected (from Stenberg et al., 2010).

2.3.4.2 Near-infrared spectroscopy and soil organic carbon content

SOC content produces a spectral signature, combination bands, and overtones of fundamental spectral features detected in the NIR region (Cambule et al., 2012). These overtones and combination bands in the NIR region are due to the stretching and bending of N-H, C-H, and C-O groups (Ben-Dor et al., 1999; Stenberg et al., 2010). Bands around 1100, 1600, 1700 to 1800, 2000 and 2200 to 2400 nm have been identified as being particularly important for SOC content calibration (Krishnan et al., 1980; Dalal & Henry, 1986; Morra et al., 1991; Henderson et al., 1992; Ben-Dor and Banin, 1995; Malley et al., 2000; Martin et al., 2002; Stenberg, 2010; Stenberg et al., 2010).

In addition to the specific NIR spectral bands mentioned, it is worth noting that the shape and position of the NIR spectrum can also provide information about SOC content (Cambule et al., 2012). The NIR spectral signature of SOC content is influenced by soil type, texture, and moisture content (Ben-Dor et al., 1999; Stenberg et al., 2010). Therefore, it is crucial to consider these factors when using NIR spectroscopy for SOC content analysis. Furthermore, the relationship between NIR spectra and SOC content may vary depending on the calibration dataset and statistical model used (Cambule et al., 2012). Therefore, careful calibration and validation of NIR models are necessary to ensure accurate SOC predictions.

2.3.4.3 Creating Calibration Algorithms Procedure

The procedure for NIR soil spectroscopic analysis is as follows (FAO, 2022): To predict soil properties, such as SOC content, using NIR spectroscopy, (i) a sample set needs to be

established, also known as a soil database. The soil database must include soils covering the entire target (study) area to account for soil variability. (ii) the spectral data (acquired by scanning samples using a NIR spectrometer) is combined with the soil property data to build the soil spectral library.

(iii) specific or various calibration models are then employed, where the models are tested, and the accuracy is validated using independent (validation) sample sets. Additionally, cross-validation techniques, spectral pre-processing, and accuracy assessment indicators can be used to ensure robust model performance and reliable SOC estimations (Terhoeven-Urselmans et al., 2010; Kock, 2021). Wadoux et al. (2021) also stress the importance of sampling designs for improving the soil spectral library. These sampling designs include simple random sampling (RS), conditioned Latin Hypercube Sampling (cLHS) and K-Means clustering sampling.

2.3.4.4 Calibration models

Different models are used for SOC content calibration and prediction, including Partial Least Squares Regression (PLSR), Random Forest (RF), Support Vector Machines (SVM), Cubist, and Artificial Neural Networks (ANN).

Previous studies have mostly applied the PLSR model, focused on a regional scale, and used the Vis-NIR region to create calibration algorithms for SOC content (Table 2-1).

Table 2-1: Previous studies creating NIR calibration algorithms for soil organic carbon content.

Authors	Country	Calibration model	Scale	Area size (km²)	*n*	Spectral range	Sampling depth (cm)	SOC content range (%)	Results	
									RMSE (%)	RPD
Nocita (2009)	South Africa	PLSR	Regional and Field	Unknown	113	Vis-NIR	20	0.18-6.03	0.36	2.27
Viscarra-Rossel et al. (2011)	Australia	PCA	National	7.68 million	4 030	Vis-NIR	100	Unknown	n.d.	n.d.
Nocita et al (2012)	Luxembourg	PLSR & PCA	Regional	420	107	Vis-NIR	5	0.9 - 5.02	0.35	2.6
McDowell et al. (2012)	Hawaii Islands	PLSR & RF	National	Unknown	305	Vis-NIR	Unknown	0.79 - 54.66	2.8	4.25
Cambule et al. (2012)	South Africa	PLSR	Regional	10 000	412	NIR	5	0.4 - 1.2	0.28	2.12
Deng et al. (2013)	Denmark	PLSR	National	43 098	2 072	Vis-NIR	50	0 - 6	0.4	2.1
Camacho-Tamayo et al. (2013)	Columbia	PLSR	Regional	51	1 240	NIR	Unknown	1.19 - 2.55	0.19	2.27
Tekin et al. (2014)	Turkey and UK	PLSR & ANN	Regional	Unknown	270	Vis-NIR	20	0.16 - 14.2	0.82	4.23
Knadel et al. (2015)	Denmark	PLSR	Field	0.264	30	Vis-NIR	Unknown	1.4 - 3.3	0.26	2.1
Clairotte et al. (2016)	France	PLSR	National	552 000	3 800	Vis-NIR and NIR		0.06 - 17.7	0.89 and 0.83	2.1 and 2.2
Kusumo et al. (2018)	Indonesia	PLSR	Regional	Unknown	305	Vis-NIR	10	0.39 - 2.28	0.15	2.01
Seidel et al. (2019)	Germany	PLSR	Field, Regional, and National	Unknown	~ 20 000	Vis-NIR	60	0.43 - 9.95	Field = 0.08; Reg = 0.16; National = 0.21	Field = 5.84; Regional =2.78; National =2.3
Zahinda (2020)	South Africa	PLSR & SVM	Regional	169 000	200	Vis-NIR	30	0.3 - 3.6	0.24	2.4
Li et al. (2020)	China	PLSR	Regional	Unknown	714	NIR	20	0.66 - 4.63	0.25	2.05
George & Kumar (2020)	India	ANN	Catchment	10	75	Vis-NIR	15	0.26 - 2.71	0.24	n.d.
Biney et al. (2020)	Czech Republic	PLSR & SVM	Field	0.22	130	Vis-NIR and NIR	20	0.6 - 2.93	0.25 for both	n.d.
Xu et al. (2021)	China	PLSR & SVM	Regional	Unknown	328	Vis-NIR	20	0.45 - 4.33	0.36	2.81
Gholizadeh et al. (2021)	Czech Republic	RF	Field	0.16	143	Vis-NIR	20	0.62 - 2.74	0.22	n.d.
Bai et al. (2022)	China	PLSR, RF & CNN	Catchment	6 256	330	Vis-NIR	20	0.09 - 2.5	0.1	3.18
Zayani et al (2023)	France	PLSR	Catchment	12	198	Vis-NIR	25	1.46 - 7.54	0.33	2.01
Singha et al. (2023)	India	PLSR & SVM	Regional	3	200	Vis-NIR	30	0.07 - 1.1	0.12	2.47
Long et al. (2023)	China	PLSR	Regional	Unknown	89	Vis-NIR	20	1.84 - 13.43	0.72	3.38
Peng et al. (2024)	China	PLSR & RF	Regional	55	50	NIR	100	0.52 - 8.49	0.75	n.d.

2.3.4.4.1 Partial Least Squares Regression

PLSR is an orthogonal data compression method that allows soil scientists to view two-dimensional spectra in multi-dimensional space (Wold, 1966; Janik et al., 2007; Nocita, 2009). The PLSR model works on finding the linear combination of input variables that best explains the variance in the output variable. In other words, the PLSR model combines a principal component analysis with a multiple regression (Wadoux et al., 2021).

It iteratively identifies latent variables that capture the maximum covariance between the input and output variables. The latent variables are created from the spectral data and soil property data, which are then divided into smaller sets called orthogonal loadings and scores. These loadings and scores are similar to the principal components of principle component analysis (PCA); however, they focus more on the soil property data than just the spectral data (Bramley & Janik, 2005; Kock, 2022). The PLSR model can be used in R (R Core Team, 2020) with the $plsr$ function in the pls package created by Mevik and Wehrens (2024). Many studies have used PLSR to create calibration algorithms for SOC content and have found great success (Table 2-1).

2.3.4.4.2 Random Forest

RF is a widely used (Table 2-1) non-linear model that operates by constructing a multitude of decision trees during training and outputting the mode of the classes (classification) or the mean prediction (regression) of the individual trees (Breiman, 2001). Data is split up in two thirds that are used to create or grow the regression tree while the rest is used for cross-validation in conjunction with the calibration step; these are known as out-of-bag samples and give an idea about how well the models perform (Shahrayini et al., 2020; Wadoux et al., 2021). When using RF for regression, a final prediction is determined by the average outputs for individual trees, while for classification, the trees are voted on by majority count (Wadoux et al., 2021). As mentioned, RF is also a widely used machine learning model for creating calibration algorithms for SOC content (Table 2-1). It is also optimised to work with large amounts of spectral data and is easy to use and tune for optimal results (Rossel and Behrens, 2010).

2.3.4.4.3 Support Vector Machines

The SVM model is based on the statistical learning theory and aims to find the hyperplane that best separates the data points while maximising the margin (Cortes & Vapnik, 1995; Vohland et al., 2011; Vapnik, 2000; Zahinda, 2020). An essential property of this model is that its solution depends only on a subset of training samples called support vectors (Xu et al., 2018; Zahinda, 2020).

2.3.4.4.4 Cubist

The Cubist model is an extension of decision trees. It builds a series of tree-based models and combines their predictions using a weighted average to provide more accurate and interpretable results (Quinlan, 1992). Cubist is a data mining tool developed from early model trees C4.5 and M5 (Peng et al., 2015). The Cubist model transforms regression into a classification problem and consists of a set of rules, and each rule consists of a linear model (Wadoux et al., 2021). Each rule has a 'condition' (i.e., reflectance at 880 nm <= 0.0144 & 1610 nm > -0.921); if the data meet this condition, the prediction is the given linear function. The program also provides statistics on each rule, referring to the predicted values range and the model's error.

The *cubist* package is used for utilising the model in R (Kuhn et al., 2024). Although the Cubist has not been commonly utilised for creating calibration algorithms for SOC content using NIR spectroscopy (Table 2-1), MIR studies have shown great promise (Minasny & McBratney, 2006; Dangal et al., 2019).

2.3.4.4.5 Artificial Neural Networks

Artificial Neural Networks (ANNs) consist of interconnected nodes that contain an activation function (Rosenblatt, 1958; Tekin et al., 2014). These nodes are also arranged in layers. They work by passing input signals through a network of interconnected nodes and adjusting the weights between nodes during training to minimise the difference between predicted and actual outputs.

2.3.4.5 Model Accuracy Assessment Indicators

Model validation is performed to assess the models created by calibration by comparing the predicted SOC content with the SOC content of the actual laboratory-measured values (Wadoux et al., 2021). There are also several indicators used in predictive soil spectroscopy (Wadoux et al., 2021), which include root mean square error (RMSE), square correlation coefficient (r^2), coefficient of determination (R^2), Lin's concordance correlation coefficient (ρ_c), the ratio of performance to deviation (RPD) and the ratio of performance to inter-quartile distance (RPIQ) (Terhoeven-Urselmans et al., 2010; Kock, 2021). Another important aspect of model validation is assessing the model's predictive power for new or independent data (Wadoux et al., 2021). This is typically done by evaluating the model's performance on a separate dataset, known as the validation or test set, which was not used for model calibration or training. This helps ensure the model does not overfit the calibration data and can generalise well to new data. Additionally, cross-validation techniques such as k-fold, cross-validation or leave-one-out (LOO) cross-

validation can further assess model performance and improve model robustness (Stenberg et al., 2010; Kock, 2021).

2.3.4.5.1 Root Mean Square Error

The RMSE indicates the mean prediction error, measuring the deviation between the observed and predicted SOC content values (Viscarra-Rossel et al., 2006; Wadoux et al., 2021). The RMSE evaluates the dispersion of the residuals, which means that this statistic explains how concentrated the data are around the line of the best fit between the observed and predicted values. RMSE values are non-negative, where 0 indicates a perfect fit between observed and predicted values.

Equation 2-1 is used to calculate RMSE, where n is the sample size and $obs_{i,}$ and $pred_i$ are vectors of the measured and predicted SOC content, respectively (Wadoux et al., 2021).

Equation 2-1: $$\boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{obs_i} - \boldsymbol{pred_i})^2}$$

2.3.4.5.2 Square Correlation Coefficient

The square correlation coefficient or Pearson's squared correlation coefficient (r^2) is commonly used to assess the dispersion around the regression line. This means that r^2 represents the strength of the linear relationship between observed and predicted values concerning the fitted regression line (Wadoux et al., 2021).

Equation 2-2 is used to calculate r^2, where n is the sample size $obsi,$ and $predi$ are vectors of the measured and predicted SOC content, respectively, obs is the total measured SOC content (Lawrence & Lin, 1989; Wadoux et al., 2021).

Equation 2-2: $$\mathbf{r^2} = \mathbf{1} - \frac{\sum_{i=1}^{n}(\mathbf{obs_i} - \mathbf{pred_i})^2}{\sum_{i=1}^{n}(\mathbf{obs_i} - \mathbf{obs})^2}$$

2.3.4.5.3 Coefficient of Determination

The coefficient of determination (R^2) is the amount of variance the model explains. The R^2 quantifies the improvement made by the model by simply using the mean of the observations as a prediction (Peng et al., 2015; Wadoux et al., 2021).

Equation 2-3 shows the calculation of R^2, where RSS refers to the sum of squares of residuals (also called the sum of the squared error), and TSS refers to the total sum of squares.

Equation 2-3: $$\mathbf{R^2} = \mathbf{1} - \frac{\mathbf{RSS}}{\mathbf{TSS}}$$

2.3.4.5.4 Lin's Concordance Correlation Coefficient

Lin's concordance coefficient (ρ_c) measures the agreement between observed values and predicted values or values from another set concerning the 1:1 line (Lawrence & Lin, 1989). The values range from 0 to 1, and if predictions perfectly agree with the observations, all the points will fall on the 1:1 line.

Equation 2-4 shows the calculation of ρ_c, where *r* is Pearson's correlation coefficient, σ is the standard deviation ($r\sigma_{pred}\sigma_{obs}$) is the covariance between observed and predicted values) and μ is the mean.

$$\textbf{Equation 2-4: } \rho c = \frac{2r\sigma_{pred}\sigma_{obs}}{\sigma_{obs}^2+\sigma_{pred}^2+(\mu_{obs}-\mu_{pred})2}$$

2.3.4.5.5 Ratio of performance to deviation

The RPD is the ratio of standard error in prediction to the standard deviation (Williams & Thompson, 1978). The RPD aims to scale the error in prediction with the standard deviation of the soil property. Since RPD has been used for different products and thus divided into various categories, depending on those products, it is essential to note when to evaluate model performance. An example would be when evaluating the model performance, an RPD > 2 would indicate adequate model performance for a soil scientist. In contrast, only an RPD > 3 would be sufficient for a plant scientist.

Equation 2-5 shows the calculation of RPD, where n is the sample size $obsi,$ and $predi$ are vectors of the measured and predicted SOC content, respectively. obs is the total measured SOC content.

$$\textbf{Equation 2-5: } \mathbf{RPD} = \frac{\sqrt{\frac{1}{n-1}}\sum_{i=1}^{n}(obs_i-obs)^2}{\sqrt{\frac{1}{n}}\sum_{i=1}^{n}(obs_i-pred_i)^2}$$

2.3.4.5.6 Ratio of performance to inter-quartile distance

The RPIQ accounts for possibly non-normal distribution of observations (Bellon-Maurel et al., 2010; Zhang et al., 2020). This statistic uses the interquartile range to represent the spread of the observations. The values obtained by RPIQ can be interpreted the same way as those obtained by RPD.

Equation 2-6 shows the calculation of *RPIQ*, where n is the sample size $obsi,$ and $predi$ are vectors of the measured and predicted SOC content, respectively, obs is the total measured SOC content, respectively. The *Q1(obs)* and *Q3(obs)* are the first (25%) and third (75%) quantiles of

the observations *(Q3(obs) − Q1(obs)* is the interquartile distance), and the denominator is the RMSE.

$$\textbf{Equation 2-6: } \mathbf{RPIQ} = \frac{(\mathbf{Q_3(obs)} - \mathbf{Q_1(obs)})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathbf{obs_i} - \mathbf{pred_i})^2}}$$

2.3.4.6 Spectral filtering and pre-processing

Spectral pre-processing is integral to creating calibration algorithms as it improves the spectral quality (Wadoux et al., 2021). Table 2-2 indicates a summary of pre-processing techniques by Wadoux et al. (2021) with references to other studies that also used these methods. The main pre-processing categories include (i) spectral noise removal, (ii) scatter correction, (iii) derivatives, (iv) centring and standardisation, (v) reduction, and (vi) other techniques, such as splice correction and continuum removal (Table 2-2).

Table 2-2: Summary of pre-processing techniques presented by Wadoux et al. (2021) with references to other studies that used these methods.

Pre-processing category	Technique	Reference
Noise Removal	Spectral Trimming	Orr et al., 2017; Wadoux et al., 2021; Long et al., 2023
	Moving Window Average (MWA)	Tekin et al., 2014; Clairotte et al., 2016; Wadoux et al., 2021
	Savitzsky-Golay (SG)	Cambule et al., 2012; McDowell et al., 2012; Stevens et al., 2013; Camacho-Tamayo et al., 2014; Tekin et al., 2014; Clairotte et al., 2016; Seidel et al., 2019; Wadoux et al., 2021;
Scatter Correction	Standard Normal Variate (SNV)	Cambule et al., 2012; Stevens et al., 2013; Camacho-Tamayo et al., 2014; Seidel et al., 2019; Wadoux et al., 2021
	Multiplicative Scatter Correction (MSC)	Cambule et al., 2012; Camacho-Tamayo et al., 2014; Wadoux et al., 2021
	Detrending	Wadoux et al., 2021
Derivatives	First- and Second Order	McDowell et al., 2012; Stevens et al., 2013; Wadoux et al., 2021
Centring and Standardization		McDowell et al., 2012; Tekin et al., 2014; Wadoux et al., 2021
Spectral or dimensions reduction	Resampling	Wadoux et al., 2021; Long et al., 2023
	Wavelets	Wadoux et al., 2021
Other	Splice Correction	Li et al., 2020; Wadoux et al., 2021
	Continuum Removal (CR)	Wadoux et al., 2021

It is important to note that there is no single best method or sequence of methods for spectral pre-processing (Wadoux et al., 2021). The most suitable methods will depend on the quality of the generated (raw) spectra and on the sensitivity of the subsequent analysis to random variation in the spectra.

2.4 Soil organic carbon stocks

SOCS is based on the (i) amount of carbon in the soil as a percentage (%) of the mass of the soil, the (ii) density of soil (soil mass per unit volume), as well as the (iii) volume of soil for which calculations are being done (Aynekulu et al., 2011; TREES Consulting, 2020; Shoch & Swails, 2020). Proper sampling techniques, appropriate laboratory procedures, and accurate calculations are essential to achieve accurate measurements.

2.4.1 Sampling, procedures, and calculations

2.4.1.1 Sampling design

SOCS changes influenced by land-use changes or management practices can be assessed using two main approaches: diachronic and synchronic (Cerri et al., 2021). The diachronic approach includes conducting field experiments in which SOCS are measured over time on the same experimental plots with different management treatments. This approach is accurate and provides repeatability of measurements. However, it is costly and might have a time limit. The synchronic approach is where samples are taken simultaneously from field plots under different management systems at known durations from an initial reference state. SOCS are compared to those from soils under this initial reference state. The accuracy and reliability of data collected using this approach depends on a reasonable selection process of the study sites, which should be comparable.

The random sampling scheme is widely used to represent a given management practice, specifically when the synchronic approach assesses SOCS changes (Cerri et al., 2021). This is the simplest way to select independent and unbiased samples, where each location has an equally probable chance of being selected (Wills et al., 2018; Cerri et al., 2021). It has been recommended that the sample be within a 3 x 3 m grid, with each point 50 m apart from the other, totalling nine points that cover an area of 1 ha. However, since this technique is costly and time-intensive, this sampling technique is not always viable and not commonly employed in South Africa.

2.4.1.2 Sampling depth

Sampling depth is also an essential factor for adequately evaluating the changes in SOCS (Smith et al., 2020; Cerri et al., 2021). The IPCC (IPCC, 2006), Verified Carbon Standard Methodology (Shoch & Swails, 2020), and Gold Standard Methodology (Aynekulu et al., 2011; TREES Consulting, 2020) recommend considering at least the top 30 cm of the soil profile as this is part of the profile is the most affected by land-use or management practices. However, as soon as deeper-rooted crops or trees are planted, usually in integrated agricultural systems (agroforestry,

silvopastoral, and crop-livestock-forest systems), deeper soil sampling is recommended to evaluate the SOCS properly changes over time (Smith et al., 2020, FAO 2020; Cerri et al., 2021). Based on this, the following depth increments are recommended for soil sampling: 0-0.1, 0.1-0.2, 0.3-0.4, 0.4-0.6, 0.6-0.8, and 0.8-1.0 m (FAO 2020; Cerri et al., 2021). It is also recommended that at least two or three samples be collected in each of these depth increments as a composite sample to further increase the accuracy of SOCS measurements.

2.4.1.3 Sampling procedure

The first step of soil sampling for the purpose of calculating SOCS is removing the vegetation or plant litter from the soil surface. Afterward, a soil pit or trench can be dug using a Tractor-Loader-Backhoe (TLB), or soil samples can be collected using a soil corer or auger (ISO, 2006; FAO, 2019). Soil samples (for SOC content) should be collected from each depth increment (cm). For collecting soil samples to determine dry bulk density (ρb), undisturbed samples should be collected using a metal cylinder with a calculated volume (~ 100 cm^3) (IUSS Working Group WRB, 2015). All samples should be placed in sealed, labelled plastic bags and delivered to a credited laboratory within 3 to 5 days (ISO, 2006; IUSS Working Group WRB, 2015).

2.4.1.4 Soil organic carbon content and dry bulk density measurement

In the laboratory, the SOC content is determined by various methods, including dry combustion (TDC) and wet-oxidation (Walkley-Black) methods (Walkley & Black, 1934; Nelson & Sommers, 1996; FAO, 2019). The choice of method depends on various factors, such as the type and quantity of soil sample, the desired accuracy and precision, and the availability of equipment and resources.

The ρb is determined in the laboratory by measuring the mass and volume of an undisturbed soil sample (Blake & Hartge, 1986). The sample is dried in an oven to remove all moisture, and its mass is measured. The ρb is then calculated by dividing the mass of the sample by its volume (Equation 2-7). It is important to note that the sample must be undisturbed to preserve its natural structure and porosity, which affects the accuracy of ρb measurement (USDA, 2018).

Equation 2-7: $\text{Dry bulk density } (\rho b)\left(\frac{g}{cm^3}\right) = \text{Dry soil (g)/Cylinder volume } (cm^3)$

2.4.1.5 Calculations of soil organic carbon stocks

When discussing SOCS calculations, it is essential to acknowledge that there might be sources of error or uncertainty in the calculations. For example, soil properties may vary across a landscape or change over time, making it challenging to estimate SOCS accurately. Therefore, these potential sources of error must be minimised. Inaccuracies in SOCS calculations can also

significantly affect carbon credit accounting and management decisions. If SOCS is overestimated, it could lead to overestimating carbon sequestration and result in too many carbon credits. On the other hand, if SOCS is underestimated, it could lead to an underestimation of carbon sequestration and result in too few carbon credits (Gavrilova, 2019).

Various tools and software are available to improve the accuracy of SOCS calculations. These can help standardise the sampling and analysis process, reduce errors, and provide more reliable estimates of SOCS (Paustian et al., 2016). Examples of such tools include the Agricultural and Environmental Data Interpretation System (AEDIS) (USDA, 2018) the Soil Organic Carbon Estimation Calculator (SOCEst) and other modelling software. The SOCS is usually calculated by multiplying the SOC content (% or g/kg) of soil with the ρb (g/cm³) of the soil with the coarse fragments (1 – Sm (v%)) and the depth at which the sample was taken (Equation 2-8).

Equation 2-8: $$\mathbf{SOCS\ (t/ha) = 10 \times SOC\ content\ (g/kg) \times \rho b\ (Mg/m^3) \times (1 - Sm\ (v\%)) \times layer\ thickness\ (cm)}$$

2.4.2 Mapping of soil organic carbon stocks

Many geostatistical models, such as ordinary kriging (OK), regression kriging, co-kriging and inverse distance weighted regression, have been used in SOC content and SOCS mapping, because of their prediction accuracy and error minimisation (Farooq et al., 2022). Digital soil mapping (DSM) has been used in the past two decades to create more accurate and precise SOCS maps (Nenkam et al., 2024). This method predicts soil properties, such as SOCS, by correlating the relationship between soil properties and environmental covariates utilising machine learning and other methods (Pouladi et al., 2023).

2.4.2.1 Ordinary kriging

OK has been one of the most popular geostatistical models in recent years (Pouladi et al., 2023). OK interpolation is an early method for soil mapping; however, this geostatistical method is based on spatial point data for interpolation and cannot fully utilise the surface data of environmental covariates (Zhu et al., 2022).

OK uses an estimated mean of a particular soil property (i.e., SOCS) at a known location to predict the value at an unsampled location (Yao et al., 2020). The predicted/interpolated value ($Z(x_0)$) at unsampled locations are calculated as presented by Equation 2-9, where n is the number of the adjacent observations used for the predictions and λ_j is the kriging weight for the $Z(x_i)$ values (Yao et al., 2020). Maps created using OK are usually evaluated using the validation statistics described and discussed in 2.3.4.5.

Equation 2-9: $$\mathbf{Z(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i)}$$

2.4.2.2 Digital soil mapping with machine learning models

The recent development of DSM technologies enabled soil scientists to map SOCS from local to continental scales, using the scorpan factors, including soil, climate, organisms, and parent materials (Minasny et al., 2013). The SOCS depends on the balance between carbon inputs and outputs and the environmental conditions (scorpan factors) that directly and indirectly affect the processes controlling the spatial distribution of SOCS (Gomes et al., 2019). It is well understood that climatic conditions, such as precipitation and temperature, influence SOCS distribution, but other factors, such as soil texture, also influence SOCS, and clay amounts are critical to SOC protection due to interactions between the SOC and reactive surfaces of clay minerals (Gomes et al., 2019). Vegetation also favours the input of decaying plant biomass, especially in the topsoil (0-30 cm) (Bui et al., 2009), and vegetation indexes, such as normalised difference vegetation index (NDVI), have been used to predict SOCS. Terrain attributes, such as elevation, slope, and curvature, also have been used to predict SOCS. The combination of these different environmental factors creates unique local conditions, leading to increasing or decreasing SOCS, and are used to predict the spatial distribution of SOCS in DSM. As mentioned, land use and management practices also significantly impact SOCS. Therefore, it is crucial to consider these aspects when mapping SOCS to capture the spatial distribution of SOCS accurately and to design effective management strategies for maintaining or increasing SOCS.

To accurately estimate SOCS, DSM relies on various input data sources, including remote sensing, topographic, and soil data. Remote sensing data, such as satellite imagery, provide information on land cover, vegetation properties, and landscape characteristics, which can be used to estimate SOCS (Luo et al., 2010; Gomes et al., 2019). Topographic data, such as elevation and slope, can provide information on soil erosion and deposition, affecting SOCS. Soil data, such as soil texture, OM content, and ρb, are essential for accurate SOCS estimation, as they are direct indicators of SOCS (Luo et al., 2010).

Several ML models, including RF, Cubist, Support Vector Machines (SVM), and kriging-based models, are used to estimate the spatial distribution of SOC (Gomes et al., 2019). RF uses decision trees to classify data, while Cubist can handle continuous and categorical data. Kriging is a geostatistical interpolation method that estimates unknown values at unobserved locations (Krige, 1951). These models help to create more accurate and detailed soil maps by incorporating various data sources to estimate the spatial distribution of SOCS. The different ML models are also described and discussed in 2.3.4.4. Previous studies (Brevik et al., 2016; Shi et al., 2021) show the versatility and effectiveness of different ML models in mapping SOCS through DSM.

However, the accuracy and resolution of the input data used in DSM can impact the accuracy of the resulting SOC maps. Furthermore, the availability and quality of ground-truth data for model calibration and validation can also be a limiting factor (Grunwald et al., 2016). Another challenge is the need to balance the trade-off between model complexity and accuracy, as more complex models may result in overfitting and decreased prediction accuracy. Finally, it is crucial to consider the temporal dynamics of SOC, as changes in land use and management practices can result in rapid changes in SOCS, which static SOC content maps may not capture (Luo et al., 2010). Validation statistics for evaluating the reliability and accuracy of maps created by DSM with ML are also described and discussed in 2.3.4.5.

2.4.2.3 Soil organic carbon stocks mapping of South Africa

Several studies have focused on mapping SOC content and SOCS across South Africa (Shulze & Schutte, 2020; Venter et al., 2021; Kotze & van Tol, 2023; Odebiri et al., 2024). Venter et al. (2021) mapped SOCS from 0-30 cm under natural vegetation over South Africa at 30 m spatial resolution, while Schulze & Schutte (2020) mapped SOC content for areas across South Africa at a spatial detail that were not previously achieved. On the other hand, Kotze & van Tol (2023) and Odebiri et al. (2024) mapped SOCS at smaller scales, with Kotze & van Tol (2023) mapping at a catchment scale using DSM and Universal Kriging and Odebiri et al. (2024) mapping SOCS on a more regional scale. Although these studies did not directly focus on comparing conventional mapping with DSM, they provide the cornerstone of SOCS mapping in South Africa. These studies also support the importance of creating accurate and reliable SOCS maps for South African soils.

2.4.3 Re-measuring and modelling approaches

There is an ongoing debate about the most appropriate approach for quantifying changes in SOC content and SOCS (Paustian et al., 2016; Smith et al., 2019; FAO, 2019). Some argue that modelling approaches are more accurate and cost-effective than direct measurements (Smith et al., 2019), while others argue that direct measurements are the most accurate and reliable approach (Lal, 2005).

Modelling approaches involve using numerical models, such as Rothamsted Carbon (RothC), CENTURY and DayCent, to simulate how management practices or land use changes affect SOCS. These models are commonly used for modelling SOC sequestration, while other techniques, such as ML, are used for estimating the actual SOC content or SOCS. Direct measurements involve physically sampling soils and analysing them in a laboratory to quantify changes in SOC content and SOCS. Importantly, the viability of the modelling and measuring approaches are greatly dependent on land management practices, affordability and accuracy.

2.4.3.1 Re-measuring approaches

Direct re-measurements involve physically sampling soils (usually with a soil auger or core sampler) and analysing the samples in a laboratory to quantify changes in SOC content and SOCS. However, the approach does require many samples and relies on appropriate study designs and protocols (Smith et al., 2019). Because the analysis also involves the re-measurement of SOC content and ρb. Combustion methods, such as TDC, can make large-scale measurements very expensive. In some cases, it may be of interest to quantify and account for the coarse fraction of belowground SOC content that is affected by grasslands or forest soils (FAO, 2019; Smith et al., 2019). Since changes in management practices influence SOC content, the ρb will also be affected. Therefore, an "equivalent mass basis" approach is recommended when comparing SOCS across land uses and management regimes (Wendt & Hauser, 2013; Smith et al., 2019). Another two disadvantages of accurate direct re-measurements are the requirement of large numbers of soil samples (Vanguelova et al., 2016; Smith et al., 2019) and sufficient sampling depth, which is crucial for adequately evaluating changes in SOC content (IPCC, 2006; Smith et al., 2019).

2.4.3.2 Modelling approach using Rothamsted Carbon (RothC) model

RothC is classified as a level 2 'soil' model (IPCC, 2006; FAO, 2019) for the turnover of SOC content in non-waterlogged topsoil that allows for the inclusion of the effects of soil type, temperature, moisture content, and plant cover on the turnover process, with a monthly time step (Coleman & Jenkinson, 1996; FAO, 2019). This model is a widely used model for estimating SOCS and is a process-based model that simulates the turnover of SOM in the soil and estimates the changes in SOCS over time (Jenkinson et al., 1991). The model has been extensively calibrated and validated for various soil types and land-use systems. Advantages of the RothC model include its ability to account for the effects of different management practices on SOCS, and this model has relatively simple input requirements. However, the model may not be suitable for all soil types and may require site-specific calibration to improve its accuracy (Falloon and Smith, 2002). The model also induces room for error, as the SOC inputs from plant residues and animal excreta must be estimated and based on scientific literature, which might not be readily available or inaccurate (Poeplau, 2016; FAO, 2019). Parameters, such as plant SOC input and fertiliser SOC input are often problematic due to high uncertainty and model sensitivity (Poeplau, 2016; FAO, 2019). Inaccurate estimates would lead to inaccurate SOCS quantification, leading to inaccurate carbon credits.

The RothC model has four active carbon pools (Afzali et al., 2019; FAO, 2019), as seen in Table 2-3. The decomposable plant material (DPM) is the most labile carbon pool in the soil. It consists

of recently added plant residues and other organic materials easily decomposed by microorganisms. The resistant plant material (RPM) is a carbon pool that contains less easily decomposed organic materials than DPM and includes more recalcitrant plant residues, such as cellulose and lignin-rich compounds. The humified organic matter (HUM), or humus, is a highly stable carbon pool that consists of organic materials that have undergone substantial chemical alterations, resulting in more excellent resistance to microbial decomposition. The inert organic matter (IOM) pool represents the soil's most stable and recalcitrant carbon pool (Afzali et al., 2019). It consists of highly complex organic compounds that are resistant to microbial decomposition. In addition to these carbon pools, RothC modelling also considers microbial biomass (BIO), as microbes play a crucial role in decomposing organic materials and transforming carbon between different pools.

Table 2-3: The four active Carbon pools and microbial biomass used for RothC modelling.

Carbon active pool name	Acronym
Decomposable plant material	DPM
Resistant plant material	RPM
Microbial biomass	BIO
Humified organic matter	HUM
Inert organic matter	IOM

Figure 2-2 indicates the structure of the carbon pools and flows in the RothC model, initially created by Coleman & Jenkinson (1996) and Falloon & Smith (2002). These include major factors controlling the fluxes (a = multiplier for effects of temperature, b = multiplier for effects of moisture, c = multiplier for effects of soil cover, DPM/RPM = Decomposable/resistant plant material ratio).

The incoming plant carbon is split between DPM and RPM, depending on the DPM/RPM ratio of the incoming plant material (Coleman & Jenkinson, 2014). The default DPM/RPM ratio is 1.44 for most agricultural lands and improved grassland, with 59% of plant material DPM and 41% RPM (Coleman & Jenkinson, 2014; FAO, 2019). The default DPM/RPM ratio of 0.25 is used for deciduous or tropical woodland, with 20% plant material DPM and 80% RPM. All incoming plant material only passes through these compartments or pools once (Coleman & Jenkinson, 2014).

DPM and RPM will decompose with time to form carbon dioxide, BIO, and HUM (Figure 2-2). Notably, the proportion of carbon dioxide and BIO + HUM is determined by the soil's clay content (%) (Coleman & Jenkinson, 2014). The BIO + HUM is then split into 46% BIO and 54% HUM, and both decompose further to form more carbon dioxide, BIO, and HUM. The model derives IOM for the 0-20 cm of the soil layer by using Equation 2-10 (Falloon et al., 1998):

Equation 2-10: $$IOM = 0.049\,(total\ SOC)^{1.139}$$

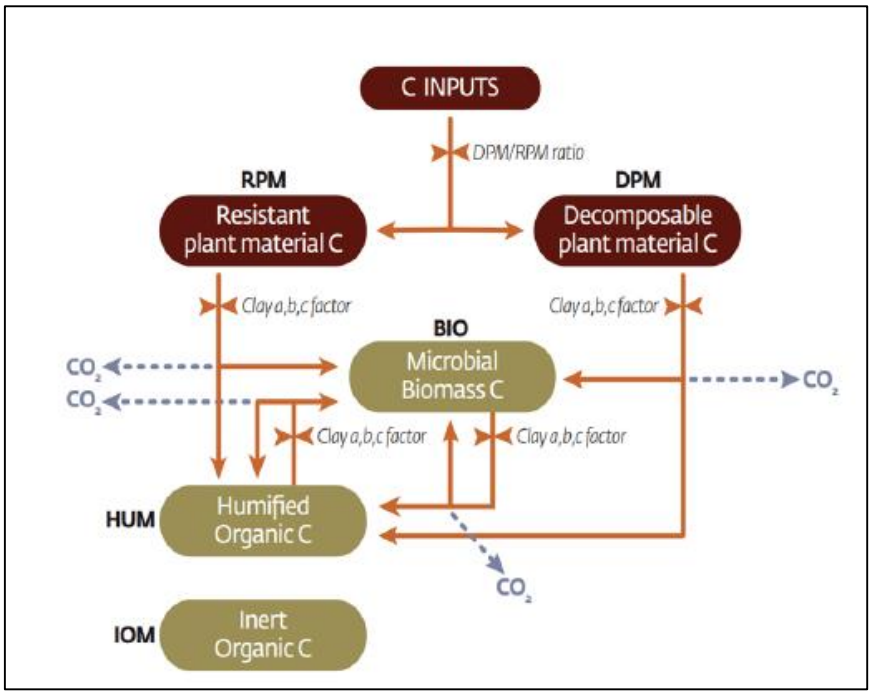


Figure 2-2: Structure of the carbon pools and flows of carbon in the RothC model, including major factors controlling the fluxes (a = multiplier for effects of temperature, b = multiplier for effects of moisture, c = multiplier for effects of soil cover; DPM/RPM = Decomposable/resistant plant material ratio) (FAO, 2019).

2.4.3.2.1 Input data requirements

Table 2-4 indicates the minimum input data requirements for running the RothC model. Input data to run the RothC model include data regarding climate (mean monthly precipitation, mean monthly air temperature and monthly evaporation), soils (initial SOCS and initial SOCS of different carbon pools), and land management/vegetation data (monthly soil cover, monthly carbon inputs from plant residues, organic fertilizers, and manure, and DPM to RPM ratio).

Table 2-4: RothC model minimum input data requirements (FAO, 2019).

Climate data	Soil data	Land management data
Mean monthly precipitation (mm)	Total initial 0-30 SOCS (t C/ha)	Monthly soil cover (bare vs vegetated)
Mean monthly air temperature (°C)	Initial SOCS of pools (t C/ha): DPM, RPM, BIO, HUM, and IOM	Irrigation (mm)
Monthly open-pan evaporation (mm)	Clay content (%) at simulation depth (cm)	Monthly C inputs from plant residue (t C/ha)
		Monthly C inputs from organic fertilizers and manure (t C/ha)
		DPM: RPM ratio (estimate of decomposability)

Precipitation and potential evapotranspiration (PET) are used to calculate topsoil moisture deficit (TSMD), as it is easier to do this than obtain monthly measurements of the actual topsoil water deficit (Coleman & Jenkinson, 2014). If PET data is unavailable, monthly potential evapotranspiration can be calculated with adequate accuracy using Müller's (1982) collection of meteorological data from around the world. Thornthwaite's equation could also be used to calculate the PET (Thornthwaite, 1948). Notably, air temperature is used rather than soil temperature as it is more obtainable for most sites. As mentioned, clay content (%) affects the way SOM decomposes and is used to calculate the amount of plant-available water the topsoil can hold (Coleman & Jenkinson, 2014). A crucial land management input factor is the soil cover status, as it is necessary to indicate whether the soil is vegetated because decomposition is faster in fallow soils than in cropped soil, even when the cropped soil is not allowed to dry out (Sommers et al., 1981; Coleman & Jenkinson, 2014). The plant residue input is calculated as the amount of carbon put into the soil per month (t C/ha), including carbon released from roots during crop growth. Notably, this input is rarely known, and the model is often run in "inverse" mode, generating input from known soil, site, and weather data (Coleman & Jenkinson, 2014). The amount of fertiliser and farmyard manure (FYM) is input separately because it is treated slightly differently from inputs of fresh plant residues.

CHAPTER 3 THE CARBON CREDIT CONUNDRUM: WHICH ANALYTICAL METHOD SHOULD BE USED FOR DETERMINING SOIL ORGANIC CARBON CONTENT IN SOUTH AFRICA

3.1 Introduction

The rise in atmospheric carbon dioxide and other greenhouse gas emissions poses a significant challenge both locally and globally, contributing to climate change that endangers environmental and social well-being (Lal, 2009; Zhou et al., 2019). A process to reduce emissions, such as carbon dioxide and methane, is through carbon sequestration, which involves the long-term storage of carbon in oceans, soils, vegetation, and geological formations (Olson & Al-Kaisi, 2014). Recent studies have indicated that the sequestration of carbon in the soil is a promising strategy for decreasing carbon emissions (Nair, 2012; Rhodes, 2017; Baurov, 2021; Zhang et al., 2021). Farmers and local land managers are uniquely positioned to sequester carbon and monetise this effort by selling carbon credits to the government or businesses, providing a source of income (Sharma et al., 2021). The demand for these carbon credits is rapidly increasing in South Africa.

Accurate quantification of soil organic carbon (SOC) content is essential to account for these carbon credits. Currently, South Africa predominantly uses the Verified Carbon Standard Methodology (Shoch & Swails, 2020) and the Gold Standard (Aynekulu et al., 2011; TREES Consulting, 2020) as frameworks for determining SOC content. Unfortunately, the Verified Carbon Standard Methodology (Shoch & Swails, 2020) does not specify which analytical method should be used for determining SOC content, while the Gold Standard (TREES Consulting, 2020) states that only the total dry combustion (TDC) method may be used. This uncertainty and variability in SOC content quantification introduces a level of risk, potentially affecting the credibility and reliability of carbon credit calculations. Thus, for credible carbon credits, SOC content needs to be determined by a single, reliable analytical method or methods with reliable conversion factors to other methods.

The three common analytical methods used for determining SOC content are the TDC (Nelson & Sommers, 1996; Kumar et al., 2019; Roper et al., 2019; FAO, 2019), loss-on-ignition (LOI) (Roper et al., 2019; FAO, 2019) and Walkley-Black wet-oxidation (WB) (Walkley & Black, 1934; Kumar et al., 2019; Roper et al., 2019; FAO, 2019) methods. The TDC method works by combusting SOC at very high temperatures (around 1 200 ℃). The carbon dioxide released during decomposition is quantified using an elemental analyser (also called CN analyser) (Pribyl, 2010; Roper et al., 2019; FAO, 2019). This high-throughput method has been shown to have a higher

accuracy compared to other methods and has been widely accepted as the standard method for SOC content determination (Aynekulu et al., 2011; TREES Consulting, 2020; FAO, 2019). The TDC method measures the total soil carbon, unless the inorganic fraction is removed using an acidic reagent or no inorganic carbon is present in the sample, typically in soils with a pH less than 7 (Konen et al., 2002; Pribyl, 2010; Roper et al., 2019; FAO, 2019).

The LOI method is a rapid and simple method that also works on the principle of thermal decomposition. Soil samples are combusted in a muffle furnace at around 550 ℃ with the gravimetric loss measured as soil organic matter (SOM). This method uses a correction factor of 1.724 to convert SOM to SOC content, by assuming that 58% of the SOM is comprised of SOC. Although there has been debate on this conversion factor, with SOC content reportedly ranging from 50 to 66% of SOM, the correction factor is still the most acceptable to date (Pribyl, 2010; Hoogsteen et al., 2015; Roper et al., 2019; FAO, 2019).

Finally, the WB method is a rapid method that requires minimum equipment that uses dichromate to oxidise SOC in concentrated sulphuric acid and afterward titrating the remaining dichromate with ferrous sulphate (Walkley & Black, 1934; Konen et al., 2002; De Vos et al., 2005; Pribyl, 2010; Roper et al., 2019; FAO, 2019). The amount of dichromate consumed during the reaction is then used to calculate the SOC content in the soil sample. However, incomplete oxidation can lead to SOC content underestimation and the use of dichromate might also be chemically harmful to the environment (Walkley & Black, 1934; Konen et al., 2002; De Vos et al., 2005; Pribyl, 2010; Roper et al., 2019; FAO, 2019).

Many previous studies have been conducted on the relationship between these analytical methods (De Vos et al., 2005; Abella & Zimmer, 2007; Roper et al., 2019; Visconti et al., 2022; El Mouridi et al., 2023). However, only two studies have been conducted in Africa (Yerokun et al., 2007; McCarty et al., 2010) and none have focused on South African soils. For regions or countries where certain analytical methods are unavailable, pedotransfer functions have been developed to convert SOC content values from one method to another (Fernandes et al., 2015; Gessesse & Khamzina, 2018; El Mouridi et al., 2023). Despite this, there are concerns that the unique characteristics of South African soils and environments may affect their applicability (Kock et al., 2024). Furthermore, there has been a lack of studies exploring the transfer of these functions between different countries or regions. As mentioned above, the WB method could lead to incomplete oxidation and consequently to SOC content underestimation. Calculating the recovery factor (RF) of the WB method, is essential in determining the extent to which the WB method underestimates SOC content to the TDC method.

The accurate assessment of carbon credits is crucial for local land managers and the industry, as identifying the most reliable method can not only enhance efficiency but also significantly reduce costs, ultimately benefiting all stakeholders involved. To address this gap, this study aimed to determine which analytical method should be used for determining SOC content for the assessment of carbon credits in South Africa. Secondly, it determined whether pedotransfer functions could be used for transferring SOC content values between methods.

3.2 Materials and methods

3.2.1 Study area and site description

For the study area, five catchments (Figure 3-1a) were selected based on their ecological and/or economic importance in South Africa as part of the HYDROSOIL project (van Zijl et al., 2024). The five catchments included several different soil-forming factors, such as climate, land use and geology, which led to the representation of almost all South African soils.

The Goukou catchment, located in the Western Cape Province, experiences a median annual precipitation (MAP) ranging from 400 to 600 mm (Shulze, 2007). The area is primarily composed of calcareous sedimentary rocks, such as sandstone, mudstone, shale, siltstone, conglomerate, and quartz (Council for Geoscience, 2019). Land use in the catchment mainly consists of commercially cultivated areas and low shrubland (fynbos) (GeoTerra Image, 2020). The soils in the Goukou catchment are characterised by sand dunes (Regosols) in the southern region, while the northern region features duplex soils (Solonetz and Planosols) and the crests feature shallow, rocky soils (Leptosols) (Land Type Survey Staff, 1972-2006). Soils in the Western Cape, specifically in the Goukou catchment, also have a higher soil inorganic carbon (SIC) fraction (carbonatic soils) compared to the other catchments (Sharififar et al., 2023). Some smaller areas within the catchment are also characterised by saturated soils (Gleysols) and located close to wetlands (Land Type Survey Staff, 1972-2006).

The Olifants catchment, located in the Gauteng Province, is 1 567 km² in size and receives a MAP of 691 mm in the north and 787 mm in the south (Shulze, 2007). The catchment area primarily consists of sandstones, shales, quartzites, and dolomites (Council for Geoscience, 2019). Land use in the catchment is mainly coal mines, commercial cultivated areas, and natural grassland (GeoTerra Image, 2020). The soils of the Olifants catchment are characterised by sandy topsoils overlying plinthic subsoils (Plinthosols), with shallow soils (Leptosols) occurring over large areas of the catchment (Land Type Survey Staff, 1972-2006). Some areas of the catchment also include deep, sandy soils (Acrisols) (Land Type Survey Staff, 1972-2006).

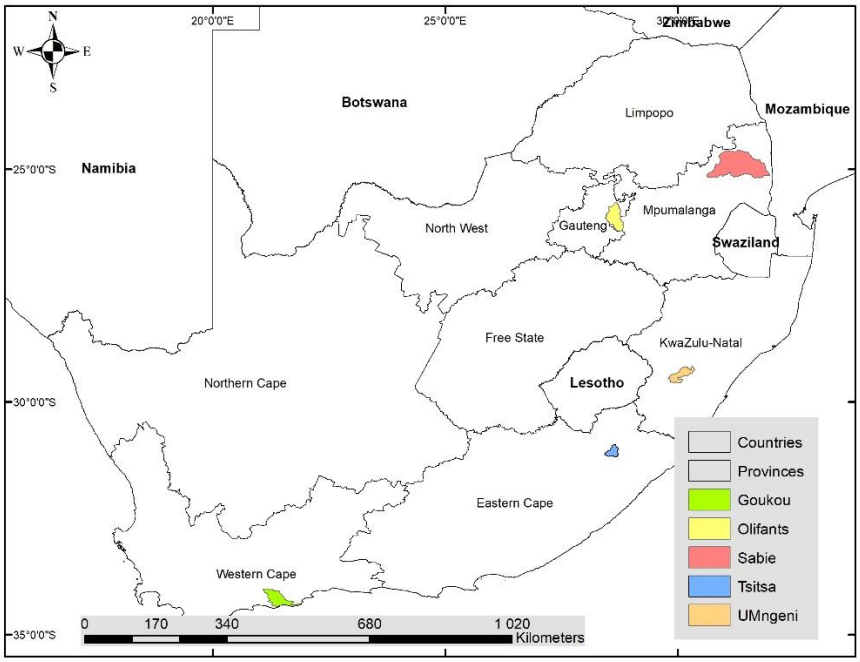
The Sabie catchment, located in the Mpumalanga Province, is 5 790 km² and receives a MAP of 540 mm in the eastern part and 1 200 mm in the western part of the catchment (Shulze, 2007).

The geology of the catchment includes quartzites, granites, basalts, conglomerates, andesites, gneiss and shales (Council for Geoscience, 2019). The land use varies from natural vegetation in the Kruger National Park and Sabie-Sands Game Reserve in the east, commercial irrigated agricultural areas in the lower lying areas and forest plantations in the west (GeoTerra Image, 2020). Large areas of this catchment contain soils that are sandy and deep (Acrisols), while small parts of the catchment contain shallow soils (Leptosols) (Land Type Survey Staff, 1972-2006). In some areas, soils overlying relatively impermeable bedrock (Arenosols) can also be found.

The Tsitsa catchment, located in the Eastern Cape Province, is 494 km² in size and receives a MAP of approximately 800 mm in the north and approximately 950 mm in the south (Shulze, 2007). The geology of the catchment includes sandstones, mudstones, dolomites, and shales (Council for Geoscience, 2019). The land use varies from being predominantly built-up land (villages), natural grassland, commercial cultivated areas, and fallow lands, while there are also small distributions of forested land within the catchment area (GeoTerra Image, 2020). Large areas of this catchment also contain soils that are sandy and deep (Acrisols), while small parts of the catchment contain shallow soils (Leptosols) (Land Type Survey Staff, 1972-2006). Soil erosion is a major land degradation problem across South Africa but is particularly severe in the Tsitsa catchment and ultimately leads to a decrease of SOC content in the area (Ighodaro et al., 2013; Parwada & Van Tol, 2016; du Plessis et al., 2020).

The UMngeni catchment, located in the KwaZulu-Natal Province, is 986 km² in size (Figure 3-1b) and receives a MAP of more than 1 200 mm (Shulze, 2007). The geology of the catchment includes dolerite, sandstones, mudstones, and shales (Council for Geoscience, 2019). Large areas of this catchment also contain soils that are sandy and deep (Acrisols), while small parts of the catchment contain shallow soils (Leptosols) (Land Type Survey Staff, 1972-2006). The soil within the catchment is generally well-drained and high in SOM (Histosols) (Fey, 2010). This catchment supplies water to nearly 15% of South Africa's population (Summerton & Schulze, 2009), with the land use being predominantly forested land (GeoTerra Image, 2020).

(a)



(b)

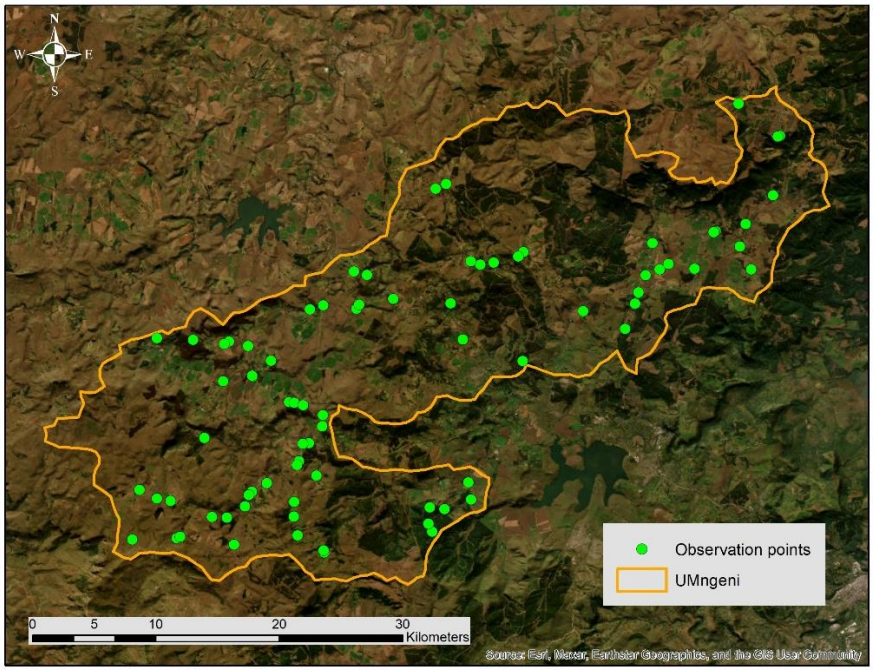


Figure 3-1: (a) The study area indicating the five catchments selected where soil samples were collected, as well as (b) an example of the observation points that were distributed throughout the UMngeni catchment.

3.2.2 Soil sampling

The sampling locations in each catchment were selected using the Conditioned Latin Hypercube Sampling (cLHS) design (Minasny & McBratney, 2006), using slope, elevation, geology, MAP and Normalized Difference Vegetation Index (NDVI) as covariates (Figure 3-1). This approach ensured that the attribute space of the different catchments was covered and should represent the SOC diversity accordingly. A total of 220 cLHS locations were identified and topsoil (0-30 cm) collected using a soil auger (Table 3-1). Only topsoil samples were taken, as per the requirement of the Verified Carbon Standard Methodology (Shoch & Swails, 2020) and Gold Standard (TREES Consulting, 2020). Sampling for the different catchments took place at various stages throughout the year and were conducted between December 2021 and November 2022. Although the sampling was done at different stages, SOC content was not expected to change over this period (Wuest, 2014).

Table 3-1: The number of samples (*n*) from each catchment for soil organic carbon (SOC) content and pH analyses.

Catchment	SOC content analysis (n)	pH analysis (n)
Goukou	33	8
Olifants	53	18
Sabie-Sands	43	25
Tsitsa	56	8
Umngeni	35	15
TOTAL	220	74

The 220 collected samples were analysed for SOC content with the three methods: WB, TDC and LOI methods (incorporating the correction factor of 1.724) (FAO, 2019). No samples were treated with any acidic reagent to remove the SIC content. Only 74 samples were selected and analysed for pH (in a 1:2.5 KCl solution) (Table 3-1) due to budget constraints. The 74 samples represented the main soil forms found in the study area, as mentioned in section 2.1.

3.2.3 Statistical analysis

3.2.3.1 SOC content distribution

The SOC content distribution for each site was shown through box-and-whisker plots created using the “ggplot” function in RStudio (Wickham, 2016). Given that the assumption of normality was not met, a Kruskal-Wallis’s test (Kruskal and Wallis, 1952) was performed in RStudio (R Core

Team, 2020) using the “kruskal.test” function, which was done to determine if there were overall significant statistical differences between the values of the methods. Afterwards, a pairwise means comparison test (“pairwise.wilcox.test” function) was also done in RStudio to determine the significance differences between each pair of methods.

3.2.3.2 Creation of pedotransfer functions

The “lm” and “poly” functions were used to fit a linear and polynomial model to the data (R Core Team, 2020). The combination of the “lm”, “poly” and “ggplot” functions in RStudio (Wickham, 2016) were used to create scatter plots and extract and visualise the pedotransfer functions between the methods. Ultimately, the best performing models were shown in this paper.

3.2.3.3 Determination of SOC content percent recovery and factor

The percent recovery (%R) for the WB method was determined by using Equation 3-1 (Mikhailova et al., 2003), while Equation 3-2 (Matus et al., 2009) was used to determine the recovery factor (RF).

Equation 3-1: $\%R = (SOC_{WB}/SOC_{TDC}) \times 100$

where SOC_{WB} refers to the SOC content by the WB method and SOC_{TDC} refers to the SOC content determined by the TDC method, with all quantities expressed as percentages.

Equation 3-2: $RF = (100/\%R)$

3.2.3.4 Carbon credit assessment

The implications of the different measurement methods for carbon credit assessment were demonstrated by selecting four specific samples that were measured with all three analytical methods. The implication of using the RF for the WB method was also added for each specific sample. The first sample selected had a very high SOC content, while the second sample selected had a very high SIC content, as indicated by the large difference between the WB and TDC measurements. The third and fourth samples selected, had a very low RF (1.1) and very high RF (3.5), respectively. Equation 3-3 was used to calculate the soil organic carbon stocks (SOCS) for each sample, assuming a dry bulk density of 1.3 g/cm^3 to a depth of 30 cm.

Equation 3-3: SOCS = SOC content (%) × dry bulk density (g/cm^3) × layer thickness (cm)

Since the price of carbon credits is trading between \$15 - \$25 per tCO_2e (Orizon Agriculture, 2024), the value of a carbon credit was deemed to be \$20 for this demonstration. The amount of 1-ton carbon dioxide equivalence (CO_2e) was calculated by multiplying the SOCS with 3.76 (molecular fraction).

3.3 Results and discussion

3.3.1 Soil organic carbon content distribution

Overall, the three methods presented varied measurements for SOC content, with statistically significant differences observed between the methods (Figure 3-2). The three methods also presented varied measurements for SOC content across the different catchments (Figure 3-3). Statistically significant differences between methods were observed only in the Olifants-, Tsitsa-, and Sabie catchments (Figure 3-3). In these catchments, the LOI method showed the highest SOC content distribution, while the WB method showed the lowest SOC content distribution (Figure 3-3). For the Goukou catchment, the TDC method showed slightly higher SOC content compared to the other two methods, although no statistical differences were found between the methods for this catchment (Figure 3-3). The pH distribution of the catchments showed that the Goukou had the highest pH, while the UMngeni had the lowest pH (Figure 3-4). However, except for the Goukou catchment, all other catchments were significantly indifferent from each other (Figure 3-4).

The Goukou catchment exhibited a higher pH distribution (median = 6.35), suggesting nearly alkaline soil and the potential presence of SIC compared to other catchments. The elevated pH may be attributed to the proximity of the samples to the coastal dunes, which are rich in carbonates from sea-spray deposits (Sharififar et al., 2023), or to the influence of carbonates derived from calcareous weathered rocks (Council for Geoscience, 2019). Soils with a higher pH (> 6.35) typically have increased saturation with basic cations (Ca^{2+} and Mg^{2+}), which react with dissolved bicarbonates and carbonates. This interaction can lead to the precipitation of SIC once critical ion concentrations are achieved (Batool et al., 2024). This shows that the TDC method may have overestimated the SOC content, by measuring both SOC- and SIC content. SIC can be removed by the addition of an acidic reagent, such as HCl, which has been done so that only SOC content is measured (Roper et al., 2019; FAO, 2019). Since the Goukou catchment is the only catchment likely to contain SIC, the TDC method should still be considered the preferred method for determining SOC content in soils with a pH < 6.35.

The results indicated that the UMngeni catchment had the lowest pH distribution (Figure 3-4) but also the highest SOC content across all three methods (Figure 3-3). The SOC content measured by the WB method was consistently lower for all catchments compared to the other methods. Although not statistically significant, the SOC content from the WB method (median = 4.22 % for the UMngeni catchment) was nearly 1.5 % lower than that of the TDC method (5.68 %). This lower SOC content from the WB method was likely due to incomplete oxidation (Walkley & Black, 1934; FAO, 2019), with recent studies suggesting that only 60-86% of SOC is oxidized (Kumar et al., 2019; FAO, 2019). The incomplete oxidation of SOC content, likely led to the method

underestimating the SOC content for the UMngeni catchment. The incomplete oxidation of this method is a common occurrence, and studies have tried to incorporate correction factors (Meersmans et al., 2009; Roper et al., 2019), however, no standardised correction factors are currently available. Although re-sampling was not done for this study, previous studies have shown that improved precision can be achieved at field scale by using the TDC method, rather than the WB method (Bowman et al., 2002; Goidts et al., 2009). Therefore, if the soil is expected to have a high SOC content, the WB method should be avoided.

The LOI method demonstrated the highest overall distribution of SOC content (Figure 3-2) and showed the greatest SOC content for the Olifants, Sabie, and Tsitsa catchments (Figure 3-3). Previous research supports these findings, indicating that the LOI method typically yields higher SOC content than the WB method (Fernandes et al., 2015; Gessesse & Khamzina, 2018; Kumar et al., 2019; Visconti et al., 2022). Jensen et al. (2018) also found that the LOI method reported significantly higher SOC content compared to the TDC method, consistent with the results (Figure 3-2 and Figure 3-3). It is important to note that the LOI method employs a correction factor of 1.724 (assuming 58 % of SOM consists of SOC), which has been shown to overestimate SOC content (Hoogsteen et al., 2015; Jensen et al., 2018; FAO, 2019). Some studies have suggested that the correction factor should be closer to 50% of SOM (Pribyl, 2010; Hoogsteen et al., 2015; Roper et al., 2019; FAO, 2019). Hoogsteen et al. (2015) have also reported a clay correction factor (0.01 to 0.09), however, there are currently no standardised correction factors to be used for the LOI method. Overall, the higher distribution of SOC content observed in the LOI method indicates less consistent values compared to the WB and TDC methods. The LOI method is still a widely used method for determining SOM, however, this method should be avoided when determining SOC content.

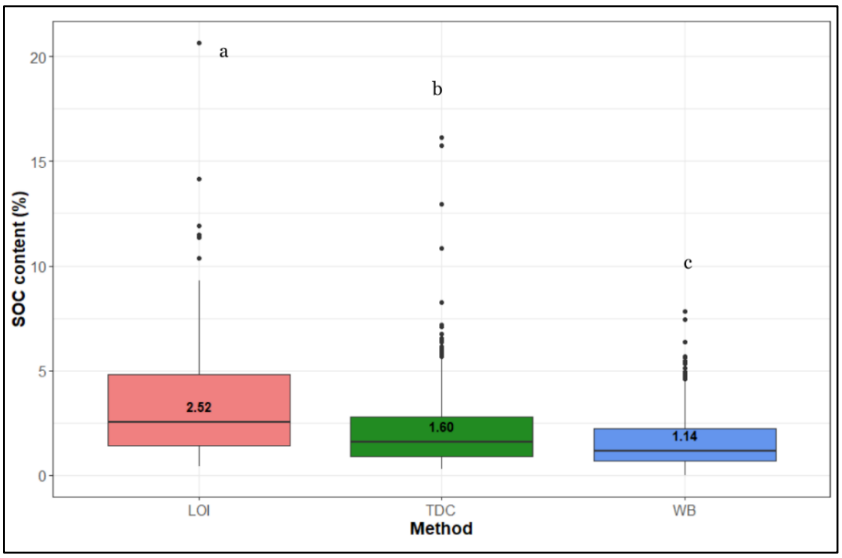


Figure 3-2: The SOC content distribution of the three methods, including total dry combustion (TDC), Walkley-Black wet-oxidation (WB) and loss-on-ignition (LOI). The Kruskal-Wallis (KW) chi-squared (X^2) value, as well as superscript is seen to display the significant differences between the methods.

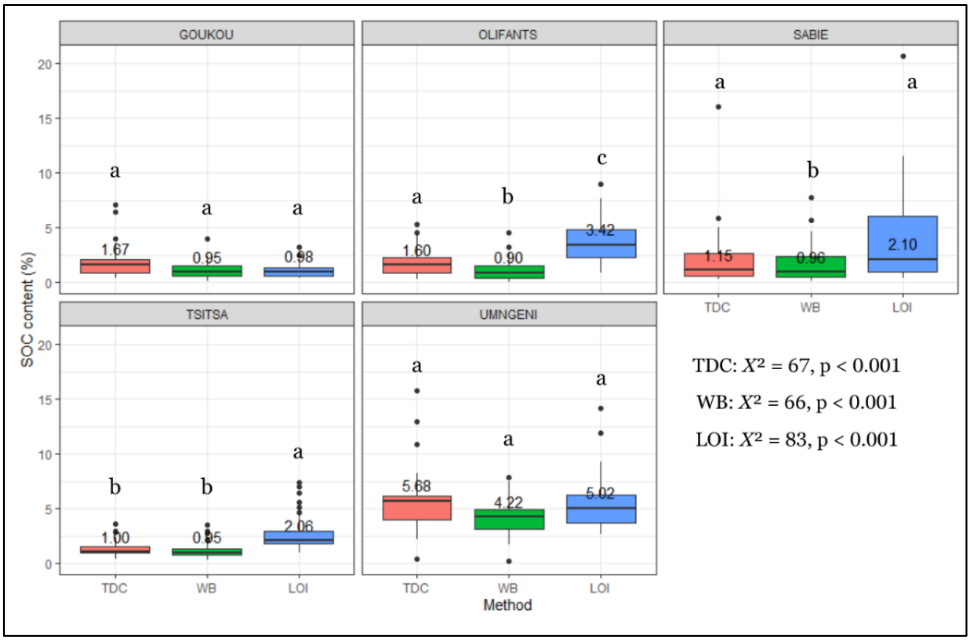


Figure 3-3: The SOC content distribution of the five catchments for the three methods including total dry combustion (TDC), Walkley-Black wet-oxidation (WB) and loss-on-ignition (LOI). The Kruskal-Wallis (KW) chi-squared (X^2) value, as well as superscript is seen to display the significant differences between the methods.

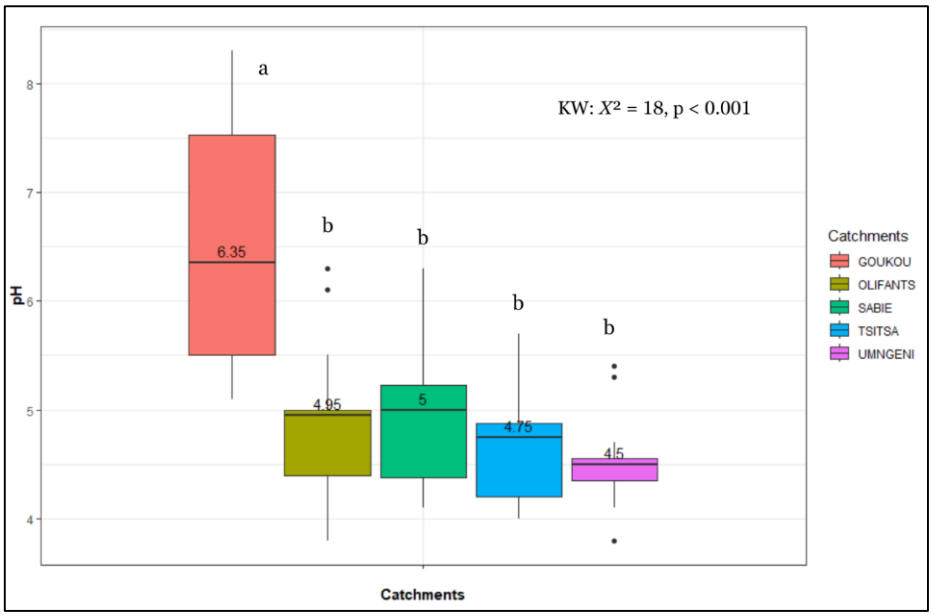


Figure 3-4: The pH distribution of the five catchments. The Kruskal-Wallis (KW) chi-squared (X^2) value, as well as superscript is seen to display the significant differences between the methods.

3.3.2 Soil organic carbon content relationships and pedotransfer functions

The SOC content relationships between the different methods for the five study sites clearly showed that the WB and TDC methods had the strongest relationships, with R^2 values larger than 0.9, except for the Goukou catchment (Table 3-2 and Figure 3-6). Overall, the WB and TDC methods had a very strong relationship, with an R^2 of 0.91.

Table 3-2: The relationships between the three analytical methods from the five different catchments, expressed as R^2.

Catchment	TDC vs LOI		WB vs LOI		WB vs TDC	
	R^2	p-value	R^2	p-value	R^2	p-value
Goukou	0.37	0.004	0.75	<0.001	0.56	<0.001
Umngeni	0.62	<0.001	0.39	0.002	0.96	<0.001
Olifants	0.7	<0.001	0.72	<0.001	0.90	<0.001
Sabie	0.93	<0.001	0.87	<0.001	0.97	<0.001
Tsitsa	0.88	<0.001	0.83	<0.001	0.97	<0.001

Previous studies also show this very strong relationship between the WB and TDC methods, with R^2 values > 0.86 (Jankauskas et al., 2005; Slepetiene et al., 2008; Gessesse & Khamzina, 2018;

Kumar et al., 2019). Only Roper et al. (2019) found a very similar R^2 of 0.55 between the TDC and WB methods, however, the reason for this moderate relationship was unclear for their study.

From the results, it was determined that the likely reason for the moderate relationship between WB and TDC for the Goukou catchment was due to the TDC method also measuring the SIC fraction of the soil, while WB only measured SOC content. The comparison between the WB and TDC methods revealed that at a SOC content of 2.5%, the deviation between the 1:1 line and the regression line was 0.2%, while at 5%, the deviation increased to 1 %. This indicates that the discrepancy between the two methods became more pronounced with increasing SOC content, caused by the incomplete oxidation of SOC content in the WB method (Wakley & Black, 1934; Kumar et al., 2019; FAO, 2019). Thus, the pedotransfer function (Equation 3-4) may be applicable for soils with less than 2.5% SOC content, as many of the values lie along the 1:1 line in this range (Figure 3-5). However, for higher values, it becomes increasingly inaccurate and should therefore not be considered as reliable. Therefore, the pedotransfer function between WB and TDC could only be used for soils with up to 2.5% SOC content.

Equation 3-4: $\mathbf{SOC_{WB} = -0.157 + 0.895\ x\ SOC_{TDC} - 0.0149\ x\ SOC_{TDC}^2 - 0.000606\ x\ SOC_{TDC}^3}$**,** where SOC_{WB} = SOC content determined by the Walkley-Black method and SOC_{TDC} = SOC content determined by the total dry combustion method.

Some studies have found a very strong relationship ($R^2 > 0.8$) between the WB and LOI methods (Jankauskas et al., 2005; Brunetto et al., 2006; Yerokun et al., 2007), as seen with the Sabie and Tsitsa catchments, however, overall, these two methods tend to have relatively poor to moderate relationships (Figure 3-6). This is usually attributed to the LOI method overestimating the SOC content (Hoogsteen et al., 2015; Roper et al., 2019), while the WB method tends to underestimate SOC content in some cases due to incomplete oxidation (Pribyl, 2010; Roper et al., 2019; FAO, 2019). This was also seen from Figure 6, with very high SOC values for the LOI method compared to the WB method. Therefore, the pedotransfer function (Equation 3-5) between the WB and LOI methods are not reliable enough to be used in South Africa.

Equation 3-5: $\mathbf{SOC_{WB} = 0.145 + 0.441\ x\ SOC_{LOI} + 0.0126\ x\ SOC_{LOI}^2 - 0.00075\ x\ SOC_{LOI}^3}$, where SOC_{WB} = SOC content determined by the Walkley-Black method and SOC_{LOI} = SOC content determined by the loss-on-ignition method.

Some studies have also found strong relationships between the TDC and LOI methods (De Vos et al., 2007; Abella & Zimmer, 2007; Fernandes et al., 2015; Jensen et al., 2018), as found at the Sabie and Tsitsa catchments. However, recent studies (Roper et al., 2019; Kumar et al., 2019) have also found poor to moderate relationships ($R^2 < 0.6$) between these two methods (Figure 3-7), due to the overestimation of SOC content with the LOI method. This overestimation was

also evident from Figure 3-8. Therefore, the pedotransfer function (Equation 3-6) between the TDC and LOI methods are also not reliable enough to be used in South Africa.

Equation 3-6: $\mathbf{SOC_{TDC} = 0.736 + 0.295\ x\ SOC_{LOI} + 0.0416\ x\ SOC_{LOI}^{2} - 0.000861\ x\ SOC_{LOI}^{3}}$, where SOC_{TDC} = SOC content determined by the total dry combustion method and SOC_{LOI} = SOC content determined by the loss-on-ignition method.

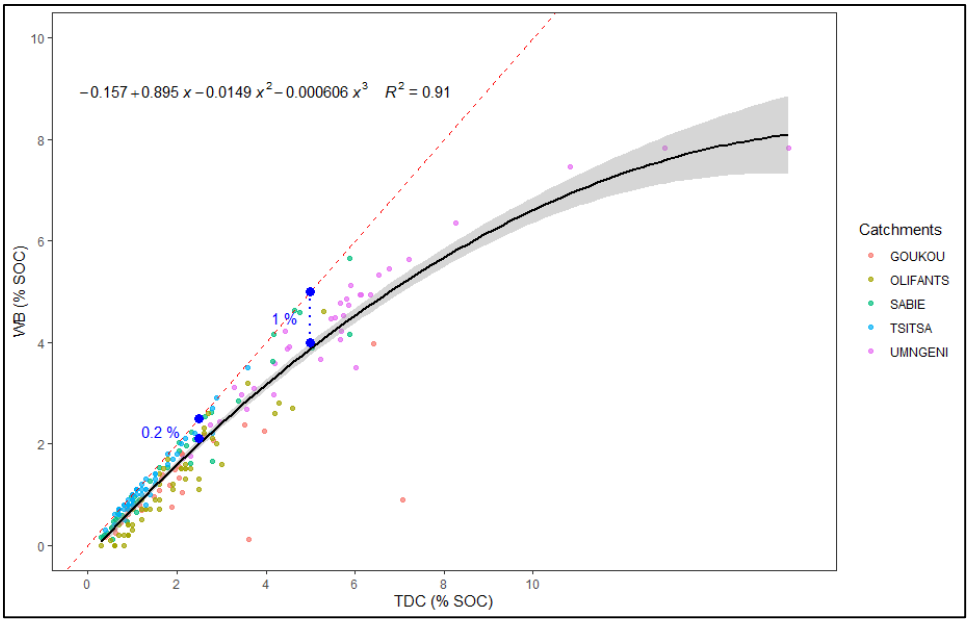


Figure 3-5: Scatter plot indicating the relationship between the Walkley-Black wet-oxidation (WB) method and the total dry combustion (TDC) method for SOC content. The red line indicates the 1:1 line.

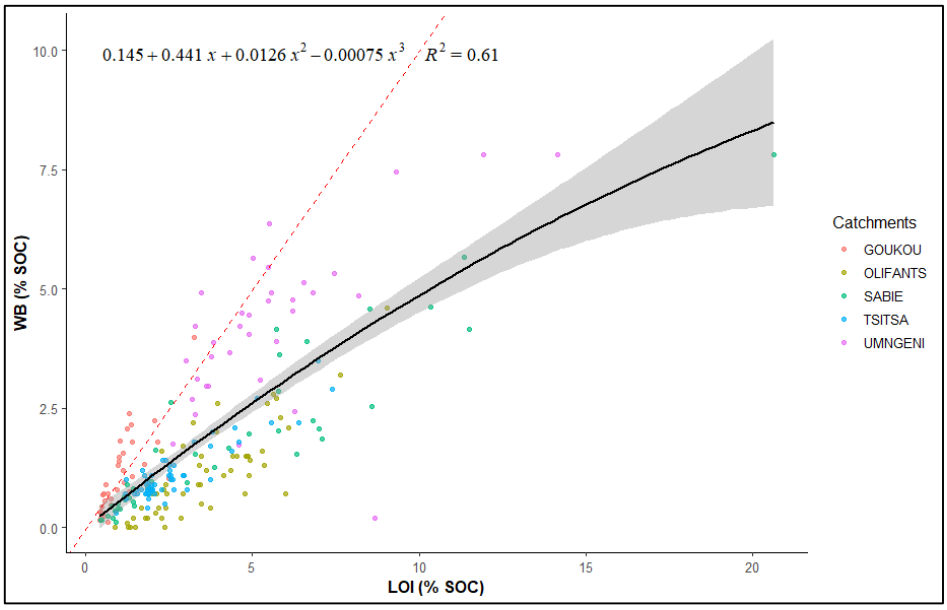


Figure 3-6: Scatter plot indicating the relationship between the Walkley-Black wet-oxidation (WB) method and the Loss-on-ignition (LOI) method for SOC content. The red line indicates the 1:1 line.

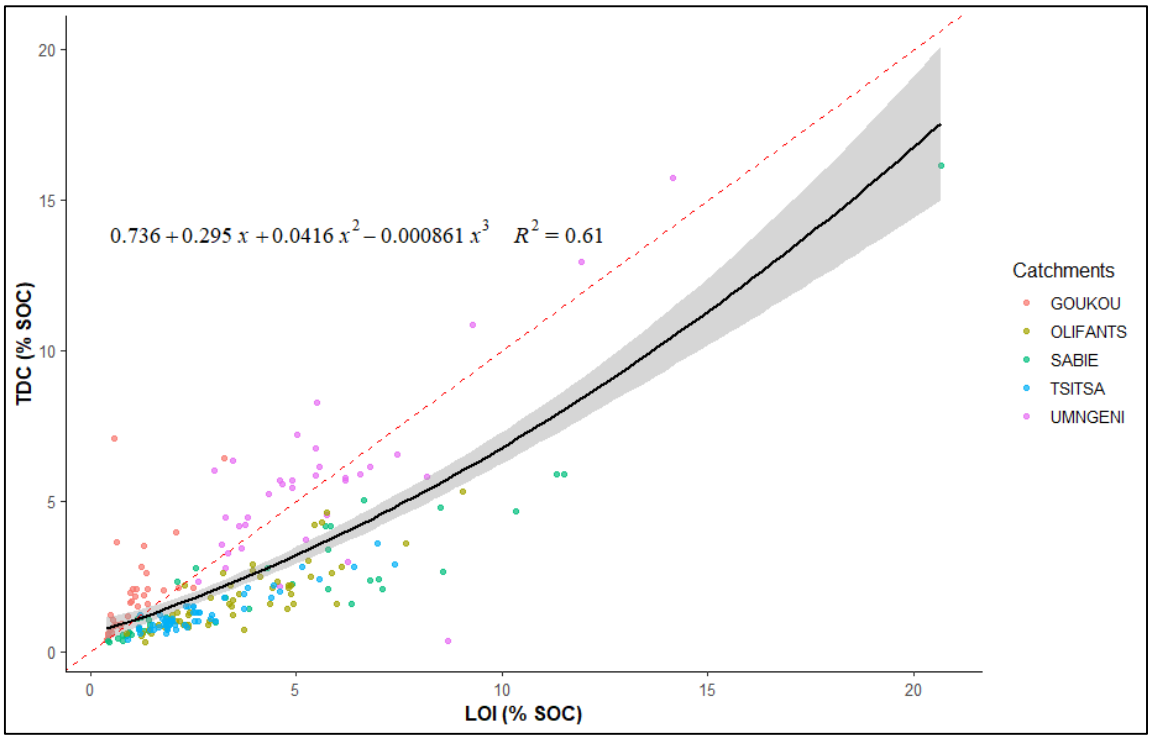


Figure 3-7: Scatter plot indicating the relationship between the total dry combustion (TDC) method and the loss-on-ignition (LOI) method for SOC content. The red line indicates the 1:1 line.

3.3.3 SOC content recovery factor

The samples (n = 220) had an RF median of 1.27, with statistically significant differences observed between the catchments (Figure 3-8). This value of 1.27, was remarkably close to the original RF (1.32) that was introduced by Walkley (1947) for quantifying the SOC content when external heat was not applied. The RF was also significantly lower compared to Sato et al. (2011) and Enang et al. (2018), which found RF ranges between 1.55 and 4.66. However, there were high variations of RF between all the catchments, with two outlier samples from the Goukou having an RF of 8 and 30, respectively (Figure 3-8). As mentioned, the TDC method likely measured the SIC content for the Goukou catchment, which led to the RF presenting these high RF values.

Except for the Goukou catchment, the variation in RF for the other catchments can likely be explained by the SOC content pools (active and resistant) and factors influencing the decomposition of organic material, such as vegetation, land use and climate. Bahadori and Tofighi (2017) concluded that soils with generally low SOC content (~ 1%) are distributed more in the resistant SOC fraction. This might be true for the Sabie and Tsitsa catchments, as these two catchments presented lowest SOC content (Figure 3-3) and the lowest RF of 1.26 and 1.11, respectively. It might therefore be assumed that soils with high SOC contents, such as the UMngeni catchment, probably have a larger active pool that increases the efficiency of the WB method (Kamara et al., 2007; Jha et al., 2014). However, the UMngeni catchment had a relatively

low RF (1.24) compared to the other catchments, which might be attributed to some organic materials, such as phenolic- and lignin compounds that resist oxidation at the temperatures obtained with sulphuric acid in the WB procedure (Díaz-Zorita, 1999; Bahadori & Tofighi, 2017). The large RF variation of the Olifants catchment was likely attributed to disturbances of agricultural practices, such as tillage, that have been known to increase the active SOC pool on crop fields and consequently lead to high RF values (Bahadori & Tofighi, 2017). Overall, the variation and significant differences between each catchment, indicate that the RF of 1.27 for the WB method cannot be used as a standard for South African soils.

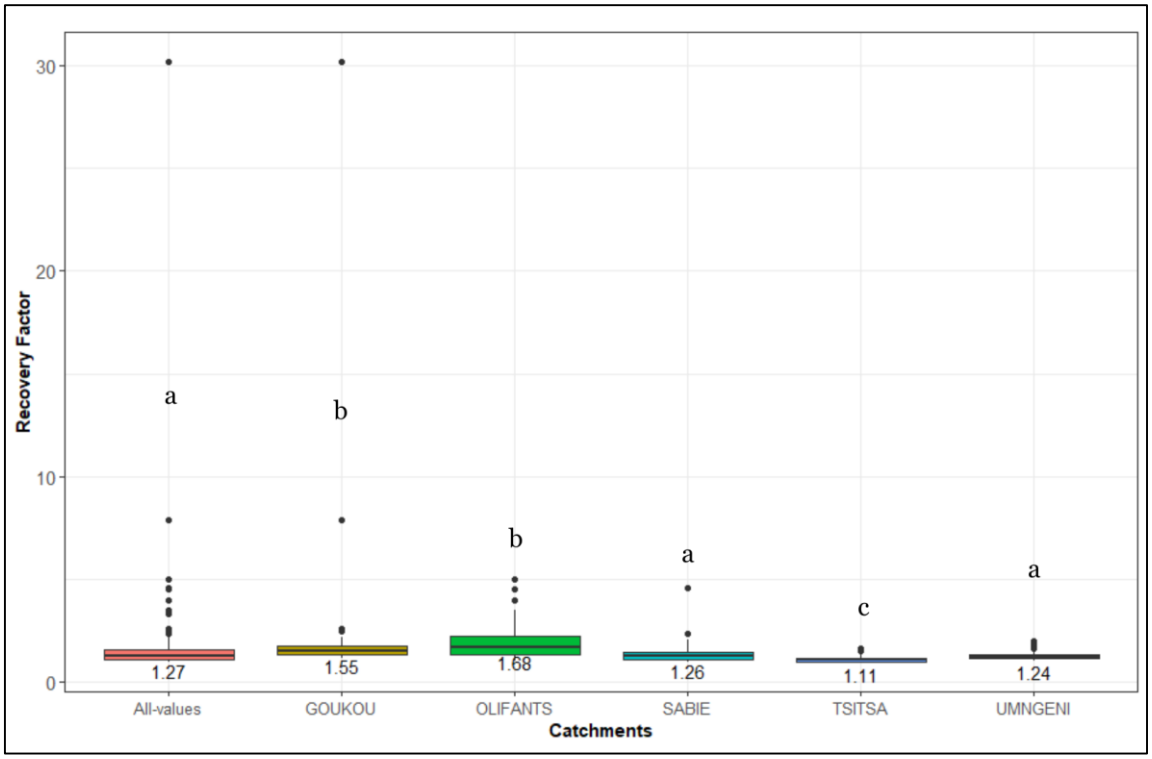


Figure 3-8: The SOC content recovery factor (RF) for the WB method of the five catchments. Superscript is seen to display the significant differences between the methods.

3.3.4 Implications for carbon credits

Figure 3-9 indicates the significant financial impact the SOC content measuring method can have on the carbon credit value of a 1-ha field. Overall, the LOI method did not provide a consistent trend when evaluated with the other two methods, with some samples indicating a higher carbon credit value (High RF) and some samples indicating a lower carbon credit value (High SIC and Low RF) (Figure 3-9). This is due to the inconsistent SOC measurement by the LOI method, attributed to the inconsistent correction factor of 1.724. Concerning the TDC method, the great overestimation of SOC content was due to the inclusion of SIC content and is evident for the high SIC sample, where a grossly inflated carbon credit value of $20 267 was calculated compared to a value of $3 263 for the WB method when utilising the RF (Figure 3-9). This clearly shows the

vast financial implications of using an analysis method outside of the intended setting. Therefore, when the TDC method is to be used, the SIC content must be removed from the soil sample, before determining the SOC content. In contrast, the high SOC sample shows the effect of the underestimation of the WB method. The TDC method presented a carbon credit value of $45 057, while the WB method presented a carbon credit value of $22 528 and a value of $28 597 when the RF was applied. This shows the implication of the incomplete oxidation when using the WB method. Both, the high and low RF samples indicated the need for a specific RF. In the high RF scenario, the TDC method had a considerably higher carbon credit value ($2 003) than both the WB ($572) and WB with the RF ($715), while in the low RF scenario, the TDC method presented a higher carbon credit value ($5 982) compared to the WB method ($5 210), but a lower value than the WB with the RF ($6 526) (Figure 3-9). This difference in carbon credit value of approximately $600/ha, shows the magnitude of the financial implications of working on a general RF, rather than the actual value measured with the TDC method.

As mentioned, the WB method can lead to the underestimation of SOC content, while the standard RF of 1.27 for the WB method is not suitable for South African soils. Therefore, the TDC method should be considered the preferred method for determining SOC content for carbon credit assessments. However, the SIC content needs to be removed through an acidic reagent, otherwise it could lead to gains or losses of up to thousands of dollars.

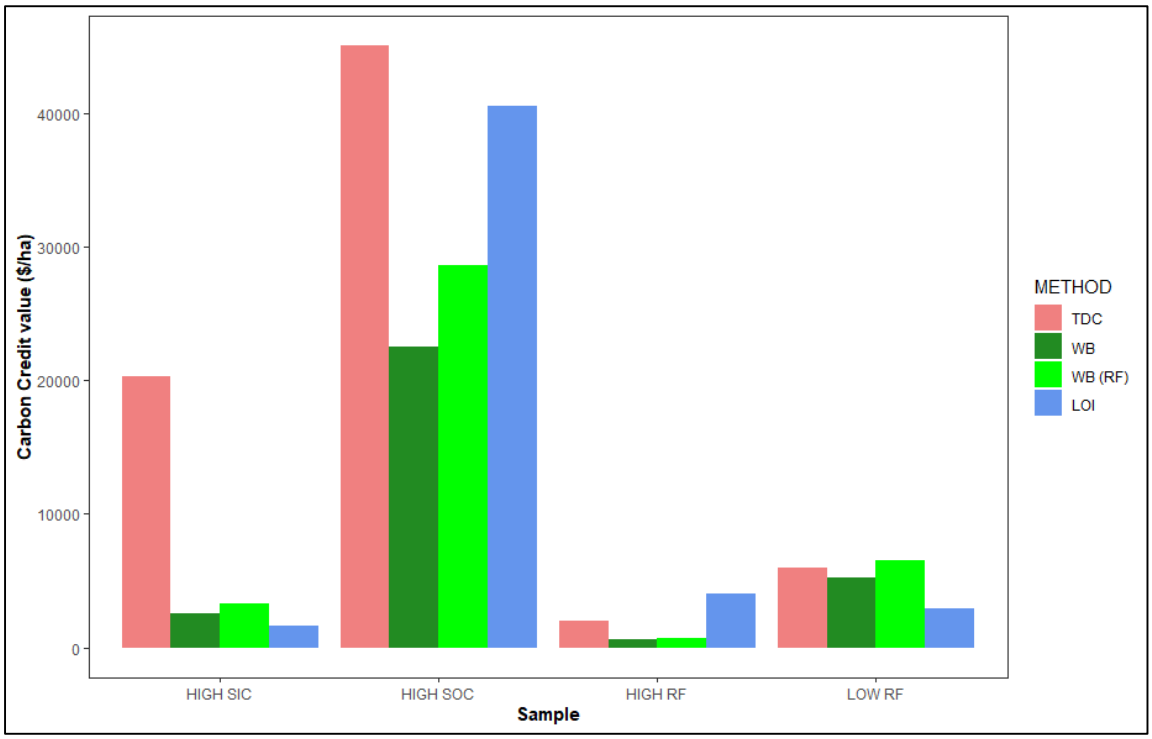


Figure 3-9: Implications of carbon credit value ($) when using the four specific samples with the three analytical methods.

3.3.5 Applicability to South African soil conditions

Overall, the study found that for South African conditions, the TDC method should be selected for determining SOC content and for accurate carbon credit assessment. The TDC method does unfortunately have a high cost and if the soil has a pH of 6.35 or higher, an acidic reagent (HCl) is needed to remove any free carbonates and prevent the measurement of SIC content. The next option should be the WB method, however, only for soils with an estimated SOC content < 2.5%, since after this point the method might underestimate the SOC content due to the incomplete oxidation. Although the median RF of 1.27 achieved, was very close to the original RF determined by Walkley (1947), the values did indicate variability, and it was clear that no standard RF could be applied for South African soils. The WB method does also have a high cost and has the highest environmental impact due to the use of hazardous chemicals (FAO, 2019). The TDC and LOI methods does however have a considerable energy usage, however, the TDC method has shown to have lower uncertainty and errors compared to the LOI and WB methods (Goidts et al., 2009). The study found that the LOI method provided the least reliable SOC content values, however, this method was the cheapest and is relatively simple to perform. As mentioned, the LOI method is still a widely used method for determining SOM, however, this method should be avoided when determining SOC content in South Africa for carbon credit assessment. The pedotransfer function created between the WB and TDC methods was acceptable for soils with a pH < 6.5, however, should be avoided when soils exhibit a higher pH or high SOC content.

3.4 Conclusion

The study found that the WB, TDC and LOI methods measured different SOC content values for the same samples. Measuring SOC content with different analytical methods could lead to significant financial implications and therefore a standard or consistent method is needed for determining SOC content for the assessment of carbon credits in South Africa.

The study also reached the second aim by successfully creating pedotransfer functions between all three methods. However, only the WB and TDC methods had a very strong relationship (R^2 = 0.91) and showed that accuracy start to decrease significantly after 2.5% SOC content. Therefore, this pedotransfer function (SOC_{WB} = -0.157 + 0.895 x SOC_{TDC} – 0.0149 x SOC_{TDC}^2 - 0.000606 x SOC_{TDC}^3) could be used for transferring SOC content values with SOC content up to 2.5%. The pedotransfer function between the LOI and WB methods (SOC_{WB} = 0.145 + 0.441 x SOC_{LOI} + 0.0126 x SOC_{LOI}^2 - 0.00075 x SOC_{LOI}^3), as well as the pedotransfer function between the TDC and LOI methods (SOC_{TDC} = 0.736 + 0.295 x SOC_{LOI} + 0.0416 x SOC_{LOI}^2 - 0.000861 x SOC_{LOI}^3) were found to be unreliable to be used in South Africa.

The LOI method should not be used for determining SOC content, due to the inconsistent correction factor of 1.724. Although the median RF of 1.27 was achieved, when the RF was applied, there were significant financial implications regarding carbon credit values. The variability of RF for the WB method between the different catchments were likely attributed to the different land use, vegetation and distribution of SOC pools. The study also found that the TDC method should still be considered the preferred method for determining SOC content for the assessment of carbon credits in South Africa. However, only in soils with a pH less than 6.35. Adding an acidic reagent to remove SIC should be considered an important step for determining reliable SOC content values. The WB method should be avoided if a soil is expected to have a high SOC content, while the LOI method could still be used for determining SOM, however, this method should be avoided when determining SOC content.

Future research should focus on improving the RF for the WB method for South African soils, as well as evaluating the conversion factor of the LOI method for South African soils. This study successfully provided guidelines for selecting the appropriate analytical method when determining SOC content in South Africa.

CHAPTER 4 CREATING A NEAR-INFRARED SPECTROSCOPY CALIBRATION ALGORITHM FOR SOIL ORGANIC CARBON CONTENT IN SOUTH AFRICA

4.1 Introduction

To mitigate carbon emissions into the atmosphere, soil carbon sequestration has been implemented globally (Lal, 2009) and has become a significant focus in recent years (Nair, 2012; Rhodes, 2017; Baurov, 2021; Zhang et al., 2021; Batool et al., 2024). Local land managers or farmers have the unique opportunity not only to sequester carbon in the soil, but to sell units of sequestered carbon, to the government or other businesses as carbon credits (Sharma et al., 2021).

Accurate quantification of soil organic carbon (SOC) content is an essential part of the accounting of these carbon credits. Conventional methods that are used for the quantification of SOC content, such as the Walkley-Black wet-oxidation (WB) and total dry combustion (TDC) methods, are costly, labour- and time-intensive (Walkley & Black, 1934; Nelson & Sommers, 1996; FAO, 2019). Alternatively, the use of infrared light reflectance as an advanced technology has shown promise as an easy, rapid and cost-effective alternative for SOC analysis (Soriana-Disla et al., 2014; Orr et al., 2017). Research on the use of near-infrared (NIR) diffuse reflectance spectroscopy in soil science has rapidly increased, with a focus on SOC content, texture, structure, and microbial activity (Wetterlind et al., 2008; Stenberg et al., 2010; Orr et al., 2017; FAO, 2019; Biney et al., 2020; Li et al., 2020; Long et al., 2023).

The NIR spectral signature of SOC is influenced by factors such as soil texture and moisture content (Ben-Dor et al., 1999; Stenberg et al., 2010; Tekin et al., 2014). Previous studies have shown that it is possible to scan soil samples in situ (also called field state), as opposed to dried and sieved soils as prepared for conventional laboratory analysis (Nocita et al., 2013; Tekin et al., 2012; Tekin et al., 2014; Biney et al., 2020). However, these studies reported significantly improved results for prediction of SOC content using laboratory prepared samples, as compared to field state samples.

Different parameters, such as sample moisture state, spectral pre-processing, sampling selection and machine learning models have been shown to affect the accuracy of NIR calibration algorithms. In recent years, simple random sampling (RS), K-means clustering and conditioned Latin Hypercube Sampling (cLHS) have gained more attention as the sampling selection algorithm (Hair et al., 2013; Wadoux et al., 2019; Li et al., 2020; Long et al., 2023). Overall, partial least squares regression (PLSR) has been the mostly used model for developing NIR calibration

algorithms of SOC content (Cambule et al., 2012; McDowell et al., 2012; Stevens et al., 2013; Deng et al., 2013; Li et al., 2020; Long et al., 2023). However, Artificial Neural Networks (ANN), Random Forest (RF), Support Vector Machines (SVM) and Cubist have also been used successfully for developing NIR calibration algorithms of SOC content (McDowell et al., 2012; Stevens et al., 2013; Tekin et al., 2014; George & Kumar, 2019).

The largest freely available spectral library is the Open Soil Spectral Library (OSSL), which has an estimation service for utilising calibrated prediction models from open-source mid-infrared (MIR), Vis-NIR and NIR spectral data (Safanelli et al., 2023). Currently, prediction models for SOC content using the MIR and Vis-NIR are outperforming the NIR prediction models (Safanelli et al., 2023). However, the prediction of SOC content has shown to be very accurate using the NIR prediction models (root mean square error = 0.27% and Lin's concordance correlation coefficient = 0.97), on samples entered in the spectral library. Currently, South African soils are highly underrepresented (n = 64) in the OSSL global dataset. These samples are from two distinct regions (western, arid region and humid coastal region), leading to inaccurate predictions (Kock et al., 2024). The gap must be addressed to determine whether the OSSL can be used to predict SOC content for South African soils.

In South Africa, only two previous studies have focused on developing NIR spectroscopy calibration algorithms for SOC content (Nocita, 2009; Zahinda, 2020). Importantly, these algorithms have been developed on small areas and are only relevant and usable for those specific areas. Currently, there is no standard NIR calibration algorithm that could be used for predicting SOC content for South African soils. The aim of the study was to develop a NIR spectroscopy calibration algorithm of SOC content that could be used for South Africa.

The objectives of the study included:

- To determine whether the OSSL could be used for SOC content prediction of South African soils.
- To determine the effect of parameters, such as sample moisture state, spectral pre-processing, sampling selection and machine learning models on the creation of NIR calibration algorithms.
- To assess the use of a NIR spectrometer as a field instrument by comparing samples scanned in situ (called field state) with samples scanned after laboratory preparation (called laboratory prepared).
- To evaluate the performance of the developed NIR calibration algorithms at a regional scale and at local scales.

4.2 Materials and methods

4.2.1 Study area and site description

Soil samples were collected from five different study sites across South Africa to provide a diversity of soil properties, including SOC content (Figure 4-1, Table 4-1). The calibration algorithms were developed at the field scale for the Ottosdal and Vrede sites, at the catchment scale for the Olifants-, Sabie-Sands- and Tsitsa catchments, and at the regional scale, where all collected data were combined. The three catchments included several different and unique soil-forming factors, such as climate, land use and geology, which led to the representation of almost all South African soils. The unique soil-forming factors mainly influencing South African soils included were arid and semi-arid to humid climates (Schulze, 2007) and varying parent material. At field scale, two fields from farms, near the towns of Ottosdal and Vrede were selected. The Ottosdal farm receives a median annual precipitation (MAP) of 565 mm (Schulze, 2007), is overlain by the Klipriviersberg Group (Council for Geoscience, 2019) and the soils are typically sandy (Acrisols). In contrast, the Vrede farm receives 795 mm MAP (Shulze, 2007), is overlain by the Balfour Formation (Council for Geoscience, 2019) and the soils are typically clayey (Gleysols and Planosols). Both farms work on dryland, two-year crop rotation of soybeans and yellow maize.

At the catchment scale, three different catchments were selected, as part of a project funded by the Water Research Commission (C2020/2021-00455). The Olifants catchment is 1 567 km² in size and receives a low MAP of 691 mm in the north and 787 mm in the south (Shulze, 2007). The catchment area consists of various bedrock lithologies including sandstones, shales, quartzites, and dolomites (Council for Geoscience, 2019). The land use is predominantly coal mines, commercial cultivated areas, and natural grassland (GeoTerra Image, 2020). The soils of the Olifants catchment are generally sandy topsoil overlying a plinthic subsoil (Plinthosols), with shallow soils (Leptosols) occurring over large areas of the catchment (Land Type Survey Staff, 1972-2006).

The Sabie-Sands catchment is 5 790 km² and receives a MAP of 540 mm in the eastern part and 1 200 mm in the western part of the catchment. The catchment area consists of various bedrock lithologies including, quartzites, granites, basalts, conglomerates, andesites, gneiss and shales (Council for Geoscience, 2019). The land use varies from natural vegetation in the Kruger National Park and Sabie-Sands Game Reserve in the east, commercial irrigated agricultural areas in the lower lying areas and forest plantations in the west (GeoTerra Image, 2020). Large areas of this catchment contain soils that are sandy and deep (Acrisols), while small parts of the catchment contain shallow soils (Leptosols) (Land Type Survey Staff, 1972-2006). In some areas, soils overlying relatively impermeable bedrock (Arenosols) can also be found.

The Tsitsa catchment is 494 km² in size and receives a MAP of approximately 800 mm in the north and approximately 950 mm in the south (Shulze, 2007). The catchment area consists of various bedrock lithologies including sandstones, mudstones, dolomites, and shales (Council for Geoscience, 2019). The land use varies from being predominantly built-up land (villages), natural grassland, commercial cultivated areas, and fallow lands, while there are also small distributions of forested land within the catchment area (GeoTerra Image, 2020). Large areas of this catchment also contain soils that are sandy and deep (Acrisols), while small parts of the catchment contain shallow soils (Leptosols) (Land Type Survey Staff, 1972-2006). Soil erosion is a major land degradation problem across South Africa but is particularly severe in the Tsitsa catchment, which has led to a decrease in soil and consequently SOC content in the catchment (Ighodaro et al., 2012; Parwada & Van Tol, 2016; du Plessis et al., 2020).

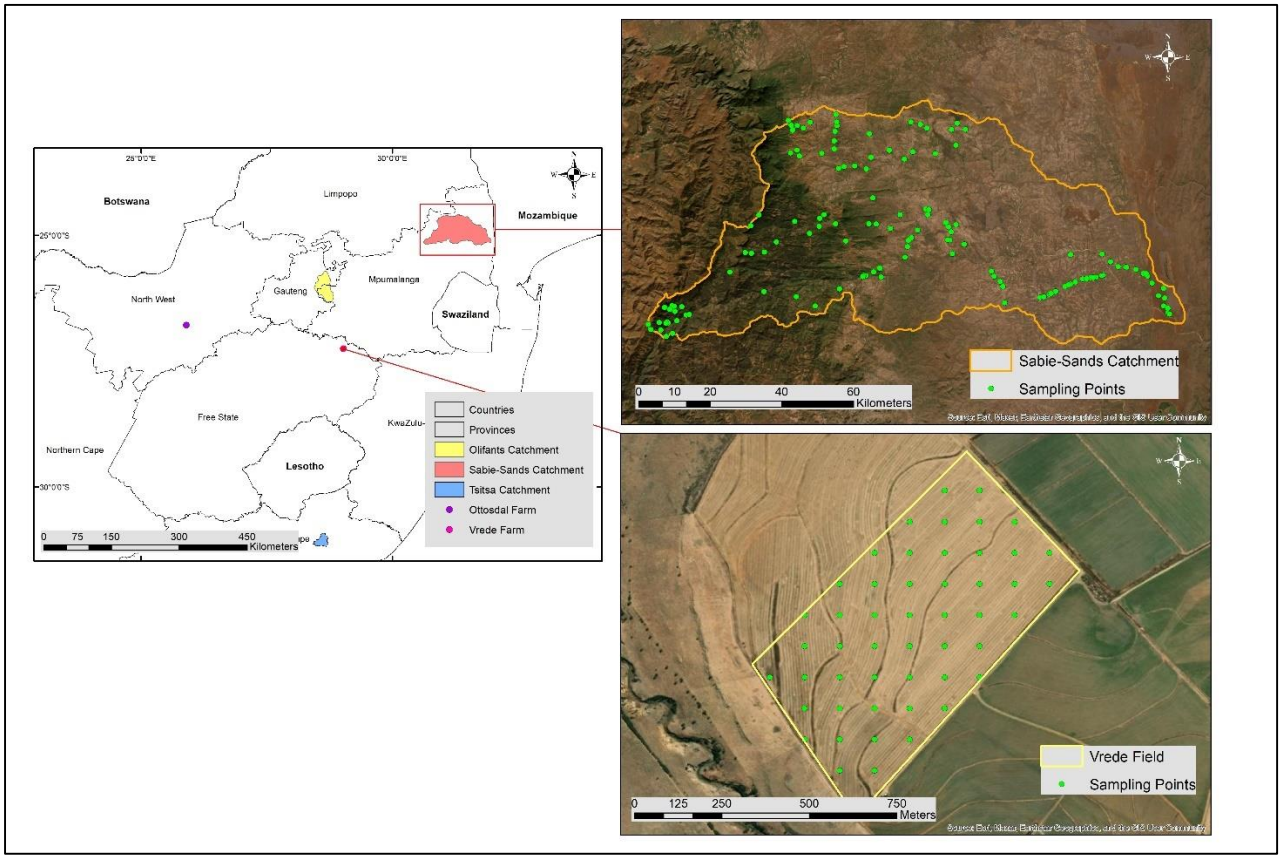


Figure 4-1: **Location of the five different study sites and an example of the sampling points within the Sabie-Sands catchment and Vrede field.**

4.2.2 Soil sampling and analysis

At the field scale, 50 sampling points were placed on a 1 ha grid to cover an area of 50 ha and the sampling was conducted after harvesting, between July and October 2022 (Figure 4-1). Samples were collected with a hand auger at the following depths: 0–5 cm, 5–15 cm, 15–30 cm, and 30–60 cm (Table 4-1).

At the catchment scale, the sampling locations in each catchment were selected using cLHS (Minansy & McBratney, 2006), using slope, elevation, geology, median annual precipitation (MAP) and Normalized Difference Vegetation Index as covariates (Figure 4-1). This approach ensured that the attribute space of the different catchments was covered and should represent the soil diversity accordingly. Sampling for the different catchments took place at various stages throughout the year and were conducted between December 2021 and November 2022. For the catchments, only topsoil samples (0-15 cm) were collected also using a hand auger (Table 4-1).

All samples were scanned (once with a scan time of 14 seconds) in a field state, with a NeoSpectra hand-held NIR Analyzer developed by Si-Ware, with a spectral range of 1 350 – 2 550 nm. Afterwards, the samples were brought to the laboratory, air-dried and passed through a 2 mm sieve. The samples were then analysed for SOC content with the total dry combustion (TDC) method using a CN elemental analyser (Nelson & Sommers, 1996). After the TDC analysis, all the samples were scanned again in the prepared state.

Table 4-1:The total number of soil samples (*n*) taken from each study site, as well as the total area of each study site (km²) and respective scale.

Study site	*n*	Area (km²)	Scale
Ottosdal field	200	0.5	Field
Vrede field	200	0.5	Field
Olifants catchment	51	1 567	Catchment
Sabie-Sands catchment	40	5 790	Catchment
Tsitsa catchment	47	494	Catchment

4.2.3 OSSL method

To compare how local data will perform using global prediction models, the global spectral dataset from the OSSL was used. The methodological outline for the OSSL can be found in Safanelli et al. (2025), where the best performing model was selected and statistical metrics utilised, including root mean square error (RMSE) (Viscarra Rossel et al., 2006; Wadoux et al., 2021), Lin's concordance coefficient (ρc) (Wadoux et al., 2021), ratio of performance to inter-quartile distance (RPIQ) (Zhang et al., 2020) and coefficient of determination (R^2) (Peng et al., 2015; Wadoux et al., 2021).

The online server for OSSL predictions was utilised and the entire laboratory prepared dataset was uploaded to the estimator and was specified as NIR in the Web application. The "Carbon, Total NCS" option was selected and the prediction values for each sample were obtained. The

OSSL ran a principal component analysis test to determine whether the spectral signatures of the uploaded samples were represented in the OSSL database. The predictions made by the OSSL models were exported and compared with the actual observed values from the soil analysis. To evaluate the OSSL predictions, the following statistical validation metrics were used: RMSE, ρc, RPIQ and R^2.

4.2.4 Creation of calibration algorithms on regional and local scale

4.2.4.1 Soil spectral library and algorithm development

A soil spectral library was established by combining all the NIR spectral data (Figure 4-2) with the SOC content data acquired from the analysis. The integration of the spectral data and SOC content data was done in R Studio using the R programming language (R Core Team, 2020).

On a regional scale, all the samples from the three catchments were used, while only the topsoil samples from the two field sites were used, to avoid skewing the dataset (n = 238). At the catchment scale, all the available samples for each catchment were used (Tsitsa, n = 47; Olifants, n = 51; Sabie-Sands, n = 40). For the field scale all the samples at every depth collected for each field (n = 200) were used, as all the samples came from the same sites.

Four different parameters were tested, including:

(1) the field state vs prepared state,

(2) the different pre-processing methods of spectral data vs no pre-processing of spectral data,

(3) sample selection algorithms to divide the spectral library into training and validation datasets, and

(4) using different machine learning models.

Spectral cleaning and pre-processing included outlier removal (sample discarded) and mean centering standardization (column-wise centering) (Wadoux et al., 2021), which was done to maintain the integrity of spectral features and their relationships and to avoid potential distortions of other pre-processing techniques. To divide the spectral library into a training and validation dataset, RS (De Gruijter et al., 2006), K-Means clustering (K-Means) (Hair et al., 2013), and conditioned Latin Hypercube Sampling (cLHS) (Minasny & McBratney, 2006) were applied, using the spectral data as input. For RS, the “sample” (R Core Team, 2020) function was used, while for K-means and cLHS, the “kmeans” (Kassambara & Mundt, 2020) and “clhs” (Roudier et al., 2022) functions were used, respectively. The spectral library was split with all three methods into a 70:30, calibration to validation ratio datasets. The validation dataset was used to evaluate the

calibration models (Bai et al., 2022; Kock et al., 2024). The different empirical machine learning models used included Partial Least Square Regression (PLSR) (Janik et al., 2007; Dangal et al., 2019; Comstock et al., 2019; Metzger et al., 2020; Wadoux et al., 2021), RF (Breiman, 2001; Knox et al., 2015; Wadoux et al., 2021), Support Vector Machines (SVM) (Peng et al., 2014; de Santana et al., 2021), Cubist (Quinlan, 1992; McBratney et al., 2003; Peng et al., 2015; Dangal et al., 2019; Wadoux et al., 2021), and Artificial Neural Networks (ANN) (Rosenblatt, 1958; Tekin et al., 2014; George & Kumar, 2020). No special tuning was done for any machine learning model and standard conditions, rules and parameters were used for each model (Wadoux et al., 2021). With the testing of these four parameters, 60 calibration algorithms were developed for the regional scale, using the entire soil spectral library training dataset. For ease of reference, each unique parameter configuration was allocated a code (Table 4-2).

Table 4-2: An example of the code for the different calibration algorithms used in the study.

	n	Sample state	Processing	Sampling design[1]	ML models[2]	Code
Options		Field	Pre-processing	RS	PLSR	
		Prepared	Raw	cLHS	RF	
				K-Means	SVM	
					Cubist	
					ANN	
Examples	1	F	Pre	RS	PLSR	F-Pre-RS-PLSR
	2	P	Raw	cLHS	ANN	P-Raw-cLHS-ANN

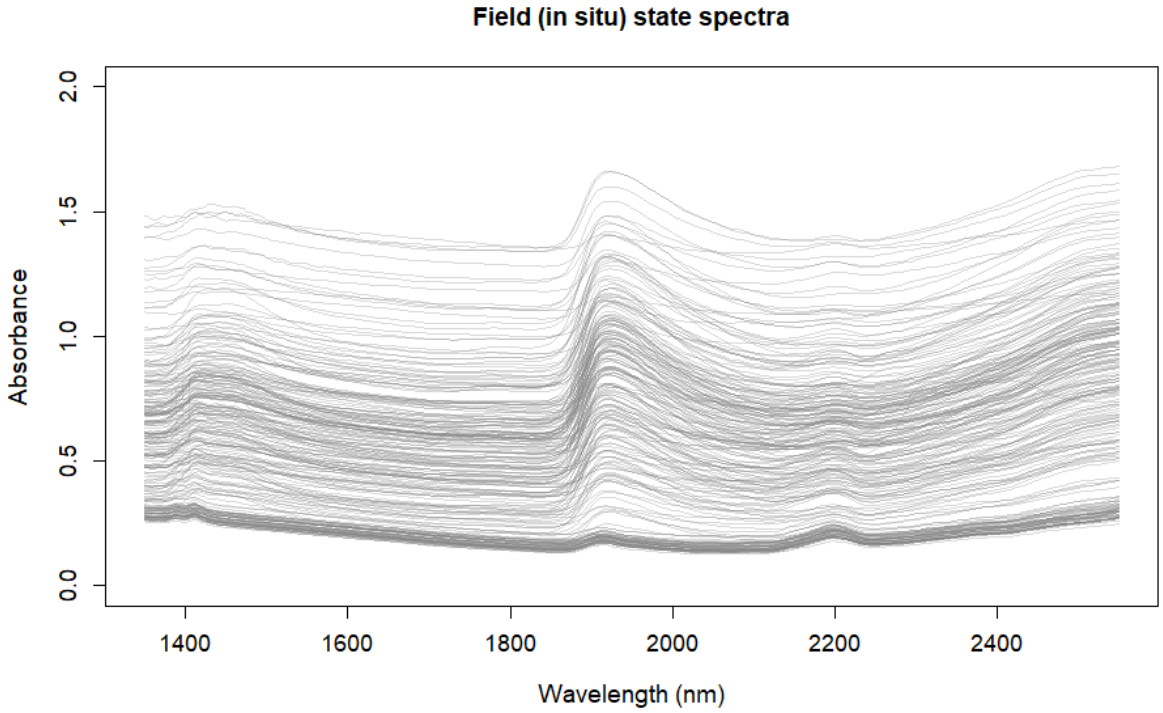
[1] RS = stratified random sampling, cLHS = conditioned Latin Hypercube Sampling, K-Means = K-Means clustering.
[2] PLSR = Partial Least Squares Regression, RF = Random Forest, SVM = Support Vector Machines, ANN = Artificial Neural Networks.

4.2.4.2 Calibration algorithm evaluation

The different calibration algorithms were evaluated on both the calibration and validation datasets by computing the RMSE, ρc, RPIQ and R^2. Additionally, scatterplots showing the observed and predicted values were also developed for the best algorithms to visually display the accuracy of the results. To assess the performance of the developed calibration algorithms at catchment- and field scale, the overall best performing algorithm at the regional scale (also called standard) was applied to the different study sites and validated as mentioned above. Additionally, the best performing algorithm for each individual study site was also determined and validated as mentioned above.

Although there are currently none, specific metrics for assessing if calibration algorithms are reliable, previous studies (Cambule et al., 2012; Camacho-Tamayo et al., 2013; Li et al., 2020; Wadoux et al., 2021) have found that calibration algorithms presenting an $R^2 > 0.7$ and an RPIQ > 1.4 were deemed reliable. However, the calibration algorithms had to have an RMSE < 0.3% on the regional scale and RMSE < 0.2% on catchment and field scale to be considered accurately usable for future SOC content prediction.

(a)



(b)

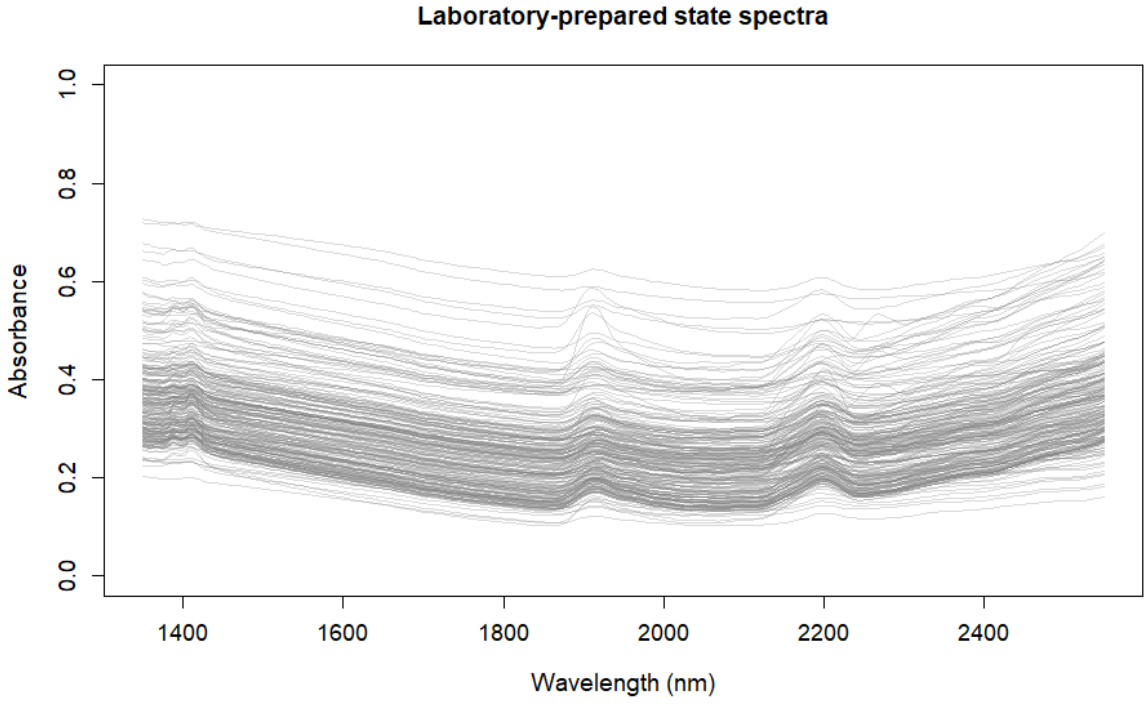


Figure 4-2: The near-infrared (NIR) raw spectral data obtained from scanning the soil samples in the (a) field and (b) prepared state with the NeoSpectra handheld instrument.

4.2.4.3 Evaluation of methodological parameters

To determine the effect of the four methodological parameters, a Kruskal-Wallis test (Kruskal & Wallis, 1952) was performed on the RMSE validation values, as normality was not assumed. Box-and-whisker plots (with error bars) were developed to visually determine the effect of the four methodological parameters on the RMSE validation values.

4.3 Results

4.3.1 Soil organic carbon content distribution and near-infrared spectra

The SOC content distribution between the study sites was relatively high, varying between 0.14 and 5% (Figure 4-3). Overall, the SOC content at the field scale was the lowest (medians < 1%) compared to the catchment and regional scales. The Tsitsa catchment also had the lowest SOC content distribution (median = 1.04%) for all catchments, while the Ottosdal study site had the overall lowest SOC content distribution (median = 0.47%).

The absorbance values for the study sites in the field state were generally higher (less negative) compared to the values in the laboratory-prepared state (Figure 4-4a and b). The spectral peaks of the study sites for the laboratory-prepared state were also less pronounced compared to the peaks for the field state. The most pronounced peaks were seen around 1400-, 1900-, and 2200 nm wavelengths, with the 1900 nm wavelength from the field state presenting the most pronounced peak. The mean raw spectral data, scanned in the field state and in a laboratory-prepared state, indicated various absorbance values and spacing for each study site. For the field state, the Tsitsa study site presented the highest absorbance and the Ottosdal study site presented the lowest absorbance (Figure 4-4a), while for the laboratory-prepared state the Olifants study site presented the highest absorbance and the Ottosdal study site presented the lowest absorbance (Figure 4-4b). The spectral peaks for both states were generally aligned across all study sites.

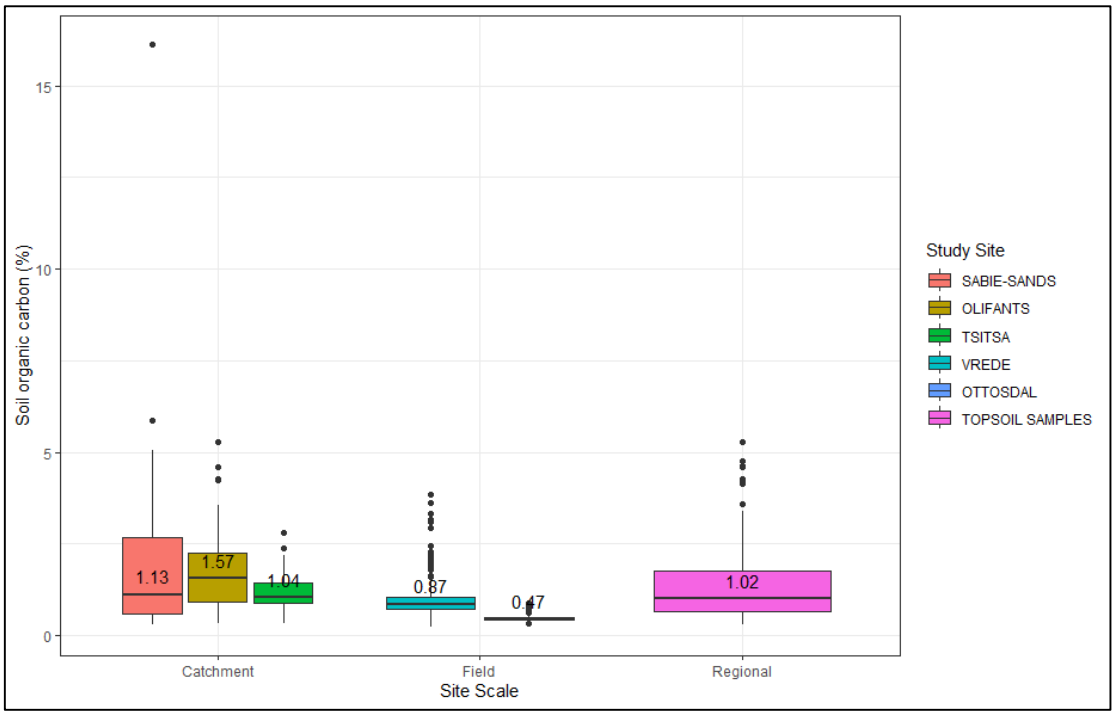
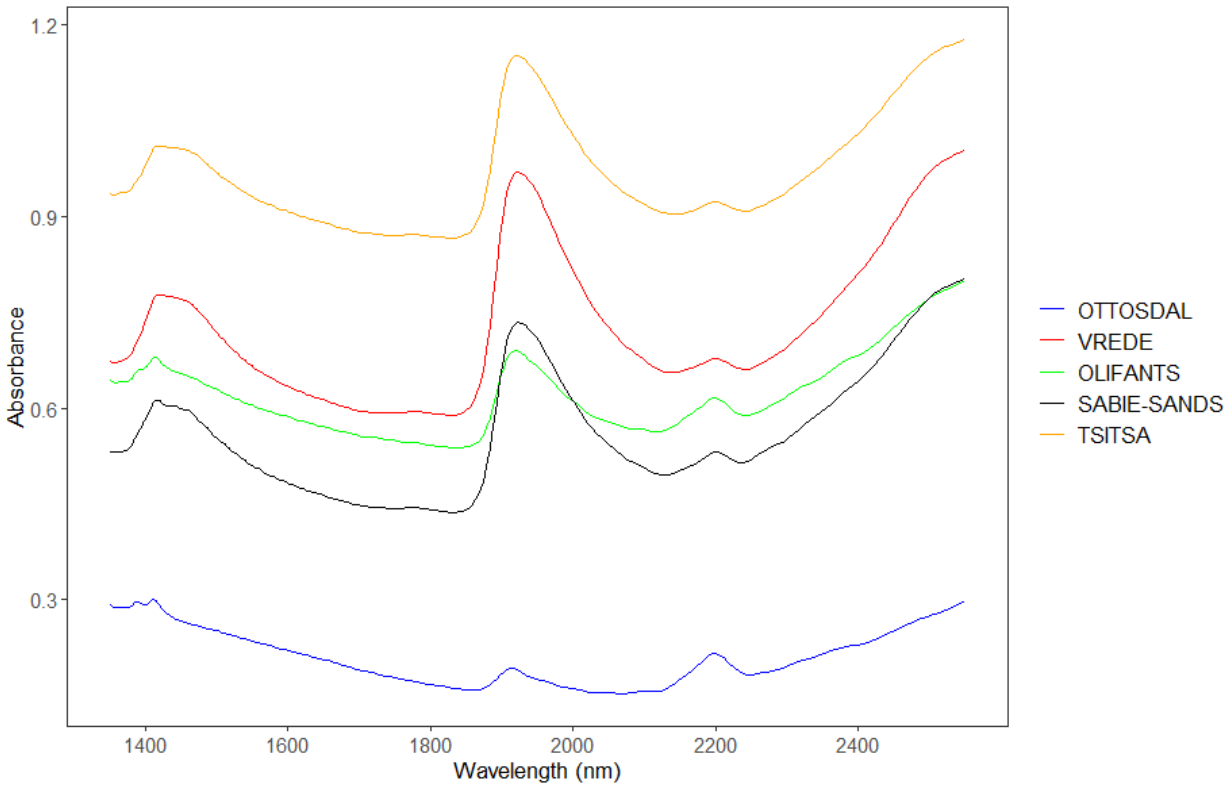


Figure 4-3: The soil organic carbon (SOC) content (%) distribution at the three scales.

(a)



(b)

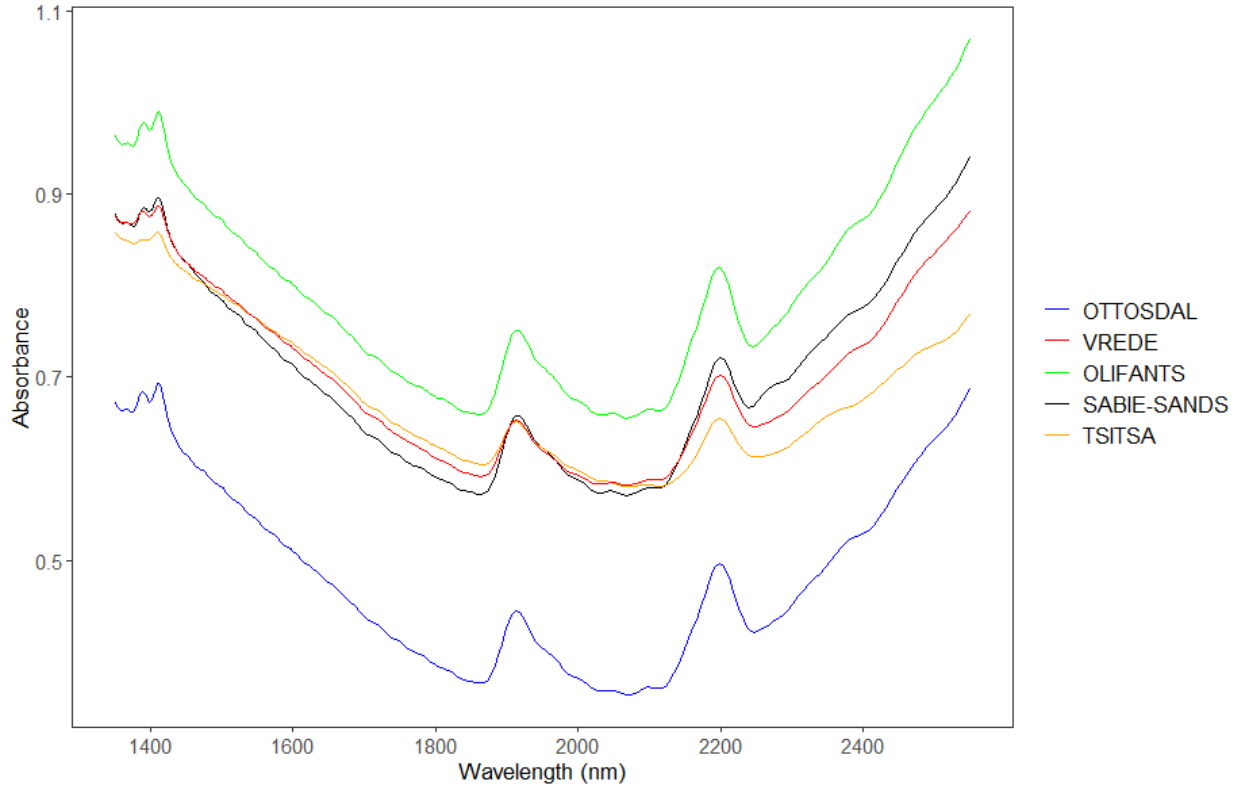


Figure 4-4: The NIR mean raw spectral data for each study site obtained from scanning the soil samples in the (a) field and (b) laboratory-prepared state with the NeoSpectra handheld instrument.

4.3.2 OSSL distribution

From the OSSL library distribution, most of the samples were unrepresented (indicated as False in Figure 4-5) in the dataset, which led to only a few samples being accurately predicted (indicated as True in Figure 4-5). As most of the sample points are situated above the 1:1 line ($\rho c = 0.21$), the OSSL calibration algorithm significantly underpredicted the SOC content, with observed SOC content higher than predicted SOC content (Figure 4-6). The OSSL prediction also presented RPIQ values < 1, R^2 values < 0.3 and an RMSE of 1.23% (Figure 4-6).

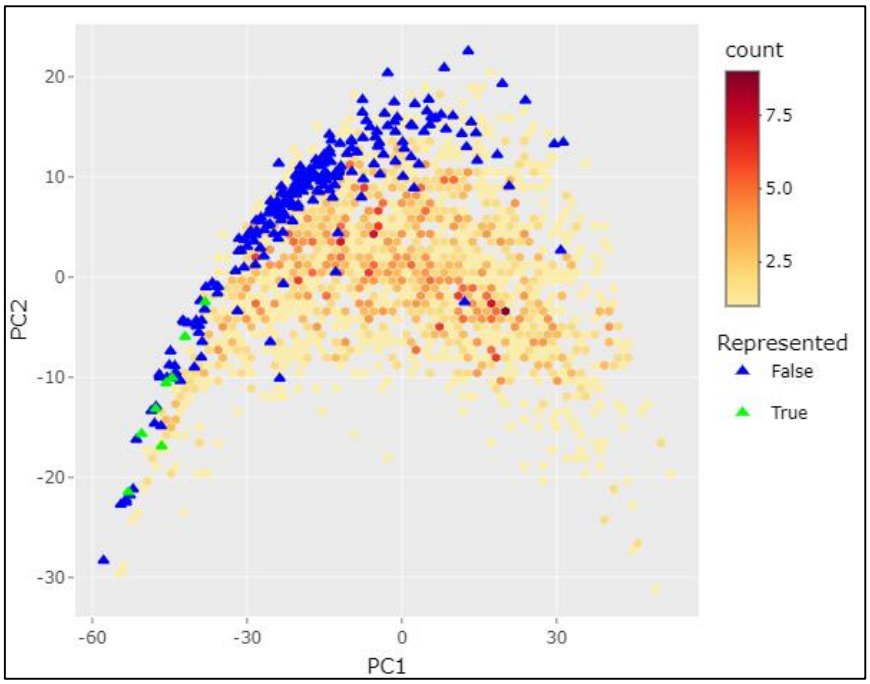


Figure 4-5: Scatter plot indicating the principal component analysis (PCA) of NIR Neospectra Open Soil Spectral Library (OSSL) and the representativeness of the SOC content from the laboratory prepared dataset.

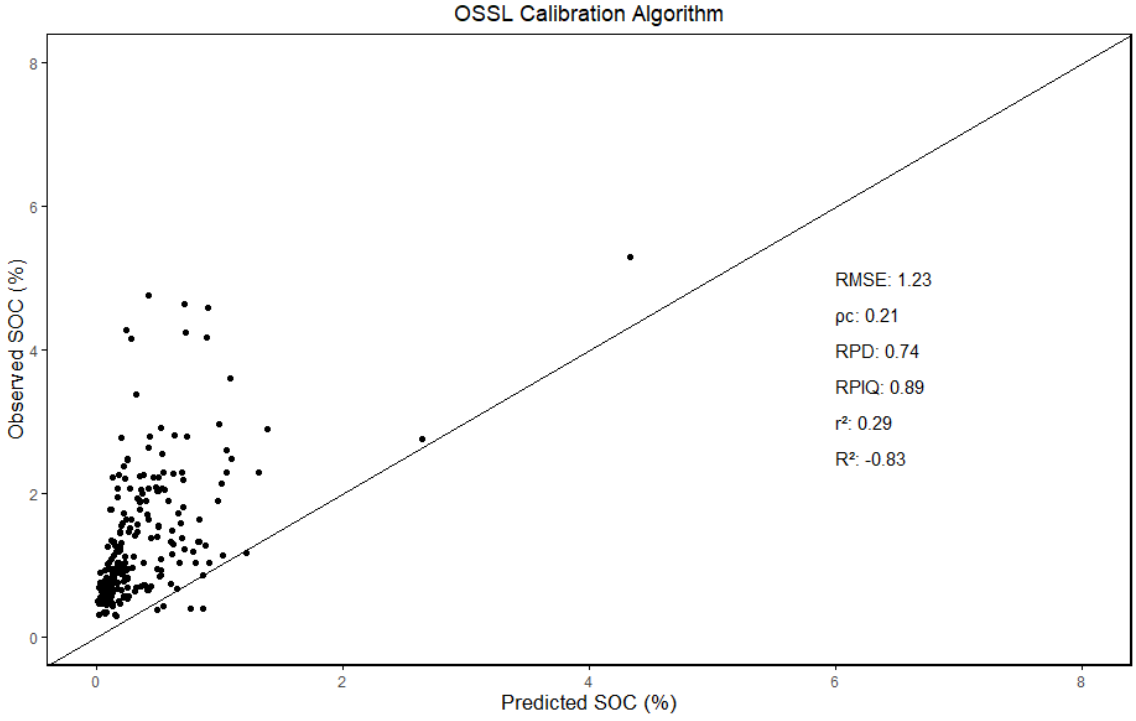


Figure 4-6: Scatter plots indicating the Open Soil Spectral Library (OSSL) calibration algorithm that was used as model prediction for soil organic carbon (SOC) content. The black line represents the 1:1 line.

4.3.3 Regional scale calibration algorithm

4.3.3.1 Developed calibration algorithms

Overall, the developed calibration algorithms at the regional scale for the field state indicated poor RMSE values (> 0.5%), poor to moderate (0–0.5) level of explanatory power (indicated by R^2) and weak to moderate (0–0.6) positive relationships (ρc) (Table 4-3, Figure 4-7). The three best performing calibration algorithms at the regional scale for the field state (algorithms 8, 9 and 19 in Table 4-3) had RMSE values between 0.69 and 0.75%, while the best performing algorithm had a ρc of 0.74, and RPIQ values of 1.39, respectively.

The calibration algorithms for the laboratory-prepared state showed similar results (Table 4-4, Figure 4-8). However, the best performing algorithms (algorithms 34, 39, 49 and 54) showed moderate RMSE values (< 0.5%), moderate to high (0.6–0.8) level of explanatory power, and strong (> 0.8) positive relationships (ρc). The overall best calibration algorithm (49) had an RMSE of 0.39% and RPIQ > 2.

The developed calibration algorithms for the field state and laboratory-prepared state also presented variable results when comparing the calibration statistics with the validation statistics, with some presenting optimal calibration statistics but not validation statistics and vice versa. Overall, there was no consistent trend found when comparing the calibration and validation statistics for the developed calibration algorithms.

Table 4-3: The calibration algorithms developed for the field state (n = 238) for SOC content in South Africa. The green highlighted calibration algorithms indicate the best performing algorithms.

Algorithm number	Code	Validation statistics			
		RMSE	R^2	ρc	RPIQ
1	F-Pre-RS-PLSR	0.9	0.18	0.41	1.07
2	F-Pre-RS-RF	0.8	0.34	0.62	1.19
3	F-Pre-RS-SVM	0.78	0.38	0.51	1.22
4	F-Pre-RS-Cubist	0.77	0.4	0.63	1.24
5	F-Pre-RS-ANN	0.91	0.15	0.44	1.05
6	F-Pre-cLHS-PLSR	0.86	0.02	0.34	1.08
7	F-Pre-cLHS-RF	0.84	0.08	0.49	1.11
8	F-Pre-cLHS-SVM	0.69	0.38	0.53	1.34
9	F-Pre-cLHS-Cubist	0.72	0.32	0.56	1.29
10	F-Pre-cLHS-ANN	0.89	-0.02	0.46	1.05
11	F-Pre-K-Means-PLSR	1.89	0.08	0.13	0.62
12	F-Pre-K-Means-RF	1.76	0.2	0.28	0.67
13	F-Pre-K-Means-SVM	1.83	0.14	0.21	0.65
14	F-Pre-K-Means-Cubist	1.83	0.14	0.22	0.65
15	F-Pre-K-Means-ANN	1.84	0.13	0.22	0.64
16	F-Raw-RS-PLSR	0.93	0.15	0.36	1.12
17	F-Raw-RS-RF	0.9	0.21	0.46	1.16
18	F-Raw-RS-SVM	0.92	0.18	0.29	1.14
19	F-Raw-RS-Cubist	0.75	0.45	0.74	1.39
20	F-Raw-RS-ANN	1.93	-2.62	0.41	0.54
21	F-Raw-cLHS-PLSR	0.9	0.15	0.34	1.35
22	F-Raw-cLHS-RF	0.97	0.02	0.39	1.25
23	F-Raw-cLHS-SVM	0.91	0.13	0.27	1.33
24	F-Raw-cLHS-Cubist	0.8	0.33	0.59	1.52
25	F-Raw-cLHS-ANN	0.98	0	0	1.24
26	F-Raw-K-Means-PLSR	1.82	0.09	0.15	0.64
27	F-Raw-K-Means-RF	1.85	0.05	0.14	0.63
28	F-Raw-K-Means-SVM	1.85	0.06	0.12	0.63
29	F-Raw-K-Means-Cubist	1.71	0.19	0.31	0.68
30	F-Raw-K-Means-ANN	1.91	0	0	0.61

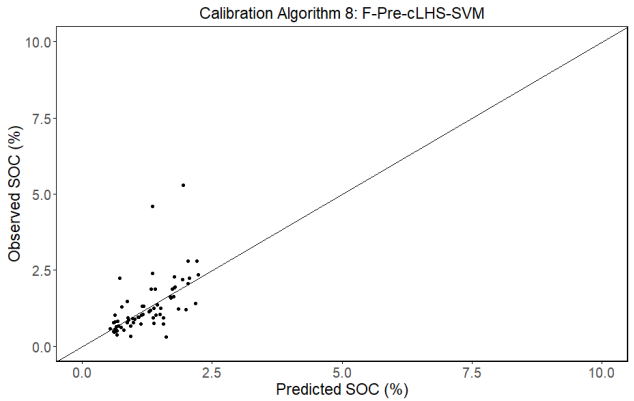
F = field state, Pre = preprocessing, RS = simple random sampling, cLHS = conditioned Latin Hypercube Sampling, K-Means = K-Means clustering, PLSR = Partial Least Squares Regression, RF = Random Forest, SVM = Support Vector Machines, ANN = Artificial Neural Networks.

Table 4-4: The calibration algorithms developed for the laboratory-prepared (air and dried) state (n = 238) for SOC content in South Africa. The green highlighted calibration algorithms indicate the best performing algorithms.

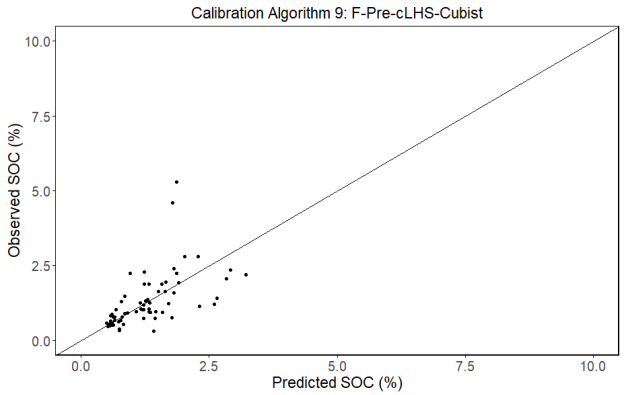
Algorithm number	Code	Validation statistics			
		RMSE	R^2	ρc	RPIQ
31	P-Pre-RS-PLSR	0.71	0.25	0.39	1.41
32	P-Pre-RS-RF	0.72	0.24	0.51	1.39
33	P-Pre-RS-SVM	0.69	0.3	0.47	1.45
34	P-Pre-RS-Cubist	0.48	0.65	0.78	2.07
35	P-Pre-RS-ANN	0.78	0.09	0.43	1.28
36	P-Pre-cLHS-PLSR	0.77	0.3	0.42	1.32
37	P-Pre-cLHS-RF	0.71	0.4	0.56	1.43
38	P-Pre-cLHS-SVM	0.75	0.33	0.48	1.35
39	P-Pre-cLHS-Cubist	0.57	0.62	0.75	1.8
40	P-Pre-cLHS-ANN	0.78	0.28	0.48	1.3
41	P-Pre-K-Means-PLSR	0.97	0.14	0.3	1.18
42	P-Pre-K-Means-RF	1.02	0.05	0.26	1.12
43	P-Pre-K-Means-SVM	1.01	0.08	0.27	1.14
44	P-Pre-K-Means-Cubist	0.75	0.49	0.6	1.53
45	P-Pre-K-Means-ANN	0.99	0.11	0.27	1.16
46	P-Raw-RS-PLSR	0.71	0.24	0.38	1.4
47	P-Raw-RS-RF	0.68	0.31	0.55	1.47
48	P-Raw-RS-SVM	0.69	0.28	0.46	1.44
49	P-Raw-RS-Cubist	0.39	0.78	0.86	2.57
50	P-Raw-RS-ANN	0.83	-0.02	NA	1.2
51	P-Raw-cLHS-PLSR	0.78	0.3	0.43	1.46
52	P-Raw-cLHS-RF	0.74	0.37	0.52	1.54
53	P-Raw-cLHS-SVM	0.81	0.24	0.42	1.4
54	P-Raw-cLHS-Cubist	0.57	0.62	0.76	2
55	P-Raw-cLHS-ANN	0.94	-0.03	NA	1.21
56	P-Raw-K-Means-PLSR	0.91	0.06	0.25	1.06
57	P-Raw-K-Means-RF	0.99	-0.1	0.22	0.98
58	P-Raw-K-Means-SVM	0.94	0	0.24	1.02
59	P-Raw-K-Means-Cubist	0.7	0.45	0.64	1.38
60	P-Raw-K-Means-ANN	0.95	-0.02	0	1.01

P = laboratory-prepared state, Pre = preprocessing, RS = simple random sampling, cLHS = conditioned Latin Hypercube Sampling, K-Means = K-Means clustering, PLSR = Partial Least Squares Regression, RF = Random Forest, SVM = Support Vector Machines, ANN = Artificial Neural Networks.

(a)



(b)



(c)

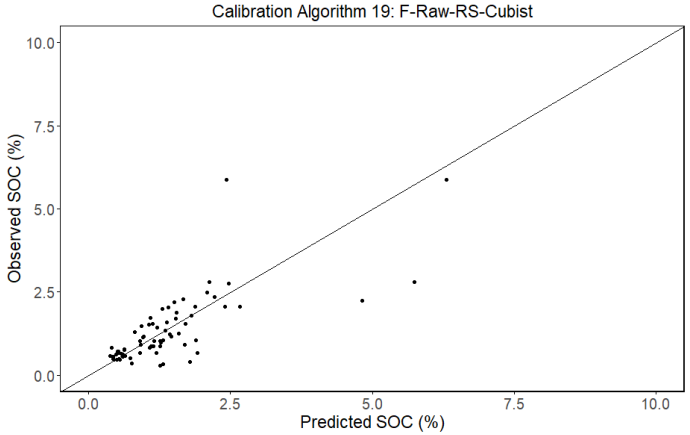


Figure 4-7: Scatter plots indicating the top three calibration algorithms for validation model prediction for SOC content of the field state. The three calibration algorithms included (a) calibration algorithm 8, (b) calibration algorithm 9 and (c) calibration algorithm 19. The black line represents the 1:1 line.

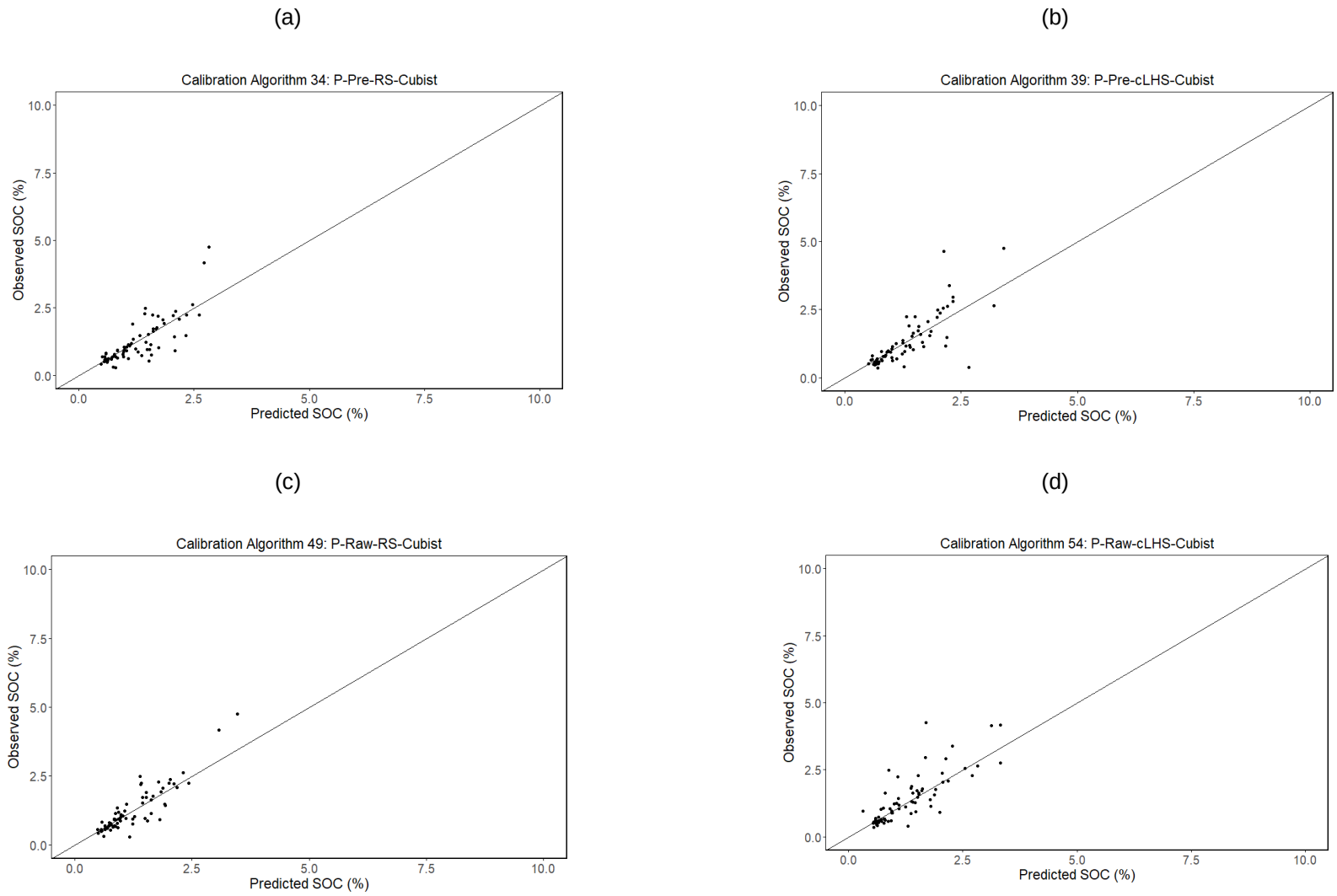


Figure 4-8: Scatter plots indicating the top four calibration algorithms for validation model prediction for SOC content of the laboratory-prepared state. The four calibration algorithms included (a) number 34, (b) number 39, (c) number 49 and (d) number 54. The black line represents the 1:1 line.

4.3.3.2 Effect of methodological parameters

The Kruskal-Wallis tests were undertaken on the different methodological parameters that were used to develop the regional calibration algorithms (Table 4-5). Overall, the sample state and sample design were the most influential parameters influencing RSME), indicated by their high χ^2 -values and low p-values (Table 4-5 and Figure 4-9). The sample design parameter had the highest impact on RMSE and was statistically significant (Table 4-5). The cLHS sample design was the most accurate, with the lowest error and RMSE distribution, while the K-Means were the least accurate sample design (Figure 4-9). The processing parameter was not as influential, nor statistically significant (Table 4-5). The distribution and error bars were very similar (Figure 4-9), with only the median of the pre-processing algorithms being slightly less than the raw spectral algorithms.

Although the machine learning models were not the most influential on RMSE (Table 4-5), the results indicated that the Cubist model might perform better than other models for developing SOC content calibration algorithms (Figure 4-9). Cubist did not have the smallest error of all the models, but it did have the lowest distribution (median = 0.71 %).

Table 4-5: Kruskal-Wallis test statistics of the methodological parameters.

Methodological parameter	χ^2 -value[1]	p-value
Sample state	13.35	<0.005
Pre-Processing	0.88	0.35
Sample design	24.28	<0.005
ML models[2]	10.5	0.03

[1] χ^2 = chi-squared
[2] ML = machine learning

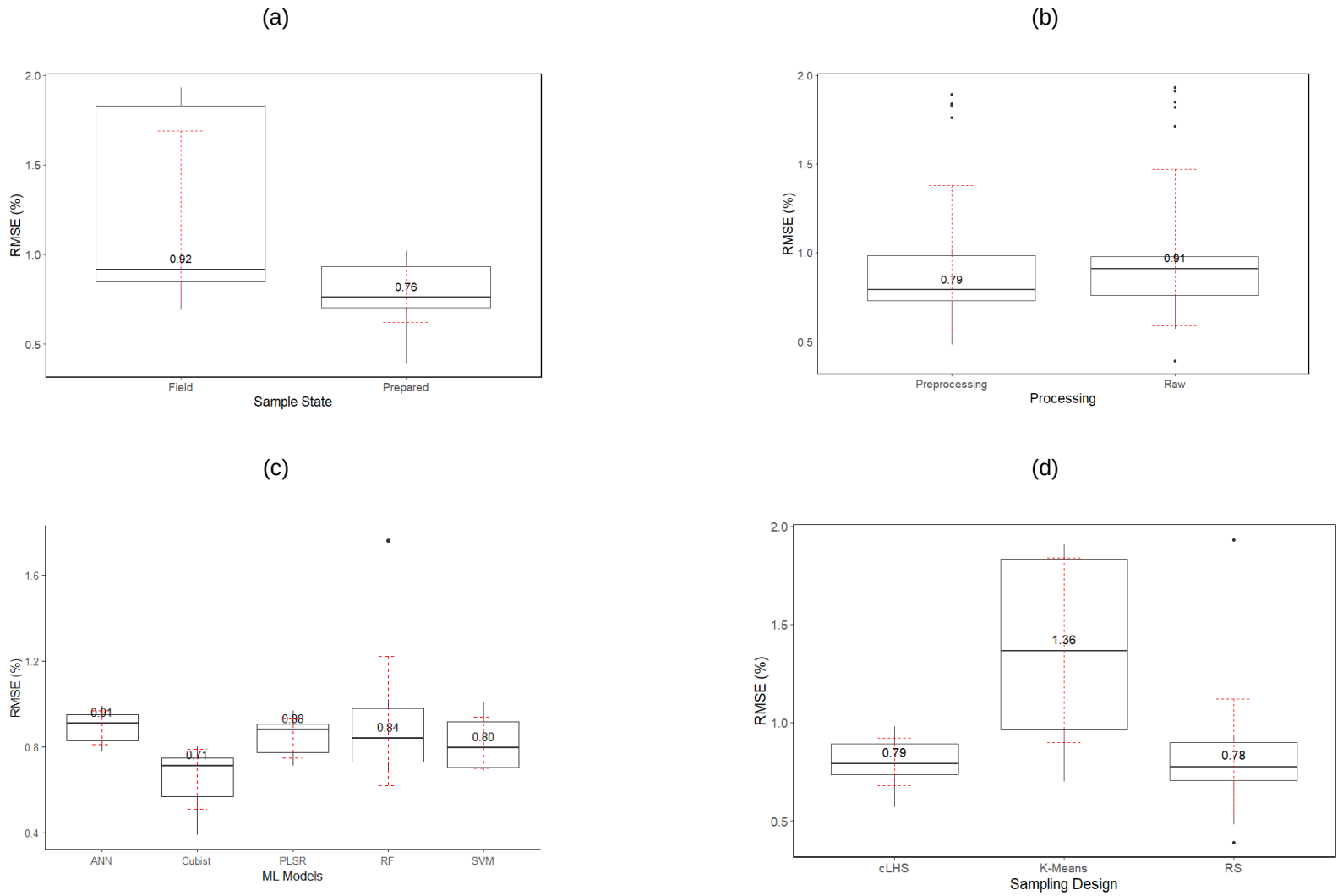


Figure 4-9: The distribution of root mean square error (RMSE) validation values for the four methodological parameters used for creating calibration algorithms of soil organic carbon (SOC) content for South Africa.

4.3.4 Local scale calibration algorithms

The laboratory-prepared state slightly outperformed the field state for almost all calibration algorithms (Table 4-6 and Table 4-7). Overall, the most optimal regional calibration algorithm (P-Pre-cLHS-Cubist) from section 4.3.3.1 was found to be the most optimal calibration algorithm for only the Tsitsa catchment for the field and laboratory-prepared states (Table 4-6 and Table 4-7).

The best overall calibration algorithm was achieved for the Ottosdal Field, with an RMSE of 0.04% and RPD of 1.87. The Tsitsa catchment and Ottosdal field sites had the best performing developed calibration algorithms for both field and laboratory-prepared states (Figure 4-10). However, all other study sites had different methodological configurations for their respective best calibration algorithms.

Table 4-6: The standard calibration algorithms and best calibration algorithms developed for the five study sites for the field state. The green highlighted calibration algorithms indicate the best performing algorithms.

Study site	Calibration algorithm	Validation statistics			
		RMSE	R^2	ρc	RPIQ
Standard calibration algorithms					
Regional	F-Pre-cLHS-Cubist	0.72	0.32	0.56	1.3
Olifants Catchment	F-Pre-cLHS-Cubist	0.7	0.3	0.53	1.6
Sabie-Sands Catchment	F-Pre-cLHS-Cubist	1.49	0.42	0.54	1
Tsitsa Catchment	F-Pre-cLHS-Cubist	0.1	0.95	0.97	3.8
Ottosdal Field	F-Pre-cLHS-Cubist	0.09	-0.38	0.27	0.9
Vrede Field	F-Pre-cLHS-Cubist	0.82	-0.2	-0.03	0.7
Best calibration algorithms					
Regional	F-Pre-cLHS-Cubist	0.72	0.32	0.56	1.3
Olifants Catchment	F-Pre-cLHS-SVM	0.57	0.53	0.68	1.9
Sabie-Sands Catchment	F-Pre-RS-PLSR	1.05	0.6	0.71	1.8
Tsitsa Catchment	F-Pre-cLHS-Cubist	0.1	0.95	0.97	3.8
Ottosdal Field	F-Raw-cLHS-Cubist	0.08	0.59	0.72	1.4
Vrede Field	F-Raw-cLHS-RF	0.55	0.06	0.2	0.7

F = field state, Pre = preprocessing, RS = simple random sampling, cLHS = conditioned Latin Hypercube Sampling, K-Means = K-Means clustering, PLSR = Partial Least Squares Regression, RF = Random Forest, SVM = Support Vector Machines, ANN = Artificial Neural Networks.

Table 4-7: The standard calibration algorithms and best calibration algorithms developed for the five study sites for the laboratory-prepared state. The green highlighted calibration algorithms indicate the best performing algorithms.

Standard calibration algorithms					
Study site	**Calibration algorithm**	**Validation statistics**			
		RMSE	**R^2**	**ρc**	**RPIQ**
Regional	P-Pre-cLHS-Cubist	0.57	0.62	0.75	1.8
Olifants Catchment	P-Pre-cLHS-Cubist	1.28	-0.05	0.04	1.2
Sabie-Sands Catchment	P-Pre-cLHS-Cubist	1.03	0.16	0.44	1.4
Tsitsa Catchment	P-Pre-cLHS-Cubist	0.09	0.97	0.99	5.4
Ottosdal Field	P-Pre-cLHS-Cubist	0.1	-0.03	0.29	0.8
Vrede Field	P-Pre-cLHS-Cubist	0.64	-0.02	0.14	0.9
Best calibration algorithms					
Regional	P-Pre-cLHS-Cubist	0.57	0.62	0.75	1.8
Olifants Catchment	P-Pre- K-Means-Cubist	0.68	0.5	0.6	1.9
Sabie-Sands Catchment	P-Raw-cLHS-Cubist	0.8	0.38	0.55	1.8
Tsitsa Catchment	P-Pre-cLHS-Cubist	0.09	0.97	0.99	5.4
Ottosdal Field	P-Raw-cLHS-Cubist	0.05	0.71	0.83	1.1
Vrede Field	P-Raw-K-Means-Cubist	0.24	0.81	0.89	1.6

P = laboratory prepared state, Pre = preprocessing, RS = simple random sampling, cLHS = conditioned Latin Hypercube Sampling, K-Means = K-Means clustering, PLSR = Partial Least Squares Regression, RF = Random Forest, SVM = Support Vector Machines, ANN = Artificial Neural Networks.

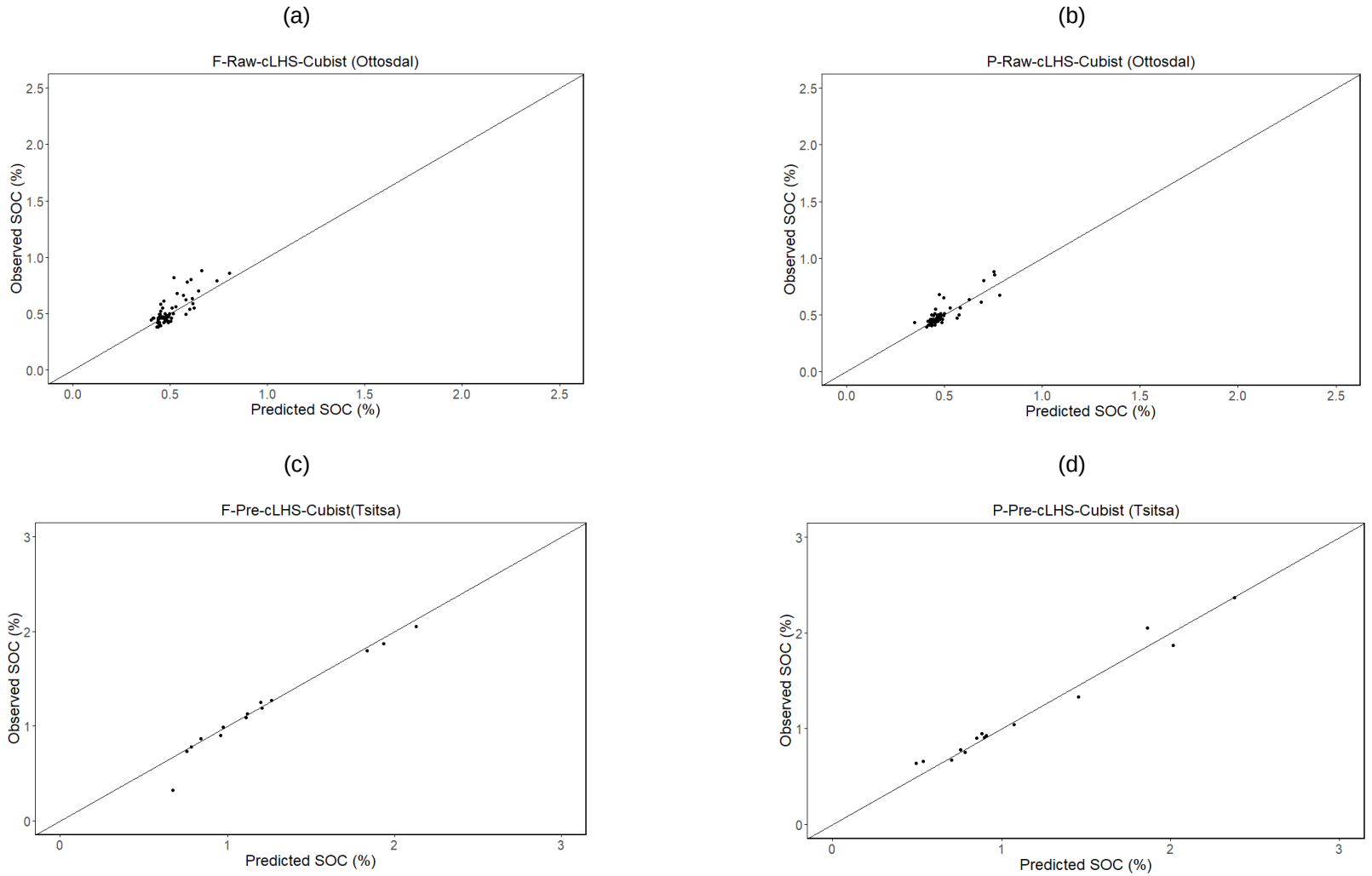


Figure 4-10: The best calibration algorithms created for soil organic carbon (SOC) content for the field- and prepared state at the catchment- and field scale. The black line represents the 1:1 line.

4.4 Discussion

4.4.1 Soil organic carbon content distribution and near-infrared spectra

The difference in SOC content between the different study sites was likely attributed to the different soil types and land use. The low SOC content distribution of the Tsitsa study site was likely due to the SOC being artificially decreased because of severe degradation of the area (du Plessis et al., 2020), while the Ottosdal study site presented low SOC content due to the soils being predominantly sandy, cultivated land (section 4.2.1). In contrast, the Olifants and Sabie-Sands study sites had soils with higher organic matter content, as these samples were collected in forested areas. These two catchments also had a higher distribution of clayey soils (section 4.2.1) and were able to retain more SOC, due to the soils' ability to bind organic matter and protect it from decomposition.

Overall, the absorbance values for the field state were generally higher compared to the laboratory-prepared state, likely because various forms of water strongly absorb NIR light (Stenberg et al., 2010). The pronounced peaks, located at the 1400 and 1900 nm wavelengths were likely attributed to the presence of water (Hummel, 2001; Stenberg et al., 2010; Gao, 2014). Previous studies have also indicated that peaks located at the 1400, 1900 and 2200 nm wavelengths could also be associated with clay minerals (Viscarra-Rossel et al., 2006; Stenberg et al., 2010; Conforti et al., 2015). Although this study did not investigate the importance of specific spectral bands for SOC content prediction, previous studies have identified the 2000 to 2400 nm wavelengths as particularly important for SOC content prediction (Stenberg, 2010; Stenberg et al., 2010). The differences in absorbance between the study sites for both field state and laboratory-prepared state, could likely be attributed to differences in soil texture as all the study sites, except for the Ottosdal study site, had a strong representation of South African soils (section 4.2.1). As mentioned, the Ottosdal study site was characterised as very sandy (section 4.2.1), which explains the very low absorbance values for this study site. Conforti et al. (2015) and Zhang et al. (2024) has demonstrated that sandy soils have a higher reflectance and lower absorbance compared to clayey soils, likely attributed to the smaller surface area and larger particle size, as well as high quartz content.

4.4.2 OSSL distribution

The OSSL has successfully predicted SOC content with their developed NIR calibration algorithms (Safanelli et al., 2025), however, as South African soil spectral data are underrepresented in the current OSSL database (Kock et al., 2024), the OSSL prediction for local data was found to be poor and unusable (RPIQ < 1, RMSE of 1.23% and ρc of 0.21). This study concluded that since only 64 samples were represented in the OSSL dataset, any spectral data within South Africa cannot be used with the global model. Therefore, future work should focus on

contributing local data to global databases, such as the OSSL, which can significantly improve their performance for South African users.

4.4.3 Regional scale calibration algorithm

4.4.3.1 Developed calibration algorithms

As mentioned, the best overall calibration algorithm (49) achieved an RMSE of 0.39% and RPIQ > 2. However, compared to recent studies the RMSE of the algorithm is slightly too high to be used for predicting SOC content (Camacho-Tamayo et al., 2013 and Li et al., 2020 achieved an RMSE < 0.3%). This is also substantiated in the scatter plots (Figure 4-6), where several samples plotted far away from the 1:1 line.

When comparing the validation statistics of the best calibration algorithms for field- and laboratory-prepared state to previous studies that also developed algorithms for SOC content at the regional scale, it is evident that sample size and spectral range might also be important factors for accurate and robust algorithms.

Studies using only the NIR range (Cambule et al., 2012; Camacho-Tamayo et al., 2013; Li et al., 2020), outperformed the best developed calibration algorithms from this study, with RMSE values < 0.3%. Only one recent study (Peng et al., 2024) performed poorer, with an RMSE of 0.75%. The main difference was the sample size, with Peng et al. (2024) using only 50 samples, while the other studies had more than 400 samples. Li et al. (2020) achieved more accurate predictions compared to this study (RMSE < 0.3%), despite having a larger SOC content range compared to the results found for this study. As RMSE values are highly dependable on the range of SOC content, this likely impacted prediction accuracy on a regional scale. The variable SOC content range of the three catchments (Figure 4-3), likely resulted in the calibration algorithms predicting relatively high RMSE values compared to Cambule et al. (2012) and Camacho-Tamayo et al. (2013). However, an increase in sample size could improve the calibration algorithms at regional scale.

A larger sample size enables more accurate modeling of complex relationships between spectral data and SOC content (Yang et al., 2019) and therefore an increase in sample size should improve the calibration performance. However, there are currently no guidelines as to how much improvement can be expected and what the minimum number of samples used is to be effective (Ng et al., 2020).

Regarding the effect of spectral range, studies that included the visible part of the spectrum (Nocita et al., 2012; Kusumo et al., 2018; Xu et al., 2021; Singha et al., 2023), all outperformed this study's best developed calibration algorithms, with RMSE values < 0.3%. The only recent

studies that underperformed had RMSE values higher than 0.72% (McDowell et al., 2012; Tekin et al., 2014; Long et al., 2023). These studies had a sample size smaller than 100, making it difficult to accurately model the relationship between the spectral data and the SOC content (Yang et al., 2019).

Specific spectral bands around 1 100, 1 600, 1 700 to 1 800 and 2 200 to 2 400 nm have been identified as being particularly important for SOC calibration (Stenberg et al., 2010). However, the visible range (400-780 nm) includes information related to color, which can be influenced by soil mineralogy, iron oxides, and soil organic matter. Previous studies have found obvious relationships between SOC and different textures (Viscarra-Rossel et al., 2006; Nocita et al., 2013). There are also absorption features responsible for the prediction of SOC content (comprising of NH, CH and CO bonds) around the bands within the NIR range and also the visible range (Stenberg et al., 2010). Thus, the visible range can provide additional information on soil constituents that affect SOC indirectly, improving the robustness and accuracy of SOC predictions.

As mentioned, there was no consistent trend found when comparing the calibration and validation statistics for the developed calibration algorithms. Along with several parameters, such as sampling selection techniques and machine learning models that were utilised, the study concluded that no definitive reason was found to explain the inconsistencies between the calibration and validation statistics. However, the likely reason for these inconsistencies was attributed to the relatively small sample size, which led to the validation performances changing significantly depending on which samples were included. The small sample size might have led to the calibration or validation samples being under- or overrepresented of the overall population, which resulted in inflating or underestimating the predictive accuracies.

In summary, the developed calibration algorithms developed were neither robust nor accurate enough to be considered for predicting SOC content, when compared to other national and international studies on a regional scale. Therefore, these algorithms need to be improved before being used to predict SOC content in South Africa using NIR spectroscopy.

4.4.3.2 Effect of methodological parameters

As mentioned, the sample design parameter had the highest impact on RMSE and could be explained by the way these parameters operate.

As K-Means clustering aims to partition data into clusters, the clustering process might not have selected samples that capture the full range of SOC content variability, leading to less representative calibration algorithms. On the other hand, the cLHS design excels in capturing

representative sampling across the entire range of the dataset, leading to the more accurate and robust sampling design (Minasny & McBratney, 2006). Even though RS did not perform better than cLHS, it did perform better than K-Means clustering. The RS design likely captured a more diverse and representative subset of the data, which ensured better generalization in the calibration algorithms.

As mentioned, the difference in RMSE values between the two sample states was statistically significant and indicated that the sample state was the methodological parameter that had the second highest impact on RMSE (Table 4-3). The laboratory-prepared state had a much smaller error and distribution compared to the field state, hence, the laboratory-prepared state showed a higher accuracy. This is because the presence of water molecules introduces variability and interference with the spectra during the scanning of field samples (Nocita et al., 2013; Tekin et al., 2014; Biney et al., 2020). Previous studies have also confirmed the decrease of accuracy of calibration algorithms with the increase of water content from dryness to saturation (Ge et al., 2022). The field state introduces additional complexity in the spectral readings, impacting the accuracy of the algorithms in predicting SOC content (Wadoux et al., 2021).

As mentioned, the distribution and error bars were very similar for the processing parameters. The two main reasons for this finding were due to insufficient pre-processing and/or quality of spectral data. Firstly, only two pre-processing techniques were utilised, which might not have been significant enough to alter the spectral data in a way that impacts the calibration algorithms. Other methods, such as Savitzky-Golay smoothing, baseline correction or detrending, might alter the spectral data and impact the calibration algorithms developed (Wadoux et al., 2021). Secondly, if the raw spectral data were of high quality and consistently well-calibrated, the impact of pre-processing might be minimal.

The main reason for Cubist outperforming the other machine learning models, was likely due to the model's unique combination of rule-based regression and instance-based learning, which allows it to capture both global and local patterns effectively (McBratney et al., 2003; Peng et al., 2015; Kock, 2022). Another reason might be due to the limitations of the other models for creating calibration algorithms for SOC content. RF and PLSR have the common limitation of not capturing certain non-linear relationships (Sharififar et al., 2019; Bai et al., 2022), and tend to overfit data, while ANN and SVM require careful tuning of hyperparameters (Sharififar et al., 2019; Bai et al., 2022). The combined effect of Cubist being able to handle complex data and the limitations of the other models, likely led to Cubist being the most accurate model.

In summary, the sample state and sample design had the largest effect on the developed calibration algorithms. The laboratory-prepared state outperformed the field state, as the water

molecules introduce complexity in the spectral readings, impacting the accuracy of the calibration algorithms. The cLHS design outperformed the other sampling designs, due to capturing the full range of SOC content variability.

4.4.4 Local scale calibration algorithm

The likely reason for other calibration algorithms outperforming the standard calibration algorithm at the different study sites was that each study site had a unique SOC content, SOC content distribution and a different number of samples. This variability likely required a different methodological configuration for optimal performance for each specific study site. Therefore, the local conditions of each specific study site should be taken into consideration.

Comparing the catchment study sites (including Olifants, Sabie-Sands and Tsitsa) to other studies that developed calibration algorithms for SOC content at a catchment scale, only the Tsitsa catchment performed better. Previous studies returned RMSE values between 0.1 and 0.33% and RPD values between 2 and 4 (Bai et al., 2022; Zayani et al., 2023). The calibration algorithms for the Tsitsa catchment were more impressive, as the previous studies used a Vis-NIR instrument and had up to 330 samples to be used for the calibrations.

The possible reasons for the Tsitsa catchment outperforming all other calibration algorithms and other studies were likely due to the low SOC range, representative sampling, effective range and optimal model fitting. Even with fewer samples, the Tsitsa catchment might have achieved a more representative sampling range, which leads to less variability for the model to learn from (Wadoux et al., 2021). When comparing the scatter plots of the Tsitsa catchment with the Ottosdal field, the Tsitsa catchment found a better 1:1 line fit compared to the Ottosdal field, where more than 10 samples were overpredicted (Figure 4-9).

Comparing the field study sites (Ottosdal and Vrede), only the Ottosdal field outperformed previous studies that developed calibration algorithms for SOC content at field scale (Nocita, 2009; Knadel et al., 2015; Biney et al., 2020; Gholizadeh et al., 2021). The previous studies had RMSE values between 0.2 and 0.35% (Nocita, 2009; Knadel et al., 2015). Although the number of samples (200) did improve the RMSE (0.08% for field state and 0.05% for prepared state) for the Ottosdal field, RPIQ values (< 2) were still lower than the previous studies.

Since the developed calibration algorithms for both the Tsitsa catchment and Ottosdal field had RMSE < 0.1% and RPIQ > 1.5, these calibration algorithms were concluded as usable for predicting SOC content within the specific areas.

4.4.5 Hand-held NIR spectrometer as a field instrument

As mentioned, there was no concise difference found between the laboratory prepared state and field state for regional scale, however, on the smaller scales the laboratory prepared state outperformed the field state. Even though this might be the case, the Ottosdal and Tsitsa study sites had successful calibration algorithms for both states, indicating that the hand-held NIR spectrometer might be considered a field instrument for measuring SOC content in situ at smaller scales. Although, laboratory prepared spectral data has been proven to provide accurate SOC content predictions, in situ spectroscopy has the advantage of being a more time- and labour-saving method (Xie et al., 2011; Yang et al., 2024; Loria et al., 2024). However, the removal of moisture is crucial for improving model prediction accuracy and is an ongoing problem that must be solved for in situ spectroscopic detection of SOC content (Yang et al., 2024). Therefore, more research is needed on the removal of moisture interference when utilising in situ measurements, which could likely lead to improved South African calibration algorithms at regional scales.

4.5 Conclusion

The study achieved the aim of developing a NIR spectroscopy calibration algorithm of SOC content that could be used for South Africa. However, this were only true for two local scale calibration algorithms. The best overall calibration algorithm (P-Pre-cLHS-Cubist) on regional scale, achieved an RMSE of 0.39% and RPIQ > 2. Although 60 calibration algorithms were developed for SOC content on a regional scale, none (including the best overall calibration algorithm) were deemed acceptable to be used for SOC content prediction in South Africa. The study concluded that the small sample size (n = 238) was likely the limiting factor for unusable calibration algorithms at regional scale and that a larger sample size would enable more accurate modeling of complex relationships between spectral data and SOC content. Currently, only the Tsitsa catchment and Ottosdal field have successfully developed calibration algorithms that can be used for future SOC content prediction.

The study found that the OSSL global prediction model poorly predicted SOC content using local data (RMSE = 1.23% and ρc = 0.21). This was attributed to South African samples being underrepresented in the global dataset, which led to the conclusion that the OSSL cannot be used for SOC content prediction of South African soils.

The study also found that sample state and sample design were the most influential methodological parameters influencing RMSE. The cLHS sampling design outperformed the other sampling designs, as it captured the representative sampling across the entire range of the dataset, leading to more accurate and robust sampling design. Although the choice of machine learning model did not influence RMSE on a regional scale, Cubist did outperform the other

models, likely due to the combined effect of Cubist being able to handle complex data and the limitations of the other models.

Finally, the third objective was also reached, with the study finding that a NIR spectrometer could be used for measuring SOC content in situ. However, only the Tsitsa catchment and Ottosdal field have successfully developed calibration algorithms and therefore the NIR spectrometer can currently only be used for predicting SOC content at smaller scales.

Since sample size was likely the limiting factor, to develop a national calibration algorithm, effort should be placed on developing accurate calibration algorithms for smaller areas that could be added to the national spectral library. In this way, site specific accurate soil spectral calibration would provide incentive to gather the data sufficient for a national calibration algorithm. The use of local models, such as locally weighted regression (LWR) and memory-based learning (MBL) could be explored to further improve the developed calibration algorithms. Spectral bands associated with SOC content for South African soils could also be explored. Finally, future work should focus on contributing local data to the OSSL global dataset, which can significantly improve their performance for South African users.

CHAPTER 5 HOW TO MAP SOIL ORGANIC CARBON STOCKS AT FIELD SCALE IN SOUTH AFRICA?

5.1 Introduction

Soil organic carbon (SOC) is one of the most important pools in the global carbon cycle, with a total quantity of 1 200 to 1 500 gigatons of organic carbon in 1 m depth ranges (Santacruz et al., 2014; Hartemink & McSweeney, 2014) and 2 400 gigatons of organic carbon to 2 m depth ranges (Batjes, 1996; Paustian et al., 2019). To mitigate emissions, carbon sequestration, which refers to the long-term storage of carbon in oceans, soils, vegetation, and geological formations, has gained significant attention in recent years (Baurov, 2021; Zhang et al., 2021a; Batool et al., 2024). Besides combating climate change through carbon sequestration, South African land managers also have the opportunity to sell units of carbon (called carbon credits) to the government or other businesses to generate an income (Sharma et al., 2021). Given its relevance to the South African industry, appropriate methods for mapping and monitoring soil organic carbon stocks (SOCS) are becoming crucial for accurate and reliable carbon credit assessment.

Currently, South African industry utilises conventional methods for mapping soil properties, such as SOC content and SOCS at field scale. In the case of South African industries, conventional methods refer to soil samples being collected from the topsoil (0-30 cm depth) (Aynekulu et al., 2011; TREES Consulting, 2020) on a 1-hectare (ha) or 2-ha grid basis (Verster et al., 2022) using geostatistical models, such as ordinary kriging (OK), regression kriging, co-kriging and inverse distance weighted regression to map the soil properties. These methods are used due to their prediction accuracy and error minimisation (Farooq et al., 2022). From the mentioned geostatistical models, OK has been the most popular technique used for mapping SOC content and SOCS in recent years (Verster et al., 2022; Li et al., 2022; Shoumik & Khan, 2023; Pouladi et al., 2023). However, OK utilises spatial point data for predicting values between points (Zhu et al., 2022). The use of OK at field scale has shown to have several disadvantages, with very little research having been done on the accuracy of the 1-ha grid layout, and the inability of the method to fully utilise the surface data of environmental covariates has shown to yield less accurate maps (Huang et al., 2022; Zhu et al., 2022; Pouladi et al., 2023).

In the past two decades, digital soil mapping (DSM) with machine learning (ML) has been increasingly used to create more accurate and precise SOCS maps (Nenkam et al., 2024). This method predicts soil properties, such as SOCS, at unsampled locations by correlating the relationship between soil properties and environmental covariates utilising ML and other methods (Pouladi et al., 2023). These ML techniques, such as Support Vector Machines (SVM), Random Forest (RF) and Cubist, among others, have also been able to capture non-linear relationships

between soil properties and environmental covariates (derived from remote sensing imagery and elevation models), leading to better predictions and more accurate results (Farooq et al., 2022; Kotzé & van Tol, 2023). This method has been shown to save a lot of labour and time and is free from the limitation of expert knowledge limited to a certain area (Zhu et al., 2022). DSM with ML has been used to map SOC content and SOCS extensively internationally (Pouladi et al., 2023) and in South Africa (Seboko et al., 2021; Venter et al., 2022; Kotze & van Tol, 2023).

Reliable and accurate SOCS maps are important for accurately assessing carbon credits. The accurate assessment of carbon credits is crucial for local land managers and the industry, as identifying the most reliable method can not only enhance efficiency but also significantly reduce costs, ultimately benefiting all stakeholders involved. Currently, South Africa predominantly uses the Verified Carbon Standard Methodology (Shoch & Swails, 2020) and the Gold Standard (Aynekulu et al., 2011; TREES Consulting, 2020) as frameworks for determining SOC content and SOCS for carbon credit assessment. Unfortunately, neither of these methods specify which method/s should be used for mapping SOCS at field scale. Although previous studies have utilised OK or DSM with ML for mapping SOCS (Wiese et al., 2016; Flynn et al., 2019; Venter et al., 2021; Flynn et al., 2022; Nenkam et al., 2024), South Africa does not currently have a standard method for mapping SOCS at field scale. To address this gap, the study aimed to determine whether conventional mapping using OK or DSM with ML should be used for mapping SOCS at field scale in South Africa.

Objectives of the study included:

- To determine if the conventional method using OK is sufficient for mapping SOCS at field scale.
- To assess three ML models for performing DSM with ML of SOCS at field scale.
- To compare the conventional method using OK with DSM with ML for mapping SOCS at field scale.
- To compare a calculate-first approach with a map-first approach when using the conventional method of OK or using DSM with ML.

5.2 Materials and methods

5.2.1 Study area and site description

Two study sites were selected from two different ecotopes, with an ecotope defined as a homogeneous area with a unique combination of climate, topographic and soil characteristics (Mzezewa et al., 2009). Within each ecotope, a field from a farm was selected, with the farms located near the towns of Ottosdal (located within the North West Province) and Vrede (located within the Free State Province) (Figure 5-1). The Vrede study site receives a median annual

precipitation (MAP) of 795 mm (Shulze, 2007), is overlain by the Balfour Formation, which consists of mudstone and sandstone (Council for Geoscience, 2019) and the soils found in the area are typically clayey. In contrast, the Ottosdal study site receives a MAP of 565 mm (Schulze, 2007), is overlain by the Klipriviersberg Group, which consists of basalt, andesite and tuff (Council for Geoscience, 2019) and the soils found in the area are typically sandy. The dominant soil forms found at the Ottosdal study site are Clovelly and Hutton, while at the Vrede study site the dominant soil form is typically Tukulu (Soil Classification Working Group, 2018). The Ottosdal soil forms are typically deep and sandy, while the Vrede study site is typically has a clay loamy texture. Both farms utilised rotational cropping with maize and soybean. Both farms also implemented no-tillage practices and incorporated residues after harvest. In addition, the Vrede farm utilises other regenerative agricultural practices, including winter cover crops, manure input and daily rotational grazing of sheep.

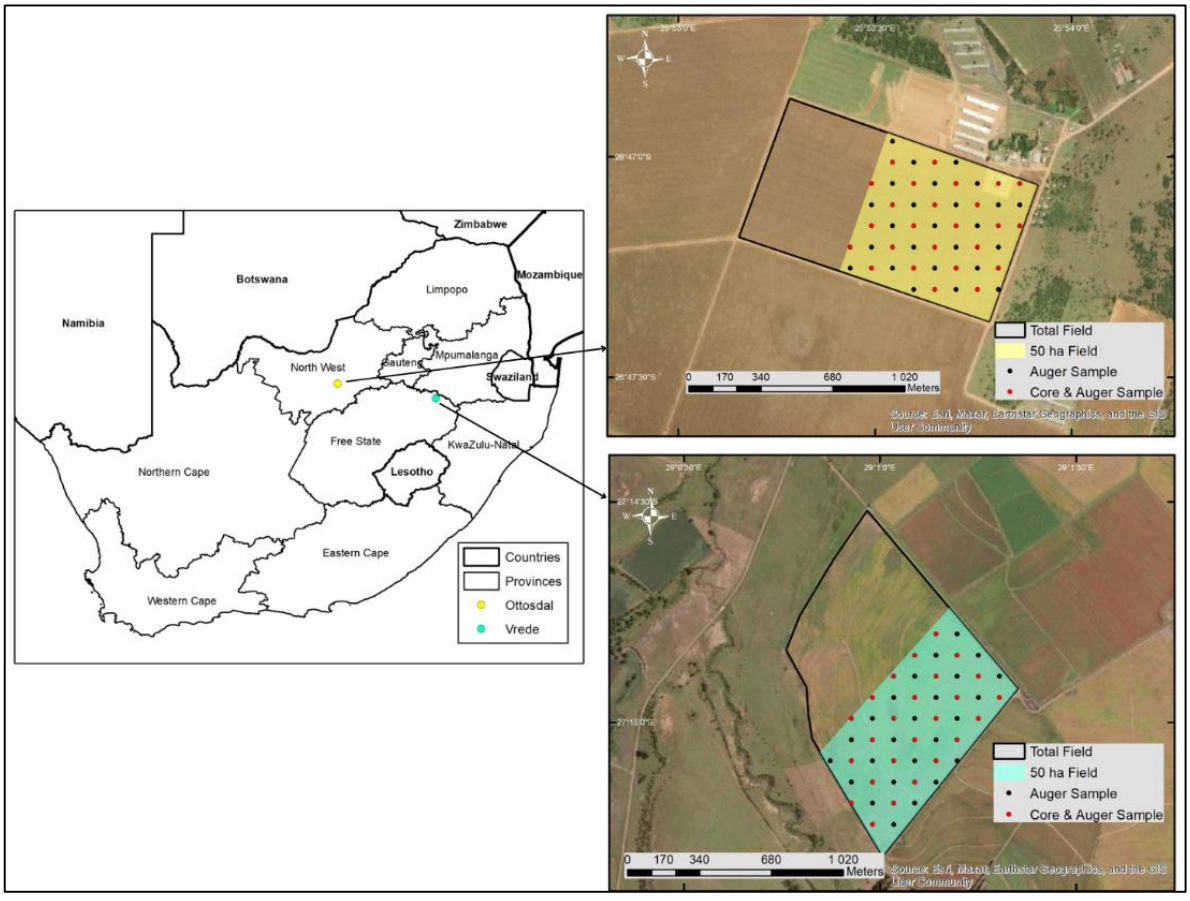


Figure 5-1: The location of the Ottosdal- and Vrede study sites, as well as the location of the fields and samples collected for the study.

5.2.2 Soil sampling analysis

Soil sampling was carried out on the two fields, with 50 sampling locations laid out on a 1 ha grid to cover an area of 50 ha (Figure 5-1), which was done to simulate how South African industry

would collect soil samples (Verster et al., 2022). The sampling was conducted after harvesting, between July and October 2022. For both the SOC content and dry bulk density (ρb) sampling, samples were collected at the following depth increments: 0-5-, 5-15-, 15-30 cm. Due to the tedious method of collecting samples for ρb, it was only carried out at every second observation, at all three depths (*n* = 75 / site). This data was then used to create a pedo-transfer function (PTF) to predict the remaining ρb values at the unsampled locations.

At all 50 sampling locations, a soil hand auger was used to collect disturbed samples at three depths (*n* = 150 per site). The disturbed soil samples were analysed for SOC content with the total dry combustion (TDC) method using a CN analyser (Nelson & Sommers, 1996; FAO, 2019). At every second profile (*n* = 25 per site), a hand core sampler with a volume of 114 cm³ was used to the collect an undisturbed soil sample (*n* = 150 per site). The undisturbed core samples were oven-dried at 105 °C and ρb determined using Equation 5-1. The disturbed soil samples from the ρb observations, were also analysed for the three fractions of soil texture.

Equation 5-1: $\textbf{Dry bulk density } (\boldsymbol{\rho b}) \left(\frac{\mathbf{g}}{\mathbf{cm^3}}\right) = \textbf{Dry soil (g)/Cylinder volume (cm}^3\textbf{)}$

The PTF for predicting the ρb at unsampled locations were created using the RF model. The 150 samples were divided into a training (75%) and evaluation (25%) dataset using conditioned Latin Hypercube Sampling (cLHS) (Minasny & McBratney, 2006; Roudier et al., 2022), with sand, silt, clay and SOC content as input variables. Using the "randomForest" (Breiman et al., 2024) and "predict" (Fox & Weisberg, 2019) functions in RStudio, the RF model was fitted on the training dataset with the sand, silt, clay and SOC content as predictor variables. The accuracy and reliability of the model was tested on the calibration and evaluation datasets with the root mean square error (RMSE), Lin's concordance correlation coefficient (ρc) (Lin, 1989), coefficient of determination (R^2), and ratio of performance to inter-quartile distance (RPIQ) (Bellon-Maurel et al., 2010). Although RPIQ is not a common PTF evaluation statistic, RPIQ provide a good indication on the reliability of the model used for predicting ρb values, with RPIQ < 1.8 indicating less reliable models, while RPIQ > 1.8 indicating reliable models. A RMSE < 0.2 g/cm³ would also indicate sufficient accuracy for predicting ρb values (Abdelbaki, 2018; Yanti et al., 2021; Bryk & Kołodziej, 2023; Chen et al., 2024).

The SOC content, ρb values from the sampled locations and predicted ρb values from the pedo-transfer function of unsampled locations were ultimately used for calculating the SOCS for each observation for both study sites (Equation 5-2). Typically for accurate SOCS calculations, coarse fragments (1 – Sm (v%)) are incorporated into the equation (Equation 5-2), however, the laboratory texture analyses revealed it to be zero.

Equation 5-2: SOCS = 10 × SOC content (g/kg) × ρb (Mg/m³) × (1 – Sm (v%)) × depth (cm)

The SOC content and ρb distribution for each study site was shown through box-and-whisker plots created using the “ggplot” function in RStudio (Wickham, 2016). Given that the assumption of normality was not met, a Kruskal-Wallis’s test (Kruskal and Wallis, 1952) was performed in RStudio (R Core Team, 2020) using the “kruskal.test” function, which was done to determine if there were overall significant statistical differences between the values of the methods. Afterwards, a pairwise means comparison test (“pairwise.wilcox.test” function) was also done in RStudio to determine the significance differences between the study sites across different depths. For each study site across the different depths, QQ-plots were also plotted to compare the sample quantiles of SOC content and ρb to a theoretical normal distribution.

5.2.3 Mapping analysis

5.2.3.1 Geostatistical mapping

For the geostatistical approach, OK was used with all 50 observations of SOC content and ρb (using both measured and predicted ρb values) to create continuous SOC content and ρb maps at all three depths measured. To calculate the SOCS with the map-first approach, the created SOC content and ρb maps were imported into System for Automated Geoscientific Analyses Geographic Information System (SAGA-GIS) (Conrad et al., 2015) and the “raster calculator” tool were used to calculate the SOCS for each mapped pixel at each depth increment using Equation 5-2, and using the calculated pixel values, the total SOCS for the field was calculated. For the calculate-first approach, the SOCS for each observation was calculated using Equation 5-2, and afterwards interpolated with OK.

The OK method involved several steps: firstly, a semi-variogram was constructed to model spatial dependence (Douzouné et al., 2024). A global model was used, i.e. all 50 observation points were used for semi-variogram construction and the semi-variograms were manually fitted with the “fit.variogram” function (part of the gstat package) in RStudio (Pebesma, 2025). The models included spherical, exponential, Gaussian, Matern and Stein’s parameterization (Tziachris et al., 2017; Douzouné et al., 2024), with all models utilised and only the most optimal models shown in this paper. The most optimal model was selected based on the lowest cross-validation error and best fit to the empirical semi-variogram. Secondly, the semi-variogram models were used to estimate the kriging weights and finally these weights were applied to predict (interpolate) the SOC content, ρb and SOCS at unsampled locations. This step was conducted with the “gstat” and “predict” functions from the gstat R package (Pebesma, 2025), and no moving window was utilised (R Core Group, 2020).

For assessing the model evaluation and accuracy of the OK interpolated maps, leave-on-out cross-validation was used (Guan et al., 2017; Yao et al., 2020). The cross-validation statistics

could only be computed using observations/points for the calculate-first approach, since the map-first approach utilised already created maps. The cross-validation uses a single observation from the original sample as the testing data in which a prediction is made and the remaining observations are used as training data (Tziachris et al., 2017). This process is then repeated for all the sampling points. For the maps created using OK, commonly used accuracy indicators were used to assess the accuracy (Pouladi et al., 2023).

5.2.3.2 Machine learning approach

The mapping using DSM with ML was also done using 50 SOC content and ρb observations at each depth, as well as selected environmental covariates, which included the 30 m digital elevation model (DEM) (USGS, 2024), as well as the secondary derivatives derived from Sentinel-2 satellite images taken on 28 February 2022 and 1 July 2022. The selection of these two dates was intentional, representing different seasons: one capturing the growth of vegetation during the planting season, and the other showing soil characteristics during the dry season. For the map-first approach, the continuous SOC content and ρb maps for the three depths were created and thereafter imported into SAGA-GIS (Conrad et al., 2015). The “raster calculator” tool was used to calculate the SOCS for each mapped pixel at each depth increment using Equation 5-2, and using the calculated pixel values, the total SOCS for the field was calculated. For the calculate-first approach, the SOCS for each observation was calculated using Equation 5-2, and afterwards the SOCS predicted as discussed below.

From the DEM, the following topographic layers were derived using the “basic terrain analysis” tool in SAGA-GIS (Conrad et al., 2015): slope percentage, profile curvature, planform curvature, multi resolution index of valley bottom flatness (MRVBF), relative slope position, valley depth, channel network distance, channel network base level, LS factor and the topographical wetness index. From the satellite images, secondary derivatives (Flynn et al., 2019; Odebiri et al., 2023) were derived using the “raster calculator” tool in SAGA-GIS (Table 5-1). These secondary derivatives were selected as they are found to be some of the main environmental covariates used for SOC content and ρb mapping at field scale (Huang et al., 2022).

Table 5-1: Spectral indices derived from the satellite imagery.

Bands	Wavelength (μm)	Band number	Symbol
Blue	0.45-0.51	2	B
Green	0.53-0.59	3	G
Red	0.64-0.67	4	R
Red edge	0.68-0.71	5	RE 1
Red edge	0.73-0.78	7	RE 3
Near infrared (NIR)	0.78-0.9	8	NIR
Short wave infrared (SWIR)	1.02-1.12	11	SWIR 1
Short wave infrared (SWIR)	1.56-1.66	12	SWIR 2
Indices	Equation		Property
Brightness index (BI)	$(R^2+G^2+B^2)/3\text{^}0.5$		Reflectance
Colouration index (CI)	(R-G)/(R+G)		Soil colour
Redness index (RI)	$(R^2/B*G^3)$		Hematite
Saturation index (SI)	(R-B)/(R+B)		Spectral slope
Normalised difference vegetation index (NDVI)	(NIR-R)/(NIR+R)		Vegetation
Ratio vegetation index (RVI)	NIR/R		Vegetation
Normalised difference moisture index (NDMI)	(NIR-SWIR1)/(NIR+SWIR1)		Vegetation
Maccioni index	(RE3-RE1)/(RE3-R)		Vegetation
Green vegetation moisture index (GVMI)	(NIR-SWIR2)/(NIR+SWIR2)		Vegetation
Bare soil index (BSI)	((SWIR2-R)-(NIR-B))/((SWIR2-R)+(NIR-B))		Land use

In RStudio, the 50 observations were divided into a calibration (75%) and independent evaluation (25%) dataset using conditioned Latin hypercube sampling (cLHS) (Minasny & McBratney, 2006; Kwon & Oh, 2021), with the spectral indices employed to select the most representative observations. Although data-slitting has been shown to be problematic in studies with relatively few samples (Piikki et al., 2021), selecting an independent dataset over cross-validation ensured that the evaluation was not biased. Furthermore, cLHS ensured that the most representative samples were selected (Zizala et al., 2024), providing a more reliable and robust assessment of predictive accuracy.

Cubist, RF and SVM was the ML models used for training the calibration dataset (*n* = 37) and then testing the model on the independent evaluation dataset (*n* = 13). The "caret" package (Kuhn, 2008) was utilised to optimise the performance of all the ML models. For the SOC content and ρb maps created using DSM with ML, RMSE, R^2 and ρc were used to assess the accuracy.

As mentioned, for the calculate-first approach the SOCS were calculated using Equation 5-2. Afterwards, cLHS were used again to divide the data into a calibration (*n* = 37) and independent evaluation datasets (*n* = 13) and finally Cubist was used for training the calibration dataset and then testing the model on the evaluation dataset. Finally, the accuracy of the created SOCS (0-30 cm) maps was assessed by the same evaluation statistics, including RMSE, R^2 and ρc.

5.2.4 Calculate-first approach vs map-first approach

For evaluating the DSM with ML maps for SOCS, the independent evaluation dataset (*n* = 13) was used, and the predicted values compared to the measured values. The RMSE and ρc were calculated for each approach and a scatter plot indicating the observed versus predicted values were creating for the DSM with ML approaches. A pairwise means comparison test ("pairwise.wilcox.test" function) was also done in RStudio to determine the significance differences between the approaches for the two study sites.

5.3 Results and discussion

5.3.1 Pedo-transfer function and descriptive statistics

The accuracy and reliability of the PTF created for ρb is shown in Figure 5-2. The PTF compared favourably to other PTFs created for ρb at small scales, with Yanti et al. (2021) achieving a R^2 of 0.43 and an RMSE of 0.108 g/cm³, while Palladino et al. (2022) achieved a R^2 of 0.17 and RMSE of 0.17 g/cm³. It also compared well with another South African PTF (although on regional scale), where van Zijl et al. (2024) achieved a R^2 of 0.87 and RMSE of 0.11. This PTF was shown to be acceptably accurate and reliable, with the validation statistics indicating an RMSE of 0.06 g/cm³, R^2 of 0.87, ρc of 0.92 and RPIQ of 3.76 (Figure 5-2).

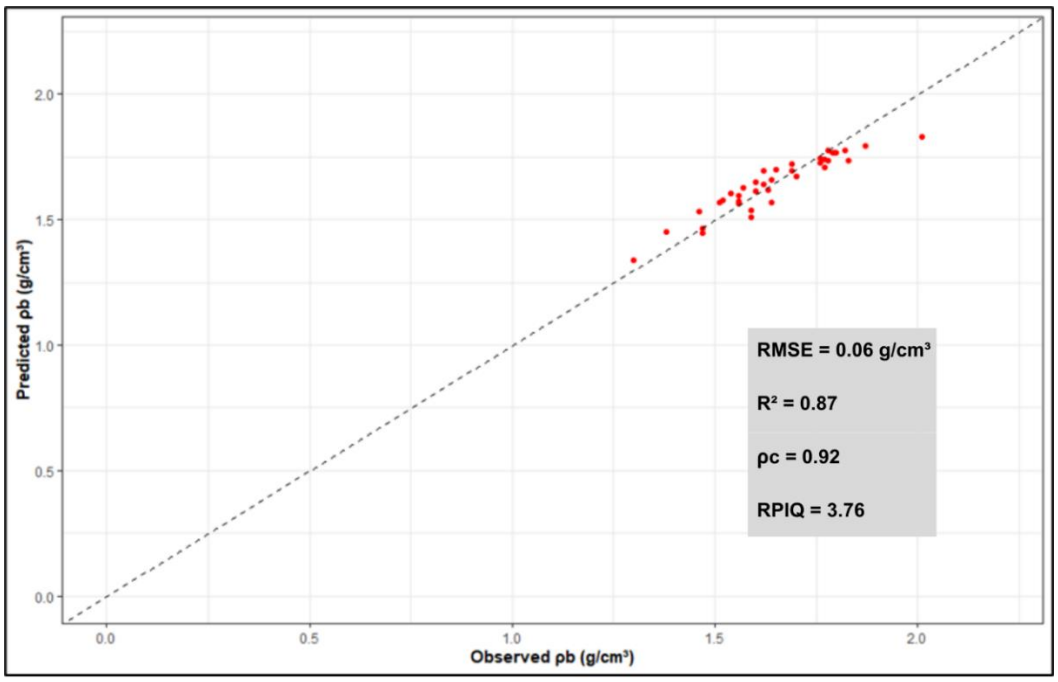


Figure 5-2: A scatter plot indicating the observed and predicted dry bulk density (ρb) values (red dots) of the validation dataset (n = 38) from the created pedo-transfer function. The dashed, black line indicates a 1:1 line.

The Vrede and Ottosdal study sites presented statistically significant differences between the two locations across different depths. The Vrede study site had higher SOC content values for all three depths compared to the Ottosdal study site, with median values of 1.37, 0.76 and 0.83% for the Vrede study site and median values of 0.6%, 0.46% and 0.46% for the three depths of the Ottosdal study site (Figure 5-3). Regarding the ρb, the Ottosdal study site showed higher values compared to Vrede study site for all three depths, with median values of 1.69, 1.75 and 1.65 g/cm³ for the three depths (Figure 5-3). The Vrede study site showed median values of 1.67, 1.5 and 1.37 g/cm³ for the three depths, respectively (Figure 5-3).

The higher SOC content and lower ρb values for the Vrede study site could likely be attributed to a combination of several factors. Firstly, the Vrede study site receives a significantly higher MAP (795 mm) compared to the Ottosdal study site (565 mm), which leads to improved moisture retention in the soil, supporting higher organic matter levels. As mentioned, the Vrede study site was characterised by clayey soils (section 5.2.1), which generally have a greater capacity for storing SOC compared to sandy soils, as those found at the Ottosdal study site. Clayey soils also generally have lower ρb compared to sandy soils, due to the clay particles having a larger surface area that holds more water and organic matter, leading to more pore space in the soil, which reduces its overall density (Brady & Weil, 2017). The regenerative agricultural practices employed at the Vrede study site, such as the use of winter cover crops, manure inputs and daily rotational grazing of sheep, likely further contributed to enhanced SOC content and improved soil structure.

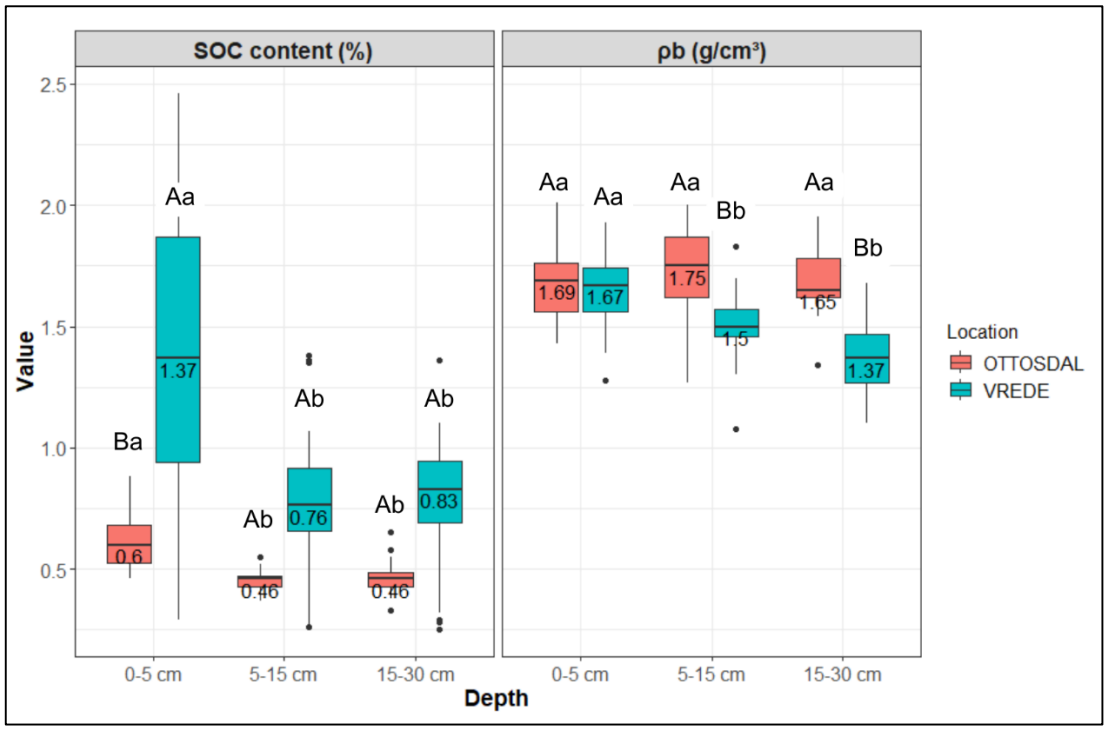


Figure 5-3: The soil organic carbon (SOC) content and dry bulk density (ρb) distribution of the three depths for the two study sites.

Overall, the QQ-plots of SOC content for the Ottosdal study site across all three depths showed strong deviation from normality (Figure 5-4). The flattened pattern of the 5-15 cm and 15-30 cm depths likely suggests a highly skewed dataset, likely caused by the values concentrated within a narrow range (as indicated by Figure 5-3). The QQ-plots for the Vrede study site showed that only the 0-5 cm depth appeared closer to a normal distribution, although some deviation existed at the tails (Figure 5-5). The QQ-plots of ρb for the Ottosdal and Vrede study sites showed relatively good alignment with the theoretical quantiles, with only minor deviations at the lower and upper tails.

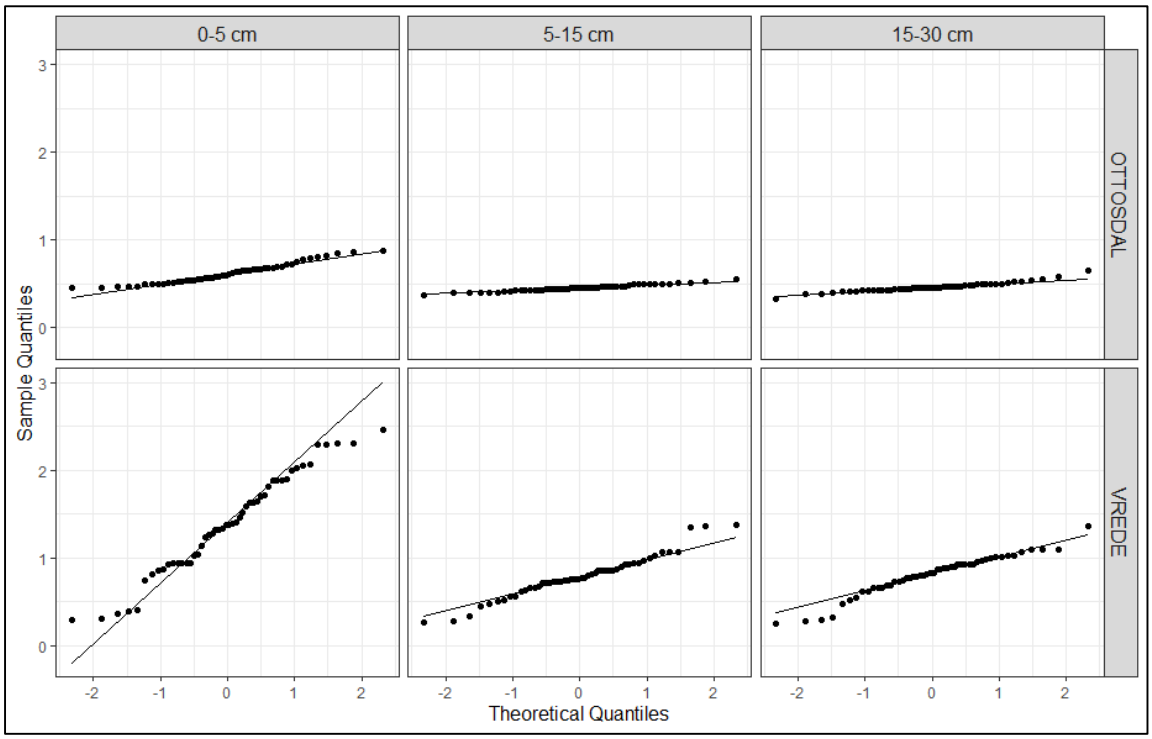


Figure 5-4: QQ-plots presenting the normality of soil organic carbon (SOC) content at different depths (0-5-, 5-15- and 15-30 cm) for the Ottosdal and Vrede study sites.

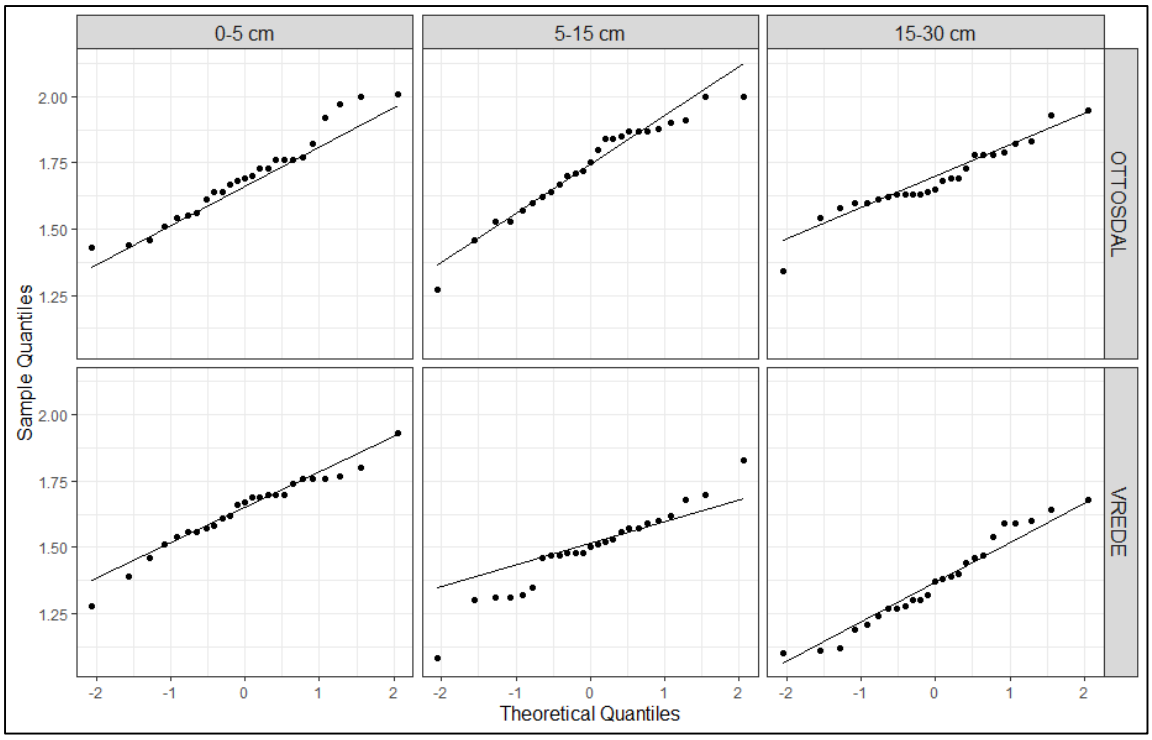


Figure 5-5: QQ-plots presenting the normality of dry bulk density (ρb) at different depths (0-5, 5-15 and 15-30 cm) for the Ottosdal and Vrede study sites.

5.3.2 Conventional mapping with ordinary kriging

5.3.2.1 Soil organic carbon content and dry bulk density

At both the Ottosdal and Vrede study sites, conventional mapping with OK did not produce statistically acceptable results (Table 5-2). Although the RMSE values for the SOC content (0.05 – 0.5%) slightly outperformed Chabala et al. (2017) and John et al. (2021), who achieved RMSE values of 0.64% and 0.58%, respectively, and was similar to that achieved by Barbouchi et al. (2020), which found an RMSE of 0.19. The R^2 of 0.6, while the R^2 and ρc values for the Ottosdal and Vrede study sites were very low, indicating that the SOC content maps did not provide an accurate representation of reality. Although the RMSE values for the ρb maps were comparable to those reported in other studies (e.g., Mouazen et al., 2006; Mousavi et al., 2017), the R^2 and ρc values for the Ottosdal and Vrede study sites were significantly lower, also indicating that the ρb maps maps did not represent reality.

Table 5-2: The cross-validation statistics of soil organic carbon (SOC) content and dry bulk density (ρb) for the two study sites.

SOC content				ρb		
Ottosdal						
Depth (cm)	**RMSE (%)**	**R^2**	**ρ_c**	**RMSE (g/cm³)**	**R^2**	**ρ_c**
0-5	0.12	-0.28	-0.01	0.12	-0.16	-0.06
5-15	0.04	-0.46	-0.08	0.14	0.05	0.03
15-30	0.05	0.2	0.12	0.11	-0.02	-0.02
Vrede						
Depth (cm)	**RMSE (%)**	**R^2**	**ρ_c**	**RMSE (g/cm³)**	**R^2**	**ρ_c**
0-5	0.5	0.48	0.39	0.12	-0.03	-0.01
5-15	0.2	0.54	0.46	0.12	0.09	0.01
15-30	0.21	0.41	0.31	0.13	0.36	0.25

The provided maps (Figure 5-10a – d) are examples of the same spatial patterns observed for all the created SOC content and ρb maps. The inadequate representation of the maps could be attributed to the minimal semi-variance increase with distance (Figure 5-6 – Figure 5-9), suggesting very weak spatial dependency in the data. This weak structure implies that the spatial autocorrelation is low or that the spatial structure could not be well captured by the very few point-based observations (Zhu & Lin, 2010). This likely led to inaccurate, patchy SOC content and ρb maps for the Ottosdal study site, as seen in Figure 10a and Figure 10c. Another effect of weak spatial dependency can also lead to some maps appearing overly smooth due to the averaging effect (Yamamoto, 2005), seen for both the SOC content map (Figure 5-10b) and ρb map (Figure 5-10d). However, the SOC content maps from the Vrede study site did have stronger spatial dependency compared to the Ottosdal study site, which likely led to improved evaluation statistics

(Table 5-2). Considering the poor cross-validation statistics and the either patchy-looking or overly smoothed maps, the use of OK for mapping SOC content was considered as inaccurate and is not recommended at field scale.

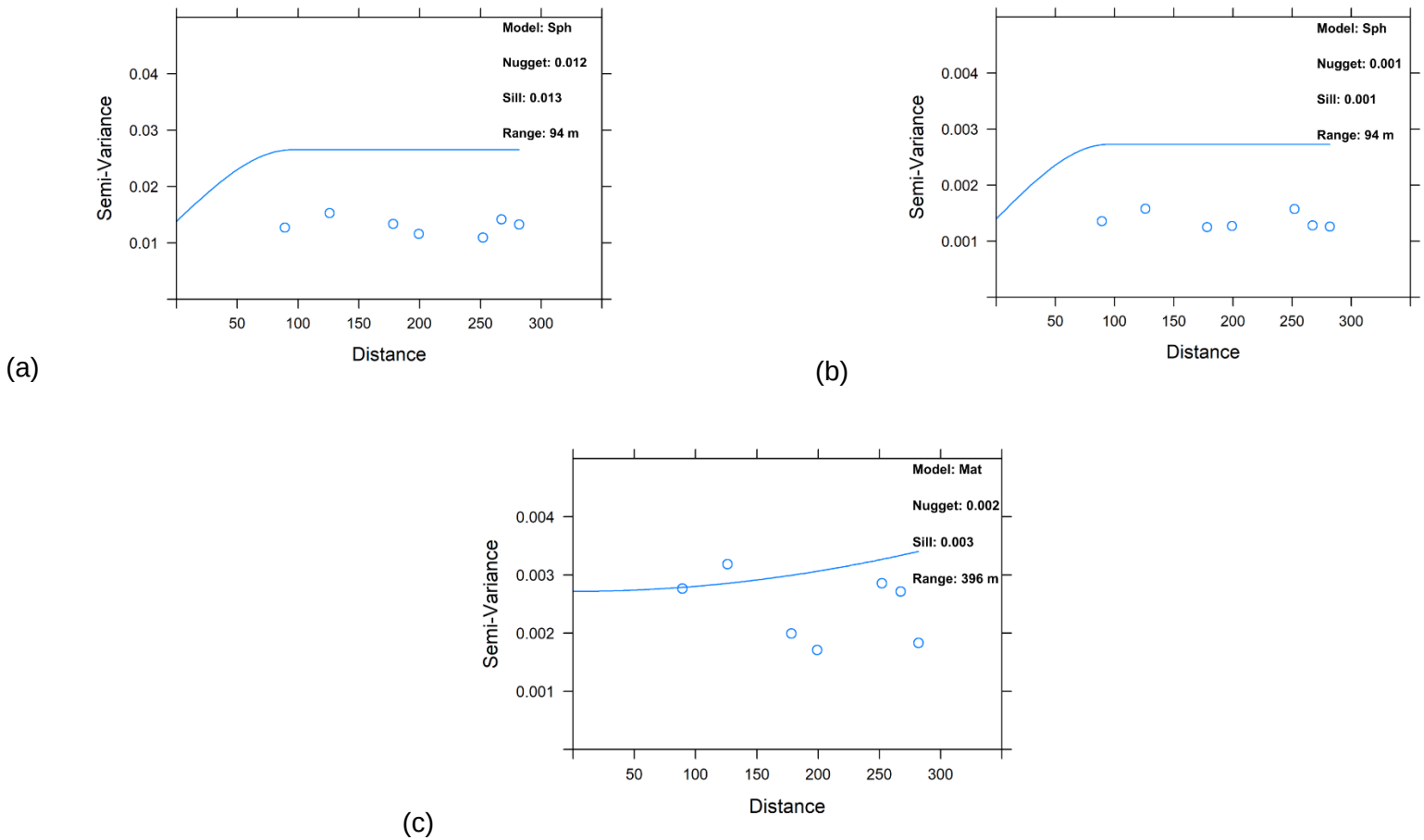


Figure 5-6: The constructed semi-variograms for the Ottosdal study site of soil organic carbon (SOC) content at (a) 0-5 cm, (b) 5-15 cm and (c) 15-30 cm.

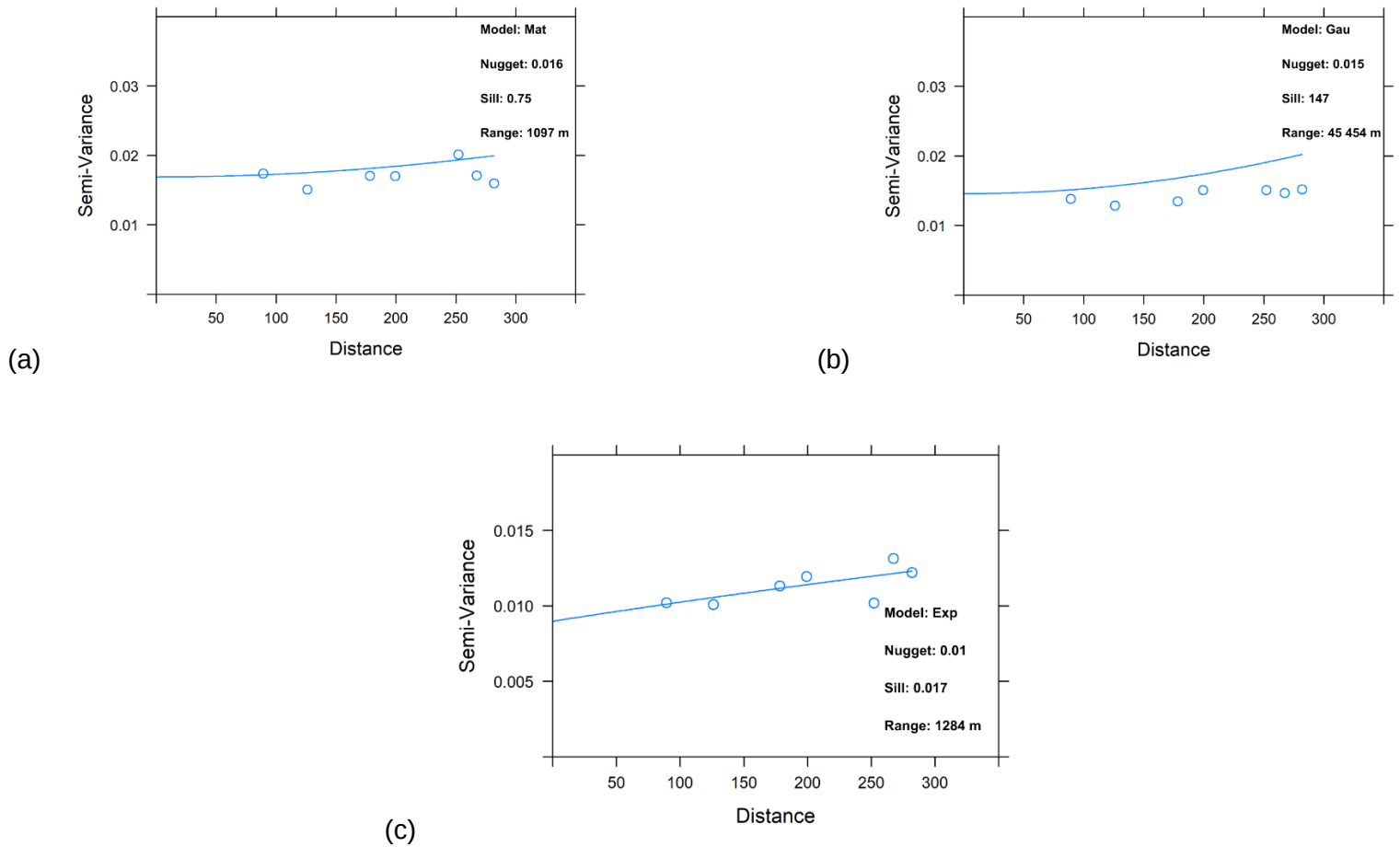


Figure 5-7: The constructed semi-variograms for the Ottosdal study site of dry bulk density (ρb) at (a) 0-5 cm, (b) 5-15 cm and (c) 15-30 cm.

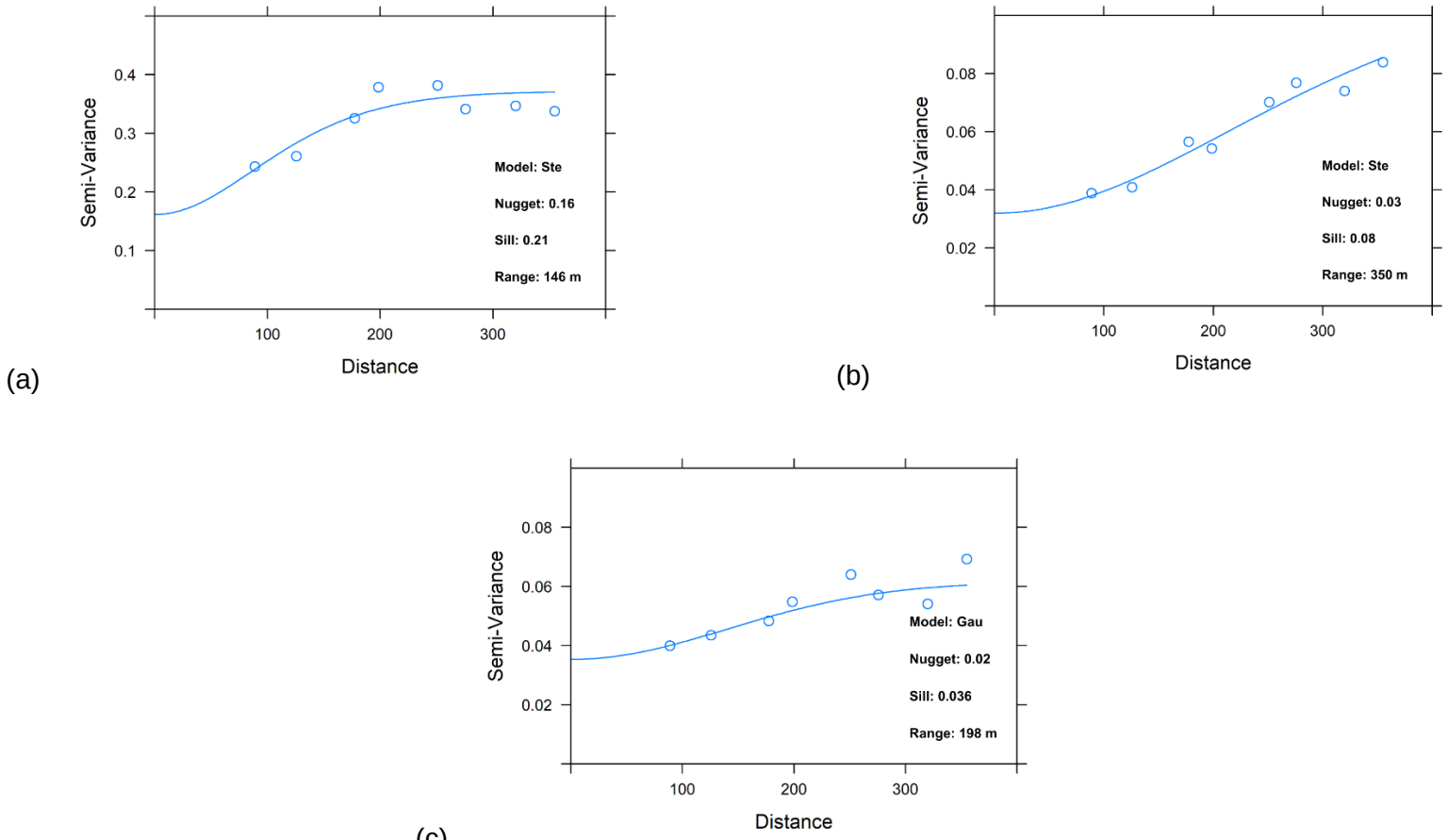


Figure 5-8: The constructed semi-variograms for the Vrede study site of soil organic carbon (SOC) content at (a) 0-5 cm, (b) 5-15 cm and (c) 15-30 cm.

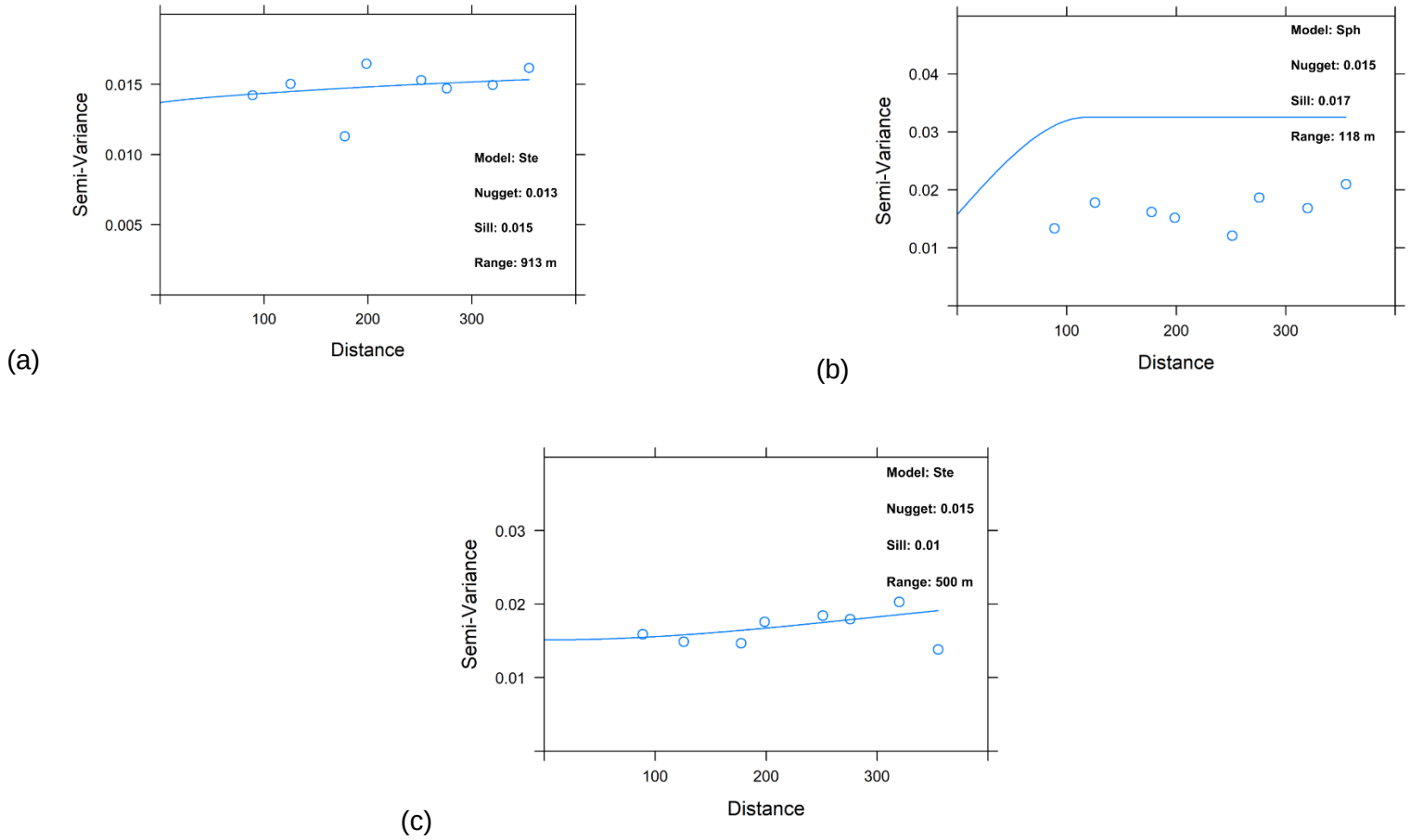


Figure 5-9: The constructed semi-variograms for the Vrede study site of dry bulk density (ρb) at (a) 0-5 cm, (b) 5-15 cm and (c) 15-30 cm.

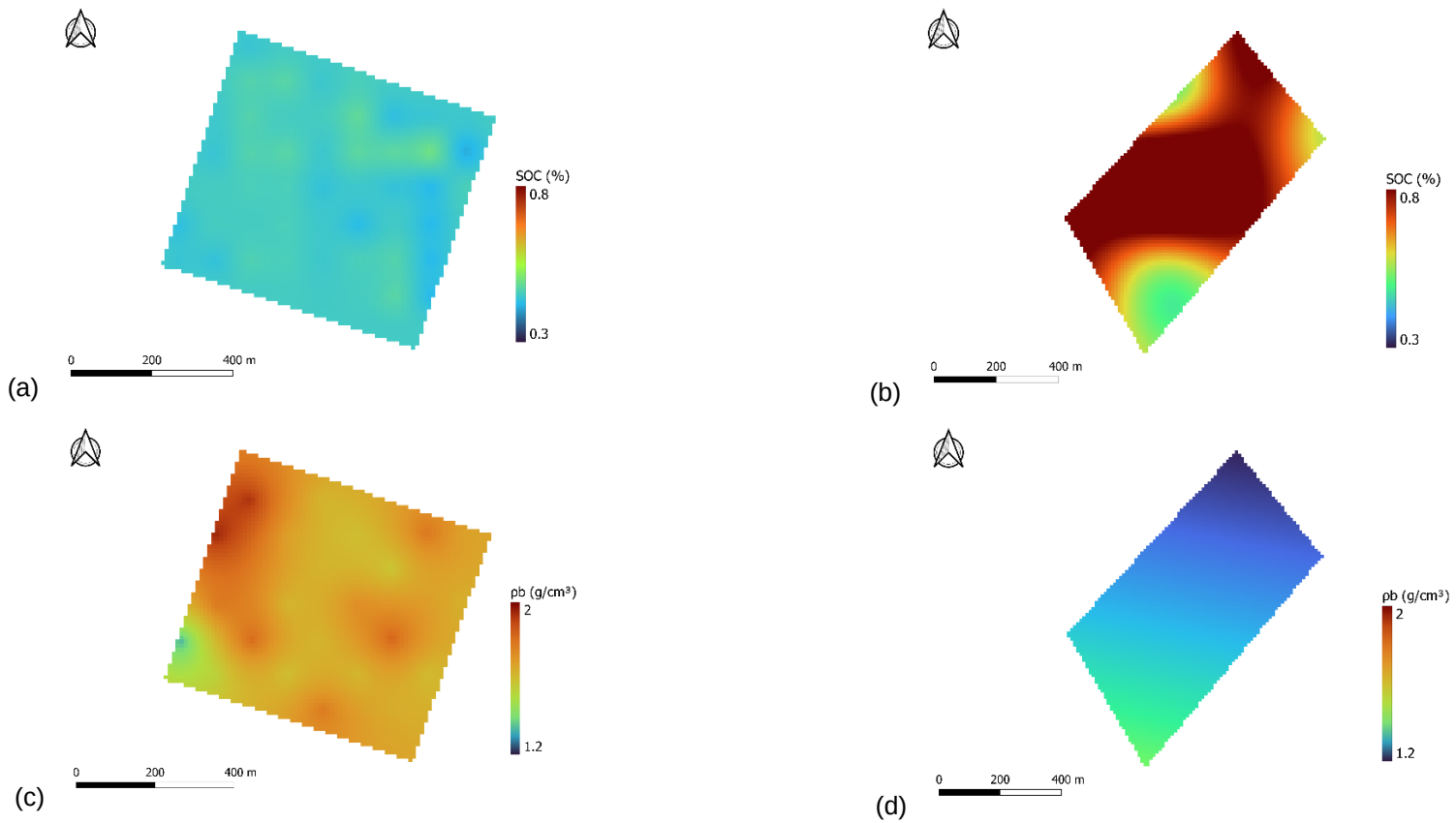


Figure 5-10: The maps created using ordinary kriging for (a) 5-15 cm soil organic carbon (SOC) content from the Ottosdal study site, (b) 5-15 cm soil organic carbon content (SOC) content from the Vrede study site, (c) 15-30 cm dry bulk density (ρb) from the Ottosdal study site and (d) 15-30 cm dry bulk density (ρb) from the Vrede study site.

5.3.2.2 Soil organic carbon stocks

The calculate-first approach and the map-first approach using OK produced very similar total SOCS. For the Ottosdal study site, the calculate-first approach produced a value of 1 000 tons SOCS, while the map-first approach produced a value of 1 004 tons SOCS. For the Vrede study site the same trend of a negligible increase in SOCS with the map-first approach was observed, as the calculate-first approach produced a value of 1 533 tons SOCS, and the map-first approach produced a value of 1 553 tons SOCS.

At both the Ottosdal and Vrede study sites, conventional mapping SOCS with OK did not produce statistically acceptable results (Table 5-3). In terms of RMSE, the Ottosdal study site (RMSE = 2.27 t/ha) outperformed Li et al. (2023), Sherpa et al. (2016) and Kumar & Lal (2011), with values of 2.9, 17.21 and 3.56 t/ha, respectively. However, the R^2 and ρc values were very low for the Ottosdal study site compared to the other studies (Hoffman et al., 2014; Duan et al., 2020) with $R^2 > 0.38$. In contrast, the Vrede study site had a very poor RMSE (8.24 t/ha) and moderate R^2 and ρc values (Table 5-3), which outperformed the other mentioned studies.

Table 5-3: The calculate-first approach cross-validation statistics of the soil organic carbon stocks (SOCS) (0-30 cm) for the Ottosdal and Vrede study sites.

Cross-validation statistics of SOCS			
Ottosdal			
Depth (cm)	**RMSE (t/ha)**	**R^2**	**ρc**
0-30	2.27	-0.36	-0.05
Vrede			
Depth (cm)	**RMSE (t/ha)**	**R^2**	**ρc**
0-30	8.24	0.62	0.58

Although the two study sites did perform better regarding some evaluation statistics, the semi-variogram parameters indicate that mapping with OK also showed weak spatial dependency for the Ottosdal study site (Figure 5-11b). The same patchy effect can also be seen for the Ottosdal study site map (Figure 5-11a), which indicate that the spatial structure could not be well captured by the point-based observations (Zhu & Lin, 2010). On the other hand, the SOCS for the Vrede study site (Figure 5-11c) did not provide the over-smoothing effect as seen from the previous maps (Figure 5-10b and Figure 5-10d), however, the poor RMSE value still led to the conclusion that the map provides an inaccuracy too large to be used.

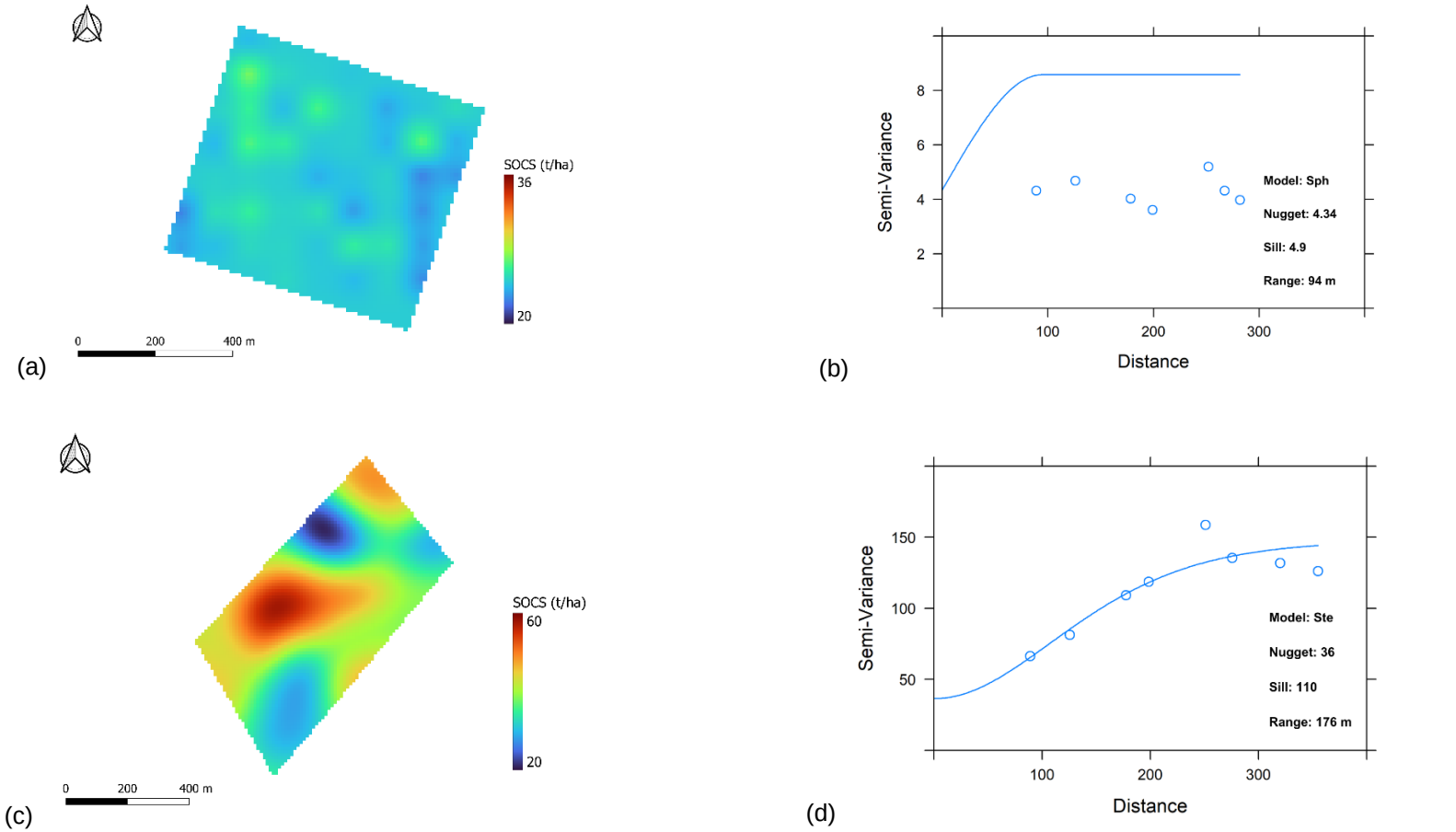


Figure 5-11: The (a) soil organic carbon stocks (SOCS) map for the Ottosdal study site and the (b) constructed semi-variogram for the Ottosdal site, as well as the (c) soil organic carbon stocks (SOCS) map for the Vrede study site and the (d) constructed semi-variogram for the Vrede study site.

5.3.3 Digital soil mapping with machine learning

5.3.3.1 Soil organic carbon content and dry bulk density

Overall, the maps produced from DSM with ML (Table 5-4) outperformed the maps produced with the conventional method using OK in terms of evaluation statistics (Table 5-2). However, the overall evaluation statistics still indicate relatively poor performance and need for improved maps. The only exception was for the SOC content at the 5-15 cm depth for the Vrede study site, where the OK performed slightly better RMSE (0.2% compared to 0.21%) and R^2 (0.54 compared to 0.46). For ρb, the 15-30 cm depth from the Vrede study site also had a slightly better RMSE (0.11% compared to 0.12%) using OK. The provided maps (Figure 5-12a – d) are examples of maps created using DSM with ML and are also the same depth maps created using OK (Figure 5-10a – d). The visual finding is that the maps created using DSM with ML does not indicate the same patchy or overly smooth look as with the maps created using OK. However, some maps, such as Figure 5-12a, did also provide overly smoothed maps, likely due to an averaging effect applied by the Cubist model to reduce prediction error and smooth the results (Khaledian & Miller, 2020). Some maps (Figure 5-12b) presented the appearance of specific dots on the map and could likely be attributed to one or more covariates that were heavily utilised by the Cubist model for predictions in these locations. It is important to note that these dots are different to the patchy appearance of Ottosdal study site maps created using OK, where the patches appear at the location of the datapoints, artificially increasing the accuracy of the maps. From the maps created using DSM with ML, the dots were likely due to a covariate layer with which the SOC content was correlated.

OK relies on the assumption that spatial autocorrelation (similarity of values at nearby locations) is consistent and stationary (Keskin et al., 2019). However, this might be effective when this assumption holds true, but OK may struggle to capture complex relationships between soil properties (SOC content and ρb) and environmental factors. In contrast, ML models, such as Cubist, were likely able to capture these non-linear, complex relationships by utilising the environmental covariates (Akpa et al., 2016; Taghizadeh-Mehrjardi et al., 2021; Taghipour et al., 2022). Previous studies (Minasny et al., 2013) have found that for small scale areas (resolution < 100 m), spatial topsoil SOC content correlates well with vegetation and local terrain attributes (i.e. slope, aspect, curvatures), which was likely also the case for this study, resulting in DSM with ML outperforming OK.

Cubist was the overall best performing ML model for mapping SOC content and ρb from both study sites (Table 5-4). However, RF did outperform Cubist regarding the 0-5 cm and 5-15 cm ρb maps of the Vrede study site. Cubist performed better than RF and SVM likely due to the model effectively capturing non-linear relationships and threshold effects in the data, by using a

combination of “if-then” rules and linear regression adjustments (Peng et al., 2015; Kock, 2022). Cubist is also one of the most common ML models used for accurately mapping SOC content and have shown to outperform other ML models, such as RF and SVM (Lamichhane et al., 2019; Huang et al., 2022). RF, while robust and flexible, may not have captured these patters due to its reliance on non-linear decision trees, without the benefit of explicit rule-based splitting and linear corrections within segments (Lamichhane et al., 2019; Sharififar et al., 2019; Bai et al., 2022). RF and SVM are also two ML models that are susceptible to overfitting data (Hernandez et al., 2009; Keskin et al., 2019). This result was also in contrast to a recent study done by Nozari et al. (2024), which found that on regional scale RF outperformed Cubist slightly for creating SOC content maps.

Since OK need reliable semi-variograms and dense observations, DSM with ML is recommended for mapping SOC content and ρb at field scale in South Africa. However, both methods need to be improved to be sufficiently accurate to be used at field scale. For improving the OK maps, soil samples need to be taken much closer to one another, which would increase the spatial dependency. For the Vrede study site, the sampling distance would be approximately 73 m, which is 50% of the shortest range of the semi-variograms (146 m for 0-5 cm), following the norm of Flatman & Yfantis (1984). To improve the DSM with ML maps, more sampling and covariate data is needed. The ML models usually require large datasets to work optimally (Sarker, 2021) and the few observations (n = 50 / study site) included for this study was suboptimal to be fully utilised. Since mapping SOC content and ρb at is challenging (Huang et al., 2022), a viable option to increase data, but decrease costs are to map SOC content at farm scale rather than field scale, as it provides the opportunity to collect measurements at more locations, while decreasing the sampling density. The most substantial improvement will be gained by incorporating crop yield, as measured by the combine harvester, as the biomass produced should be correlated to the SOC content (Zhang, 2021b). Other field scale covariates, such as fine-resolution DEM, as determined through precision equipment on a tractor (Yang et al., 2019), electromagnetic induction (Doolittle & Brevik, 2014) and gamma-radiometric measurements (Dierke & Werban, 2013) would greatly improve the DSM with ML maps.

Table 5-4: The validation statistics of soil organic carbon (SOC) content and dry bulk density (ρb) for the two study sites. The green highlighted cells show the best performing models.

SOC content							ρb					
	Ottosdal			Vrede			Ottosdal			Vrede		
Random Forest (RF)												
Depth (cm)	**RMSE (%)**	**R^2**	**ρc**	**RMSE (%)**	**R^2**	**ρc**	**RMSE (g/cm³)**	**R^2**	**ρc**	**RMSE (g/cm³)**	**R^2**	**ρc**
0-5	0.1	0.17	0.21	0.4	0.42	0.37	0.11	0.01	-0.07	0.14	0.35	-0.18
5-15	0.03	0.21	0.23	0.2	0.46	0.39	0.11	0.11	0.16	0.07	0.21	0.37
15-30	0.04	0.07	0.23	0.18	0.43	0.32	0.15	0.04	-0.02	0.13	0.44	0.24
Support Vector Machines (SVM)												
Depth (cm)	**RMSE (%)**	**R^2**	**ρc**	**RMSE (%)**	**R^2**	**ρc**	**RMSE (g/cm³)**	**R^2**	**ρc**	**RMSE (g/cm³)**	**R^2**	**ρc**
0-5	0.11	0.07	0.13	0.4	0.43	0.33	0.11	< 0.001	< 0.001	0.1	0.003	-0.03
5-15	0.03	0.1	0.19	0.21	0.36	0.32	0.11	0.12	0.23	0.09	0.18	0.27
15-30	0.05	0.06	0.23	0.18	0.36	0.33	0.16	0.02	-0.05	0.14	0.21	0.19
Cubist												
Depth (cm)	**RMSE (%)**	**R^2**	**ρc**	**RMSE (%)**	**R^2**	**ρc**	**RMSE (g/cm³)**	**R^2**	**ρc**	**RMSE (g/cm³)**	**R^2**	**ρc**
0-5	0.09	0.56	0.66	0.44	0.76	0.58	0.12	0.11	-0.29	0.16	0.02	-0.1
5-15	0.03	0.42	0.51	0.21	0.46	0.48	0.11	0.12	0.24	0.18	0.001	0.02
15-30	0.04	0.27	0.41	0.16	0.5	0.6	0.17	0.11	-0.15	0.12	0.57	0.36

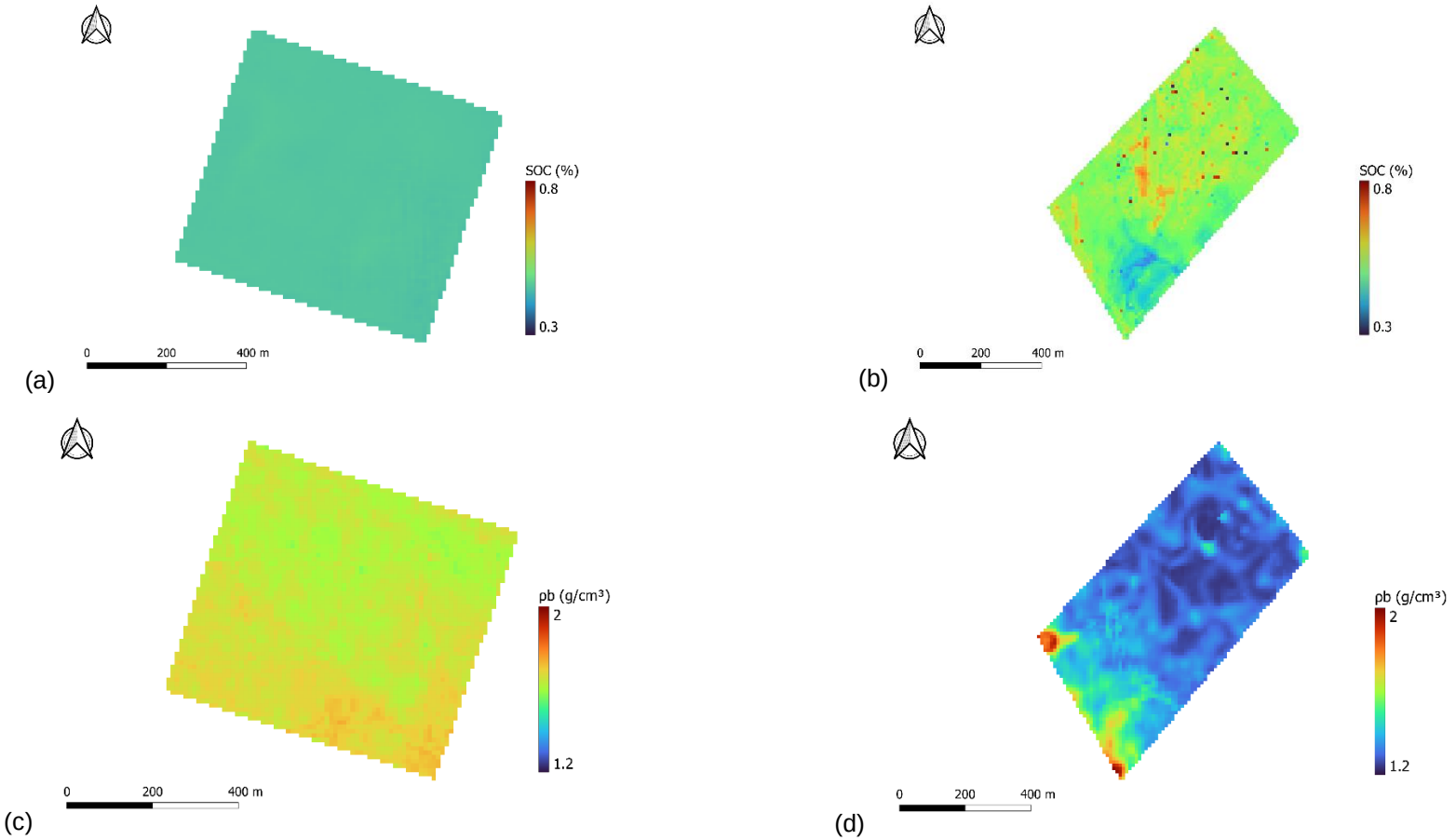


Figure 5-12: The maps created using digital soil mapping with machine learning for (a) 5-15 cm soil organic carbon (SOC) content from the Ottosdal study site, (b) 5-15 cm soil organic carbon (SOC) content from the Vrede study site, (c) 15-30 cm dry bulk density (ρb) from the Ottosdal study site and (d) 15-30 cm dry bulk density (ρb) from the Vrede study site.

5.3.3.2 Soil organic carbon stocks

The calculate-first approach and the map-first approach using DSM with ML produced very similar total SOCS, however in contrast to OK, the calculate-first approach had negligibly higher SOCS values. For the Ottosdal study site, the calculate-first approach produced a value of 1 001 tons SOCS, while the map-first approach produced a value of 990 tons SOCS. For the Vrede study site the calculate-first approach produced a value of 1 574 tons SOCS, and the map-first approach produced a value of 1 562 tons SOCS.

Overall, the created SOCS maps (Figure 5-13 and Figure 5-14) were improved from the maps created using OK, however, the evaluation statistics indicate that the map accuracy need to be improved. Although the created SOCS maps using DSM with ML only utilised the evaluation dataset (n = 13), it produced better evaluation statistics (Table 5-5) compared to the SOCS maps created using OK (Table 5-3). All the evaluation statistics were better using DSM with ML, only the R^2 value for the Vrede study site, were slightly better using OK (R^2 = 0.64) compared to DSM with ML (R^2 = 0.52). There was also a greatly improved RMSE of the map created using DSM with ML for the Vrede study site, with a RMSE of 6.56 t/ha, compared to the 8.24 t/ha using OK. Visually, the SOCS map created for the Ottosdal study site using DSM with ML (Figure 5-13a) did not show any of the patchy-looking areas as with the SOCS map created using OK (Figure 5-11a). Although the evaluation statistics for the Vrede study site using DSM with ML performed better compared to OK (Table 5-3 and Table 5-5), visually the DSM with ML map also presented the appearance of specific dots on the map (Figure 5-14a) and could likely be attributed to one or more covariates that were heavily utilised by the Cubist model for predictions in these locations, as was previously discussed. Therefore, OK produced a better-looking map compared to DSM with ML.

Table 5-5: The validation statistics of soil organic carbon stocks (SOCS) for the two study sites using the calculate-first approach.

SOCS			
Ottosdal			
Depth (cm)	**RMSE (t/ha)**	**R^2**	**ρc**
0-30	2.2	0.19	0.35
Vrede			
Depth (cm)	**RMSE (t/ha)**	**R^2**	**ρc**
0-30	6.56	0.52	0.61

The reason for the patchy-looking maps was likely the same as with the SOC content and ρb, where OK relies on the assumption that spatial autocorrelation is consistent and stationary

(Keskin et al., 2019; Farooq et al., 2022). The Cubist model was likely able to capture the non-linear, complex relationships between the SOCS and environmental covariates, leading to more accurate maps. Since OK need reliable semi-variograms and dense observations for accurate maps, DSM with ML is recommended for mapping SOCS at field scale in South Africa. As mentioned, for improving SOCS mapping using OK as the conventional method, observations could be taken closer together, improving the spatial dependency. For improving SOCS mapping using DSM with ML, more detailed datasets, such as, such as finer-resolution DEMs, yield data, electromagnetic induction and gamma-ray data could be acquired, and focus could be shifted to farm scale instead of field scale.

Overall, the two study sites had better RMSE values compared to other studies, such as Deng et al (2018), Hinge et al. (2018), Gomes et al. (2019) and Negassa et al. (2023) that achieved RMSE values > 10 t/ha. Deng et al. (2018) did achieve a better R^2 of 0.76, with the other mentioned studies achieving relatively the same performance compared to the Ottosdal and Vrede study sites, with R^2 values between 0.3 and 0.5. Two recent studies done by Taghizadeh-Mehjardi et al. (2020) and Odebiri et al. (2023) found that the use of deep learning models has emerged as innovative analytical strategies for improving SOCS mapping. Odebiri et al. (2023) utilised deep learning models, such as Concrete Autoencoder-Deep Neural Network, Boruta-Deep Neural Network, and Deep Neural Network and achieved RMSE values of 7.91–10.71 t/ha and ρc values of 0.79 – 0.85. This might indicate that the use of deep learning models can further improve mapping SOCS at field scale in South Africa.

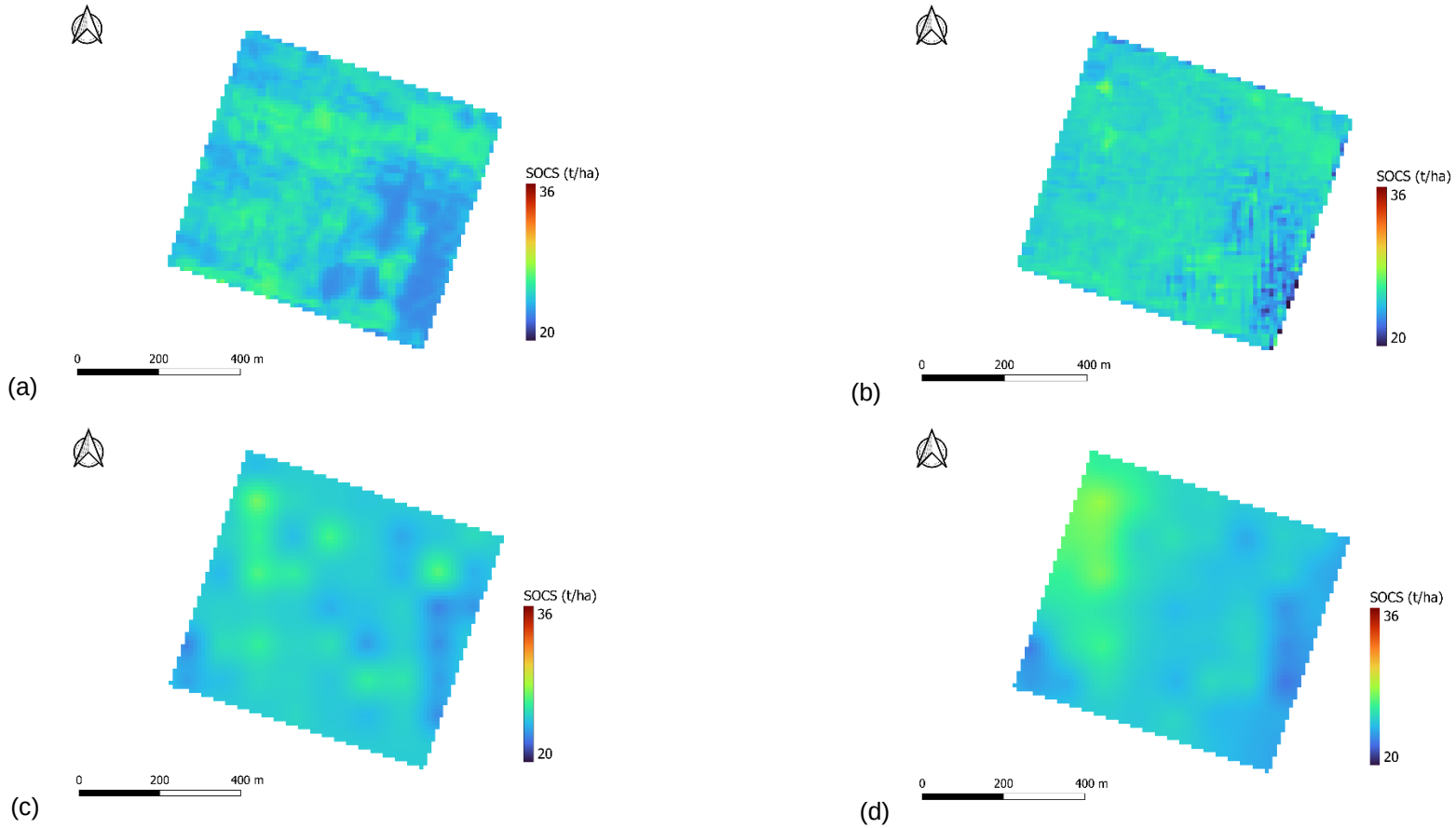


Figure 5-13: The soil organic carbon stocks (SOCS) maps created for the Ottosdal study site using (a) the calculate-first approach using digital soil mapping (DSM) with machine learning (ML) (b) the map-first approach using digital soil mapping (DSM) with machine learning (ML) (c) the calculate-first approach using ordinary kriging (OK) (d) the map-first approach using ordinary kriging (OK).

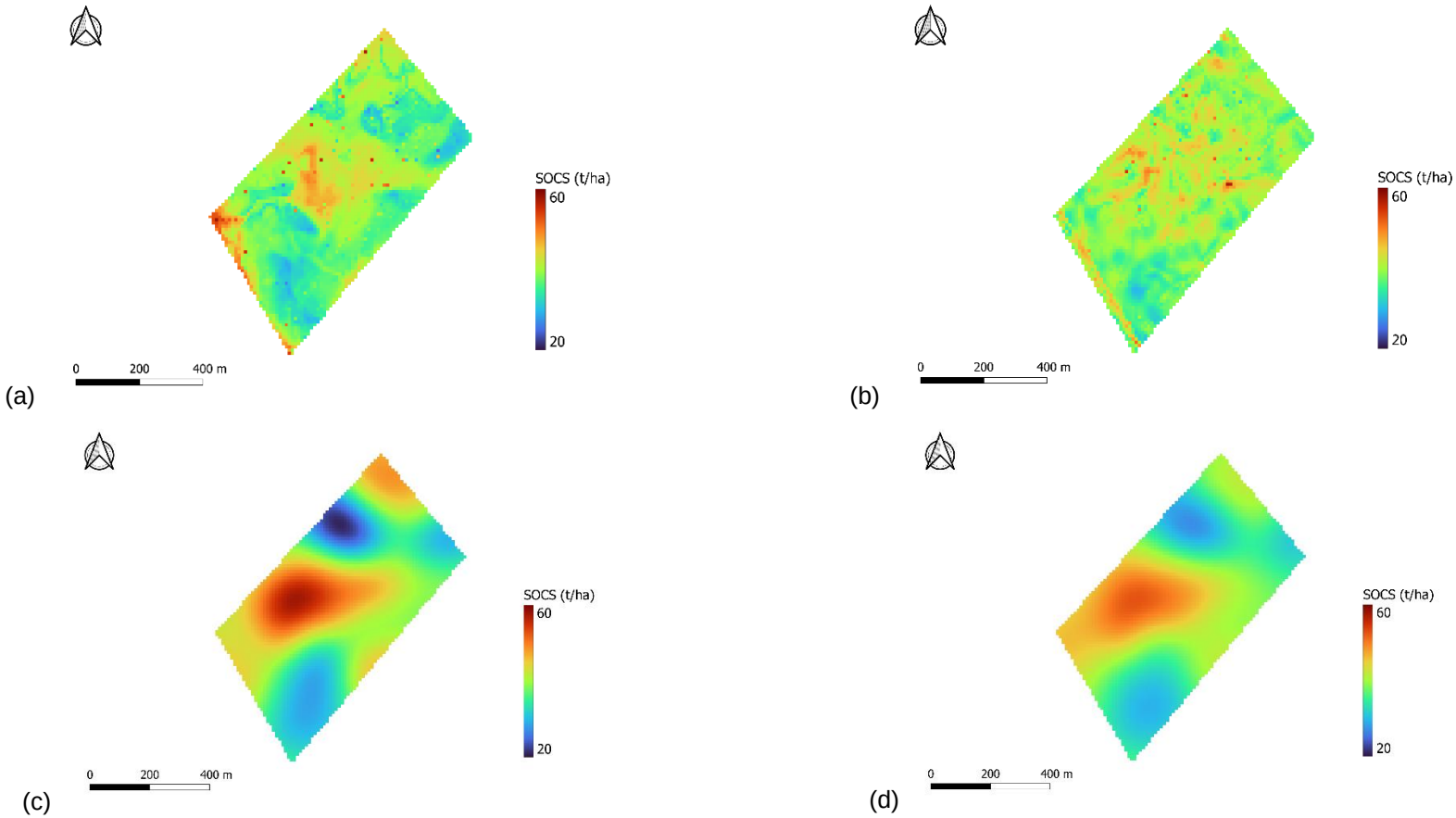


Figure 5-14: The soil organic carbon stocks (SOCS) maps created for the Vrede study site using (a) the calculate-first approach using digital soil mapping (DSM) with machine learning (ML) (b) the map-first approach using digital soil mapping (DSM) with machine learning (ML) (c) the calculate-first approach using ordinary kriging (OK) (d) the map-first approach using ordinary kriging (OK).

5.3.4 Calculate-first approach vs map-first approach

Overall, the two approaches for both OK and DSM with ML yielded very similar total SOCS for each study site (Figure 5-15). However, using OK, the map-first approach yielded higher total SOCS, while using DSM with ML, the calculate-first approach yielded higher total SOCS (Figure 5-15).

When comparing the calculate-first approach with the map-first approach, the latter achieved slightly better performance regarding ρc values for both study sites (Figure 5-16), while the RMSE where slightly lower for the Map-first approach at the Ottosdal study site. There were also no statistically significant differences found between the two approaches for both study sites, with p-values > 0.05. Although the calculate-first approach achieved a RMSE of 2.27 t/ha and the map-first approach achieved a ρc of 0.21 t/ha for the Ottosdal study site, the results from the Vrede study site showed that the calculate-first approach was slightly more accurate (ρc = 0.41) and with a lower RMSE (7.89 compared to 8.18 t/ha).

Overall, it was inconclusive which approach should be selected when utilising the conventional method of OK or DSM with ML. The calculate-first approach is able to utilise ML and capture complex relationships between SOCS and environmental covariates (Lamichhane et al., 2019; Sharififar et al., 2019; Bai et al., 2022), which could lead to more accurate maps. This approach also incorporates both SOC content and ρb field measurements, which could lead to more reliable and accurate SOCS estimations. However, collecting field data for both SOC content and ρb can be labour-intensive, costly and logistically challenging over large areas or difficult terrain. In contrast, the map-first approach may be faster and suitable in areas where accurate and reliable maps of SOC content and ρb already exist, and the relationship between these variables and the environmental factors are relatively simple and linear. However, the map-first approach is dependent on the accuracy of the previously created maps, which may introduce errors and likely lead to less accurate and reliable maps. As the results were inconclusive, more research is needed to determine which of these two approaches should be selected when utilising OK or DSM with ML. However, the study concluded that the calculate-first approach should be endorsed, due to this approach being simpler and might lead to more reliable and accurate SOCS estimations.

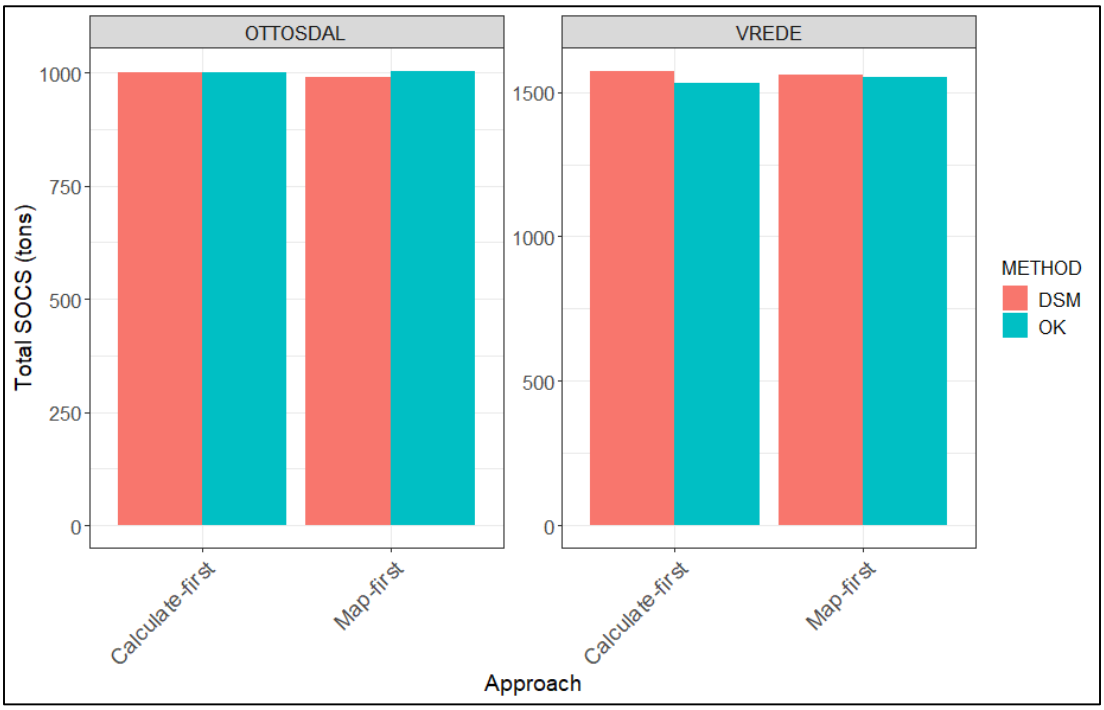


Figure 5-15: The total soil organic carbon stocks (SOCS) when using OK and DSM with ML with the calculate-first and map-first approaches.

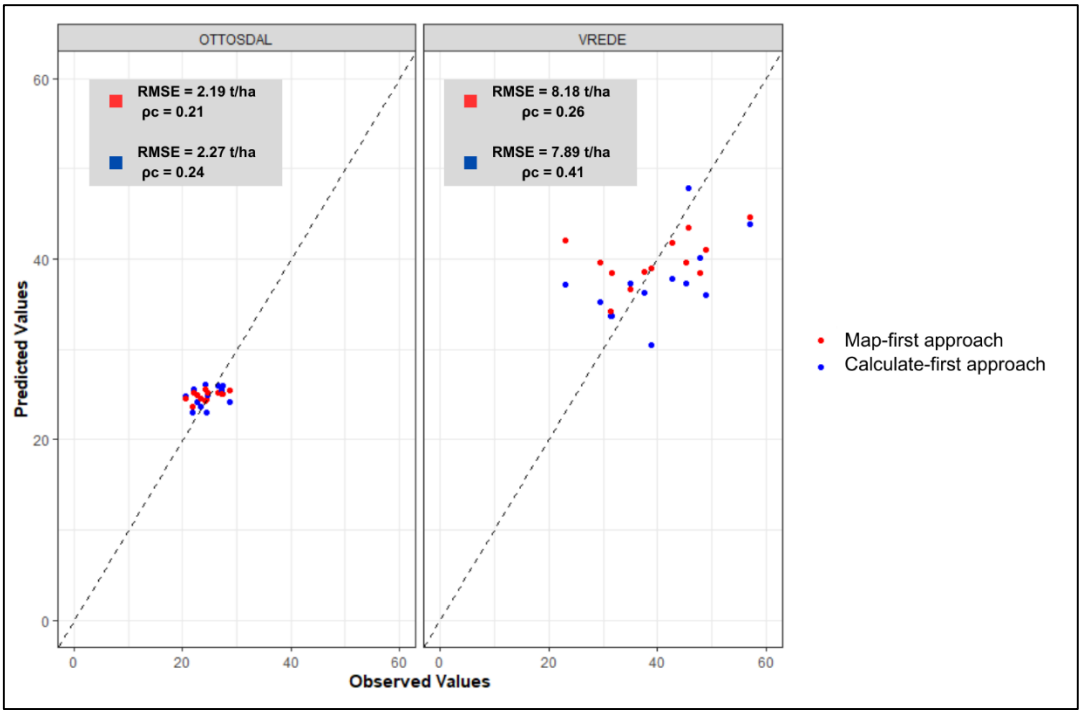


Figure 5-16: Two scatter plots indicating the observed vs predicted values using the calculate-first approach and map-first approach for the Ottosdal and Vrede study site.

5.4 Conclusion

The findings in this study demonstrated that DSM with ML should be used rather than OK for mapping SOCS at field scale. The results also showed the conventional method using OK for mapping SOCS were not sufficient for mapping SOCS at field scale. However, the results indicated that the SOC content, ρb and SOCS maps for DSM with ML needs to be improved. It was shown that the conventional method using OK yielded less accurate evaluation statistics and inaccurate maps, however, the SOCS map for the Vrede study site using DSM with ML did present specific dots, which made the OK map visually better-looking. Since, OK relies on the assumption that spatial autocorrelation (similarity of values at nearby locations) is consistent and stationary, the spatial structure could not be well captured by the point-based observations and let to either patchy-looking or overly smoothed SOC content, ρb maps. The results were inconclusive regarding whether the calculate-first approach or the map-first approach should be used when utilising DSM with ML. Although both approaches have several advantages and disadvantages, more research is needed to determine which approach should be used when utilising OK or DSM with ML. However, the study concluded that the calculate-first approach should be endorsed, due to this approach being simpler and might lead to more reliable and accurate SOCS estimations.

To increase the accuracy of maps generated using DSM with ML, additional soil samples could be collected, and field scale covariates should be incorporated. This could include finer-resolution DEMs, yield data, electromagnetic induction and gamma-ray data. Moreover, since mapping SOC content and ρb at field scale is challenging efforts should focus on mapping these properties at farm scale instead. The use of deep learning models could also further improve mapping SOCS at field scale in South Africa. Finally, future studies should also focus on utilising the most applicable environmental covariates for optimal soil property mapping.

CHAPTER 6 MODEL VS MEASURE: COMPARING SOIL ORGANIC CARBON STOCKS QUANTIFICATION APPROACHES AT FIELD SCALE IN SOUTH AFRICA

6.1 Introduction

Soil organic carbon (SOC) plays a crucial role in the global carbon cycle, with an estimated 1 500 to 2 400 gigatons of organic carbon in the top meter of soils worldwide (Hartemink & McSweeney, 2014; Paustian et al., 2019; Smith et al., 2020) To mitigate carbon emissions, carbon sequestration (the long-term storage of carbon in oceans, soils, vegetation, and geological formations) has garnered increasing attention in recent years (Baurov, 2021; Zhang et al., 2021; Batool et al., 2024). The trading of carbon credits to the government and businesses has become a major focus in recent years (Sharma et al., 2021). In South Africa, farmers and land managers have the opportunity to generate income by trading these carbon credits. Given its relevance, appropriate methods for modelling and monitoring SOC content and soil organic carbon stocks (SOCS) are also becoming very important.

There is an ongoing debate about the most appropriate approach for quantifying changes in SOCS (Paustian et al., 2016; Smith et al., 2020; FAO, 2019). Some authors argue that modelling approaches are more cost-effective than direct re-measurements (Smith et al., 2020), while others argue that direct measurements are the most accurate and reliable approach. There is a need to evaluate the costs of sampling against the value of SOCS sequestered and search for trade-offs between costs involved and alternative SOCS estimation methods (Smith et al., 2020). Accuracy is usually the key focus of scientific studies and carbon credit standards that follows a direct re-measurements approach. For carbon credit standards that follows a modelling approach, the key focus is on the certainty that temporal (e.g. annual) changes in SOCS can be predicted over time. This includes the extent that the certainty criterion is met for: (i) equilibrium baseline conditions, (ii) “historic” re-measured or modelled changes in SOCS, and (iii) changes in SOCS predicted at 5 years initially, proceeded by successive 5-year intervals (at 10, 15 and 20 years). Finally, the determined or predicted carbon credits (with either re-measure or modelling approach) must be reduced if the certainty criterion is not met, based on the prescribed rules of a standard methodology.

Modelling approaches involve using numerical models, such as Rothamsted carbon (RothC) CENTURY and DayCent, to predict how changes in management practices or land-use affect SOCS over time. The RothC model is classified as a level 2 “soil” model, which refers to process-oriented models that take SOC dynamics into account, but do not simulate other complex

processes, such as plant above- or belowground biomass growth or nutrient dynamics (IPCC, 2006; FAO, 2019). RothC is commonly used for predicting changes in SOCS over time in non-waterlogged topsoil and allows for the including of the effects of soil type, climate, moisture content and plant cover on the turnover process (Coleman & Jenkinson, 1996; Coleman & Jenkinson, 2014; FAO, 2019). Advantages of the RothC model include its ability to account for the effects of different management practices on SOCS and this model has relatively simple input requirements. Importantly, the model may not be suitable for all soil types and may require site-specific calibration to improve its accuracy (Falloon & Smith, 2002). The model also introduces room for error, as the SOC inputs from plant residues and animal excreta need to be estimated and based on scientific literature, which might also not be easily available or might be inaccurate (Poeplau, 2017; FAO, 2019). Parameters, such as plant SOC input and fertiliser SOC input are often problematic due to high uncertainty and high model sensitivity (Poeplau, 2017; FAO, 2019). Fertiliser input is also not a requirement for the Fortran, python and Windows versions of RothC, which is another limitation of the model. Inaccurate estimates would lead to inaccurate quantification of SOCS, which would lead to inaccurate carbon credits. This would ultimately either decrease or increase the carbon credits and ultimately affect all stakeholders involved.

Direct re-measurements involve physically sampling soils and analysing them in a laboratory to quantify changes in SOC content and stocks, however, does require many samples and rely on appropriate study designs and protocols (Smith et al., 2020). The calculation of SOCS includes measuring SOC content and dry bulk density (ρb) at specific depth intervals and incorporating coarse fragments (1 – Sm (v%)) (Equation 2-8). Advantages of the re-measuring approach includes the reliability of the actual direct soil measurements at different scales.

In South Africa, carbon credits typically rely on the Verified Carbon Standard methodology (Shoch & Swails, 2020) or the Gold Standard methodology (Aynekulu et al., 2011; TREES Consulting, 2020) as standard methodologies for quantifying SOCS (Bodenstein, personal communication, 2023). The re-measure approach is not commonly used, as this approach takes time and has a high operation cost (Smith et al., 2020). Therefore, South African companies typically employ the modelling approach, commonly using the RothC model (Bodenstein, personal communication, 2023; Mills, personal communication, 2025). The aim of this study was:

- to compare two SOCS quantification approaches, modelling and re-measure, on field scale in South Africa and
- to evaluate the cost against the value of SOCS sequestered.

6.2 Materials and methods

6.2.1 Study area and site description

For this study, one field from each of two farms was selected, located near the towns of Ottosdal in the North West Province and Vrede in the Free State Province (Figure 6-1). The Ottosdal farm receives a median annual precipitation (MAP) of 565 mm (Schulze, 2007), with basalt, andesite and tuff as the predominant lithology (Council for Geosciences, 2019), and the soils in the area are generally sandy (clay < 10%) and classified as Hutton and Clovelly soil forms (Soil Classification Working Group, 2018). In contrast, the Vrede farm receives 795 mm MAP (Schulze, 2007), with mudstone and sandstone as the dominant lithology (Council for Geosciences, 2019), and the soils are primarily silty (silt > 15%) and classified as Tukulu soil form (Soil Classification Working Group, 2018).

Agricultural management practices, including rotational cropping with maize and soybean as well as implementing no-tillage practices and incorporating residues after harvest, were implemented at both farms. The Vrede farm implemented a double cropping system, whereas the Ottosdal farm implemented a single cropping system. The Vrede farm utilised winter cover crops, while the Ottosdal farm did not. The winter cover crops at the Vrede farm included oats, rye, vetch, Japanese radish, forage rape and barley. Since, the Ottosdal farm only implemented a single cropping system, the SOC input and SOC sequestration potential was likely lower compared to the Vrede farm. According to the farmers, a total of 1 000 sheep (400 ewes and 600 lambs) grazed daily for a total of 300 days on the Vrede farm from June 2022 to July 2024, while a total of 200 cattle (150 cows and 50 juveniles) grazed a total of 56 days (4 weeks after harvest) on the Ottosdal farm from June 2022 to July 2024.

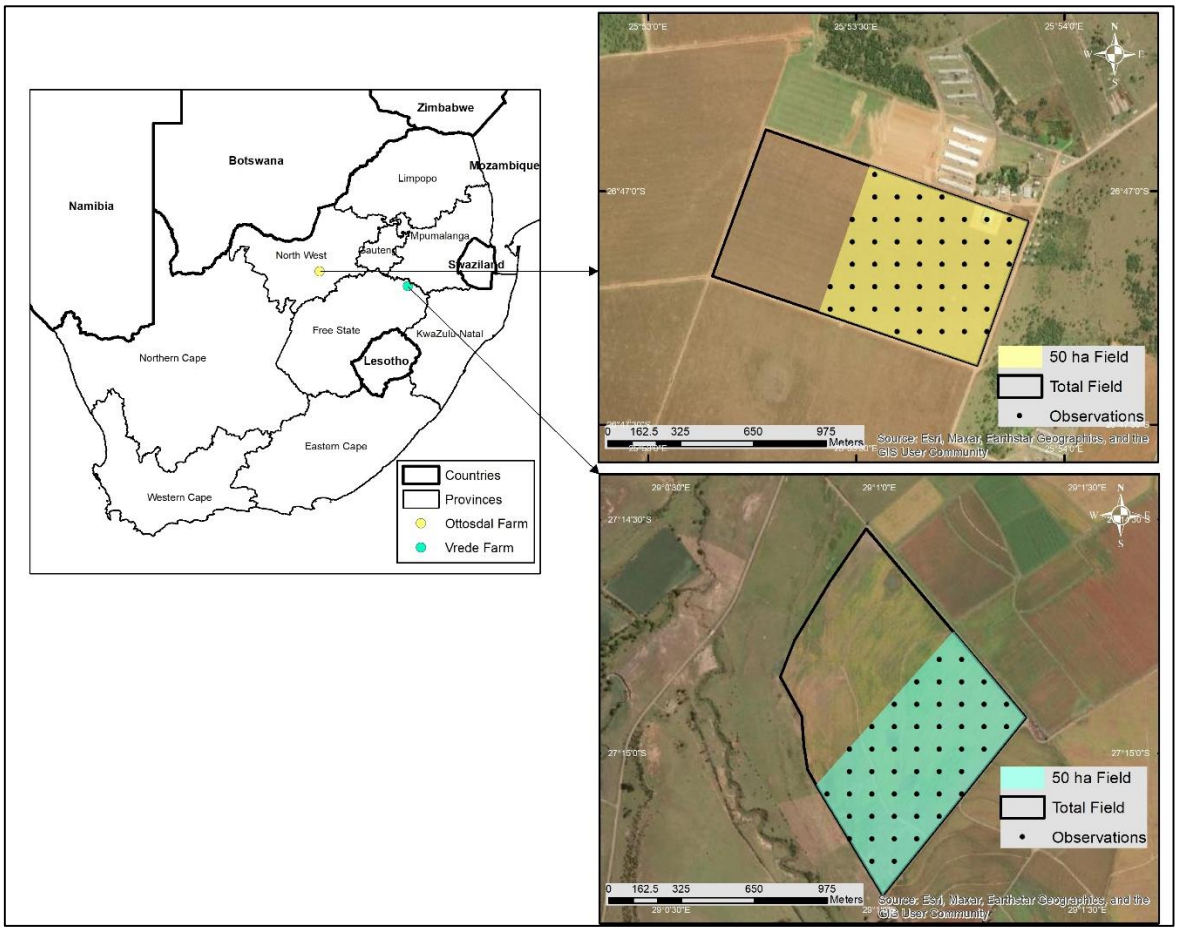
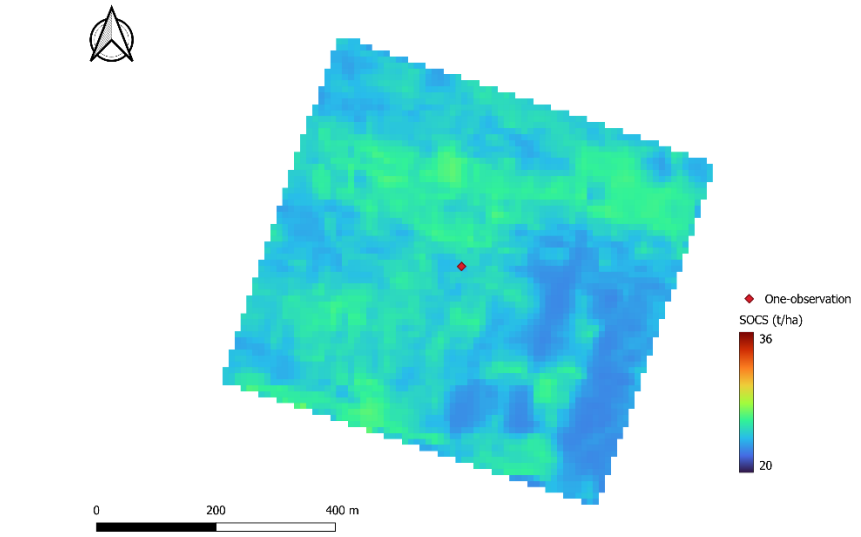


Figure 6-1: The location of the Ottosdal Farm and Vrede Farm, as well as the location of the fields and augered (disturbed) samples collected for the study.

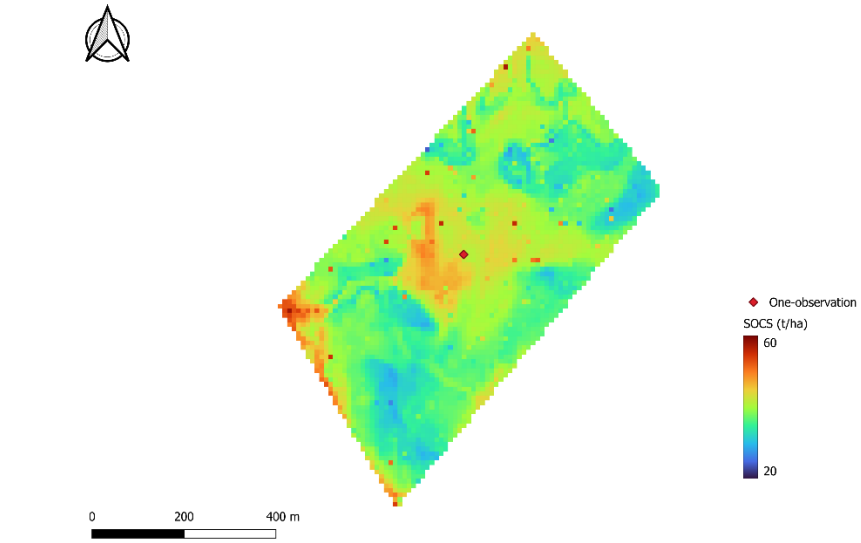
6.2.2 Soil sampling and analysis

6.2.2.1 Baseline assessment

The baseline assessment (for 2022) was done in Chapter 5, by utilising the SOC content and ρb values and calculating the SOCS (t/ha) for the 50 observations using Equation 2-8 and then afterwards summing all observations for finally calculating the total SOCS. Since the texture analysis in Chapter 5 determined the coarse fragments as zero, the equation was slightly amended for calculating the total SOCS. Box-and-whisker plots were created using “ggplot” in RStudio (Wickham, 2016), which presented the baseline SOCS (2022) at the three depths for the two study sites. In Chapter 5, SOCS maps were also created for both study sites (Figure 6-2). For this study, soil mapping would not be used, as Chapter 5 indicated that improved mapping is needed for accurate SOCS calculation and mapping.



(a)



(b)

Figure 6-2: The soil organic carbon stocks (SOCS) map created using digital soil mapping and machine learning for the (a) Ottosdal study site and (b) Vrede study site (from Chapter 5). The "One-observation" that was selected close to the centre of the field is also shown.

6.2.2.2 Re-measure approach

For the re-measurement approach, disturbed soil samples from all 50 observations and three depth increments, (0-5, 5-15 and 15-30 cm; n = 150 / site) from both sites were collected in July 2024 (after harvest) using a soil auger (Figure 6-1). The disturbed soil samples from both sites were air-dried and passed through a 2-mm sieve, before it was scanned with a NeoSpectra hand-held near-infrared (NIR) spectrometer, developed by Si-Ware. The SOC content of soil samples from the Vrede study site was determined by using a LECO CN analyser (Nelson & Sommers,

1996; FAO, 2019), while SOC content from the Ottosdal study site was determined by using a NIR calibration algorithm (RMSE = 0.05, RPD = 1.87) that was created in Chapter 4. The ρb values for all 50 observations at three depths for both study sites were determined by using the pedo-transfer function created in Chapter 5. Finally, Equation 2-8 was used for calculating the SOCS for all 50 observations at each study site.

6.2.3 Modelling approach using RothC

6.2.3.1 RothC model description

For the modelling approach, the RothC model was utilised and the "RothCModel" function in the "SoilR" package (Sierra et al., 2012; FAO, 2019; Jordon, 2021) was used to predict the SOCS for 2024. The RothC model simulates SOC dynamics through time by partitioning incoming plant residues, and other SOC inputs such as manure, into four active compartments/pools: decomposable plant material (DPM), resistant plant material (RPM), microbial biomass (BIO) and humus (HUM) (Coleman & Jenkinson, 2014). The model also includes an inert organic matter (IOM) pool, which does not take part in SOC content turnover. Apart from IOM, each soil organic matter pool decomposes and has a specific decomposition rate. The decomposition rate for each pool is affected by soil moisture, temperature, plant cover and clay content (Kaonga & Coleman, 2008; Coleman & Jenkinson, 2014; FAO, 2019). The RothC model computes the total organic carbon (t/ha) in a monthly time-step (Geremew et al., 2024) and requires the input data described in the following section to achieve this.

6.2.3.2 Input data requirements

The minimum data required to run the RothC model relates to climate (Table 6-1), soil, land management (Table 6-2 – Table 6-5), monthly vegetation cover and an estimated DPM/RPM ratio based on the land use (Coleman & Jenkinson, 2014). Furthermore, the model utilises a few default inputs, including the DPM/RPM ratio of 1.44, (Coleman & Jenkinson, 2005; Setia et al., 2011; Coleman and Jenkinson, 2014; Geremew et al. 2024) and the decomposition rate constants (Ks) for DPM (10), RPM (0.3), BIO (0.66) and HUM (0.02) (Coleman & Jenkinson, 2005; Coleman & Jenkinson, 2014; Sierra & Mueller, 2023; Scherstjanoi & Dechow, 2024).

Regarding the required input soil data, the clay content was acquired from the texture analysis in Chapter 5, while the SOCS for 2022 was calculated in section 6.2.2. The temperature-based Thornthwaite empirical equation was applied to estimate the potential evapotranspiration (PET) and used for calculating the environmental modifying rate (xi) for each study site (Equation 6-1 and Equation 6-2). Mean monthly temperature and total monthly precipitation data from 1990 to 2022 was collected from NASA Prediction of Worldwide Energy Resources (POWER) (NASA,

2024), while data from July 2022 until July 2024 (25 months) was collected from the South African Weather Service (SAWS, 2024). The South African Weather Service would also have been used for collecting data from 1990 to 2022, however, was unavailable due to maintenance.

For calculating the carbon input through farmyard manure (FYM), the volatile solid (VS) excretion rates and total carbon content in VS was acquired from scientific literature (Table 6-2). The final FYM value was calculated by using Equation 6-3 – Equation 6-7 (Table 6-2). The amount carbon lost through grazing (In_{out}) was calculated using Equation 6-8 – Equation 6-11 and utilising the daily dry matter intake (DMI) from literature (Table 6-3). Since it was uncertain the exact weight of the cows at the Ottosdal farm, a conservative DMI value of 12 kg was selected based on typical values reported by Mokolobate et al. (2015) for South Africa.

For calculating the carbon input through plant residues (In), the input from cash crops (Table 6-4) and from cover crops (Table 6-5) were calculated separately. For calculating the carbon input from cash crops, the harvest index (HI), shoot-to-root ratios (S/R) and yield data was utilised in Equation 6-12 – Equation 6-16 (Table 6-4). For calculating the carbon input from cover crops, the above-ground biomass (AGB), below-ground biomass (BGB), HI and S/R was utilised in Equation 6-17 – Equation 6-20 (Table 6-5) (see also Appendix F and Appendix G).

Table 6-1: Baseline soil organic carbon stocks (C0) and environmental modifying rate (xi) and accompanying equations as RothC model input variables for soil organic carbon stocks prediction.

RothC model input required	Input variable	Symbol	Equation	Equation description and constants	References
Environmental modifying rate (xi)	Total monthly precipitation and mean monthly temperature	Potential evapotranspiration (PET) (mm/month)	Equation 6-1: $PET = 16 ((10 \times T) / I)^{a}$	PET refers to the potential evapotranspiration T is the mean monthly temperature (°C) Index (I) is the heat index, calculated as the sum of $(T/5)^{1.514}$ a is a climate factor and is calculated as: $a = 0.000000675 \times I^3 - 0.0000771 \times I^2 + 0.01792 \times I + 0.49239$	Thornthwaite, 1948; Coleman et al. (1997); Coleman & Jenkinson (2005); Geremew et al. (2024)
	Soil cover modifying factor	RC_{factor} (unitless)		If the soil cover is bare, it returns 1 If the soil cover is vegetated, it returns 0.6 For other values, it returns NA	Coleman et al. (1997); Coleman & Jenkinson (2005); Coleman & Jenkinson, 2014; Morais et al. (2019); Scherstjanoi & Dechow (2024); Coleman et al. (2024)
	xi value (between 0 and 2)	xi	Equation 6-2: $xi = RC_{factor} \times fT \times fW$	RC_{factor} is the calculated soil cover fT refers to the temperature modifier and calculated using the fT.RothC function fW refers to the moisture modifier and calculated using the fW.RothC function	

Table 6-2: Farmyard manure input (FYM) and accompanying equations as RothC input variables for soil organic carbon stocks prediction.

RothC model input required	Input variable	Symbol	Equation	Equation description and constants	References
Farmyard manure input (FYM)	Annual volatile solids (VS) excretion	VS_{animal}	Equation 6-3: $VS_{animal} = VS_{rate} \times (TOTAL_{animals}/1000) \times 365$	VS_{animal} (tDM VS/annum) refers to the volatile solid excretion VS_{rate} (tDM VS/day/animal) refers to the dry volatile solid excretion rate; where for sheep it was 0.4 kg VS/1000 kg/day, where for cattle it was 5.8 kg VS/1000 kg/day (2.65 kg VS/453 kg/day)	
	VS excretion for grazing period	$VS_{animal/period}$	Equation 6-4: $VS_{animal/period} = VS_{animal} \times$ (days grazed/365)	$VS_{animal/period}$ (tDM VS) refers to the volatile solid excretion for the specific period grazed	
	Carbon content in excreted VS	$C_{/animal}$	Equation 6-5: $C_{/animal} = VS_{animal/period} \times VS_{carbon\ content}$	$C_{/animal}$ (tDM C/animal) refers to the carbon content excreted for each animal $VS_{animal/period}$ refers to the volatile solid excretion for the specific period grazed VS_{carbon} refers to the carbon content of the volatile solid/manure, which was 0.35 or 35% of dry matter	IPCC, 2006; Kimura et al. (2011); Roy & Kashem (2014); Gavrilova et al. (2019)
	Total carbon input from all animals	C_{TOTAL}	Equation 6-6: $C_{TOTAL} = animals_{TOTAL} \times C_{/animal}$	C_{TOTAL} refers to the total carbon input from all animals $animals_{TOTAL}$ refers to the total number of animals grazed $C_{/animal}$ refers to the carbon content in the excreted volatile solid/manure	
	Total carbon input per hectare (FYM)	FYM	Equation 6-7: $FYM = C_{TOTAL}$/field area	FYM refers to the total carbon input from all animals at a unit field area C_{TOTAL} refers to the total carbon input from all animals Field area was 50 ha	

Table 6-3: Farmyard manure output (In_{out}) and accompanying equations as RothC input variables for soil organic carbon stocks prediction.

RothC model input required	Input variable	Symbol	Equation	Equation description and constants	References
Farmyard manure output (In_{out})	Daily dry matter intake (DMI)	DMI	Equation 6-8: $DMI = animals_{TOTAL} \times animal_{weight} \times 0.03/0.02$	DMI refers to the dry matter intake for animals $animals_{TOTAL}$ refers to the total number of animals grazed $animal_{weight}$ refers to the weight of the animals, where the following weights were used: Adult sheep = 60 kg Juvenile sheep/lambs = 25 kg Adult cow = 450 kg Juvenile cow = 200 kg DMI for cows = 12 kg of dry matter per day, while DMI for sheep = 1.8 kg of dry matter per day (3% of body weight)	Nicol & Brookes (2007); Kimura et al. (2011); Roy & Kashem (2014); Mokolobate et al. (2015); Gavrilova et al. (2019)
	Total dry matter intake (DMI) over grazing period	$DMI_{/period}$	Equation 6-9: $DMI_{/period} = DMI \times days\ grazed$	$DMI_{/period}$ refers to the total dry matter intake over the grazing period DMI refers to the dry matter intake for animals	
	Carbon content loss from grazing	C_{lost}	Equation 6-10: $C_{lost} = DMI_{/period} \times (C_{content}/1000)$	C_{lost} refers to the carbon content loss from grazing $DMI_{/period}$ refers to the total dry matter intake over the grazing period $C_{content}$ refers to the carbon content of crops, which was 0.45	
	Carbon content lost from grazing by total field area	In_{out}	Equation 6-11: $In_{out} = C_{lost}$ / field area	In_{out} refers to the total carbon content lost from grazing C_{lost} refers to the carbon content loss from grazing Field area was 50 ha	

Table 6-4: Carbon plant inputs (In) and accompanying equations as RothC input variables for soil organic carbon stocks prediction.

RothC model input required	Input variable	Symbol	Equation	Equation description and constants	References
Carbon plant inputs (In)	Carbon input from cash crops (roots)	Cr	Equation 6-12: $Cr = (yield / HI \times SR)) \times C_{factor}$	Cr refers to the carbon input from the roots of the crop HI refers to the harvest index of the crop SR refers to the shoot-to-root ratio of the crop C_{factor} refers to the carbon content fraction of the crop, which was 0.45	Kuzyakov & Domanski (2000); Bolinder et al. (2007); Unkovich et al. (2010); Luo et al. (2015); Patel et al. (2015); Hergoualc'h et al. (2019); Dechow et al. (2019); Löwik (2023)
	Carbon input from cash crops (stubble and chaff)	Cs	Equation 6-13: $Cs = S_{factor} \times (yield / HI) \times C_{factor}$	Cs refers to the carbon input from the stubble and chaff of the crop S_{factor} refers to the below-ground biomass-to-above-ground biomass of the crop C_{factor} refers to the carbon content fraction of the crop, which was 0.45	
	Carbon input from cash crops (root exudates)	Ce	Equation 6-14: $Ce = E_{factor} \times (yield / HI) \times C_{factor}$	Ce refers to the carbon input from the root exudates of the crop E_{factor} refers to the carbon exudate factor C_{factor} refers to the carbon content fraction of the crop, which was 0.45	
	Carbon input from cash crops (residues from weeds)	Cw	Equation 6-15: $Cw = W_{factor} \times (yield / HI) \times C_{factor}$	Cw refers to the carbon input from weed residues W_{factor} refers to the carbon weed factor C_{factor} refers to the carbon content fraction of the crop, which was 0.45	
	Total carbon input from cash crops	$In_{cashcrops}$	Equation 6-16: $In_{cashcrops} = Cr + Cs + Ce + Cw$	$In_{cashcrops}$ refers to the total carbon input from cash crops Cr refers to the carbon input from roots of the crop Cs refers to the carbon input from stubble and chaff of the crop Ce refers to the carbon input from the root exudates of the crop Cw refers to the carbon input from the weed residues	

Table 6-5: Carbon plant inputs (In) and accompanying equations as RothC input variables for soil organic carbon stocks prediction (continued).

RothC model input required	Input variable	Symbol	Equation	Equation description and constants	References
Carbon plant inputs (In)	Biomass estimation for each cover crop	$Biomass_{covercrops}$	Equation 6-17: $Biomass_{covercrops} = AGB_{fraction} \times BGB_{fraction}$	$Biomass_{covercrops}$ refers to the total biomass of all cover crops $AGB_{fraction}$ refers to the above-ground biomass fraction for the cover crop $BGB_{fraction}$ refers to the below-ground biomass fraction for the cover crop	Kuzyakov & Domanski (2000); Cowie & Swift (2004); Bolinder et al. (2007); Unkovich et al. (2010); Subedi et al. (2018); Hergoualc'h et al. (2019); Dechow et al. (2019)
	Carbon input for cover crops	$In_{covercrops}$	Equation 6-18: $In_{covercrops} = (Biomass_{covercrops} / HI \times SR) \times 0.45$	$In_{covercrops}$ refers to the total carbon input from cover crops $Biomass_{covercrops}$ refers to the total biomass of cover crops HI refers to harvest index of cover crop SR refers to the shoot-to-root ratio of cover crop The 0.45 refers to the carbon fraction in cover crops	
	Total carbon plant inputs	In_{TOTAL}	Equation 6-19: $In_{TOTAL} = In_{cashcrops} + In_{covercrops}$	In_{TOTAL} refers to the total carbon plant input $In_{cashcrops}$ refers to the total carbon input from cash crops $In_{covercrops}$ refers to the total carbon input from cover crops	
	Final carbon plant inputs	In	Equation 6-20: $In = In_{TOTAL} - In_{out}$	In refers to the final carbon plant input used in the model In_{TOTAL} refers to the total carbon plant input In_{out} refers to the total carbon content lost from grazing	

6.2.3.3 Model spin-up and calibration

Before the spin-up procedure, a single observation (referred to as “One-observation”), close to the centre of the field was also selected to represent the entire field (similar to how SOCS would be calculated by industry) and the average for all 50 observations were also calculated (referred to as “composite”), which was done to represent a composite sample from the entire 50 ha field (Figure 6-3).

The spin-up procedure entails an initialisation process in which the model runs for a predetermined period (1 000 years in this case) to achieve the SOC equilibrium state across the five pools. The process includes agricultural land management and land use data from 2017-2022 at the two study sites and long-term monthly average climate data (1990 – 2022) (Coleman et al., 2024; Geremew et al., 2024).

The agricultural land management and monthly average climate data was used in the model spin-up and the carbon quantity in each pool is initially set to zero. However, the IOM pool needs to be known, as this pool do not change over time in RothC and is dependent on the model’s dynamics. The initial calculated SOCS (from section 6.2.2) was used to calculate the IOM pool for each observation. Equation 6-21 presents the empirical function proposed by Falloon et al. (1998) that was used to determine the size of this pool based on the SOCS for all 50 observations measured in 2022.

Equation 6-21: $$\mathbf{IOM} = \mathbf{0.049} \times \mathbf{SOCS}\,(\mathbf{2022})^{\mathbf{1.139}}$$

The DPM/RPM ratio of 1.44 were used for both study sites as input for agricultural land management (Coleman & Jenkinson, 1996; Coleman et al., 2024).

After the spin-up, the model needed to be calibrated as most of the measured SOCS values from 2022 did not match the modelled spin-up values for 2022 (Figure 6-3). The calibration procedure was written in the R environment (R Core Group, 2020) and the kinetic constants (k for DPM, RPM, BIO and HUM) and FYM (Table 6-2) simultaneously calibrated until the modelled SOCS fell within the 95% quantile range of the measured SOCS. The calibration procedure involved slightly increasing or decreasing the values of the mentioned kinetic constants and total FYM. The kinetic rate constants were altered in a range between 80 and 120% of the RothC default values (Cagnarini et al., 2019), while the total FYM were changed between 0 and 1.5 tons.

The values were calibrated using two different strategies: the first involved individually calibrating each of the 50 observations (referred to as “individual”), while the second involved calibrating the average SOCS (referred to as “average”) and applying it to all 50 observations.

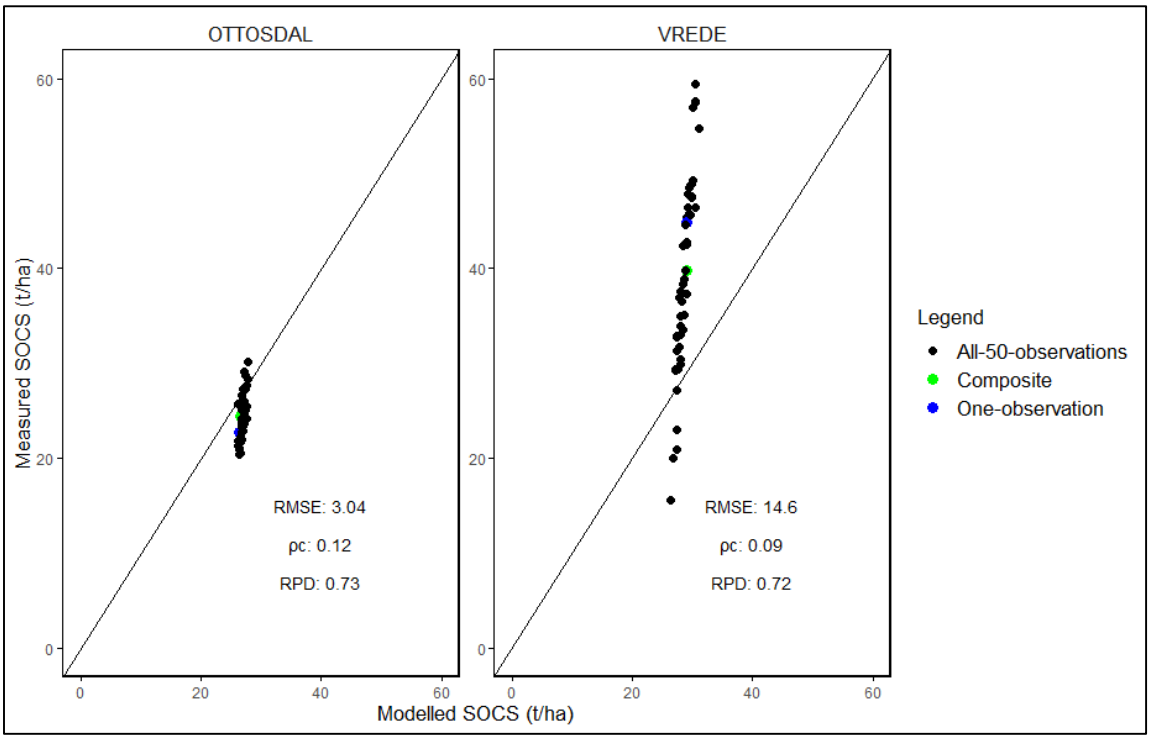


Figure 6-3: The measured vs modelled soil organic carbon stocks (SOCS) for the Vrede and Ottosdal study sites, after the spin-up procedure. The black line represents the 1:1 relationship, while the observations marked in green, blue and black denote the single observation (industry standard), average (composite sample) and all other 50 observations, respectively. RMSE = root mean square error; ρc = Lin's concordance correlation coefficient; RPD = ratio of performance to deviation.

6.2.3.4 Model prediction for 2024

Following the spin-up and calibration, the equilibrium carbon pool values were extracted using the integrated function (*getC*) part of the "SoilR" package (Sierra et al., 2012). These extracted values were then used as the initial pool values for predicting the SOCS changes from 2022 to 2024. The hydro-climatic data, land management and monthly vegetation cover from 2022 to 2024 was utilised in the model to predict the SOCS.

6.2.4 Data analysis

6.2.4.1 Re-measure approach vs modelling approach using RothC

A scatter plot was created using the "ggplot" function in RStudio (R Core Group, 2020) and the re-measure approach values and modelling approach values compared with each other for each study site. Validation statistics, including root mean square error (RMSE), Lin's concordance correlation coefficient (ρc) (Lin, 1989), and ratio of performance to deviation (RPD) were used for evaluating the relationship between the two approaches.

The total SOCS (for 2024) were calculated for each approach by summing all 50 observation values together, since the unit of the values were in t/ha. These values were referred to as “All-50-observations” for each approach. Importantly, the “All-50-observations” value for the re-measured approach represented the actual total SOCS for the two study sites, since 50 observations were utilised (1 observation / ha) for determining the total SOCS for the two fields.

Secondly, the single observation taken close to the middle of the field, (“One-observation”), and the average for all observations (“compsite”) was multiplied by 50 (since the field was 50 ha), which resulted in the total SOCS for each field (in tons). Since, two strategies were used for calibrating the values, the best performing calibration values were determined (section 6.3.2), and a bar graph was created using the “ggplot” function in RStudio (R Core Group, 2020) for visualising the differences between approaches. Overall, the “individual” calibration strategy showed the best validation statistics, however, the “average” calibration strategy was also applied to predict the SOCS (for each observation) for 2024 (Figure 6-8). The “individual” calibration strategy was utilised for predicting the total SOCS for 2024 (Figure 6-9).

6.2.4.2 Soil organic carbon stocks and financial implications

For calculating the number of SOCS sequestered after the 2-year period, the baseline SOCS (2022) was subtracted from the calculated SOCS of 2024. R programming was again utilised for creating a bar graph using the “ggplot” function in RStudio (Wickham, 2016) indicating the number of SOCS sequestered. Since the price of carbon credits is trading between $15 - $25 (R275 – R460) per tCO2e, the amount of 1 ton carbon dioxide equivalence (CO_2e) was calculated by multiplying the SOCS with 3.76 (molecular fraction) and thereafter the carbon credit value calculated by multiplying with R300 as a standard carbon credit value for 1 tCO_2e.

For evaluating the cost of sampling with the value of SOCS sequestered using the re-measure approach, the cost of SOC content and ρb analysis was calculated. Operating costs, including sample collection and fuel, were estimated to be R 5 000 / site. The re-measure approach utilised 50 observations at three depths, which resulted in a total of 150 disturbed samples at each study site that were analysed for SOC content (section 6.2.2). Although the SOC content from the Ottosdal study site was predicted using a site specific created NIR calibration algorithm, there is currently no accurate, readily available NIR calibration algorithm that can be used for the entire South Africa (Chapter 4). Therefore, local land managers and farmers need to determine SOC content either through the TDC method (Chapter 3) or through the Walkley-Black wet-oxidation method (Chapter 3). Local soil laboratories price between R 80 and R 120 for a sample determined with either the TDC or WB methods. Soil samples were priced as R 100 / sample and the total cost calculated. Although ρb is usually not re-measured, since it does not change

significantly over time, local laboratories price ρb measurements at R 70 / sample. Finally, the total costs were calculated and compared to the difference in potential carbon credit income as calculated with the modelling vs re-measure approaches.

6.3 Results and discussion

6.3.1 Descriptive statistics

Overall, the Ottosdal study site showed an increase in SOC content and ρb from 2022 to 2024 for all three depths (Figure 6-4), and also an increase in SOCS for all three depths (Figure 6-5). The Vrede study site showed an increase in SOC content, ρb (Figure 6-4) and SOCS (Figure 6-5) from 2022 to 2024 for only the 5-15 cm depth. The 5-15 cm depth for the Vrede study site, especially showed a tremendous increase in SOCS from 2022 to 2024. Both study sites also had an increase in SOCS after the 2-year period, which was more pronounced at the Ottosdal study site compared to the Vrede study site (Figure 6-6).

The increase in SOCS after the 2-year period could likely be attributed to the regenerative agricultural practices incorporated, as mentioned in section 2.2. These practices, including crop rotations and the incorporation of organic matter, were employed at both sites, leading to enhanced soil organic carbon stocks by promoting microbial activity and increasing carbon inputs.

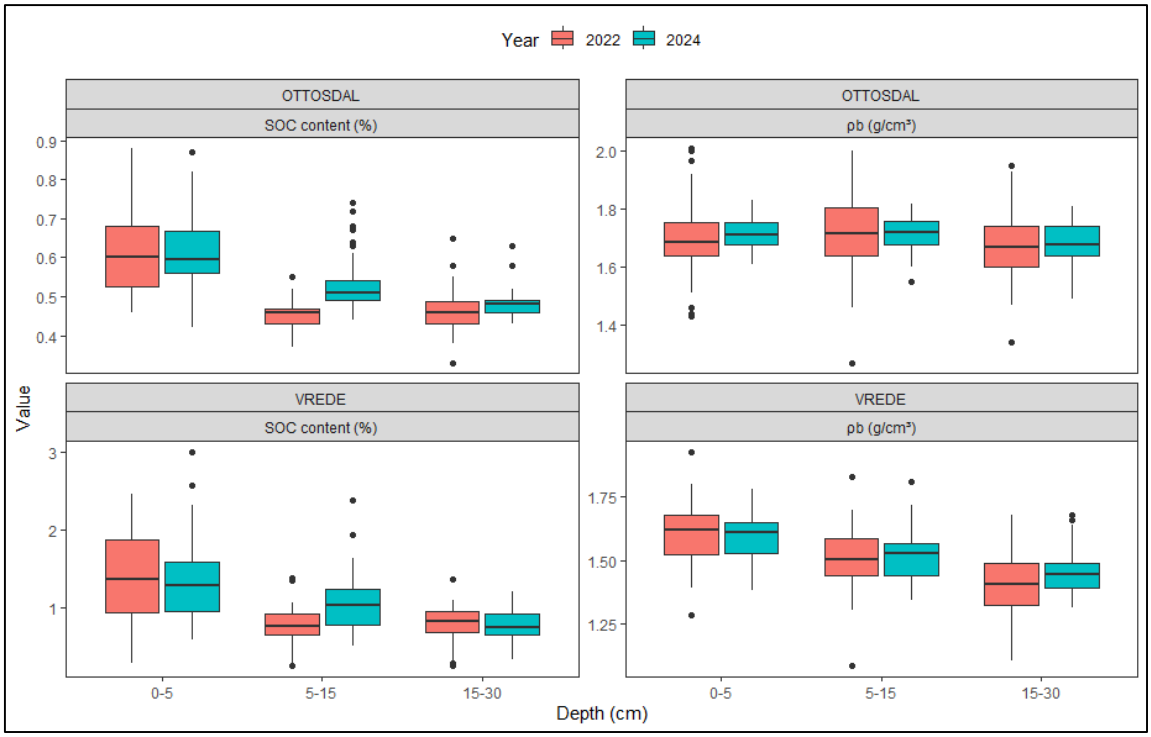


Figure 6-4: Box- and whisker plots presenting the distribution of soil organic carbon (SOC) content and dry bulk density (ρb) in 2022 and 2024, for the Ottosdal and Vrede study sites.

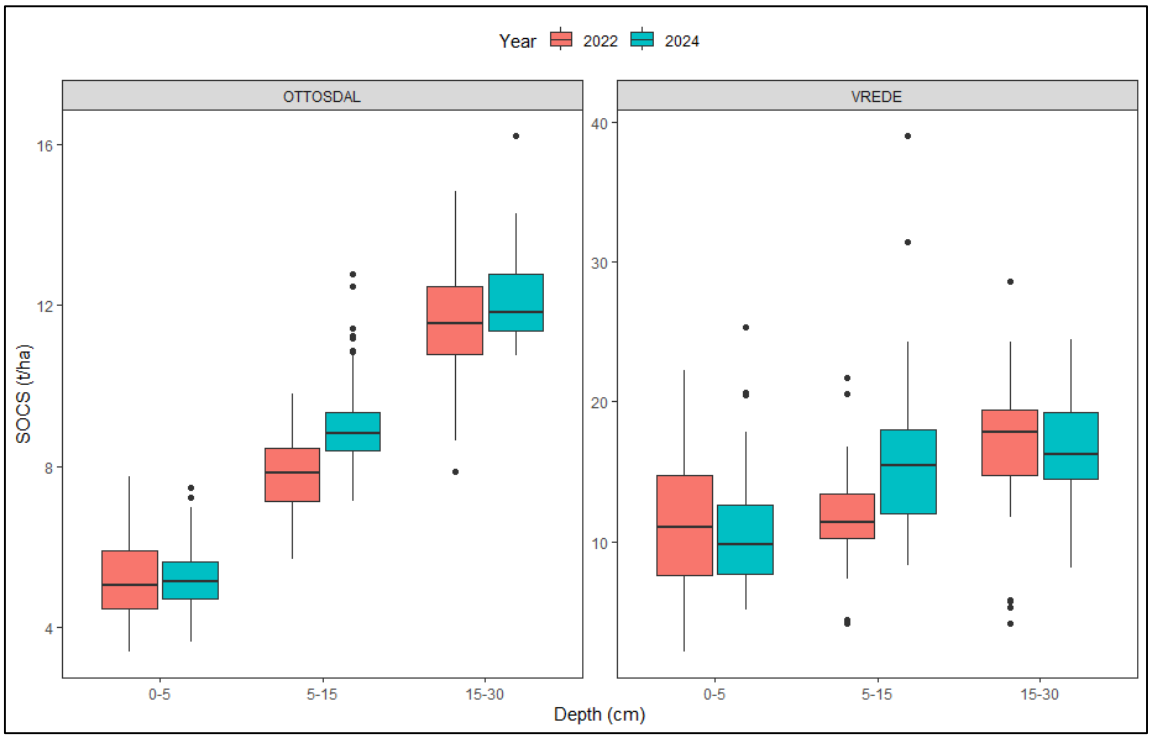


Figure 6-5: Box- and whisker plots presenting the distribution of soil organic carbon stocks (SOCS) at the three depths in 2022 and 2024, for the Ottosdal and Vrede study sites.

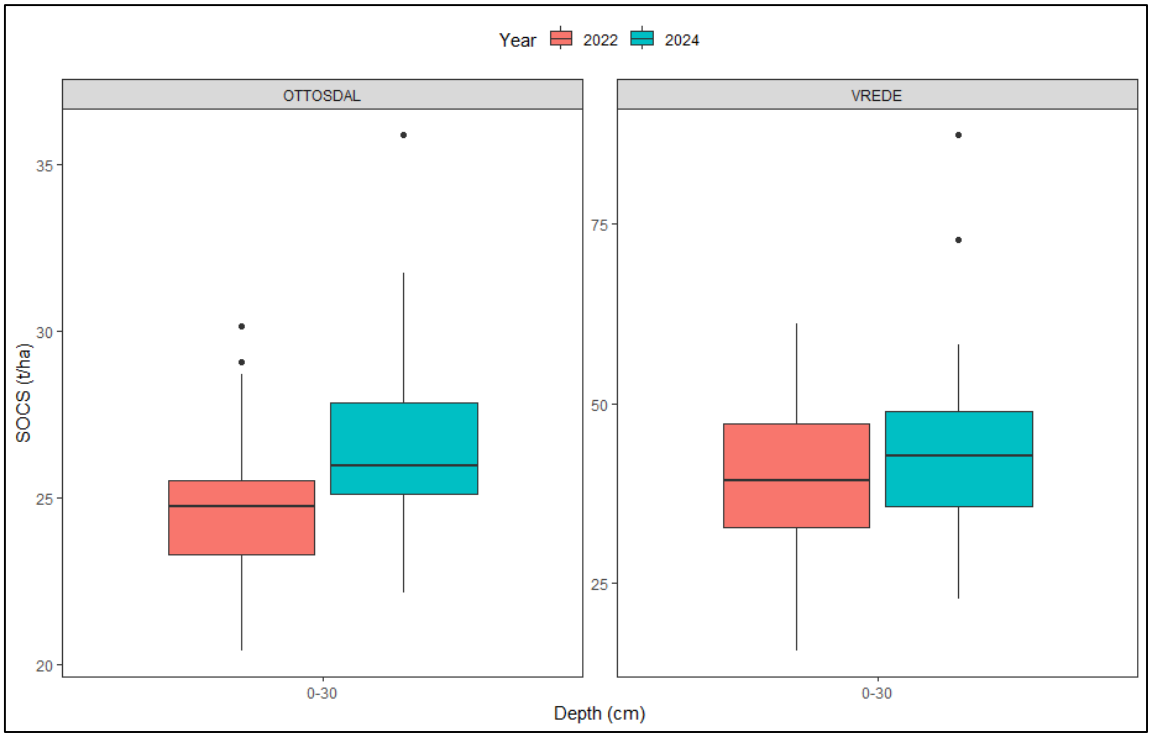
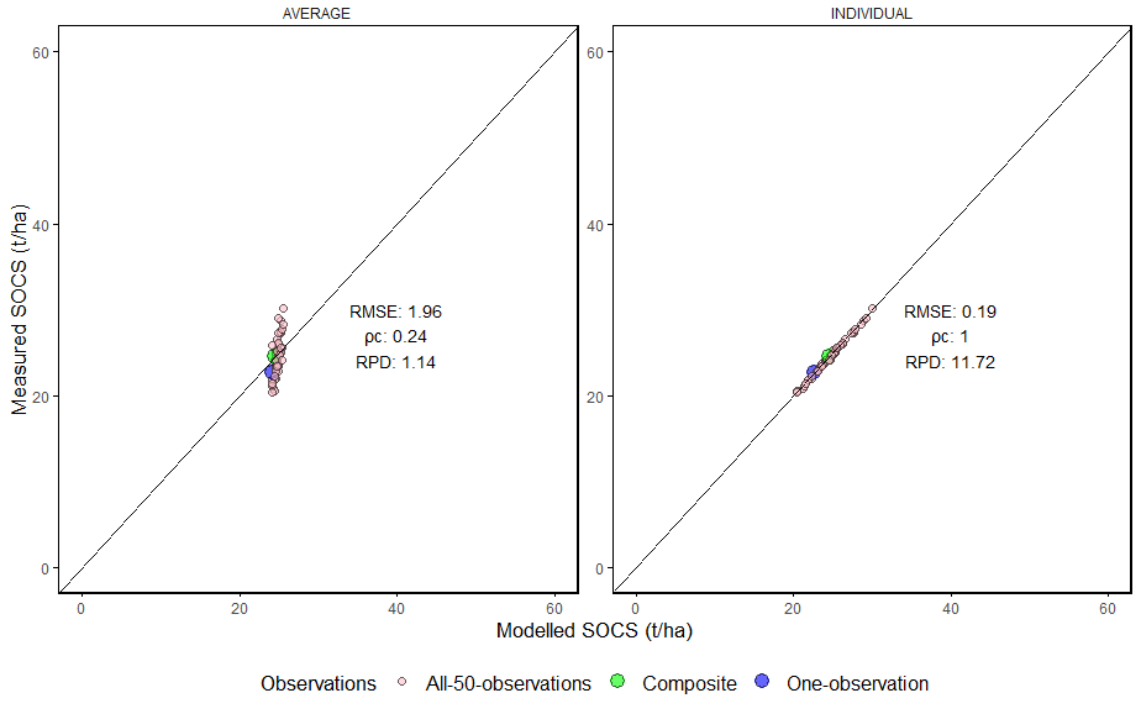


Figure 6-6: Box- and whisker plots the distribution of total soil organic carbon stocks (SOCS) in 2022 and 2024, for the Ottosdal and Vrede study sites.

6.3.2 Model calibration for 2022

For both study sites, the “average” calibration oversimplified the data and resulted in a poor fit (high RMSE and low ρ_c and RPD values), while the “individual” calibration accounted for the variability in the data, which resulted in a very good fit (low RMSE, high ρ_c and RPD) (Figure 6-7). Even though the “individual” calibration outperformed the “average” calibration, the smaller SOCS variability at the Ottosdal study site led to a slightly improved performance of the “average” calibration strategy, as evidenced by the lower RMSE (1.96 t/ha compared to 9.44 t/ha) and a more condensed data distribution compared to the Vrede study site (Figure 6-7). The “average” calibration technique led that the “One-observation” was slightly underpredicted for the Ottosdal study site, while for the Vrede study site it was slightly overpredicted.

(a)



(b)

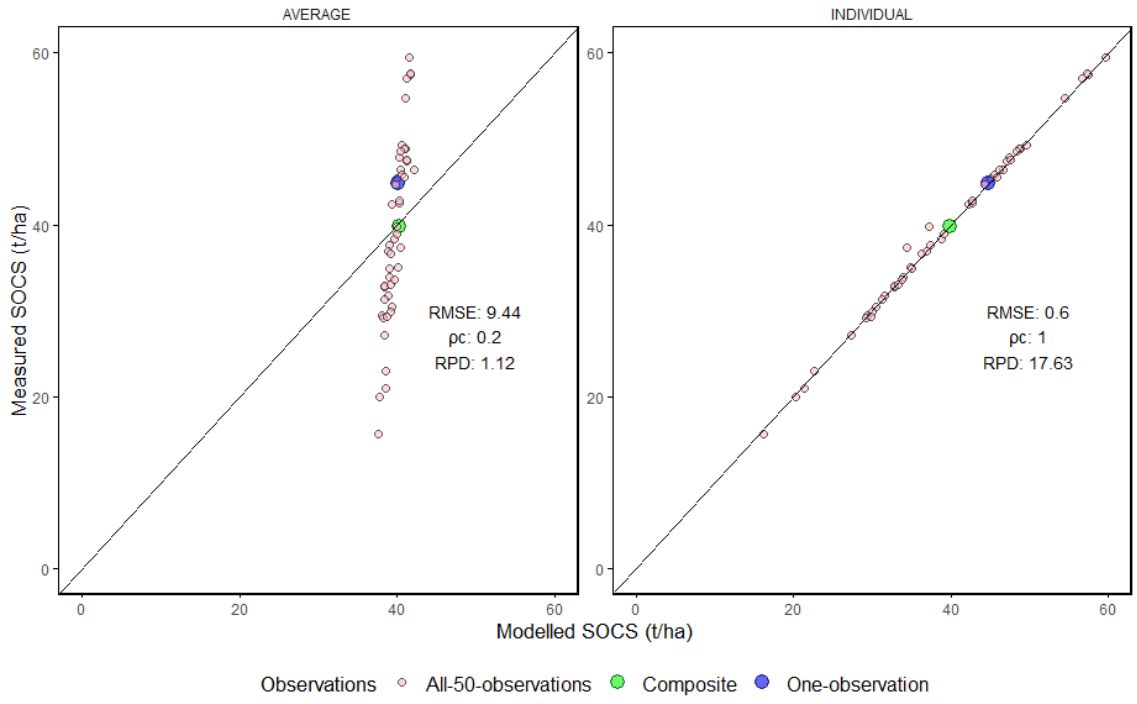


Figure 6-7: The measured vs modelled soil organic carbon stocks (SOCS) of 2022 for (a) the Vrede and (b) Ottosdal study sites, using two calibration approaches: (left) AVERAGE - calibration based on the average SOCS across all 50 observations, and (right) INDIVIDUAL - individual calibration of each observation. The black line represents the 1:1 relationship, while the observations marked in blue, green and red denote the single observation (industry standard), composite sample and all other 50 observations, respectively. RMSE = root mean square error; ρc = Lin's concordance correlation coefficient; RPD = ratio of performance to deviation.

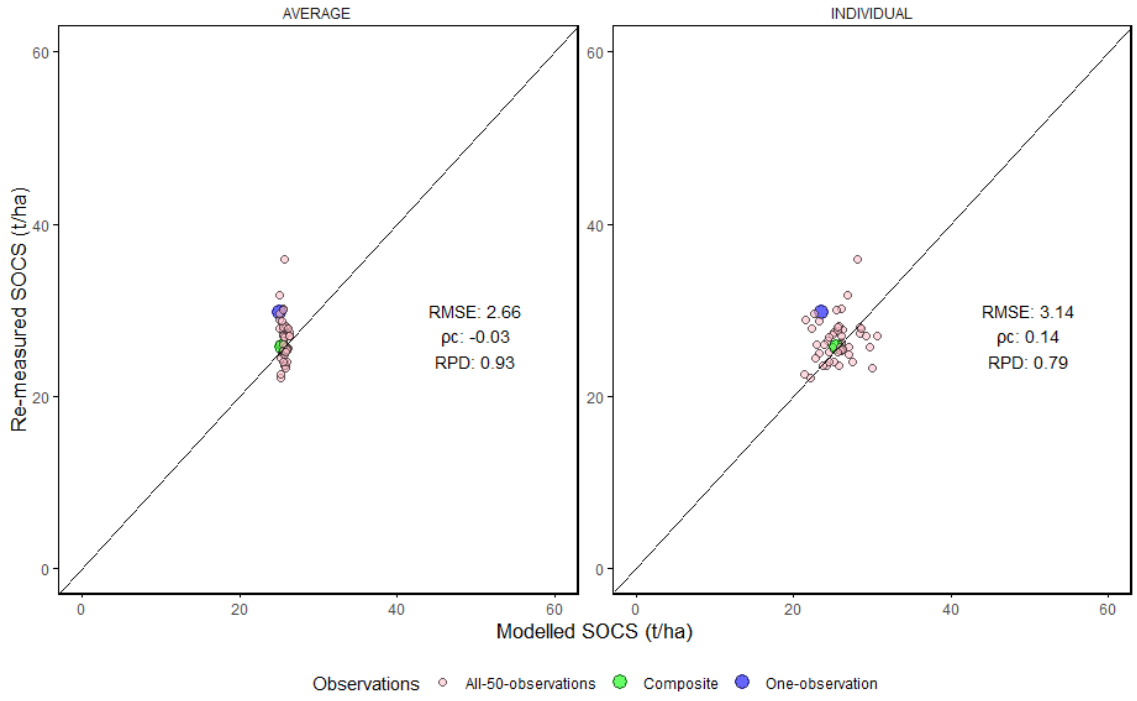
6.3.3 Re-measure approach vs modelling approach using RothC

Overall, both study sites presented poor performance for both calibration strategies, with high RMSE values and low ρc, R^2, RPIQ and RPD values (Figure 6-8). Although both calibration strategies presented poor performance (Figure 6-8), the Ottosdal study site exhibited a tighter clustering of observations, likely caused by the smaller SOCS variability compared to the Vrede study site. The spatial variability of SOC at field scale have also been found by Gonzalez-Molina et al. (2011) and was attributed to soils being heterogeneous and uncertainty regarding the number of domestic ruminants and intensity of grazing. The SOC spatial variability was ultimately one of the main factors affecting the accuracy and reliability of measured and modelled values for the two study sites. This spatial variability also had financial implications, as seen from Table 6-6 and

Table 6-7.

Overall, it was inconclusive if a modelling or re-measure approach should be used for quantifying SOCS on field scale in South Africa. As mentioned, utilising the re-measured approach for all 50 samples (“All-50-observations”), were regarded as the closest to the true value of SOCS and SOCS sequestered from 2022 to 2024. From Figure 6-9 and Figure 6-10 it was clear that the “One-observation” showed no consistent trend using both approaches for both study sites, which is concerning as it is intended to replicate industry practices where a composite sample is typically taken close to the centre of a field to represent the entire area. The “composite” observation showed the closest total SOCS to the re-measured true value (“All-50-observations”) (Figure 6-9 and Figure 6-10), since this sample was calculated as the average from all 50 samples and represented a composite sample from the entire field. The results indicated that the spatial variability of SOCS must be taken into account at the onset of measuring the SOC content in year one (2022 in this case), independent of whether the SOCS would be modelled or re-measured subsequently. The results also indicated that when “All-50-observations” were utilised, the modelling approach using RothC slightly underpredicted the SOCS values, which is in contrast to what previous studies have found (Senapati et al., 2014; Brilli et al., 2017; Davoudabadi et al., 2024). This slight underprediction was likely due to the conservative nature of the input values used in the RothC model, such as estimates for organic matter content, where the model is designed to provide more cautious predictions of carbon turnover and sequestration (Coleman & Jenkinson, 2014). However, since the modelling approach only slightly underestimated (3 tons) the total SOCS of the Vrede study site (Figure 6-9 and Figure 6-10), the SOC spatial variability was likely better captured compared to the Ottosdal study site.

(a)



(b)

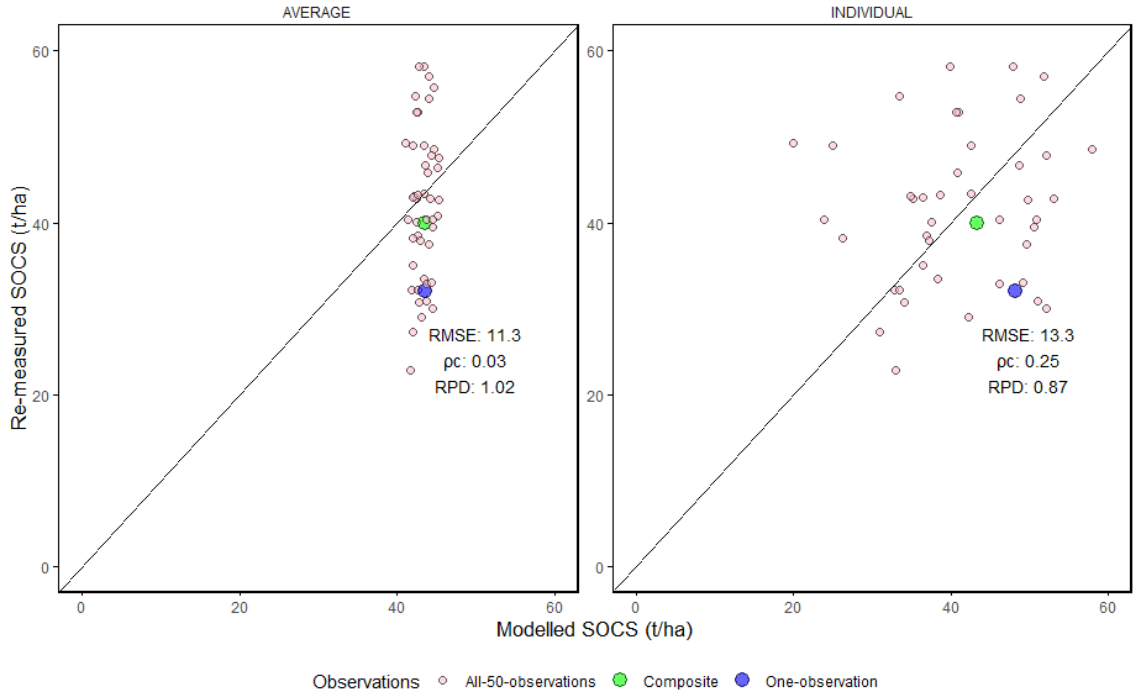


Figure 6-8: The measured vs modelled soil organic carbon stocks (SOCS) of 2024 for (a) the Vrede and (b) Ottosdal study sites, using two calibration approaches: (left) AVERAGE - calibration based on the average SOCS across all 50 observations, and (right) INDIVIDUAL - individual calibration of each observation. The black line represents the 1:1 relationship, while the observations marked in blue, green and red denote the single observation (industry standard), composite sample and all

other 50 observations, respectively. RMSE = root mean square error; ρc = Lin's concordance correlation coefficient; RPD = ratio of performance to deviation.

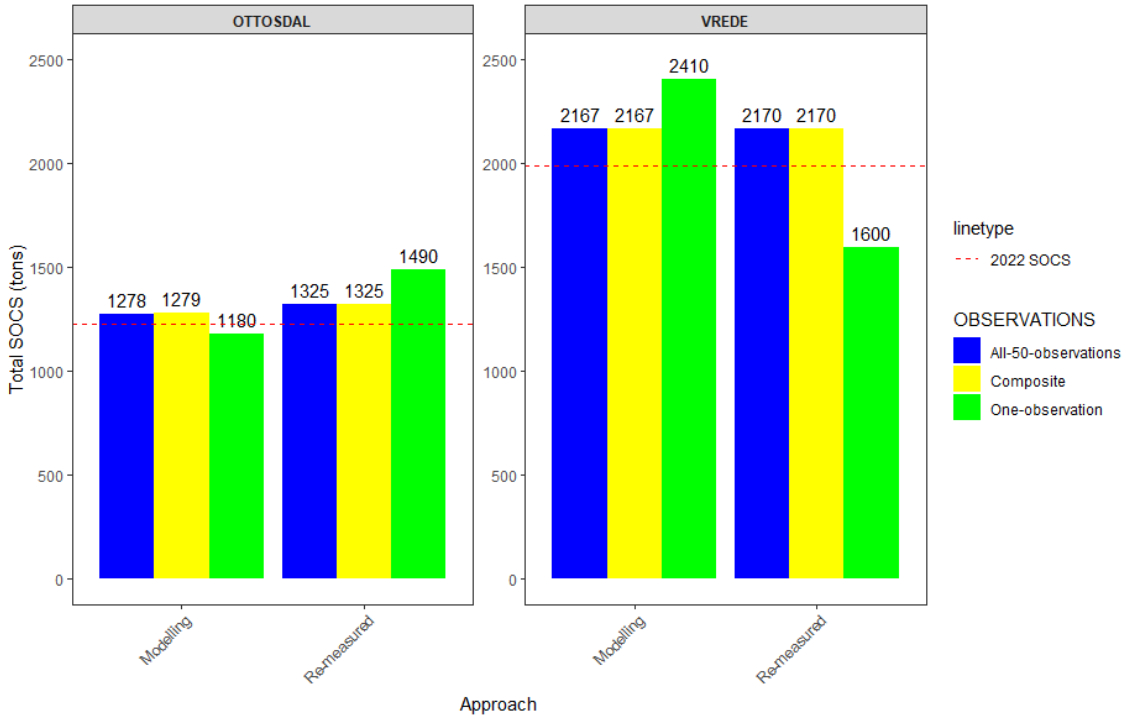


Figure 6-9: Comparison of total soil organic carbon stocks (SOCS) between modelled (blue) and re-measured (green) approaches for the Ottosdal and Vrede study sites, using different observation datasets: "All-50-observations" and "One-observation".

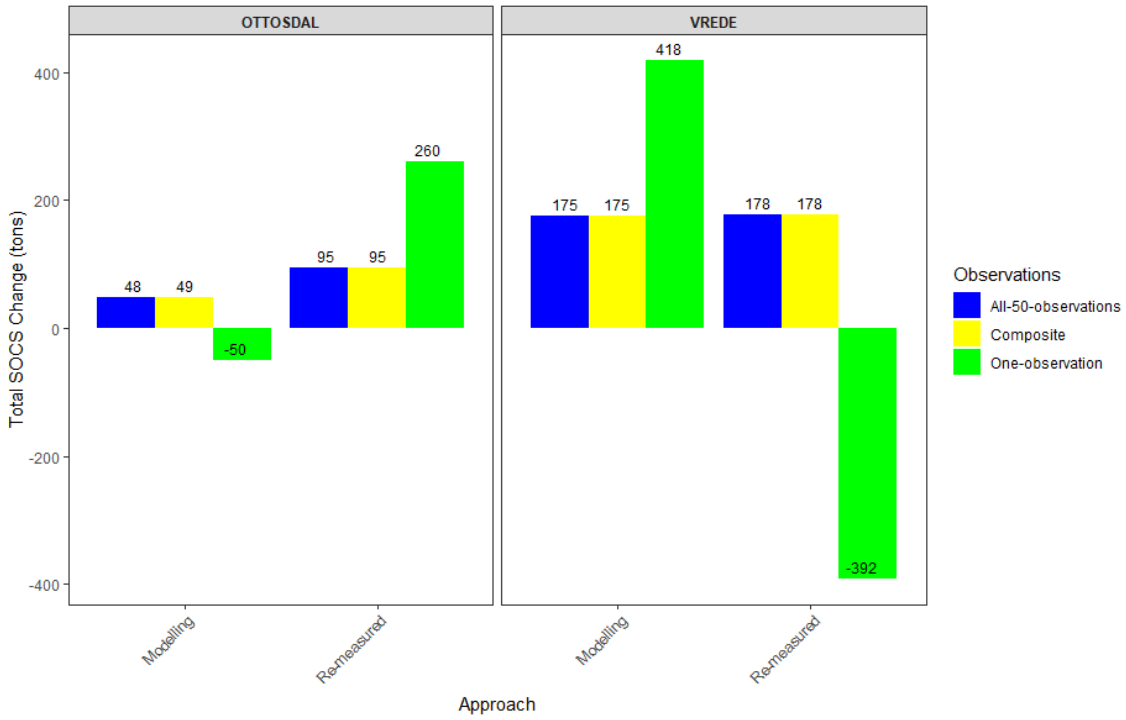


Figure 6-10: The total soil organic carbon stocks change (sequestered) from 2022 to 2024 for the Ottosdal and Vrede study sites, using the modelling and re-measure approaches and two types of observation techniques.

Table 6-6 and

Table 6-7 indicates the SOCS sequestered and total SOCS with their accompanying carbon credit values for the different approaches. The final cost when utilising the re-measure approach resulted in a total of R 30 500 for each study site. The financial comparison between the carbon credit values and the associated final costs reveals distinct differences across the two study sites between the true SOCS value ("All-50-observations") and the two approaches (

Table 6-7).

Table 6-6: The soil organic carbon stocks (SOCS) sequestered from 2022 to 2024, carbon dioxide equivalence (CO_2e) and the accompanying carbon credit value for the Ottosdal and Vrede study sites, when utilising two types of observation techniques for the re-measured and modelling approach.

Ottosdal				
Observation dataset	**Approach**	**SOCS sequestered (tons)**	**CO_2e**	**Carbon credit value**
All-50-observations (TRUE VALUE)	Re-measure	95	348.65	R 104 595.00
All-50-observations	Modelling	48	176.16	R 52 848.00
Composite	Re-measure	95	348.65	R 104 595.00
Composite	Modelling	49	179.83	R 53 949.00
One-observation	Re-measure	260	954.2	R 286 260.00
One-observation	Modelling	-	-	-
Vrede				
All-50-observations (TRUE VALUE)	Re-measure	178	653.26	R 195 978.00
All-50-observations	Modelling	175	642.25	R 192 675.00

Composite	Re-measure	178	653.26	R 195 978.00
Composite	Modelling	175	642.25	R 192 675.00
One-observation	Re-measure	-	-	-
One-observation	Modelling	418	1534.06	R 460 218.00

Table 6-7: The soil organic carbon stocks (SOCS) difference between the true SOCS value for 2024 and the different approaches, carbon dioxide equivalence (CO_2e) and the accompanying carbon credit value for the Ottosdal and Vrede study sites, when utilising two types of observation techniques.

Ottosdal				
Observations	**Approach**	**Difference (tons)**	**CO_2e**	**Carbon credit value**
All-50-observations	Modelling	47	172.49	R51 747.00
Composite	Modelling	46	168.82	R50 646.00
One-observation	Re-measured	145	532.15	R159 645.00
One-observation	Modelling	165	605.55	R181 665.00
Vrede				
All-50-observations	Modelling	3	11.01	R3 303.00
Composite	Modelling	3	11.01	R3 303.00
One-observation	Re-measured	570	2 091.90	R627 570.00
One-observation	Modelling	240	880.80	R264 240.00

The study found that different sequestered SOCS values were calculated when utilising different observation techniques for both Ottosdal and Vrede study sites (Table 6-6). This had an effect on the carbon credit value for the number of sequestered SOCS after the 2-year period, with the true value having almost twice the carbon credit value (~ R 104 000) compared to the modelling approach (~ R 52 000) at the Ottosdal study site. Since the “composite” values were derived from each study site, they had the same carbon credit value as when all 50 observations were utilised. On the other hand, the “One-observation” collected with the re-measured approach had a carbon credit value of approximately R 290 000 at the Ottosdal study site, while the modelling approach presented a carbon loss and zero carbon credit value (Figure 6-10 and Table 6-6). In contrast, the true value of total SOCS sequestered and the modelling approach utilising 50 observations (“All-50-observations”) only had a difference of R 3 000, and the “One-observation” collected with the modelling approach had a carbon credit value of approximately R 460 000, while the re-measure approach presented a carbon loss and zero carbon credit value (Figure 6-10 and Table 6-6). This led to the “One-observation” using the modelling approach presenting approximately R 260 000 more than the true carbon credit value of the field.

There was also a substantial difference in SOCS and potential carbon credits generated between using the modelling and re-measure approaches (

Table 6-7). For the Ottosdal study site, there was a difference of R 51 474 between the true value of SOCS and the SOCS using the modelling approach when utilising “All-50-observations”. This exceeded the final cost (R 30 500) by R 20 974. There was also a substantial difference of approximately R 160 000 and R 180 000 between the “One-observation” and the true SOCS, when using the re-measured and modelling approaches, respectively.

For the Vrede study site, there was only a R 3 000 difference between the true SOCS value and the SOCS value when utilising “All-50-observations”. This led to a loss of R 27 197 when incorporating the operating costs of R 30 500, which would have been taking into account when 50 samples had to be collected. Concerningly, when utilising the “One-observation” using either the re-measured or modelling approach, the difference was over R 250 000 compared to the true SOCS value. While there is some uncertainty regarding the operational costs, the potential value of carbon credits is several orders of magnitude higher than these costs when utilising the “One-observation”. Therefore, it is crucial that this observation be accurate and reliable at the initial sampling (2022 for this study) and that soil variability taken into account as any discrepancies could have significant financial implications.

As mentioned before, it was inconclusive if a modelling or re-measure approach should be used for quantifying SOCS on field scale in South Africa. However, the results indicated that utilising more samples either in the re-measure or modelling approach would provide more reliable carbon credit values. Given that the Ottosdal study site showed a difference of approximately R 21 000, and the Vrede study site experienced a loss after accounting for operating costs, utilising RothC at field scale has significant financial implications. Recent studies have shown that a combination of direct measurements (at plot scale) and modelling afterwards (at larger spatial scales) can greatly improve predicting SOCS and SOCS sequestration (Smith et al., 2020). By applying this approach and utilising sampling designs together with machine learning models the RothC model might be able to capture the SOCS variability and lead to more reliable carbon credit values.

Approaches, such as strategic soil sampling at the initial measuring stage using sampling designs with machine learning models could capture SOC spatial variability and increase carbon credit reliability. Recent studies have shown how different sampling designs together with machine learning models are able to capture SOC content variability at smaller scales (Zizala et al., 2024; Bougiouklis et al., 2025; Sharma et al., 2025). Sampling designs, including conditioned Latin Hypercube Sampling (cLHS) (Minasny & McBratney, 2006), K-means clustering (Hair et al., 2013)

and Feature Space Coverage Sampling (FSCS) (Brus et al., 2007) are able to capture the representativeness of SOC content in a small area and ensure that the sampled data accurately reflects the heterogeneity of the field. In combination with machine learning models, such as Random Forest and Cubist, this approach can be further enhanced by using algorithms to analyse the spatial patterns of SOC content and divide the field into distinct SOC zones. As the RothC model does not include a sub-model for plant growth and yield (Coleman & Jenkinson, 2014; Jordon et al., 2022; Geremew et al., 2024), incorporating spatial covariates, such as normalized difference vegetation index (NDVI) and crop yield data, into the model would further enhance its predictive accuracy by providing additional insights into plant growth, which can help capture the dynamic relationship between SOC and crop productivity. Maas and Lal (2023) have found that vegetation indices, including NDVI, were able to estimate soybean yields exceptionally well (R^2 = 0.91) and were utilised in the RothC model. Although utilising sampling designs together with machine learning models would likely require more composite samples being collected throughout the field to capture the SOC variability, it would ensure that the SOCS are modelled more accurately.

A few limitations of this study need to be addressed. Firstly, the RothC model is typically designed for use at a more generalised scale (Falloon & Smith, 2002), and applying it to a field scale, variability might not be accounted for as seen by the results. The model also assumes certain steady-state conditions and simplified processes for carbon turnover, which may not always reflect the complexity of field scale dynamics, including microbial activity and decomposition rates. Also, the RothC model may be sensitive to short-term changes in SOCS due to its design focusing on long-term cycling. RothC is usually used for predicting SOCS after 10 years or more, which has shown to yield more consistent results (Gonzalez-Molina et al., 2011). Future research could improve modelling SOCS using RothC at field scale, by incorporating field scale dynamics and calibrating the model for short-term cycling. The SOC sequestration industry must take these considerations of the model limitations into account, and find ways, such as an improved sampling design, to address these inaccuracies of RothC at field scale.

6.4 Conclusion

In South Africa, companies commonly employ the modelling approach using the RothC model for predicting SOCS over time rather than utilising a re-measure approach by using direct measurements. The main finding from the study indicated that it was inconclusive if a modelling or re-measure approach should be used for quantifying SOCS on field scale in South Africa. The spatial variation of SOC was the main factor influencing the total SOCS, sequestered SOCS and carbon credit values.

The results indicated that the modelling approach using the RothC model slightly underestimated the total SOCS and sequestered SOCS when utilising all 50 observations for both study sites. The industry-represented, “One-observation”, did not have a consistent trend and showed overprediction and underprediction for both study sites using the different approaches. This had significant financial implications using both approaches. The study also found that the spatial variation of the SOC must be taken into account at the onset of measuring the SOC content in year one, independent of whether the SOCS would be modelled or re-measured. The carbon sequestration industry must take the model limitations (such as field scale dynamics and sensitivity to short-term changes) into account, and find ways, such as an improved sampling design to address these accuracies of RothC at field scale. By also incorporating vegetation indices, such as NDVI, and crop yield data into the RothC model would further enhance RothC’s predictive accuracy and help capture the dynamic relationship between SOC and crop productivity. Finally, a model sensitivity analysis could also help identify the rank of the most important model inputs that impacts the predicted SOCS for that specific model setup. This would also provide the modeller the opportunity to focus their efforts on more important model input values for improving region/site specific values.

CHAPTER 7 CONCLUSION

7.1 Conclusions

The study endeavoured to address gaps in terms of the measuring, quantification, mapping and modelling of SOC that the Verified Carbon Standard methodology and the Gold Standard methodology are not being clear on or specific about. These gaps focused on (i) the measurement of SOC content, (ii) predicting SOC content with NIR spectroscopy, (iii) mapping of SOCS, and finally (iv) the two quantifying approaches of SOCS. The study aimed to address these gaps by refining the methods for determining SOCS in South Africa.

The first objective was assessed by collecting soil two-hundred-and-twenty topsoil (0-30 cm) samples and analysing the samples for SOC content with the WB wet-oxidation method, LOI method and TDC method. Polynomial models were fitted to the data and scatter plots created, while transfer functions were also created for transferring SOC content values from one method to another. The findings of the study suggested that the TDC method should be considered the preferred method for determining SOC content for the assessment of carbon credits in South Africa. The WB method should be avoided if the soil is expected to have a high SOC content, while the LOI method could still be used for determining SOM, however, this method should be avoided when determining SOC content. Transfer functions between all three methods were successfully created, however, only the WB and TDC methods had a very strong relationship (R^2 = 0.91) and showed that accuracy starts to decrease significantly after 2.5% SOC content. Therefore, the pedotransfer function (SOC_{WB} = -0.157 + 0.895 x SOC_{TDC} – 0.0149 x SOC_{TDC}^2 – 0.000606 x SOC_{TDC}^3) could be used for transferring SOC content values with SOC content up to 2.5%.

The second objective was assessed by collecting soil samples from five different study sites across South Africa and analysing these samples with the TDC method and scanning the samples with a NIR spectrometer. Sixty (60) NIR calibration algorithms were developed on a regional scale, with additional algorithms tailored for each site at various scales. The OSSL model was also used to compare how local data will perform using global prediction models. The impact of methodological parameters, such as sample state, sampling design, processing and ML models, on the RMSE of the validation statistics was also assessed. Although 60 regional-scale calibration algorithms were created, none were suitable for SOC content prediction. Findings of the study demonstrated that the OSSL prediction model cannot be used for predicting SOC content for South African soils. Sample state and sampling design were the most influential factors affecting RMSE. The cLHS design outperformed other sampling designs by capturing representative samples across the dataset, leading to more accurate and robust calibration algorithms. Prepared

samples yielded better results compared to field samples due to reduced variability and spectral interference from water molecules. Notably, the Tsitsa catchment and Ottosdal field were the only sites with viable calibration algorithms for both field and prepared states.

The third objective was assessed by collecting fifty (50) samples from three depths, 0-5, 5-15 and 15-30 cm and analysing the soil for SOC content using the TDC method. The ρb at unsampled locations was determined by creating a pedo-transfer function. Maps for SOC content, ρb and SOCS were created using OK and DSM with ML and assessed with validation statistics including RMSE and ρc. A calculate-first approach and map-first approach were used for calculating the total SOCS for each study site. Results of the study indicated that DSM with ML should be used rather than OK for mapping SOCS at field scale and therefore, using DSM with ML will lead to more accurate and reliable SOCS maps. However, the results also indicated that the SOC content, ρb and SOCS maps for DSM with ML need to be improved. Finally, the calculate-first and map-first approaches yielded very similar total SOCS, however, the calculate-first approach is simpler and decrease the room for error and therefore, is recommended for mapping SOCS at field scale.

The fourth and final objective was assessed by establishing a baseline assessment for 2022 and collecting 50 samples from each of three depth increments, 0-5, 5-15 and 15-30 cm in 2024. The SOCS was calculated for two study sites (for 2024) using the re-measure approach and the modelling approach using the RothC model. Finally, sampling cost was estimated against the value of sequestered SOCS and compared between the two approaches. The results were inconclusive regarding which approach should be used for quantifying SOCS at field scale in South Africa. The spatial variation of SOC was the main factor influencing the total SOCS, sequestered SOCS and carbon credit values. The results indicated that the modelling approach using the RothC model slightly underestimated the total SOCS and sequestered SOCS when utilising all 50 observations for both study sites. The industry-represented, “One-observation”, did not have a consistent trend and showed overprediction and underprediction for both study sites using the different approaches. This had significant financial implications using both approaches. The study also found that the spatial variation of the SOC must be taken into account at the onset of measuring the SOC content in year one, independent of whether the SOCS would be modelled or re-measured.

7.2 Significance of the study

From Chapter 3, the study found that the TDC method is recommended for determining SOC content, as there is no standard method being used when determining SOC content for carbon credits in South Africa. The study also found that the WB method should be avoided if the soil is

expected to have a high SOC content, while the LOI method could still be used for determining SOM, however, this method should be avoided when determining SOC content. The transfer function created can also now be used for transferring SOC content values between the WB and TDC methods. Therefore, the study successfully focused on addressing the gap regarding the measurement of SOC content

From Chapter 4, the research provided the first step to acquire a national calibration algorithm that could be used for predicting SOC content. Importantly, the study found that global prediction models, such as the OSSL, cannot be used for predicting SOC content for South African soils. The study also contributed to available soil spectroscopy data in previously underrepresented areas. Therefore, the study successfully focused on addressing the gap regarding predicting SOC content with NIR spectroscopy.

From Chapter 5, the research showed that using the conventional method of OK is not sufficient for mapping SOCS at field scale and that DSM with ML should rather be used. This research also showed that the calculate-first approach is simpler and decrease the room for error and that the calculate-first approach is recommended for mapping SOCS at field scale. Therefore, the research successfully focused on addressing the gap regarding how to map SOCS at field scale in South Africa.

From Chapter 6, the research showed that spatial variation of SOC was the main factor influencing the total SOCS, sequestered SOCS and carbon credit values. The research highlighted that the spatial variation of the SOC must be taken into account at the onset of measuring the SOC content in year one, independent of whether the SOCS would be modelled or re-measured. The carbon sequestration industry must also take the model limitations (such as field scale dynamics and sensitivity to short-term changes) into account, and find ways, such as an improved sampling design to address these accuracies of RothC at field scale.

7.3 Limitations of the study and recommendations for future research

From Chapter 3, the findings suggests that future research should focus on validating the current, created transfer functions and help develop new transfer functions for soils with SOC content higher than 2.5%. Increased sampling should be done, and regional and small-scale calibration and validation studies could be conducted to account for the diversity of South African soils, including focus on incorporating soils with a higher SIC fraction compared to SOC fraction.

From Chapter 4, the low number of samples was a limitation for creating the calibration algorithm at regional scale. To acquire a national calibration algorithm, effort should be placed on creating useful calibration algorithms for smaller areas that would be added to the national spectral library.

In this way, site specific accurate soil spectral calibration would provide incentive to gather the data sufficient for a national calibration algorithm. Further refinement of the standard calibration algorithm may include incorporating a larger and more diverse set of soil samples, and investigating other methodological parameters, such as spectral pre-processing and machine learning models that were not used for this study.

From Chapter 5, the low number of samples was also a limitation to the accuracy of maps generated using DSM with ML. Therefore, additional soil samples could be collected, and field scale covariates should be incorporated. These could include finer-resolution DEMs, yield data, electromagnetic induction and gamma-ray data. Moreover, since mapping SOC content and ρb at field scale is challenging efforts should focus on mapping these properties at farm scale instead. The use of deep learning models could also further improve mapping SOCS at field scale in South Africa.

From Chapter 6, the fine scale likely led to the variability of SOC not being accounted for and led to varying results. The model may also be sensitive to short term changes in SOCS, due to its design in focusing on long-term cycling. Future research could improve modelling SOCS using RothC at field scale, by incorporating field scale dynamics and calibrating the model for short-term cycling. The SOC sequestration industry must take these considerations of the model limitations into account, and find ways, such as an improved sampling design, to address these inaccuracies of RothC at field scale. Finally, more research should also focus on evaluating the costs of sampling against the value of SOCS sequestered and search for trade-offs between costs involved and alternative SOCS estimation methods.

REFERENCES

Abella, S.R., & Zimmer, B.W. 2007. Estimating organic carbon from loss-on-ignition in northern Arizona forest soils. *Soil Science Society of America Journal*, 71:545–550. https://acsess.onlinelibrary.wiley.com/doi/full/10.2136/sssaj2006.0136

Adbelbaki, A.M. 2018. Evaluation of pedotransfer functions for predicting soil bulk density for U.S. soils. *Ain Shams Engineering Journal*, 9:1611–1619. http://dx.doi.org/10.1016/j.asej.2016.12.002

Afzali, S.F., Azad, B., Golabi, M.H., & Francaviglia, R. 2019. Using RothC model to simulate soil organic carbon stocks under different climate change scenarios for the rangelands of the arid regions of southern Iran. *Water (Switzerland)*, 11(10). https://doi.org/10.3390/w11102107

Akpa, S.I., Odeh, I.O., Bishop, T.F., Hartemink, A.E., & Amapu, I.Y. 2016. Total soil organic carbon and carbon sequestration potential in Nigeria. *Geoderma*, 271:202–215. https://doi.org/10.1016/j.geoderma.2016.02.021

Alvarez, C., Alvarez, C.R., Costantini, A., & Basanta, M. 2014. Carbon and nitrogen sequestration in soils under different management in the semi-arid Pampa (Argentina). *Soil and Tillage Research*, 142:25–31. https://doi.org/10.1016/j.still.2014.04.005

Atkins, P., & de Paula, J. 2010. Atkins' physical chemistry. 9th ed. Oxford University Press: London, England.

Aynekulu, E., Vagen, T.G., Shephard, K., & Winowiecki, L. 2011. A protocol for modeling, measurement and monitoring soil carbon stocks in agricultural landscapes. Version 1.1. World Agroforestry Centre, Nairobi.

Bai, Z., Xie, M., Hu, B., Luo, D., Wan, C., Peng, J., & Shi, Z. 2022. Estimation of Soil Organic Carbon Using Vis-NIR Spectral Data and Spectral Feature Bands Selection in Southern Xinjiang, China. *Sensors*, 22(16). https://doi.org/10.3390/s22166124

Bao, Y., Yao, F., Meng, X., Zhang, J., Liu, H., & Mouazen, A. 2023. Predicting soil organic carbon in cultivated land across geographical and spatial scales: Integrating Sentinel-2A and laboratory Vis-NIR spectra. *ISPRS Journal of Photogrammetry and Remote Sensing*, *203*. https://doi.org/10.1016/j.isprsjprs.2023.07.020

Barbouchi, M., Bahri, H., Souissi, H., Mhammed, C., & Annabi, M. 2020. Soil organic carbon mapping in North of Tunisia: Comparison between different interpolation methods. 1st African Conference on Precision Agriculture, 8-10 December.

Barwari, V.I.H., Hashim, F.A., & Mohammed, B.H. 2017. Comparison between Walkley-Black and Loss-On-Ignition methods for organic carbon estimation in soils from different locations. *Kufa Journal for Agricultural Sciences*, 2:292–306. https://www.researchgate.net/publication/349064091_Comparison_between_Walkley-Black_and_Loss-On-Ignition_methods_for_organic_carbon_estimation_in_soils_from_different_locations?enrichId=rgreq-27d6468eb8caf85e17c1ad4e13355f91-XXX&enrichSource=Y292ZXJQYWdlOzM0OTA2NDA5MTtBUzo5ODc5NjQ4NTEwOTc2MDBAMTYxMjU2MDU4NDY4Mg%3D%3D&el=1_x_3&_esc=publicationCoverPdf

Batey, T. 2009. Soil compaction and soil management – a review. *Soil Use and Management*, 25:335–345. https://doi.org/10.1111/j.1475-2743.2009.00236.x

Batjes, N.H. 1996. Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science*, 47(2):151-163. https://doi.org/10.1111/j.1365-2389.1996.tb01386

Batool, M., Cihacek, L.J., & Alghamdi, R.S. 2024. Soil Inorganic Carbon Formation and the Sequestration of Secondary Carbonates in Global Carbon Pools: A Review. *Soil Systems*, 8(1). https://doi.org/10.3390/soilsystems8010015

Baurov, A.A. 2021. Methods of Carbon Sequestration (Review). *Polymer Science, Series D*, 14:603–605. https://doi.org/10.1134/S1995421222010038

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., & McBratney, A. 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *Trends in Analytical Chemistry*, 29(9):1073–1081. https://doi.org/10.1016/j.trac.2010.05.006

Ben-Dor, E., & Banin, A. 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Science Society of America Journal*, 59:364–372. https://doi.org/10.2136/sssaj1995.03615995005900020014x

Ben-Dor, E., Irons, J. R., & Epema, G.F. 1999. Soil reflectance. In ‘‘Remote Sensing for the Earth Sciences: Manual of Remote Sensing’’ (A. N. Rencz, Ed.), 3rd edn: 111–188). Wiley, New York.

Biney, J. K. M., Boruvka, L., Agyeman, P. C., Němeček, K., & Klement, A. 2020. Comparison of field and laboratory wet soil spectra in the vis-NIR range for soil organic carbon prediction in the absence of laboratory dry measurements. *Remote Sensing*, *12*(18). https://doi.org/10.3390/RS12183082

Blake, G.R. & Hartge, K.H. 1986. Bulk Density Methods of Soil Analysis, Part 1, Physical and Mineral Method. Ed. by A. Klute, 363–382.

Bodenstein, D. 2023. Conversation regarding RothC (Personal communication).

Bolinder, M.A., Janzen, H.H., Gregorich, E.G., Angers, D.A., VandenBygaart, A.J. 2007. An approach for estimating net primary productivity and annual carbon inputs to soil for common agricultural crops in Canada. *Agriculture, Ecosystems and Environment*, 118(1–4):29–42. https://doi.org/10.1016/j.agee.2006.05.013

Bramley, R.G.V., & Janik, L.J. 2005. Precision agriculture demands a new approach to soil and plant sampling and analysis - Examples from Australia. *Communications in Soil Science and Plant Analysis*, 36(1–3):9–22. https://doi.org/10.1081/CSS-200042958

Breiman, L. 2001. Random Forests. *Machine Learning*, 45(1):5–32. http://dx.doi.org/10.1023/A:1010933404324

Brevik, E.C., Calzolari, C., Miller, B.A., Pereira, P., Kabala, C., Baumgarten, A. & Jordán, A. 2016. Soil mapping, classification, and pedologic modeling: history and future directions. *Geoderma*, 264:256–74. https://doi.org/10.1016/j.geoderma.2015.05.017

Brilli, L., Bechini, L., Bindi, M., Carozzi, M., Cavalli, D., Conant, R., Dorich, C.D., Doro, L., Ehrhardt, F., Farina, R., Ferrise, R., Fitton, N., Francaviglia, R., Grace, P., Iocola, I., Klumpp, K., Léonard, J., Martin, R., Massad, R.S., Recous, S., Seddaiu, G., Sharp, J., Smith, P., Smith, W.N., Soussana, J-F., & Bellocchi, G. 2017. Review and analysis of strengths and weaknesses of agro-ecosystem models for simulating C and N fluxes. *Science of the Total Environment*, 598:445–470. https://doi.org/10.1016/j.scitotenv.2017.03.208

Brunetto, G., Melo, G.W., Kaminski, J., Furlanetto, V., & Fialho, F.B. 2006. Evaluation of the loss-on-ignition method in the organic matter analysis in soils of the Serra Gaucha of the Rio Grande do Sul. *Ciência Rural, Santa Maria*, 36(6):1936–1939. https://www.researchgate.net/publication/262737370_Evaluation_of_the_loss-on-ignition_method_in_the_organic_matter_analysis_in_soils_of_the_Serra_Gaucha_of_the_Rio_Grande_do_Sul

Brus, D.J., de Gruijter, J.J., & van Groenigen, J.W. 2007. Designing Spatial Coverage Samples Using the k-means Clustering Algorithm. *Developments in Soil Science*, 31:183–192. https://doi.org/10.1016/S0166-2481(06)31014-8

Bryk, M., & Kołodziej, B. 2023. Pedotransfer Functions for Estimating Soil Bulk Density Using Image Analysis of Soil Structure. *Sensors*, 23(1852):1–19. https://doi.org/10.3390/s23041852

Bui, E., Henderson, B., & Viergever, K. 2009. Using knowledge discovery with data mining from the Australian Soil Resource Information System database to inform soil carbon mapping in Australia. *Global Biogeochemical Cycles*, 23(4):1–15. https://doi.org/10.1029/2009GB003506

Camacho-Tamayo, J.H., Rubiano, Y.S., & Hurtado, M.S. 2014. Near-infrared (NIR) diffuse reflectance spectroscopy for the prediction of carbon and nitrogen in an Oxisol. *Agronomia Colombiana*, 32(1):96–94. https://www.cabidigitallibrary.org/doi/pdf/10.5555/20143202549

Cambule, A.H., Rossiter, D.G., Stoorvogel, J.J., & Smaling, E.M.A. 2012. Building a near infrared spectral library for soil organic carbon estimation in the Limpopo National Park, Mozambique. *Geoderma*, 183–184:41–48. https://doi.org/10.1016/j.geoderma.2012.03.011

Cerri, C.E.P., Cherubin, M.R., Damian, J.M., Mello, F.F.C., & Lal, R. 2021. Soil carbon sequestration by sustainable management practices: Potential and opportunity for the Americas. Inter-American Institute for Cooperation on Agriculture.

Chabala, L.M., Mulolwa, A., & Lungu, O. 2017. Application of Ordinary Kriging in Mapping Soil Organic Carbon in Zambia. *Pedosphere*, 27(2):338–343. https://doi.org/10.1016/S1002-0160(17)60321-7

Chan, Y. 2008. Increasing soil organic carbon of agricultural land. *PrimeFacts*, 375:1–5. www.dpi.nsw.gov.au/primefacts

Chang, C.W., Laird, D.A., Mausbach, M.J., & Hurburgh, C.R. 2001 Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. *Soil Science Society of America Journal*, 65:480–490. https://doi.org/10.2136/sssaj2001.652480x

Chatterjee, A., Lal, R., Wielopolski, L., Martin, M.Z. & Ebinger, M.H. 2009. Evaluation of different soil carbon determination methods. *Critical Reviews in Plant Sciences*, 28:164–178. doi:10.1080/07352680902776556

Chen, J., Luo, Y., van Groenigen, J.K., Hungate, B.A., Cao, J., Zhou, X., & Wang, R. 2018. A keystone microbial enzyme for nitrogen control of soil carbon storage. *Science Advances*, 4(8):eaaq168. https://pmc.ncbi.nlm.nih.gov/articles/PMC6105232/

Chen, S., Chen, Z., Zhang, X., Luo, Z., Schillaci, C., Arrouays, D., Richer-de-Forges, A.C., & Shi, Z. 2024. European topsoil bulk density and organic carbon stock database (0–20cm) using machine-learning-based pedotransfer functions. *Earth System Science Data*, 16(5):2367–2383. https://doi.org/10.5194/essd-16-2367-2024

Christensen, B.T., & Malmros, P. 1982. Loss-on-ignition and carbon content in a beech forest soil profile. *Holarctic Ecology*, 5:376–380. https://www.jstor.org/stable/3682222

Clairotte, M., Grinand, C., Kouakoua, E., Thébault, A., Saby, N.P.A., Bernoux, M., & Barthès, B.G. 2016. National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy. *Geoderma*, 276:41–52. https://doi.org/10.1016/j.geoderma.2016.04.021

Coleman, K., & Jenkinson, D.S. 1996. RothC-26.3 - A Model for the turnover of carbon in soil. In: Powlson, D.S., Smith, P., Smith, J.U. (eds) Evaluation of Soil Organic Matter Models. NATO ASI Series, volume 38. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-61094-3_17

Coleman, K., & Jenkinson, D.S. 2005. Rothc-26.3. A model for the turnover of carbon in soil. Model Description and Windows Users' Guide. Institute of Arable Crops Research, Rothamsted, UK.

Coleman, K., & Jenkinson, D.S. 2014. RothC-A model for the turnover of carbon in soil Model description and users guide (Windows version). http://www.rothamsted.ac.uk/sustainable-soils-and-grassland-

Coleman, K., Prout, J.M., Milne, A.E., Coleman, K., Prout, J., & Milne, A. 2024. RothC-A model for the turnover of carbon in soil Model Description. Rothamsted Research Harpenden, Herts. https://www.rothamsted.ac.uk/sites/default/files/Documents/RothC_description.pdf

Comstock, J.P., Sherpa, S.R., Ferguson, R., Bailey, S., Beem-Miller, J.P., Lin, F., Lehmann, J., & Wolfe, D.W. 2019. Carbonate determination in soils by mid-IR spectroscopy with regional and continental scale models. *PLoS ONE*, 14(2):1–19.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8:1991–2007. doi:10.5194/gmd-8-1991-2015.

Cortes, C., & Vapnik, V. 1995. Support-vector networks. *Machine Learning*, 20**:**273–297 (1995). https://doi.org/10.1007/BF00994018

Council for Geoscience. 2019. Geological Data 1:1 000 000. Council for Geoscience. Pretoria, South Africa.

Dakhlalla, A.O., Parajuli, P.B., Ouyang, Y., & Schmitz, D.W. 2016. Evaluating the impacts of crop rotations on groundwater storage and recharge in an agricultural watershed. *Agricultural Water Management*, 163:332–343. https://doi.org/10.1016/j.agwat.2015.10.001

Dalal, R. C., & Henry, R. J. 1986. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. *Soil Science Society of America Journal*, 50:120–123. https://doi.org/10.2136/sssaj1986.03615995005000010023x

Dangal, S.R.S., Sanderman, J., Wills, S., & Ramirez-Lopez, L. 2019. Accurate and precise prediction of soil properties from a large mid-infrared spectral library. *Soil Systems*, 3(1):1–23. https://doi.org/10.3390/soilsystems3010011

Davoudabadi, M.J., Pagendam, D., Drovandi, C., Baldock, J., & White, G. 2024. Innovative approaches in soil carbon sequestration modelling for better prediction with limited data. *Scientific Reports*, 14(1). https://doi.org/10.1038/s41598-024-53516-z

De Freitas, P.L., & Landers, J.N. 2014. The transformation of agriculture in Brazil through development and adoption of zero tillage conservation agriculture. *International Soil & Water Conservation Research,* 2(1):35–46. https://doi.org/10.1016/S2095-6339(15)30012-5

De Gruijter, J.J., Brus, D.J., Bierkens, M.F.P., & Knotters, M. 2006. Sampling for natural resource monitoring. Springer Science & Business Media, Dordrecht.

De Gryze, S., Wolf, A., Kaffka, S.R., Mitchell, J., Rolston, D.E., Temple, S.R., Lee, J., & Six, J. 2010. Simulating greenhouse gas budgets of four California cropping systems under conventional and alternative management. *Ecological applications*, 20(7):1805–1819. https://doi.org/10.1890/09-0772.1

De Santana, F.B., Otani, S.K., de Souza, A.M., & Poppi, J. R. 2021. Comparison of PLS and SVM models for soil organic matter and particle size using vis-NIR spectral libraries. *Geoderma Regional*, 27:e00436. https://doi.org/10.1016/j.geodrs.2021.e00436

De Vos, B., Lettens, S., Muys, B., & Deckers, J. A. 2007. Walkley-Black analysis of forest soil organic carbon: Recovery, limitations and uncertainty. *Soil Use and Management*, *23*(3): 221–229. https://doi.org/10.1111/j.1475-2743.2007.00084.x

De Vos, B., Vandecasteele, B., Deckers, J., & Muys, B. 2005. Capability of Loss-on-Ignition as a Predictor of Total Organic Carbon in Non-Calcareous Forest Soils. *Communications in Soil Science and Plant Analysis*, 36(19-20):2899–2921. https://doi.org/10.1080/00103620500306080

Dechow, R., Franko, U., Kätterer, T., & Kolbe, H. 2019. Evaluation of the RothC model as a prognostic tool for the prediction of SOC trends in response to management practices on arable land. *Geoderma*, 337:463–478. https://doi.org/10.1016/j.geoderma.2018.10.001

Deng, F., Minasny, B., Knadel, M., McBratney, A., Heckrath, G., & Greve, M.H. 2013. Using Vis-NIR spectroscopy for monitoring temporal changes in soil organic carbon. *Soil Science*, 178(8):389–399. https://doi.org/10.1097/SS.0000000000000002

Deng, X., Chen, X., Ma, W., Ren, Z., Zhang, M., Grieneisen, M.L., Long, W., Ni, Z., Zhan, Y., & Lv, X. 2018. Baseline map of organic carbon stock in farmland topsoil in East China. *Agriculture, Ecosystems and Environment*, 254:213–223. https://doi.org/10.1016/j.agee.2017.11.022

Dierke, C., & Werban, U. 2013. Relationships between gamma-ray data and soil properties at an agricultural test site. *Geoderma*, 199:90–98. https://doi.org/10.1016/j.geoderma.2012.10.017

Doolittle, J.A., & Brevik, E.C. 2014. The use of electromagnetic induction techniques in soils studies. *Geoderma*, 223–225:33–45. http://dx.doi.org/10.1016/j.geoderma.2014.01.027

Douzouné, K., Oloukoi, J., Gongnet, E.E., Sèdjro, T., & Affossogbe, A. 2024. Application of Ordinary Kriging in Mapping Soil Organic Carbon in Chad using SoilGrids data. *South African Journal of Geomatics*, 13(2):397–408. https://dx.doi.org/10.4314/sajg.v13i2.11

Du Plessis, C., van Zijl, G., van Tol, J., & Manyevere, A. 2020. Machine learning digital soil mapping to inform gully erosion mitigation measures in the Eastern Cape, South Africa. *Geoderma*, 368. https://doi.org/10.1016/j.geoderma.2020.114287

Duan, L., Li, Z., Xie, H., Yuan, H., Li, Z., & Zhou, Q. 2018. Regional pattern of soil organic carbon density and its influence upon the plough layers of cropland. *Land Degradation and Development*, 31(16):2461–2474. https://doi.org/10.1002/ldr.3610

El Mouridi, Z., Ziri, R., Douaik, A., Bennani, S., Lembaid, I., Bouharou, L., Brhadda, N., & Moussadek, R. 2023. Comparison Between Walkley-Black and Loss On Ignition Methods for Organic Matter Estimation in Different Moroccan Soils. *Ecological Engineering & Environmental Technology*, 24(4):253–259. http://www.ecoeet.com/Comparison-between-Walkley-Black-and-Loss-on-Ignition-Methods-for-Organic-Matter,163121,0,2.html

Falloon, P., & Smith, P. 2002. Simulating SOC changes in long-term experiments with RothC and CENTURY: model evaluation for a regional scale application. *Soil Use and Management*, 18:101–111. https://doi.org/10.1079/SUM2001108

Falloon, P.D. 2001. Large scale spatial modelling of soil organic carbon dynamics. Nottingham: The University of Nottingham. (Thesis – PhD).

Falloon, P.D., Smith, P., Coleman, K., & Marshall, S. 1998. Estimating the size of the inert organic matter pool from total soil organic carbon content for use in the Rothamsted carbon model. *Soil Biology and Biochemistry,* 30(8/9):1207–1211. https://doi.org/10.1016/S0038-0717(97)00256-3

FAO (Food and Agriculture Organization for the United Nations). 2019. Measuring and modelling soil carbon stocks and stock changes in livestock production systems: Guidelines for assessment. Rome: Livestock Environmental Assessment and Performance (LEAP) Partnership.

FAO (Food and agriculture Organization for the United Nations). 2022. A primer on soil analysis using visible and near-infrared (vis-NIR) and mid-infrared (MIR) spectroscopy. Rome, FAO. https://doi.org/10.4060/cb9005en

Farooq, I., Bangroo, S.A., Bashir, O., Shah, T.I., Malik, A.A., Iqbal, A.M., Mahdi, S.S., Wani, O.A., Nazir, N., & Biswas, A. 2022. *Land*, 11(2180):1–15. https://doi.org/10.3390/land11122180

Fernandes, R.B.A., de Carvalho Junior, I.A., Ribeiro Junior, E.S., & de Sá Mendonça. 2015. Comparison of different methods for the determination of total organic carbon and humic substances in Brazilian soils. *Revista Ceres*, 62(5):496–501. https://www.scielo.br/j/rceres/a/hR6p9wZ38DsFQLwTgqVbzMj/

Fey, M. 2010. Soils of South Africa. Cambridge University Press, Cape Town, South Africa.

Flatman, G.T., & Yfantis, A.A. 1984. Geostatistical strategy for soil sampling: The survey and the census. *Environmental Monitoring and Assessment*, 4(4):335–349. https://pubmed.ncbi.nlm.nih.gov/24257861/

Flynn, T., de Clercq, W., Rozanov, A., & Clarke, C. 2019. High-resolution digital soil mapping of multiple soil properties: an alternative to the traditional field survey? *South African Journal of Plant and Soil*, 36:237–247. https://doi.org/10.1080/02571862.2019.1570566

Flynn, T., Wiese, L., Rozanov, A. 2022. Soil carbon stock assessment using depth and spatial models on afforested arable lands. *South African Journal of Plant and Soil*, 39(4):235–247. https://doi.org/10.1080/02571862.2022.2079741

Fox, J., & Weisberg, S. 2019. *An R Companion to Applied Regression*, Third Edition, Sage Publications Ltd.

Franzluebbers, A.J. 2002. Water infiltration and soil structure related to organic matter and its stratification with depth. *Soil and Tillage Research*, 66:197–205. https://doi.org/10.1016/S0167-1987(02)00027-2

Gavrilova, O., Leip, A., Dong, H., Douglas MacDonald, J., Alfredo Gomez Bravo, C., Amon, B., Barahona Rosales, R., del Prado, A., Aparecida de Lima, M., Oyhantçabal, W., John van der Weerden, T., Widiawati, Y., Bannink, A., Beauchemin, K., & Clark, H. 2019. Chapter 10: Emissions from Livestock and Manure Management, in: 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories.

Ge, Y., Wadoux, A., & Peng, Y. 2022. A primer on soil analysis using visible and near-infrared (vis-NIR) and mid-infrared (MIR) spectroscopy. In *Soil spectroscopy training material*. FAO. https://doi.org/10.4060/cb9005en

George, J., & Kumar, S. 2020. Soil organic carbon prediction using visible-near infrared reflectance spectroscopy employing artificial neural network modelling. *Current Science*. https://www.jstor.org/stable/27229874.

GeoTerra Image. 2020. https://geoterraimage.com/

Geremew, B., Tadesse, T., Bedadi, B., Gollany, H.T., Tesfaye, K., Aschalew, A., Tilaye, A., & Abera, W. 2024. Evaluation of RothC model for predicting soil organic carbon stock in north-west Ethiopia. *Environmental Challenges*, 15(100909). https://doi.org/10.1016/j.envc.2024.100909

Gessesse, T.A., & Khamzina, A. 2018. How reliable is the Walkley-Black method for analyzing carbon-poor, semi-arid soils in Ethiopia? *Journal of Arid Environments*, 153:98–101. https://www.researchgate.net/publication/323846790_How_reliable_is_the_Walkley-Black_method_for_analyzing_carbon-poor_semi-arid_soils_in_Ethiopia

Gholizadeh, A., Neumann, C., Chabrillat, S., van Wesemael, B., Castaldi, F., Borůvka, L., Sanderman, J., Klement, A., & Hohmann, C. 2021. Soil organic carbon estimation using VNIR–

SWIR spectroscopy: The effect of multiple sensors and scanning conditions. *Soil and Tillage Research*, 211. https://doi.org/10.1016/j.still.2021.105017

Gomes, L.C., Faria, R.M., de Souza, E., Veloso, G.V., Ernesto, C., Schaefer, G.R., & Filho, E.I.F. 2019. Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, 340:337–350. https://doi.org/10.1016/j.geoderma.2019.01.007

González-Molina, L., Etchevers-Barra, J.D., Paz-Pellat, F., Díaz-solis, H., Fuentes-Ponce, M.H., Covaleda-Ocón, S., & Pando-Moreno, M. 2011. Performance of the RothC-26.3 model in short-term experiments in Mexican sites and systems. *The Journal of Agriculture Science*, 149(4):415–425. https://doi.org/10.1017/S0021859611000232

Gozubuyuk, Z., Sahin, U., Adiguzel, M.C., Ozturk, I., & Celik, A. 2015. The influence of different tillage practices on water content of soil and crop yield in vetch–winter wheat rotation compared to fallow–winter wheat rotation in a high altitude and cool climate. *Agricultural Water Management*, 160:84–97. https://doi.org/10.1016/j.agwat.2015.07.003

Grunwald, S., Mizuta, K., Ceddia, M.B., Pinheiro, É.F.M., Kay Kastner-Wilcox, R., Gavilan, C.P., Ross, C.W., & Clingensmith, C.M. 2016. In: McBratney, A.B., Morgan, C.L.S., & Field, D. (Eds.), The meta soil model: an integrative multi-model framework for soil security. Global Soil Security, Springer Publ., New York, NY.

Guan, F., Xia, M., Tang, X., & Fan, S. 2017. Spatial variability of soil nitrogen, phosphorus and potassium contents in Moso bamboo forests in Yong'an City, China. *Catena*, 150:161–172. https://doi.org/10.1016/j.catena.2016.11.017

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. 2013. Multivariate data analysis. Pearson Education Limited, Upper Saddle River.

Hartemink A.E., & McSweeney, K. 2014. Soil Carbon: Progress in Soil Science. WI: Springer International Publishing.

Henderson, T.L., Baumgardner, M.F., Franzmeier, D.P., Stott, D.E., & Coster, D.C. 1992. High dimensional reflectance analysis of soil organic-matter. *Soil Science Society of America Journal*, 56:865–872. https://doi.org/10.2136/sssaj1992.03615995005600030031x

Hernández, N., Kiralj, R., Ferreira, M.M.C., Talavera, I., 2009. Critical comparative analysis, validation and interpretation of SVM and PLS regression models in a QSAR study on HIV-1 protease inhibitors. *Chemometrics and Intelligent Laboratory Systems*, 98:65–77. https://doi.org/10.1016/j.chemolab.2009.04.012.

Hinge, G., Surampalli, R.Y., & Goyal, M.K. 2018. Prediction of soil organic carbon stock using digital mapping approach in humid India. *Environmental Earth Sciences*, 77:172. https://doi.org/10.1007/s12665-018-7374-x

Hoffmann, U., Hoffmann, T., Jurasinski, G., Glatzel, S., & Kuhn, N.J. 2014. Assessing the spatial variability of soil organic carbon stocks in an alpine setting (Grindelwald, Swiss Alps). *Geoderma*, 232–234:270–283. http://dx.doi.org/10.1016/j.geoderma.2014.04.038

Hoogsteen, M.J.J., Lantinga, E.A., Bakker, E.J., Groot, J.C.J., & Tittonell, P.A. 2015. Estimating soil organic carbon through loss on ignition: effects of ignition conditions and structural water loss. *European Journal of Soil Science*, 66:320–328. https://bsssjournals.onlinelibrary.wiley.com/doi/10.1111/ejss.12224

Huang, H., Yang, L., Zhang, L., Pu, Y., Yang, C., Wu, Q., Cai, Y., Shen, F., & Zhou, C. 2022. A review on digital mapping of soil carbon in cropland: progress, challenge, and prospect. *Environmental Research Letters*, 17(123004). https://doi.org/10.1088/1748-9326/aca41e

Ighodaro, I.D., Lategan, F.S., & Yusuf, S.F.G. 2013. The Impact of Soil Erosion on Agricultural Potential and Performance of Sheshegu Community Farmers in the Eastern Cape of South Africa. *Journal of Agricultural Science*, 5(5). https://doi.org/10.5539/jas.v5n5p140

IPCC (Intergovernmental Panel on Climate Change). 2006. IPCC guidelines for national greenhouse gas inventories. Prepared by the National Greenhouse Gas Inventories Programme, in: Eggleston, H.S., Buendia, L., Miwa, K., Ngara, T. & Tanabe, K. (Eds.). IGES, Japan.

ISO (International Organization for Standardization). 2006. ISO:14044:2006 Environmental management - Life cycle assessment- Requirements and guidelines. International Standardization Organization for Standardization (ISO). Geneva.

IUSS (International Union of Soil Sciences) Working Group WRB. 2019. Digital Soil Mapping. www.digitalsoilmapping.org

Janik, L.J., Skjemstad, J., Shepherd, K., & Spouncer, L. 2007 The prediction of soil carbon fractions using mid-infrared-partial least square analysis. *Soil Restoration*, 45:73–81. https://doi.org/10.1071/SR06083

Jaskauskas, B., Jankauskiene, G., Slepetiene, A., Fullen, M.A., & Booth, C.A. 2005. International Comparison of Analytical Methods of Determining the Soil Organic Matter Content of Lithuanian Eutric Albeluvisols. *Communications in Soil Science and Plant Analysis,* 37:707–

720. https://www.tandfonline.com/doi/full/10.1080/00103620600563499

Jebari, A., Alvaro-Fuentes, J., Pardo, G., Almagro, M., & del Prado, A. 2021. Estimating soil organic carbon changes in managed temperate moist grasslands with RothC. *PLoS ONE*, 16. https://doi.org/10.1371/journal.pone.0256219

Jensen, J.L., Christensen, B.T., Schjønning, P., Watts, C.W., & Munkholm, L.J. 2018. Converting loss-on-ignition to organic carbon content in arable topsoil: pitfalls and proposed procedure. *European Journal of Soil Science*, 69**:**604–612. https://pmc.ncbi.nlm.nih.gov/articles/PMC6109958/

Jin, H., Hongwen, L., Rasaily Rabi, G., Qingjie, W., Guohua, C., Yanbo, S., Xiaodong, Q., & Lijin, L. 2011. Soil properties and crop yields after 11 years of no tillage farming in wheat–maize cropping system on North China Plain. *Soil and Tillage Research*, 113:48–54. https://doi.org/10.1016/j.still.2011.01.005

John, K., Abraham, I.I., Kebonye, N.M., Agyeman, P.C., Ayito, E.O., & Kudjo, A.S. 2021. *Journal of the Saudi Society of Agricultural Sciences*, 20(6):379–389. https://doi.org/10.1016/j.jssas.2021.04.005

Jordon, M.W. 2021. Data and R code for RothC-26.3 simulations of management changes on Great Britain arable land (1.0.0) [Data Set]. *Zenodo*. https://doi.org/10.5281/zenodo.5734400.

Kaonga, M.L., & Coleman, K. 2008. Modelling soil organic carbon turnover in improved fallows in eastern Zambia using the RothC-26.3 model. *Forest Ecology and Management*, 256(5):1160–1166. https://doi.org/10.1016/j.foreco.2008.06.017

Kasozi, G.N., Nkedi-Kizza, P., & Harris, W.G. 2009. Varied carbon content of organic matter in Histosols, spodosols, and carbonatic soils. *Soil Science Society of America Journal*, 73:1313–1318. https://doi.org/10.2136/sssaj2008.0070

Kassambara, A., & Mundt, F. 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7.999. https://github.com/kassambara/factoextra.

Kenkam, A.M., Wadoux, A., Minasny, B., Silatsa, F.B.T., Yemefack, M., Ugbaje, S.U., Akpa, S., van Zijl, G., Bouasria, A., Bouslihim, Y., Chabala, L.M., Ali, A., & McBratney, A.B. 2024. *Geoderma*, 449(117007). https://doi.org/10.1016/j.geoderma.2024.117007

Keskin, H., Grunwald, S., & Harris, W.G. 2019. Digital mapping of soil carbon fractions with machine learning. *Geoderma*, 339:40–58. https://doi.org/10.1016/j.geoderma.2018.12.037

Khaledian, Y., & Miller, B.A. 2020. Selecting appropriate machine learning methods for digital soil mapping. *Applied Mathematical Modelling*, 81:401–418. https://doi.org/10.1016/j.apm.2019.12.016

Kimura, S.D., Mishima, S.I., &Yagi, K. 2011. Carbon resources of residue and manure in Japanese farmland soils. *Nutrient Cycling in Agroecosystems*, 89(2):291–302. https://doi.org/10.1007/s10705-010-9394-0

Knadel, M., Arthur, E., Weber, P., Moldrup, P., Greve, M.H., Chrysodonta, Z.P., & de Jonge, L.W. 2018. Soil Specific Surface Area Determination by Visible Near-Infrared Spectroscopy. *Soil Science Society of America Journal*, 82(5):1046–1056. https://doi.org/10.2136/sssaj2018.03.0093

Knadel, M., Thomsen, A., Schelde, K., & Greve, M.H. 2015. Soil organic carbon and particle sizes mapping using vis-NIR, EC and temperature mobile sensor platform. *Computers and Electronics in Agriculture*, 114:134–144. https://doi.org/10.1016/j.compag.2015.03.013

Knox, N.M., Grunwald, S., McDowell, M.L., Bruland, G.L., Myers, D.B., & Harris, W.G. 2015. Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. *Geoderma*, 239–240:229–239. https://doi.org/10.1016/j.geoderma.2014.10.019

Kock, A. 2022. *Creation of mid-infrared spectroscopy calibration algorithms for soil property predictions*. Potchefstroom: North-West University. (Thesis – MSc). https://repository.nwu.ac.za/bitstream/handle/10394/39565/Kock%20A%2025901966.pdf?sequence=1&isAllowed=y

Kock, A., Ramphisa-Nghondzweni, P.D., & van Zijl, G.M. 2024. Development of soil spectroscopy models for the Western Highveld region, South Africa: Why do we need local data? *European Journal of Soil Science*, 75(6):e70014.

Konen, M.E., P.M., Jacobs, C.L., Burras, B.J., Talaga, & Mason, J.A. 2002. Equations for predicting soil organic carbon using loss-on-ignition for north central US soils. *Soil Science Society of American Journal*, 66:1878. https://acsess.onlinelibrary.wiley.com/doi/10.2136/sssaj2002.1878

Kotze, J., & van Tol, J. 2023. Extrapolation of Digital Soil Mapping Approaches for Soil Organic Carbon Stock Predictions in an Afromontane Environment. *Land*, 12(520):1–18. https://doi.org/10.3390/land12030520

Krige, D.G. 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical Metallurgical & Mining Society of South Africa*, 52(6):119–139. https://journals.co.za/doi/pdf/10.10520/AJA0038223X_4792

Krishnan, P., Alexander, D.J., Butler, B., & Hummel, J.W. 1980. Reflectance technique for predicting soil organic matter. *Soil Science Society of America Journal*, 44:1282–1285. https://doi.org/10.2136/sssaj1980.03615995004400060030x

Kruskal, W.H., & Wallis, W.A. 1952. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47:583–621. https://doi.org/10.2307/2280779

Kuhn, M. 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5):1–26. https://doi.org/10.18637/jss.v028.i05

Kuhn, M., Weston, S., Keefer, C., Coulter, N., & Quinlan, R. 2024. Cubist: Rule- and Instance-Based Regression Modeling. R package version 0.4.4. https://cran.r-project.org/package=Cubist

Kumar, S., & Lal, R. 2011. Mapping the organic carbon stocks of surface soils using local spatial interpolator. *Journal of Environmental Monitoring*, 13:3128–3135. https://pubs.rsc.org/en/content/articlelanding/2011/em/c1em10520e

Kumar, S., Ghotekar, Y.S., & Dadhwal, V.K. 2019. C-equivalent correction factor for soil organic carbon inventory by wet oxidation, dry combustion and loss on ignition methods in Himalayan region. *Journal of Earth Sciences and Environment*, 128(62):1–10. https://link.springer.com/article/10.1007/s12040-019-1086-9

Kusumo, B.H., Sukartono, & Bustan, B. 2018. The rapid measurement of soil carbon stock using near-infrared technology. *IOP Conference Series: Earth and Environmental Science*, 129(1). https://doi.org/10.1088/1755-1315/129/1/012023

Kuzyakov, Y., & Domanski, G. 2000. Carbon input by plants into the soil. Review. *Journal of Plant Nutrition and Soil Science*, 163(4):421–431. https://doi.org/10.1002/1522-2624(200008)163:4<421::AID-JPLN421>3.0.CO;2-R

Lal, R. 2009. Challenges and opportunities in soil organic research. *European Journal of Soil Science*, 60(1):158-169. https://doi.org/10.1111/j.1365-2389.2008.01114.x

Lamichhane, S., Kumar, L., & Wilson, B. 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma*, 352:395–413. https://doi.org/10.1016/j.geoderma.2019.05.031

Land Type Survey Staff. 1972-2006. Land Types of South Africa: Digital Map (1: 250 000 Scale) and Soil Inventory Datasets. Agriculture Research Council Institute for Soil, Climate and Water. Pretoria, South Africa

Lawrence, I., & Lin, K. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45:255–268. https://doi.org/10.2307/2532051

Lessmann, M., Ros, G.H., Young, M.D., & de Vries, W. 2022. Global variation in soil carbon sequestration potential through improved cropland management. *Global Change Biology*, 28(3):1162–1177. https://doi.org/10.1111/gcb.15954

Li, H., Jia, S., & Le, Z. 2020. Prediction of soil organic carbon in a new target area by near-infrared spectroscopy: Comparison of the effects of spiking in different scale soil spectral libraries. *Sensors (Switzerland)*, 20(16):1–14. https://doi.org/10.3390/s20164357

Li, L., Yue, Y., Qin, F., Dong, X., Sun, C., Liu, Y., & Zhang, P. 2022. Multi-Scale Characterization of Spatial Variability of Soil Organic Carbon in a Semiarid Zone in Northern China. *Sustainability*, 14(15):9390. https://doi.org/10.3390/su14159390

Li, T., Hasegawa, T., Yin, X., Booth, K., Adam, M., Bregaglio, S., Buis, S., Confalonieri, R., Fumoto, T., Gaydon, D., Marcaida, M., Nakagawa, H., Oriol, P., Ruane, A.C., Ruget, F., Singh, B., Singh, U., Tang, L., Tao, F., Wilkens, P., Yoshida, H., Zhang, Z., & Bouman, B. 2014. Uncertainties in predicting rice yield by current cropmodels under a wide range of climatic conditions. *Global Change Biology*, 21:1328–1341. https://doi.org/10.1111/gcb.12758

Li, Y., Wang, X., Chen, Y., Gong, X., Yao, C., Cao, W., & Lian, J. 2023. Application of predictor variables to support regression kriging for the spatial distribution of soil organic carbon stocks in native temperate grasslands. *Journal of Soils and Sediments*, 23:700–717. https://doi.org/10.1007/s11368-022-03370-1

Liu, D., Chan, K.Y., Conyers, M.K., Li, G., Poile, G.J. 2011. Simulation of soil organic carbon dynamics under different pasture managements using the RothC carbon model. *Geoderma*, 165:69–77. https://doi.org/10.1016/j.geoderma.2011.07.005

Liu, D.L., Chan, K.Y., & Conyers, M.K. 2009. Simulation of soil organic carbon under different tillage and stubble management practices using the Rothamsted carbon model. *Soil and Tillage Research*, 104:65–73. https://doi.org/10.1016/j.still.2008.12.011

Long, M., Yue, T., Xu, Z., Guo, J., Luo, J., Guo, X., & Zhao, X. 2023. Improved Soil Organic Carbon Prediction in a Forest Area by Near-Infrared Spectroscopy: Spiking of a Soil Spectral Library. *Forests*, 14(1). https://doi.org/10.3390/f14010118

Löwik, E. 2023. Estimating soybean harvest index and nitrogen concentrations of grain and residue using globally available data. Netherlands: Wageningen University. (Thesis – MSc).

Luo, X., Ma, C., Yue, Y., Hu, K., Li, Y., Duan, Z., Wu, M., Tu, J., Shen, J., Yi, B., & Fu, T. 2015. Unravelling the complex trait of harvest index in rapeseed (Brassica napus L.) with association mapping. *BMC Genomics*, 16(1). https://doi.org/10.1186/s12864-015-1607-0

Luo, Z., Wang, E., & Sun, O.J. 2010. Can no-tillage stimulate carbon sequestration in agricultural soils? A meta-analysis of paired experiments. *Agriculture, Ecosystems and Environment*, 139(1–2):224–231. https://doi.org/10.1016/j.agee.2010.08.006

Maas, E.D.V.L. & Lal, R.A. 2023. A case study of the RothC soil carbon model with potential evapotranspiration and remote sensing model inputs. *Remote Sensing Applications: Society and Environment*, 29:100876. https://doi.org/10.1016/j.rsase.2022.100876

Malley, D.F., Martin, P.D., McClintock, L.M., Yesmin, L., Eilers, R.G., & Haluschak, P. 2000. Feasibility of analysing archived Canadian prairie agricultural soils by near infrared reflectance spectroscopy. In ‘‘Near Infrared Spectroscopy: Proceedings of the 9th International Conference’’ (A. M. C. Davies & R. Giangiacomo, Eds.), 579–585. NIR Publications, Chichester, UK.

Martin, P.D., Malley, D.F., Manning, G., & Fuller, L. 2002. Determination of soil organic carbon and nitrogen at the field level using near-infrared spectroscopy. *Canadian Journal of Soil Science*, 82:413–422. https://doi.org/10.4141/S01-054

McBratney, A.B., Mendonça Santos, M.L., & Minasny, B. 2003. On digital soil mapping. *Geoderma*, 117:3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

McCarty, G.W., Reeves, J.B., Yost, R., Doraiswamy, P.C., & Doumbia, M. 2010. Evaluation of methods for measuring soil organic carbon in West African soils. *African Journal of Agriculture Research*, 5(16):2169–2177. https://www.researchgate.net/publication/266244106_Evaluation_of_methods_for_measuring_soil_organic_carbon_in_West_African_soils

McDowell, M.L., Bruland, G.L., Deenik, J.L., Grunwald, S., & Knox, N.M. 2012. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. *Geoderma*, 189–190:312–320. https://doi.org/10.1016/j.geoderma.2012.06.009

Meersmans, J., Van Wesemael, B., De Ridder, F., Fallas Dotti, M., De Baets, S., & Van Molle, M. 2009. Changes in organic carbon distribution with depth in agricultural soils in northern Belgium, 1960–2006. *Global Change Biology*, 15:2739–2750. https://doi.org/10.1111/j.1365-2486.2009.01855.x

Merante, P., Dibari, C., Ferrise, R., Sánchez, B., Iglesias, A., Lesschen, J. P., Kuikman, P., Yeluripati, J., Smith, P., & Bindi, M. 2017. Adopting soil organic carbon management practices in soils of varying quality: implications and perspectives in Europe. Italy: University of Florence.

Metzger, K., Zhang, C., Ward, M., & Daly, K. 2020. Mid-infrared spectroscopy as an alternative to laboratory extraction for the determination of lime requirement in tillage soils. *Geoderma*, 364(114171). https://doi.org/10.1016/j.geoderma.2020.114171

Mevik, B.H., & Wehrens, R. 2024. Introduction to the pls Package. Help section of the “pls” package of RStudio software. (Section 7):1–23. https://cran.r-project.org/web/packages/pls/vignettes/pls-manual.pdf

Mills, A. 2025. Conversation regarding RothC (Personal communication).

Minasny, B. & McBratney, A.B. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences*, 32(9):1378–1388. https://doi.org/10.1016/j.cageo.2005.12.009

Minasny, B., McBratney, A.B., Malone, B.P., & Ichsani, W. 2013. Digital mapping of soil carbon. *Advances in Agronomy*, 118:1–47. https://doi.org/10.1016/B978-0-12-405942-9.00001-3

Mokolobate, M.C., Scholtz, M.M., Neser, F.W.C., & Buchanan, G. 2015. Approximation of forage demands for lactating beef cows of different body weights and frame sizes using the Large Stock Unit. *Applied Animal Husbandry & Rural Development*, 8:34–38. https://www.sasas.co.za/AAH&RD/approximation-of-forage-demands-for-lactating-beef-cows-of-different-body-weights-and-frame-sizes-using-the-large-stock-unit/

Morais, T.G., Teixeira, R.F.M., Domingos, T. 2019. Detailed global modelling of soil organic carbon in cropland, grassland and forest soils. *PLoS ONE*, 14(9). https://doi.org/10.1371/journal.pone.0222604

Morra, M.J., Hall, M.H., & Freeborn, L.L. 1991. Carbon and nitrogen analysis of soil fractions using near-infrared reflectance spectroscopy. *Soil Science Society of America Journal*, 55:288–291. https://doi.org/10.2136/sssaj1991.03615995005500010051x

Mouazen, A.M., & Ramon, H. 2006. Development of on-line measurement system of bulk density based on on-line measured draught, depth and soil moisture content. *Soil and Tillage Research*, 86:218–229.

Mousavi, S.R., Sarmadian, F., Dehghani, S., Sadikhani, M.R., & Taati, A. 2017. Evaluating inverse distance weighting and kriging methods in estimation of some physical and chemical properties of soil in Qazvin Plain. *Eurasian Journal of Soil Science*, 6(4):327–336. https://dergipark.org.tr/tr/download/article-file/310579

Müller, M.J. 1982. Selected climatic data for a global set of standard stations for vegetation science: The Hague: Dr. W. Junk, Publishers.

Nair, P.K.R. 2012. Carbon sequestration studies in agroforestry systems: a reality-check. *Agroforest Systems*, 86:243-253. https://doi.org/10.1007/s10457-011-9434

Negassa, M.K., Haile, M., Feyisa, G.L., Wogi, L., & Merga, F. 2023. Modeling and mapping spatial distribution of baseline soil organic carbon stock, a case of West Hararghe, Oromia Regional State, Eastern Ethiopia. *Geology, Ecology, and Landscapes*:1–6. https://doi.org/10.1080/24749508.2023.2167632

Nelson, D.W., L.E., & Sommers. 1996. Total carbon, organic carbon, and organic matter. In: Methods of soil analysis. Part 3 Chemical methods. (D.L. Sparks, editor) 961–1010. SSSA and ASA, Madison, WI.

Nenkam, A.M., Wadoux, A.M.J.C., Minasny, B., Silatsa, F.B.T., Yemefack, M., Ugbaje, S.U., Akpa, S., van Zijl, G., Bouasria, A., Bouslihim, Y., Chabala, L.M., Ali, A., & McBratney, A.B. 2024. Applications and challenges of digital soil mapping in Africa. *Geoderma*, 449(117007) https://doi.org/10.1016/j.geoderma.2024.117007

Ng, W., Minasny, B., de Sousa Mendes, W., & Melo Demattê, J.A. 2020. The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data. *Soil*, 6(2):565–578. https://doi.org/10.5194/soil-6-565-2020

Nicol, A.M., & Brookes, I.M. 2007. Pasture and Supplements for Grazing animals. New Zealand Society of Animal Production. https://www.campusbooks.nz/pasture-and-supplements-for-grazing-animals-9780473052362

Nocita, M. 2009. *Soil spectroscopy as a tool to assess organic carbon, iron oxides, and clay content in the Subtropical Thicket Biome of the Eastern Cape province of South Africa*. Wageningen University and Research Centre: GIRS-2009-13.

Nocita, M., Stevens, A., Noon, C., & van Wesemael, B. 2013. Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. *Geoderma*, 199:37–42. https://doi.org/10.1016/j.geoderma.2012.07.020

Nozari, S., Pahlavan-Rad, M.R., Brungard, C., Heung, B., & Borůvka, L. 2024. *Soil and Water Research*, 19(1):32–49. https://doi.org/10.17221/119/2023-SWR

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

Odebiri, O., Mutanga, O., Odindi, J., & Naicker, R. 2023. Mapping soil organic carbon distribution across South Africa's major biomes using remote sensing-topo-climatic covariates and Concrete AutoencoderDeep neural networks. *Science of the Total Environment*, 865(161150). https://doi.org/10.1016/j.scitotenv.2022.161150

Odebiri, O., Mutanga, O., Odindi, J., Slotow, R., Mafongoya, P., Lottering, R., Naicker, R., Matongera, T.N., & Mngadi, M. 2024. Mapping Sub-surface Distribution of Soil Organic Carbon Stocks in South Africa's Arid and Semi-Arid Landscapes: Implications for Land Management and Climate Change Mitigation. *Geoderma Regional*, 37:e00817. https://www.sciencedirect.com/science/article/pii/S2352009424000646

Olson, K.R., & Al-Kaisi, M.M. The importance of soil sampling depth for accurate account of soil organic carbon sequestration, storage, retention and loss. *Catena*, 125:33–37. https://www.sciencedirect.com/science/article/pii/S034181621400280X

Orizon Agriculture. 2024. What are Carbon Credits? https://orizonagriculture.com/faqs/ (accessed 1 November 2024).

Orr, R., McBeath, A.V., Dieleman, W.I.J., Bird, M.I., & Nelson, P.N. 2017. Estimating organic carbon content of soil in Papua New Guinea using infrared spectroscopy. *Soil Research*, 55(8):735–742. https://doi.org/10.1071/SR16227

Palladino, M., Romano, N., Pasolli, E., & Nasta, P. 2022. Developing pedotransfer functions for predicting soil bulk density in Campania. *Geoderma*, 412(115726):1–13. https://doi.org/10.1016/j.geoderma.2022.115726

Parwada, C., & van Tol, J. 2016. The nature of soil erosion and possible conservation strategies in Ntabelanga area, Eastern Cape Province, South Africa. *Acta Agriculturae Scandinavica Section B: Soil and Plant Science*, 66(6): 544–552. https://doi.org/10.1080/09064710.2016.1188979

Patel, S., Sawyer, J.E., Lundvall, J.P., & Hall, J. 2015. Root and shoot biomass and nutrient composition in a winter rye cover crop. North Central Extension-Industry Soil Fertility Conference, Volume 31. Des Moines, IA.

Paustian, K., Larson, E., Kent, J., Marx, E., Swan, A. 2019. Soil C Sequestration as a Biological Negative Emission Strategy. *Frontiers in Climate*, 1(8):1–11. https://doi.org/10.3389/fclim.2019.00008

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G.P., Smith, P. 2016. Climate-smart soils. *Nature*, 532:49–57. https://www.nature.com/articles/nature17174

Pebesma, E. 2025. gstat: Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation. R package version 2.1-3. https://cran.r-project.org/web/packages/gstat/index.html

Peng, S., Bao, N., Wang, S., Gholizadeh, A., Saberioon, M., & Peng, Y. 2024. Mapping vertical distribution of SOC and TN in reclaimed mine soils using point and imaging spectroscopy. *Ecological Indicators*, *158*. https://doi.org/10.1016/j.ecolind.2023.111437

Peng, X., Shi, T., Song, A., Chen, Y. & Gao, W. 2014. Estimating soil organic carbon using VIS/NIR Spectroscopy with SVMR and SPA methods. *Remote Sensing*, 6:2699–2717. https://doi.org/10.3390/rs6042699

Peng, Y., Xiong, X., Adhikari, K., Knadel, M., Grunwald, S., & Greve, M.H. 2015. Modeling soil organic carbon at regional scale by combining multi-spectral images with laboratory spectra. *PLoS ONE*, 10(11). https://pubmed.ncbi.nlm.nih.gov/26555071/

Poeplau, C., Vos, C. & Don, A. 2017. Soil organic carbon stocks are systematically overestimated by misuse of the parameters bulk density and rock fragment content. *Soil*, 3:61–66. https://doi.org/10.5194/soil-3-61-2017

Pouladi, N., Gholizadeh, A., Khosravi, V., & Borůvka, L. 2023. Digital mapping of soil organic carbon using remote sensing data: A systematic review. *Catena*, 232(107409). https://doi.org/10.1016/j.catena.2023.107409

Pouladi, N., Møller, A.B., Tabatabai, S., & Greve, M.H. 2019. Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging. *Geoderma*, 342:85–92. https://doi.org/10.1016/j.geoderma.2019.02.019

Pribyl, D.W. 2010. A critical review of the conventional SOC to SOM conversion factor. *Geoderma*, 156:75–83. https://www.sciencedirect.com/science/article/pii/S0016706110000388

Quinlan, J.R. 1992. Learning with continuous classes. Proceedings of the 5th Australian Joint Conference On Artificial Intelligence: 343–348.

R Core Team., 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria.

Rhodes, C.J. 2017. The imperative of regenerative agriculture. *Science Progress*, 100(1):80–129. https://doi.org/10.3184/003685017X14876775256165

Roper, W.R., Robarge, W.P., Osmond, D.L., & Heitman, J.L. 2019. Comparing Four Methods of Measuring Soil Organic Matter in North Carolina Soils. *Soil Science Society of American Journal*, 83:466–474. https://acsess.onlinelibrary.wiley.com/doi/full/10.2136/sssaj2018.03.0105

Rosenblatt, F. 1958. The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain. *Psychological Review*, 65 (6): 386–408. https://www.ling.upenn.edu/courses/cogs501/Rosenblatt1958.pdf

Rossel, R.A.V. & Behrens, T. 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158(1–2):46–54. https://doi.org/10.1016/j.geoderma.2009.12.025

Roudier, P., Brugnard, C., Beaudette, D., Louis, B., Daust, K., & Clifford, D. 2021. Conditioned Latin Hypercube Sampling – package 'clhs'. R package version 0.9.0. https://cran.r-project.org/web/packages/clhs/clhs.pdf

Roy, S., & Kashem, M.A., 2014. Effects of Organic Manures in Changes of Some Soil Properties at Different Incubation Periods. *Open Journal of Soil Science*, 4(3):81–86. https://doi.org/10.4236/ojss.2014.43011

Safanelli, J. L., Hengl, T., Parente, L., Minarik, R., Bloom, D. E., Todd-Brown, K., Gholizadeh, A., Mendes, W. de S., & Sanderman, J. 2023. Open Soil Spectral Library (OSSL): Building reproducible soil calibration models through open development and community engagement. *BioRxiv*, 2023.12.16.572011. https://doi.org/10.1101/2023.12.16.572011

Santacruz, A. 2011. Optimización evolutiva de redes espaciales de muestreo: Una aplicación de los algoritmos genéticos y la geoestadística al monitoreo del carbono orgánico del suelo. Editorial Académica Española, Saarbrücken.

Sarker, I.H. 2021. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2(160):1–21. https://doi.org/10.1007/s42979-021-00592-x

Sato, J.H., de Figueiredo, C.C., Marchão, R.L., Madari, B.E., Benedito, L.E.C., Busato, J.G., & de Souza, D.M. Methods of soil organic carbon determination in Brazilian savannah soils. *Scientia Agricola*, 71(4):302–308. https://www.scielo.br/j/sa/a/HwNzvFLPJ7xZSTBnNfQg4LQ/?lang=en

SAWS (South African Weather Service). 2024. Climate data. https://www.weathersa.co.za/home/subscribe/(accessed 1 October 2024).

Scherstjanoi, M., & Dechow, R. 2024. Extended R documentation of function sorcering() from package 'sorcering'. Thünen Institute of Climate-Smart Agriculture, Braunschweig, Germany.

Schulze, R.E. 2007. South African Atlas of Climatology and Agrohydrology, WRC Report 1489/1/06, Section 1.3. Water Research Commission, Pretoria, South Africa.

Schulze, R.E., & Schutte, S. 2020. Mapping soil organic carbon at a terrain unit resolution across South Africa. *Geoderma*, 373:114447. https://doi.org/10.1016/j.geoderma.2020.114447

Seboko, R.K., Kotze, E., van Tol, J., & van Zijl, G.M. 2021. Characterization of Soil Carbon Stocks in the City of Johannesburg. *Land*, 10(83):1–12. https://doi.org/10.3390/land10010083

Seidel, M., Hutengs, C., Ludwig, B., Thiele-Bruhn, S., & Vohland, M. 2019. Strategies for the efficient estimation of soil organic carbon at the field scale with vis-NIR spectroscopy: Spectral libraries and spiking vs. local calibrations. *Geoderma*, 354(113856). https://doi.org/10.1016/j.geoderma.2019.07.014

Senapati, N., Hulugalle, N.R., Smith, P., Wilson, B.R., Yeluripati, J. B., Daniel, H., Ghosh, S., & Lockwood, P. 2014. Modelling soil organic carbon storage with RothC in irrigated Vertisols under cotton cropping systems in the sub-tropics. *Soil and Tillage Research*, 143:38–49. https://doi.org/10.1016/j.still.2014.05.009

Setia, R., Smith, P., Marschner, P., Baldock, J., Chittleborough, D., & Smith, J. 2011. Introducing a decomposition rate modifier in the rothamsted carbon model to predict soil organic carbon stocks in saline soils. *Environmental Science and Technology*, 45(15):6396–6403. https://doi.org/10.1021/es200515d

Shahrayini, E., Noroozi, A.A., & Eghbal, M.K. 2020. Prediction of Soil Properties by Visible and Near-Infrared Reflectance Spectroscopy. *Eurasian Soil Science*, 53(12):1760–1772. https://link.springer.com/article/10.1134/S1064229320120108

Sharififar, A., Minasny, B., Arrouays, D., Boulonne, L., Chevallier, T., van Deventer, P., Field, D.J., Gomez, C., Jang, H.J., Jeon, S.H., Koch, J., McBratney, A.B., Malone, B.P., Marchant, B.P., Martin, M.P., Monger, C., Munera-Echeverri, J.L., Padarian, J., Pfeiffer, M., Richer-de-Forges, A.C., Saby, N.P.A., Singh, K., Song, X.D., Zamanian, K., Zhang, G.L., & van Zijl, G.M. 2023. Soil inorganic carbon, the other and equally important soil carbon pool: Distribution, controlling factors, and the impact of climate change. *Advances in Agronomy* (178):165–231. https://www.sciencedirect.com/science/article/pii/S0065211322001122

Sharififar, A., Singh, K., Jones, E., Ginting, F.I., & Minasny, B. 2019. Evaluating a low-cost portable NIR spectrometer for the prediction of soil organic and total carbon using different calibration models. *Soil Use and Management*, 35(4):607–616. https://doi.org/10.1111/sum.12537

Sharma, M., Kaushal, R., Kaushik, P., & Ramakrishna, S. 2021. Carbon Farming: Prospects and Challenges. *Sustainabilit*y, 13(1):1–15. https://doi.org/10.3390/su131911122

Sherpa, S.R., Wolfe, D.W., & van Es, H.M. 2016. Sampling and Data Analysis Optimization for Estimating Soil Organic Carbon Stocks in Agroecosystems. *Soil Science Society of America Journal*, 80:1377–1392. https://doi.org/10.2136/sssaj2016.04.0113

Shi, T., Hu, X., Guo, L., Su, F., Tu, W., Hu, Z., Liu, H., Yang, C., Wang, J., Zhang, J., & Wu, G. 2021. Digital mapping of zinc in urban topsoil using multisource geospatial data and random forest. *Science of the Total Environment*, 792(148455). https://doi.org/10.1016/j.scitotenv.2021.148455

Shirato, Y., & Taniyama, I. 2003. Testing the suitability of the Rothamsted Carbon model for long-term experiments on Japanese non-volcanic upland soils. *Soil Science and Plant Nutrition*, 49(6):921–925. https://doi.org/10.1080/00380 768.2003.10410357

Shoch, D., & Swails, E. 2020. Verified Carbon Standard Methodology (VM0042): Methodology for improved agriculture land management.

Shoumik, B.A.A., & Khan, M.Z. 2023. Spatio-temporal dynamics of soil organic carbon and total nitrogen: evidenced from 2000 to 2020 in a mixed ecosystem. *Environmental Earth Sciences*, 82(3):84. https://doi.org/10.1007/s12665-023-10756-y

Sierra, C.A., Müller, M., & Trumbore, S.E. 2012. Models of soil organic matter decomposition: The SoilR package, version 1.0. *Geoscientific Model Development*, 5(4):1045–1060. https://doi.org/10.5194/gmd-5-1045-2012

Singh, B.P., Setia, R., Wiesmeier, M., & Kunhikrishnan, A. 2018. Agricultural management practices and soil organic carbon storage. In B. Singh (ed.), *Soil Carbon Storage: Modulators, Mechanisms and Modeling*. Academic Press. https://doi.org/10.1016/B978-0-12-812766-7.00007-X

Singha, C., Swain, K.C., Sahoo, S., & Govind, A. 2023. Prediction of soil nutrients through PLSR and SVMR models by VIs-NIR reflectance spectroscopy. *Egyptian Journal of Remote Sensing and Space Science*, 26(4). https://doi.org/10.1016/j.ejrs.2023.10.005

Skjemstad, J.O., Spouncer, L.R., Cowie, B., & Swift, R.S. 2004. Calibration of the Rothamsted organic carbon turnover model (RothC ver. 26.3), using measurable soil organic carbon pools. *Australian Journal of Soil Research*, 42(1):79–88. https://doi.org/10.1071/SR03013

Slepetiene, A., Slepetys, J., & Liaudanskiene, I. 2008. Standard and modified methods for soil organic carbon determination in agricultural soils. *Agronomy Research*, 6(2):543–554. https://agronomy.emu.ee/vol062/p6211.pdf

Smith, P., Soussana, J.F., Angers, D., Schipper, L., Chenu, C., Rasse, D.P., Batjes, N.H., van Egmond, F., McNeill, S., Kuhnert, M., Arias-Navarro, C., Olesen, J.E., Chirinda, N., Fornara, D., Wollenberg, E., Álvaro-Fuentes, J., Sanz-Cobena, A., & Klumpp, K. 2020. How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. *Global Change Biology*, 26(1):219–241. https://doi.org/10.1111/gcb.14815

Soil Classification Working Group. 2018. Soil Classification: A Natural and Anthropogenic System for South Africa. 3rd ed. South Africa, PTA: Agriculture Research Council Institute for Soil, Climate and Water.

Sommers, L.E., Gilmour, C.M., Wilding, R.E. & Beck, S.M. 1981. Effect of water potential and decomposition processes in soils. Water Potential Relations, in: *Soil Microbiology,* 97-117. Special publication of the Soil Science Society of America No. 9, Madison.

Soriano-Disla, J.M., Janik, L.J., Viscarra-Rossel, R.A., Macdonald, L.M., McLaughlin, M.J. 2014. The performance of visible, near-, and midinfrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Applied Spectroscopy Reviews*, 49:139–186. https://doi.org/10.1080/05704928.2013.811081

Stenberg, B. 2010. Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. *Geoderma*, 158(1–2):15–22. https://doi.org/10.1016/j.geoderma.2010.04.008

Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., & Wetterlind, J. 2010. Visible and Near Infrared Spectroscopy in Soil Science. *Advances in Agronomy*, 107:163–215. https://doi.org/10.1016/S0065-2113(10)07005-7

Stevens, A., Nocita, M., Tóth, G., Montanarella, L., & van Wesemael, B. 2013. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. *PLoS ONE*, 8(6). https://doi.org/10.1371/journal.pone.0066409

Stockmann, U., Padarian, J., McBratney, A., Minasny, B., de Brogniez, D., Montanarella, L., Hong, S.Y., Rawlins, B.G., & Field, D.J. 2015. Global soil organic carbon assessment. *Global Food Security*, 6:9–16. https://doi.org/10.1016/j.gfs.2015.07.001

Subedi, S., Srivastava, A., Sharma, M., & Shah, S. 2018. Effect of organic and inorganic nutrient sources on growth, yield and quality of radish (Raphanus sativus L.) varieties in Chitwan, Nepal. *Journal of Agriculture*, 16(1):61–69. https://doi.org/10.3329/sja.v16i1.37423

Taghipour, K., Heydari, M., Kooch, Y., Fathizad, H., Hueng, B., & Taghizadeh-Mehrjardi, R. 2022. Assessing changes in soil quality between protected and degraded forests using digital soil mapping for semiarid oak forests, Iran. *Catena*, 213(106204):1–13. https://doi.org/10.1016/j.catena.2022.106204

Taghizadeh-Mehrjardi, R., Fathizad, H., Ali Hakimzadeh Ardakani, M., Sodaiezadeh, H., Kerry, R., Heung, B., & Scholten, T. 2021. Spatio-temporal analysis of heavy metals in arid soils at the catchment scale using digital soil assessment and a random forest model. *Remote Sensing*, 13(9):1698. https://doi.org/10.3390/rs13091698

Tekin, Y., Tumsava, Z., & Mouazen, A.M. 2012. Effect of moisture content on prediction of organic carbon and pH using visible and near-infrared spectroscopy. *Soil Science Society of America Journal*, 76:188–198. https://doi.org/10.2136/sssaj2011.0021

Tekin, Y., Tümsavas, Z., & Mouazen, A.M. 2014. Comparação entre rede neural artificial e regressão por mínimos quadrados parciais na predição do carbono orgânico do solo e do ph, em diferentes níveis de umidade do solo utilizando espectroscopia no vis-Ivp. *Revista Brasileira de Ciencia Do Solo*, 38(6):1794–1804. https://doi.org/10.1590/S0100-06832014000600014

Terhoeven-Urselmans, T., Vagen, T.G., Spaargaren, O. & Shepherd, K.D. 2010. Prediction of Soil Fertility Properties from a Globally Distributed Soil Mid-Infrared Spectral Library. *Soil Science Society of America Journal*, 74(5): 1792–1799. https://doi.org/DOI 10.2136/sssaj2009.0218

Teutscherova, N., Vazquez, Sotelo, M.E., & Villegas, D.M. Intensive short-duration rotational grazing is associated with improved soil quality within one year after establishment in Colombia. *Applied Soil Ecology*, 159:103835. https://www.researchgate.net/publication/346908176_Intensive_short-duration_rotational_grazing_is_associated_with_improved_soil_quality_within_one_year_after_establishment_in_Colombia

Thompson, M. 2008. CHNS Elemental Analysers. The Royal Society of Chemistry. ISSN 1757–5958.

Thornthwaite, C. 1948. An Approach toward a Rational Classification of Climate. *Geographical Review*, 38:55–94. https://www.jstor.org/stable/210739

TREES Consulting. 2020. Gold Standard for the global goals: Soil Organic Carbon Framework Methodology.

Tziachris, P., Metaxa, E., Papadopoulos, F., & Papadopoulou, M. 2017. Spatial Modelling and Prediction Assessment of Soil Iron Using Kriging Interpolation with pH as Auxiliary Information. *International Journal of Geo-information*, 6(283):1–16. https://doi.org/10.3390/ijgi6090283

Unkovich, M., Baldock, J., & Forbes, M. 2010. Variability in Harvest Index of Grain Crops and Potential Significance for Carbon Accounting: Examples from Australian Agriculture. *Advances in Agronomy*, 105:173–219. https://doi.org/10.1016/S0065-2113(10)05005-4

USDA (United States Department of Agriculture). 2018. Soil bulk density/moisture/aeration. *Nutrition & Food Science*, 42(4):11–14. doi: 10.1108/nfs.2012.01742daa.005

USGS (United States Geological Survey Earth Explorer). Date of access: 2023. https://earthexplorer.usgs.gov/

Van Zijl, G.M., van Tol, J., Smit, I.E., Sehlapelo, M., Kock, A., Cloete, W.H., Faul, J., Le Roux, J., Riddell, E., Jacobs, A., Verwey, E., Cooke, V., de Clercq, W., Manyevere, A., & Lorentz, S. 2024. Towards a Hydrological soil map of South Africa (HYDROSOIL) – Developing a protocol and showcasing its uses. Water Research Commission Report (Project No. C2020/2021-00455).

Vapnik, V.N. 2000. The nature of statistical learning theory. In: Jordan, M., Lauritzen, S.L., Lawless, J.F., Nair, V. (Eds.). Statistics for Engineering and Information Science. Springer Verlag, New York, pp. 1-314.

Venter, Z.S., Hawkins, H-J., Cramer, M.D., & Mills, A.J. 2021. Mapping soil organic carbon stocks and trends with satellite-driven high resolution maps over South Africa. *Science of the Total Environment*, 771(145384). https://doi.org/10.1016/j.scitotenv.2021.145384

Viscarra Rossel, R.A. & Behrens, T. 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, 158:46–54. https://doi.org/10.1016/j.geoderma.2009.12.025

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

Visconti, F., Jiménez, M.G., de Paz, J.M. 2022. How do the chemical characteristics of organic matter explain differences among its determinations in calcareous soils? *Geoderma*, 406(115454). https://www.sciencedirect.com/science/article/pii/S0016706121005346

Vohland, M., Besold, J., Hill, J. & Fründ, H. 2011. Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near-infrared spectroscopy. *Geoderma*, 166:198–205. https://doi.org/10.1016/j.geoderma.2011.08.001

Wadoux, A., Hevelink, G., Lark, M., Lagacherie, P., Bouma, J., Mulder, V., Libohova, Z., Yang, L., McBratney, A. 2021. Ten challenges for the future of pedometrics. *Geoderma*, 401:115–155. https://doi.org/10.1016/j.geoderma.2021.115155

Wadoux, A., Malone, B., Minasny, B., Fajardo, M. & McBratney, A. 2021. Soil Spectral Inference with R: Analysing Digital Soil Spectra Using the R Programming Environment. AG: Springer International Publishing.

Walkley, A. 1947. A Critical Examination of a Rapid Method for Determining Organic Carbon in Soils: Effect of Variations in Digestion Conditions and of Inorganic Soil Constituents. *Soil Science*, 63:251–264.
http://dx.doi.org/10.1097/00010694-194704000-0000

Walkley, A., & Black, I.A. 1934. An examination of the Degtjareff method for determining organic carbon in soils: Effect of variations in digestion conditions and of inorganic soil constituents. *Soil Science*, 63:251–263. https://doi.org/10.1097/00010694-194704000-00001

Wang, X., Wang, J., & Zhang, J. 2012. Comparisons of Three Methods for Organic and Inorganic Carbon in Calcareous Soils of Northwestern China. *PLoS ONE*, 7(8):e44334. doi:10.1371/journal.pone.0044334

Wetterlind, J., Stenberg, B., & Jonsson, A. 2008. Near infrared reflectance spectroscopy compared with soil clay and organic matter content for estimating within-field variation in N uptake in cereals. *Plant and Soil*, 302:317–327. https://link.springer.com/article/10.1007/s11104-007-9489-9

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer International Publishing. https://ggplot2.tidyverse.org

Wiese, L., Ros, I., Rozanov, A., Boshoff, A., de Clercq, W., & Seifert, T. 2016. An approach to soil carbon accounting and mapping using vertical distribution functions for known soil types. *Geoderma*, 263:264–273. https://doi.org/10.1016/j.geoderma.

Williams, P.C., & Thompson, B.N. 1978. Influence of whole meal granularity on analysis of HRS wheat for protein and moisture by near infrared reflectance spectroscopy (NRS). *Cereal Chemistry*, 55:1014–1037. https://www.cerealsgrains.org/publications/cc/backissues/1978/Documents/chem55_1014.pdf

Wills, S., Roecker, S., Williams, C., & Murphy, B. 2018. Soil sampling for soil health assessment. In: Reicosky, D (ed). Managing soil health for sustainable agriculture Volume 2: monitoring and management. London, Burleigh Dodds Science Publishing.

Wold, H. 1966. Estimation of principal components and related models by iterative least squares. In P.R. Krishnaiaah (Ed.). Multivariate Analysis. (pp.391-420) New York: Academic Press.

Wuest, S. 2014. Seasonal Variation in Soil Organic Carbon. Soil Science Society of American Journal, 78:1442–1447. https://www.ars.usda.gov/ARSUserFiles/6233/seasonalVariationInSoilOrganic.pdf

Xu, M., Chu, X., Fu, Y., Wang, C., & Wu, S. 2021. Improving the accuracy of soil organic carbon content prediction based on visible and near-infrared spectroscopy and machine learning. *Environmental Earth Sciences*, 80(8). https://doi.org/10.1007/s12665-021-09582-x

Xu, S., Zhao, Y., Wang, M. & Shi, X. 2018. Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis-NIR spectroscopy. *Geoderma*, 310:29–43. https://doi.org/10.1016/j.geoderma.2017.09.013

Yamamoto, J.K. 2005. Correcting the Smoothing Effect of Ordinary Kriging Estimates. *Mathematical Geology*, 37(1):69–94. https://doi.org/10.1007/s11004-005-8748-7

Yang, J., Xu, J.F., Zhang, X.L., Wu, C.Y., Lin, T., & Ying, Y.B. 2019. Deep learning for vibrational spectral analysis: Recent progress and a practical guide. *Analytica Chimica Acta*, 1081:6–17. https://doi.org/10.1016/j.aca.2019.06.012, 2019.

Yang, L., Guo, H., Yang, S., Hoshino, Y., Suzuki, S., Gao, D., & Cao, Y. 2019. Generation of a High-Precision Digital Elevation Model for Fields in Mountain Regions Using RTK-GPS. *International Journal of Automation Technology*, 13(5):671–678. https://www.jstage.jst.go.jp/article/ijat/13/5/13_671/_article/-char/en

Yanti, E.D., Mulyono, A., Djuwansah, M.R., Narulita, I., Putra, R.D., & Surinati, D. 2021. Development of pedotransfer functions for predicting soil bulk density: A case study in Indonesian small island. *Journal of Water and Land Development*, 51(10–12):181–187. https://creativecommons.org/licenses/by-nc-nd/3.0/)

Yao, X., Yu, K., Deng, Y., Liu, J., & Lai, Z. 2020. Spatial variability of soil organic carbon and total nitrogen in the hilly red soil region of Southern China. *Journal of Forestry Research*, 31(6):2385–2394. https://doi.org/10.1007/s11676-019-01014-8

Yerokun, O.A., Chikuta, S., & Mambwe, D. 2007. An Evaluation of Spectroscopic and Loss on Ignition Methods for Estimating Soil Organic Carbon in Zambian Soils. *International Journal of Agriculture Research*, 2(11):965–970. https://scialert.net/abstract/?doi=ijar.2007.965.970

Zahinda, F.M. 2020. *Detecting soil properties in agricultural lands using field spectroscopy and regression models*. Johannesburg: University of the Witwatersrand. (Thesis – MSc). https://wiredspace.wits.ac.za/items/6192d688-61e5-4156-a574-a7e170062bcb

Zayani, H., Fouad, Y., Michot, D., Kassouk, Z., Lili-Chabaane, Z., & Walter, C. 2023. Detecting the temporal trend of cultivated soil organic carbon content using visible near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 31(5):241–255. https://doi.org/10.1177/09670335231193113

Zhang, H., Hobbie, E.A., Feng, P., Zhou, Z., Niu, L., Duan, W., Hao, J., & Hu, K. 2021b. Responses of soil organic carbon and crop yields to 33-year mineral fertilizer and straw additions under different tillage systems. *Soil and Tillage Research*, 209(104943). https://www.sciencedirect.com/science/article/pii/S0167198721000131?via%3Dihub

Zhang, X.L., Xu, J.F., Yang, J., Chen, L., Zhou, H.B., Liu, X.J., Li, H.F., Lin, T., & Ying, Y. B. 2020. Understanding the learning mechanism of convolutional neural networks in spectral analysis, *Analytica Chimica Acta*, 1119:41–51, https://doi.org/10.1016/j.aca.2020.03.055

Zhang, Y., Lavallee, J.M., Robertson, A.D., Even, R., Ogle, S.M., Paustian, K. & Cotrufo, M.F. 2021. Simulating measurable ecosystem carbon and nitrogen dynamics with the mechanistically defined MEMS 2.0 model. *Biogeosciences*, 18:3147-3171. https://doi.org/10.5194/bg-18-3147-2021

Zhou, W., Han, G., Liu, M., & Li, X. 2019. Effects of soil pH and texture on soil carbon and nitrogen in soil profiles under different land uses in Mun River Basin, Northeast Thailand. *PeerJ* 7:e7880. https://peerj.com/articles/7880/

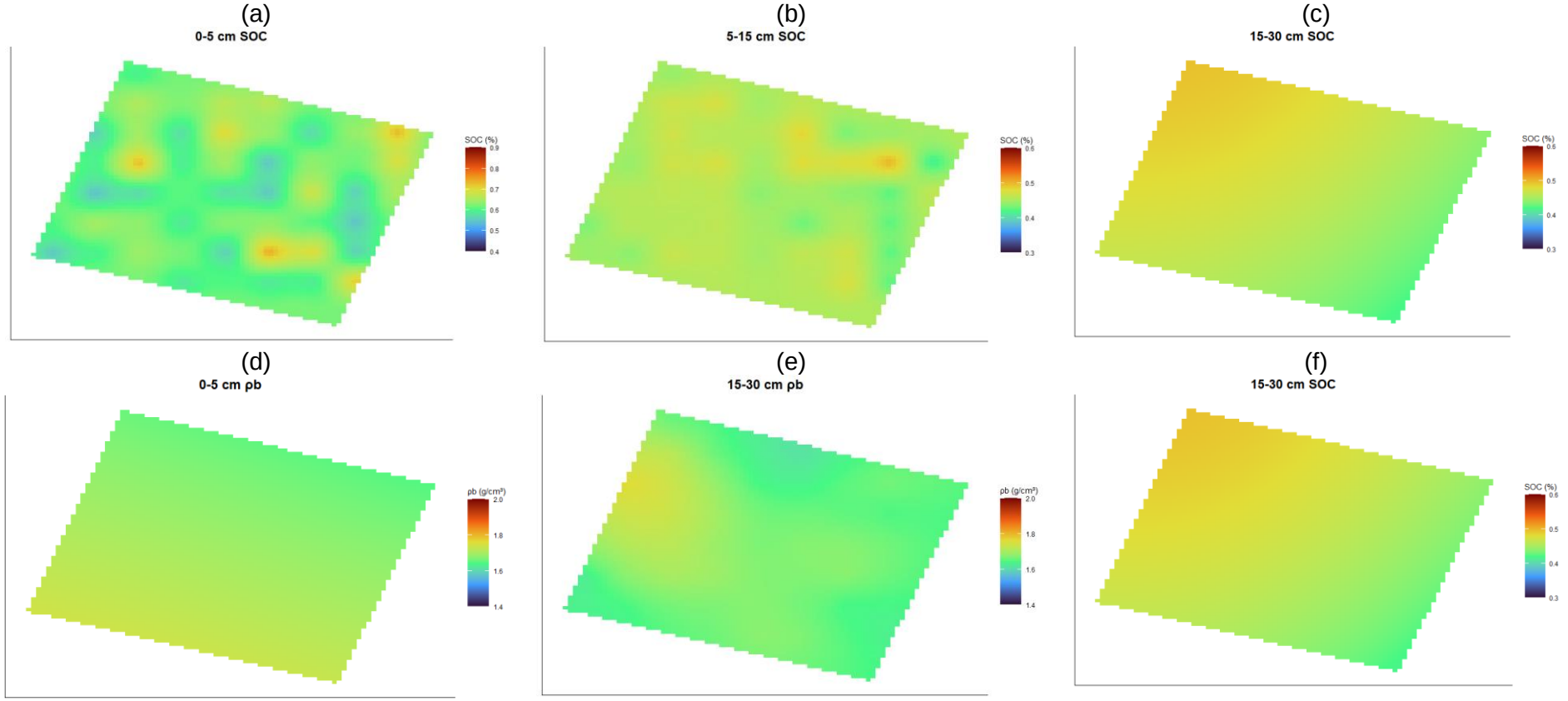
Zhu, C., Wei, Y., Zhu, F., Lu, W., Fang, Z., Li, Z., & Pan, J. 2022. Digital Mapping of Soil Organic Carbon Based on Machine Learning and Regression Kriging. *Sensors*, 22(8997). https://doi.org/10.3390/s22228

Zhu, Q., & Lin, H.S. 2010. Comparing Ordinary Kriging and Regression Kriging for Soil Properties in Contrasting Landscapes. *Pedosphere*, 20(5):594–606. https://doi.org/10.1016/S1002-0160(10)60049-5

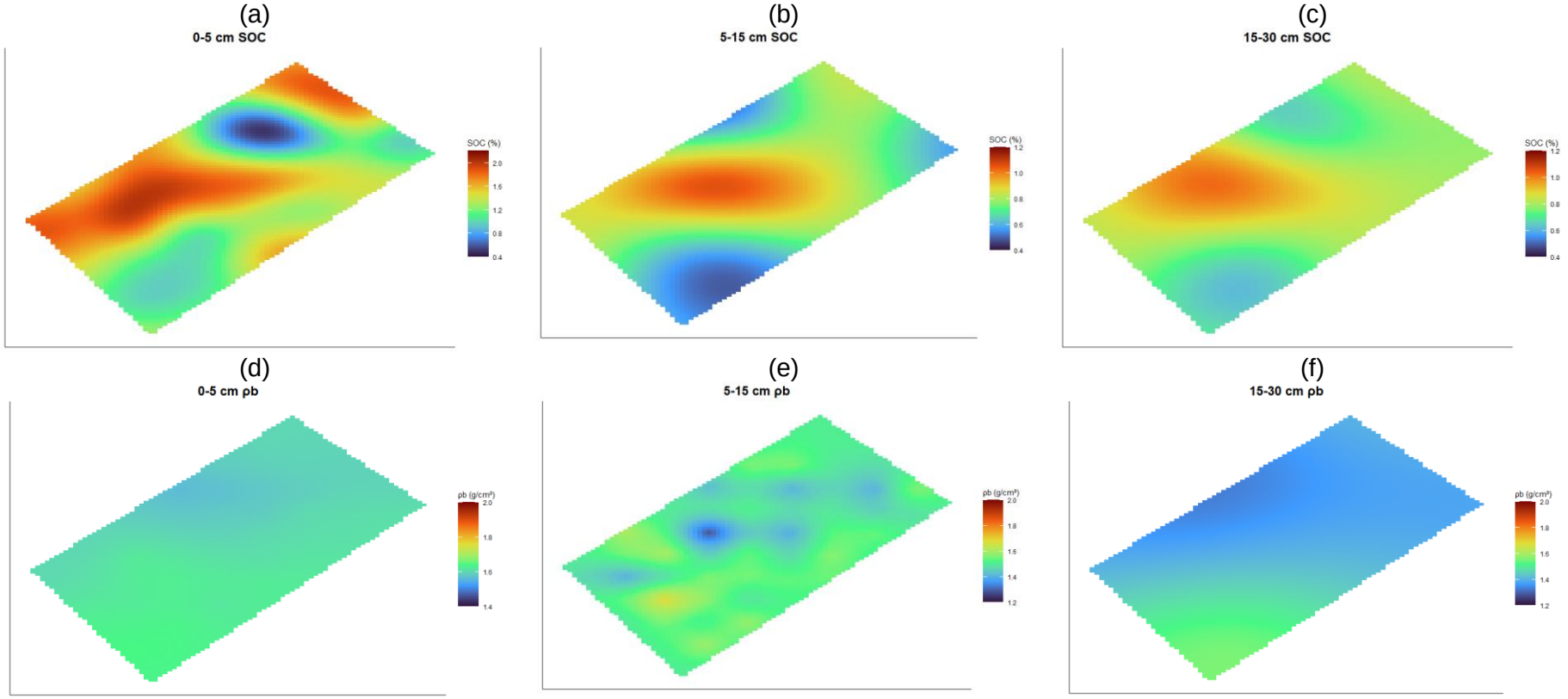
Žížala, D., Princ, T., Skála, J., Juřicová, A., Lukas, V., Bohovic, R., Zádorová, T., & Minařík, R. 2024. Soil sampling design matters – Enhancing the efficiency of digital soil mapping at the field scale. Geoderma Regional, 39:e00874. https://doi.org/10.1016/j.geodrs.2024.e00874

ANNEXURES

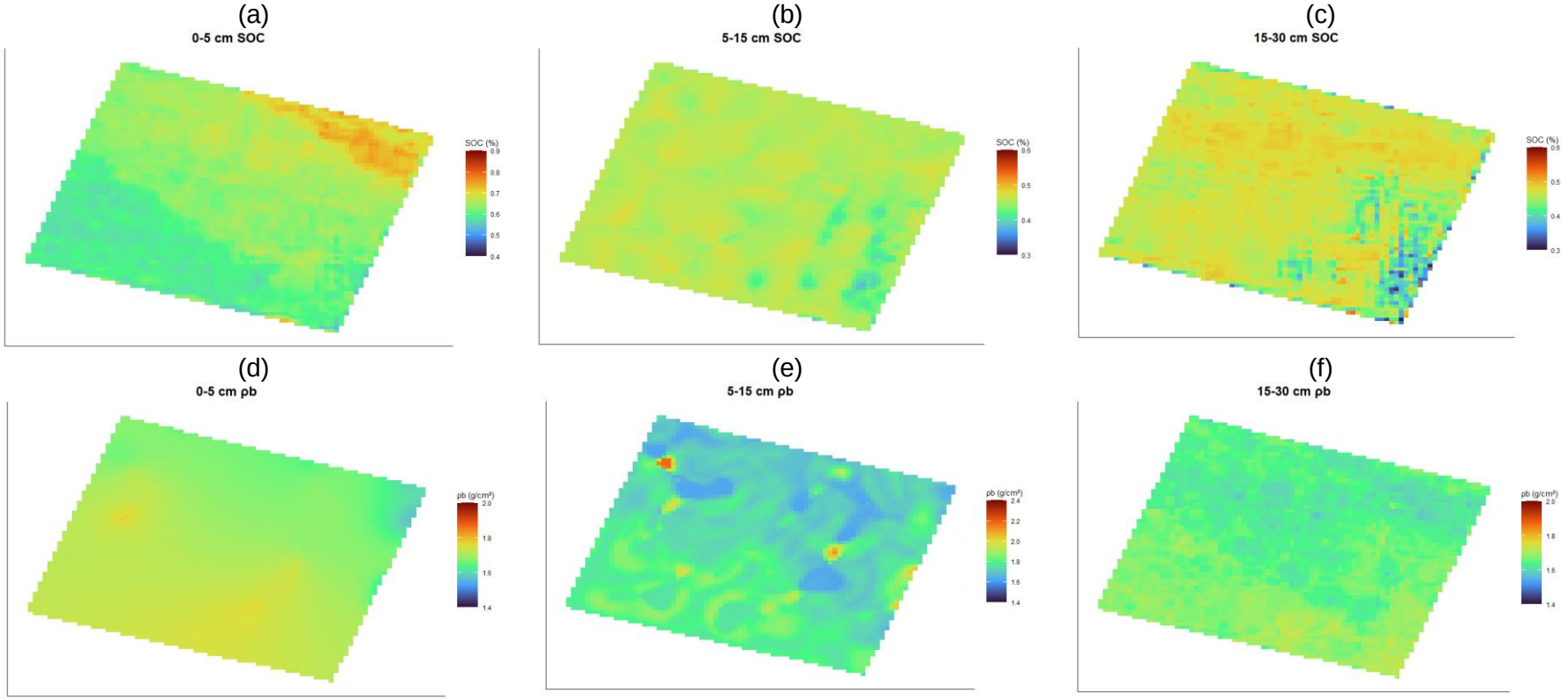
Appendix A: The maps created using ordinary kriging from the Ottosdal study site for (a) 0-5 cm soil organic carbon (SOC) content, (b) 5-15 cm soil organic carbon content (SOC) content, (c) 15-30 cm soil organic carbon content (SOC) content (d) 0-5 cm dry bulk density (ρb), (e) 5-15 cm dry bulk density (ρb) and (f) 15-30 cm dry bulk density (ρb).



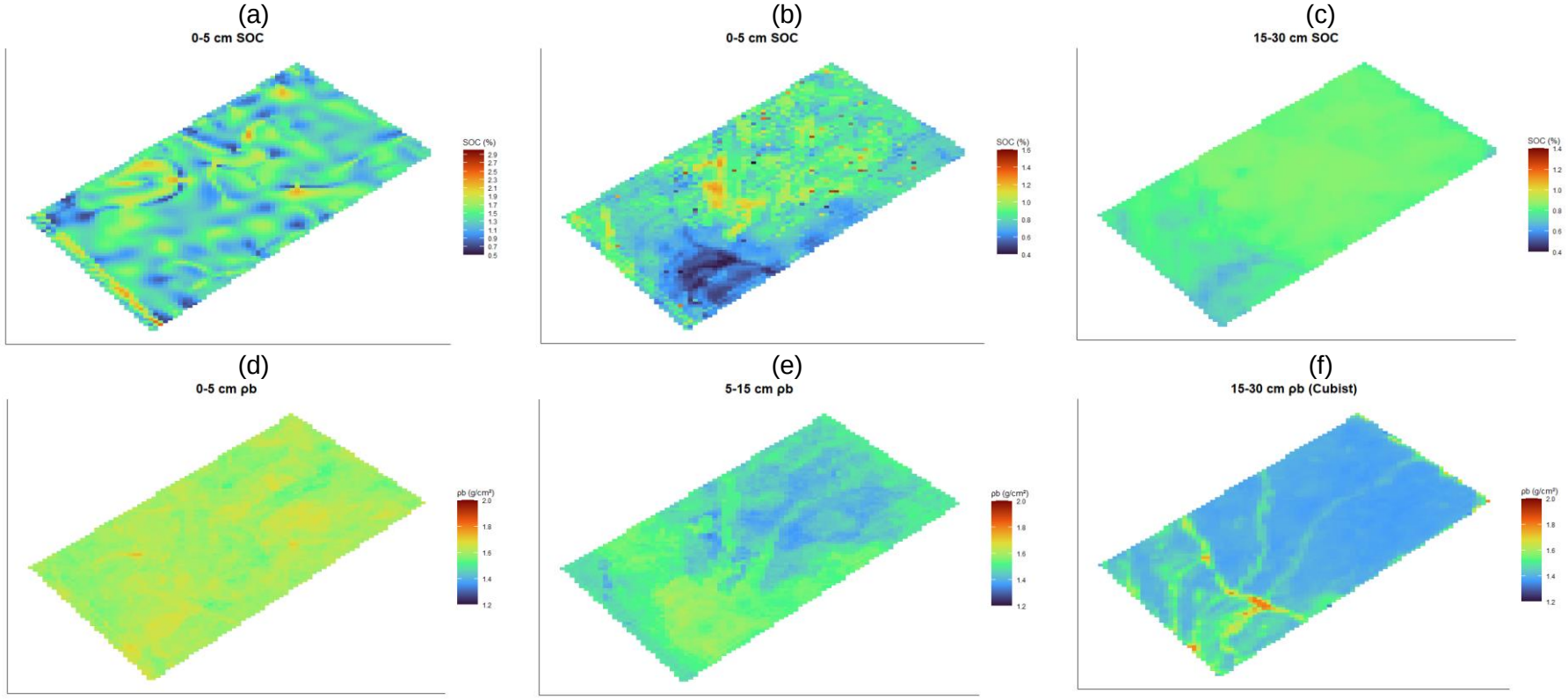
Appendix B: The maps created using ordinary kriging from the Vrede study site for (a) 0-5 cm soil organic carbon (SOC) content, (b) 5-15 cm soil organic carbon content (SOC) content, (c) 15-30 cm soil organic carbon content (SOC) content (d) 0-5 cm dry bulk density (ρb), (e) 5-15 cm dry bulk density (ρb) and (f) 15-30 cm dry bulk density (ρb).



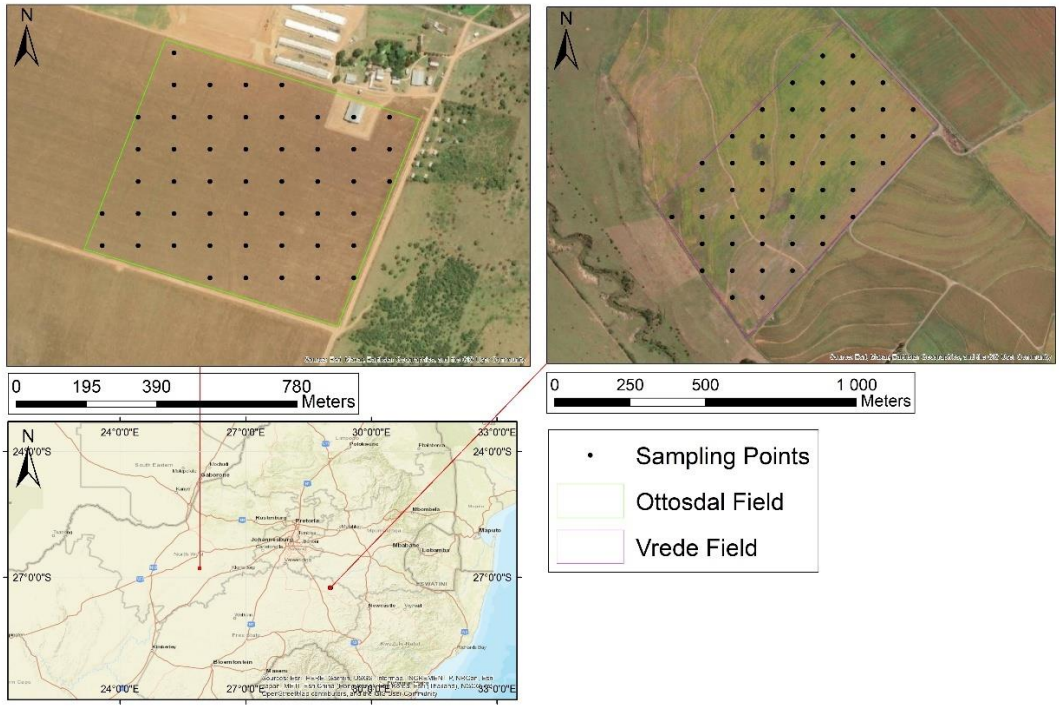
Appendix C: The maps created using digital soil mapping (DSM) with machine learning (ML) from the Ottosdal study site for (a) 0-5 cm soil organic carbon (SOC) content, (b) 5-15 cm soil organic carbon content (SOC) content, (c) 15-30 cm soil organic carbon content (SOC) content (d) 0-5 cm dry bulk density (ρb), (e) 5-15 cm dry bulk density (ρb) and (f) 15-30 cm dry bulk density (ρb).



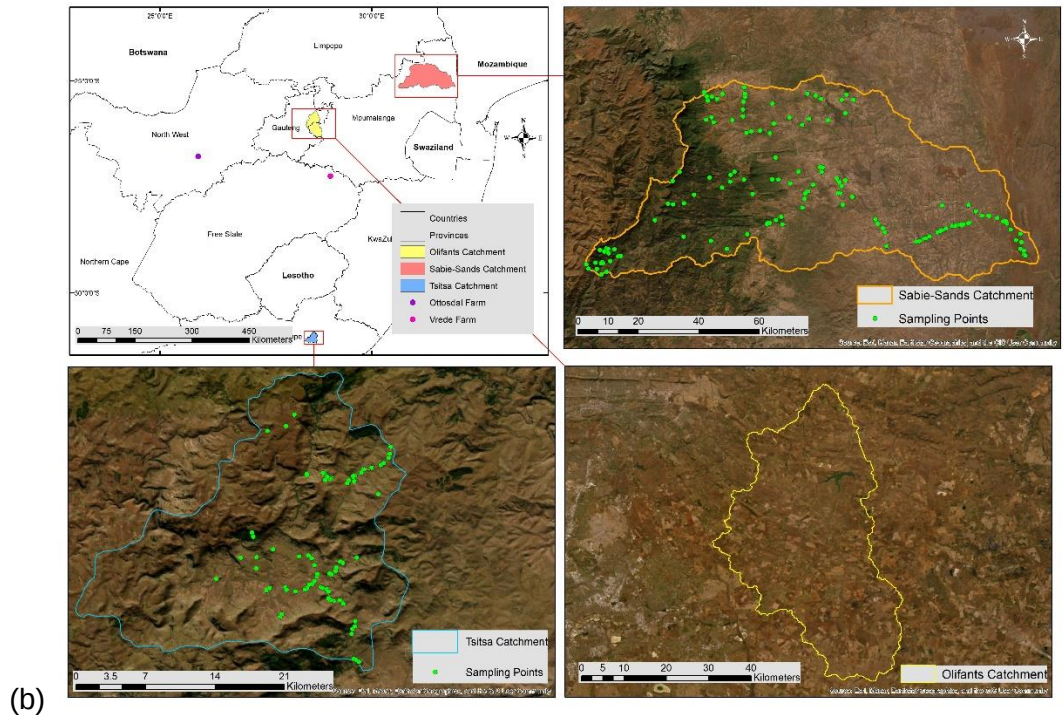
Appendix D: The maps created using digital soil mapping (DSM) with machine learning (ML) from the Vrede study site for (a) 0-5 cm soil organic carbon (SOC) content, (b) 5-15 cm soil organic carbon content (SOC) content, (c) 15-30 cm soil organic carbon content (SOC) content (d) 0-5 cm dry bulk density (ρb), (e) 5-15 cm dry bulk density (ρb) and (f) 15-30 cm dry bulk density (ρb).



Appendix E: The (a) field study areas and (b) catchment study areas with their sampling locations that was selected and used for this thesis.



(a)



(b)

Appendix F: Cash crop information for each study site, indicating crop type, harvest date, crop yield, harvest index (HI), shoot-to-root ratio (SR), below-ground biomass-to-above-ground biomass ratio (S_{factor}), carbon exudate factor of the roots (E_{factor}) and carbon factor from weed residues (W_{factor}).

Study site	Crop type	Harvest date	Crop yield (t/ha)	HI	SR	S_{factor}	E_{factor}	W_{factor}
Ottosdal	Maize	July 2017	9	0.52	5.6	0.1	0.09	0.07
	Soybeans	May 2018	2.7	0.38	5.2	0.1	0.09	0.07
	Maize	July 2019	4.8	0.52	5.6	0.1	0.09	0.07
	Maize	July 2020	5.16	0.52	5.6	0.1	0.09	0.07
	Maize	July 2021	7.24	0.52	5.6	0.1	0.09	0.07
	Soybeans	May 2022	3.8	0.38	5.2	0.1	0.09	0.07
	Maize	June 2023	8	0.52	5.6	0.1	0.09	0.07
	Maize	June 2024	1.6	0.52	5.6	0.1	0.09	0.07
Vrede	Maize	June 2017	6.7	0.52	5.6	0.1	0.09	0.07
	Soybeans	April 2018	3.1	0.38	5.2	0.1	0.09	0.07
	Maize	June 2020	1.7	0.52	5.6	0.1	0.09	0.07
	Maize	June 2021	7.5	0.52	5.6	0.1	0.09	0.07
	Soybeans	April 2022	3.4	0.38	5.2	0.1	0.09	0.07
	Maize	June 2023	8	0.52	5.6	0.1	0.09	0.07
	Soybeans	April 2024	1.65	0.38	5.2	0.1	0.09	0.07

Appendix G: Cover crop information for the Vrede study site, indicating crop type, seed rate, harvest index (HI), shoot-to-root ratio (SR), above-ground biomass (AGB) and below-ground biomass (BGB) fractions.

Crop type	Seed rate (kg/ha)	HI	SR	AGB	BGB
Oats	8	0.42	2.5	1.3	0.33
Rye	14	0.27	2.2	1.6	0.56
Vetch	3	0.38	2.8	0.2	0.11
Radish	0.5	0.3	0.5	0.64	0.08
Forage rape	0.5	0.25	5.0	1.8	0.54
Barley	8	0.38	3.0	1.4	0.28

Appendix H: The soil texture and soil organic carbon (SOC) values from the TDC, WB and LOI methods of the Goukou study site.

Study site	Soil texture	TDC (% C)	WB (% C)	LOI (% SOM)	LOI (% C)
GOUKOU	Sand	0.85	0.48	1.28	0.74
GOUKOU	Sand	0.88	0.71	1.12	0.65
GOUKOU	Sand	0.91	0.60	1.29	0.75
GOUKOU	Sandy Loam	2.06	1.55	1.91	1.11
GOUKOU	Sand	1.48	0.95	2.00	1.16
GOUKOU	Loamy Sand	1.16	0.79	1.59	0.92
GOUKOU	Sandy Loam	1.67	1.39	1.72	1
GOUKOU	Loam	3.51	2.38	2.26	1.31
GOUKOU	Sand	1.84	1.19	1.90	1.1
GOUKOU	Sand	2.11	1.03	4.31	2.5
GOUKOU	Sand	1.60	1.07	2.41	1.4
GOUKOU	Sand	6.41	3.98	5.59	3.24
GOUKOU	Sand	1.03	0.71	0.95	0.55
GOUKOU	Sand	0.57	0.32	0.76	0.44
GOUKOU	Sandy Loam	1.64	1.31	1.67	0.97
GOUKOU	Sandy Clay Loam	2.09	1.82	1.78	1.03
GOUKOU	Sand	0.59	0.32	0.72	0.42
GOUKOU	Sandy Loam	2.82	2.06	2.14	1.24

Appendix I: The soil texture and soil organic carbon (SOC) values from the TDC, WB and LOI methods of the Goukou and UMngeni study sites.

Study site	Soil texture	TDC (% C)	WB (% C)	LOI (% SOM)	LOI (% C)
GOUKOU	Sand	0.63	0.24	0.91	0.53
GOUKOU	Sandy Loam	3.62	0.12	1.12	0.65
GOUKOU	Clay Loam	2.05	1.32	3.05	1.77
GOUKOU	Sandy Loam	2.62	2.16	2.34	1.36
GOUKOU	Sand	0.40	0.16	0.71	0.41
GOUKOU	Sand	1.88	0.76	2.24	1.3
GOUKOU	Sand	0.74	0.44	0.84	0.49
GOUKOU	Sandy Loam	2.12	1.80	3.71	2.15
GOUKOU	Sand	0.44	0.20	0.74	0.43
GOUKOU	Sand	1.95	1.48	1.69	0.98
GOUKOU	Sand	1.20	0.68	0.86	0.5
GOUKOU	Sand	0.84	0.56	0.98	0.57
GOUKOU	Sandy Loam	3.96	2.24	3.57	2.07
GOUKOU	Sand	2.09	1.80	2.40	1.39
GOUKOU	Sand	7.08	0.90	1.02	0.59
UMGENI	Loam	5.68	4.05	8.43	4.89
UMGENI	Loam	5.75	4.53	10.69	6.2
UMGENI	Clay Loam	2.30	1.76	4.52	2.62

Appendix J: The soil texture and soil organic carbon (SOC) values from the TDC, WB and LOI methods of the UMngeni study site.

Study site	Soil texture	TDC (% C)	WB (% C)	LOI (% SOM)	LOI (% C)
UMGENI	Sandy Clay Loam	6.35	4.93	5.98	3.47
UMGENI	Clay	0.37	0.20	14.95	8.67
UMGENI	Clay	15.74	7.82	24.40	14.15
UMGENI	Silty Clay Loam	5.56	4.49	8.05	4.67
UMGENI	Clay	6.14	4.93	11.72	6.8
UMGENI	Silty Clay Loam	3.73	3.09	9.00	5.22
UMGENI	Loam	3.28	3.12	5.76	3.34
UMGENI	Sandy Loam	4.17	2.96	6.21	3.6
UMGENI	Clay Loam	4.47	3.87	6.60	3.83
UMGENI	Clay Loam	6.12	4.93	9.57	5.55
UMGENI	Silty Clay Loam	12.96	7.82	20.57	11.93
UMGENI	Clay Loam	10.84	7.46	16.03	9.3
UMGENI	Clay	5.85	4.74	9.43	5.47
UMGENI	Sandy Clay Loam	3.56	2.68	5.50	3.19
UMGENI	Sandy Loam	5.45	4.46	8.43	4.89
UMGENI	Sandy Clay Loam	4.20	3.59	6.48	3.76
UMGENI	Clay	8.26	6.36	9.48	5.5
UMGENI	Silty Clay	7.20	5.64	8.66	5.02
UMGENI	Silt Loam	5.69	4.22	7.95	4.61
UMGENI	Loam	6.76	5.45	9.41	5.46

Appendix K: The soil texture and soil organic carbon (SOC) values from the TDC, WB and LOI methods of the UMngeni and Sabie study sites.

Study site	Soil texture	TDC (% C)	WB (% C)	LOI (% SOM)	LOI (% C)
UMGENI	Clay Loam	2.75	2.37	5.64	3.27
UMGENI	Clay Loam	5.22	3.67	7.45	4.32
UMGENI	Sandy Loam	6.54	5.33	12.84	7.45
UMGENI	Loam	5.90	5.13	11.28	6.54
UMGENI	Loamy Sand	3.44	2.96	6.34	3.68
UMGENI	Loam	4.44	4.22	5.66	3.28
UMGENI	Sandy Loam	5.80	4.86	14.10	8.18
UMGENI	Sandy Loam	5.68	4.78	10.67	6.19
UMGENI	Sandy Clay Loam	4.53	3.91	9.88	5.73
UMGENI	Sandy Loam	6.01	3.50	5.17	3
SABIE	Loamy Sand	0.68	0.54	2.47	1.43
SABIE	Loamy Sand	1.08	0.65	2.00	1.16
SABIE	Sandy Loam	1.78	1.54	5.66	3.28
SABIE	Loamy Sand	0.77	0.58	2.03	1.18
SABIE	Loamy Sand	2.31	1.62	3.62	2.1
SABIE	Loamy Sand	0.62	0.50	1.64	0.95
SABIE	Loamy Sand	0.52	0.35	1.50	0.87
SABIE	Sandy Loam	5.88	4.16	19.83	11.5
SABIE	Sandy Loam	2.34	2.23	11.72	6.8
SABIE	Loam	4.64	4.62	17.84	10.35

Appendix L: The soil texture and soil organic carbon (SOC) values from the TDC, WB and LOI methods of the Sabie study site.

Study site	Soil texture	TDC (% C)	WB (% C)	LOI (% SOM)	LOI (% C)
SABIE	Sand	0.53	0.35	1.53	0.89
SABIE	Sandy Loam	4.15	3.62	10.03	5.82
SABIE	Sand	16.12	7.81	35.62	20.66
SABIE	Sand	0.62	0.42	1.64	0.95
SABIE	Loamy Sand	0.57	0.39	1.74	1.01
SABIE	Loam	2.64	2.54	14.78	8.57
SABIE	Sandy Loam	4.17	4.15	9.84	5.71
SABIE	Sandy Clay Loam	2.06	2.02	9.97	5.78
SABIE	Sand	0.35	0.15	0.76	0.44
SABIE	Sandy Loam	1.15	0.89	2.16	1.25
SABIE	Sandy Clay Loam	1.60	1.54	10.93	6.34
SABIE	Sandy Loam	2.40	2.08	12.05	6.99
SABIE	Sandy Loam	0.97	0.95	5.22	3.03
SABIE	Sandy Clay Loam	2.06	1.85	12.24	7.1
SABIE	Sand	0.57	0.46	1.36	0.79
SABIE	Loamy Sand	0.37	0.19	1.38	0.8
SABIE	Sandy Loam	4.76	4.58	14.67	8.51
SABIE	Sandy Loam	1.04	0.96	3.38	1.96
SABIE	Loamy Sand	0.88	0.46	2.48	1.44
SABIE	Clay Loam	3.38	2.85	9.95	5.77

Appendix M: The soil texture and soil organic carbon (SOC) values from the TDC, WB and LOI methods of the Sabie and Tsitsa study sites.

Study site	Soil texture	TDC (% C)	WB (% C)	LOI (% SOM)	LOI (% C)
SABIE	Sandy Clay Loam	2.22	1.96	8.43	4.89
SABIE	Loamy Sand	1.40	1.27	6.64	3.85
SABIE	Sandy Clay Loam	1.03	0.81	2.43	1.41
SABIE	Sandy Loam	2.77	2.62	4.41	2.56
SABIE	Loamy Sand	0.68	0.62	2.24	1.3
SABIE	Sand	0.55	0.12	1.55	0.9
SABIE	Sand	0.43	0.23	1.16	0.67
SABIE	Sand	0.52	0.35	1.50	0.87
SABIE	Sand	0.30	0.15	0.81	0.47
SABIE	Sandy Loam	1.11	0.85	3.17	1.84
SABIE	Sandy Clay Loam	2.79	1.66	7.40	4.29
TSITSA	Sandy Clay Loam	0.75	0.5	4.07	2.36
TSITSA	Sandy Clay Loam	1.04	0.8	5.22	3.03
TSITSA	Sandy Clay Loam	2.05	2.0	6.58	3.81
TSITSA	Sandy Loam	1.48	1.2	4.33	2.51
TSITSA	Sandy Clay Loam	1.31	1.1	4.30	2.50
TSITSA	Clay	2.81	2.2	11.01	6.39
TSITSA	Clay Loam	1.87	1.7	6.46	3.74
TSITSA	Sandy Clay Loam	1.25	1.0	2.07	1.20
TSITSA	Clay	2.04	1.8	7.92	4.60

Appendix N: The soil texture and soil organic carbon (SOC) values from the TDC, WB and LOI methods of the Tsitsa study site.

Study site	Soil texture	TDC (% C)	WB (% C)	LOI (% SOM)	LOI (% C)
TSITSA	Sandy Clay Loam	0.64	0.4	3.18	1.84
TSITSA	Sandy Clay	1.19	1.1	5.07	2.94
TSITSA	Sandy Clay Loam	1.04	0.9	3.42	1.98
TSITSA	Sandy Loam	0.96	0.8	3.38	1.96
TSITSA	Sandy Clay Loam	0.96	0.8	3.44	2.00
TSITSA	Sandy Clay Loam	1.54	1.4	4.11	2.38
TSITSA	Sandy Loam	0.90	0.7	3.07	1.78
TSITSA	Clay Loam	1.78	1.6	7.57	4.39
TSITSA	Sandy Loam	1.29	0.8	4.56	2.64
TSITSA	Clay	2.37	2.2	9.59	5.56
TSITSA	Sandy Loam	1.40	1.0	6.43	3.73
TSITSA	Loam	0.93	0.8	2.62	1.52
TSITSA	Clay	0.73	0.5	3.76	2.18
TSITSA	Sandy Clay	1.32	1.2	5.97	3.46
TSITSA	Loam	1.13	1.0	3.17	1.84
TSITSA	Sandy Loam	0.66	0.7	2.67	1.55
TSITSA	Clay Loam	0.73	0.6	6.61	3.84
TSITSA	Sandy Loam	0.67	0.6	2.03	1.18
TSITSA	Sandy Loam	1.09	1.1	3.01	1.74
TSITSA	Loam	1.27	1.1	4.40	2.55

Appendix O: The soil texture and soil organic carbon (SOC) values from the TDC, WB and LOI methods of the Tsitsa study site (continued).

Study site	Soil texture	TDC (% C)	WB (% C)	LOI (% SOM)	LOI (% C)
TSITSA	Clay Loam	0.95	0.7	3.27	1.90
TSITSA	Clay Loam	1.04	1.0	4.51	2.61
TSITSA	Sandy Loam	0.63	0.6	2.11	1.22
TSITSA	Sandy Loam	0.78	0.7	2.04	1.18
TSITSA	Sandy Clay Loam	1.04	0.9	3.13	1.81
TSITSA	Sandy Clay Loam	1.46	1.4	3.98	2.31
TSITSA	Sandy Loam	0.32	0.2	1.83	1.06
TSITSA	Sandy Loam	0.91	0.8	3.93	2.28
TSITSA	Sandy Clay Loam	0.90	0.8	3.35	1.94
TSITSA	Sandy Clay Loam	1.18	1.2	2.91	1.69
TSITSA	Clay Loam	2.92	2.9	12.72	7.38
TSITSA	Loam	3.60	3.5	12.01	6.96
TSITSA	Loam	1.02	1.0	4.38	2.54
TSITSA	Clay Loam	1.13	1.1	3.41	1.98
TSITSA	Sandy Loam	0.99	0.9	3.27	1.89
TSITSA	Sandy Loam	0.82	0.8	2.86	1.66
TSITSA	Sandy Clay Loam	0.90	0.9	3.54	2.06
TSITSA	Sandy Clay Loam	1.12	1.1	5.03	2.92
TSITSA	Clay Loam	0.90	0.8	3.17	1.84
TSITSA	Clay Loam	2.20	2.1	7.69	4.46

Appendix P: The soil texture and soil organic carbon (SOC) values from the TDC, WB and LOI methods of the Tsitsa and Olifants study sites.

Study site	Soil texture	TDC (% C)	WB (% C)	LOI (% SOM)	LOI (% C)
TSITSA	Clay	2.80	2.7	8.85	5.13
TSITSA	Sandy Clay Loam	0.67	0.7	3.57	2.07
TSITSA	Sandy Clay Loam	1.33	1.3	4.56	2.64
TSITSA	Clay Loam	1.80	1.8	5.60	3.25
TSITSA	Sandy Loam	0.66	0.7	2.49	1.45
TSITSA	Sandy Loam	0.70	0.7	3.24	1.88
TSITSA	Sandy Clay Loam	0.78	0.8	3.21	1.86
TSITSA	Sandy Clay Loam	1.55	1.3	4.31	2.50
TSITSA	Sandy Loam	0.87	0.7	3.16	1.83
TSITSA	Sandy Loam	1.26	1.0	4.25	2.46
TSITSA	Sandy Loam	0.61	0.5	2.57	1.49
TSITSA	Clay	1.79	1.5	7.51	4.36
OLIFANTS	Sandy Clay	4.59	2.7	9.88	5.73
OLIFANTS	Loamy Sand	0.75	0.4	6.44	3.73
OLIFANTS	Clay	2.19	1.5	8.38	4.86
OLIFANTS	Loam	2.55	2.3	10.09	5.85
OLIFANTS	Sandy Clay Loam	2.75	2.6	6.81	3.95
OLIFANTS	Clay Loam	0.92	0.8	4.02	2.33
OLIFANTS	Loam	1.58	1.4	8.49	4.92
OLIFANTS	Sandy Loam	2.23	1.3	5.84	3.39

Appendix Q: The soil texture and soil organic carbon (SOC) values from the TDC, WB and LOI methods of the Olifants study site.

Study site	Soil texture	TDC (% C)	WB (% C)	LOI (% SOM)	LOI (% C)
OLIFANTS	Loamy Sand	0.63	0.0	2.56	1.48
OLIFANTS	Sandy Loam	0.89	0.4	4.10	2.38
OLIFANTS	Clay	4.27	2.8	9.68	5.62
OLIFANTS	Sandy Clay Loam	1.55	1.2	7.51	4.35
OLIFANTS	Loamy Sand	2.90	2.0	6.78	3.93
OLIFANTS	Sandy Loam	0.73	0.2	2.46	1.42
OLIFANTS	Loamy Sand	0.57	0.0	2.18	1.27
OLIFANTS	Sandy Clay Loam	0.92	0.2	4.92	2.85
OLIFANTS	Loamy Sand	1.17	0.9	4.13	2.39
OLIFANTS	Sandy Clay Loam	2.82	2.1	10.50	6.09
OLIFANTS	Sand	0.60	0.0	1.49	0.87
OLIFANTS	Sandy Loam	2.61	2.2	5.53	3.21
OLIFANTS	Sandy Loam	1.34	0.7	3.66	2.12
OLIFANTS	Clay Loam	2.48	1.3	9.24	5.36
OLIFANTS	Sandy Clay Loam	2.07	1.5	8.30	4.82
OLIFANTS	Sandy Clay Loam	2.24	1.6	3.93	2.28
OLIFANTS	Loamy Sand	1.30	0.7	4.17	2.42
OLIFANTS	Sandy Loam	4.25	2.6	9.40	5.45
OLIFANTS	Loamy Sand	0.98	0.3	3.58	2.08
OLIFANTS	Loamy Sand	0.53	0.1	2.14	1.24

Appendix R: The soil texture and soil organic carbon (SOC) values from the TDC, WB and LOI methods of the Olifants study site (continued).

Study site	Soil texture	TDC (% C)	WB (% C)	LOI (% SOM)	LOI (% C)
OLIFANTS	Clay Loam	2.26	1.5	7.83	4.54
OLIFANTS	Sandy Loam	1.93	1.2	6.24	3.62
OLIFANTS	Sandy Loam	1.03	0.7	4.99	2.90
OLIFANTS	Sandy Clay Loam	0.90	0.2	3.90	2.26
OLIFANTS	Sandy Clay Loam	0.93	0.4	2.88	1.67
OLIFANTS	Sandy Loam	5.29	4.6	15.57	9.03
OLIFANTS	Sandy Loam	1.20	0.7	3.42	1.99
OLIFANTS	Sandy Clay Loam	1.44	0.7	8.22	4.77
OLIFANTS	Sandy Clay Loam	3.57	3.2	13.20	7.65
OLIFANTS	Sandy Clay Loam	2.25	1.5	8.29	4.81
OLIFANTS	Sandy Loam	1.73	1.5	5.98	3.47
OLIFANTS	Sandy Clay Loam	1.63	0.7	10.33	5.99
OLIFANTS	Sandy Clay Loam	1.47	0.9	5.89	3.42
OLIFANTS	Sandy Loam	1.18	0.5	5.96	3.46
OLIFANTS	Loamy Sand	1.77	1.7	5.02	2.91
OLIFANTS	Sandy Loam	1.95	1.1	8.41	4.88
OLIFANTS	Sandy Clay Loam	1.57	0.9	5.81	3.37
OLIFANTS	Loamy Sand	0.34	0.0	2.29	1.33
OLIFANTS	Sandy Loam	2.08	1.5	7.61	4.41
OLIFANTS	Sandy Loam	0.95	0.4	3.76	2.18

Appendix S: The soil texture and soil organic carbon (SOC) values from the TDC, WB and LOI methods of the Olifants study site (continued).

Study site	Soil texture	TDC (% C)	WB (% C)	LOI (% SOM)	LOI (% C)
OLIFANTS	Sandy Loam	0.95	0.4	3.76	2.18
OLIFANTS	Sandy Loam	0.79	0.0	4.09	2.37
OLIFANTS	Sandy Loam	0.82	0.2	3.10	1.80
OLIFANTS	Sandy Clay Loam	2.96	1.6	9.14	5.30
OLIFANTS	Sandy Loam	2.48	1.1	7.11	4.12
OLIFANTS	Sand	0.87	0.2	3.18	1.85

Appendix T: The soil organic carbon (SOC) content of the Ottosdal study site measured with the TDC method.

Sample number	OTTOSDAL SOC content (%)		
	0-5 cm	5-15 cm	15-30 cm
1	0.54	0.46	0.51
2	0.58	0.45	0.42
3	0.54	0.46	0.41
4	0.55	0.5	0.45
5	0.82	0.39	0.33
6	0.47	0.43	0.43
7	0.56	0.42	0.48
8	0.67	0.48	0.48
9	0.62	0.47	0.46
10	0.49	0.44	0.49
11	0.88	0.47	0.44
12	0.79	0.49	0.55
13	0.49	0.39	0.4
14	0.58	0.41	0.43
15	0.68	0.46	0.47
16	0.65	0.46	0.5
17	0.54	0.47	0.47
18	0.64	0.44	0.52
19	0.65	0.4	0.38
20	0.59	0.45	0.42
21	0.46	0.39	0.43
22	0.46	0.46	0.47
23	0.51	0.47	0.47
24	0.59	0.47	0.44
25	0.52	0.42	0.47
26	0.47	0.44	0.5
27	0.75	0.42	0.46
28	0.49	0.39	0.43
29	0.7	0.47	0.38
30	0.56	0.43	0.45

Appendix U: The soil organic carbon (SOC) content of the Ottosdal study site measured with the TDC method (continued).

Sample number	OTTOSDAL SOC content (%)		
	0-5 cm	5-15 cm	15-30 cm
31	0.85	0.49	0.47
32	0.55	0.5	0.52
33	0.68	0.44	0.46
34	0.47	0.51	0.48
35	0.65	0.51	0.44
36	0.61	0.55	0.65
37	0.78	0.37	0.41
38	0.49	0.44	0.5
39	0.68	0.49	0.49
40	0.52	0.46	0.46
41	0.8	0.46	0.58
42	0.67	0.52	0.46
43	0.51	0.41	0.43
44	0.66	0.43	0.46
45	0.86	0.44	0.41
46	0.7	0.49	0.46
47	0.63	0.5	0.5
48	0.72	0.44	0.44
49	0.72	0.47	0.44
50	0.56	0.43	0.54

Appendix V: The soil organic carbon (SOC) content of the Vrede study site measured with the TDC method.

Sample number	VREDE SOC content (%)		
	0-5 cm	5-15 cm	15-30 cm
1	0.81	0.56	0.62
2	0.93	0.47	0.47
3	1.32	0.85	0.77
4	1.04	0.56	0.66
5	0.94	0.51	0.61
6	0.94	0.5	0.55
7	1.7	0.81	0.92
8	0.94	0.62	0.69
9	1.13	0.45	0.51
10	1.03	0.65	0.69
11	1.89	0.73	1
12	2.3	0.94	0.8
13	1.37	0.77	0.74
14	2.46	0.73	1.09
15	0.74	0.65	0.65
16	0.87	0.71	0.73
17	1.4	0.82	0.81
18	1.64	0.9	0.79
19	1.9	0.87	1.01
20	2.3	1	0.9
21	1.59	1.06	1.06
22	2.06	1.38	0.97
23	0.94	0.72	0.77
24	0.94	0.77	0.88
25	1.39	0.8	0.92
26	2	1.35	1.1
27	2.29	1.36	1.36
28	1.63	0.92	0.87
29	2.29	1.07	1.01
30	1.46	1.06	0.83

Appendix W: The soil organic carbon (SOC) content of the Vrede study site measured with the TDC method (continued).

Sample number	VREDE SOC content (%)		
	0-5 cm	5-15 cm	15-30 cm
31	1.52	0.67	0.66
32	1.89	0.76	0.98
33	1.32	0.97	1.03
34	1.81	0.86	0.93
35	1.23	0.71	0.95
36	1.63	0.63	0.87
37	1.26	0.75	0.92
38	1.27	0.74	0.93
39	1.72	0.33	1.09
40	0.39	0.85	0.28
41	0.29	0.92	0.29
42	0.3	0.85	1.02
43	1.33	1.02	0.32
44	0.4	0.27	0.72
45	0.86	0.26	0.25
46	0.36	0.71	0.88
47	1.37	0.74	0.79
48	2.03	0.76	0.89
49	1.89	0.86	0.8
50	2.05	0.94	0.83

Appendix X: The clay (%) distribution of the Ottosdal study site.

Sample number	OTTOSDAL clay (%)		
	0-5 cm	5-15 cm	15-30 cm
1	5.14	6.91	7.96
2	5.92	7.65	7.42
3	7.17	7.69	7.47
4	6.61	6.38	6.70
5	6.35	6.78	7.00
6	7.12	6.67	7.41
7	5.81	6.78	7.61
8	6.66	8.88	9.34
9	5.91	7.01	6.92
10	4.40	6.88	7.89
11	7.06	6.86	8.00
12	5.90	6.43	7.34
13	6.66	6.75	6.53
14	5.16	7.90	8.03
15	6.56	7.64	7.16
16	7.33	7.45	7.21
17	6.33	6.03	7.43
18	6.70	7.62	6.70
19	3.76	6.88	7.62
20	6.33	6.73	7.53
21	6.50	5.86	6.80
22	6.48	6.25	9.50
23	6.47	7.11	8.38
24	6.35	6.64	7.27
25	6.51	4.78	6.89
26	6.42	6.71	7.92
27	7.18	5.32	7.61
28	6.27	6.26	6.92
29	5.24	6.77	6.68
30	6.77	6.52	7.79

Appendix Y: The clay (%) distribution of the Ottosdal study site (continued).

Sample number	OTTOSDAL clay (%)		
	0-5 cm	5-15 cm	15-30 cm
31	6.75	7.34	7.65
32	6.86	6.06	7.19
33	5.55	6.51	7.03
34	6.91	7.71	7.58
35	6.84	7.97	7.90
36	6.78	6.53	6.43
37	6.49	6.97	7.76
38	6.37	6.68	7.07
39	6.62	8.08	8.37
40	6.66	7.33	8.48
41	7.00	8.55	8.75
42	7.65	7.81	8.06
43	6.23	6.75	7.22
44	6.38	6.40	7.57
45	3.21	5.96	6.86
46	6.40	8.10	8.57
47	6.85	8.08	8.80
48	6.61	8.04	8.05
49	6.64	8.13	10.09
50	5.85	7.66	7.61

Appendix Z: The clay (%) distribution of the Vrede study site.

Sample number	VREDE clay (%)		
	0-5 cm	5-15 cm	15-30 cm
1	5.87	6.73	9.54
2	6.39	6.50	6.78
3	5.18	5.34	6.85
4	5.32	6.08	8.19
5	6.47	5.60	6.11
6	5.77	5.63	5.92
7	6.01	6.57	6.49
8	5.55	5.44	6.29
9	5.21	6.50	6.27
10	6.27	5.69	9.99
11	5.11	7.48	7.65
12	9.26	9.36	9.14
13	6.60	6.48	10.47
14	5.64	6.58	7.80
15	5.33	5.93	5.76
16	5.36	6.22	7.69
17	6.95	6.57	8.71
18	5.39	5.67	8.08
19	7.88	7.06	7.43
20	5.36	6.14	9.61
21	6.91	7.21	7.20
22	6.42	5.65	7.15
23	6.14	6.86	7.19
24	5.45	5.78	6.92
25	5.84	6.29	7.66
26	6.12	6.59	7.23
27	5.55	5.88	7.29
28	5.87	6.35	6.24
29	5.26	8.13	8.05
30	6.01	6.84	7.98

Appendix AA: The clay (%) distribution of the Vrede study site (continued).

Sample number	VREDE clay (%)		
	0-5 cm	5-15 cm	15-30 cm
31	4.99	5.33	7.18
32	6.29	7.66	7.49
33	5.22	6.84	9.64
34	6.47	7.84	8.97
35	6.70	6.57	6.83
36	5.94	7.60	7.70
37	5.75	7.76	7.75
38	6.03	6.05	9.17
39	7.56	5.58	6.27
40	7.57	6.16	8.91
41	5.13	6.66	7.83
42	6.06	7.21	5.99
43	5.16	7.19	7.75
44	6.62	6.11	8.77
45	6.07	6.37	7.82
46	6.40	7.57	7.95
47	6.17	8.49	9.65
48	6.35	7.68	8.87
49	6.71	7.94	8.44
50	5.55	6.66	7.66