An overview of the Soil Organic Matter content present within the Emakhosaneni area, KwaZulu-Natal, South Africa using Remote Sensing technologies

by

TESSNIKA SEWPERSAD

Submitted in fulfilment of the academic requirements for the degree of

Master of Science in Environmental Science

In the Discipline of Geography,

School of Agriculture, Earth and Environmental Sciences

University of KwaZulu-Natal,

Westville campus,

South Africa

June 2021

As the candidate's supervisor I have approved this thesis/dissertation for submission

Signed: _____ Name: Prof. Michael Gebreslasie Date: 01 July 2021

PREFACE

The research contained in this thesis was completed by the candidate while based in the Discipline of Geography, School of Agricultural, Earth and Environmental Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Westville Campus, South Africa. The School of Agriculture, Earth, and Environmental Sciences financially supported the research.

The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.

Signed: Miss Tessnika Sewpersad	Signed: Prof. Michael Gebreslasie
Date: 18 June 2021	Date: 18 June 2021

DECLARATION: PLAGIARISM

I, Tessnika Sewpersad, declare that:

(i) the research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;

(ii) this dissertation has not been submitted in full or in part for any degree or examination to any other university;

(iii) this dissertation does not contain other persons' data, pictures, graphs, or other information, unless specifically acknowledged as being sourced from other persons;

(iv) this dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written but the general information attributed to them has been referenced;

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

(v) where I have used material for which publications followed, I have indicated in detail my role in the work;

(vi) this dissertation is primarily a collection of material, prepared by myself, published as journal articles or presented as a poster and oral presentations at research conferences. In some cases, additional material has been included;

(vii) this dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Signed: Miss Tessnika Sewpersad

Date: 18 June 2021

ACKNOWLEDGEMENTS

I would like to thank God Almighty for giving me the knowledge and strength to complete my thesis. Thank you for guiding me through the process of conducting my research during this very tough year. Your blessings have encouraged me to complete my research work.

To my parents, Mr and Mrs Sewpersad, thank you for the support and encouragement during my academic years. The sacrifices you have made over the years have helped me become the person I am, and for that, I am eternally grateful. Your support has given me the strength to complete my thesis.

To my supervisor and mentor, Prof. Michael Gebreslasie, thank you for supervising this research. You have shown a great deal of expertise and professionalism over the past years. The criticism that you have provided regarding this research was highly valuable and assisted in shaping the research. I am incredibly grateful to you for your professional support and personal guidance throughout my undergraduate and postgraduate studies as both my supervisor and mentor. Your passion, love, and enthusiasm for Geographic Information Systems and Remote Sensing Sciences have and always will motivate me to pursue a future in the geospatial sciences.

A great deal of thanks to Dr Ebhuoma for assisting in the field research. Also, thank you for assisting in the collection of soil samples, which played an integral part in this research. Your guidance and expertise were highly valuable in this research.

A big thank you to Mr Xolani Gumede for assisting and guiding me through the laboratory analysis. Your insightful comments, encouragement, and suggestions have been quite pivotal in this research.

A great deal of appreciation goes out to those local residents of the area that assisted in the collection of soil samples. Your assistance is much appreciated as these samples supported the research.

Lastly, I would like to extend my gratitude to those individuals who assisted in executing this research. And, to my dearest pet, Trixie, your demise during the course of this research has been somewhat difficult to cope with, but the emotional support you provided has supported and motivated me to keep moving forward. And, for that, I am extremely grateful.

ABSTRACT

Soil Organic Matter (SOM) is one of the fundamental constituents of soil and plays significant roles in the overall fertility, productivity, and quality of soil. It consists of decaying plant and animal material at various stages of decomposition, substances released by plant roots, and soil organisms. Additionally, it is responsible for supporting many physical, chemical, and biological functions within the soil. And, these functions influence the provision of ecosystem services to humans, plants, and animals. However, this SOM is under threat as 25% of the earth's surface has become degraded, with 12 million hectares of topsoil being lost every year and, hence SOM. South Africa is one of those countries that are impacted by arable soil loss. Therefore, accurate measurement of SOM at different spatial scales is vital in providing information for planning a recovery strategy. However, traditional methods of soil analysis can be time-consuming, costly, and labour extensive. On the other hand, remote sensing is an efficient method that is time effective, low-cost, non-destructive, and has rapid data acquisition. Thus, offering an alternative to traditional methods of soil analysis. Hence, this research aimed to examine the SOM content within the Emakhosaneni area, KwaZulu-Natal, South Africa, by measuring the reflectance of soil using laboratory and remote sensing analysis. The first study examined the percentage of SOM content present within the study area, and its four major land uses using a laboratory technique. The results indicated that the area has a relatively low average percentage of SOM content (2.79%) present within its soils. Also, out of the four major land use types the agricultural land use had the highest average percentage of SOM content, followed by rangeland, built-up, and eroded land uses. These results further indicated the influence of land use activities on the SOM content within the study area. Overall, this study revealed the SOM content within the area is very low. It also highlighted the severity and consequences of the depleting SOM content within the area's soils. The second study examined the relationship between the SOM content and spectral reflectance of soils within the study area using laboratory spectroscopy. Results suggested that the spectra obtained was influenced by soil colour and thus, established a relationship between SOM content and reflectance of soil samples as SOM content is linked to soil colour. Further assessment of this relationship by Partial Least Squares Regression analysis revealed a fair performance. With the models created using the pre-processed spectra and SOM content performing better than those with no pre-processed spectra. Overall, this study highlighted the importance and capability of Visible and Near Infrared (400 nm to 2400 nm) spectroscopy in examining SOM content compared to conventional laboratory approaches and its influence on the spectral reflectance of the soils. The third study assessed the relationship between SOM content and Sentinel-2 satellite remote sensing data of soils in the area. Both geostatistical methods and hybrid geostatistical methods were used to predict SOM content. Results showed that the hybrid geostatistical methods performed better than the geostatistical methods in predicting SOM content due to the contribution of auxiliary information. Therefore, this study emphasizes the potential of auxiliary remote sensing data such as Sentinel-2 imagery in predicting SOM content in KwaZulu-Natal, South Africa, while also making critical inferences regarding the spatial and temporal variability of SOM within an area. The overall findings presented in this research are encouraging and show that different remote sensing techniques can be successfully used in the estimation, assessment, and prediction of SOM content, especially within the Emakhosaneni area, KwaZulu-Natal, South Africa, with accuracy levels that are acceptable.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF ABBRIEVATIONS

°C	Degree Celsius
µm	Micrometre
ANOVA	Analysis of Variance
ASD	Analytical Spectral Device
BI	Brightness Index
BI2	Second Brightness Index
BOA	Bottom of Atmosphere
CEC	Cation Exchange Capacity
CI	Colour Index
cm	Centimetre
cm^2	Square centimetre
DN	Digital Number
DVI	Difference Vegetation Index
e.g.	Example
EMR	Electromagnetic radiation
EPA	Environmental Protection Agency
ESA	European Space Agency
etc.	Etcetera
EU	European Union
EVI	Enhanced Vegetation Index
FAO	Food and Agriculture Organization
FDT	First-Order Derivative transformation
FOV	Field of view
g	Gram
GNDVI	Green Normalized Difference Vegetation Index
GPS	Global Positioning System
Ha	Hectares
HI	Hue Index
HM	Humic matter
Km	Kilometre
Km^2	Square kilometre

KZN	KwaZulu-Natal
LOI	Loss on Ignition
LOOCV	Leave-one-out cross-validation
m	Metre
ME	Mean Error
ml	Millilitre
MLR	Multiple Linear Regression
mm	Millimetre
MSAVI	Modified Soil Adjusted Vegetation Index
MSAVI2	Second Modified Soil Adjusted Vegetation Index
MSI	Multi-Spectral Instrument
NDVI	Normalized Differences Vegetation Index
NIR	Near infrared
nm	Nanometre
OK	Ordinary Kriging
PC	Principal Component
PCA	Principal Component Analysis
PLSR	Partial Least Squares Regression
r	Correlation coefficient
R	Raw spectra
R^2	Coefficient of determination
RDVI	Renormalized Difference Vegetation Index
RI	Redness Index
RK	Regression Kriging
RMSE	Root Mean Square Error
RMSEC	Root Mean Square Error Calibration
RMSEP	Root Mean Square Error Prediction
ROK	Regression Ordinary Kriging
RSK	Regression Simple Kriging
SANSA	South African National Space Agency
SAVI	Soil Adjusted Vegetation Index
SAWS	South African Weather Services
SDGs	Sustainable Development Goals

SG	Savitzky–Golay
SI	Saturation Index
SK	Simple Kriging
SNAP	Sentinel Application Platform
SOC	Soil Organic Carbon
SOM	Soil Organic Matter
SPSS	Statistical Package for Social Science
SSA	Statistics South Africa
SWIR	Shortwave infrared
TOA	Top of Atmosphere
TVI	Transformed Vegetation Index
V	Vegetation
VIS	Visible
VNIR	Visible Near Infrared
W	Watt
WB	Walkley-Black

CHAPTER ONE

INTRODUCTION

1.1 Introduction

The natural environment embodies all biotic and abiotic components that occur naturally on earth. This environment can be further divided into four different spheres encompassing the water, air, land, and living organisms. These subsystems or spheres of the earth consists of the lithosphere, hydrosphere, atmosphere, and biosphere (Earth eclipse, 2018). The lithosphere consists of all the hard and solid land mass on the earth's surface (crust), semi-solid rocks found underneath the earth's crust (mantle), and the liquid rocks found in the inner core of the earth (core) (Osman, 2013). The crust comprises of several layers, one of which includes the pedosphere. The pedosphere can be explained as an interface where the lithosphere, hydrosphere, atmosphere, and biosphere interconnect (Montgomery *et al*., 2000). More specifically, it consists of all soil present on earth and is where the many soil-forming processes are operational. Therefore, soil can be defined as "The upper layer of the earth which may be dug, ploughed, and specifically, the loose surface material of the earth in which plants grow" (Thompson, 1957).

According to the Food and Agriculture Organization (FAO) (2020), soil is considered "as a natural body consisting of layers (soil horizons) that are composed of weathered mineral materials, organic material, air, and water. Soil is the end product of the combined influence of climate, topography, organisms (flora, fauna, and human) on parent materials (original rocks and minerals) over time. As a result, soil differs from its parent material in texture, structure, consistency, colour, chemical, biological and physical characteristics". Soils are complex materials that can vary based on their physical, chemical, and biological composition. The composition of soil is largely influenced by its formation process. This process is a result of rocks and minerals that are found on the earth's surface becoming weathered over time. Subsequently, the environmental conditions of the atmosphere and lithosphere are known to largely impact the soil-forming process. Therefore, soils are primarily dependent on five formative factors which include; climate, vegetation, fauna, topography, and parent materials (Ben-Dor *et al*., 1999). These factors are responsible for allowing soils to exist in three phases, namely, solid, liquid, and gas. The solid phase of soil is comprised of organic and inorganic matter in the form of weathered rocks and decayed matter. The liquid phase consists of soil that is dissolved in water, where water fills the pore spaces between soil particles. The gaseous

phase of soil includes air being trapped between the solid and liquid parts of the soil (Montgomery *et al*., 2000). Each of these phases is responsible for influencing the soil's ability to support both biotic and abiotic components. For example (e.g.), the liquid phase of soil serves as a medium for dissolved mineral nutrients and is a key factor in determining how soils can support vegetation and plant growth. Subsequently, soil in its gaseous phase allows for important gases to be produced, absorbed, released, and accumulated within the soil (Smagin and Scalenghe, 2006). According to Glinski (2011), these phases play a pivotal role in the formation of soil, its physical properties, and soil processes.

The properties of soils are usually determined by the parent material that has become weathered over time (Doula and Sarris, 2016). These properties are essential in the identification, determination, classification, and management of the soil. Understanding soil properties can assist farmers and scientists in making important decisions regarding soil, such as its management (Schulte *et al*., 2005). Soil properties can be divided into three categories; physical, chemical, and biological (Foster *et al*., 2013). These properties are interdependent and work together collectively to produce a particular type of soil. Physical soil properties are associated with both the arrangement and size of solid soil particles and how the movement of liquids and gases in soils are impacted by their particles (Erich, 2001). Physical soil properties include soil colour, texture, structure, porosity, and permeability (Osman, 2012a). Chemical properties of soil include those that are related to the chemical make-up of the soil. This property is known to affect the fertility of the soil and its physical properties (Natural Resources Conservation Service, 2014). Chemical soil properties include cation exchange capacity, sodium adsorption ratio, exchangeable sodium potential, dispersivity, and soil pH. Biological soil properties refer to the number of biological organisms that are active within the soil. These usually include microorganisms such as bacteria and fungi as well as animals, namely; earthworms and insects (FAO, 2019). Other biological organisms include plants. Each of these organisms plays an important role in performing functions that influence the development of the soil, its structure, and its productivity (McCauley *et al*., 2005). Subsequently, these properties give an insight into the biological activity of the soil along with its biodiversity and the amount of biomass present. The properties of the soil are therefore important to understand in order to ensure that soil remains productive and viable for different uses.

Soil is a dynamic material that is composed of a combination of various types of materials and substances. These materials and substances exist in varying amounts that work together in influencing the properties of the soil and its composition. There are four main components of

soil that result in its final composition; organic matter, mineral particles, water, and air (Munna, 2017). Each of these components is responsible for influencing the productivity and fertility of the soil. Furthermore, these components exist in different percentages within a particular soil, and different soil types would have different amounts of each of these components. An average soil sample consists of 45% minerals, 25% water, 25% air, and 5% organic matter (Brady and Weil, 2002).

Minerals are solid components consisting of atoms with a particular order and regular arrangement (Munna, 2017). They are the largest component of the soil and can be divided into primary and secondary minerals. Primary minerals include minerals found in sand and silt, which contain soil materials similar to its parent material from which it was formed. In contrast, secondary minerals are a result of weathered primary minerals that are released in the form of important ions that go on to further form stable mineral forms such as clay (DeGomez *et al*., 2015). Clay is one of the richest minerals and contains important nutrients such as calcium, potassium, and iron.

Water comprises a relatively large volume of a soil. It plays an essential role in the transportation of nutrients to plants and organisms and facilitates processes within the soil, such as biological decomposition. Water within the soil acts as a medium in which nutrients can be dissolved in and used for plant growth. Subsequently, it is responsible for maintaining the soil's texture, arrangement, and compactness of soil particles (Biology Discussion, n.d). Different soil types and quality will influence the soil's ability to retain water.

The air component of the soil refers to the pores or spaces between soil particles. Subsequently, the air component also occupies approximately the same volume as water within a soil. These pores allow gases from the atmosphere to enter the soil as well allows gases to exit the soil and enter the atmosphere. Gases such as oxygen allow for root and microbe respiration in the soil, which in turn facilitates plant growth (DeGomez *et al*., 2015). The pores also allow water to enter the soil. When soils become waterlogged, it prevents the entry of gases into the soil and can thus impact the growth rate of plants.

In addition, organic matter is the smallest portion among other components that contributes towards the soil's composition. Organic matter is comprised of the remains of dead plant and animal material as well as the vast number of microorganisms that inhabit the soil (DeGomez *et al*., 2015). The remains of dead plants and animals are decomposed into organic compounds by microorganisms which results in the formation of humus. This component is responsible for

providing plants with the required essential elements and water that is needed for plant growth. Subsequently, it also has the ability to retain large amounts of carbon (FAO, 2017a). It is also an indicator of the productivity and fertility of soil. A particular soil will have a specific type and arrangement of these four components, therefore, creating a soil that has unique properties and characteristics. It is therefore essential to understand the components of a soil in order to understand its other underlying properties and features.

One such critical component is Soil Organic Matter (SOM). According to FAO (2017a), SOM refers to "the organic constituents in soil in various stages of decomposition such as tissues from dead plants and animals, materials less than 2 mm in size, and soil organisms". SOM plays a pivotal role in the functioning of soil ecosystems and global warming. It is also known for the stabilization of a soil's structure, the ability to retain and release plant nutrients, and maintenance of water-holding capacity in a soil (FAO, 2017a). Therefore, this makes SOM a critical indicator for agricultural productivity and fertility and, more importantly, resilience to environmental changes. A key characteristic of SOM is that it has the ability to retain carbon. This is known as Soil Organic Carbon (SOC) and is essentially the carbon that is held within the organic parts of soil, such as the products of decomposing plants and animals along with soil microbial biomass (i.e. SOM) (Gahlod *et al*., 2019). SOM is known to retain approximately 58% carbon, which is more than both the atmosphere and terrestrial vegetation combined (Donovan, 2013). This, coupled with the fact that it can retain carbon for longer periods of time, suggests that it can be used as a means to mitigate and adapt to climate change, improve agricultural productivity, and ecosystem functioning. However, over the years, the poor use and mismanagement of soil have led to the degradation of SOM and hence SOC. Despite large proportions of carbon being retained in soil, a total of 33% of the world's soil has been degraded over the years, and thus SOM has been lost (FAO, 2017b). It is therefore crucial for humans to understand the properties, composition, and functions of soil. The studying of soil educates us on the proper uses of soil, its limitations, and how it should be protected for future generations to come.

When it comes to studying soil there are different techniques that must be used in order to test for a specific reason. For example, a general assessment on the health of soils can be made by farmers assessing the quality and stand of the crops that are grown on these soils. In contrast, a scientific assessment provides a detailed physical, chemical, and biological analysis of soils (Motsara and Roy, 2008). The scientific analysis of soil assists us in determining the composition, nutrient content, and other important characteristics of soil. This procedure can

either be carried out in the field or in a laboratory. The collecting of soil samples in the field and bringing them back to a laboratory for analysis has been one of the most widely used and common methods that is still used to this very day. The reason for this is that a laboratory is considered as a controlled environment, thus preventing any confounding factors that can impact the results of the analysis. These methods can sometimes also be quite costly and time-consuming. Despite these common methods being used regularly, new techniques and technologies are being developed and incorporated into soil analysis. Such a case is the use of remote sensing in soil assessment, management, protection, and analysis.

Remote sensing refers to the process of capturing information about features on earth using satellites in outer space (Campbell and Wynne, 2011). It can also be referred to as a "technique to observe the earth surface or the atmosphere from out of space using satellites (space borne) or from the air using aircrafts (airborne)" (Aggarwal, 2004). The definition of remote sensing is known to differ based on the context in which it is being used and its application in various sciences. At times it may be defined as being a process, method or technique, or even a science itself. Despite its many definitions and applications, they all require and rely on the electromagnetic spectrum to acquire information about the earth's surface without being in contact with it. Information is acquired by sensing and recording reflected or emitted energy and processing, analysing, and applying that information.

The process at which features are remotely sensed goes through various steps before the output image can be used for application purposes by its users. However, different features return different amounts of energy in different bands of the electromagnetic spectrum. This energy can either be absorbed, transmitted, or reflected by a feature. However, remote sensing is mostly interested in measuring the reflectance from features. The return of energy from features is highly dependent on the property of the material, surface roughness, angle of incidence, intensity, and wavelength of the radiant energy (Aggarwal, 2004). It is therefore important to have a good understanding of the electromagnetic spectrum. Remote sensing usually involves the use of sensors in order to detect the electromagnetic energy, quantify it, and record it so that it can be transmitted to devices found on the ground. These sensors can be further subdivided into passive and active sensors.

Remote sensing has reached many milestones over the years and is still continuing to do so. With the growth of the digital revolution coupled with the development of user's skills and knowledge of remote sensing, this technique is becoming more popular with regards to data

acquisition, monitoring, mapping, and managing. It is widely being practised in many disciplines other than geography. This may be due to the fact that it allows users to capture data of features over large areas over long periods of time; this can be helpful when trying to compare geographical features such as fertile areas and how they have changed over time.

Remote sensing techniques allow for the estimation of different soil types and their spatial distribution at a much reasonable cost and has a higher accuracy when assessing large study areas (Kumar *et al*., 2013). These benefits are a result of conventional soil testing methods being too costly and time-consuming as many soil samples need to be collected and analysed manually or mechanically. Furthermore, using remote sensing for soils allows one to examine the spatial and temporal variability of soil within an area.
Each feature on earth is known to have its own unique spectral reflectance as they reflect or absorb the sun's energy in different ways. Similarly, soil has its own spectral reflectance that is unique to its properties, characteristics, and composition. This spectral reflectance allows one to distinguish between different features that are being examined over the earth's surface with the aid of remote sensing techniques. Therefore, an explicit understanding of the spectral reflectance of soils is required in any application of remote sensing regarding soils (Ravisankar and Sreenivas, 2011). Soil reflectance data is obtained either in the laboratory or in the field using a remote sensing instrument. This data gives an insight on the relationship between the physical and chemical properties of a soil and its reflectance. According to Wadodkar *et al*., (2014), "spectral reflectance of soils plays an important role in extracting information on different types of soils and in subsequent applications in soil mapping, land degradation mapping and monitoring, soil fertility management and watershed management". Subsequently, the spectral reflectance of vegetation can be used to infer soil conditions such nutrient differences and eroded locations. According to Krishan *et al*., (2014) "most of the passive remote sensors can provide information about soils from reflectance spectra in the visible (0.40 µm to 0.70 µm), near infrared (0.70 to1.10 µm) and shortwave infrared (1.10 to 2.50 µm) regions of electromagnetic spectrum. Besides this, thermal infrared regions (3.0 to 5.0 µm and 8.0 to 12.0µm) do provide diagnostic information about soils". Therefore, the resulting spectral reflectance curve is determined by the physical and chemical properties of the soil in the above-mentioned regions of the electromagnetic spectrum. Furthermore, understanding this curve allows one to examine soils in a non-destructive manner.

Due to the advanced capabilities of remote sensing, such as its ability to produce rapid and accurate results at lower costs, the examination of soils has become less tedious and much

easier. Many countries worldwide are adopting remote sensing when it comes to examining their soils, and such a case is South Africa. South Africa is considered as being a semi-arid region that receives less than 500 millimetres (mm) of rainfall (South African Weather Services [SAWS], 2019). Relatively low rainfall coupled with climate change and population growth make soil a critical natural resource. This natural resource differs in its distribution, properties, classification, genesis, and land use across South Africa (Fey, 2010). For example, each region of the country will be characterised by a specific soil type based on the climate of the area. There are three major soil regions of the country, where most of the drier areas of the country are comprised of sandy loam soils, while the Western and Eastern Cape contain grey sandy and sandy loam soils (Encyclopaedia Britannica, 2019). However, the soils in South Africa are usually not characterized by high fertility and become easily degraded with the exception of certain soils. This being an indication of degraded SOM. South Africa is known to cover approximately 121.9 million hectares (ha) of land, and only 13% of this land is suitable for activities such as agriculture (Mohamed, 2000). This small amount of soil is considered as being fertile enough to support the growth of crops. An increase in SOM results in soil fertility which in turn results in a greater yield of crops being produced. However, the fertility of soil along with SOC has decreased over the years due to improper use and management. According to Paterson *et al*., (2015), "the situation is likely to become more critical in the foreseeable future, with the continually increasing population of South Africa requiring more food, fibre and energy to be produced, against the background of potentially changing climate patterns across the country and the wider southern African region". Therefore, it is imperative to assess SOM to ensure that it can support different uses and activities and the plants that require nutrients from the soil for their growth.

Similarly, the current research will aim to examine the soil nutrients in the Emakhosaneni area located in KwaZulu-Natal (KZN), South Africa, using remote sensing techniques. The area is comprised of a small population that has a dispersed settlement pattern. Upon examining the area using satellite imagery, it is clearly evident that agriculture is practised on both a commercial and subsistence level (Google Earth, 2020). Such activities are known to cause soil erosion and loss of nutrients within the soil. Therefore, using remote sensing techniques when it comes to examining SOM on such a scale will be advantageous. Furthermore, this would provide important information regarding the health and status of the soil within the Emakhosaneni area.

1.2 Motivation

South Africa is one of the many countries in the African continent that is known to be quite abundant when it comes to natural resources such as minerals, soil; on the other hand, is a limiting resource that is unequally distributed spatially and temporally. The country is known for its characteristic semi-arid nature as it receives an annual rainfall of less than 500 mm, which is below the world average (SAWS, 2019). It is the east and south coast of South Africa that receives most of its summer rainfall, while the west coast is considerably drier due to the fact that it receives most of its rainfall during the winter season. This unequal concentration of water in South Africa influences the status, health, fertility, and types of soils that occur within South Africa. Soil is a critical resource that is needed by the 55.7 million residents in South Africa, either directly or indirectly (Statistics South Africa [SSA], 2011).

According to the Soil Classification Working Group (1991), soils of South Africa are categorised into 73 soil forms that are characterised by the nature of the topsoil and the sub-soil horizons. Each of these varies across the country and are defined by the climate, rainfall pattern, flora, fauna, etcetera (etc.) within a particular region (Jenny, 1994; Van Der Merwe *et al*., 2002; Fey, 2010). Despite the diversity in soil forms, they all play an important role to flora, fauna, and humans. For instance, soil plays an important role in the ecological cycles of the environment. Some of these cycles include the carbon, nitrogen, oxygen, water, and nutrient cycles (Bot and Benites, 2005). Each cycle is unique; however, certain elements may be similar in more than one cycle. They are responsible for their contribution to providing a variety of ecosystem services. For instance, carbon plays a critical role in the functioning of life on earth. Carbon is constantly being cycled through the various spheres in the environment as either a solid, liquid, or gas. According to Bot and Benites (2005), "carbon cycling is the continuous transformation of organic and inorganic carbon compounds by plants and micro- and macro-organisms between the soil, plants and the atmosphere". However, it is soil that plays the most significant role in the carbon cycle. Soil is known for storing more carbon than the atmosphere and terrestrial vegetation combined (Lefevre *et al*., 2017). This carbon is then used by microorganisms present within the soil for processes such as respiration and decomposition. More importantly, since soil serves as a giant sink for carbon, it plays a critical role in climate change and global warming. Therefore, soil is responsible for absorbing excess carbon that is present within the atmosphere due to the release of greenhouse gases via anthropogenic activities. In doing so, soil is able to alleviate some of the effects and impacts of climate change.

Without soil, ecological cycles such as the above mentioned would not be able to function normally, which, in turn, can negatively influence the environment.

According to FAO (2015a), "soils are a key enabling resource, central to the creation of a host of goods and services integral to ecosystems and human well-being. The maintenance or enhancement of global soil resources is essential if humanity's overarching need for food, water, and energy security is to be met. In particular, the projected increases in food, fibre, and fuel production required to achieve food and energy security will place increased pressure on the soil". Soil is an important natural resource that is used each and every day by humans, either directly or indirectly, plants, animals, as well as other living organisms. It performs many functions for its users in a variety of ways. This could either be in the form of facilitating certain processes, supporting many life forms, or serve as a form of protection. Regarding the importance of soil from the human aspect, it provides an array of ecosystem services that are critical to human life (Soil Science Society of America, 2002). Soil has also been identified as one of the key factors when it comes to achieving a few of the seventeen Sustainable Development Goals (SDGs) of the 2030 Agenda for Sustainable Development. More specifically, sustaining and increasing the SOC storage to assist in achieving these goals. In doing so, the full potential of soil ecosystem services can be unlocked and thus enable the support, maintenance or improvement of soil fertility and productivity (SDG 2 "Zero Hunger" and SDG 3 "Good Health and Well Being"), as well as to store and supply water that is cleaner (SDG 3 and SDG 6 "Clean Water and Sanitation"), maintain biodiversity (SDG 15 "Life on Land"), and to increase ecosystem resilience to cope with a changing climate (SDG 13 "Climate Action") (FAO, 2017a). Ensuring that the SOM content in soils (and thus SOC) remain at an equilibrium or increased levels would assist in restoring the soil and improving the overall quality of soil. Important information can also be gathered through analysis and can further be used to assist farmers and the general public when it comes to the proper use and management of soil.

Subsequently, soil serves as a medium on which humans can build structures and live. More specifically, soils provide a base for the construction of buildings, roads, parks, settlements, etc. (Nortcliff *et al.*, 2000). Furthermore, soils are also used in the creation of building material such as cement and bricks. It is upon soils along which civilizations or settlements have arisen as they provide a stable medium that can support the structures that humans have built over the years. Soils allow for the establishment and development of a community by allowing

important transport networks to be created, as well as the creation of shelters and office buildings.

The soil in an area is subject to the climatic conditions, vegetation, rainfall patterns, and the disturbance of soil in its natural state (Jenny, 1994; Van Der Merwe *et al*., 2002). When soil is disturbed by various anthropogenic activities, the physical, chemical, and biological properties of the soil begin to change. This, coupled with the changes in the environment, can negatively influence the fertility, quality, and health of the soil. Similar challenges are being faced within the Emakhosaneni area that was selected for this research. It is evident that the area is known for partaking in agricultural activities based on its inspection from satellite imagery (Google Earth, 2020). The scale of farming ranges from subsistence to commercial, where large land areas are allocated for this activity. This suggests that individuals living within the area rely on the soil to carry out agricultural practices. Soil plays an essential role in providing a medium for plant growth. The fertility and quality of the soil will determine the productivity of the soil and hence the yield of crop produced. Agricultural practices that occur on a commercial scale usually involves the use of fertilizers, pesticides, and herbicides in order to increase the yield of crops grown to meet the demands of the people (Rother *et al*., 2008). However, the extensive use of pesticides and nutrients over prolonged periods of time can impact the soil by changing its composition. It can also runoff along the different crops being grown in the nearby area or into the nearby river during rainfall periods. This water that contains pesticides and/or herbicides is used by local and subsistence farmers to grow crops that are both consumed and sold in farmers markets, thus hindering the quality of crops and impacting on the health of those individuals that consume these crops (Ongley, 1996; Gerba and Choi, 2006). In addition, with high population growth, there is an increased demand for food; this now means that crop yield needs to increase, which inadvertently indicates that more resources such as water and good fertile soil are needed (Pimentel *et al*., 1997). However, agricultural practices involve using the soil for prolonged periods of times in order to meet the growing food demands of the world's population. This prevents the soil from replenishing to a healthier state before it can be used once again. Such prolonged use of soil can result in its degradation (Karlen and Rice, 2015). This can negatively impact the quality of crops and their yield, thus directly impacting farmers as a loss of income is seen. Furthermore, a lack of food can result in food insecurity, and this could inadvertently exacerbate the poverty cycle.

Subsequently, individuals within the area are partaking in poor land use practices such as monoculture, extensive cultivation, and grazing of livestock over prolonged periods of time

over the same area. The outcomes of this are that it gives rise to soil erosion, which lowers the soil moisture levels and, therefore, leads to the degradation of SOM. More specifically, groundwater sources are subjected to drying up as they are dependent on the infiltration of rainfall through the soil (Sehmi and Kundzewicz, 1997). With the soil becoming more degraded by land use practices, not enough time is given to replenish the soil moisture content.

Each of these uses of soil has a cyclic impact on the soil and its users. This can reduce the fertility of soil over time along with its functioning. Furthermore, in conducting this particular research, it would provide useful information regarding the fertility of soil within the Emakhosaneni area. It would give an insight into the SOM content present within the soil of the study area. The information gathered from the research would be essential in assisting local farmers by providing them with the evidence that can help improve the soil's productivity along with their crop yield. Subsequently, using remote sensing techniques for such research can allow larger areas to be studied at a reduced price and time. It also allows for the studying of soil in a non-destructive manner. Lastly, research such as the above mentioned can assist in identifying the proper requirements in order to ensure and improve the standard and quality of soil within the area.

1.3 Aim of the research

Given the above discussion, the general aim of this research is to examine the Soil Organic Matter content within the Emakhosaneni area, KwaZulu-Natal, South Africa, by measuring the reflectance of soil using laboratory and remote sensing analysis.

1.4 Objectives of the research

1. To determine the Soil Organic Matter content of soils within the Emakhosaneni area using laboratory-based techniques.
2. To examine the relationship between Soil Organic Matter content and laboratory-based spectral reflectance in the Emakhosaneni area.
3. To assess the relationship between Soil Organic Matter content and satellite remote sensing data in the Emakhosaneni area.

1.5 Thesis structure

Chapter One includes an introduction to remote sensing and its techniques in addressing soil and the degradation of SOM within South Africa. Subsequently, this chapter outlines the research aims and objectives and provides a general overview of the thesis structure. Chapter Two includes an examination of the total SOM content present within the soil of the Emakhosaneni area using a laboratory-based technique. In this study, the percentages of SOM content are calculated for each of the four land use types in which soil samples were collected. An Analysis of Variance is also conducted to determine any statistically significant differences in SOM percentages within the land use types. The chapter also focuses on the risks imposed by the deterioration of SOM on people and their livelihoods. Chapter Three includes a critical analysis of the relationship between the SOM found within the Emakhosaneni area, particularly within its major land use types and the spectral reflectance from its soils. It also highlights the importance of Visible and Near Infrared spectroscopy in the measurement and prediction of SOM content as compared to the conventional laboratory approaches through the use of correlation analysis and partial least squares regression. Chapter Four provides an assessment of the relationship between SOM content and satellite remote sensing data within the soil of the Emakhosaneni area. Focus is placed on the prediction of SOM content by using satellite remote sensing data coupled with geostatistical methods and hybrid geostatistical methods. Chapter Five provides a conclusion of the overall research conducted and appropriate recommendations for future studies.

1.6 Conclusion

This research focuses on the fertility of soil within the Emakhosaneni area and key aspects such as SOM and remote sensing. This is done by examining the SOM content present within the soil in a non-destructive manner through remote sensing. Soil being a critical resource, is highly subjected to degradation. It is the most relied on medium by humans for basic needs, specifically those in low-income households. It also plays a global role in the storage of carbon and, therefore, mitigating and alleviating climate change. Such research can be beneficial in trying to conserve and manage this limited resource for future generations. It can also assist in predicting future risks and patterns and to allow for appropriate management strategies to be put into action at an early stage. The use of remote sensing would also reduce the cost and time taken to analyse SOM whilst providing acceptably accurate and reliable results. This chapter outlines the major aims and objectives of the research and highlights the research's motivation. Lastly, an overview is given of the chapters that are to follow.

CHAPTER TWO

A comprehensive estimation of Soil Organic Matter content within soils of the Emakhosaneni area, KwaZulu-Natal using a laboratory-based approach

Abstract

Soil is a critical resource that is required by every individual on earth, along with living and non-living organisms. Four major components exist in varying amounts that result in the final composition of soil; Mineral particles (45%), Water (25%), Air (25%), and Organic Matter (5%). Soil Organic Matter (SOM) consists of decaying plant and animal material at various stages of decomposition, substances released by plant roots, and soil organisms. It is responsible for supporting many physical, chemical, and biological functions within the soil. However, human activities coupled with population growth has led to the degradation of SOM. Similarly, soils within South Africa have low fertility due to the degradation in SOM over the years caused by the improper use and management of soil. Therefore, this study aimed to examine the total SOM content present within the soil of the Emakhosaneni area, KwaZulu-Natal, South Africa using laboratory techniques. Thirteen random soil samples were obtained from each of the four major land uses (Agriculture, Built-up (Residential), Eroded, and Rangeland); hence, a total of 52 soil samples was obtained. Samples were then oven dried overnight (105℃), crushed, and sieved using a 425 µm sieve. Samples were then analysed using the Walkley-Black method, and the percentage of SOM was calculated. Descriptive statistics revealed that the area had an overall average SOM content of 2.79%, which can be supported by the generally low soil organic carbon content (<0.5%to>4%) found in South African topsoils. The results also revealed that agricultural land use had the highest average percentage of SOM, followed by rangeland, built-up, and eroded land use. This is due to the presence of vegetation which supplies organic material to its soil organisms, hence increasing the biological activity within the soil and facilitating the SOM formation process and thus SOM content. Similarly, extensive grasses within the rangeland land use served the same role. Human activities like construction, walking, and driving within the built-up land use removed and compacted topsoil, which negatively impacted the soil structure and SOM content. Due to ongoing sheet, rill, and gully erosion, the SOM content was reduced within the eroded land. The one-way Analysis of Variance revealed that the percentage of SOM within the major land uses was statistically significant (p<0.05). Furthermore, the post-hoc Tukey's test revealed that the percentage of SOM within the agriculture land use and eroded land were statistically significant (p<0.05). This results from the vegetation layer responsible for promoting biological activity, which can generate and enhance SOM content and provide the soil with protection from environmental and human factors. SOM plays significant roles in the overall fertility, productivity, and quality of soil and should be managed and protected in a sustainable manner for future generations to come.

Keywords: Soil Organic Matter, Land Use, Vegetation, Microorganisms, Walkley-Black

2.1 Introduction

Soil is a critical resource that is required by every individual on earth, including living and non-living organisms (Bhattacharyya and Pal, 2015). Its formation is dependent on soil-forming factors that work together to produce soil. There are four significant components that exist in varying amounts, which results in the final composition of soil; Mineral particles (45%), Water (25%), Air (25%), and Organic Matter (5%) (Brady and Weil, 2002). SOM consists of decaying plant and animal material at various decomposition stages, substances released by plant roots and soil organisms (Brady and Weil, 2008). SOM is responsible for supporting many physical, chemical, and biological functions within the soil. These functions influence the ecosystem services provided to humans, plants, and animals. Despite the small amount of SOM found in topsoil, it plays a significant role in the overall fertility, productivity, and quality of soil. However, human activities coupled with an increasing growth in population has led to the degradation of SOM within soils. Similarly, the soils in South Africa have low levels of fertility due to degradation in SOM over the years caused by the improper use and management of soil. It is, therefore, important to monitor, assess, measure and manage the SOM content present within soil. Thus, the current study focuses on SOM estimation using an approach that is laboratory-based. The estimation and/or measurement of SOM gives an insight into the content found within soil. This quantification of SOM can be used to further help understand soil and enhance its overall health, quality, and fertility. It can also be used to assist in determining appropriate management practices to ensure soil functioning. The use of a laboratory-based approach allows for SOM content to be estimated accurately under controlled conditions.

2.1.1 Pedosphere

The pedosphere is referred to as the outermost layer of the earth that is comprised of soil and is subject to soil-forming processes (Wang *et al*., 2019). It exists at the interface of the atmosphere (air above and in the soil), biosphere (living plants, animals, and microorganisms), hydrosphere (water in, on, and below the soil), and lithosphere (unconsolidated regolith and consolidate bedrock) (Jacobson *et al*., 2000; Van Breemen and Buurman, 2002). The pedosphere is the earth's skin; it develops when there is a dynamic interaction between the four spheres (atmosphere, biosphere, lithosphere, and the hydrosphere) (Lal *et al*., 1997; Van Breemen and Buurman, 2002). Thus, the pedosphere serves as an interface that allows the water, air, land, and living organisms present on earth to interact and assist in the formation and functioning of the soil (Lal *et al*., 1997). Furthermore, it functions as the earth's biogeomembrane, which facilitates and regulates the exchange of substances and fluxes of

energy amongst the biota, atmosphere, hydrosphere, and lithosphere (Targulian *et al*., 2018) (Figure 2.1). This thin layer acts as a cover of the earth's surface, which has an enormous diversity of soils regarding their compositions and properties.

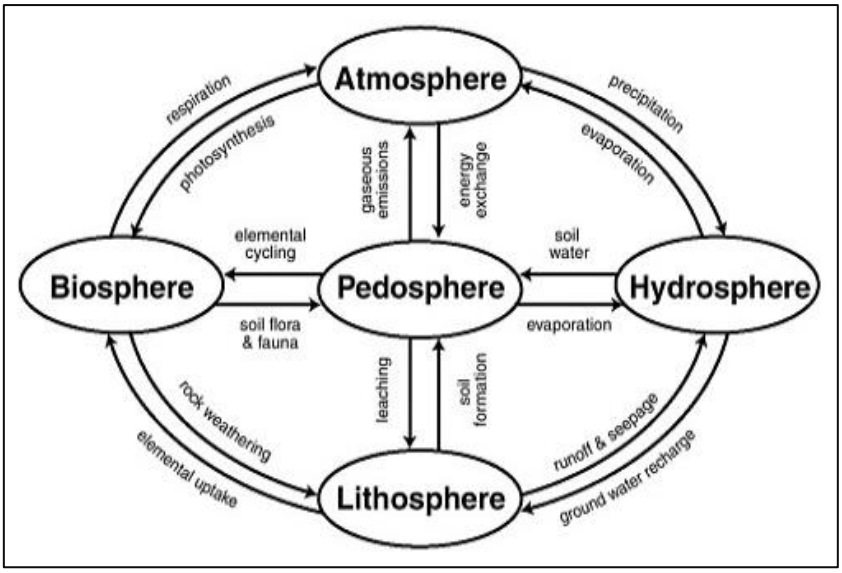


Figure 2.1: The pedosphere seen as the intersection among the lithosphere, hydrosphere, atmosphere, and biosphere (Source: Lal *et al*., 1997)

2.1.2 Soil

According to the FAO (2020), soil can be defined as "a natural body consisting of layers (soil horizons) that are composed of weathered mineral materials, organic material, air and water. Soil is the end product of the combined influence of climate, topography, organisms (flora, fauna and human) on parent materials (original rocks and minerals) over time. As a result, soil differs from its parent material in texture, structure, consistency, colour, chemical, biological and physical characteristics". This definition by the FAO focuses on the composition of soil along with the soil-forming processes. It is specific in that it mentions the factors that act together in order to produce and influence soil. However, the definition of soil is known to differ based on the context in which it is being used (Nortcliff *et al*., 2006). For instance, within the discipline of geology soil is considered as an accumulation of loose material that is formed by the process of rocks becoming weathered both mechanically and chemically over time (Kalev and Toor, 2018). However, to a farmer, the soil is seen as a natural medium located on the surface of the earth in which plants can grow. More specifically, it can be noted that a modern definition of soil usually consists of six key elements, namely (Brevik and Arnold, 2015):

1. Soils are natural bodies.
2. Soils are both spatial and temporal.
3. Soils form at the surface.
4. Soils are the result of complex biogeochemical and physical processes.
5. Soils are capable of supporting life.
6. Soils can be mapped at appropriate scales.

Defining soil in a detailed manner allows for a more holistic understating of soil and its many processes, characteristics, and properties.

Soils are complex materials that are largely influenced by their formation process. Since soil is a complex mixture comprising of different components, soil formation is more complex and is dependent on soil-forming factors (Balasubramanian, 2017). Soil is distinct from the parent material from which it has originated in that it has undergone considerable physical and chemical weathering of primary and secondary minerals (Birkeland, 1999). According to Jenny (1994), five factors were deemed essential in soil formation and can be expressed as a fundamental equation of soil formation, $s = f(cl, o, r, p, t)$, where the five independent factors of soil formation include climate (*cl*), organisms (*o*), topography (*r*), parent material (*p*), and time (*t*). These five factors are interconnected and work together to form soil. Subsequently, they are also responsible for influencing the properties of soil. Soil properties can be divided into three categories; physical, chemical, and biological (Foster *et al*., 2013). These properties are interdependent and work together collectively to produce a certain type of soil.

Soil is a critical resource required by every individual on earth, along with living (e.g. flora and fauna) and non-living organisms (e.g. water, air, minerals, etc.) (Bhattacharyya and Pal, 2015). Its dynamic nature allows for a composition that is composed of a combination of various types of materials and substances. These materials and substances exist in varying amounts that work together in influencing the properties of the soil and its composition. According to Kalev and Toor (2018), there are four major components that result in the final composition of the soil, which include; Mineral particles, Water, Air, and Organic Matter. Subsequently, these components can exist in varying percentages within a particular soil, and therefore, different soil types would have varying amounts of each of these components. An average soil sample consists of 45% minerals, 25% water, 25% air, and 5% organic matter (Brady and Weil, 2002) (Figure 2.2).

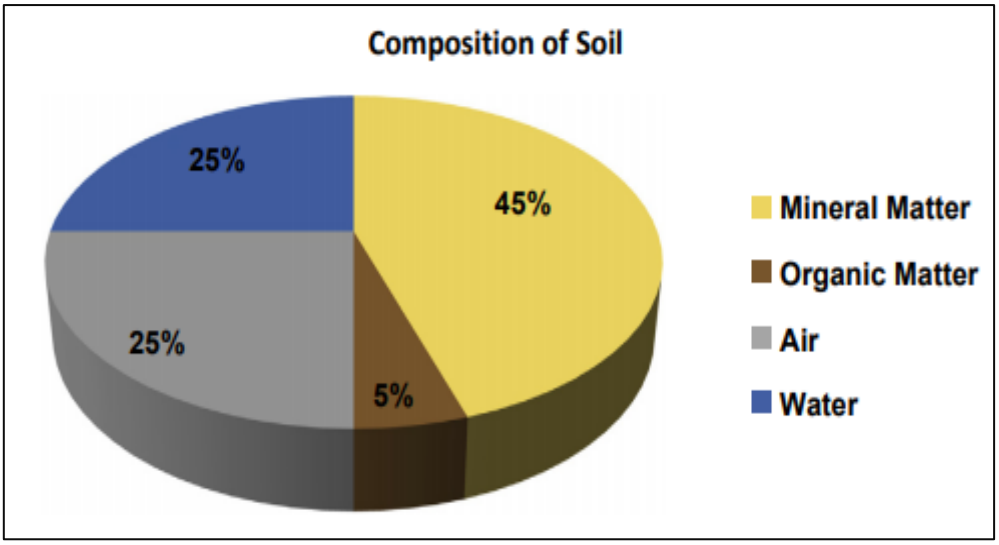


Figure 2.2: The four major components of soil in varying proportions (Source: Kalev and Toor, 2018)

2.2 Soil Organic Matter

According to Jenny (1994), soil may be considered as an open system with components entering and/or leaving the soil, including SOM. Changes in the dynamics of the five soil-forming factors can have an influence on both the accumulation and/or release of SOM (Simpson, 2010). SOM primarily originates from plant litter and microbial biomass and consists of various different compounds that vary in structure, content, and resistance. SOM contributes the smallest amount towards the soil's composition (Figure 2.2). According to the FAO (2017a), SOM refers to "the organic constituents in soil in various stages of decomposition such as tissues from dead plants and animals, materials less than 2 mm in size, and soil organisms". Subsequently, it can be defined as "all organic materials found in soil that are part of or have been part of living organisms" (Chenu *et al*., 2015). SOM is a range of materials at different stages of transformation due to the influence of abiotic and biotic processes. It is responsible for contributing a small portion towards the solid phase of soil. The formation of SOM is primarily a biological process that is directly or indirectly influenced by the flora and fauna present within the soil (Bot and Benites, 2005).

Organic material such as plants, animals, and microorganisms contribute to the organic matter in soil. However, the major input towards organic matter in soil is plants, either in the form of aboveground litter (twigs, leaf litter, and stems) or belowground material (exudates, root litter, and mycorrhizal hyphae) (Wetterstedt, 2010). These organic materials usually enter and mix with the soil, where they undergo decomposition. According to McClaugherty and Berg (2011),

decomposition “is the breakdown, by physical and biological mechanisms, of organic substances found in the soil”. Decomposers such as bacteria, earthworms, and fungi are responsible for the breaking down process (Francioli *et al*., 2020). The organic material provides a source of energy and nutrients for these organisms that live in the soil. The organic material that is broken down by these organisms is then converted into new substances that are added back into the soil. This, in turn, contributes to the overall organic matter present in soil. For example, the remains of plants and animals contain sugars, proteins, lignins, celluloses, hemicelluloses, waxes, and lipids (Osman, 2012b). They also contain essential elements such as carbon, hydrogen, oxygen, nitrogen, phosphorous, and sulphur. When this organic material undergoes decomposition, its products and newly synthesized material accumulate in the soil and produce a more complex form of SOM (Osman, 2012b). This, in turn, provides the soil and its organisms with nutrients and food (Figure 2.3).

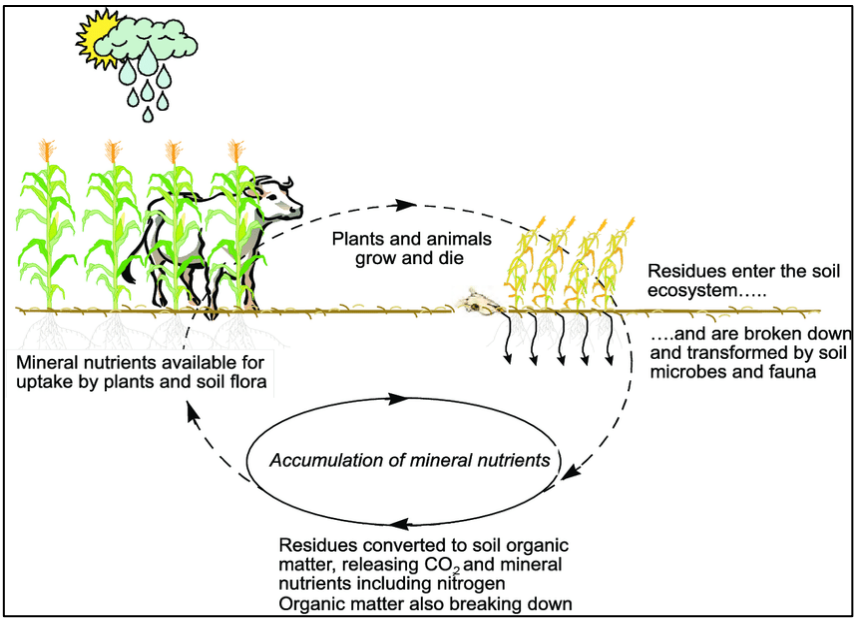


Figure 2.3: General overview of SOM cycle (Source: Dalgliesh and Foale, 1998)

Additionally, the rate of SOM formation is dependent on factors such as temperature and precipitation (Ismail-Meyer *et al*., 2018). These factors influence the decomposition process and the activity of organisms present within the soil. For instance, temperature is known to have an effect on both the growth of microorganisms along with the biological activity of soils (Pietikäinen *et al*., 2005). More specifically, the growth and activity of decomposers is temperature dependent. Any changes in temperature will have a direct influence on the overall functioning of bacteria, earthworms, and fungi present in soil. For instance, higher temperatures will increase microbial activities in soil and will allow for more rapid

decomposition of organic materials. The converse will be seen under lower temperatures, where microbial activities are reduced, and therefore the rate of decomposition will decrease, and so too the organic matter (Blume *et al*., 2016). However, these decomposers cannot withstand and function under extremely hot and cold temperatures, with some even dying (Balasubramanian, 2017). Therefore, any abrupt changes in temperature can impact the rate of decomposition and hence the amount of organic matter present in soil. Subsequently, precipitation in the form of rainfall allows for the leaching of minerals and soluble substances from fresh litter to take place (McClaugherty and Berg, 2011). Leaching also allows for the transportation of chemical compounds like dissolved substances or larger materials such as decomposing plant and animal materials throughout soil (Richardson, 2016). The minerals and materials found on the surface become dissolved in the water and is transported to the lower reaches of the soil by the physical movement of water. It is then used by decomposers as a source of food and energy which is used during the decomposition process. Furthermore, rainfall allows for a greater supply of moisture to organisms present in soil (Osman, 2012b). Thereby, this plays a vital role in the growth of microorganisms and the rate of decomposition. However, increased moisture in soil reduces the availability of oxygen in soil required by these organisms. In contrast, decreased moisture content can lead to the desiccation of organisms and reduce the rate of decomposition (Reddy, 2016).

2.2.1 Components of SOM

SOM is a heterogeneous material that is comprised of different components. The major components include; plant and animal residue, soil organisms, stable organic matter (humus), and actively decomposing matter (Reddy, 2016) (Figure 2.4). These components exist in varying proportions and work together to influence the overall SOM content of soil. The components are also subject to soil properties and environmental factors.

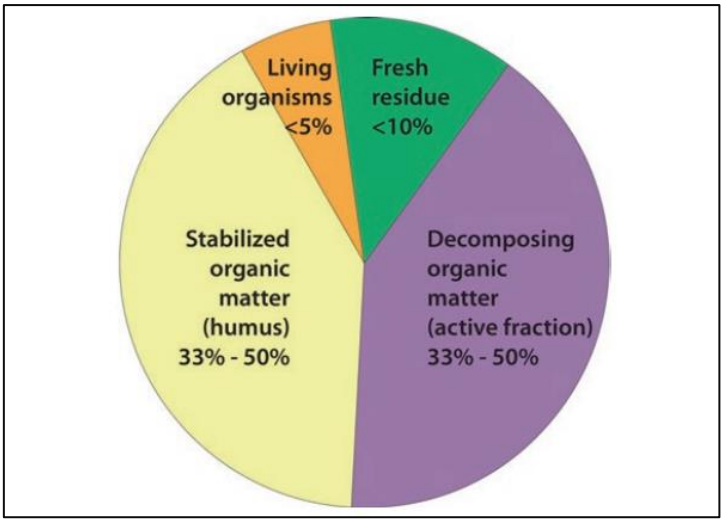


Figure 2.4: The four major components of SOM in varying proportions (Source: Reddy, 2016)

2.2.1.1 Plant and animal residue

The plant and animal residue component include; freshly fallen leaves, twigs, branches, flowers, fruits, roots, stems, and dead animal tissue that has been incorporated into the soil (Osman, 2012b). Once these materials are incorporated into soil, they immediately undergo decomposition. They are easily identifiable in soil as their structures are still morphologically recognizable (Blume *et al*., 2016).

2.2.1.2 Soil organisms

Soil organisms contribute the smallest amount to the composition of SOM (Figure 2.4). They can be divided into living macroorganisms and microorganisms. Macroorganisms in SOM consist of; small animals and predators (ants, insects, mites, beetles, and spiders), herbivores, and detritivores (millipedes, woodlice, earthworms, slugs, snails, rabbits, mice, springtails, ants, beetles, and grubs) (Reddy, 2016). They play an essential role in aerating the soil by burrowing, which allows the exchange of gases and water in soil. Some of these organisms are responsible for breaking up large materials into smaller fragments so that they can be easily decomposed by microorganisms (Blume *et al*., 2016). Macroorganisms also mix, mechanically alter, and add to SOM. Microorganisms, on the other hand, include; detritivores, predators or parasites (nematodes, protozoa, and rotifers), bacteria, actinomycetes, fungi (mushrooms, yeasts, and molds), and algae (Reddy, 2016). They function as a catalyst during the transformation of SOM to nutrients that differ from plant litter (McClaugherty and Berg, 2011). Bacteria and fungi are responsible for decomposing minerals and complex organic compounds that are usually highly resistant. Macroorganisms and microorganisms also work together in

order to increase rates of decomposition as well as allow for the transportation of organic materials throughout the soil.

2.2.1.3 Stable organic matter (humus)

Stable organic matter (humus) contributes the most to SOM. According to Osman (2012b), humus is "a brown to black, amorphous, colloidal organic matter that has undergone decomposition to such an extent that further decomposition is very slow". This state of organic matter is stable but is the most reactive and important component of SOM. It is considered as organic matter that has been fully decomposed to a point where the tissue structure of plant and animal remains can no longer be seen. Humus consists of stable natural organic complexes formed by the products of decomposition and resynthesis (Osman, 2012b). Humus is said to have a low turnover rate, meaning a longer residence time in soil with little to no decomposition (Blume *et al*., 2016). Humus acts as a coating on soil particles, which results in the aggregation of soil particles and fragments. Since humus comprises of proteins that are nitrogenous compounds, it stores and slowly releases nitrogen into the soil (Reddy, 2016). This, in turn, improves the overall fertility of soil. It also has the ability to absorb a large volume of water. Humus may be considered as the final product of decomposition.

2.2.1.4 Actively decomposing matter

Actively decomposing matter is actively being used and transformed by the living plants, animals, and microbes in soil (Reddy, 2016). It includes organic material that has fairly or partially undergone decomposition. According to Osman (2012b), the "fresh, undecomposed and partially decomposed organic matter" collectively form the active SOM. The structure of organic material may still be identifiable. This component of SOM undergoes decomposition much faster whilst also providing a source of food and energy for organisms present in the soil. These components work together to constitute and provide SOM with its characteristic features.

SOM is responsible for supporting many functions in the pedosphere and its interacting spheres. These functions can be grouped as physical, chemical, and biological. These groups also have dynamic interactions between them. An important physical function of SOM includes the formation and enhancement of aggregates in soil (Aoyama, 2015). A soil aggregate is "a group of primary soil particles that cohere to each other more strongly than to other surrounding particles" (Nimmo, 2005). SOM assists soil particles to clump together and form aggregates. According to Fageria (2012), binding agents such as silicate clays, oxides of iron and aluminium, and decomposition products of organic matter (humus) are responsible for

aggregate formation. A unique combination of these binding agents can result in the formation of microaggregates (<250 micrometres [µm]) and macroaggregates (>250 µm) (Li *et al*., 2017). Soil aggregates improve the stability of soil, and hence the overall structure of soil. They form the basis of soil structure, which results in the orderly arrangement of soil particles. Therefore, soil aggregates play an important role in maintaining the porosity and aeration of soil, plant and microbial growth, infiltration of water, and determining both the intensity and rate of erosion (Silva *et al*., 2016). A well-structured soil will allow for the exchange of important gases and water in soil via pores. These gases and water play an important role in sustaining the soil environment, along with the structure of the soil. It also influences the drainage of soil and its ability to retain water, especially during high rainfall periods. According to Finch *et al*., (2014), soils that have a particularly high organic matter content will likely have a more stable soil structure. This, in turn, will determine and influence the many functions that soils are able to support.

Subsequently, SOM plays an important chemical function when it comes to cation exchange capacity. Cation exchange capacity (CEC) refers to the soil's ability to attract, retain, and release various elements and compounds (Tomašic *et al*., 2013). More specifically, it focuses on the manner in which negatively charged soil particles (anions) bind to positively charged atoms (cations) (Brady and Weil, 2008). These positively charged nutrients are released and used by plants. However, a characteristic feature of SOM is that it contains both positive and negative charges (Astera, 2010). This means that it can attract and retain both cations and anions. More specifically, humus colloids contain dissociable groups such as; amino ($-NH_2$), hydroxyl (-OH), and carboxyl (-COOH) (Osman, 2012b). This means that humus contains electrically charged particles. In addition, these humus colloids have a larger surface area than compared to clay particles in soil (Astera, 2010). Therefore, allowing for a greater number of negative charges along its surface to which positively charged nutrients are attracted and retained. Subsequently, these colloids contain positive charges, to which negatively charged nutrients are absorbed and exchanged. These capabilities, therefore, suggests that humus has a very high CEC. According to Astera (2010), humus colloids can retain and exchange many times its weight in regards to water and plant nutrients. As a result, this improves the overall fertility of soil and its ability to sustain the growth and development of plants. Other chemical functions of SOM include; buffering, pedochemical weathering, chelation, and the translocation of substances within soil (Osman, 2012b).

Additionally, SOM's biological functions include; providing food sources for soil organisms and a reservoir of nutrients (Osman, 2012b). Regarding the provision of food for soil organisms, when fresh organic material is combined into the soil, it is immediately attacked by soil microorganisms and detritivores. They feed on and break down organic material, which serves as a source of carbon and energy. More specifically, the supply of carbon by organic material allows these organisms to build new cells and tissue (Hoorman and Islam, 2010). In addition, the organic material also provides them with energy that is required during their growth and development in the soil.

In contrast, SOM is responsible for providing plants with the nutrients required for growth and development. It serves as both a source and sink of plant nutrients (Aoyama, 2015). These nutrients are obtained and stored in soil via nutrient cycling. According to Osman (2012c), "nutrients are taken up, some proportion is retained in biomass, some proportion is returned to the soil as litter, and nutrients are released by decomposition of litter materials". This then forms a nutrient pool in the soil, from which plants can absorb and use these organic nutrients. According to Osman (2012b), SOM is responsible for providing more than 90% of nitrogen and more or less 50-60% of phosphorous and sulphur. These are considered as macronutrients, whilst micronutrients include; iron, manganese, zinc, copper, boron, molybdenum, etc. Plants require sixteen essential elements that are required for their growth and development (Uchida, 2000). However, the nutrients are released at a slow pace to ensure that the roots of plants are able to keep hold of these nutrients, so that nutrient loss is relatively low in an organic rich soil. Subsequently, SOM has the ability to retain SOC in its decomposing plant and animal material (Gahlod *et al*., 2019). SOM can retain approximately 58% carbon, which is more than both the atmosphere and terrestrial vegetation combined (Donovan, 2013). This carbon is responsible for improving the activity of microorganisms in soil and providing plants with their nutrients, thus, increasing the biodiversity of the soil environment. It can also be used as a means to mitigate and adapt to climate change and improve ecosystem functioning.

The functions mentioned above of SOM are also responsible for directly or indirectly influencing the ecosystem services provided to humans, plants, and animals. SOM is usually located in the A Horizon of soil (Appendix A), where topsoil is commonly found. This outermost layer of soil (13-25 centimetres [cm]) contains the highest concentration of SOM. More specifically, it contains roughly 1-6% of SOM, whereas the subsoil contains less than 2% of SOM (Adams *et al*., 2011). Despite the small amount of SOM found in topsoil, it plays

significant roles in the overall fertility and productivity of soil. However, human activities such as agriculture coupled with the increasing growth in the human population have led to the degradation of fertile topsoil. More specifically, approximately 25% of the earth's surface has become degraded, with 12 million ha of topsoil being lost every year (FAO, 2015b). Therefore, indicating a loss in the total SOM content present in soil. This pattern will continue into the future until soils are no longer able to support the many activities carried out by humans along with supporting ecosystem services and functions.

Due to SOM being an integral part of the overall functioning, fertility, and productivity of soil, it is essential that SOM content be measured and monitored. Therefore, SOM measurements have also been commonly used as the basis for evaluating the quality of soils (Roper *et al*., 2019). Over the years, several techniques have been established and used in the analysis of SOM content (Motsara and Roy, 2008). Such a case includes laboratory-based approaches, which allows for SOM content to be estimated accurately under controlled conditions. Laboratory-based techniques are still commonly used and include; mass lost on ignition (Loss on Ignition [LOI]), volumetric measurements (Walkley-Black [WB]), and colorimetric measurements (Humic matter [HM]). Dry combustion methods such as the LOI method involve soil samples being burned at high temperatures, and the organic matter content is determined from the weight lost (Ivezić *et al*., 2016). In contrast, wet combustion methods such as the WB and HM method involve the oxidation of organic matter by using acids and measuring the evolved carbon dioxide by gravimetric, titrimetric, or manometric methods (Chatterjee *et al*., 2009). Each of these methods has its own strengths and weaknesses when it comes to determining SOM content. However, the most widely used methods include the LOI and WB method, with the WB method being more favoured. This maybe a result of the LOI method under-or-overestimating the amount of organic matter present within the soil, which can result in unreliable and erroneous results (Ivezić *et al*., 2016). The WB method, on the other hand, is a simple and inexpensive that produces accurate results (Matus *et al*., 2009; Wang *et al*., 2012).

Similarly, the soils in South Africa are usually not characterized by high fertility and become easily degraded with the exception of certain soils. South Africa is known to cover approximately 121.9 million ha of land, and only 13% of this land is suitable for activities such as agriculture (Mohamed, 2000). This being an indication of degraded SOM over the years. This degradation in SOM over the years is attributed to the improper use and management of soil. The situation is more likely to worsen in the near future if no measures are put into place.

It is therefore imperative to assess SOM to ensure that it is able to support and facilitate different uses and activities for all its users in the various spheres. Therefore, the current study will aim to examine the total SOM content present within the soil of the Emakhosaneni area located in KwaZulu-Natal (KZN), South Africa, using laboratory techniques.

2.3 Materials and methods

2.3.1 Description of the study area

South Africa comprises of nine provinces. The selected province in which the study will be conducted is the KZN province, located along the east coast of South Africa and is characterised by varying levels of diversity. The KZN province occupies an area of 94 361 Square kilometres (km^2), which is 7.7% of South Africa's land area and has a population of 10 267 300 people (SSA, 2011). The distribution of household types in KZN includes formal (71.6%), traditional (19%), and informal (8.3%) (SSA, 2011). The province of KZN consists of 12 district municipalities that can further be subdivided into 55 local municipal areas. The study was conducted in the Emakhosaneni area that is located within the Okhahlamba Municipality, KZN.

The Okhahlamba Municipality (Figure 2.5) is located in the western region of KZN and is approximately 242 Kilometres (km) from the city centre in the eThekwini Metropolitan Municipality. It is situated in the mountainous regions of KZN between Lesotho, Free State, Emnambithi, and Mtshezi (SSA, 2011). The total population of individuals present within the municipal area is 132 068 and are distributed over an area of 3,970.98 km^2 with an average density of 33.26 individuals per km^2 (Frith, 2011). The area comprises mainly of traditional/tribal households with a small percentage of formal housing. The area contains a diversity of race groups that contribute to the growing population and include Black African (97.1%), Indian (0.4%), White (2.1%), and Coloured (0.2%) individuals (Frith, 2011). The local municipality gives rise to many towns and settlement areas, but Emakhosaneni was selected as the study area for this particular study.

The Emakhosaneni area is located in the central region of the Okhahlamba municipality (28°46'45.1"S; 29°14'42.4"E) (Figure 2.5). The area has a total population of 1 938 individuals that are distributed over an area of 13.72 km^2 and has an average density of 141.30 individuals per km^2 (SSA, 2011). Despite being a large area, there are only 380 households found within the area that are sparsely dispersed. All households within the area are traditional/tribal. The area also mainly comprises of Black African individuals (99.9%) (SSA, 2011). It is situated approximately 17.5 km from the nearest town, Bergville.

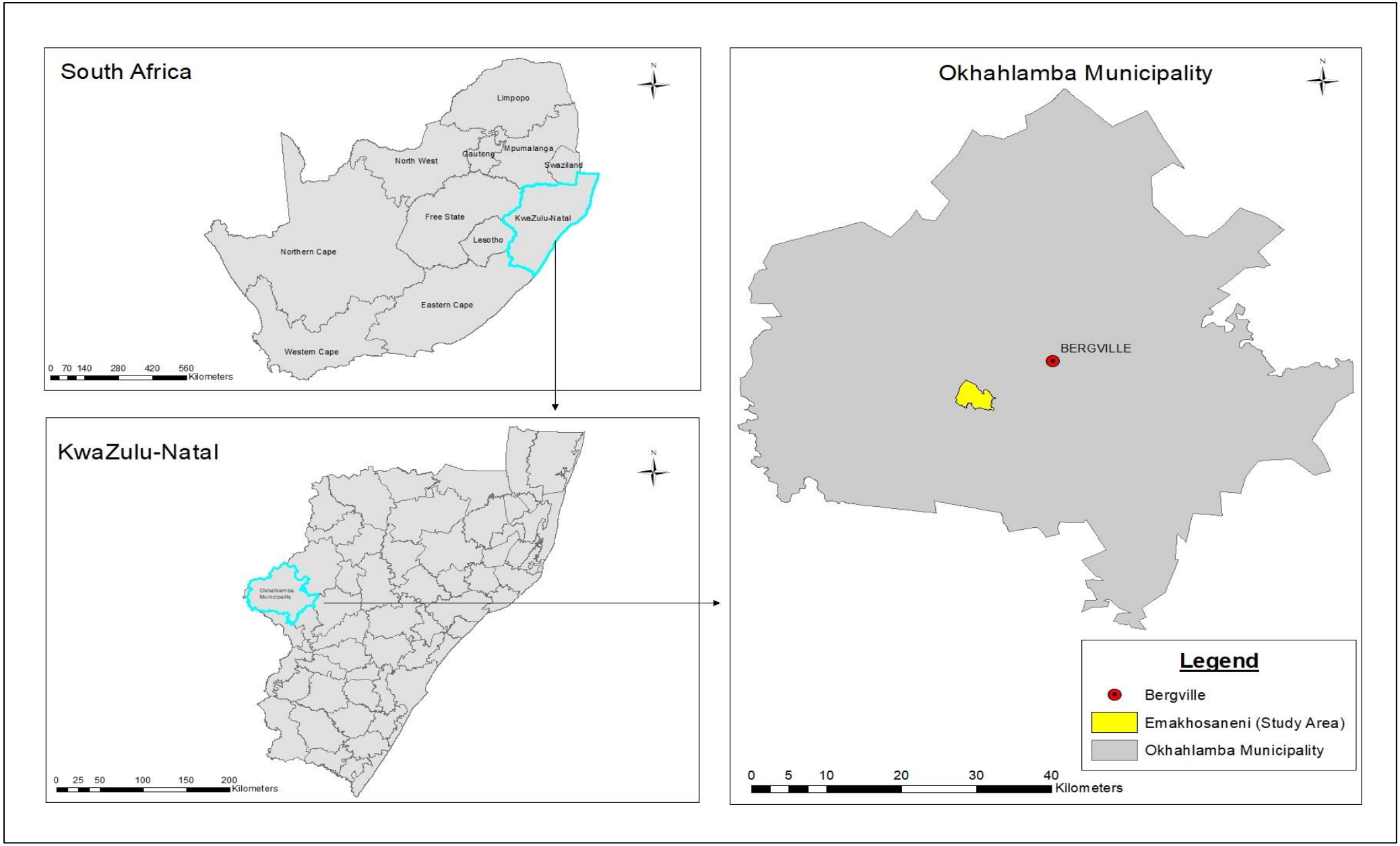


Figure 2.5: The selected study area map within the Okhahlamba Municipality of KZN (Author, 2020)

The climate in the study area is considered as being warm and temperate. The monthly average midday temperatures range from 19 degrees Celsius (°C) in the winter months like July to 29°C in the summer months like February (Weather Atlas, 2020). However, rainfall is experienced throughout the year by receiving an average of 644 mm of rain per year, with the highest rainfall experienced during the month of January (120 mm) and the lowest rainfall is during the month of July (5 mm) (Weather Atlas, 2020). Climatic conditions such as temperature and rainfall are influenced by the warm Agulhas and Mozambique oceanic currents. The wind speeds vary over the months, with the winter and spring months experiencing the highest wind speeds, while the summer and autumn months experience the lowest wind speeds.

Furthermore, the area is located in a mountainous region. As one moves down the mountainous landscape, the gradient begins to become less steep. Along the base of the mountain, the gradient begins to decrease gradually until a gentle to no gradient is seen. This extends for several kilometres. Also, an increase in the number of households is seen as one moves away from the mountainous regions of the area.

Additionally, the area is surrounded by various water sources. The Woodstock dam is located north of the study area, while to the eastern regions, the Tugela River and Drieldam are found (Google Earth, 2020). The river and dams serve as a reliable source of water that can be easily accessed and therefore supports its surrounding land use types. Two types of geology that constitute the study area are; Adelaide and Tarkastad (South African National Space Agency [SANSA], 2015). The area is dominated by agricultural practices on both a large and small scale. Large areas of land are cleared from their natural vegetation in order to make land available for the expansion of the settlement and for agricultural practices. Subsequently, large areas of land are exposed to erosion by both wind and water as a result of natural vegetation being cleared. With such extensive farming occurring in the area, the appearance of land fragmentation is seen. There are a few sparse trees that remain after the land was cleared from its natural vegetation around the area, but the highest densities are seen in those parts that are inhabited by community members. Lastly, there are a number of road networks found throughout the area that aid access to various parts into and out of the area.

2.3.2 Soil sample collection

The selected study area covers a large area. Therefore, a thorough examination of the study area was conducted before sample collection. This was done using satellite imagery via Google Earth (Google Earth, 2020) and an aerial photograph of the study area followed by field reconnaissance. This led to the identification of four major land uses/land cover that constitute the study area; Agriculture, Built-up (Residential), Eroded, and Rangeland.

Soil samples were collected from the 16th July to 19th July 2019, from 08:00 am to 15:00 pm. There was no extreme weather (e.g. Strong winds or rainfall) experienced during the sampling process. A stratified random sampling method was used to determine the sampling points within the study area (Figure 2.6). This was done by firstly identifying the major land use types within the study area and then randomly determining sampling points within these land use types. It is a commonly used method that is designed to ensure that the samples taken are representative of the area and eliminates any biasedness (Worsham *et al*., 2012). These sampling points were also selected placed on their accessibility within the study area. Thirteen random soil samples were obtained from each of the four major land uses (Agriculture, Built-up (Residential), Eroded, and Rangeland), which amounts to a total of 52 soil samples being obtained from the study area.

These soil samples were collected from the soil's surface to a depth of 15 cm using a soil auger. The collected soil sample was then carefully placed into a clear plastic bag that is airtight in order to prevent any spillage or drying of the sample (Figure 2.7a). Careful attention was taken during these steps in order to avoid any contamination of the sample. After collection, the sample was labelled appropriately, and the Global Positioning System (GPS) coordinates were recorded. Subsequently, any other observations regarding the sampling point were recorded. However, at certain sampling points, the soil auger could not penetrate the soil's surface, and instead, a spade was used to obtain the soil sample. The above-mentioned procedure was replicated for the collection of soil samples from the 52 sampling points. Due to the study area being quite far from the laboratory, the soil samples were stored in a cool and dry cooler box, away from direct exposure to sunlight. Once all 52 soil samples were obtained, they were transported to and stored in the laboratory, where the analysis would begin.

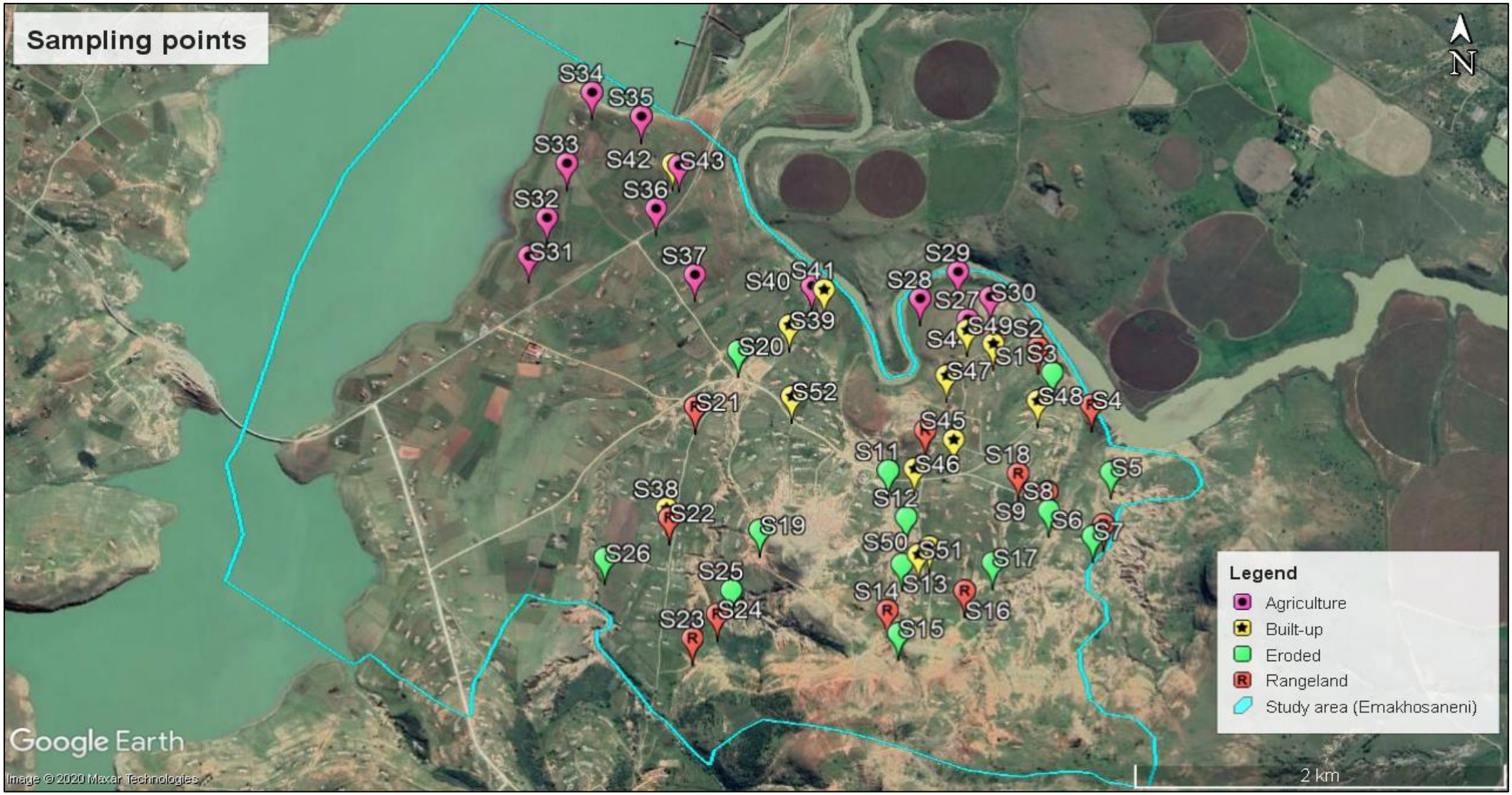


Figure 2.6: The sampling points from which soil samples were obtained within the four major land uses (Author, 2020)

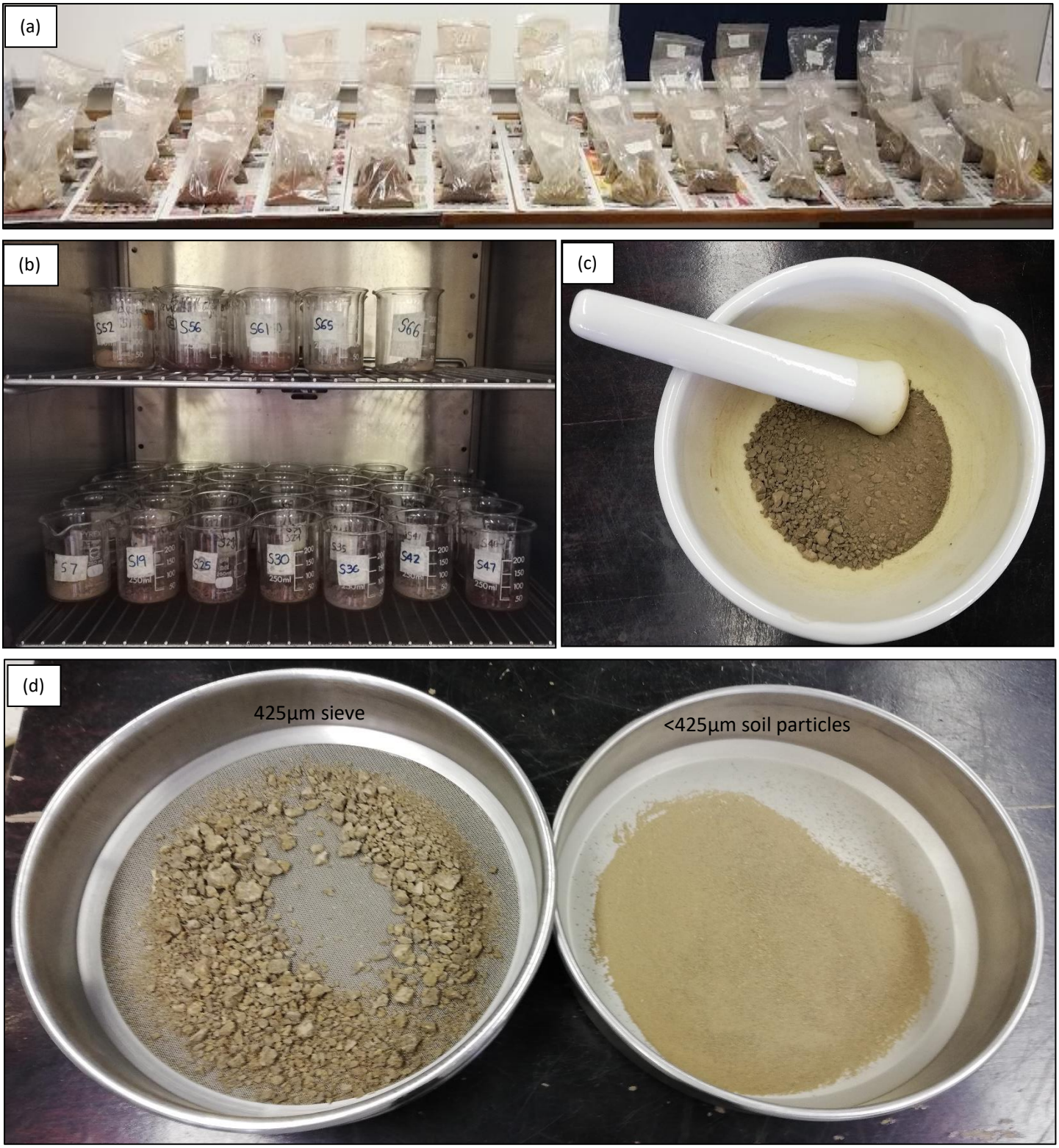


Figure 2.7: Storage of samples (a) and the soil sample preparation, which involved (b) oven drying of soil, followed by (c) crushing of dried samples, and (d) the sieving of samples (Author, 2020)

2.3.4 Soil sample preparation

The collected samples were first placed into their respective labelled glass beakers (20 grams [g] of soil) and placed into the oven for drying. They were oven dried overnight at a temperature of 105°C (Figure 2.7b). This process ensures that no moisture and microbial activity is present in the soil during analysis. Once oven dried, the samples were crushed into fine soil particles using a porcelain mortar and pestle (Figure 2.7c). The crushing of soil samples assists in breaking down soil aggregates or large fragments into smaller fragments or particles. The crushed samples were then passed through a 425 µm stainless steel sieve. The material that remained within the sieve was discarded, while the soil particles that passed through the sieve (<425 µm) was stored for analysis (Figure 2.7d). Sieving ensures that large and foreign materials such as roots and stones are not included in the analysis and allows for a more homogenous sample to be analysed. The mortar and pestle, and sieve were thoroughly cleaned in between samples in order to prevent any contamination between samples. The sieved samples were then transferred into their respective labelled glass beakers and covered with Parafilm M until analysis (Appendix B).

2.3.4 Soil analysis

Various methods are used to measure and estimate SOM content in soils (Motsara and Roy, 2008). However, this study focused on using one of the most common methods to determine the total SOM content in the soil samples, known as; the WB method. The WB is a wet oxidation method that involves the use of acids to determine the amount of oxidizable carbon in soil; this amount is then converted to SOM using a conversion factor (Walkley and Black, 1934).

The original procedure developed by Walkley and Black (1934) was used, however, with some minor modifications. Approximately 0.3 g of oven dried soil was placed in a 500 millilitres (ml) conical flask, to which 10 ml of potassium dichromate (1N $K_2Cr_2O_7$) was added and swirled gently (Figure 2.8a and b). Thereafter, 20 ml of concentrated sulphuric acid (H_2SO_4) was carefully added and swirled gently. This was then left to stand for 30 minutes in a fume hood (Figure 2.8c). During this reaction, heat is generated as a result of the H_2SO_4 mixing with the $K_2Cr_2O_7$, along with harmful fumes (Bianchi *et al*., 2008). After 30 minutes, 200 ml of distilled water was added to the conical flask and swirled (Figure 2.8d). This was then left to stand for a few minutes. This step slowed down the reaction and assisted in cooling down the solution. Thereafter, 10 ml of concentrated orthophosphoric acid (H_3PO_4) and 1 ml diphenylamine (indicator, $(C_6H_5)_2NH$) was added to the conical flask. This had changed the

colour of the solution to a dark black (Figure 2.8e). The indicator was added prior to titration to prevent deactivation caused by adsorption onto clay surfaces (Schulte and Hoskins, 1995). Once the indicator was added, the solution was immediately titrated against a standard solution of ferrous ammonium sulphate (0,5N $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$) in small increments, whilst swirling the flask until the solution changed colour from black to green (Figure 2.8e). This colour change indicated the endpoint of the titration. The total volume of ferrous ammonium sulphate used was recorded (S in equation 1). Prior to titration, a reagent blank (i.e. no soil) was run in order to standardize the ferrous ammonium sulphate using the above procedure, and the volume of it required to produce the colour change was recorded (B in equation 1). The above procedure was carried out for all soil samples, and the amount of SOM was calculated using the following equation (equation 1);

$$\text{SOM \%} = \frac{10 \text{ x } \left(1-\frac{S}{B}\right) \text{ x } 0.39 \text{ x } 1.72}{W}$$

Where,

10 = Volume of potassium dichromate used

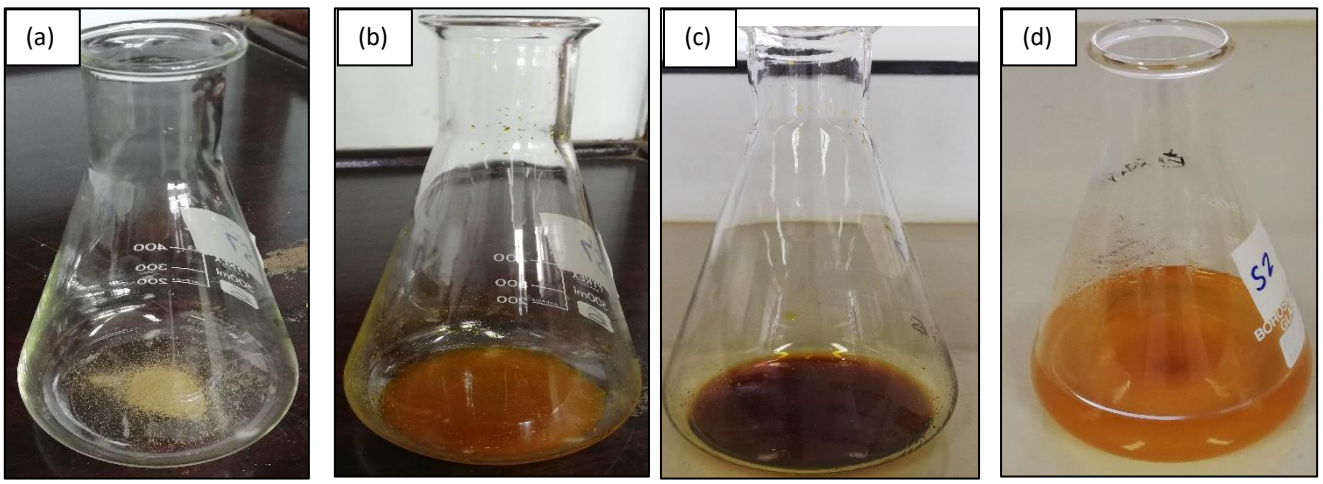
S = Total volume of ferrous ammonium sulphate used in titration with soil

B = Total volume of ferrous ammonium sulphate used in blank titration

0.39 = 12/4000 = (0.003) milliequivalent weight of C in g;
1.3 = oxidation correction factor, therefore,
0.003 x 100 x 1.3 = 0.39

1.72 = Conversion factor (organic matter contains 58% organic C, hence 100/58 = 1.72)

W = Weight of soil (g)



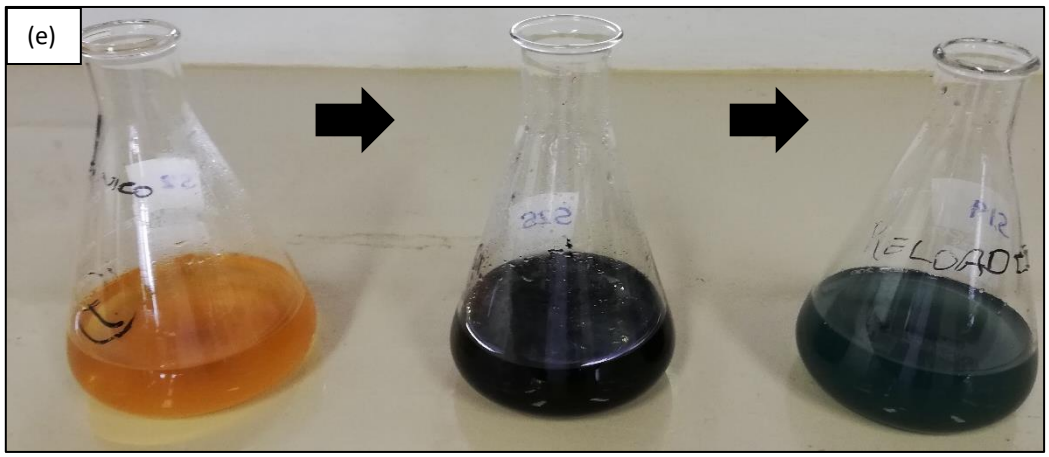


Figure 2.8: WB method showing (a) dry soil sample, (b) $K_2Cr_2O_7$ mixed with soil (c) $K_2Cr_2O_7$ and H_2SO_4 with soil, (d) diluted solution, and (e) final outcome of WB method with colour changes being evident (Author, 2020)

2.3.4 Statistical analysis

Once the percentage of SOM was derived, descriptive statistics, such as the means and standard deviations, were calculated and used to summarize the data. Inferential statistics, such as a one-way Analysis of Variance (ANOVA) was conducted in order to determine whether there are any statistically significant differences in the percentages of SOM within the four major land use types. The fixed factor for the ANOVA was land use type with four categories; agriculture, built-up, eroded, and rangeland, while the percentage of SOM was the dependent variable. Before carrying out the one-way ANOVA, a set of assumptions, such as the data being normally distributed and having an equal variance across all treatments, were fulfilled. This was done using a Shapiro-Wilk test and Levene's test, respectively. Furthermore, a post-hoc Tukey's test was conducted in order to determine exactly where the differences in SOM lie in each of the land use types. A significance level of 0.05 was used for this study.

The Statistical Package for Social Science (SPSS) version 25 software was used for this study. This software assisted in the performance of appropriate statistical tests along with descriptive statistics. The creation of tables was done using the SPSS software.

2.4 Results and Discussion

The descriptive statistics of the SOM content within the study area and its major land uses are displayed in Table 2.1. In Table 2.1, it is evident that the agriculture land use has the highest average SOM content (3.42%). While eroded land contains the lowest average SOM content (2.19%). An average difference of 1.23% in SOM exists between these land uses. The average SOM content in land uses such as rangeland and built-up lie between the highest and lowest percentages of SOM. On the other hand, the standard deviation exhibits a different trend; in that eroded land has the highest standard deviation (1.26) while built-up land use has the lowest (0.98). With regards to the minimum and maximum percentages of SOM found in each of the different land uses, agriculture had SOM percentages that were higher compared to the others. The highest minimum percentage of SOM was 2.23% (agriculture), and the lowest was 0.00% (eroded). However, agriculture (5.96%) had the highest percentage SOM, while rangeland (4.62%) had the lowest. Regarding the overall average of SOM content in the study area, an amount of 2.79% was seen. The lowest percentage was 0.00%, and the highest was 5.96%.

Table 2.1: Descriptive statistics indicating the overall percentage of SOM in the study area and its major land uses (N= Number of samples, Std= Standard deviation)

	N	Mean	Std. deviation	Minimum	Maximum
Overall	52	2.79	1.19	0.00	5.96
Agriculture	13	3.42	1.03	2.23	5.96
Built-up (residential)	13	2.50	0.98	0.74	4.76
Eroded	13	2.19	1.26	0.00	4.91
Rangeland	13	3.05	1.18	1.34	4.62

From the results obtained, the average percentage of SOM within the Emakhosaneni area is moderately low compared to the average amount that is usually found in soils. The concentration of SOM usually ranges from 1-5% in most mineral soils (Kalev and Toor, 2018). SOM concentrations are known to be the highest in the topsoil layer (13-25 cm) and decrease with depth. The soil samples analysed for this study were obtained from the same layer, thus indicating that the topsoil within the Emakhosaneni area contains approximately 2.79% of SOM (Table 2.1). According to Du Preez *et al*., (2011), soils within South Africa are characterised by low organic matter levels, which may be attributed to the low SOC content found in topsoils that usually range from <0.5% to >4%. This low supply of carbon in the soil can directly influence the SOM content present within the soil, as SOC constitutes 58% of SOM (Donovan, 2013). Therefore, lower SOC levels present within the soils of the study area

may have resulted in a lower SOM concentration. Furthermore, the soils in different regions or areas are highly influenced by their soil-forming factors; climate, organisms, relief (topography), parent material, and time (Jenny, 1994). For instance, South Africa is considered a semi-arid region that receives less than 500 mm of rainfall (SAWS, 2019). Under such conditions, the supply of moisture to the soil and its organisms is lowered (Osman, 2012b). This may reduce microbial activity within the soil and slow the decomposition process down, hindering SOM formation and accumulation. Similarly, the soil samples were collected during the month of July, when rainfall in the area is at its lowest. This, coupled with relatively warm temperatures, may have resulted in lower SOM concentrations in the study area during sampling. Therefore, resulting in the moderately low SOM content that was obtained during analysis.

Regarding the SOM content found within each of the major land uses, it is evident that they show varying percentages. The agriculture land use had the highest average percentage of SOM (Table 2.1). One of the reasons for this is that crops are regularly grown within agricultural land uses; therefore, there is a regular supply of organic material in the form of leaf litter and stems that is added to the soil (Behera and Prasad, 2020). As mentioned previously, these materials are one of the major sources of organic matter in soils. Therefore, a relatively good supply of these materials will facilitate the decomposition process and thus the formation of SOM. Similarly, one of the major crops grown within the study area is maize. These crops are cultivated on a small scale within the agricultural land use. As these crops grow, maize litter such as; culms, leaves, and sheaths fall off and are usually left on the soil's surface or incorporated into the soil (Kamota *et al*., 2014). These litters then undergo decomposition at distinct environmental conditions within the soil, which results in the formation of SOM. However, these litters break down at varying rates and release important nutrients such as carbon into the soil over time. The decomposition of litter is also dependent on the activity and community of decomposers that are present within the soil (Hou *et al*., 2019). This, in turn, will also influence the rate at which SOM is formed and accumulated in the soil. Subsequently, the living roots of the maize crops are responsible for stimulating the decomposition of SOM. According to Kumar *et al*., (2016), labile carbon compounds are secreted by roots into the soil, enhancing both microbial growth and activity, thus leading to greater extracellular enzyme activities within the soil and, therefore, increasing SOM decomposition. Subsequently, maize plants have a fibrous root system that is located close to the surface of the soil, where it forms a dense network of roots.

When these maize plants have reached senescence, the dead roots contain nutrients (carbon, nitrogen, phosphorous, etc.) that are then released into the soil, thus, increasing the microbial activity and rate at which organic material is broken down and converted into SOM. They also serve as a source of organic material for SOM formation. Another reason for the relatively high percentage of SOM seen within the agricultural land use is crop residues. Crop residues consist of materials that are left behind after the crops have been harvested (Cherubin *et al*., 2018). Like plant litter, it also includes the remains of crops such as leaves, stems, stalks, seed pods, etc., and plays a similar role in influencing and enhancing SOM. Crop residues are usually retained on the soil's surface, where they act as a buffer against changes in the environment, provide a food source for soil organisms, and limit evaporation from the soil surface (Nowatzki, 2019). According to Lal (2009), crop residues are considered a rich source of carbon, and their addition to soil is known to enhance SOM content. Similarly, within certain agricultural sampling points, maize crop residues were retained on the soil's surface (Appendix G). This may have increased microbial activity in the soil, thus, increasing the rate of SOM formation. The residue also served as a protective layer on the soil's surface to reduce the effects of erosion, retain moisture within the soil, and ensure that temperature changes were not too extreme on soil microorganisms. Therefore, ensuring that SOM content is retained during the fallow period where cool, dry soil conditions are experienced (May to August) (Kamota *et al*., 2014).

Additionally, the relatively high percentage of SOM within the agricultural land use can be a result of the manner in which the soil is treated. The soil within the agricultural land use is tended to or taken care of regularly to ensure a greater yield of crops produced. For instance, the soil receives enough moisture during irrigation, which facilitates the breakdown of organic materials on the soil's surface and its transportation within the soil. It also prevents the desiccation of soil organisms and ensures the decomposition process is facilitated (Osman, 2012b). Thus, allowing for the production and maintenance of SOM within the soil. Subsequently, many farmers add an additional source of organic matter into the soil via the application of manure (Appendix E). Manure is an organic material that is produced from animal waste; and contains important nutrients derived from plant material (Mani, 2011). It is responsible for releasing these important nutrients back into the soil, as well as providing food for the microorganisms in the soil. More importantly, it plays an important role in enhancing SOM content within the soil (Bot and Benites, 2005). These additional treatments of soil ensure proper maintenance and accumulation of SOM content.

In contrast, the average percentage of SOM within the built-up land use was relatively low (2.50%) (Table 2.1). The low percentage can be supported by the fact that natural vegetation is cleared from the area in order for homes to be constructed. Natural vegetation or any plant grown along the soil's surface plays an important role in acting as a protective layer against environmental and anthropogenic factors. Such ground cover generates and enhances SOM by promoting biological activity and protecting the soil from wind, rain, and temperature extremes (Hoyle, 2013). Therefore, removing this plant layer during the construction of homes puts the soil at risk of losing SOM. Also, during the clearance of natural vegetation, topsoil is removed in order to level the land so that a proper foundation can be built for homes (Davis and Luttrell, 2016). During this process, SOM is being removed, and its content is reduced within the soil.

Furthermore, the removal of both vegetation and topsoil leaves the soil's surface exposed and susceptible to erosion and climatic factors, which further leads to the degradation of SOM. Anthropogenic factors may also influence the low percentage of SOM seen within the built-up land use. Within built-up land uses such as residential areas, there is heavy traffic from humans walking and driving vehicles along the soil's surface, thus putting physical pressure onto the soil and causing it to compact (Appendix I). Soil compaction is a process whereby soil particles are compressed together, thereby reducing the pore spaces between them (Idowu and Angadi, 2013). The reduced porosity in the soil can negatively influence the entry of gases, nutrients, and water into the soil. Soil compaction changes the soil's structure, which in turn influences the physical, chemical, and biological processes of soil (Nawaz *et al*., 2013). It also affects the carbon cycle, which alters biological activity in the soil and causes the SOM content to decrease (Rai *et al*., 2018).

Additionally, the regular traffic of humans and animals along the soil's surface can also result in the degradation and removal of ground cover (Appendix I). This can leave the soil bare and vulnerable to rainfall, resulting in soil crusting and sealing, which is quite common in South African soils (Laker and Nortjé, 2019). Infiltration of water is restricted under such conditions, and therefore, a loss or decrease in SOM content is seen (McKenzie, 2010). These reasons, coupled with population growth, can further result in the loss of SOM within the built-up land use.

Moreover, eroded land had the lowest average SOM percentage, where an amount of 2.19% was observed (Table 2.1). Soil erosion is defined as "the wearing away of the land surface by physical forces such as rainfall, flowing water, wind, ice, temperature change, gravity or other

natural or anthropogenic agents that abrade, detach and remove soil or geological material from one point on the earth's surface to be deposited elsewhere" (Doula and Sarris, 2016). Soil erosion is a natural process; however, it is exacerbated by anthropogenic activities. During the process of erosion, large amounts of topsoil are either removed and deposited or lost. More specifically, soil erosion by water or wind results in the loss of organic-rich topsoil (Arriaga *et al*., 2017). For instance, the study area, on average, experiences 644 mm of rainfall throughout the year, with some months experiencing a higher rainfall and some experiencing a lower rainfall (Weather Atlas, 2020). During rainfall periods, raindrops strike the soil's surface and dislodge soil particles, which are then transported over small distances. This can break up soil aggregates or fragments into smaller particles. Rainwater, in the form of surface runoff, picks up these soil particles and keeps them in suspension, and as it continues to flow, more soil particles are picked up from the soil's surface (Osman, 2014a). This process is exacerbated under high rainfall conditions, which increases the rate of topsoil loss. According to Phuong *et al*., (2017), soil erosion leads to the loss of fertile topsoil and SOM. The topsoil is where biological activity and decomposition is most concentrated; therefore, a loss in this layer would mean a decrease in SOM formation and accumulation (Biswas and Naher, 2019). According to Le Roux (2014), an average of 12,6 tons/ha/year of soil is lost in South Africa annually.

Furthermore, soil erosion is influenced by the topography of an area and the presence of vegetation. As mentioned previously, ground cover plays an important role in acting as a protective layer for the soil's surface. And, within the study area, it was observed that the majority of the eroded land had little to no ground cover (Appendix J). This means that there is a lack of organic material for SOM formation as well as protection from topsoil loss and hence, the maintenance of SOM. Subsequently, since the study area is located in a mountainous region, a relatively steep to gentle slope was seen (Appendix J). This can influence the rate of runoff during rainfall and, thus, the rate at which topsoil and SOM are lost. Another reason for the low SOM content found within the eroded land is a result of overgrazing. The rearing of livestock such as cattle, goats, and chickens was quite common within the study area, together with overgrazing being evident (Appendix E). Overgrazing is one of the many factors that accelerate the rate of soil erosion (Osman, 2014a). Overgrazing results in the destruction and removal of ground cover through eating and trampling by livestock, disturbance of the root system by scuffling, and the compaction of the soil surface, thus, reducing infiltration rates and increasing runoff and soil erosion (Sharma, 1997). These effects of overgrazing directly lead to the degradation and loss of SOM, which can support the low percentage of SOM found

within the eroded land. Also, it was evident that different types of erosion such as; sheet, rill, and gully erosion were quite distinct within this land use and also had an influence on the percentage of SOM (Appendix J).

Furthermore, within the rangeland land use, the average SOM percentage was fairly high (3.05%) (Table 2.1). This type of land use is quite dominant within the study area. These types of lands consist of natural vegetation, which mainly includes; grasses, grass-like plants, forbs, or shrubs that is suitable for grazing or browsing purposes (Environmental Protection Agency [EPA], 2020). The presence of extensive vegetation cover like grass is responsible for the fairly high percentage of SOM found within the rangeland land use (Appendix H). According to Haynes *et al*., (2003), the turnover of dense grass root systems in rangelands is one of the main sources of organic material. Grasses have a thin, fibrous root system that is located close to the surface of the soil, where it forms a dense network of fine roots. Since these roots are found below the soil's surface, they become readily available to soil organisms for decomposition. Given the structure of these root systems, they can easily be decomposed within the soil by microorganisms and therefore contribute to the SOM content. Seeing as there is a relatively good supply of organic material coupled with an intensified rooting system within this land use, it is, therefore, able to support an extensive microbial community in the rhizosphere (Haynes *et al*., 2003). Thus, facilitating the SOM formation and maintenance process.

Subsequently, the turnover of roots signifies the direct input of organic matter into the soil and the ability to contribute significantly to SOM accumulation (Hoyle, 2013). More specifically, when these grasses reach senescence, they leave behind their dead roots, which are then broken down and converted into SOM. Furthermore, the roots of these grasses contribute to SOC pools via rhizodeposition within soils and have a longer residence time of carbon than compared to shoot tissue (McGranahan *et al*., 2014). Thus, they have the ability to store large amounts of carbon, which can influence aggregate stability and soil structure and, therefore, the formation and maintenance of SOM (Zhou *et al*., 2020). Given the fact that the fibrous root system of grasses contains several main roots that branch off to form a compact network of intermeshed lateral roots, which play an important role in binding soil particles to form aggregates (Kafle and Balla, 2008). The physical pressure of roots causes soil particles to bind. This, coupled with root exudates, such as mucilage, assist in binding soil particles to form aggregates (McNear, 2013). Therefore, improving the overall soil structure and SOM content of the rangeland land use. Another reason for the fairly high percentage of SOM may be attributed to the grazing of livestock. When these animals graze, they leave behind animal dung which can

influence the SOM content, as mentioned previously (Haynes *et al*., 2003) (Appendix E). Additionally, these soils are usually left undisturbed from anthropogenic factors; therefore, there is a lower chance of SOM degradation and loss within the rangeland land use.

The results obtained from the one-way ANOVA revealed that the percentage of SOM within each of the land uses were statistically significant ($F=3.180$, $p<0.05$). This indicates that the percentage of SOM within each of the land uses varies and that they are different from one another. These results also suggest that the land use type had an influence on the amount of SOM present within the study area at the time of sampling. The above-mentioned reasoning can further support these results. However, the results from the one-way ANOVA did not indicate where the significant difference lies. More specifically, it did not reveal which of the land uses were significantly different from each other. Therefore, a post-hoc Tukey's test was conducted in order to determine where the difference lies in each of the land use types. The results obtained from the post-hoc Tukey's test are presented in Table 2.2.

In Table 2.2, when the rangeland land use was compared to the other land uses, their *p* values were all greater than 0.05 (i.e., $p>0.05$). Comparing rangeland to agriculture had the highest significance value ($p=0.824$), while rangeland to eroded had the lowest value ($p=0.219$). However, the comparisons between the eroded land and other land uses had a *p* value less than 0.05. This was evident between the eroded and agriculture land use ($p=0.034$). Whereas the comparison between eroded and built-up land use had the highest overall value ($p=0.897$). Furthermore, the comparison between built-up land use and the other land uses had a similar trend to the rangeland land use, where all land uses had values of $p>0.05$. The comparison between built-up and agriculture land use had the lowest value ($p=0.161$). Lastly, the comparisons between the agriculture land use and other land uses had a *p* value less than 0.05 (i.e., $p<0.05$). And, this was evident between the agriculture land use and eroded land ($p=0.034$).

Table 2.2: The post-hoc Tukey's test showing the multiple comparisons of SOM across the different land use types (*p<0.05; std.=standard error; sig.=significance)

(I) Land use	(J) Land use	Mean Difference (I-J)	Std. Error	Sig.
Rangeland	Eroded	0.85769	0.43830	0.219
	Built-up	0.55154	0.43830	0.593
	Agriculture	-0.37769	0.43830	0.824
Eroded	Rangeland	-0.85769	0.43830	0.219
	Built-up	-0.30615	0.43830	0.897
	Agriculture	-1.23538*	0.43830	0.034
Built-up	Rangeland	-0.55154	0.43830	0.593
	Eroded	0.30615	0.43830	0.897
	Agriculture	-0.92923	0.43830	0.161
Agriculture	Rangeland	0.37769	0.43830	0.824
	Eroded	1.23538*	0.43830	0.034
	Built-up	0.92923	0.43830	0.161

The results obtained from the post-hoc Tukey's test revealed that the majority of the comparisons were not statistically significant (Table 2.2). For instance, the land use types rangeland and built-up were not statistically significant with each other and the other two land use types. This means that differences in the percentage of SOM between the built-up and rangeland land use are not statistically significant. However, it was found that only the land use types agriculture and eroded were statistically significant with each other and not with the other two land use types. More specifically, it is an indication that the percentage of SOM between the agriculture land use and eroded land is statistically significant. The reasons for this difference can also be supported by the previously mentioned reasons above.

However, the difference in the percentage of SOM between these land use types can be attributed to the presence of vegetation within these land use types. As mentioned above, vegetation plays a pivotal role in acting like a protective layer for the soil's surface. According to Hoyle (2013), this layer is responsible for promoting biological activity that can generate and enhance SOM content and provide the soil with protection from environmental factors (e.g., wind, rain, and temperature extremes). However, within the eroded land, the presence of vegetation was little to none (Appendix J). Therefore, the maintenance and SOM formation process were hindered. Subsequently, within the eroded land, it was evident that water played a critical role in accelerating the erosion rates within the land use. The presence of sheet, rill, and gully erosion was quite distinct within the study area and may have resulted from a

combination of environmental, anthropogenic, and animal factors. These types of erosion left the soil bare and exposed which then resulted in the drying-out of the soil. Moreover, since sampling was done during the drier months (June-August), with little rainfall, the occurrence of wind erosion may have been prevalent. Wind erosion involves the movement and transportation of light soil particles via suspension, surface creep, or saltation over a distance that ranges from a few centimetres to hundreds of kilometres (Bullock, 2005). It is responsible for removing soil particles that are dry and loose (Osman, 2014b). This, in turn, may have reduced the SOM as part of the topsoil was being lost. Despite samples being collected within the drier months, the agriculture land use still had a fairly high percentage of SOM. This may be attributed to the supply of organic material from the crops grown within the land use. Furthermore, sampling was done during the fallow period, where agricultural soils are left uncropped or weed-free for a period of time in order to improve the overall fertility and health of the soil (Feng and Balkcom, 2017) (Appendix D). However, during this period, crop residues were placed along the soil's surface to reduce the risk of erosion but also provided soil organisms with organic material for decomposition, as mentioned above (Appendix G). This, in turn, also ensured SOM content was maintained within these land use types while also facilitating the formation of SOM. Subsequently, it was observed that during this period, natural vegetation such as grasses began to grow within the agricultural land use types (Appendix D). These grasses assisted in improving soil structure, microbial activity, and reduced the impacts of erosion which ensured a fairly high percentage of SOM the agricultural land use.

Additionally, vegetation plays an important role in governing the properties of soil organisms. The presence of vegetation has the ability to attract biota in the soil via root signals (Jacoby *et al*., 2017). Therefore, there is a greater community of soil biota within the agricultural land use than the eroded land. Subsequently, their activity within the soil can be explained by the amount of SOM that is formed. Furthermore, SOM is comprised of a significant number of living microorganisms and their dead fractions (Ramesh *et al*., 2019). The greater community of soil biota improves the overall soil structure by forming soil aggregates and increasing porosity, which plays an important role in the entry of gases and water into the soil.

Similarly, within the agricultural land use, the presence of molehills was evident (Appendix F). This is an indication that there is a good supply of organisms present within the soil for them to feed on. More importantly, moles are known for playing an important role in the mixing and aerating of soil. When these animals create tunnels under ground, they cause the soil from

the surface to mix with soil from the subsurface (Kuhn and Edge, 2000). This allows for the mixture of SOM, nutrients, organisms, etc., from the topsoil to the subsoil layers. Furthermore, the creation of tunnels by the moles helps to aerate soils by improving porosity. These factors are also responsible for influencing the SOM content within the agricultural land use. However, within the eroded land, the presence of molehills was not evident. According to Zhao *et al*., (2019), "soil microbial biomass, community structure, and physiological activities are sensitive to aboveground vegetation and belowground conditions". Due to the lack and absence of vegetation within this land use, the presence of soil microorganisms may be limited. Furthermore, the conditions of erosion within this land use may also have an influence on the presence of soil organisms. However, these soil organisms may be present within these soils, but their activity may be limited due to the unfavourable conditions of the eroded land. This, in turn, has an influence on the SOM content within these soils.

Additionally, the occurrence of grazing within the agriculture land use was more evident than compared to the eroded land due to the presence of vegetation (Appendix E). The animal dung from the grazing of livestock along these soils also provided a source of organic material for soil organisms and the formation of SOM.

The results of the study indicate that the land use types had a significant influence on the percentage of SOM within these soils. It also revealed that the amount of SOM within the soil of the study area is quite low, and therefore, the overall fertility and quality of soil is hindered.

2.5 Conclusion

The study has provided a critical analysis of the amount of SOM that is found within the Emakhosaneni area, particularly within its major land use types. Due to the semi-arid nature of the area, the soil within the study area is characterised by low SOM content. The major land use types also had a relatively low percentage of SOM. This, coupled with the ongoing erosion in the area, has negatively impacted the fertility and quality of the soil within the area. The low amounts of SOM, especially within the agricultural land use, make it difficult for subsistence farmers to increase their yield in the crops that are grown, which may lead to food insecurity. Furthermore, the low amounts of SOM can reduce the biodiversity within the soil environment. Therefore, it is important for studies such as the above to be carried out so that proper management and protection strategies are put into place so that SOM content increases within these soils. More specially, SOM governs the physical, chemical, and biological properties of soil. Therefore, any changes towards SOM would bring about changes in the functioning,

quality, and productivity of the soil. As a result, it is imperative that SOM and soil are managed and protected in a sustainable manner for future generations to come.

CHAPTER THREE

Using laboratory-based spectroscopy in the assessment of Soil Organic Matter within the Emakhosaneni area, KwaZulu-Natal, South Africa

Abstract

Soil Organic Matter (SOM) consists of decaying plant and animal material at various stages of decomposition, substances released by plant roots, and soil organisms. It is responsible for supporting physical, chemical, and biological functions within soil. Techniques such as laboratory spectroscopy provide a new approach in analysing SOM content as it is cost and time effective. SOM has a unique spectral reflectance resulting from its composition and its physical and chemical properties. This relationship has been used in the estimation and prediction of SOM. Therefore, this study aimed to examine the relationship between SOM content and laboratory-based spectral reflectance within the soil of the Emakhosaneni area, KwaZulu-Natal, South Africa. Thirteen random soil samples were obtained from each of the four major land uses (Agriculture, Built-up (Residential), Eroded, and Rangeland); hence, a total of 52 soil samples was obtained. Samples were oven dried overnight (105°C), crushed, sieved (1 mm), and analysed using the Walkley-Black method to determine SOM content. Furthermore, the Analytical Spectral Device FieldSpec3 was used to collect the spectral reflectance of the soil samples and was pre-processed to reduce noise before analysis. Descriptive statistics revealed that the area had an overall average SOM content of 2.79%. Results also showed that the agriculture land use had the highest average percentage of SOM, followed by rangeland, built-up, and eroded land use. The average reflectance of spectra revealed that eroded land had the highest average reflectance, followed by rangeland, built-up, and agriculture land use. This was due to the fact that soil colour is linked to SOM. The correlation analysis indicated a negative moderate relationship between SOM content and spectral reflectance for the overall data along the covered spectral range (400 nm to 2400 nm). The partial least squares regression analysis revealed that the models using pre-processed spectra performed better than the model which used raw data. Despite the influence of noise on the performance of the models, they did a fair job in predicting SOM content. Overall, the study indicated that an increase in SOM content will result in a decreased reflectance of the soil samples and that the converse relationship will be seen as SOM content decreases. Such information can be used in the estimation, prediction, and quantification of SOM content so that it can be managed and protected sustainably for future generations to come.

Keywords: Soil Organic Matter, Land Use, Reflectance, Correlation, Walkley-Black

3.1 Introduction

The outermost layer of the earth comprises of soil and is commonly referred to as the pedosphere (Wang *et al*., 2019). Within this layer, many soil-forming processes are operational and contain the greatest diversity of soils when it comes to their composition and properties. According to the FAO (2020), soil can be defined as "a natural body consisting of layers (soil horizons) that are composed of weathered mineral materials, organic material, air, and water. Soil is the end product of the combined influence of climate, topography, organisms (flora, fauna, and humans) on parent materials (original rocks and minerals) over time. As a result, soil differs from its parent material in texture, structure, consistency, colour, chemical, biological, and physical characteristics". Soils are complex materials responsible for influencing and facilitating a myriad of processes and functions on earth, thus making this material an integral element essential to life on earth. More specifically, it is a critical resource required by every individual on earth, along with living and non-living organisms (Bhattacharyya and Pal, 2015). The dynamic nature of the soil is a result of its unique composition of materials and substances. According to Kalev and Toor (2018), there are four major components that result in the final composition of soil; mineral particles (45%), water (25%), air (25%), and organic matter (5%).

SOM contributes the smallest amount towards the soil's composition. According to the FAO (2017a), SOM refers to "the organic constituents in soil in various stages of decomposition such as tissues from dead plants and animals, materials less than 2 mm in size, and soil organisms". SOM formation is primarily a biological process that is directly or indirectly influenced by the flora and fauna present within the soil (Bot and Benites, 2005). Plants, either in the form of aboveground litter (twigs, leaf litter, and stems) or belowground material (exudates, root litter, and mycorrhizal hyphae), usually enter and mix with the soil where they undergo decomposition (Wetterstedt, 2010). The by-products of this process, along with newly synthesized material, accumulate in the soil and produce a more complex form of SOM (Osman, 2012b). Therefore, SOM components usually include; plant and animal residue, soil organisms, stable organic matter (humus), and actively decomposing matter (Reddy, 2016). Despite its small contribution in soil, SOM is responsible for supporting many physical, chemical, and biological functions in the pedosphere and its interacting spheres. These functions assist in providing ecosystem services to humans and maintain soil functioning for plants and animals. More specifically, it plays a significant role in the quality of soil and hence, the overall fertility and productivity of the soil.

Various methods are used to measure and estimate SOM content in soils (Motsara and Roy, 2008). Commonly used methods are laboratory-based and include; the Walkley-Black, Loss On Ignition, Automated dry combustion, and Humic matter methods (Roper *et al*., 2019). However, these methods can sometimes be rather costly and take long periods to analyse, especially when study areas are large with extensive soil sampling. According to Wang *et al*., (2017), the low-cost, high efficiency, large-scale, non-destructive, and rapid data acquisition of remote sensing has proven to be a valuable tool for strengthening or improving traditional methods. For instance, techniques such as spectroscopy provide a new approach in determining SOM content (Hong *et al*., 2018a).

Spectroscopy focuses on investigating and measuring the spectra produced when materials or objects interact with or emit electromagnetic radiation (EMR) (van der Meer, 2018). It aims to understand the interaction between objects of interest and its EMR. This interaction is responsible for carrying information about the object of interest or the processes within it (Milton, 2004). Spectroscopy can either be applied in the laboratory or the field. Field spectroscopy focuses on the spectral measurements retrieved under natural conditions within the field, where sensors are usually attached to vehicles for on-the-go measurements (Escribano *et al*., 2017). Laboratory spectroscopy focuses on obtaining spectral measurements of objects; however, it is done under controlled conditions. The impact of environmental factors is minimized under such conditions. Field spectroscopy measurements are obtained under solar illumination, whereas laboratory spectroscopy measurements are obtained under artificial illumination. The type of spectroscopy selected is dependent on its application and the material or objects of interest. However, when it comes to the analysis of soil, laboratory spectroscopy is highly favoured. According to Wang *et al*., (2017), this is due to the "lofty high spectral resolution, convenience, and controllability" of laboratory spectroscopy.

The spectral information obtained from different objects returns different amounts of energy that can be absorbed, transmitted, or reflected in different bands of the electromagnetic spectrum. However, interest is placed on measuring and quantifying the reflectance of objects since it is unrelated to time, location, illumination intensity, atmospheric conditions, and weather (Peddle *et al*., 2001). According to Aggarwal (2004), reflected energy from objects are highly dependent on the property of the material, surface roughness, angle of incidence, intensity, and wavelength of the radiant energy. Thus, allowing earth's objects to have their own unique spectral reflectance as they reflect the sun's energy in different ways. Likewise, the soil has its own spectral reflectance signature unique to its properties, characteristics, and

composition (Appendix C). A typical spectral reflectance curve of soil gives insight into the relationship between soil's physical and chemical properties and its reflectance (Liu *et al*., 2008). It depicts this information by showing reflectance spectra in the visible (VIS, 400-700 nanometres [nm]), near infrared (NIR, 700-1100 nm) and, shortwave infrared (SWIR, 1100-2500 nm) regions of the electromagnetic spectrum (Krishan *et al*., 2014). Within these regions, soil reflectance increases with increasing wavelength due to iron oxide absorption at the shorter wavelengths (Huete, 2004). Subsequently, factors such as moisture content, soil texture (proportion of sand, silt, and clay), surface roughness, presence of iron oxide, and organic matter content are known to influence the reflectance of soils (Mujumdar and Kumar, 2012). These factors are complex in nature and are interrelated.

Similarly, the SOM within soil has its own unique spectral reflectance signature. This signature is a result of the composition of SOM together with its physical and chemical properties. For example, the remains of plants and animals contain sugars, proteins, lignins, celluloses, hemicelluloses, waxes, and lipids (Osman, 2012b). These biochemical materials are responsible for influencing the reflectance in the VIS, NIR, and SWIR region of the electromagnetic spectrum (Bartholomeus *et al*., 2008). However, several studies have revealed that SOM has a unique spectral response in the VIS and NIR regions (Wang *et al*., 2017). This may result from electronic transitions of atoms that occur within the VIS region (Luce *et al*., 2014). Subsequently, the bending and stretching of functional groups (e.g., NH, CH, and CO) found within SOM results in overtones and combination bands within the NIR region (Stenberg *et al*., 2010). Further, this results in the signature of SOM having a shape that depicts a raising curve with an increase in the wavelength. However, absorptions can also exist within the VIS and NIR regions from the SOM and influence the shape of the signature. According to Rossel and Behrens (2010), absorptions in the VIS region are influenced by chromophores and the darkness of humic acid, while absorptions in the NIR region are influenced by the overtones and combination absorption of functional groups (e.g., NH, CH, and CO). Such absorption can sometimes be difficult to identify due to the broad and overlapping bands within the VIS-NIR region. Furthermore, soil colour is known to be influenced by the presence of SOM. This may result from the decomposition of humus, which causes the soil to become much darker in colour (Bot and Benites, 2005). Therefore, a darker soil will be indicative of a high SOM content compared to a lighter-coloured soil. Studies by Daniel *et al.*, (2001) and He *et al*., (2009) shows a strong relationship between the VIS bands and the SOM, therefore allowing SOM to be easily identified within the VIS region of the electromagnetic spectrum. Furthermore, humic

substances such as humins and humic acid (darkest pigment of SOM) are responsible for reducing the spectral reflectance of SOM over the VIS to SWIR spectral range, while fluvic acid is known to have no influence on the SOM reflectance (Mohamed *et al*., 2018). Similarly, the factors that influence the spectral reflectance of soil will inadvertently also influence the spectral reflectance of SOM.

Additionally, the quantity and quality of SOM are known to significantly affect the shape and nature of a SOM reflectance spectrum (Amin *et al*., 2020). As mentioned previously, the higher the SOM content, the darker the soil, which then lowers the reflectance. The converse trend will be depicted for a lower content of SOM. Soils that typically have a low SOM content (<1%) can see a significant decrease in their reflectance due to the smallest increase in SOM content (Cierniewski and Kuśnierek, 2010). However, the converse will be seen for soils that have a high SOM content unless a small decrease in SOM content occurs, which can then change its reflectance. Subsequently, with 2% SOM content as the threshold, when SOM exceeded this threshold, the SOM played a significant role in concealing the spectral reflectance, whereas SOM content that was below this threshold was less effective (Wang *et al*., 2017). The influence of SOM content on spectral reflectance can be seen in Figure 3.1. Subsequently, the stage of decomposition, formation, and accumulation of SOM can greatly alter the shape of its reflectance spectrum. For example, the study conducted by Ben-Dor and Banin (1995) revealed that the organic material in one group of soils (SOM=4%) comprised mostly of humic substances, while another group (SOM>4%) comprised mostly of decomposed litter, which can vary more than well-decomposed humus. These different stages contain varying amounts of SOM, which can either cause an increase or decrease in the reflectance and thus the shape and slope of the spectral reflectance curve.

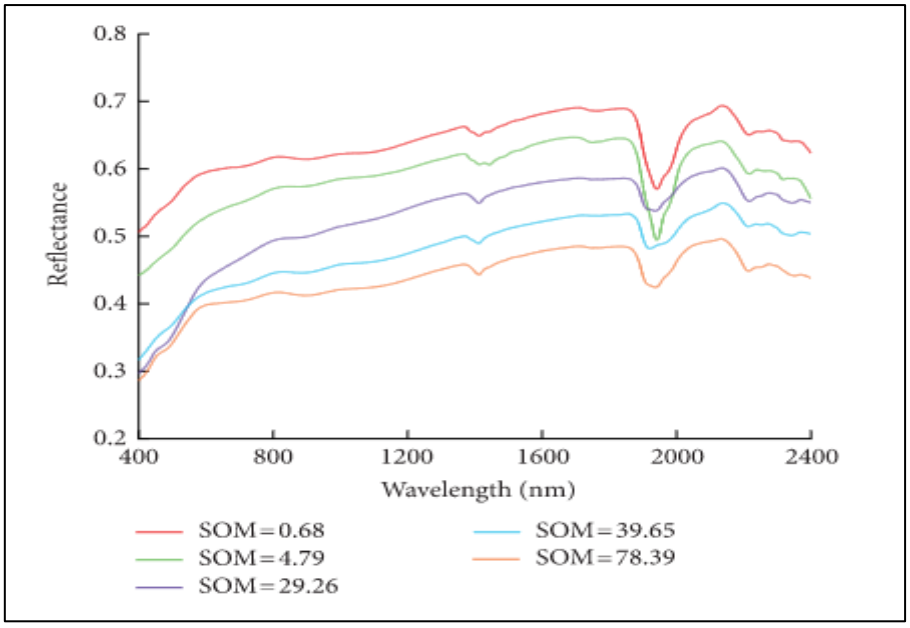


Figure 3.1: Spectral reflectance of soils with different organic matter content (Source: Wang *et al*., 2017)

Certain spectral wavebands have a sensitive relationship with SOM, and this has been reported by several studies (Amin *et al*., 2020). One of the first studies to examine this relationship was conducted by Krishnan *et al*., (1980), in which they observed that the slope of the spectral curve increased with increasing SOM content around the 800 nm wavelength band. Subsequently, Viscarra Rossel *et al*., (2006) discovered a good correlation between SOM and wavelengths such as; 410 nm, 570 nm, and 660 nm along the VIS region. Additionally, Wang *et al*., (2017) stated that these five bands (440 nm, 560 nm, 625 nm, 740 nm, and 1336 nm) are the optimal spectral bands for estimating SOM under laboratory conditions. Liu *et al*., (2009) confirmed that reflectance between the range 620 nm to 810 nm was related to SOM and that the highest correlation coefficient was at 710 nm. Furthermore, Bartholomeus *et al*., (2008) found the highest correlation between SOC and the wavelength range of 640 nm to 690 nm along the absorption spectrum. The amount of SOC is closely related to SOM, as it constitutes 58% of SOM (Donovan, 2013). These sensitive spectral bands further support the fact that SOM can be easily identified and therefore estimated using its reflectance within the VIS and NIR regions.

Such studies have paved the way in which SOM can be estimated and quantified using spectroscopy, especially under laboratory conditions. More specifically, the non-destructive nature of this technique allows repeatable measurements that can be taken simultaneously, which gives it a significant advantage over traditional laboratory methods (Angelopoulou *et al*., 2020). It is also a more cost-effective and less laborious method. In using such techniques

to estimate SOM content, results can be achieved at a much faster rate, thus allowing for accurate information to be achieved over shorter periods of time. This can be especially useful to farmers within the agricultural sector as it provides an insight into the status of fertility and productivity of the soil. Furthermore, a significant understanding of SOM dynamics can be achieved using such techniques, which can then provide valuable information regarding management practices that can either maintain or increase SOM content.

Human activities coupled with the increasing growth in the human population has led to the degradation of fertile topsoil and hence, SOM. The rate at which SOM is being degraded exceeds that at which it can be produced, especially over larger scales. For instance, the soils within South Africa are usually not characterized by high fertility and become easily degraded with the exception of certain soils. According to Du Preez *et al*., (2011), soils within South Africa are characterised by low organic matter levels, which may be attributed to the low SOC content found in topsoils. This degradation in SOM over the years is attributed to the improper use and management of soil. However, the situation is more likely to worsen in the near future if no measures are put into place. Therefore, the use of efficient laboratory spectroscopy can be used to examine SOM content to ensure that it is able to support and facilitate different uses and activities for all its users on earth. As a result, the current study will aim to examine the relationship between SOM content and laboratory-based spectral reflectance within the soil of the Emakhosaneni area located in KwaZulu-Natal, South Africa.

3.2 Materials and methods

3.2.1 Description of the study area

South Africa comprises of nine provinces. The selected province in which the study will be conducted is the KZN province located along the east coast of South Africa and occupies an area of 94 361 km^2, which is 7.7% of South Africa's land area and has a population of 10 267 300 people (SSA, 2011). The KZN province consists of 12 district municipalities that can further be subdivided into 55 local municipal areas. The study was conducted in the Emakhosaneni area that is located within the Okhahlamba Municipality, KZN.

The Okhahlamba Municipality (Figure 3.2) is located in the western region of KZN and is approximately 242 km from the city centre of the eThekwini Metropolitan Municipality. It is situated in the mountainous regions of KZN between Lesotho, Free State, Emnambithi and Mtshezi (SSA, 2011). The total population of individuals present within the municipal area is 132 068 and are distributed over an area of 3,970.98 km^2 with an average density of 33.26 individuals per km^2 (Frith, 2011). The local municipality gives rise to many towns and settlement areas, but Emakhosaneni was selected as the study area for this particular study.

The Emakhosaneni area is located in the central region of the Okhahlamba municipality (28°46'45.1"S; 29°14'42.4"E) (Figure 3.2). The area has a total population of 1 938 individuals that are distributed over an area of 13.72 km^2 and has an average density of 141.30 individuals per km^2 (SSA, 2011). Despite being a large area, there are only 380 households found within the area that are sparsely dispersed. All households within the area are traditional/tribal. The area also mainly comprises of Black African individuals (99.9%) (SSA, 2011). It is situated approximately 17.5 km from the nearest town, Bergville.

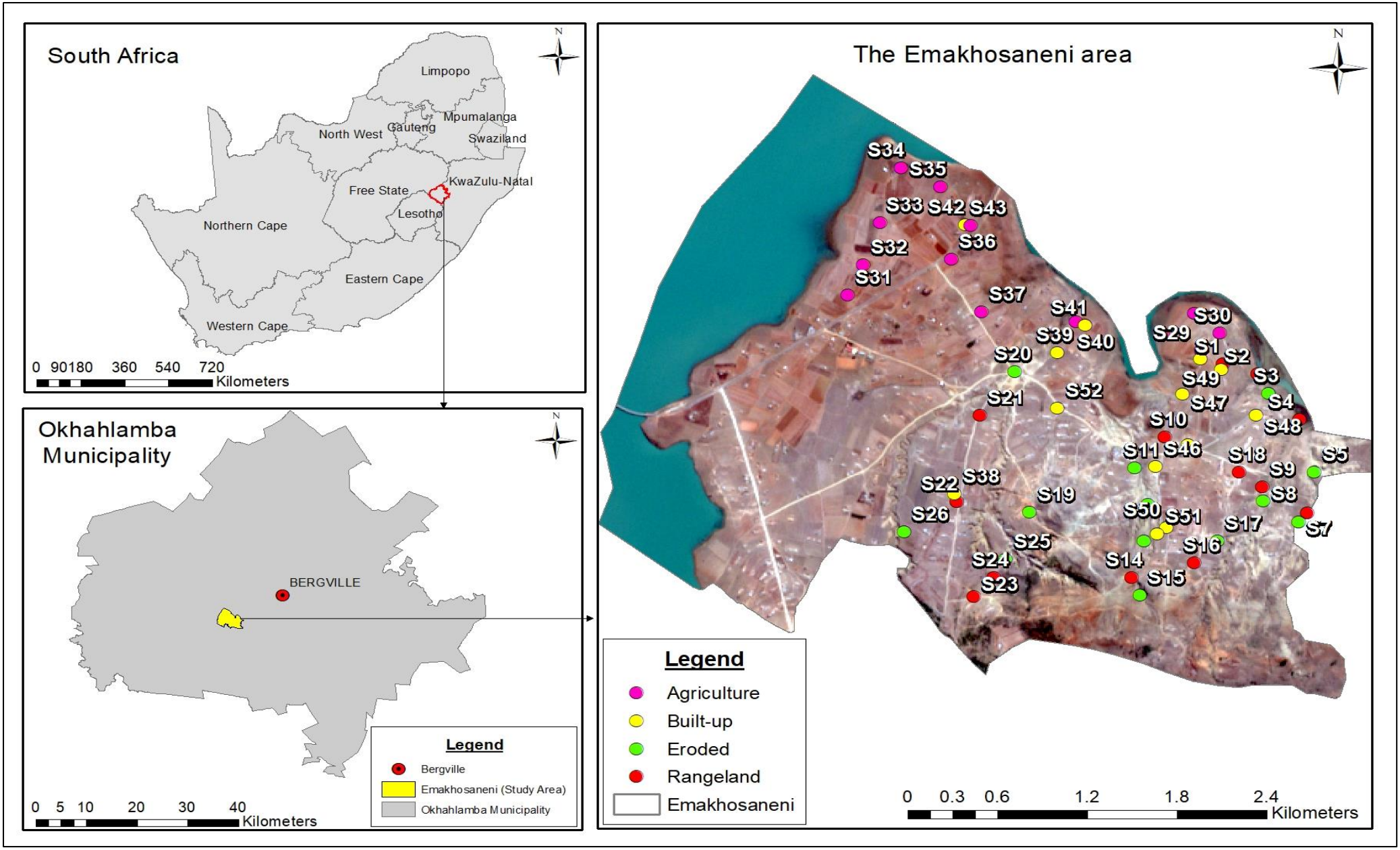


Figure 3.2: The selected study area and sampling points from which soil samples were obtained within the four major land uses (Author, 2020)

The study area experiences a warm and temperate climate. The monthly average midday temperatures range from 19°C in the winter months like July to 29°C in the summer months like February (Weather Atlas, 2020). On the other hand, the area receives an average of 644 mm of rain per year, with the highest rainfall experienced during the month of January (120 mm) and the lowest rainfall is during the month of July (5 mm) (Weather Atlas, 2020). The wind speeds vary over the months, with the winter and spring months experiencing the highest wind speeds, while the summer and autumn months experience the lowest wind speeds.

Furthermore, the area is located in a mountainous region. As one moves down the mountainous landscape, the gradient begins to become less steep. Along the base of the mountain, the gradient begins to decrease gradually until a gentle to no gradient is seen. This extends for several kilometres. Also, an increase in the number of households is seen as one moves away from the mountainous regions of the area.

In addition, the area is surrounded by various water sources. The Woodstock dam is located north of the study area, while to the eastern regions, the Tugela River and Drieldam are found (Google Earth, 2020). These water sources can be easily accessed and therefore supports their surrounding land use types. Two types of geology constitute the study are; Adelaide and Tarkastad (SANSA, 2015). The area is dominated by agricultural practices on both a large and small scale. Large areas of land are cleared from their natural vegetation in order to make land available for the expansion of the settlement and for agricultural practices. With such extensive farming occurring in the area, the appearance of land fragmentation is seen. There are a few sparse trees that remain after the land was cleared from its natural vegetation around the area, but the highest densities are seen in those parts that are inhabited by community members. Lastly, there are a number of road networks found throughout the area that aid access to various parts into and out of the area.
A more detailed description of the study area can be found in chapter two.

3.2.2 Soil sample collection

A thorough examination of the study area was conducted before sample collection. This was done using satellite imagery via Google Earth (Google Earth, 2020) and an aerial photograph of the study area followed by field reconnaissance. This led to the identification of four major land uses/land cover that constitute the study area; Agriculture, Built-up (Residential), Eroded, and Rangeland.

Soil samples were collected from the 16th July to 19th July 2019, from 08:00 am to 15:00 pm. A stratified random sampling method was used to determine the sampling points within the study area (Figure 3.2). It is a commonly used method that is designed to ensure that the samples taken are representative of the area and eliminates any biasedness (Worsham *et al*., 2012). Thirteen random soil samples were obtained from each of the four major land uses (Agriculture, Built-up (Residential), Eroded, and Rangeland), which amounts to a total of 52 soil samples being obtained from the study area.

Soil samples were collected from the soil's surface to a depth of 15 cm using a soil auger. The collected sample was then placed into a labelled, airtight plastic bag to prevent any spillage or drying of the sample, and the GPS coordinates were recorded. Subsequently, any other observations regarding the sampling point were recorded. However, at certain sampling points, the soil auger could not penetrate the soil's surface, and instead, a spade was used to obtain the soil sample. The above-mentioned procedure was replicated for the collection of soil samples from the 52 sampling points. Due to the study area being quite far from the laboratory, the soil samples were stored in a cool and dry cooler box, away from direct exposure to sunlight. Once all 52 soil samples were obtained, they were transported to and stored in the laboratory, where the analysis would begin.

3.2.3 Soil sample preparation and analysis

Approximately 500 g of each soil sample was oven dried overnight at a temperature of 105°C. Once oven dried, the samples were crushed into fine soil particles using a porcelain mortar and pestle. The crushed samples were then passed through a 1 mm stainless steel sieve. Sieving ensures that large and foreign materials such as roots and stones are not included in the analysis and allows for a more homogenous sample to be analysed. The sieved samples were then transferred into their respective labelled glass beakers and covered with Parafilm M until analysis. A portion of these samples was used for analysing the SOM content using a laboratory-based approach, while the remaining amount was used for spectral analysis.

Various methods are used to measure and estimate SOM content in soils (Motsara and Roy, 2008). However, this study used one of the most common methods known as; the Walkley-Black method to determine the percentage of SOM content in the soil samples. It is a wet oxidation method that involves the use of acids to determine the amount of oxidizable carbon in soil; this amount is then converted to SOM using a conversion factor (Walkley and Black, 1934). For this study, a small amount of the sieved soil sample was placed in a conical flask

together with potassium dichromate and sulphuric acid and was left to stand for 30 minutes. Thereafter, distilled water, orthophosphoric acid, and the indicator were added to the conical flask, which changed the colour of the solution to a dark black. The solution was then titrated against ferrous ammonium sulphate in small increments until the solution changed colour from black to green, indicating the titration's endpoint. The total volume of ferrous ammonium sulphate used in the titration was recorded, and the percentage of SOM content was calculated. This procedure was repeated for all soil samples.

A detailed description explaining the manner in which SOM content was analysed using laboratory analysis and calculated can be found in chapter two.

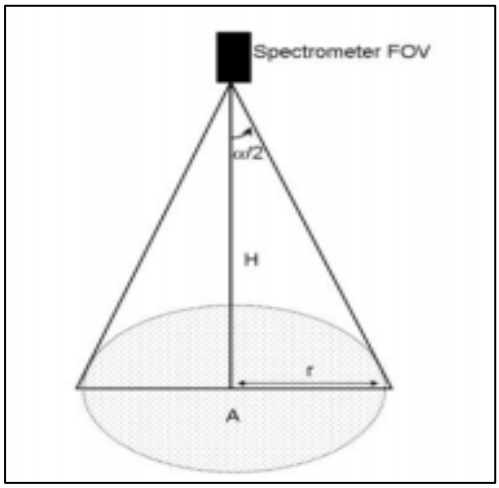
3.2.4 Spectral data collection

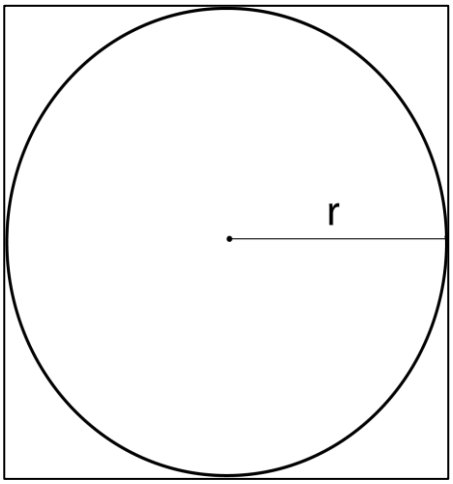
The spectral measurements of soil samples were taken from 19th to 20Th December 2019. These measurements were taken in a dark laboratory environment under diffused light conditions. The temperature in the room was maintained at 24°C. A portable spectroradiometer known as the ASD FieldSpec3 (Analytical Spectral Devices Inc., Boulder, CO, USA) was used to collect the reflectance spectra of the soil samples. This instrument covers a spectral range from 350 nm to 2500 nm. Along the instrument's spectral region 350 nm to 1000 nm, the sampling interval is 1.4 nm. The spectral resolution is 3 nm, and along the spectral region 1001 nm to 2500 nm, the sampling interval is 2 nm, and the spectral resolution is 10 nm. The spectroradiometer measures the optical energy that is obtained through a bundle of specially formulated optical fibers that is precisely cut, polished, and sealed for optimum energy collection.

Approximately 500 g of oven dried soil was placed in a black plastic dish that had a diameter of 22 cm and was 2 cm deep (Figure 3.3). A black plastic dish was used to ensure that there was no interference from it during the collection of spectra. The soil sample was then levelled to a smooth surface using a stainless-steel spatula (Figure 3.3). This step ensured a maximum diffuse reflection and a high signal-to-noise ratio during the collection of spectra (Mouazen *et al*., 2010). A smoother surface also reduces the amount of scattering. The plastic dish with the soil sample was then placed on a table with a black tablecloth for analysis.

Figure 3.3: A soil sample that was levelled off to a smooth surface in a black plastic dish (Author, 2020)

It is essential to ensure that the area of the black plastic dish is wider than the area covered by the ASD. Thus, before the setup of the instrument, the estimated sample area of the fiber and the plastic dish was calculated using the following approach (Danner *et al*., 2015);





Area covered by ASD

$r = Tan\ (\frac{\propto}{2})$ x d

= $Tan\ (\frac{25}{2})$ x 40

= 8.87 cm

$A = \pi r^2$

= $\pi(8.87)^2$

= 247.05 cm^2

Area of Plastic dish

$A = \pi r^2$

= $\pi(11)^2$

= 380.13 cm^2

A 50 watt (W) halogen lamp (Analytical Spectral Devices, Inc. USA) was used as a source of illumination for the soil sample. The lamp was placed on a tripod and held 60 cm above the table at a 30° illumination zenith angle (Figure 3.5a). A horizontal distance of 48 cm existed between the lamp and the centre of the soil sample. The pistol grip of the ASD FieldSpec3 spectroradiometer was also attached to a separate tripod and held 40 cm above the soil sample

at a 0° zenith angle (i.e., it was pointed vertically towards the sample) (Figure 3.5a). The pistol grip contained a bare optic fiber that had a 25° field of view (FOV) and covered a circular area of 247.05 Square centimetres (cm^2) was used to measure the spectral reflectance of the soil sample (Figure 3.5c).

Before any spectral measurements could be taken, the instrument and the lamp were warmed up for 1 hour and 30 minutes respectively to ensure optimization and stabilization as the three spectrometer arrays need to reach thermal equilibrium. After warming up, a fiber optic check was performed for each of the bands (VIS, NIR, and SWIR) to ensure that no fibers were damaged or broken, as this could impact the accuracy of the measurements by contributing to the noise. After that, a dark current measurement was taken, followed by an optimization of the instrument and then a white reference measurement. A white spectralon (Labsphere, Inc., Sutton, NH, USA) reflectance panel (30 cm x 30 cm) with a 100% reflectance was used to take a white reference measurement for the calibration of the instrument (Figure 3.4). Once the instrument was calibrated, the spectral reflectance of the soil sample was measured using the inbuilt software (RS^3, version 6.3). A total of 10 spectral measurements were then taken for the soil sample along 2151 bands (i.e., covered spectral range: 350 nm to 2500 nm). These spectral measurements were then saved as raw digital numbers (DN) and were later converted to reflectance using the ViewSpec Pro software version 6.2. The converted reflectance spectra were then averaged (using ViewSpec Pro software) and used as the reflectance spectrum for the soil sample. The above procedure was carried out for all 52 soil samples. The instrument was optimized every 30 minutes, and a white reference was collected every 15 minutes (Figure 3.4).

Figure 3.4: The white spectralon used to take a white reference measurement (Author, 2020)

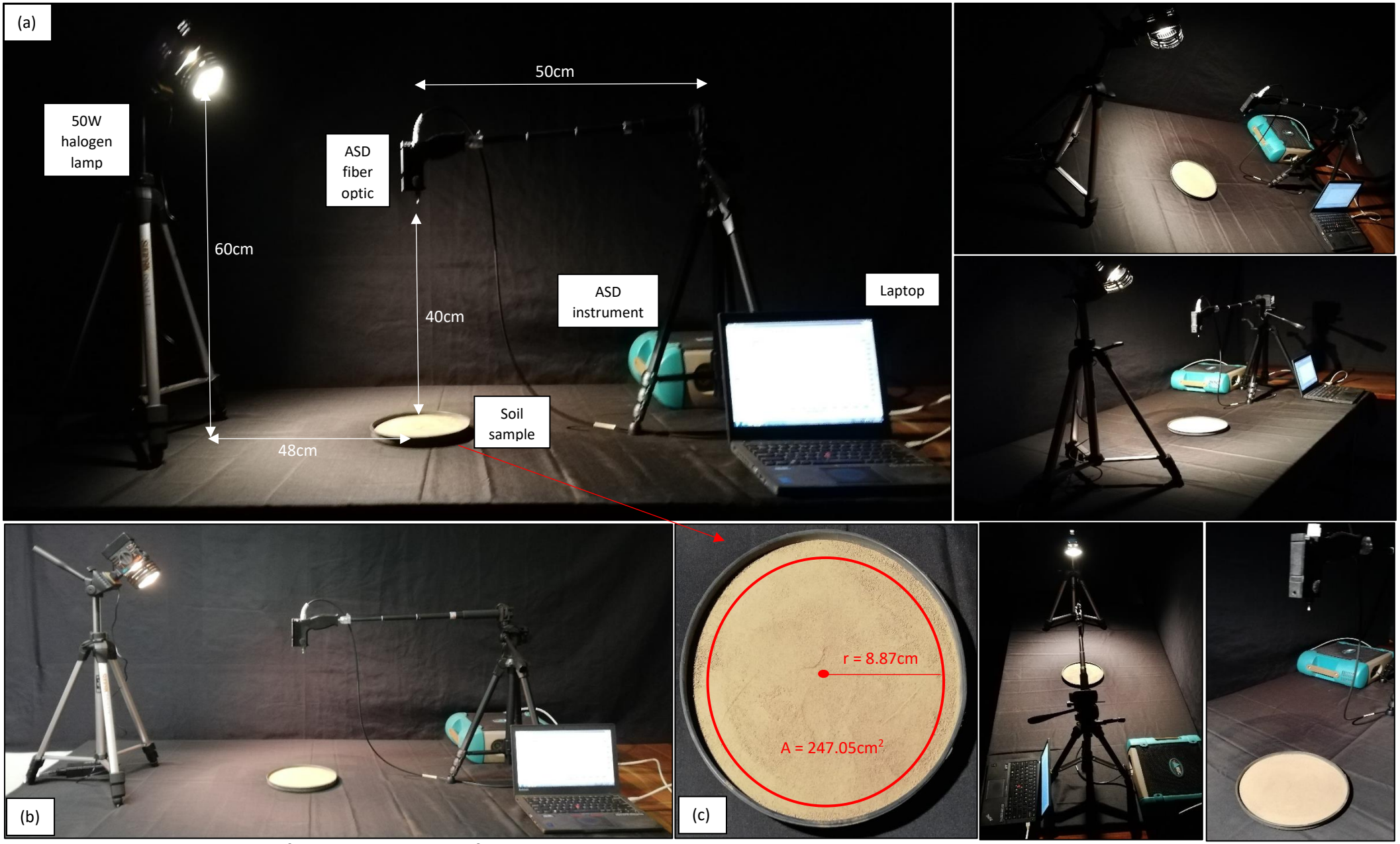


Figure 3.5: Overview of experimental setup of instrument under (a) dark and (b) light conditions, and (c) area covered by the ASD (Author, 2020)

3.2.4.1 Spectral data pre-processing

The raw spectra collected from the ASD instrument usually contain some noise (Figure 3.6a). To reduce this noise that resulted from the instrument, light scattering effects, and energy responses that varied along the different wavelengths, certain wavelengths were removed (Gao *et al*., 2017; Wang *et al*., 2017; Li *et al*., 2019). These wavelengths included; the noisy ends from the beginning (350 nm to 399 nm) and end regions (2401 nm to 2500 nm). This process reduced the number of bands from 2151 to 2001. Further, the remaining bands (that range between 400 nm to 2400 nm) were then resampled using a 10 nm interval; this step further reduced the noise and the number of bands to 201 (Figure 3.6b). It also assisted in smoothing the spectral curves, which depicted curve features such as absorption peaks and valleys of the soil samples. After resampling, the spectra had undergone several smoothing methods such as; Moving average, Gaussian filter, Median filter, and Savitzky–Golay (SG); however, for this particular study, the SG method was selected as it was the most effective in smoothing the spectra. The SG smoothing was applied using a second-order polynomial with 11 smoothing points (five smoothing points on the left and five on the right) (Figure 3.6c). This method reduces the noise from spectra (e.g., baseline-drift, tilt, reverse, etc.) and increases the signal to noise ratio while preserving the number of variables (Savitzky and Golay, 1964; Agrawal and Deshmukh, 2018). The mathematical formula of the SG smoothing method is expressed as (equation 1) (Savitzky and Golay, 1964);

$$A_i, SG = \frac{\sum_{j=-m}^{j=m} C_j \,.\, A_i + \, j}{N}$$

Where, *Ai, SG* is the smoothed, and *Ai* is the original reflectance value, *Cj* is the filter coefficients, *j* is the running index of the original data in the original data table, and *N* is the number of convoluting integers.

Once the spectra had undergone smoothing, they were transformed using the First-Order Derivative transformation with SG smoothing (FDT) (Figure 3.6d). This method assists in correcting baseline effects found in raw spectra and emphasizing small spectral variations that are not visible in the raw spectra (The Unscrambler, 2014). These pre-treatment methods are commonly found to be effective in similar studies as they smooth the spectra and reduce particle size effects and noise that are created during the process of spectral data collection (Bilgili *et al*., 2010; Srivastava *et al*., 2016; Vibhute *et al*., 2018). The pre-processed spectra

were then used for further analysis. All pre-processing methods were carried out using the Unscrambler X, CAMO, Norway, version 10.4 software.

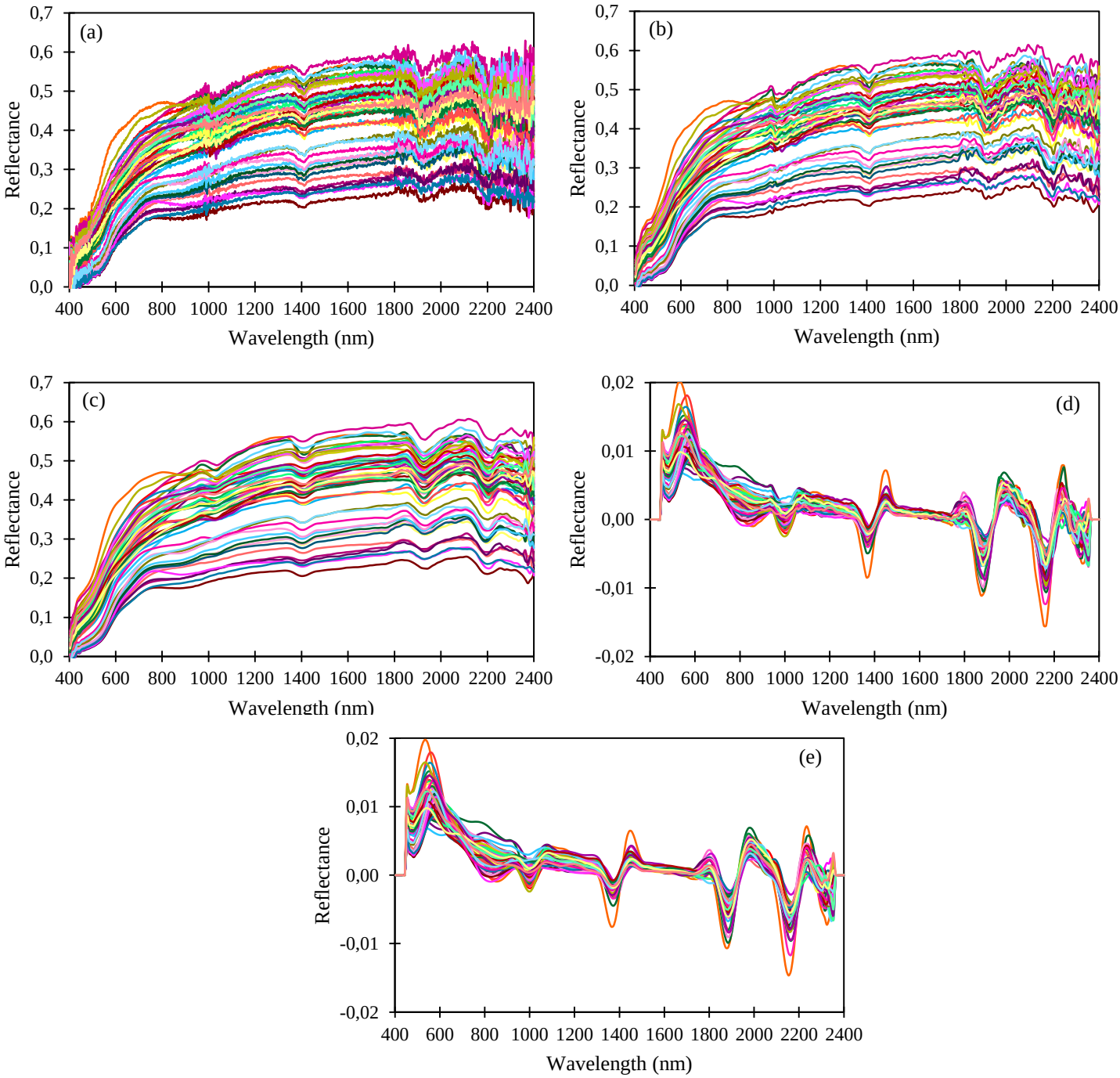


Figure 3.6: The reflectance curves showing the (a) raw spectra, (b) the reduced spectra, (c) the SG smoothened spectra, (d) the FDT transformed spectra of the soil samples, and (e) the combined pre-processed spectra (SG+FDT) (Author, 2020)

Finally, the pre-processed spectra were then averaged based on the different land uses from which the soil samples were collected (i.e., Agriculture, Built-up (Residential), Eroded, and

Rangeland). An overall average of all pre-processed spectra was also obtained using the Unscrambler X software. These averaged spectra were then represented graphically.

3.2.5 Data analysis

3.2.5.1 Soil

Once the percentage of SOM was calculated, descriptive statistics, such as the means and standard deviations, were calculated and used to summarize the data. This was done using SPSS version 25 software. This software assisted in the performance of appropriate descriptive statistics. The creation of tables was also done using the SPSS software.

3.2.5.2 Spectral reflectance

A correlation analysis was carried out in order to examine the relationship between SOM content and the reflectance of soil samples. According to Qiao *et al*., (2017), correlation analysis is a reliable method that is used for analysing the correlativity between dependent and independent variables. The strength of the relationship between the two variables is expressed as the correlation coefficient (*r*) and was calculated using the following formula (equation 2) (Virgawati *et al*., 2019):

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}}$$

Where, *n* is the number of pairs of scores, Σxy is the sum of the products of paired scores, Σx is the sum of *x* scores, Σy is the sum of *y* scores, Σx^2 is the sum of squared *x* scores, and Σy^2 is the sum of squared *y* scores.

The correlation coefficient value can range from +1 to -1. An *r* value of 0 suggests that there is no association or relationship between the variables, while a value greater than 0 indicates a positive relationship where; an increase in the value of one variable causes an increase in the value of the other variable. However, an *r* value that is less than 0 indicates a negative relationship where; an increase in the value of one variable causes in decrease in the value of the other variable.

The correlation coefficient values were calculated at each wavelength band from 400 nm to 2400 nm across SOM content. This was done for both the raw and pre-processed spectra for the overall dataset and within each land use type using only the SG pre-processed data (agriculture, built-up, eroded, and rangeland). The correlation coefficient values were then represented graphically. The above procedure was carried out using Unscrambler X software.

In order to interpret the correlation coefficients, values in Table 3.1 were used to understand the strength of the relationship between the SOM content and spectral reflectance.

Table 3.1: Levels of correlation coefficient values (Source: Virgawati *et al*., 2019)

Coefficient interval (+/-)	Level of correlation
0	No correlation
0.00-0.25	Very weak correlation
0.25-0.50	Moderate correlation
0.50-0.75	Strong correlation
0.75-0.99	Very strong correlation
1	Perfect correlation

3.2.5.3 Partial Least Squares Regression

Partial Least Squares Regression (PLSR) is a widely popular modelling technique that is used in chemometrics. This method is also frequently used to conduct quantitative spectral data analyses (Rossel and Behrens, 2010). PLSR has proven to be a robust and reliable approach when it comes to quantitative spectral analysis. This may result from its advantages in terms of dimension reduction, synthesis, and solving collinearity problems that may exist amongst independent variables (Martens and Martens, 2000; Farifteh *et al*., 2007; Yu *et al*., 2016; Wang *et al*., 2017). The PLSR method is similar to that of multiple linear regression, in which it is used to relate the predictor (X) and response (Y) variables. However, it selects successive orthogonal factors known as latent variables that increase the covariance between the predictor (X- soil reflectance spectra) and the response (Y- SOM content) variables. A more detailed discussion regarding PLSR can be found in Martens and Naes (1989).

For this particular study, a PLSR analysis was adopted in order to examine the relationship between SOM content and the collected reflectance spectra. The analysis was performed using the soil reflectance as the predictor variable and the SOM content as the response variable. To determine the optimal number of latent variables to use when creating the PLSR model, the leave-one-out cross-validation (LOOCV) method was used. This process minimizes the over or under-fitting of data when creating the PLSR model (Nawar *et al*., 2016). Furthermore, the data was mean centred before applying a kernel PLS algorithm available in the software used. Lastly, outliers were identified and removed before proceeding with the analysis. The PLSR analysis was also carried out using the Unscrambler X software.

3.2.5.4 PLSR model development and validation

Before the development of the PLSR models, the 52 sampling sites were divided into two groups. The first group was the calibration samples used for creating the models (n=36, 70% of sampling sites), while the second group was used for the validation of the models (n=16, 30% of sampling sites).

The calibration dataset was used to create four PLSR models based on the following:

1. Raw spectra and SOM content (R)
2. Pre-processed SG spectra and SOM content (SG only)
3. Pre-processed FDT spectra and SOM content (FDT only)
4. Overall pre-processed spectra and SOM content (SG+FDT)

These four PLSR models were created using the above-mentioned method.

The validation dataset was used to validate the accuracy of the PLSR prediction models. Firstly, to test the prediction accuracy and strength of the PLSR models, the coefficient of determination (R^2) of the SOM predictions was calculated using the following (equation 3):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_{measured} - Y_{predicted})^2}{\sum_{i=1}^{n}(Y_{measured} - Y_{mean})^2}$$

Where, $Y_{measured}$ is the measured SOM content, $Y_{predicted}$ is the predicted SOM content, Y_{mean} is the mean of the measured values, and *n* is the number of samples.

Secondly, in order to test the accuracy of the different prediction methods used, the Root Mean Square Error (RMSE) was calculated for the calibration (RMSEC) and validation (RMSEP) datasets using the following (equation 4 and 5):

$$RMSEC = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_c - Y_m)^2}$$

$$RMSEP = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_p - Y_m)^2}$$

Where,Y_m is the measured SOM content, Y_c is the predicted SOM content derived from the calibration dataset, Y_p is the predicted SOM content derived from validation dataset, and *N* is the number of samples.

3.3 Results and Discussion

3.3.1 SOM analysis

The descriptive statistics of the SOM content within the study area and its major land uses are displayed in Table 3.2. Descriptive statistics of the SOM content (Table 3.2) depicted that the land use agriculture has the highest average SOM content (3.42%). In contrast, the land use eroded contains the lowest average SOM content (2.19%). An average difference of 1.23% in SOM exists between these land uses. The average SOM content in land uses such as rangeland and built-up lie between the highest and lowest percentages of SOM. The standard deviation, on the other hand, exhibits a different trend, in that eroded land has the highest standard deviation (1.26) and built-up land use the lowest (0.98). Regarding the minimum and maximum percentage of SOM found in each of the different land uses, agriculture had higher SOM percentages than others. The highest minimum was 2.23% (agriculture), and the lowest was 0.00% (eroded). However, agriculture (5.96%) had the highest maximum percentage of SOM, while rangeland (4.62%) had the lowest. With regards to the overall average of SOM content in the study area, an amount of 2.79% was seen. The lowest percentage was 0.00%, and the highest was 5.96%.

Table 3.2: Descriptive statistics indicating the overall percentage of SOM within the study area and its major land uses (N= Number of samples, Std= Standard deviation)

	N	Mean	Std. deviation	Minimum	Maximum
Overall	52	2.79	1.19	0.00	5.96
Agriculture	13	3.42	1.03	2.23	5.96
Built-up (residential)	13	2.50	0.98	0.74	4.76
Eroded	13	2.19	1.26	0.00	4.91
Rangeland	13	3.05	1.18	1.34	4.62

From the above results obtained, it is evident that the average percentage of SOM within the Emakhosaneni area is relatively low compared to the average amount that is usually found in soils. This average roughly ranges from 1-6% in the topsoil, whereas the subsoil contains less than 2% of SOM (Adams *et al*., 2011). Subsequently, certain soils, such as mineral soils, gave a SOM concentration that ranges from 1-5% (Kalev and Toor, 2018). The concentrations of SOM are greatest in the topsoil layer (13-25 cm) and decrease with depth. The soil samples that were analysed for this study were obtained from the same layer, indicating that the topsoil within the Emakhosaneni area contains approximately 2.79% of SOM (Table 3.2). This may be supported by the fact that low organic matter levels characterise soils within South Africa

as a result of the low SOC content found in topsoils that usually range from <0.5% to >4%. A low supply of carbon in the soil can directly influence the SOM content, as SOC constitutes 58% of SOM (Donovan, 2013). The lower levels of SOC present within the soils of the study area may have contributed to the lower concentration of SOM during analysis. Additionally, different soil types in other regions or areas are highly influenced by their soil-forming factors; climate, organisms, relief (topography), parent material, and time (Jenny, 1994). For instance, South Africa is a semi-arid region that receives less than 500 mm of rainfall (SAWS, 2019). Under such conditions, moisture supply to the soil and its organisms is lowered (Osman, 2012b). This may have slowed down the microbial activity and decomposition process, hindering the formation and accumulation of SOM within these soils. Likewise, sampling was done during the month of July, when rainfall in the area is at its lowest. This coupled with relatively warm temperatures may have resulted in lower SOM concentrations being present within the study area during sampling and, therefore, during analysis. A detailed discussion regarding the average SOM content found within each of the sampled land uses can be found in chapter two.

3.3.2 Spectral reflectance of soil samples

The average spectral reflectance curves of the soil samples are displayed in Figure 3.7. The spectral reflectance curves for the different land uses (Figure 3.7) exhibit a similar pattern or shape along the spectrum covered (400 nm to 2400 nm). They all start at a relatively low reflectance at 400 nm, then begin to increase exponentially to about 700 nm and then continuously increase and decrease at a relatively stable rate until 2400 nm. More specifically, from 400 nm to 700 nm, the spectral curves are quite truncated together at the starting wavelength but begin to spread away from one another and begin to smoothly increase as the wavelength increases while maintaining the same S-shaped curve. From 700 nm to approximately 1344 nm, there is a gradual increase in reflectance, shortly followed by a sudden, gradual decrease in reflectance at 1400 nm. A similar trend is followed thereafter but over a much shorter distance in wavelength, where an increase in reflectance from 1454 nm to 1854 nm occurs and then a sudden decrease at 1900 nm. Shortly after, the reflectance increases to 2124 nm followed by a rapid decrease to 2200 nm, a dome or bell-shaped is exhibited along these regions. Once again, the reflectance increases to a peak and decreases afterwards to 2400 nm.

The average spectral curve for each land use type, shows that eroded land had the highest reflectance curve while the agriculture land use had the lowest reflectance curve. The

reflectance curve of rangeland and built-up land uses lie in between the eroded and agriculture land uses. Interestingly, the eroded, rangeland, and built-up curves are in close proximity to one another over the covered spectrum (400 nm to 2400 nm), whereas the agriculture land use curve is relatively much lower. The agriculture curve also begins at a much lower reflectance and ends at a low reflectance. It is also evident that the eroded and rangeland land use curves overlap in the 400 nm to 700 nm region and begins to separate after that, but they begin to overlap around 1854 nm and remain in close proximity with each other afterwards. In terms of the overall reflectance curve, it also depicts a similar shape and trend as the above mentioned; however, it is quite similar to the curve of the built-up land use.

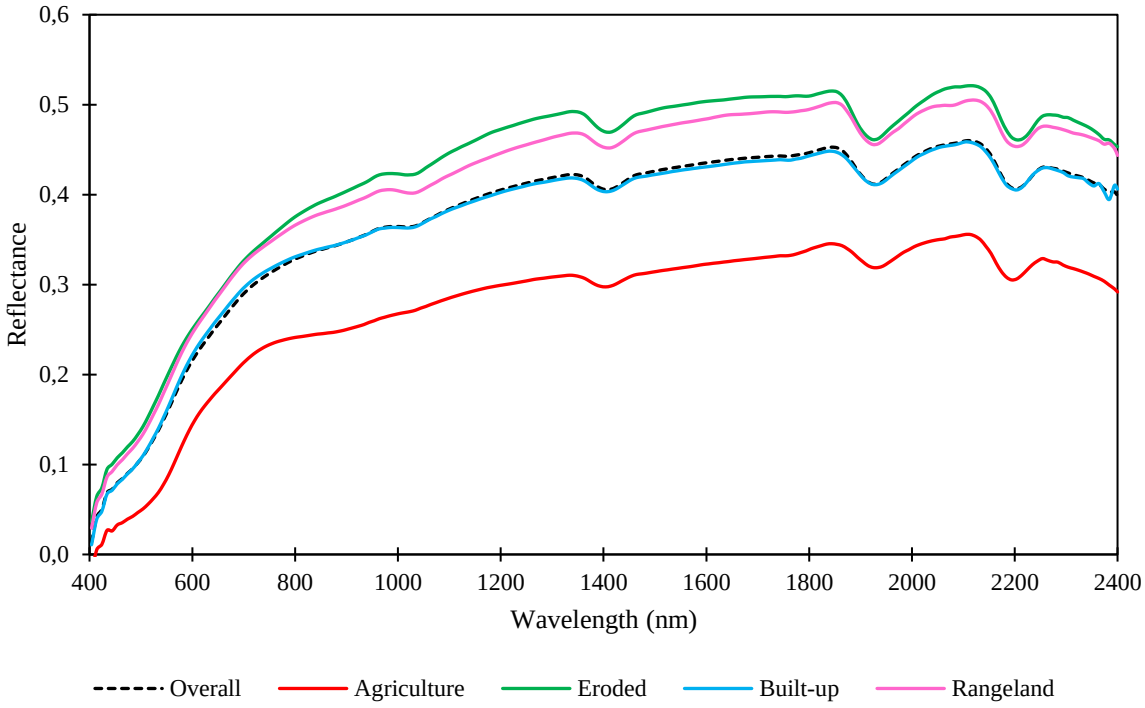


Figure 3.7: The average spectral reflectance curves of soil samples obtained within each of the major land uses of the study area and the overall average spectral reflectance

From the results obtained in Figure 3.7, it was evident that the average spectral reflectance curves for each of the land uses varied but exhibited similar trends in their slopes, shape and, pattern along the covered wavelength range (400 nm to 2400 nm). Along the 400 nm to 700 nm wavelength range, there was an increasing linear pattern in reflectance, and this may be attributed to the fact that this is the VIS region. Within this region, the reflectance is influenced by the high energy of radiation that results in the transition of electrons between molecular orbits with varying energy levels (Miller, 2001). However, low radiation energy corresponds to longer wavelengths and energy absorption, which causes vibrations in the molecular bonds

(Wetterlind *et al*., 2013). Thus, resulting in the creation of absorption bands along this region, which justifies the lower reflectance exhibited at the beginning wavelength bands in relation to the rest of the wavelength range. Apart from vibrational transitions, electronic transitions also result in creating absorption bands within this region (Fang *et al*., 2018). These absorption bands are highly related to colour, more specifically, soil colour. According to Wang *et al*., (2017), the influence of soil colour may be a result of electron excitations. As mentioned previously, a dark coloured soil will absorb more light and reflect little to no light back, whereas a light coloured soil will reflect the majority of the incoming light and absorb little to no light. Since the VIS region is comprised of different coloured bands, certain colours will return and/or absorb varying amounts of light or energy back. Therefore, supporting the low reflectance seen in the VIS region of the average spectra in Figure 3.7. Moreover, soil colour is directly linked to SOM and can be used to explain the absorption in the VIS region. Such absorptions in the VIS region are influenced by chromophores and the darkness of humic acid present in SOM (Rossel and Behrens, 2010). The different SOM content in the soils could have influenced the reflectance of the spectra along the VIS region. It could have caused the slow linear increase in reflectance from 400 nm to 700 nm. Another reason for the absorption within this region is because of iron-oxide minerals found in soil. Various kinds of iron-oxides cause electron transitions within the VIS region. Common iron-oxide minerals such as; goethite and hematite tend to show strong absorption bands between 400 nm and 660 nm (Wetterlind *et al*., 2013; Fang *et al*., 2018).

Additionally, from 700 nm and onwards to 2400 nm, there was a constant increase and slight decrease in reflectance, with both peaks and valleys being quite evident (Figure 3.7). Firstly, along this region, there were three major absorption peaks at 1400 nm, 1900 nm, and 2200 nm. Absorptions along these wavelengths may be a result from the overtones of groups such as; OH, SO_4, and CO_3 and the combinations of fundamental features of CO_2 and H_2O (Stenberg *et al*., 2010). More specifically, the absorption peak at 1400 nm is a result of the absorption band for water, which is related to the bending and stretching of the O-H bonds found in free water (Wang *et al*., 2017). Furthermore, the absorption peak at 1900 nm is also an absorption band for water. The combination band in this region involves the stretching and bending of the water molecule and is therefore specific to molecular water (Oliveira *et al*., 2013). Subsequently, the absorption peak at 2200 nm is due to combination tones caused by the bending and stretching vibrations of Al-OH and Mg-OH (Mutuo *et al*., 2006; Stenberg *et al*., 2010). Apart from the absorption peaks in this region, the reflectance around 2250 nm to 2400 nm saw a decreasing

trend (Figure 3.7). This is also due to absorption bands, but are much weaker and are related to combination vibrations that involve stretching of OH and bending of metal-OH (Wetterlind *et al*., 2013). These metals usually include Fe-OH and/or Mg-OH but can vary based on the minerals present within the soil (Fang *et al*., 2018). It was also noted that the reflectance in the NIR region was much higher compared to the VIS region (Figure 3.7). This may be due to the fact that broader bands are found within the NIR region. Subsequently, soil type may have also had an overall influence on the shape of the curves.

In terms of the average spectra obtained from each of the land uses, it was also evident that they displayed a similar shape but with a unique reflectance (Figure 3.7). One of the main factors influencing this is soil colour. The soil samples obtained from the study area differed greatly in terms of their colour. More specifically, soils within the study area depicted a distinctive colour that was related to land use. For instance, soil samples obtained from the agriculture land use were the darkest (i.e., reddish-brown) when compared to the other land uses. Soils from the eroded land were the lightest (light brown) in colour, while the rangeland and built-up soils were slightly darker when compared to eroded soils; however, their colour was still lighter than that of the agriculture soil samples. As mentioned previously, a darker soil will have a lower reflectance compared to a lighter soil. Similarly, the agriculture soil gave off a lower reflectance due to its darker soil colour, whereas the eroded soil gave off the highest reflectance due to its lighter soil colour. The lighter rangeland and built-up soils also had a higher reflectance. Additionally, the spectra obtained for each land use was also influenced by the amount of SOM present within the soil, as SOM is related to soil colour (Šestak *et al*., 2018). The agriculture land use had the highest SOM content while eroded land had the lowest SOM content, and land uses rangeland, and built-up had SOM contents that were in between these two land uses (as seen in Table 3.2). The higher SOM content present within the agriculture land use had resulted in the lowest reflectance spectrum due to the presence of dark pigments such as humic acid. Humus, which is related to stabilized SOM, causes the soil to become much darker in colour (Bot and Benites, 2005). However, within the eroded soils, the formation and accumulation of SOM were hindered and therefore, the formation of darker pigments in soil was limited. Thus, allowing for the highest reflectance spectrum. Moreover, the spectra obtained from the rangeland land use had a higher SOM content; however, its reflectance was higher than the built-up land use, which had a lower SOM content. This may be attributed to the stage of decomposition, formation, and accumulation of SOM which can also influence the reflectance spectrum (Stenberg *et al*., 2010). The fact that the spectra

obtained in this study was influenced by soil colour, further supports the fact that there is a relationship between SOM content and reflectance of soil samples.

3.3.3 Relationship between SOM content and reflectance spectra

The calculated correlation coefficients (*r*) between SOM content and spectral reflectance are illustrated in Figure 3.8. It is evident that the calculated correlation coefficients at each wavelength varied in both the raw and pre-processed spectral reflectance of soil samples. Figure 3.8a and b indicated the overall correlation coefficients for the raw and pre-processed spectra (SG+FDT), respectively. It can be clearly seen that after pre-processing, the curve is much smoother, especially from 1000 nm to 2400 nm, while the shape differed between the curves. In terms of the shape of the curves, Figure 3.8a starts off high at 400 nm and then drastically decrease to about 600 nm; from 600 nm to 800 nm, the coefficient values become stable before increasing and decreasing along the rest of the wavelengths. Subsequently, significant peaks in correlation coefficient values can be seen at around 1400 nm, 1900 nm, and 2200 nm. However, in Figure 3.8b, the curve starts off low and continues to decrease to about 550 nm, after which it begins to increase to a certain point which is then followed by several peaks and valleys that cover the remaining wavelength range (1000 nm to 2400 nm). The correlation coefficients along the covered range were negative throughout for the curve in Figure 3.8a, while the curve in Figure 3.8b ranged between both the positive and negative values of the covered range. In Figure 3.8a, the highest correlation coefficient was around 400 nm (r= -0.19), and the lowest correlation coefficient was around 700 nm (r= -0.39). Contrastingly, in Figure 3.8b, the highest correlation coefficient was around 2004 nm (r= -0.50), and the lowest correlation coefficient was around 1774 nm (r= -0.01).

The curves in Figure 3.8c, d, e, and f also depict similar trends and shapes to that of the raw overall correlation coefficient curve (Figure 3.8a). The correlation coefficient curves for the different land uses depicted smoother curves but with varying coefficient values. In Figure 3.8c, the shape of the curve is similar to that of the raw curve shown in Figure 3.8a, but until 1800 nm. The curve starts high at 400 nm and then decreases to about 600 nm to 800 nm, thereafter increasing until it decreases around 1800 nm, after which the coefficient values seem to be decreasing until 2400 nm. The correlation coefficients for the agriculture land use were also negative, with the highest correlation coefficient around 400 nm (r= -0.31) and the lowest correlation coefficient around 2324 nm (r= -0.58). Subsequently, in Figure 3.8d, it is evident that there is a slight decrease in coefficient values along the 400 nm region, followed by an increase from 450 nm to 600 nm and a slight decrease thereafter to 964 nm. However, there

was a turning point in values around 1044 nm, where values began to suddenly increase greatly until 1764 nm, after which values began to increase and decrease with dips around 2204 nm and 2364 nm and a peak around 2384 nm. The correlation coefficients for the built-up land use were also along the negative range, with the highest correlation coefficient around 1854 nm (r= -0.29) and the lowest correlation coefficient around 1044 nm (r= -0.52). Furthermore, in Figure 3.8e, a sharp peak is seen around 400 nm followed by a sudden decline from 444 nm and then a sudden incline to 1000 nm; thereafter, a sudden peak is seen at 1404 nm followed by a consistent incline to 2374 nm. From this point and onwards, a decreasing pattern can be seen until 2400 nm. The correlation coefficients for the eroded land were along the positive (400 nm to 524 nm and 924 nm to 2400 nm) and negative (534 nm to 914 nm) range, with the highest correlation coefficient around 2314 nm (r= 0.31) and the lowest correlation coefficient around 644 nm (r= -0.27). Similarly, Figure 3.8f depicts trends similar to that of Figure 3.8e, where the curve starts off high (400 nm) then slowly decreases only to give rise to a great dip (514 nm to 914 nm) in the curve. From 974 nm, the curve begins to increase until the first peak is seen (1414 nm); from this point onwards, the curve increases and decreases, with several peaks being evident until 2400 nm. The correlation coefficients for the rangeland land use were along the positive range only, with the highest correlation coefficient around 1984 nm (r= 0.61) and the lowest correlation coefficient around 624 nm (r= 0.10).

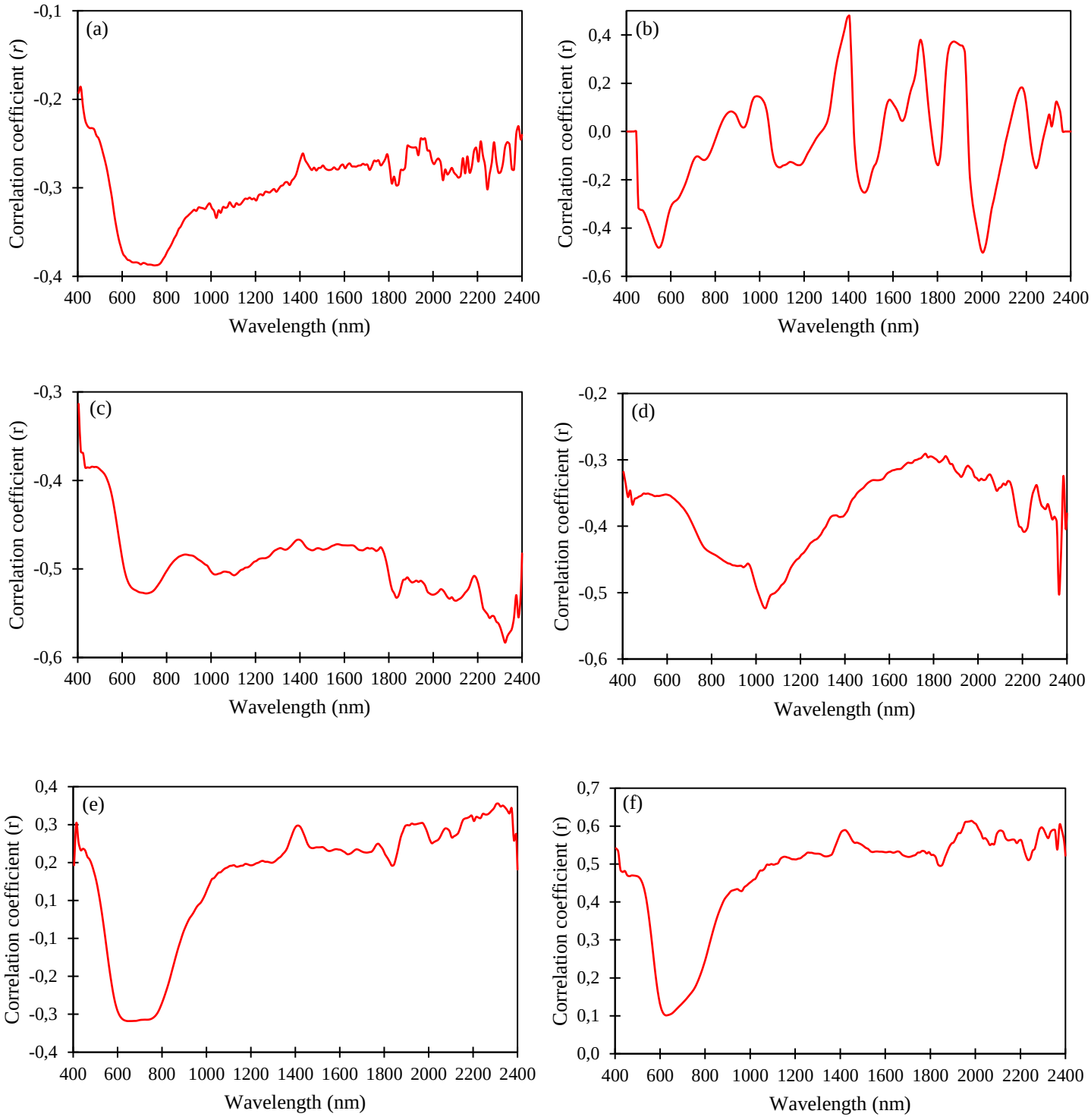


Figure 3.8: Calculated correlation coefficients between SOM content and spectral reflectance of: (a) all raw spectra, (b) all pre-processed spectra, (c) SG pre-processed agriculture spectra, (d) SG pre-processed built-up spectra, (e) SG pre-processed eroded spectra, and (f) SG pre-processed rangeland spectra

In order to examine the relationship between SOM content and the spectral reflectance collected from soil samples, correlation coefficients were calculated at each wavelength along the covered range (400 nm to 2400 nm). These correlation values were calculated for the overall SOM content and spectral reflectance as well as for the different land use types (agriculture, built-up, eroded, and rangeland). With regards to the overall correlation coefficients that were calculated, both the raw and pre-processed reflectance spectra using the SG and FDT method were used. From the results obtained in Figure 3.8a and b, it was evident that the pre-processed curve (Figure 3.8b) was much smoother than that of the raw spectra curve (Figure 3.8a). The SG smoothing method assisted with this in that it removed any noise that was caused by the instrument or by other factors. It was also evident that there was a significantly greater number of peaks and valleys along the covered spectrum compared to the raw spectra curve (Figure 3.8a). This was a result of the FDT method used during the pre-treatment process of the raw spectra. Within the raw spectra curve, it can be clearly seen that the correlation coefficients in the VIS region (400 nm to 700 nm) were higher than the NIR region (700 nm to 2400 nm). As mentioned previously, the VIS region contains the visible light that can be seen with the naked eye, which means it is related to colour, more specifically, soil colour. The absorption bands within this region are largely influenced by SOM content and may have resulted in higher correlation values compared to the NIR region (Duarte and Duarte, 2019). However, these values begin to decrease as one moves from the VIS to the NIR region. Along this region, it was mentioned that three major absorption peaks at 1400 nm, 1900 nm, and 2200 nm existed. It is along these peaks the correlation coefficients are the highest in the NIR region (Figure 3.8a). This may indicate that a correlation exists between SOM content and the absorbance in this region. However, from the results obtained, the correlation coefficients were not high; in fact, they were all negative (Figure 3.8a). This suggests that a weak negative relationship exists between the SOM content analysed and the spectral reflectance of the soil samples collected. Conversely, the pre-processed curve had higher correlation coefficients in the NIR region compared to the VIS region. This may be a result of the FDT method, which according to Kooistra *et al*., (2001), generally has little influence on the VIS region and beginning of the NIR region due to absorption features being quite broad in these regions. Such reasoning can also account for the shape of the curve in these regions compared to the rest of the NIR region, where distinct peaks and valleys or maximum and minimum correlation coefficients are located. Subsequently, the FDT method removes the baseline from the spectra whilst emphasizing absorption features (Buddenbaum and Steffens, 2012). And, as mentioned previously, the NIR region consists of absorption peaks that have now become amplified due

to the applied pre-processing methods used. More specifically, this could have influenced the higher correlation coefficients seen in this region, especially at 1404 nm and 2004 nm (Figure 3.8b). From the results obtained, the correlation coefficients were not that high and were both negative and positive, but they improved after the pre-processing of the spectra (Figure 3.8b). This suggests that a moderate relationship exists between the SOM content analysed and the pre-processed spectral reflectance of the soil samples collected. These results may have been influenced by the previously mentioned reasons for the low SOM content within the area and the spectral reflectance of the soil samples. More specifically, these results were similar to the studies by Sullivan *et al*., (2005), Gao *et al*., (2017), Hong *et al*., (2018b), and Wei *et al*., (2020), where negative correlations were also seen between SOM content and spectral reflectance for the overall raw and pre-processed spectra.

Additionally, the correlation coefficients obtained from the different land use types focused only on the relationship between pre-processed reflectance spectra using the SG method and SOM content. It can be seen that the correlation coefficients curves varied along the covered wavelength range for each of the land use types. All the land use types exhibited a similar trend in correlation values in the VIS region except for the built-up land use curve (Figure 3.8d). This may be because soil samples obtained from this land use had variations in soil colour along the study area. This variation coupled with the relatively low SOM content within this land use type may have influenced the spectral reflectance, thus resulting in the moderately weak correlation along this region. However, the weaker correlation between SOM content and reflectance for the other land use types in the VIS region may be a result of the influence from previously mentioned factors such as the presence of absorption bands, soil colour, soil type, etc. In contrast, the correlation coefficients in the NIR region were more or less high for all the land use types, with the exception of the agriculture land use (Figure 3.8c-f). This may be attributed to the fact that agriculture land use had the highest SOM content and that its reflectance was influenced by this, therefore resulting in negative correlations that are weak along this region. According to Virgawati *et al*., (2019), this indicates that SOM and reflectance of soil samples have "negative associations with strong relationships". More specifically, it means that as SOM content increases, the reflectance will decrease along the NIR region. In the case of the other land uses, the coefficients were higher, but the relationships between SOM content and reflectance were weaker. This, coupled with the number of absorption bands in the NIR region, may have caused the higher correlation values. Furthermore, correlations were negative for the agriculture, and built-up land uses, positive for the rangeland land use, and a

combination of both for the eroded land. Therefore, indicating that agriculture SOM content had a negatively moderate to strong correlation with its reflectance spectra, built-up land use had a similar trend, eroded land had a moderately positive correlation with its reflectance spectra, and rangeland land use had a positively strong correlation with its reflectance spectra. It also suggests that an increase in SOM content will result in a decreased reflectance of the soil samples and that the converse relationship will be seen as SOM content decreases.

These results indicate that a relationship exists between SOM content and the reflectance of soil samples, despite being quite a weak one. Soil types within the different land use types may have significantly influenced the percentage of SOM within these soils and, therefore, its reflectance.

3.3.4 PLSR modelling analysis

The performance results of the different PLSR models created using both raw spectral data and different pre-processing methods are displayed in Table 3.3. With regards to the R^2 values, for the calibration dataset, values range from 0.28 (lowest) to 0.49 (highest), while the validation dataset values range from 0.22 (lowest) to 0.34 (highest). However, in terms of the RMSE, the error values were higher than that of the R^2 values. For the calibration dataset, RMSEC values ranged from 0.73% to 0.98%, while in the validation dataset, RMSEP values ranged from 1.08% to 1.11%. Regarding each of the PLSR models, the R and SG models had similar results where both their R^2 and RMSE values were quite similar with little to no difference. However, the FDT and SG+FDT models had higher R^2 values with lower RMSE values in the calibration dataset compared to the validation dataset where the converse trend was seen. Subsequently, the FDT model has the highest R^2 value (0.49) and lowest RMSEC value (0.73%) in the calibration dataset, but in the validation dataset, it had the lowest R^2 value (0.22) and the highest RMSEP value (1.11%). A similar but contrasting trend can also be seen for the SG model, while the R and SG+FDT models have values and trends that fall in between the other PLSR models.

Table 3.3: The statistical summary of the different PLSR models created using raw and different pre-processed spectral data and SOM content

PLSR model type	Calibration		Validation	
	R^2	RMSEC (%)	R^2	RMSEP (%)
R	0.31	0.95	0.33	1.08
SG	0.28	0.98	0.34	1.08
FDT	0.49	0.73	0.22	1.11
SG+FDT	0.46	0.75	0.25	1.09

In Figure 3.9, the measured SOM content is plotted against the predicted SOM content that was derived from the PLSR models. With regards to the plots, it is evident that they have a more or less constant trend as the lines of best fit follow somewhat of a horizontal pattern. In terms of Figure 3.9a, it is evident that the data points are distributed randomly in close proximities to their respective best fit lines for both the calibration and validation datasets, with the exception of some points. It can also be seen that the points are distributed mainly along the lower to central region of the plot. The SG plot (Figure 3.9b) is quite similar to that of the R plot (Figure 3.9a), except for few points having slightly different values. In terms of Figure 3.9c, it is evident that the data points are more compact and closer to their respective lines of best fit for both the calibration and validation datasets. However, the data points seem to be more spread out along these best-fit lines. It can also be seen that for the calibration dataset, certain points below the line of best fit are found spread out and are deviating away as compared to those points above the line. Similarly, Figure 3.9d exhibits trends and distributions of data points that are comparable to those outlined in Figure 3.9c; however, the only difference is that values at certain points may vary slightly.

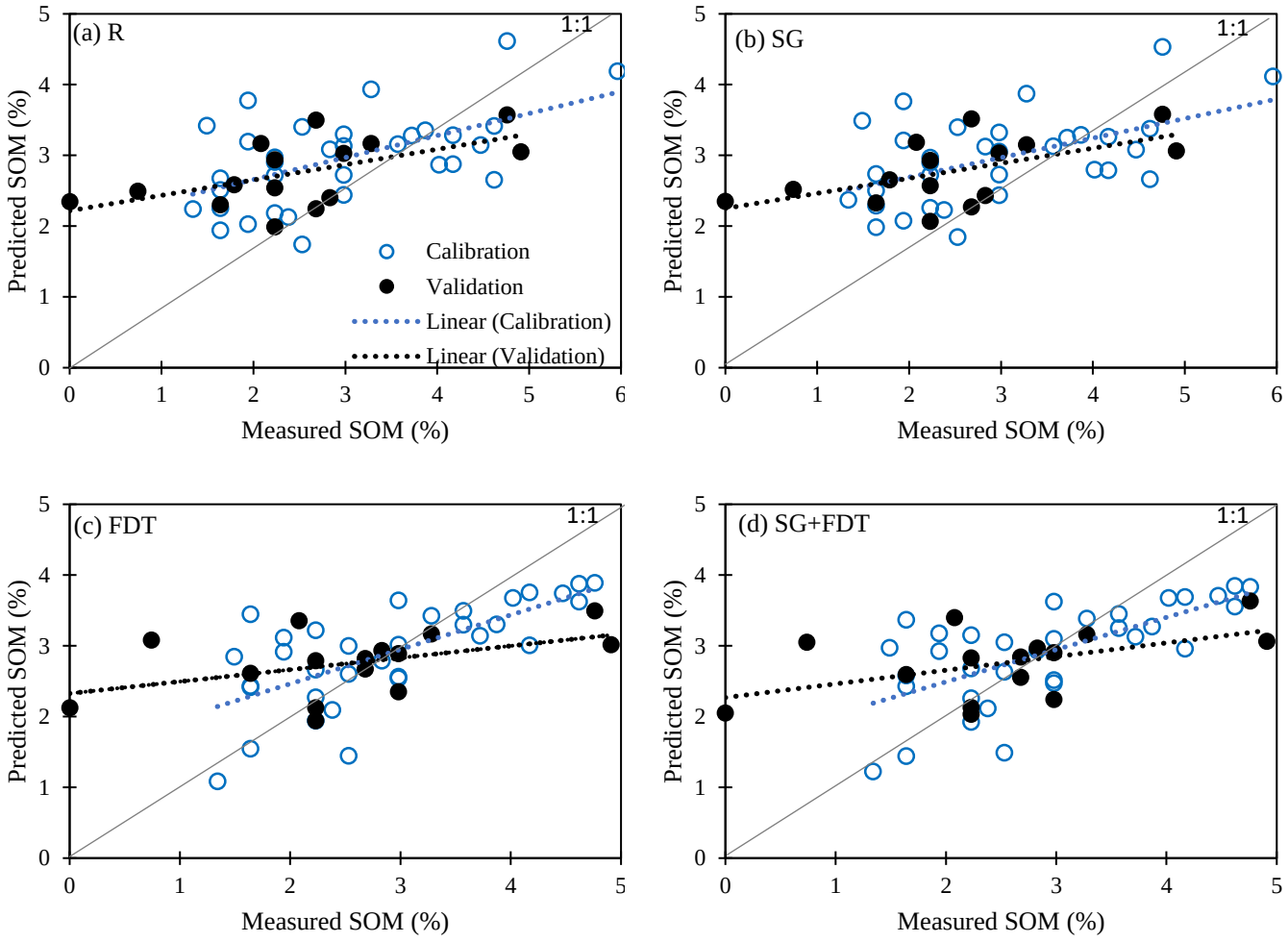


Figure 3.9: Scatterplots of measured versus predicted SOM content using the four derived PLSR models: (a) R, (b) SG, (c) FDT, and (d) SG+FDT

The above results provide an overview of the assessment and accuracy of the PLSR models derived from the use of raw spectral data and different pre-processing methods and measured SOM content in terms of predicting SOM content. With regards to the evaluation statistics outlined in Table 3.3, it was evident that each of the four PLSR models performed differently to a certain extent. The R^2 values indicated the percentage of variance of the dependent variable (SOM content) that can be explained by the independent variables (soil reflectance spectra) (Vibhute *et al*., 2018). However, the RMSE measures the square root of the mean of the square of all of the error and is a good measure of accuracy (Neill and Hashemi, 2018). For instance, the R model, which involved the use of raw spectra and measured SOM content had one of the highest R^2 values and lowest RMSE values for both the calibration and validation datasets. The R^2 values of 0.31 and 0.33 were obtained for the calibration and validation dataset, respectively. Therefore, this meant that the R model accounted for 31% and 33 % variability in SOM content

using the raw spectra for the calibration and validation dataset, respectively. These lower percentages can be attributed to the fact that the raw spectra were used in creating the model. And, as mentioned previously, raw spectral data can be contaminated with instrument noise, light scattering effects or energy responses that vary along the different wavelengths (Gao *et al*., 2017; Wang *et al*., 2017; Li *et al*., 2019). This noise present in the spectral data could have lowered the performance of the PLSR model and hence its R^2 values. The RMSEC and RMSEP values, on the other hand, were in close range with one another, with only a small difference existing between them. This can be inferred by the scatterplot shown in Figure 3.9a. The RMSEC value was lower than one, while the RMSEP value was greater than one. This suggests that there may be either an over-fitting or under-fitting of the PLSR model (Castillo *et al*., 2005). However, in this particular case, it may be over-fitting which was fairly minimal as the error for the validation dataset is larger than the error for the calibration dataset. Similar results were obtained by Bangelesa *et al*., (2020), where the RMSEP value was higher than that of the RMSEC value, which then led to the overfitting of the PLSR model that involved the use of raw spectral data. Additionally, the influence of certain percentages of measured and predicted SOM content could have resulted in the obtained RMSE values for the PLSR model using raw spectra.

With regards to the SG model, it was evident that it produced the lowest R^2 value with the highest RMSEC for the calibration dataset but had the highest R^2 value with the lowest RMSEP for the validation dataset. More specifically, R^2 values of 0.28 and 0.34 were obtained for the calibration and validation dataset, respectively (Table 3.3). This suggested that the SG model accounted for 28% variability in SOM content for the calibration dataset and 34% variability in SOM content for the validation dataset. The slight improvement in the R^2 value of the SG model is directly attributed to the fact that the spectra used in creating the PLSR model were pre-processed using the SG method. This pre-treatment method is commonly used in spectral analyses as a means to get rid of any inappropriate information that may exist within the spectral data (Gholizadeh *et al*., 2013). Also, the SG method is used to reduce any noise that may exist within the spectra (e.g., baseline-drift, tilt, reverse, etc.) and increase the signal to noise ratio (Savitzky and Golay, 1964; Agrawal and Deshmukh, 2018). The removal of noise from the raw spectra eliminated any unnecessary information and/or inconsistencies that may have existed within it. And, therefore the removal of such obstacles resulted in an increased performance of the PLSR model using SG data. In terms of the RMSE values, it was evident that the RMSEC value was lower than one while the RMSEP value was greater than one (Table

3.3). However, they were in a closer range with one another and with only a small difference existing between them. Once again, this may be a result of the raw spectra being pre-processed prior to its use in creating the PLSR model. According to Vasques *et al*., (2008), the SG smoothing method is a common pre-treatment method used to improve the SOM model results. This further suggests that the SG pre-processing method did, in fact, have an influence in improving the performance of the PLSR whilst reducing any errors.

In contrast, the FDT model produced the highest R^2 value with the lowest RMSEC for the calibration dataset but had the lowest R^2 value with the highest RMSEP for the validation dataset. The results revealed R^2 values of 0.49 and 0.22 for the calibration and validation dataset respectively (Table 3.3). To be more specific, the FDT model accounted for 49% variability in SOM content for the calibration dataset and 22% variability in SOM content for the validation dataset. This improvement in R^2 values is also related to the fact that the raw spectra were pre-processed using the FDT method before creating the PLSR model. The FDT is a transformation method that assists in the correction of baseline effects found in raw spectra. It also assists in smoothing, removing noise, and reducing particle size effects found in the raw spectra (Bilgili *et al*., 2010; Srivastava *et al*., 2016; Vibhute *et al*., 2018). Therefore, this pre-treatment method assisted in the removal of any irregularities or unnecessary information present within the raw spectra, which may have caused a hindrance to the performance of the model. Regarding the RMSE values, it was evident that the RMSEC value was relatively lower than one while the RMSEP value was greater than one (Table 3.3). Moreover, the values were in a further range from each other and with only a significant difference existing between them. This can be accounted for by the influence of noise that was present in the raw spectra, which was used to create the PLSR model using the FDT method. According to Fang *et al*., (2018), any noise present within the spectral data may be amplified when the FDT method is carried out. Therefore, the noise present within the raw spectra could have been amplified, which further resulted in the alteration of the data being used to create the PLSR model, thus accounting for the obtained RMSE values for the model. Subsequently, influence from certain percentages of measured SOM content could have resulted in the higher RMSEP value.

In terms of the SG+FDT model, the results revealed a high R^2 value with the low RMSEC for the calibration dataset and a low R^2 value with a high RMSEP for the validation dataset. The R^2 values of 0.46 and 0.25 were obtained for the calibration and validation dataset, respectively (Table 3.3). These values indicate that the SG+FDT model accounted for 46% variability in SOM content for the calibration dataset and 25% variability in SOM content for the validation

dataset. For this particular model, both of the pre-processing methods were used. Firstly, the SG method was applied to the raw spectra, and then the FDT method was applied to the SG pre-processed spectra. This data was then used to create the PLSR model. Due to the combination of the two pre-processing methods, the results of the model were improved. As outlined by Li *et al*., (2015), such a combination of SG smoothing and FDT has proven to be quite superior in improving PLSR models. Additionally, the use of pre-processing methods assists in the removal and/or correction of irregularities and spectral noise that may exist within the spectral data, and therefore such reasoning can be used to explain the improved performance of this particular model. The results of the RMSE values revealed that the RMSEC value was lower than one while the RMSEP value was greater than one (Table 3.3). Since this model involved the use of two pre-processing methods, its results are also influenced by the above-mentioned reasoning for the previous PLSR models. For instance, when the SG method was carried out on the raw spectra, some of the noise may not have been removed entirely, and when the FDT was carried out using this data, any existing noise may have been amplified and thus caused the model performance to decrease while the error increases. Subsequently, influence from certain percentages of measured and predicted SOM content could have resulted in the higher RMSEP value. Moreover, the trends in the results for this model were similar to those witnessed in the FDT model.

In addition, when comparing the results of the calibration dataset with that of the validation dataset, it was clearly seen that the calibration dataset performed better than the validation dataset (Table 3.3). This may be a result of the number of samples used in each dataset. For instance, a smaller number of samples may either over or underestimate the value that is being calculated compared to a greater number of samples being used for the calculation. According to Kuang and Mouazen (2012), an increase in the number of samples is known to improve the accuracy of the prediction and its RMSEP. And, therefore due to the calibration dataset having a greater number of samples than the validation dataset, its performance was greater when compared to the validation dataset. Furthermore, the results of these datasets revealed that the calibration dataset had higher R^2 values with lower RMSEC values while the validation dataset depicted the converse trend (Table 3.3). These results fall in line with those reported by Shepherd and Walsh (2002) and Kuang and Mouazen (2012), where R^2 values increased, and the RMSE for the calibration dataset decreased due to a greater number of samples being used in the calibration dataset compared to that of the validation dataset. Subsequently, the PLSR

method itself and the manner in which it calculates the predicted values could have influenced the values of the evaluation statistics as well.

Moreover, the results had revealed that the performance and prediction accuracy varied for each of the different PLSR models. According to the results, it was evident that models which involved the use of pre-processing methods outperformed the model that involved the use of raw spectra. More specifically, it was the FDT PLSR model that produced the most accurate model, followed by the SG+FDT model, SG model, and then lastly, the R model, which was the least accurate. These results fall in line with the findings from other studies, which also found an increase in the prediction accuracy of SOM content using reflectance spectra that have been pre-processed (Kuang and Mouazen, 2012; Nawar *et al*., 2016; Wang *et al*., 2017; Šestak *et al*., 2018; Vibhute *et al*., 2018; Bangelesa *et al*., 2020). Overall, it can be deduced that R and SG models did not do an efficient job in predicting SOM content compared to the FDT and SG+FDT models, which did a fair job in the prediction of SOM content. This can further be inferred from the scatterplots seen in Figure 3.9, which provide a graphical representation of the relationship between the measured and predicted SOM content. Thus, due to previously mentioned reasons coupled with the different soil properties in the varying land uses may have implicated the results of models created using PLSR analysis.

3.4 Conclusion

The study has provided a critical analysis of the relationship between the SOM that is found within the Emakhosaneni area, particularly within its major land use types and the spectral reflectance from its soils. This relationship revealed that as the SOM content increases within a soil sample, the spectral reflectance is lowered. This was particularly relevant for the agriculture and eroded soil samples, where the highest and lowest SOM content was seen, respectively. Furthermore, the calculated correlation coefficients depicted negative correlations with a moderately strong relationship between SOM content and spectral reflectance along the covered wavelength range. Subsequently, the results from the PLSR revealed that those models in which pre-processed data were used performed better than the model in which raw spectral data was used. More specifically, it was revealed that each of the different pre-processing methods handled spectral noise, baseline offsets, absorptions, and any other irregularities present within the spectral data differently and also highlighted its importance and influence on the performance and accuracy of the PLSR models. Therefore, it was concluded that pre-treatment by the FDT method produced a positive response on the robustness and performance of the PLSR model when it comes to the prediction of SOM content for this particular study.

The FDT PLSR model's performance and accuracy were followed by the SG+FDT model, then the SG model, and lastly, the R model. Despite the fair performances of these models, they highlighted the importance of VNIR spectroscopy in the measurement and prediction of SOM content compared to the conventional laboratory approaches, which can be rather costly and time-consuming. Since SOM governs the physical, chemical, and biological properties of soil, it is imperative that SOM and soil are managed and protected in a sustainable manner for future generations to come. It is therefore important for studies such as the above to be carried out so that useful information can be gathered efficiently regarding the estimation, prediction, and quantification of SOM content.

CHAPTER FOUR

Spatial prediction of Soil Organic Matter within the Emakhosaneni area, KwaZulu-Natal using Sentinel-2 remote sensing data

Abstract

Soil Organic Matter (SOM) consists of decaying plant and animal material at various stages of decomposition, substances released by plant roots, and soil organisms. It is responsible for supporting physical, chemical, and biological functions within soil. Techniques such as remote sensing provide a new approach in assessing SOM. Its low-cost, high efficiency, large-scale, non-destructive, and rapid data acquisition has proven to be effective and can be quite valuable in the prediction of SOM. Therefore, this study aimed to assess the relationship between SOM content and satellite remote sensing data within the soil of the Emakhosaneni area, KwaZulu-Natal, South Africa. Thirteen random soil samples were obtained from each of the four major land uses (Agriculture, Built-up (Residential), Eroded, and Rangeland); hence, a total of 52 soil samples were obtained. Samples were oven dried overnight (105°C), crushed, sieved (425 µm), and analysed using the Walkley-Black method to determine SOM content. Auxiliary variables such as the reflectance values of each band and calculated spectral indices were obtained from remote sensing data (Sentinel-2). A Principal Component Analysis was performed on the spectral indices dataset to reduce multicollinearity and redundancy. Thereafter, two Multiple Linear Regression (MLR) models were created using these variables. In terms of predicting SOM, geostatistical methods such as Ordinary Kriging (OK) and Simple Kriging (SK) together with hybrid geostatistical methods Regression Ordinary Kriging (ROK) and Regression Simple Kriging (RSK) were used. OK and SK were performed on the measured SOM data only, while ROK and RSK were performed on the residuals of the two MLR models and summed together with predictions from the two MLR models to give the final prediction of SOM. The performance of each derived model was then evaluated. The results revealed that the hybrid geostatistical methods performed better than the geostatistical methods. More specifically, the ROK and RSK using spectral bands had the highest R^2 values and lowest ME and RMSE values, followed by the ROK and RSK using principal components (PCs), and lastly, the OK and SK methods, which had the lowest R^2 values and highest ME and RMSE values. These results can be attributed to the contribution of auxiliary information, which provided additional information that boosted the predictive performance of these methods. Overall, the ROK and RSK using spectral bands accounted for 63% variability of the predicted SOM content. SOM plays a critical role in soil functioning and should be managed and protected sustainably for future generations to come. Therefore, the use of efficient remote sensing data can assess SOM content to ensure that it is able to support and facilitate different uses.

Keywords: Soil Organic Matter, Regression Kriging, Prediction, Geostatistics, Walkley-Black

4.1 Introduction

SOM contributes the smallest amount towards the soil's composition. According to the FAO (2017a), SOM refers to "the organic constituents in soil in various stages of decomposition such as tissues from dead plants and animals, materials less than 2 mm in size, and soil organisms". Its formation is primarily a biological process directly or indirectly influenced by flora and fauna present within the soil (Bot and Benites, 2005). According to Wetterstedt (2010), the source materials for the formation of SOM include; plants, either in the form of aboveground litter (twigs, leaf litter, and stems) or belowground material (exudates, root litter, and mycorrhizal hyphae), which usually enter and mix with the soil where they undergo decomposition. Products of decomposition along with newly synthesized material accumulate in the soil and produce a more complex form of SOM (Osman, 2012b). Therefore, SOM components usually include; plant and animal residue, soil organisms, stable organic matter (humus), and actively decomposing matter (Reddy, 2016). Despite the small contribution of SOM towards the soil, it supports various physical, chemical, and biological functions in the pedosphere and its interacting spheres. These functions assist in providing ecosystem services to humans and maintain soil functioning for many plants and animals. More specifically, SOM plays a significant role in soil quality and hence, the overall fertility and productivity of soil.

Over the years, various methods were developed and modified to measure and estimate SOM content in soils (Motsara and Roy, 2008). Commonly used methods are usually laboratory-based and include; the Walkley-Black, Loss On Ignition, Automated dry combustion and, Humic matter methods (Roper *et al*., 2019). However, these methods can be relatively costly and time-consuming, especially when study areas are large with extensive soil sampling. Therefore, even though common methods are still being used, new techniques and technologies such as remote sensing are being incorporated into SOM analysis. According to Wang *et al*., (2017), the low-cost, high efficiency, large-scale, non-destructive, and rapid data acquisition of remote sensing has proven to be a useful tool for strengthening or improving traditional methods.

Remote sensing can be explained by the use of EMR in order to gather information about an object, feature, or phenomenon without any physical contact (Gholizadeh *et al*., 2018). More detailed definitions of remote sensing exist and can vary based on its understanding and application. However, the focus is placed on the interaction between an object and the EMR that is either absorbed, reflected, or transmitted. This interaction is responsible for carrying information about the object of interest or the processes that are occurring within it (Milton,

2004). Such information or data is usually gathered from different remote sensing sources. According to Angelopoulou *et al*., (2019), remote sensing data sources can either be; spaceborne, airborne, or ground-based. More specifically, spaceborne systems typically include a variety of active and passive sensors mounted onto a spacecraft (e.g., satellites, rockets, and space shuttles) orbiting the earth that operates at varying wavelengths of the electromagnetic spectrum (Zhu *et al*., 2017). Similarly, airborne systems have sensors mounted onto an aircraft (e.g., balloons, airplanes, and drones) that capture information about the earth's surface from up above (Myers and Miller, 2007).

Consequently, ground-based remote sensing usually occurs along the earth's surface where sensors are mounted onto vehicles, cranes, towers, etc., or involve the use of handheld devices to capture information. Each of these remote sensing systems is unique in how they acquire information about the earth and have their strengths and weaknesses. For instance, satellite imagery from spaceborne systems has become quite efficient in obtaining information about the earth.

Satellite remote sensing has gained immense popularity over the years when it comes to observing, measuring, and recording EMR that is reflected or emitted by the earth and its environment. This may be due to the fact that satellite remote sensing covers a larger area for data acquisition due to it having a high vantage point (i.e., outer space) and a broad field of view (Sivakumar *et al*., 2004). A broader coverage in surface area allows for more information to be gathered and viewed within a single view and makes it easier to examine the spatial variability of an area. Furthermore, the repetitive coverage of satellite remote sensing ensures a timely update of areas and hence, data (Navalgund *et al*., 2007). The frequent collection of data allows information to be gathered and updated in near to real-time, which can prove to be beneficial when trying to observe changes in the environment. According to Fu *et al*., (2020), satellites in outer space can "successively observe the whole globe or an assigned part of it within a defined time period". Subsequently, this type of remote sensing can operate at various scales and resolutions. Such capabilities allow users to acquire information from a wider range that is unique and relevant to their use. Overall, satellite remote sensing has greater advantages compared to airborne and ground-based remote sensing; due to its better coverage in time and area extent (Sivakumar *et al*., 2004). According to Roy *et al*., (2017), the "synoptic view, repetitive coverage with calibrated sensors to detect changes, observations at different resolutions, provides a better alternative for natural resources management as compared to traditional methods". This, coupled with the fact that it is relatively cost-effective, allows for a

greater understanding, together with the mapping, modelling, measuring, managing, and monitoring of the earth and its environment (Lefsky and Cohen, 2003).

Satellite imagery obtained from remote sensing has been used to gather information regarding the soil present on earth. Like any other natural resource, feature, material, etc., on earth, soil has its very own reflectance that is unique to its properties, characteristics, and composition. The EMR of soils that is captured by remote sensing satellites is a result of the physical and chemical soil surface properties, such as the matrix of the topsoil (Ben-Dor *et al*., 2008; Mulder *et al*., 2011). In other words, it captures the spectral reflectance of soil which is related to its physical and chemical properties. These properties can vary on both a spatial and temporal scale across the electromagnetic spectrum. The reflectance of soils is usually evident along the VIS (400-700 nm), NIR (700-1100 nm), and SWIR (1100-2500 nm) regions of the electromagnetic spectrum (Krishan *et al*., 2014). According to Gallo *et al*., (2018), the information retrieved from these regions can be used to describe the relationship between spatio-temporal and quantitative soil information changes that occur over large areas. Subsequently, within these regions, the reflectance of soil increases with increasing wavelength. However, factors such as moisture content, soil texture (proportion of sand, silt, and clay), surface roughness, presence of iron oxide, and organic matter content, as well as other factors, are known to influence the reflectance of soils (Mujumdar and Kumar, 2012). These factors are complex in nature, variable, and are interrelated. Satellite imagery of soils provides detailed information on the spatial variability of important properties and characteristics of soil, which plays a major role in the soil survey (Žížala *et al*., 2019). Furthermore, it can be used to establish regional or national reference systems for soil properties by providing an alternative to costly techniques such as traditional sampling and analysis of soils (Castaldi *et al*., 2019a). According to Gholizadeh *et al*., (2018), several studies over the years revealed that data obtained from satellite imagery have proven to be efficient when it comes to assessing surface soil characteristics (Zhang *et al*., 2013; Danoedoro and Zukhrufiyati (2015); Castaldi *et al*., 2016; Vågen *et al*., 2016; Gomez *et al*., 2018).

Similarly, satellite imagery can be used to acquire information on the SOM present within the soil. Like soil, SOM reflectance is a result of its composition together with its physical and chemical properties. For example, the remains of plants and animals contain sugars, proteins, lignins, celluloses, hemicelluloses, waxes, and lipids (Osman, 2012b). These biochemical materials are responsible for influencing the reflectance in the VIS, NIR, and SWIR region of the electromagnetic spectrum (Bartholomeus *et al*., 2008). However, several studies have

revealed that SOM has a unique spectral response in the VIS and NIR regions of the electromagnetic spectrum (Wang *et al*., 2017). Subsequently, factors that influence the reflectance of soils from satellite remote sensing will inadvertently influence the reflectance of SOM. These factors include; soil moisture, soil roughness, vegetative cover, crop residue, atmospheric conditions, and the variable spectral and spatial resolution of different sensors (Dou *et al*., 2019). However, it is not only these factors that are responsible for influencing the reflectance of SOM. According to Zhuo *et al*., (2008), the recorded reflectance from satellite sensors are often a (mixed) result of surface components (i.e., mixed pixel problem) together with atmospheric scattering and absorption processes, which usually mask the diagnostic soil features. Therefore, the returned energy from soil can sometimes not truly depict information that is relevant to SOM. Moreover, the characteristics and properties of SOM can also govern the reflectance of EMR and hence, the type of information that is acquired by the satellite. For instance, SOM has a direct influence on soil colour. This may result from the decomposition of humus, which causes the soil to become much darker in colour (Bot and Benites, 2005). Therefore, a darker soil will be indicative of a high SOM content compared to a lighter-coloured soil. According to Daniel *et al.,* (2001), this indicates a relationship between the VIS reflectance and SOM. More specifically, it dictates the relationship between SOM content and its VIS light reflectance (He *et al*., 2009). Simply put, darker soils contain more SOM and will reflect less light back, whereas lighter soils have less SOM and will reflect more light back. This feature, therefore, allows for SOM to be easily identified within the VIS region. Furthermore, humic substances such as humins and humic acid (darkest pigment of SOM) are responsible for reducing the spectral reflectance of SOM over the VIS to SWIR spectral range, while fluvic acid is known not to influence the SOM reflectance (Mohamed *et al*., 2018). Subsequently, the stage of decomposition, formation, quality, and accumulation of SOM can influence its reflectance.

There are certain spectral wavebands that have a sensitive relationship with SOM, and this has been reported by several studies (Amin *et al*., 2020). For instance, Viscarra Rossel *et al*., (2006) revealed a good correlation of SOM along wavelengths such as; 410 nm, 570 nm, and 660 nm along the VIS region. Liu *et al*., (2009) confirmed that reflectance between the range 620 nm to 810 nm was related to SOM and that the highest correlation coefficient was at 710 nm. Furthermore, Zheng *et al*., (2016) also revealed that the wavelength range of 500 nm to 700 nm had the highest correlation with SOM. Wang *et al*., (2017) stated that these five bands (440 nm, 560 nm, 625 nm, 740 nm, and 1336 nm) are the optimal spectral bands for estimating

SOM. And Peng *et al*., (2013) discovered that reflectance between 570 nm and 630 nm increased after removing SOM from soil. These sensitive bands justify that SOM can be easily identified within the VIS and NIR regions. Furthermore, these results indicate that these response wavelengths of SOM can be covered by those band widths found in most multispectral satellites, therefore suggesting that they can be used in the assessment of SOM content (Dou *et al*., 2019).

The use of satellite imagery in the assessment of SOM can sometimes be difficult due to previously mentioned factors, which influence the reflectance of soil and SOM from spaceborne systems. This has limited the number of studies using satellite imagery for assessing SOM content over the years; however, its repetitive temporal resolution and improved spatial and spectral resolutions have made up for this limitation (Dou *et al*., 2019). The majority of these studies usually involve mapping or predicting SOM content. For example, Schreier *et al*., (1988) examined rates of change and spatial distribution of SOM in individual agricultural fields using aerial photographs and spectral measurements. Liu *et al*., (2011) used Landsat TM imagery to build an exponential model to predict SOM over large areas. Mirazee *et al*., (2016) used Landsat ETM+ imagery together with geostatistical models to improve and map the prediction of SOM. Also, Dou *et al*., (2019) created and evaluated soil spectral indices in order to model and predict SOM using MODIS images from three different years. These studies reveal that satellite remote sensing can be an important tool for acquiring data on SOM and assisting in its quantification, prediction, modelling, and mapping of its spatial and temporal variability.

Furthermore, such studies have paved the way in which SOM can be assessed using satellite remote sensing. More specifically, the non-destructive nature of this technique allows repetitive measurements to be taken simultaneously, which gives it a significant advantage over traditional laboratory methods (Angelopoulou *et al*., 2019). According to Mutanga *et al*., (2016), remote sensing has progressed over the years due to the "technological advancements that have led to the availability of efficient, relatively cheap, robust, as well as high resolution images". These images can improve the ease at which SOM information is gathered, especially over large areas. It also acquires information in those areas that may not be easily accessible. In using such techniques to assess SOM content, results can be achieved at a much faster rate, thus allowing for accurate information to be achieved over shorter periods of time. This can be quite useful to the agricultural sector as it provides an insight into the status of fertility and productivity of the soil. Subsequently, it can also be used to create models that can predict and

quantify SOM content, which can be mapped in order to examine the spatial variability of SOM content within an area. Furthermore, this allows for a significant understanding of SOM dynamics which can be achieved using such techniques, and therefore provide valuable information regarding management practices that can either maintain or increase SOM content.

Human activities coupled with population growth has led to the degradation of fertile topsoil and hence, SOM. The rate at which SOM is being degraded exceeds that at which it can be produced, especially over larger scales. For instance, the soils within South Africa are usually not characterized by high fertility and become easily degraded with the exception of certain soils. According to Du Preez *et al*., (2011), soils within South Africa are characterised by low organic matter levels. This degradation in SOM over the years is attributed to the improper use and management of soil. However, the situation is more likely to worsen in the near future if no measures are put into place. Therefore, the use of efficient satellite remote sensing data can assess SOM content to ensure that it is able to support and facilitate different uses and activities for all its users on earth. As a result, the current study will aim to assess the relationship between SOM content and satellite remote sensing data within the soil of the Emakhosaneni area located in KwaZulu-Natal, South Africa.

4.2 Materials and methods

4.2.1 Description of the study area

South Africa comprises of nine provinces. The selected province in which the study will be conducted is the KZN province located along the east coast of South Africa and occupies an area of 94 361 km^2, which is 7.7% of South Africa's land area and has a population of 10 267 300 people (SSA, 2011). The KZN province consists of 12 district municipalities that can further be subdivided into 55 local municipal areas. The study was conducted in the Emakhosaneni area that is located within the Okhahlamba Municipality, KZN.

The Okhahlamba Municipality (Figure 4.1) is located in the western region of KZN and is approximately 242 km from the city centre of the eThekwini Metropolitan Municipality. It is situated in the mountainous regions of KZN between Lesotho, Free State, Emnambithi and Mtshezi (SSA, 2011). The total population of individuals present within the municipal area is 132 068 and are distributed over an area of 3,970.98 km^2 with an average density of 33.26 individuals per km^2 (Frith, 2011). The local municipality gives rise to many towns and settlement areas, but Emakhosaneni was selected as the study area for this particular study.

The Emakhosaneni area is located in the central region of the Okhahlamba municipality (28°46'45.1"S; 29°14'42.4"E) (Figure 4.1). The area has a total population of 1 938 individuals that are distributed over an area of 13.72 km^2 and has an average density of 141.30 individuals per km^2 (SSA, 2011). Despite being a large area, there are only 380 households found within the area that are sparsely dispersed. All households within the area are traditional/tribal. The area also mainly comprises of Black African individuals (99.9%) (SSA, 2011). It is situated approximately 17.5 km from the nearest town, Bergville.

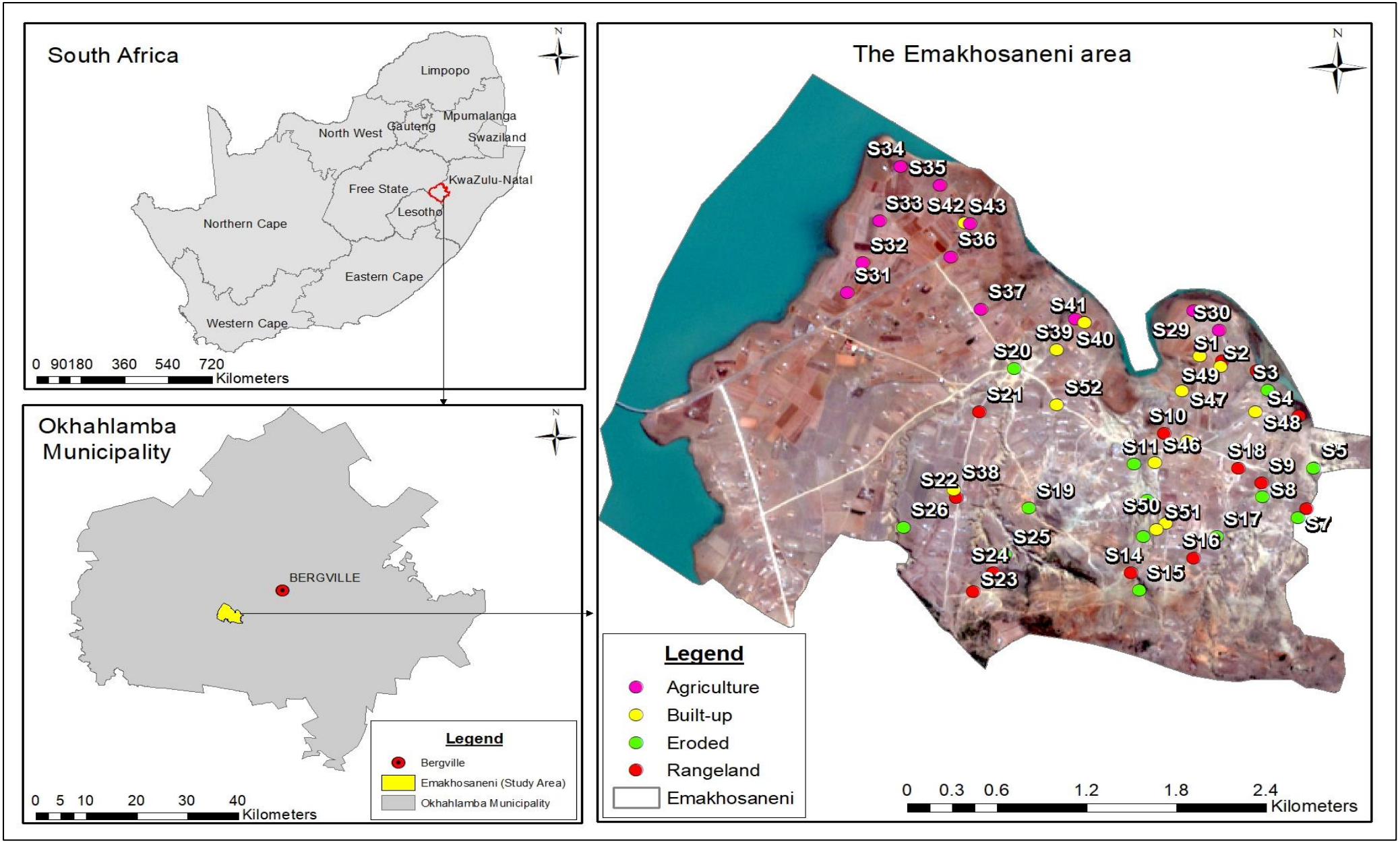


Figure 4.1: The selected study area and sampling points from which soil samples were obtained within the four major land uses (Author, 2020)

The study area experiences a warm and temperate climate. The monthly average midday temperatures range from 19°C in the winter months like July to 29°C in the summer months like February (Weather Atlas, 2020). On the other hand, the area receives an average of 644 mm of rain per year, with the highest rainfall experienced during the month of January (120 mm) and the lowest rainfall is during the month of July (5 mm) (Weather Atlas, 2020). The wind speeds vary over the months, with the winter and spring months experiencing the highest wind speeds, while the summer and autumn months experience the lowest wind speeds.

Furthermore, the area is located in a mountainous region. As one moves down the mountainous landscape, the gradient begins to become less steep. Along the base of the mountain, the gradient begins to decrease gradually until a gentle to no gradient is seen. This extends for several kilometres. Also, an increase in the number of households is seen as one moves away from the mountainous regions of the area.

In addition, the area is surrounded by various water sources. The Woodstock dam is located north of the study area, while to the eastern regions, the Tugela River and Drieldam are found (Google Earth, 2020). These water sources can be easily accessed and therefore supports their surrounding land use types. Two types of geology constitute the study are; Adelaide and Tarkastad (SANSA, 2015). The area is dominated by agricultural practices on both a large and small scale. Large areas of land are cleared from their natural vegetation in order to make land available for the expansion of the settlement and for agricultural practices. With such extensive farming occurring in the area, the appearance of land fragmentation is seen. There are a few sparse trees that remain after the land was cleared from its natural vegetation around the area, but the highest densities are seen in those parts that are inhabited by community members. Lastly, there are a number of road networks found throughout the area that aid access to various parts into and out of the area.

A more detailed description of the study area can be found in chapter two.

4.2.2 Soil sample collection

A thorough examination of the study area was conducted before sample collection. This was done using satellite imagery via Google Earth (Google Earth, 2020) and an aerial photograph of the study area followed by field reconnaissance. This led to the identification of four major land uses/land cover that constitute the study area; Agriculture, Built-up (Residential), Eroded, and Rangeland.

Soil samples were collected from the 16th July to 19th July 2019, from 08:00 am to 15:00 pm. A stratified random sampling method was used to determine the sampling points within the study area (Figure 4.1). It is a commonly used method that is designed to ensure that the samples taken are representative of the area and eliminates any biasedness (Worsham *et al*., 2012). Thirteen random soil samples were obtained from each of the four major land uses (Agriculture, Built-up (Residential), Eroded, and Rangeland), which amounts to a total of 52 soil samples being obtained from the study area.

Soil samples were collected from the soil's surface to a depth of 15 cm using a soil auger. The collected sample was then placed into a labelled, airtight plastic bag to prevent any spillage or drying of the sample, and the GPS coordinates were recorded. Subsequently, any other observations regarding the sampling point were recorded. However, at certain sampling points, the soil auger could not penetrate the soil's surface, and instead, a spade was used to obtain the soil sample. The above-mentioned procedure was replicated for the collection of soil samples from the 52 sampling points. Due to the study area being quite far from the laboratory, the soil samples were stored in a cool and dry cooler box, away from direct exposure to sunlight. Once all 52 soil samples were obtained, they were transported to and stored in the laboratory, where the analysis would begin.

4.2.3 Soil sample preparation and analysis

Approximately 20 g of each soil sample was oven dried overnight at a temperature of 105℃. Once oven dried, the samples were crushed into fine soil particles using a porcelain mortar and pestle. The crushed samples were then passed through a 425 µm stainless steel sieve. The sieved samples were then transferred into their respective labelled glass beakers and covered with Parafilm M until analysis.

Various methods are used to measure and estimate SOM content in soils (Motsara and Roy, 2008). However, this study used one of the most common methods known as; the WB method to determine the percentage of SOM content in the soil samples. It is a wet oxidation method that involves the use of acids to determine the amount of oxidizable carbon in soil, and this amount is then converted to SOM using a conversion factor (Walkley and Black, 1934). For this study, a small amount of the sieved soil sample was placed in a conical flask together with potassium dichromate and sulphuric acid and was left to stand for 30 minutes. Thereafter, distilled water, orthophosphoric acid, and the indicator were added to the conical flask, which changed the colour of the solution to a dark black. The solution was then titrated against ferrous ammonium sulphate in small increments until the solution changed colour from black to green, thus indicating the endpoint of the titration. Finally, the total volume of ferrous ammonium sulphate used in the titration was recorded, and the percentage of SOM content was calculated. This procedure was repeated for all soil samples.

A detailed description explaining the manner in which the SOM content was analysed using laboratory analysis and calculated can be found in chapter two.

4.2.4 Satellite imagery

For this particular study, the Sentinel-2 satellite was selected for acquiring satellite imagery data. The Sentinel-2 is a mission of the European Union (EU) Copernicus program which comprises of a constellation of two satellites (Sentinel-2A and Sentinel-2B) that are equipped with the Multi-Spectral Instrument (MSI); which is responsible for collecting optical imagery over land and coastal waters (European Space Agency [ESA], 2020a). Sentinel-2 is able to support the monitoring of changes that occur along the earth's surface by offering exceptional scenes. This is due to its wide swath width (290 km), spatial resolution (10–60 metre [m]), and high revisit time (ten days for each satellite or five days when two satellites are operating simultaneously) (Gholizadeh *et al*., 2018; ESA, 2020a). Subsequently, Sentinel-2 is a multispectral instrument that consists of thirteen bands with varying spatial resolutions that

range from the VIS to the SWIR region. More information regarding these bands is depicted in Table 4.1.

Table 4.1: Spectral bands and resolutions of the Sentinel-2 MSI sensor (Source: Zheng *et al*., 2018)

Band number	**Spectral band**	**Centre wavelength (nm)**	**Band width (nm)**	**Spatial resolution (m)**
B1	Coastal aerosol	443	20	60
B2	Blue	490	65	10
B3	Green	560	35	10
B4	Red	665	30	10
B5	Red-edge 1	705	15	20
B6	Red-edge 2	740	15	20
B7	Red-edge 3	783	20	20
B8	Near infrared (NIR)	842	115	10
B8a	Near infrared narrow (NIRn)	865	20	20
B9	Water vapour	945	20	60
B10	Shortwave infrared/Cirrus	1380	30	60
B11	Shortwave infrared 1 (SWIR1)	1910	90	20
B12	Shortwave infrared 2 (SWIR2)	2190	180	20

According to Castaldi *et al*., (2019a), some of the features and strengths of the Sentinel-2 include: "(i) their availability: everybody can freely explore and download the images, (ii) the short revisit time: five days under the same viewing angle, (iii) the fine spatial resolution: 10 m in the VIS and 20 m in the NIR and SWIR region, (iv) the presence of 13 bands including two bands acquiring data in the SWIR region, and (v) high quality physically-calibrated data". These features of the Sentinel-2 are therefore promising for soil applications. More detailed information regarding the Sentinel-2 mission can be found on the ESA website (https://sentinel.esa.int/web/sentinel/home).

The satellite imagery used for this study was obtained from the Copernicus open access hub (https://scihub.copernicus.eu/dhus/#/home), which is a scientific data hub for the Sentinel-2. Prior to downloading the image used for the study, the data hub was firstly searched for images from the 16th July to 19Th July 2019 within the study area. The search revealed that four images were available for download; two images collected on the 16th of July (Sentinel-2B, Level 1C and 2A) and two images collected on the 18th of July (Sentinel-2A, Level 1C and 2A). The different levels of images depict the different product types of the Sentinel-2, where Level 1C

represents images taken from Top Of Atmosphere (TOA) and Level 2A represents images taken from Bottom Of Atmosphere (BOA) (ESA, 2020b). However, for this study Level 2A images were of interest as they are already converted to reflectance data, and Level 1C images were omitted for use. Furthermore, the images collected on the 18th of July had a higher cloud cover percentage compared to the cloud-free imaged collected on the 16th of July. Therefore, for this study, one cloud-free image acquired on the 16th of July 2019 from the Sentinel-2B satellite with Level 2A product type was downloaded (Appendix K).

The downloaded image was a Level 2A product that is processed by the Sen2Cor processor. This processor is a plugin that is incorporated into the Sentinel Application Platform (SNAP) software (Castaldi *et al*., 2019a). More specifically, this product type has geometric, radiometric, and atmospheric corrections already applied to it, which are made in pre-processing by the data producer (Žížala *et al*., 2019). The entire procedure is outlined in the Sentinel-2 user handbook (ESA, 2015). Therefore, no pre-processing was required as the downloaded image was readily available for use by the data producer.

The selected image was loaded onto the SNAP software version 7.0.0, where it was resampled, and subset/clipped to the selected study area. In terms of resampling, the resampling tool in SNAP was used to resample all bands to a common resolution of 10 m by means of a nearest neighbourhood algorithm. This method is computationally efficient and is known to preserve the input pixel values of the image (Roy *et al*., 2016; Gholizadeh *et al*., 2018). Once resampled, the image was subset to the extent of the study area in order to place focus on the desired area and reduce processing time during analysis. This was done using the subset tool within SNAP; thereafter, the tool was also used to select the required bands for analysis. From the thirteen bands covered by the sensor, only nine were selected for use. These bands include; B2, B3, B4, B5, B6, B7, B8, B11, and B12 and are commonly used in the assessment of soil and its properties (Gholizadeh *et al*., 2018; Castaldi *et al*., 2019b). After the required bands were selected, a Land/Sea mask was used to mask out those pixels or areas outside the desired boundary of the study area. The processed image was then used for further analysis.

4.2.5 Band values and indices retrieval at sampling points

The reflectance values of each band at the 52 sampling points were extracted in the SNAP software using the pin manager, which allowed to filter out the reflectance values at the specific sampling points. These values were exported into a spreadsheet for analysis.

According to Gholizadeh *et al*., (2018), in remote sensing, the optical properties of soil are largely influenced by four major factors; mineral composition, soil moisture, organic matter content, and soil texture. Therefore, different groups of spectral indices such as colour indices, vegetation indices, and soil indices were calculated (Table 4.2). These indices are assumed to improve the prediction capability of remote sensing data by indirectly retrieving variables that are correlated with a target variable. A total of 16 spectral indices were calculated using SNAP, as tools for the indices are easily available in the software. The indices include; Brightness Index (BI), the Second Brightness Index (BI2), Redness Index (RI), Colour Index (CI), Hue Index (HI), Saturation Index (SI), Normalized Differences Vegetation Index (NDVI), Transformed Vegetation Index (TVI), Enhanced Vegetation Index (EVI), Green Normalized Difference Vegetation Index (GNDVI), Vegetation (V), Renormalized Difference Vegetation Index (RDVI), Difference Vegetation Index (DVI), Soil Adjusted Vegetation Index (SAVI), Modified Soil Adjusted Vegetation Index (MSAVI), and the Second Modified Soil Adjusted Vegetation Index (MSAVI2). The formulas that were used to derive these indices are displayed in Table 4.2, along with the Sentinel-2 bands used. Detailed descriptions of some of these indices can be found in Xue and Su (2017) and Silleos *et al*., (2006). After the indices were calculated for the entire study area, the calculated values were extracted and exported to a spreadsheet in the same manner as mentioned previously.

Table 4.2: Spectral bands and formulae used to calculate the spectral indices (Source: Gholizadeh *et al*., 2018)

Index	Definition	Definition based on Sentinel-2	Details	Reference
BI	$\frac{\sqrt{(\rho\text{Red} * \rho\text{Red}) + (\rho\text{Green} * \rho\text{Green})}}{2}$	$\frac{\sqrt{(\text{B4} * \text{B4}) + (\text{B3} * \text{B3})}}{2}$		Escadafal (1989)
BI2	$\frac{\sqrt{(\rho\text{Red} * \rho\text{Red}) + (\rho\text{Green} * \rho\text{Green}) + (\rho\text{NIR} * \rho\text{NIR})}}{3}$	$\frac{\sqrt{(\text{B4} * \text{B4}) + (\text{B3} * \text{B3}) + (\text{B8} * \text{B8})}}{3}$		Escadafal (1989)
RI	$\frac{\rho\text{Red} * \rho\text{Red}}{\rho\text{Green} * \rho\text{Green} * \rho\text{Green}}$	$\frac{\text{B4} * \text{B4}}{\text{B3} * \text{B3} * \text{B3}}$		Pouget *et al.*, (1990)
CI	$\frac{\rho\text{Red} * \rho\text{Green}}{\rho\text{Red} * \rho\text{Green}}$	$\frac{\text{B4} - \text{B3}}{\text{B4} + \text{B3}}$		Pouget *et al.*, (1990)
HI	$\frac{2 * \rho\text{Red} - \rho\text{Green} - \rho\text{Blue}}{\rho\text{Green} - \rho\text{Blue}}$	$\frac{2 * \text{B4} - \text{B3} - \text{B2}}{\text{B3} - \text{B2}}$		Mandal (2016)
SI	$\frac{\rho\text{Red} - \rho\text{Blue}}{\rho\text{Red} + \rho\text{Blue}}$	$\frac{\text{B4} - \text{B2}}{\text{B4} + \text{B2}}$		Mandal (2016)
NDVI	$\frac{\rho\text{NIR} - \rho\text{Red}}{\rho\text{NIR} + \rho\text{Red}}$	$\frac{\text{B8} - \text{B4}}{\text{B8} + \text{B4}}$		Rouse *et al.*, (1974)

TVI	$(\frac{\rho NIR - \rho Red}{\rho NIR + \rho Red})^{0.5} * 100$	$(\frac{B8 - B4}{B8 + B4})^{0.5} * 100$		Nellis and Briggs (1992)
EVI	$G\frac{\rho NIR - \rho Red}{\rho NIR + C1 * \rho Red - C2 * \rho Blue + L}$	$G\frac{B8 - B4}{B8 + C1 * B4 - C2 * B2 + L}$	G=2.5, C1=6, C2=7.5, L=1	Huete *et al*., (2002)
GNDVI	$\frac{\rho NIR - \rho Green}{\rho NIR + \rho Green}$	$\frac{B8 - B3}{B8 + B3}$		Gitelson *et al*., (1996)
V	$\frac{\rho NIR}{\rho Red}$	$\frac{B8}{B4}$		Jordan (1969)
RDVI	$\frac{\rho NIR - \rho Red}{\sqrt{\rho NIR + \rho Red}}$	$\frac{B8 - B4}{\sqrt{B8 + B4}}$		Roujean and Breon (1995)
DVI	$\rho NIR - \rho Red$	$B8 - B4$		Tucker (1979)
SAVI	$\frac{(\rho NIR - \rho Red) * (1 + L)}{\rho NIR - \rho Red + L}$	$\frac{(B8 - B4) * (1 + L)}{B8 - B4 + L}$	L=0.5	Huete (1988)
MSAVI	$\frac{(\rho NIR - \rho Red) * (1 + L)}{\rho NIR + \rho Red + L}$	$\frac{(B8 - B4) * (1 + L)}{B8 + B4 + L}$		Qi *et al*., (1994a)
MSAVI2	$\frac{2 * \rho NIR + 1 - \sqrt{(2 * \rho NIR + 1)^2 - 8 * (\rho NIR - \rho Red)}}{2}$	$\frac{2 * B8 + 1 - \sqrt{(2 * B8 + 1)^2 - 8 * (B8 - B4)}}{2}$	L=1–2*s* NDVI* WDVI	Qi *et al*., (1994b)

4.2.6 Principal component analysis

Principal component analysis (PCA) is a widely used statistical technique that has the ability to explore trends in multiple variables (Guo *et al*., 2018). More specifically, it can transform variables that are correlated into a number of independent variables known as principal components (PC) that are uncorrelated. These PCs are derived from linear combinations of the original variables. More details on PCA can be found by Johnson and Wichern (2002) and Gniazdowski (2017). A PCA was carried out in this study to avoid any multicollinearity between variables, reduce the dimension of the dataset, and decrease the number of input variables used for analysis (Parchami-Araghi *et al*., 2013; Mirazee *et al*., 2016). Furthermore, it was used as a preliminary step in the development of the prediction model. The use of PCs can greatly influence the prediction of soil properties (Gobin 2000; Hengl *et al*., 2007). PCA was performed using factor analysis in SPSS version 26. The calculated spectral indices were used as input variables in the PCA. In order to identify the dominant PCs, Kaiser's criterion or the eigenvalues rule (i.e., components that have eigenvalues equal to or greater than 1.0 are retained and used for further investigation) was applied. Once the dominant PCs were identified and retained, they were rotated using a varimax rotation, which assists in the interpretability of components. According to Askari *et al*., (2020), this rotation transforms the loadings of the PCs in order to maximize the correlation between PCs and variables. Thereafter, the factor loadings with a value less than 0.60 were suppressed, as stated by Tabachnick and Fidell (2014), which recommend using a value of 0.32. Once the relevant PCs were identified and extracted, their regression scores were saved and used for later analysis. The works of Beaumont (2012), Wuttichaikitcharoen and Babel (2014), and Samuels (2016) assisted in carrying out the PCA for this study. The PCA was only carried out on the derived spectral indices and not on the spectral bands, as multicollinearity was moderate and well within the threshold (i.e., <10).

4.2.7 Development of SOM prediction models

Prior to the development of models, the 52 sampling sites were divided into two groups. The first group being the training samples were used for creating the models (n=41, 80% of sampling sites), and the second group was used in the validation of the models (n=11, 20% of sampling sites).

4.2.7.1 Multiple linear regression models

A multiple linear regression (MLR) based on a multivariate linear regression analysis was adopted to create the models for SOM prediction. This regression method assumes that a linear

relationship exists between the independent and dependent variables (Liu *et al*., 2015). More specifically, it focuses on estimating the combined effect of predictor variables on a particular response variable (Kumar *et al*., 2018). In this study, the predictor variables were the PCs derived by transforming the sixteen spectral indices and the spectral bands, while the response variable was SOM. The MLR analysis was also carried out in SPSS using the training dataset. Two separate MLR models were created between the spectral bands and SOM and the spectral indices (transformed into PCs) and SOM. With regards to the first model, all nine spectral bands were used as independent variables while SOM was the dependent variable. Important descriptive statistics and the derived model were noted. Furthermore, the prediction values using the derived model and its residuals (i.e., residuals are estimated from the difference between predicted and observed SOM values) were saved for further analysis. With regards to the second model, the extracted PCA scores were used as independent variables while SOM was the dependent variable. As mentioned previously, predicted values and residuals from the derived model were saved for further analysis. For the development of both MLR models, the enter method in SPSS was used with the following MLR equation (equation 1) (Mallick *et al*., 2020):

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \ldots\ldots\ldots\ldots\ldots + b_pX_p$$

Where, x is the dependent variable, $X_1, X_2, X_3 \ldots\ldots\ldots X_p$ are the independent variables, and b_0 is the value of Y when all of the independent variables (X_1 through X_p) are equal to zero (i.e., the constant), and b_1 through b_p are the estimated regression coefficients. Each regression coefficient represents the change in Y relative to a one-unit change in the respective independent variable.

4.2.7.2 Geostatistical methods

Geostatistics is considered as a subset of statistics that specializes in the interpretation of geographically referenced data (Goovaerts, 1997). According to Hengl (2009), it is considered as the "branch of statistics that specializes in the analysis and interpretation of any spatially (and temporally) referenced data, but with a focus on inherently continuous features (spatial fields)". It focuses on the analysis of spatial data and models that explain how spatial patterns originate and develop. For this particular study, three different geostatistical methods were assessed in the prediction of SOM: Ordinary Kriging (OK), Simple Kriging (SK) and Regression Kriging (RK). In terms of OK, a weighted average with known neighbouring values that are dependent on the distance between them, their grouping, and their values are used to

estimate an unsampled location. It is similar to the interpolation method, Inverse Distance Weighting (i.e., IDW) and also intends to estimate the value of a random function *z* at one or more unsampled locations (Durdevic *et al*., 2019). OK has the ability to provide unbiased estimates that can have minimum errors. For this study, OK was carried out on the measured SOM using the training set. The general equation (equation 2) of the OK method is expressed as (Gia Pham *et al*., 2019):

$$Z(X_0) = \sum_{i=1}^{N} \lambda_i Z(X_i)$$

Where, $Z(X_0)$ is the predicted value at the unmeasured position X_0, $Z(X_i)$ is the measured value at position X_i, λ_i is the weighting coefficient from the measured position to X_0, and N is the number of positions within the neighbourhood searching.

With regards to SK, which operates in a similar manner to OK with the only difference being that it incorporates a mean value in the estimation of values at unknown locations (Webster and Oliver, 2001). More specifically, the mean is assumed to be constant throughout the region that is being sampled due to the assumption of stationarity of the SK method. The SK method is done such that at a prediction point, the predictor uses the information at each sample and the knowledge of the mean together with the covariance is used to predict at unsampled locations (Wackernagel, 2003). For the study, SK was carried out on the measured SOM using the training set. The general equation of the SK method is expressed as (equation 3) (Mirazee *et al*., 2016):

$$Z(X_0) = \sum_{i=1}^{N} \lambda_i \left(Z(X_i) - m(\mu_i)\right)$$

Where, $Z(X_0)$ is the predicted value at the unmeasured position X_0, $Z(X_i)$ is the measured value at position X_i, λ_i is the weighting coefficient from the measured position to X_0, $m(\mu_i)$ is the known stationary mean of $Z(X_i)$, and N is the number of positions within the neighbourhood searching.

RK, on the other hand, is considered a hybrid geostatistical method. Simply put, this type of method combines a non-spatial method, such as regression, with a spatial interpolation. It is based on the notion that the deterministic component of the target variable is supported by a regression model, while the residuals are expected to describe the dependent component that varies spatially (Zhang *et al*., 2012). Moreover, RK includes the (1) identification of a trend

using a MLR between the response variable (SOM) and auxiliary variables (predictors: spectral bands and PCs derived from spectral indices) and (2) attaining a local mean using regression methods and estimating the associated residuals using a variogram and kriging and then summing them up together (Kumar *et al*., 2018). According to Hengl *et al*., (2004), RK can be expressed mathematically, as seen below (equation 4):

$$Z(X_0) = \hat{m}(X_0) + \hat{e}(X_0) = \sum_{k=0}^{p} \widehat{\beta_k}.\, q_k(X_0) + \sum_{i=1}^{N} \lambda_i.\, e(X_i)$$

Where, $\hat{m}(X_0)$ is the fitted deterministic part, $\hat{e}(X_0)$ is the interpolated residual, $\hat{\beta}_k$ are the estimated deterministic model coefficients, $q_k(X_0)$ is the *k*th external explanatory variable or predictor at location X_0, p is the number of predictors, λ_i are the kriging weights determined by the spatial dependence structure of the residual, and $e(X_i)$ is the residual at position X_i.

RK separates the prediction into two parts. Therefore, the equation on the right-hand side is first represented by the regression, and the second part is represented by the kriging of the residual.

For this study, both kriging methods, OK and SK, were performed on the residuals obtained from the MLR model of the spectral bands using the training dataset. Once the residuals were interpolated, they were added together/summed with the predicted values from the MLR model. This had resulted in Regression Ordinary Kriging (ROK) and Regression Simple Kriging (RSK). The above procedure was also repeated for the spectral indices model.

Furthermore, prior to carrying out the above geostatistical methods, a trend analysis was conducted on the variables that were going to be used in the interpolation process. The trend analysis revealed that there were no trends present in the SOM and residuals datasets across the study area. However, it was more or less constant over the study area; therefore, there was no need to remove any trends prior to analysis as it was relatively minimal for both datasets across the study area. Additionally, a histogram and normal QQ plot were used on both the SOM and residuals datasets to determine their normality. This revealed that both datasets followed a normal distribution. These preliminary steps were carried out on the training set of each dataset to ensure that they were useful for analysis and would obtain results that have minimum errors.

Once preliminary steps were carried out, the two datasets were fitted with a semivariogram model. Different models such as spherical, exponential, and gaussian were optimized and used

in an experimental semivariogram fitting. Semivariogram modelling plays a critical role in providing information on the spatial autocorrelation of datasets. According to Zhang *et al*., (2012), "the semivariogram model with the smallest residual sum of squares was selected as the best fitting model" for the datasets. For this particular study, a Spherical semivariogram model was selected for both datasets as it fulfilled the above-mentioned requirement. The spherical model can be expressed as (equation 5) (Guo *et al*., 2018):

$$\gamma(h) = \begin{Bmatrix} 0 & h = 0 \\ C_0 + C\left(\frac{3h}{2h} - \frac{h^3}{a^3}\right) & 0 < h \leq a \\ C_0 & h > a \end{Bmatrix}$$

Where, $\gamma(h)$ is a variogram, a is the range of the soil samples, h is the spatial lag, C_0 is the nugget, and $C_0 + C$ is the partial sill. The semivariogram was then used in the prediction process.

The software used to carry out all of the above-mentioned geostatistical methods was ArcGIS version 10.2. More specifically, the geostatistical analyst extension in this software assisted in carrying out the trend analysis, examining the normality of the datasets, semivariogram modelling, and the prediction (OK and SK). The raster calculator tool in ArcGIS was used to perform the ROK and RSK on the spectral bands and spectral indices by adding the prediction and the residuals together. Lastly, the software was used in creating the final prediction maps. In order to predict the SOM of the study area, the generic framework by Hengl *et al*., (2004) was followed.

4.2.8 Validation of prediction models

To validate the accuracy of the prediction models, the validation dataset was used for the validation analysis. Firstly, in order to test the accuracy and strength of the MLR models, the coefficient of determination (R^2) of the SOM predictions was calculated using the following (equation 6):

$$R^2 = \left(\frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum {x_i}^2 - (\sum x_i)^2}\sqrt{n\sum {y_i}^2 - (\sum y_i)^2}}\right)^2$$

Secondly, in order to test the accuracy of the prediction methods (OK, SK, ROK, and RSK) to predict SOM, the Mean Error (ME) (equation 7) and Root Mean Square Error (RMSE) (equation 8) were calculated in the following manner:

$$ME = \frac{1}{N}\sum_{i=1}^{N}(Z(X_i) - Z^*(X_i))$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Z(X_i) - Z^*(X_i))^2}$$

Where, X_i is the estimated SOM at location *i*, y_i is the observed SOM at location *i*, *n* is the number of samples, $Z^*(X_i)$ is the predicted value, and $Z(X_i))^2$ is the measured value.

4.3 Results and Discussion

4.3.1 SOM analysis

The descriptive statistics of the SOM content within the study area and its major land uses are displayed in Table 4.3. In Table 4.3, it is evident that the agriculture land use has the highest average SOM content (3.42%). In contrast, eroded land contains the lowest average SOM content (2.19%). Thus, an average difference of 1.23% in SOM exists between these land uses. The average SOM content in land uses such as rangeland and built-up lie between the highest and lowest percentages of SOM. On the other hand, the standard deviation exhibits a different trend, in that the eroded land has the highest standard deviation (1.26) and built-up land use the lowest (0.98). With regards to the minimum and maximum percentage of SOM found in each of the different land uses, agriculture had SOM percentages that were higher compared to the others. The highest minimum was 2.23% (agriculture), and the lowest was 0.00% (eroded). At the same time, agriculture (5.96%) had the highest maximum percentage of SOM, while rangeland (4.62%) had the lowest. With regards to the overall average of SOM content in the study area, an amount of 2.79% was seen. The lowest percentage was 0.00%, and the highest was 5.96%.

Table 4.3: Descriptive statistics indicating the overall percentage of SOM within the study area and its major land uses (N= Number of samples, Std= Standard deviation)

	N	Mean	Std. deviation	Minimum	Maximum
Overall	52	2.79	1.19	0.00	5.96
Agriculture	13	3.42	1.03	2.23	5.96
Built-up (residential)	13	2.50	0.98	0.74	4.76
Eroded	13	2.19	1.26	0.00	4.91
Rangeland	13	3.05	1.18	1.34	4.62

From the above results obtained it can be seen that the average percentage of SOM within the Emakhosaneni area is relatively low when compared to the average amount that is usually found in soils. This average roughly ranges from 1-6% in the topsoil, whereas the subsoil contains less than 2% of SOM (Adams *et al*., 2011). Subsequently, certain soils, such as mineral soils, have a SOM concentration that ranges from 1-5% (Kalev and Toor, 2018). The concentrations of SOM are greatest in the topsoil layer (13-25 cm) and begins to decrease with depth. The soil samples that were analysed for this study were obtained from the very same layer, therefore indicating that the topsoil within the Emakhosaneni area contains, on average, approximately 2.79% of SOM (Table 4.3). This may be supported by the fact that soils within

South Africa are characterised by low organic matter levels as a result of the low SOC content found in topsoils that usually range from <0.5% to >4%. And a low supply of carbon in the soil can directly influence the SOM content, as SOC constitutes 58% of SOM (Donovan, 2013). The lower levels of SOC present within the soils of the study area may have contributed to the lower concentration of SOM during analysis. Additionally, different soil types in different regions or areas are highly influenced by their soil-forming factors; climate, organisms, relief (topography), parent material, and time (Jenny, 1994). For instance, South Africa is a semi-arid region that receives less than 500 mm of rainfall (SAWS, 2019). Under such conditions, the supply of moisture to the soil and its organisms is lowered (Osman, 2012b). This may have slowed down the microbial activity and decomposition process and, therefore, hindered SOM formation and accumulation within these soils. Likewise, sampling was done during the month of July, when rainfall in the area is at its lowest. This coupled with relatively warm temperatures may have resulted in lower SOM concentrations being present within the study area during sampling and, therefore, during analysis. A detailed discussion regarding the average SOM content found within each of the sampled land uses can be found in chapter two.

4.3.2 PCA of spectral indices

Table 4.4 indicates the PCs extracted from the sixteen input variables (i.e., spectral indices) used in the PCA and the percentage of variance explained by each PC. The first three extracted components accounted for 97.871% of the total data variation. It is evident that PC1 has the highest eigenvalue (9.143), followed by PC2 (4.282), and PC3, with the lowest eigenvalue (2.235). Similarly, the percentage of variance explained by each PC follows the same trend. PC1 accounted for the highest variance of 57.143%, PC2 accounted for 26.762% variance, and PC3 accounted for the lowest variance of 13.966%.

Table 4.4: Eigenvalues and percentages of variance associated with each PC

PCs	Eigenvalues	Variance (%)	Cumulative variance (%)
1	9.143	57.143	57.143
2	4.282	26.762	83.905
3	2.235	13.966	97.871

From these results, it can be seen that the sixteen spectral indices used as input variables in the PCA were reduced to only three PCs while using a varimax rotation. This may be attributed to the eigenvalue rule, where eigenvalues greater than or equal to one are retained and used for further investigation. However, components that had an eigenvalue less than one were omitted

for use, as they carried less variation (i.e., information) regarding the data. According to Wuttichaikitcharoen and Babel (2014), the order of significance of these variables is influenced by the magnitude of their eigenvalues. In other words, the highest eigenvalue is represented by the first PC, and that explains the most variance in the dataset followed by the second PC with a lower eigenvalue and so on until an eigenvalue of one or greater is seen. For instance, the PCA in this study retained only three PCs. Subsequently, these three PCs are responsible for explaining 97.871% total variance in the dataset. This percentage is relatively high, which means that these few PCs are capable enough to explain the variance in the dataset and, more specifically, the spectral indices used in the analysis (Lorenzo-Seva, 2013). Furthermore, it also implies that the PCs are able to retain over 90% of the original information used in the analysis.

Table 4.5 shows that the factor loadings have been placed in descending order to make interpretability easier, along with important or heavily loading factors printed in bold. With regards to PC1 and PC3, the majority of the loadings are positive, with only a few variables showing negativity, while in PC2, half of the loadings are positive, and the other half is negative. Additionally, in PC1, MSAVI has the highest factor loading value of 0.997, and SI has the lowest value of 0.013. It can also be seen that the first nine variables have heavy loadings on PC1. However, in PC2, CI has the highest factor loading value of 0.985, and V has the lowest value of 0.001. Subsequently, it can be seen that the five variables printed in bold have the strongest loadings on PC2. However, in PC3, BI2 has the highest factor loading value of -0.930, and SI has the lowest value of 0.022. Subsequently, it can be seen that the last two variables printed in bold have the strongest loadings on PC3.

Table 4.5: Factor loadings for the three PCs of the spectral indices data

Variables	PC1	PC2	PC3
MSAVI	**0.997**	-0.043	0.053
MSAVI2	**0.996**	-0.042	0.067
SAVI	**0.991**	-0.025	0.128
RDVI	**0.989**	-0.023	0.146
DVI	**0.986**	-0.080	-0.115
EVI	**0.969**	-0.195	0.124
TVI	**0.935**	0.029	0.349
NDVI	**0.934**	0.023	0.351
V	**0.921**	0.001	0.359
CI	-0.136	**0.985**	0.092
SI	0.013	**0.947**	0.022
RI	-0.038	**0.909**	0.358
HI	-0.269	**0.880**	0.150
GNDVI	0.599	**0.730**	0.325
BI2	-0.155	-0.329	**-0.930**
BI	-0.464	-0.350	**-0.810**

The results of the PCA component matrix with varimax rotation for the sixteen spectral indices are displayed in Table 4.5. The component matrix of the PCA aims to explain the correlations between each of the variables and estimated PCs (Salkind, 2010). The results indicated that the different PCs were influenced by the factor loadings of the different spectral indices. More specifically, it can be clearly seen that each PC is influenced by a certain group of spectral indices. For instance, it was evident that nine variables had factor loadings that were heavily loaded (factor loading ≥ 0.60) on the first PC, and these variables included the Soil (MSAVI, MSAVI2, and SAVI) and Vegetation indices (RDVI, DVI, EVI, TVI, NDVI, and V). The factor loadings of these variables were relatively high and close to a value of 1, therefore indicating a positively strong relationship between the variables and PC1 (Song *et al*., 2017). It also means that these variables are highly correlated to PC1, and therefore PC1 can be used to represent these variables. However, the other seven indices showed weaker negative relationships and were weakly correlated to PC1, with the exception of GNDVI, which had a higher factor loading but not as strong as the previously mentioned variables. Furthermore, it was also evident that five variables had factor loadings that were heavily loaded on the second PC, and these variables included the Colour (CI, SI, RI, and HI) and Vegetation indices (GNDVI). The factor loadings of these variables were also relatively high and therefore indicated a positively strong relationship between the variables and PC2. Correlations between

variables and PC2 were relatively strong and can mean strong associations exist between the two. However, within this PC (PC2), it was also evident that more variables (11 of the 16 indices) had shown very weak positive and negative relationships with PC2. Moreover, it was also evident that two variables had factor loadings that were heavily loaded on the third PC, and these variables included the Brightness indices (BI and BI2). However, in this particular PC (PC3), the two variables had factor loadings that were relatively high but showed a negatively strong relationship between the variables and PC3. Though it is a negative relationship, these variables have a great influence on PC3. However, within this PC (PC3), it was also evident that most variables (14 of the 16 indices) had shown positive relationships with PC2 but were still quite weak. Based on the factor loading results from the PCA, it can be clearly seen that each PC (PC1, PC2, and PC3) is represented by a specific group of spectral indices. For instance, PC1 is characterized by environmental variables due to the influence of vegetation and soil indices, which are also similar to the results of Mirazee *et al*., (2016), Song *et al*., (2017) and Guo *et al*., (2018). PC2 is characterized by the colouration of soil due to the influence of colour indices, and PC3 is characterized by bare soil based on the brightness indices.

4.3.3 Prediction of SOM using MLR models

Table 4.6 provides the statistics from the assessment of the MLR model created from the spectral bands. With regards to the ME values, it is evident that the training dataset had a higher value (0.71%) compared to the validation dataset (0.54%). However, the RMSE values were much higher when compared to the ME values. For instance, RMSE was 0.89% in the training dataset and 0.67% in the validation dataset; a difference of 0.22% exists between them. In contrast, the R^2 value for the training dataset was relatively lower (0.48) when compared to the validation dataset (0.60).

Table 4.6: Assessment of the MLR model for predicting SOM content by using spectral bands

	N	**R^2**	**ME (%)**	**RMSE (%)**
Training	41	0.48	0.71	0.89
Validation	11	0.60	0.54	0.67

Furthermore, the equation derived from the MLR model used to predict SOM using spectral bands is as follows (equation 9):

$$\text{SOM (\%)} = 6.284 - 6.691(\text{Band2}) + 9.122(\text{Band3}) - 18.199(\text{Band4}) + 82.835(\text{Band5}) - 92.111(\text{Band7}) + 21.850(\text{Band8}) + 36.754(\text{Band11}) - 48.614(\text{Band12})$$

The above results provide an overview of the assessment and accuracy of the MLR model derived from the use of the spectral bands in terms of predicting SOM content. Important statistics regarding the error and variance of the model were calculated. It was revealed that the validation dataset performed better than the training dataset due to the lower errors and higher variance calculated (Table 4.6). This may be a result of the number of samples used in each dataset. For instance, a smaller number of samples may either over or underestimate the value that is being calculated compared to a greater number of samples being used for the calculation. Furthermore, certain values within the dataset may have had a greater chance of influencing the calculated statistic when compared to other values in the dataset. With regards to the prediction error statistics calculated (ME and RMSE) of the MLR model, the ME was relatively low for the datasets. This model assessment technique measures the mean error between the observed and predicted values using the MLR model (Ait-Amir *et al*., 2015). In other words, it gives an estimation of how off the predicted value is from the measured one. The ME values that were calculated for the MLR were smaller than one and therefore indicated that the MLR did a fair job in predicting the SOM content using spectral bands. However, in terms of the RMSE, the error values were moderately higher than that of ME. RMSE measures the square root of the mean of the square of all of the error and is a good measure of accuracy (Neill and Hashemi, 2018). Once again, it focuses on how the predicted value deviates from the measured value using the MLR model. The RMSE value was also smaller than one and therefore indicated that the MLR did a fair job in predicting the SOM content using spectral bands. However, the results of the prediction errors for the training and validation dataset were fairly comparable, with little deviation existing between them. This can be inferred by the scatterplot shown in Figure 4.2a. According to Castillo *et al*., (2005), such a result may either indicate over-fitting or under-fitting of the MLR model. However, in this particular case, it may be under-fitting which was fairly minimal as the error for the training dataset is significantly larger than the error for the validation dataset. More importantly, the R^2 values indicated the amount of variance (of the dependent variable) that can be explained by the spectral bands (independent variables). The results revealed a fairly high R^2 value for the validation dataset (0.60).

Therefore, this meant that the MLR model accounted for 60% of the variability in SOM content using spectral bands and is a moderately accurate model.

In Figure 4.2, a graphical summary of the MLR model derived from spectral bands is represented. Figure 4.2a illustrates a scatterplot for both the training and validation datasets using the measured SOM against the predicted SOM from the MLR model. It is evident that the majority of the data points are in close proximities to their respective lines of best fit for both datasets. However, some data points are also further away from these lines, especially within the training dataset. This may result from variability that may exist within the measured SOM dataset due to soil samples being collected from different land use types. Also, the presence of extreme values may be the cause. It is also evident that the scatterplot depicts a linear trend between the measured and predicted SOM.

Figure 4.2b-d shows various diagnostics of residuals derived from the spectral bands MLR model. In Figure 4.2b, the histogram of the residuals along with the normal probability curve is displayed. It is evident that the residuals approximate a normal distribution which can be supported by the symmetric bell shape curve of the histogram. In Figure 4.2c, the normality probability plot is displayed using the residuals. The plot revealed that the majority of the residual points are aligned along the 45° line, with few points deviating from the line along the upper reaches. This also suggests that the residuals follow a normal distribution. In Figure 4.2d, the predicted SOM and residuals are plotted against each other in order to display the homogeneity of the residuals. It is evident that the data points are randomly distributed whilst still being in close proximity to the origin (i.e., zero). This random dispersion around zero indicates homogeneity exists within the residuals. According to Mirazee *et al*., (2016), such analysis of the residuals provides an insight into its distribution which is very important to spatial autocorrelation.

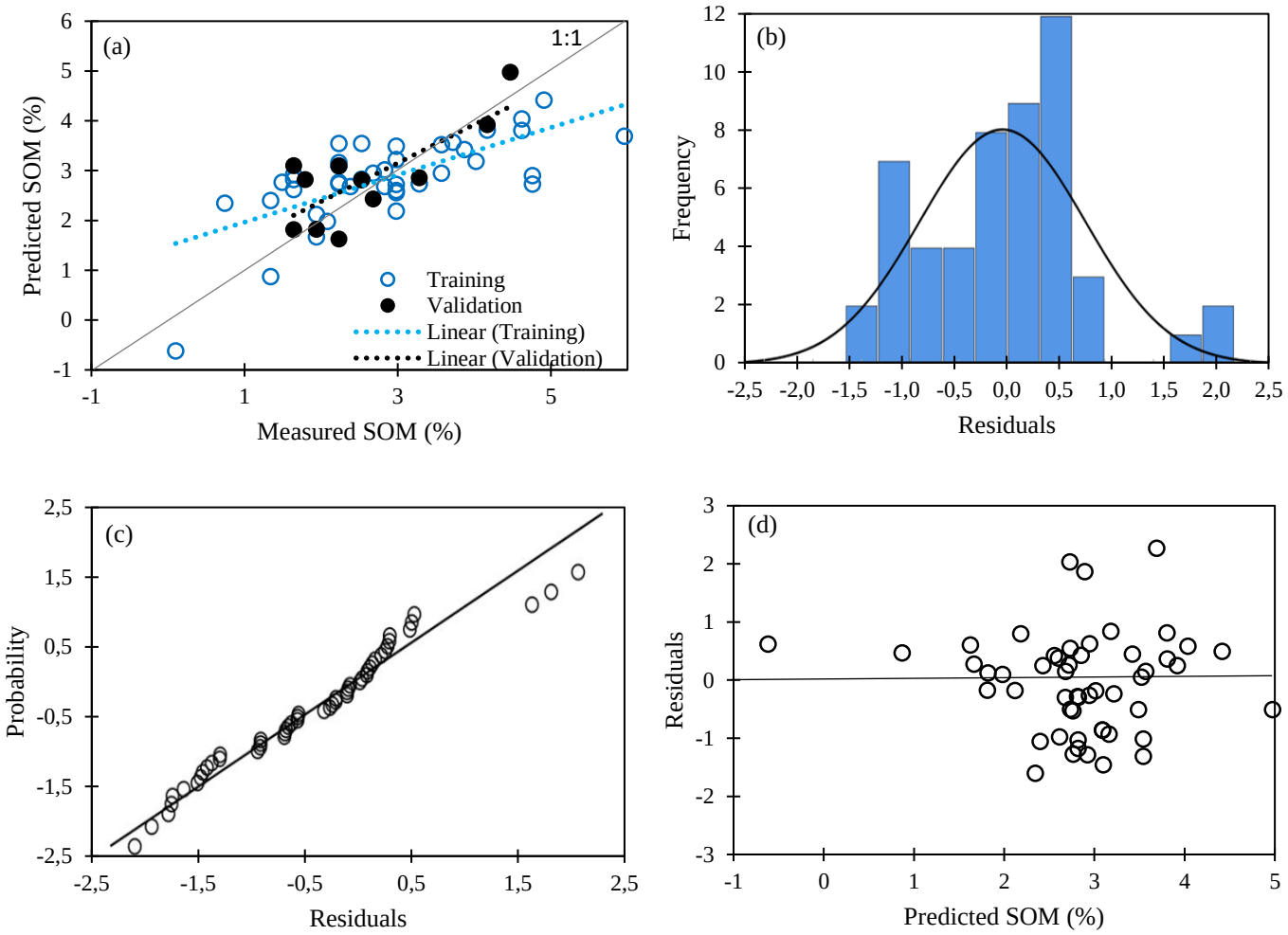


Figure 4.2: The results of the spectral bands MLR model showing the: (a) measured versus predicted SOM content, (b) histogram of the MLR residuals, (c) normal probability curve of the MLR residuals, and (d) residuals versus predicted SOM value

Table 4.7 provides the statistics from the assessment of the MLR model created from the PCs and SOM content. Regarding the ME values, it is evident that the training dataset had a higher value (0.91%) than the validation dataset (0.69%). However, the RMSE values were much higher when compared to the ME values for both datasets. In the instance of RMSE, where a value of 1.16% was seen for the training dataset and 0.81% for the validation dataset, a difference of 0.35% exists between them. In contrast, the R^2 value for the training dataset was quite low (0.10) when compared to the validation dataset, which also had a moderately low R^2 value of 0.34.

Table 4.7: Assessment of the MLR model for predicting SOM content by using PCs

	N	R^2	ME (%)	RMSE (%)
Training	41	0.10	0.91	1.16
Validation	11	0.34	0.69	0.81

Subsequently, the equation derived from the MLR model used to predict SOM using PCs is depicted below (equation 10):

$$SOM\ (\%) = 2.847 + 0.316(PC1) + 0.244(PC2) + 0.098(PC3)$$

The above results provide an overview of the assessment and accuracy of the MLR model derived from the use of the PCs and measured SOM content in terms of predicting SOM content. Similarly, important statistics regarding the prediction error and variance of the model were calculated. The MLR model using PCs also revealed that the validation dataset performed better than the training dataset due to the lower errors and higher variance calculated (Table 4.7). Once again, the number of samples used in each dataset and the values used within each of the datasets may have had a chance of influencing the calculated statistics compared to the other dataset. With regards to the prediction error statistics calculated (ME and RMSE) of the MLR model, the ME values were moderately high for both the datasets. However, these values were still smaller than one and therefore indicated that the MLR did a fairly substantial job in predicting the SOM content using the PCs. However, in terms of RMSE, the error values were significantly higher than that of ME. This can be supported by the work of Chai and Draxler (2014), which stated that values of RMSE are usually never smaller than that of the ME. Furthermore, the RMSE was also smaller than one for the validation dataset, therefore suggesting that the MLR model did a fair job in predicting the SOM content using the PCs derived from the spectral indices. In contrast, the RMSE was greater than one by not much for the training dataset. This can be attributed to the range of values in the PC scores dataset which ranged from -2 to 2. Influence from such values could have resulted in higher RMSE values for both the training and validation dataset. However, the results of the prediction errors for the training and validation dataset were not entirely comparable, as a moderate level of deviation exists between them. This can be inferred by the scatterplot shown in Figure 4.3a. Such a result may either indicate over-fitting or under-fitting of the MLR model (Castillo *et al*., 2005). However, in this particular case, it may be the case of under-fitting, which was fairly minimal as the error for the training dataset is significantly larger than the error for the validation dataset. More importantly, the R^2 values indicated the amount of variance (of the dependent variable)

that can be explained by the PCs derived from the spectral indices (independent variables). The results revealed a moderately low R^2 value for the validation dataset (0.34). Therefore, this meant that the MLR model accounted for 34% of the variability in SOM content using the three PCs and is a fairly accurate model.

Figure 4.3 depicts a graphical summary of the MLR model that was derived from the three PCs. In Figure 4.3a, a scatterplot for both the training and validation datasets by using the measured SOM against the predicted SOM from the MLR model is illustrated. From this scatterplot, it is evident that the data points are randomly distributed and are in close proximities to their respective lines of best fit for both the datasets. However, some data points are also further away from these lines for both datasets. As mentioned previously, the presence of variability may exist within the measured SOM dataset as soil samples were obtained from different land use types together with the presence of extreme values that may have contributed to such results. It is also evident that a linear trend exists between the measured and predicted SOM.

Figure 4.3b-d shows various diagnostics of the residuals derived from the PCs MLR model. In Figure 4.3b, a histogram of the residuals along with a normal probability curve is displayed. It is evident that the residuals approximate a normal distribution due to the symmetric bell shape curve of the histogram. In Figure 4.3c, a normality probability plot is displayed using the residuals. The plot revealed that the majority of the residual points are aligned along the 45° line, with only a few points deviating from the line along the upper reaches. This is also an indication that the residuals follow a normal distribution. In Figure 4.3d, the predicted SOM and residuals are plotted against each other in order to display the homogeneity of the residuals. The data points are randomly distributed along the origin (i.e., zero) whilst also being in close proximity to it. Such random dispersion around zero indicates the presence of homogeneity of the residuals. Moreover, essential analysis on the distribution of residuals is required as it is directly used in the calculation of experimental semivariograms for the spatial autocorrelation of the kriging interpolation (Song *et al*., 2017).

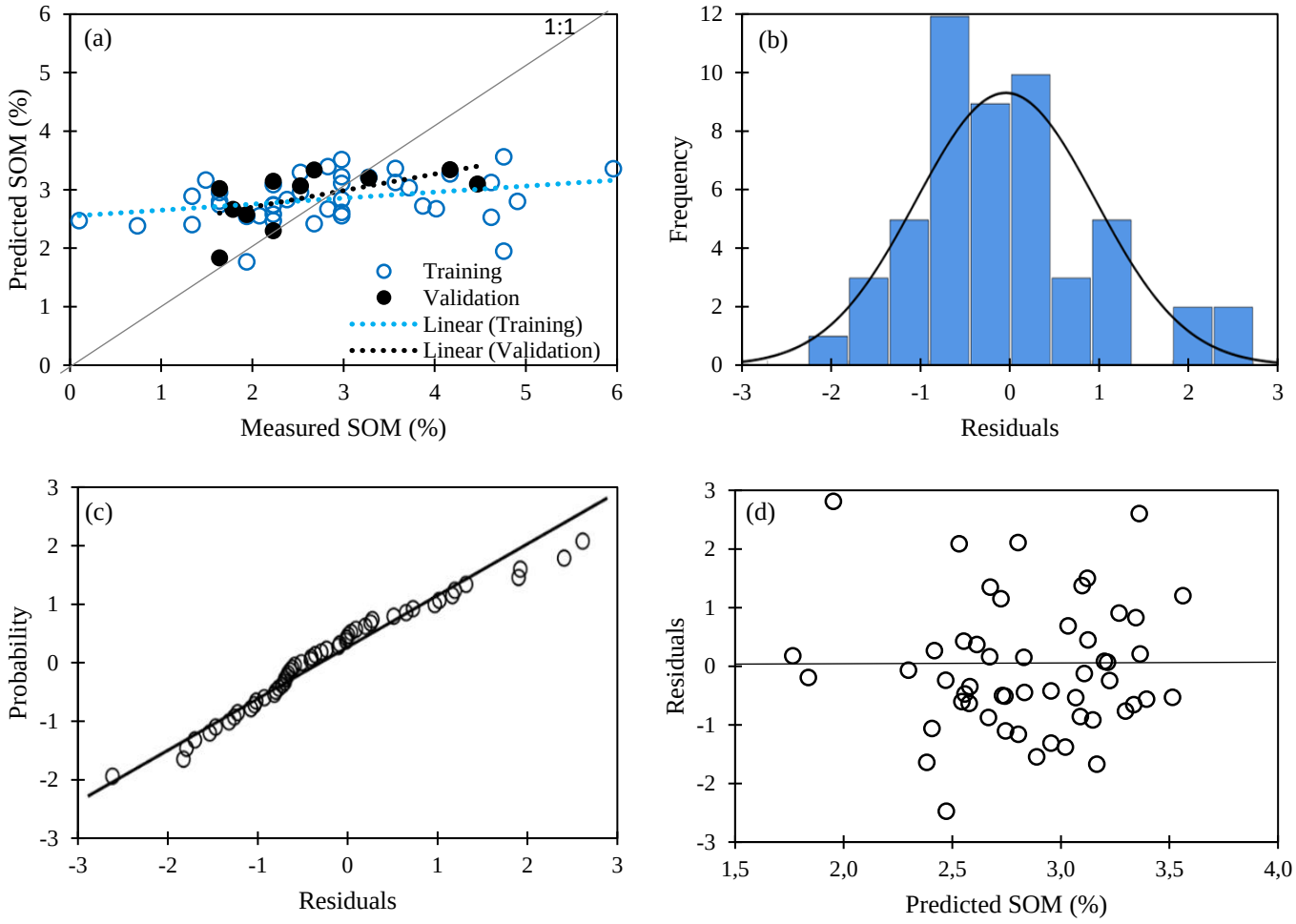


Figure 4.3: The results of the PCs MLR model showing the: (a) measured versus predicted SOM content, (b) histogram of the MLR residuals, (c) normal probability curve of the MLR residuals, and (d) residuals versus predicted SOM value

4.3.4 Prediction of SOM using geostatistical methods

In Table 4.8, an overview of the parameters used to model the semivariograms for the SOM content and residuals of the two MLR models are displayed. With regards to the range of the semivariograms, the measured SOM content had higher range values compared to the residuals of the two different MLR models. More specifically, the highest range (868.45 m) was covered by the semivariogram for SK using the SOM content, while the lowest range (648 m) was covered by the ROK using the residuals of spectral bands MLR model. In terms of the nugget, partial sill, and sill values, these ranged from 0.390 to 1.013, 0.365 to 0.736, and 0.885 to 1.446 respectively. Furthermore, the nugget/sill ratio was the highest for the measured SOM content using OK (0.735) and the lowest for the measured SOM content using SK (0.406), while the rest of the ratios ranged from 0.458 to 0.511.

Table 4.8: Semivariogram model parameters for SOM content and residuals of the MLR models derived from spectral bands and PCs, respectively

Data type	Interpolation method	Model	Range (m)	Nugget (C_0)	Partial sill (C)	Sill ($C_0 + C$)	C_0/Sill
Measured SOM	OK	Spherical	723.21	1.013	0.365	1.378	0.735
Measured SOM	SK	Spherical	868.45	0.390	0.570	0.960	0.406
Residuals of spectral bands MLR	ROK	Spherical	648.00	0.452	0.433	0.885	0.511
Residuals of spectral bands MLR	RSK	Spherical	785.29	0.570	0.604	1.174	0.486
Residuals of PCs MLR	ROK	Spherical	690.87	0.710	0.736	1.446	0.491
Residuals of PCs MLR	RSK	Spherical	814.57	0.511	0.604	1.115	0.458

The results of the variography analysis are depicted as a summary in Table 4.8. It outlines the parameters used in fitting the appropriate semivariogram models to the SOM content and residuals of the two MLR models' datasets. Firstly, these results revealed that the data for both the measured SOM content and residuals obtained from the two MLR models were well fitted by a spherical model. The selection of the semivariogram model was obtained via experimental variography analysis, as mentioned in the methodology section previously. Subsequently, these results fall in line with the works of Li (2010), Mirazee *et al*., (2016), and Gia Pham *et al*., (2019), where a spherical model was also used during variography analysis regarding the prediction of SOM. Furthermore, the results depicted that range of the measured SOM content was higher when compared to the range of the residuals of the MLR models. This can be attributed to the auxiliary information that was added to the MLR models. In this particular study, this information included the spectral bands and spectral indices. According to Li (2010), the addition of such information to regression models can lead to the shortening of the semivariogram's model range. Another important result obtained from the variography analysis is the nugget to sill ratio (i.e., Nugget/Sill). This ratio is used to describe the degree of spatial dependence and random variation for the factor being estimated (Song *et al*., 2017). Simply put, it is used to indicate the spatial structure of the estimated factors. The ratio can be classified into three categories where; 0–25% indicates strong spatial dependence, 25–75% indicates moderate spatial dependence, and >75% indicates weak spatial dependence (Jobbágy

and Jackson, 2000). From the results obtained, it is evident that the majority of these ratios lie between the range of 0.406 to 0.511, with the exception of the measured SOM content using OK semivariogram. Therefore, based on the above Nugget/Sill ratio categories, it is evident that the results indicate a moderate spatial dependency for the SOM data. Furthermore, the ratio revealed that a considerable proportion of the variation was spatially correlated based on the scale of sampling.

In Figure 4.4, the scatterplots for both the training and validation datasets using the measured SOM against the predicted SOM from the various geostatistical models are illustrated. In terms of Figure 4.4a and b, the scatterplots reveal a more or less constant trend as the lines of best fit follow a horizontal pattern. In both cases, the data points are distributed randomly in close proximities to their respective best fit lines, with the exception of some points. However, in Figure 4.4c and d, it is evident that the scatterplots of both the training and validation datasets follow a linear trend as the best fit lines are close to a 45° angle, with only a few points deviating from this line. Subsequently, it is also evident that the data points are more compact and closer to their respective lines of best fit. Moreover, Figure 4.4e and f also depicted a linear trend for both the training and validation datasets. However, more points were deviating from the lines of best fit. Subsequently, the data points were less compact and more spread out along their respective lines.

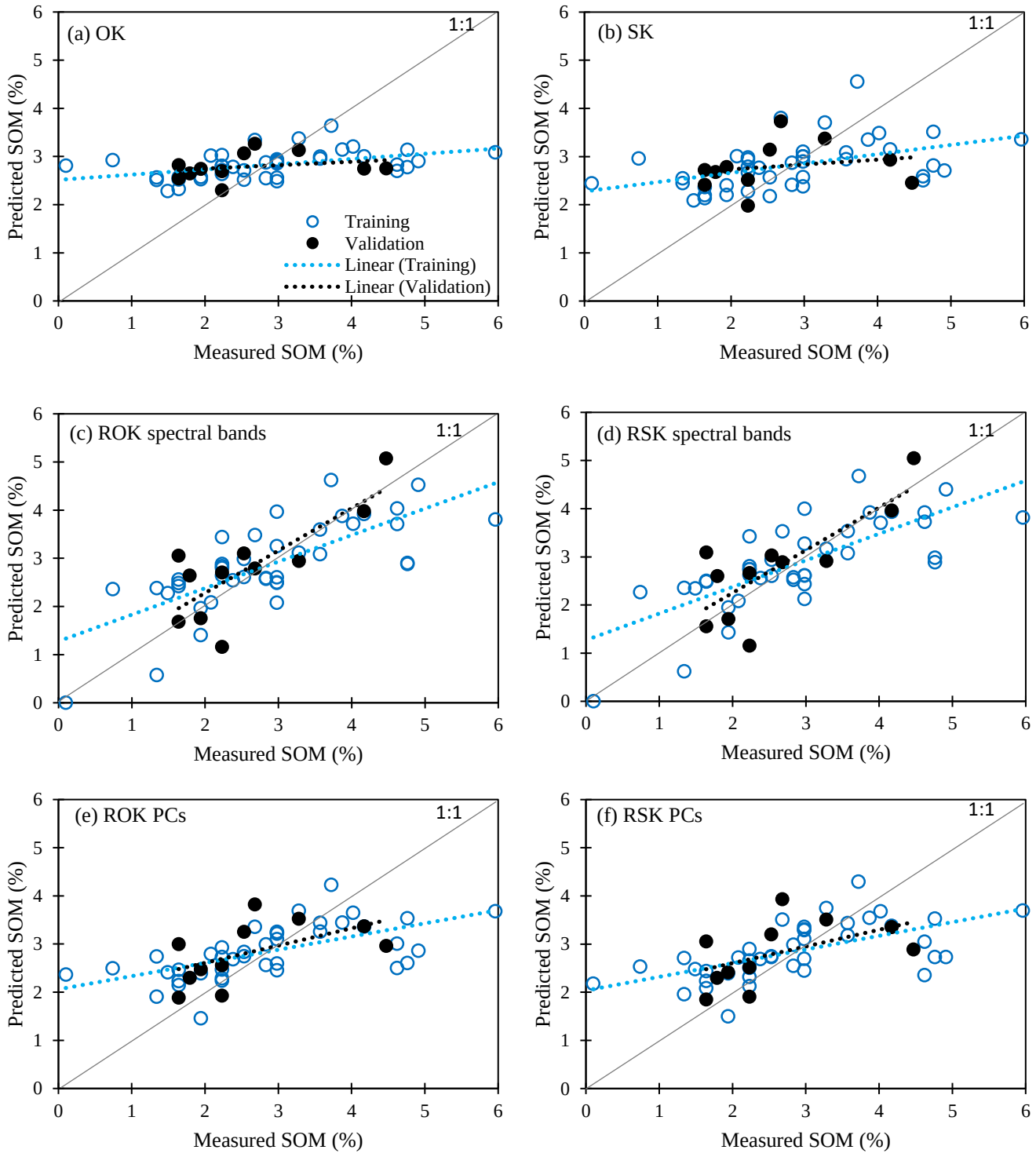


Figure 4.4: The measured versus predicted SOM content, using geostatistical (a) OK, (b) SK and hybrid geostatistical methods (c) ROK of spectral bands, (d) RSK of spectral bands, (e) ROK of PCs, and (f) RSK of PCs

The results of the measured SOM content plotted against the predicted SOM content is displayed in Figure 4.4 using the different geostatistical and hybrid geostatistical methods for both the training and validation datasets. These scatterplots provide a graphical representation of the relationship between the measured and predicted SOM content. From the results, it was evident that Figure 4.4a and b had a more or less constant trend. This may be due to the random distribution of points along the lines of best fit. Subsequently, it is also an indication that the OK predicted SOM content percentages that were either over or under-estimated when compared to the measured SOM content. Therefore, indicating that little to no relationship exists between the measured and predicted SOM content. However, the converse results were depicted for Figure 4.4c and d, where a linear trend was clearly visible. This linear trend between the measured and predicted SOM content indicates that a moderately strong linear relationship exists between them. This relationship is also positive. Furthermore, the compactness of the data points to their respective lines of best fit within these scatterplots may have influenced such a relationship. Subsequently, the use of auxiliary information (i.e., spectral bands) may have influenced the prediction of SOM, as well as the type of geostatistical method used, could have resulted in such a relationship (Dobos *et al*., 2006). Additionally, the results depicted in Figure 4.4e and f were similar to that of Figure 4.4c and d; however, the only difference being that the positive linear relationship was not as strong. This may be attributed to the use of the PCs (i.e., derived from spectral indices). According to Mirazee *et al*., (2016), a non-linear relationship exists between SOM content and PCs. Thus, resulting in a lower but positive linear relationship between the measured and predicted SOM content. Subsequently, the hybrid geostatistical methods used may have also had an influence.

In Figure 4.5, the maps display the predicted SOM content using the different geostatistical and hybrid geostatistical methods. In Figure 4.5a, it is evident that higher percentages of SOM content are found in the western region of the study area, while the lower percentages of SOM are found in the central and south eastern region. However, there is also a moderate percentage of SOM in the northeast of the lower portion of the study area. In terms of Figure 4.5b majority of the area is covered by a moderate percentage of SOM, with the exception of the central region, where the lowest percentages of SOM are seen. Conversely, on either side, the highest percentages of SOM are found together with hotspots in the northern regions of the study area. With regards to the maps derived from hybrid geostatistical methods, similar trends are evident. For instance, in Figure 4.5c, it can be seen that the western region of the study area contains moderate to high SOM percentages, whereas the eastern region contains lower percentages of

SOM content, with the exception of the hotspot in the northeast. Subsequently, in Figure 4.5d, similar trends are visible; however, the western region contains higher SOM percentages while the eastern region contains relatively lower SOM percentages. Moreover, in Figure 4.5e majority of the area is covered by a moderate percentage of SOM, with the exception of hot and cold spots of SOM in the central regions. More specifically, in the centre of the study area, the lowest percentage of SOM content is found, and to the left and right of this, the highest percentages of SOM are found with three and one hotspots, respectively. Once again, similar trends that were present in Figure 4.5e are visible in Figure 4.5f. However, the only difference being that the hot and cold spots are much larger hence covering a greater area with higher and lower percentages of SOM content, respectively.

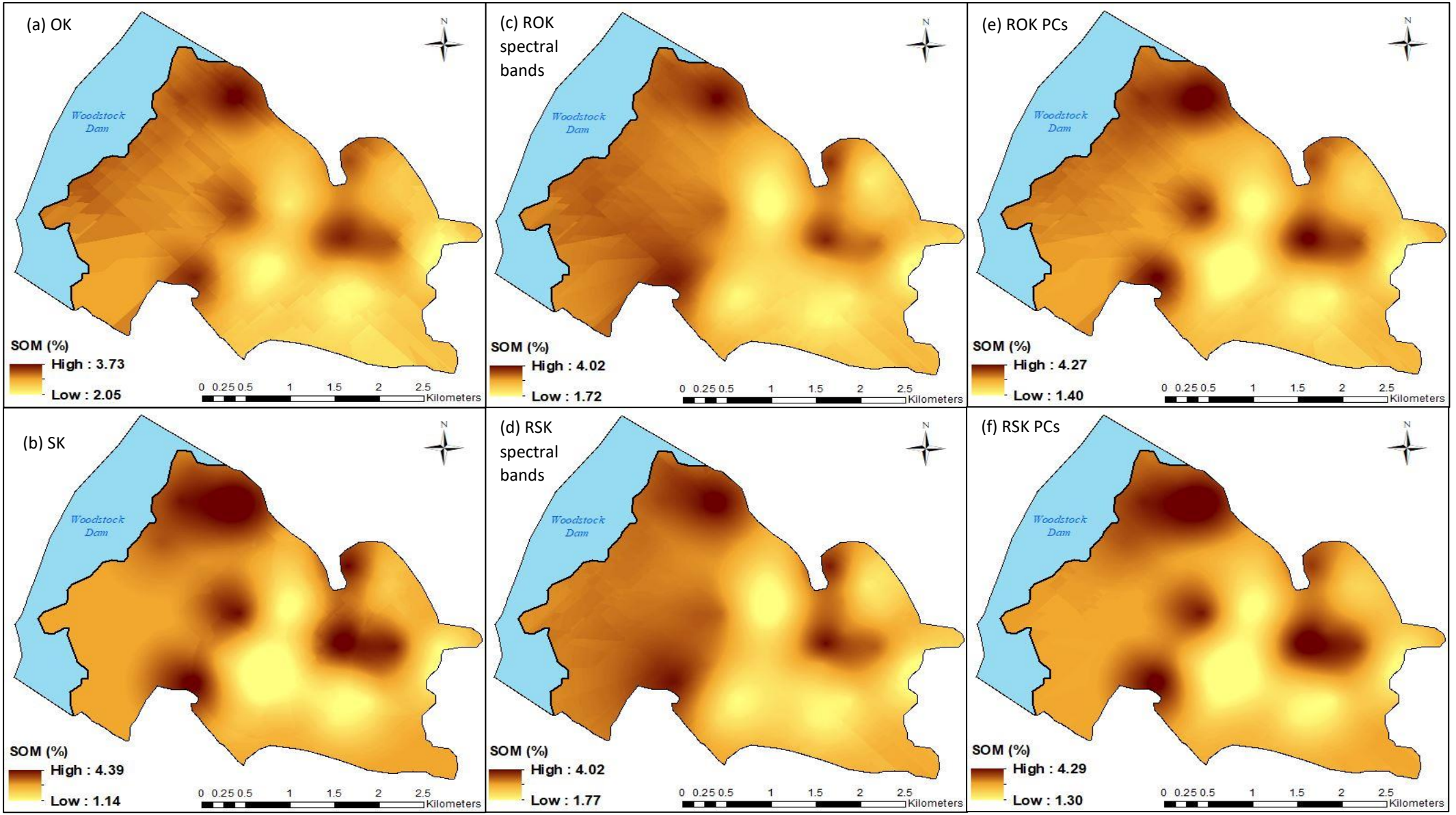


Figure 4.5: Maps of predicted SOM content (%) using geostatistical (a) OK, (b) SK and hybrid geostatistical methods (c) ROK of spectral bands, (d) RSK of spectral bands, (e) ROK of PCs, and (f) RSK of PCs

The results of the predicted SOM content percentages using different types of geostatistical and hybrid geostatistical methods are displayed spatially by the maps shown in Figure 4.5. According to these results, it was evident that the six prediction methods contained similar trends. More specifically, they had a similar distribution trend of the SOM content regarding the general spatial patterns throughout the study area. For instance, in all the maps, it was apparent that the western region of the study area contained the highest percentages of SOM content. This is a result of the agriculture land use that is quite dominant in the western region of the study area. As seen in Figure 4.1, this land use type is relatively extensive and covers the majority of the western region. These small-scale farms play an active role in influencing the formation of and maintenance of SOM content. For instance, crops are grown regularly within agricultural land uses; therefore, there is a regular supply of organic material in the form of leaf litter and stems that is added to the soil (Behera and Prasad, 2020). As mentioned previously, these materials are one of the major sources of organic matter in soils. Therefore, a relatively good supply of these materials will facilitate the decomposition process and thus the formation of SOM. Another reason for the higher SOM percentages witnessed in the western regions of the area is a result of the manner in which the soil is treated. Soil within the agricultural land use is tended to or taken care of regularly via irrigation and manure application to ensure a greater yield of crops produced. These additional treatments of soil ensure proper maintenance and accumulation of SOM content (Bot and Benites, 2005). Furthermore, it was evident that the area in the northeast region also depicted higher percentages of SOM compared to its surrounding area (Figure 4.5). This can also be accounted for by the above-mentioned reasoning due to the presence of agricultural practices on a small scale (Figure 4.1). Subsequently, other land use types such as rangeland could have resulted in higher percentages of SOM within this region. These land use types usually contain extensive vegetation, which provides the soil with organic material. This, coupled with the animal dung from the grazing of livestock, can influence the SOM content within the region (Haynes *et al*., 2003).

In contrast, the central to eastern regions of the study area contained the lowest percentages of SOM content. This can be accounted for by the ongoing severe erosion within these regions. The extensive eroded land can be clearly seen in Figure 4.1. It can also be observed that within these eroded areas, ground cover is little to none, thus leaving the soil exposed to adverse environmental factors. More specifically, this means that there is a lack of organic material for SOM formation as well as protection from topsoil loss and hence, the maintenance of SOM.

According to Phuong *et al*., (2017), soil erosion leads to the loss of fertile topsoil and SOM. Therefore, indicating lower percentages of SOM predicted within these regions.

Moreover, the presence of hot (i.e., higher percentages of SOM content) and cold (i.e., lower percentages of SOM content) spots are quite evident in the maps of Figure 4.5. These spots of high and low percentages of SOM content are mainly distributed in the central region of the study area. The creation of such spots may be attributed to the above-mentioned factors that influenced the spatial distribution of SOM content within the study area. Other reasons may include the influence of certain measured SOM content percentages at certain sampling points along these regions. For example, a higher percentage of SOM content would have a greater chance of influencing the predicted SOM content at a particular location than compared to a lower percentage. Such extreme high or low values in SOM content may have resulted in the creation of hot and cold spots. Another reason for these spots is that the percentages of SOM content measured at certain sampling points may be similar in value and therefore resulted in the predicted SOM content percentages being quite similar. The reason for similarities in SOM content between sampling points may be attributed to the factors that influence SOM content, such as land use type, soil type, soil-forming factors, moisture, vegetation, etc. (Bot and Benites, 2005).

With regards to the ranges of the predicted SOM content percentages of the different prediction methods, it was evident that they were relatively similar to one another. But, more specifically, they were relatively similar to the range of the measured SOM (0.00% to 5.96%). However, a slight difference did exist between the ranges of the measured SOM and the predicted SOM percentages, with the exception of the OK method in Figure 4.5a. This may be due to the type of geostatistical and hybrid geostatistical method used in the prediction of SOM content.

4.3.5 Validation of prediction methods

In Table 4.9, the calculated evaluation statistics are displayed for each of the interpolation methods used. With regards to the ME values, it is evident that SK had the highest value (0.83%), and the ROK of spectral bands had the lowest (0.53%) overall ME value. However, when comparing each set of interpolation methods to each other, it can be seen that OK and SK had the highest values, followed by ROK and RSK of PCs and then ROK and RSK of spectral bands having the lowest values. However, in terms of the RMSE, the error values were relatively higher than that of the ME values. Once again, it is evident that SK had the highest value (0.98%), and the RSK of spectral bands had the lowest (0.67%) overall RMSE value.

Regarding each set of the interpolation methods, the same trend as the above mentioned is evident for the RMSE values. Additionally, the R^2 values were significantly lower when compared to the other statistics overall. The values ranged from 0.04 to 0.63, with SK and ROK and RSK of spectral bands having the lowest and highest R^2 values, respectively. However, when comparing each set of interpolation methods to each other, it can be noted that ROK and RSK of spectral bands had the highest R^2 values, followed by ROK and RSK of PCs and then OK and SK having the lowest R^2 values.

Table 4.9: Evaluation statistics of the different interpolation methods using the validation data set

Interpolation method	**R^2**	**ME (%)**	**RMSE (%)**
OK	0.07	0.79	0.93
SK	0.04	0.83	0.98
ROK of spectral bands	0.63	0.53	0.68
RSK of spectral bands	0.63	0.54	0.67
ROK of PCs	0.30	0.70	0.82
RSK of PCs	0.26	0.71	0.85

The results obtained in Table 4.9 outline the evaluation statistics of the different interpolation methods that were used to predict SOM content. The validation dataset was used to calculate these evaluation statistics. These statistics were used to assess the performance and accuracy of each geostatistical and hybrid geostatistical method in terms of predicting the SOM content within the study area. With regards to the prediction error statistics calculated (ME and RMSE), the ME values were moderately low for each of the interpolation methods. These values were smaller than one, and therefore, indicated that all six methods did a fairly good job in predicting the SOM content, especially the ROK and RSK of spectral bands method. However, in terms of the RMSE, the error values were moderately higher than that of ME. As mentioned previously, this is due to the fact that RMSE values are usually never smaller than that of the ME values (Chai and Draxler, 2014). Furthermore, the RMSE values were also smaller than one for all six prediction methods, suggesting that they all did a fair job predicting the SOM content. Even though the RMSE of the OK and SK methods were relatively close to a value of one, they still did a fair job in the prediction of SOM content. Influence from certain percentages of measured SOM content could have resulted in higher RMSE values for both the OK and SK methods. Subsequently, the interpolation method itself and the manner in which it calculates the predicted values could have influenced the higher RMSE values as well. The fact

that both of the error statistics were relatively low; suggests that the interpolation methods did, in fact, do a moderately fair job in predicting the percentage of SOM content within the study area.

Moreover, according to the results, the R^2 values were significantly lower when compared to the other statistics overall. This is accounted for by the fact that R^2 values usually range from 0 to 1 (Bloomenthal, 2020). These values further indicate the variability that exists between the measured and predicted percentage of SOM content. In other words, it provides an indication of how well the prediction method performed by comparing the measured with the predicted percentage of SOM content. According to the results in Table 4.9, it was evident that the OK and SK methods had the lowest R^2 values from all interpolation methods used in this study. They each had a value of 0.07 and 0.04, respectively, which meant that each method accounted for 7% and 4% of the predicted SOM content, respectively. These low percentages indicate weak positive relationships between the measured and predicted percentages of SOM content.

In contrast, the ROK and RSK of spectral bands methods had the highest R^2 values from all interpolation methods used in this study. More specifically, they both had a value of 0.63, which meant that each method accounted for 63% of the predicted SOM content, respectively. These higher percentages indicate moderately positive relationships between the measured and predicted percentages of SOM content. Subsequently, the ROK and RSK of PCs methods had moderately lower R^2 values of 0.30 and 0.26, which meant that each accounted for 30% and 26% of the predicted SOM content, respectively. These moderately low percentages indicate weak positive relationships between the measured and predicted percentages of SOM content. These relationships can also be seen graphically from the scatterplots shown in Figure 4.4 of the validation dataset.

The major reason as to why the hybrid geostatistical methods performed better than the geostatistical methods is because of the use of auxiliary information. The auxiliary information used in this study was derived from remote sensing data. More specifically, the Sentinel-2 imagery. The data from the satellite imagery was used in two ways; namely, the reflectance's from each of the selected bands and the calculation of spectral indices using these reflectance values at the sampling points. This data was used as auxiliary information in the prediction of SOM content. This auxiliary information provided the hybrid geostatistical methods with additional information that could be used to improve the precision of the predictors during the prediction process. According to Mirazee *et al.*, (2016), information from remote sensing data

can act as an extremely important auxiliary variable for improving the estimates of SOM content and its spatial variability. Furthermore, the additional information allows various factors that influence soil and SOM formation, maintenance, and accumulation to be taken into consideration when predicting its distribution on a spatial scale. And, since the OK and SK methods (geostatistical methods) were performed using only the measured SOM content and did not have such additional information to assist in prediction, thus the prediction performance turned out to be quite poor. Subsequently, the hybrid geostatistical methods combined linear regressions with geostatistical analysis. Such combinations haven proven to be quite superior compared to regular geostatistical methods as it yields more detailed results and has higher accuracies with regards to prediction (Hengl *et al*., 2004). Subsequently, according to Song *et al*., (2017), hybrid geostatistical methods such as RK are more effective than ordinary geostatistical methods. Such combinations coupled with auxiliary data have outperformed ordinary geostatistical methods such as OK, SK, and Co-Kriging in many studies (Hengl *et al*., 2004). The absence of combining additional methods with the OK and SK methods could have influenced their prediction performance as they were solely relying on the manner in which the predicted SOM content is calculated. These reasons can account for the low prediction performances of both the OK and SK methods, as their prediction of SOM content did not match that of the measured SOM content. This, coupled with the error statistics obtained for these methods, further support their lower predictions in SOM content. Moreover, the aforementioned reasons can also be used to support the higher prediction performances of both sets of hybrid geostatistical methods. With regards to the ROK and RSK that involved the use of spectral bands, both performed the best overall when predicting SOM content, as they each accounted for 63% variability in their predictions. Despite having the best performances for the study, several factors may have implicated their prediction. Among these factors are soil water content, calcium carbonate, residue cover, and mineralogy (Hummel *et al*., 2001; Stenberg *et al*., 2010).

In contrast, the ROK and RSK that involved the use of PCs derived from the calculated spectral indices did not perform as well despite its use of auxiliary information. This can be attributed to the PCA performed using the spectral indices. During the process of reducing the dimensionality and multicollinearity for the spectral indices' dataset, some information may have been lost (Geiger and Kubin, 2012). This lost information could not have been used for analysis, which in turn affected the ability of these interpolation methods to predict SOM content.

The results of the study indicate that remote sensing data plays a crucial role in improving the prediction of SOM content when using geostatistical and hybrid geostatistical methods. More specifically, they were in line with findings from other studies, which also found that hybrid geostatistical methods have better predictive performances and are more accurate when it comes to the prediction of soil properties such as SOM content than compared to geostatistical methods (Hengl *et al*., 2004; Maynard *et al*., 2011; Zhang *et al*., 2012; Motaghian and Mohammadi, 2011; Mirazee *et al*., 2016).

4.4 Conclusion

The study has provided a critical analysis of the prediction of SOM content by using geostatistical methods and hybrid geostatistical methods. The results revealed that hybrid geostatistical methods (ROK and RSK of spectral bands and PCs) outperformed ordinary geostatistical methods (OK and SK). More specifically, it was revealed that the addition of auxiliary information such as the reflectance's from spectral bands as well as calculated spectral indices obtained from remote sensing data assisted in the prediction performance and accuracy of the interpolation methods. On the other hand, the absence of such information resulted in poor predictions that were inaccurate. Overall, the ROK and RSK of spectral bands had the most accurate prediction, followed by ROK and RSK of PCs, and lastly, the OK and SK were the least accurate. Therefore, the incorporation of auxiliary remote sensing data such as Sentinel-2 imagery can significantly improve the prediction of SOM content. This, coupled with the fact that Sentinel-2 data is readily available (i.e., open-source) with a high spatial and temporal resolution, means it can be quite beneficial for predicting SOM content. Such techniques allow for SOM content to be monitored and estimated in a sustainable manner. Furthermore, important inferences can be made regarding the spatial and temporal variability of SOM within an area. And, since SOM governs the physical, chemical, and biological properties of soil, it is imperative that SOM and soil are managed and protected in a sustainable manner for future generations to come. Therefore, it is important for studies such as the above to be carried out so that useful information can be gathered efficiently regarding the estimation, prediction, and quantification of SOM content on both local and regional scales.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

The current thesis set out to examine the SOM content within the Emakhosaneni area, KwaZulu-Natal, South Africa, by measuring the reflectance of soil using laboratory and remote sensing analysis. The execution of this aim was guided by the objectives of the thesis outlined in chapter one. They each provided insight into the SOM content within the study area through the estimation, assessment, and prediction of SOM content by using both laboratory-based and remote sensing techniques. Conducting a critical analysis of the SOM content further provides an understanding of the natural functioning of soils and, thus, SOM. It is, therefore, important to monitor, assess, and manage the overall quality and quantity of SOM as it is directly linked to soil fertility and productivity.

Traditional methods, despite their efficiency, can be rather time-consuming, costly, and labour extensive (Golicz *et al*., 2019). Remote sensing, on the other hand, is an efficient method that is time effective, low-cost, and non-destructive while also having the capability of rapid data acquisition (Nawar *et al*., 2016; Wang *et al*., 2017). Thus, offering an alternative to traditional soil analysis methods such as; the estimation, assessment, and prediction of SOM content.

A laboratory-based approach facilitated the estimation of SOM content within the four major land uses of the study area. This allowed for an insight into the deterioration of SOM and its risks imposed on the residents and their livelihoods as well as the surrounding environment. Furthermore, the laboratory-based remote sensing allowed for spectral reflectance measurements of soil samples to be taken. This provided an understanding of the relationship between the SOM content found within its major land use types and the spectral reflectance from its soils. Subsequently, satellite remote sensing data was used to predict SOM content by applying geostatistical methods and hybrid geostatistical methods.

This chapter provides recommendations that are specific to the current thesis. Firstly, the chapter opens with concluding remarks regarding the overall findings of the current thesis and the important information gathered from each of the studies. Moreover, recommendations regarding future studies are also provided in order to ensure the reliability and validity of the results produced from the study. Lastly, suggestions regarding the appropriate measures that should be taken in order to improve the deteriorating SOM content experienced within the

study area are provided. These suggestions are for alleviating the severity of the problem on the environment and the residents within the Emakhosaneni area.

5.2 Conclusions

The current thesis provided an overall critical analysis of the SOM content within the Emakhosaneni area, KwaZulu-Natal, South Africa, by measuring the reflectance of soil using laboratory and remote sensing analysis. Each of the chapters in this thesis undertook a different approach in analysing and examining the SOM content. More specifically, they each focused on the estimation, assessment, and prediction of SOM content by using both laboratory-based and remote sensing techniques. And, the concluding remarks are summarised briefly in the subsequent paragraphs.

The first study (chapter two) examined the total SOM content within the soil of the Emakhosaneni area located in KwaZulu-Natal, South Africa, using laboratory techniques. More specifically, the study focused on determining the percentage of SOM content present within the study area, and its four major land uses. For this, a total of 52 soil samples were obtained from the study area, and the SOM content was analysed using laboratory analysis (WB method). The results from the study revealed that the area has an average SOM content of 2.79% present within its soils. This lower percentage of SOM content may be due to the semi-arid nature of the area coupled with human activities and their impacts as well as environmental factors. It was also revealed that the four major land use types exhibited a similar pattern, to a certain extent. With the agricultural land use having the highest average percentage of SOM content, followed by rangeland, built-up, and eroded land use. These results further indicated the influence of land use activities on the SOM content within the study area. Furthermore, the results of the one-way ANOVA revealed that the percentage of SOM within the major land uses were statistically significant ($p<0.05$). Therefore, indicating that the percentage of SOM within each of the land uses varied and that they are different from one another. However, this difference was determined by using a post-hoc Tukey's test, and the results indicated that only the percentage of SOM within the agriculture land use and eroded land were statistically significant ($p<0.05$). Thus, suggesting that the presence of vegetation coupled with human intervention is responsible for promoting biological activity which can generate and enhance SOM content. Overall, these results suggested that the SOM content within the Emakhosaneni area is deteriorating while also highlighting the severity and consequences of the problem.

The second study (chapter three) examined the relationship between SOM content and laboratory-based spectral reflectance within the soil of the Emakhosaneni area located in KwaZulu-Natal, South Africa. More specifically, laboratory spectroscopy was used to identify and examine a relationship between the SOM content and spectral reflectance of soils within the study area. The results from the study revealed that the average reflectance of spectra within each of the major land uses differed but had similar trends in their spectral reflectance curves. For instance, the eroded land had the highest average reflectance, followed by rangeland, built-up, and agricultural land use. And the reason for this is soil colour which is directly linked to SOM content. This, in turn, suggested that the spectra obtained using the spectroradiometer was influenced by soil colour and established a relationship between SOM content and reflectance of soil samples. In addition, this relationship was further investigated through correlation analysis, and it was revealed that a negative moderate relationship existed between SOM content and spectral reflectance for the overall data. However, the land uses agriculture and built-up had negatively moderate to strong correlations while the eroded and rangeland land uses had a positively strong correlation with its reflectance spectra. This further revealed the influence of SOM content on the spectral reflectance of the soils within the area by suggesting that as the SOM content increases within a soil sample, the spectral reflectance is lowered and vice-versa. PLSR models were used to further assess this relationship. The four models created using different types of pre-processed spectra and SOM content revealed a fair performance, with the FDT model having the best performance with low errors (R^2 =0.49, RMSEC=0.73% and R^2=0.22, RMSEP=1.11%). Subsequently, the R and SG models did not do an efficient job in predicting SOM content as compared to the FDT and SG+FDT models. This was due to spectral noise, baseline offsets, absorptions, and any other irregularities present within the spectral data that influenced both performance and accuracy. Overall, the study highlighted the importance of remote sensing, especially VNIR spectroscopy, in examining SOM content compared to the conventional laboratory approaches, which can be rather costly and time-consuming.

The third study (chapter four) assessed the relationship between SOM content and satellite remote sensing data within the soil of the Emakhosaneni area located in KwaZulu-Natal, South Africa. The use of Sentinel-2 imagery was utilised in the prediction of SOM content by using geostatistical methods (OK and SK) and hybrid geostatistical methods (ROK and RSK of spectral bands and PCs). The results from the study showed that the hybrid geostatistical methods performed better than the geostatistical methods in predicting SOM content. More

specifically, the RSK using spectral bands had the overall best performance with R^2 =0.63 and RMSE=0.67%. This was followed by the performance of the ROK using spectral bands, ROK and RSK using PCs and lastly, the OK and SK methods. This fairly good performance by the ROK and RSK using spectral bands and PCs models were attributed to the contribution of auxiliary information. This information included the reflectance values from spectral bands as well as calculated spectral indices obtained from Sentinel-2 imagery. It enhanced not only the performance of the models but also the accuracy in predicting SOM content. Therefore, the incorporation of auxiliary remote sensing data such as Sentinel-2 imagery can significantly improve the prediction of SOM content. Moreover, it also highlights the potential of the Sentinel-2 imagery data in predicting SOM content in KwaZulu-Natal, South Africa, while also making critical inferences regarding the spatial and temporal variability of SOM within an area.

Overall, each of the studies conducted in this thesis produced significant results that pertain to the deteriorating SOM content within the Emakhosaneni area. They highlighted the severity of the problem while also using different approaches, techniques, and technologies to estimate, assess, and predict SOM content. Since SOM governs the physical, chemical, and biological properties of soil, it plays a critical role in the overall functioning, fertility, and productivity of soil. This is especially relevant to both the soil environment and its organisms as well as humans. It also provides a myriad of ecosystem services to plants, animals, and humans. Therefore, a disturbance in SOM content within the soil can have considerable implications on the above mentioned. It also further suggests the importance of monitoring, maintaining, and managing SOM content within the soil. This is particularly relevant to the Emakhosaneni area as both farmers and residents rely on the fertility of the soil in order to improve yield production and growth of natural vegetation for the grazing of livestock along with other activities. However, poor management practices coupled with the ongoing erosion in the area has made this extremely difficult. It is, therefore, critical that SOM content be monitored and measured regularly; however, this can be relatively expensive and time-consuming and hence, remote sensing approaches have been used as an alternative. The effectiveness of remote sensing coupled with its rapid data acquisition, low cost, convenience, non-destructive nature, and controllability have proven to be relatively efficient in examining and predicting SOM content.

5.3 Recommendations

The results and discussion produced in the previous chapters revealed many recommendations that originated from the findings of the studies. These recommendations pertain to both future studies in order to ensure the reliability and validity of the results produced from the study and residents of the Emakhosaneni area and the surrounding environment.

In terms of the recommendations that pertain to carrying out similar research studies in the future, the following can be taken into consideration:

- Regarding the number of soil samples taken from each of the four major land uses, a greater number should be obtained. This would ensure that a more representative dataset is obtained from each of the land uses for analysis. Furthermore, it will also provide more information on the soils and the surrounding environment within the Emakhosaneni area. Thus, improving the accuracy and reliability of the results obtained.

- Keeping with the idea of obtaining a greater number of soil samples, they should also be obtained during different months or seasons of the year. This will particularly be interesting to see how seasonal changes influence the SOM content within the Emakhosaneni area. Subsequently, it would provide an insight on the factors, such as climate, and how they play a role in influencing SOM content within the area on a temporal scale. Also, it would give a greater understanding of how this content varies within the different land use types throughout the year within the Emakhosaneni area.

- The laboratory method that was used to analyse the SOM content within this study area was the WB method. Despite the common use of this method, it does have shortcomings, especially when working with a large number of samples. Therefore, other methods such as the LOI method, HM, etc., should be considered when analysing SOM content. The different methods may be more accurate, efficient, cheaper, and time effective. Furthermore, it would be interesting to see how each method quantifies SOM content and the different results they produce compared to the WB method. Subsequently, a different method may bring to light new information regarding the SOM content.

- The method in which the spectral measurements was taken using the ASD spectroradiometer should be modified. This includes taking measurements that are at a

closer distance to the soil sample under laboratory conditions, as this would minimise the scattering of light and any additional noise that may contaminate the spectral measurements. Lowering this distance would, in turn, concentrate the FOV covered by the spectroradiometer on a smaller area of the soil samples and would also play a significant part in reducing the spectral noise of the samples.

- Apart from laboratory-based spectroscopy, field spectroscopy can also be used in the collection of spectral measurements. These in-situ measurements can be used in conjunction with the laboratory measurements to gain a better understanding of the reflectance of soil samples and their relationship with SOM. In contrast, a comparative analysis between the two methods can also be done in order to determine the efficiency and accuracy of the two methods when it comes to understanding the relationship between the reflectance of soil samples and SOM content, especially within different land use types.

- The pre-processing of reflectance spectra plays a critical role in determining the overall accuracy of the analysis. Therefore, a variety of pre-processing techniques should be applied. These include different smoothing methods, baseline removal, scaling, and normalization techniques such as second-derivative reflectance, multiplicative scatter correction, etc., all of which can be used to minimize the noise and any irregularities present within the spectra. According to (Wei *et al*., 2020), pre-processing techniques can also improve the correlation between SOM content and spectral reflectance, therefore making it easier to extract information about sensitive wavebands. Different pre-processing techniques can also provide critical information on any trends or patterns that may exist within the reflectance spectra obtained.

- When it comes to the statistical analysis of the spectral reflectance measurements and SOM content, a PLSR was used. In addition, different statistical methods such as the Random Forest Regression and Support Vector Machine Regression can be used in place of or in conjunction with the PLSR method to create the prediction models. These different methods may yield better accuracies and have higher predictive performances. Furthermore, they may be more effective (e.g. computationally faster, easier, less time-consuming, etc.) in analysing the relationship between the reflectance of soil samples and SOM content.

- Satellite remote sensing was used in the prediction of SOM content within the Emakhosaneni area. The prediction models were facilitated by the use of auxiliary variables obtained from the satellite imagery. However, future research should consider using a greater number and variety of auxiliary variables obtained from both remote sensing and other data sources. These variables should be included in the hybrid geostatistical methods to assist in improving the prediction performance and accuracy of the models created. Among the variables are elevation, slope, aspect, a greater variety of spectral indices, soil type, etc. Therefore, a better predictive model will be established and allow for a greater understanding of the spatial patterns of SOM content within an area.

- Moreover, different hybrid geostatistical methods such as artificial neural networks and extreme learning machine can also be considered for the prediction of SOM content. This would provide different perspectives and allow for important comparisons to be made regarding the prediction of SOM content. It can also highlight which of these methods are both effective and efficient when it comes to predicting SOM.

- Future research should consider using high resolution imagery with spatial resolutions below 5m, such as IKONOS and QuickBird. These higher resolution images will allow for a greater detail to be seen compared to the Sentinel-2 imagery. Also, other freely available imagery sources such as Landsat and SPOT can be used in the prediction of SOM content. This will be especially useful when comparing the results obtained from the Sentinel-2 imagery.

Regarding the deteriorating SOM estimated within the Emakhosaneni area, the following measures can be put into place to assist and alleviate the severity of the problem.

Soil fertility is directly linked to SOM, and a low fertility of soil is indicative of low SOM content. Therefore, soil fertility needs to be increased once again in the area through the application of fertilizers and animal manure. Such applications are rich in nutrients such as nitrates, phosphorous, potassium, ammonia, etc., all of which are essential elements to both the soil and plants. Their addition to the soil in the area will assist in maintaining nutrient supply to the soil and its living organisms. However, fertility applications should be applied in balanced proportions and sufficient quantities (Bot and Benites, 2005). Subsequently, the application of manure can be facilitated by the supply produced from the rearing of livestock

in the area. This natural fertilizer can also serve as a source of organic material for decomposition and hence SOM formation.

The study area is prone to severe erosion by wind and water, as outlined in chapter two. The ongoing erosion has made it difficult to maintain SOM content as large amounts of fertile topsoil are either blown away or washed away via surface runoff. This, in turn, leaves the soil surface exposed to harsh environmental and human factors that can further impair the remaining SOM content. Therefore, it is important that the soil's surface remains protected during this vulnerable state so that SOM equilibrium can be achieved. This application of crop residues can act as the protective layer and reduce the impacts caused by erosion. Despite certain farmers in the study area applying this technique, its results were not as effective due to the sparsity and quantity of the residues. And ensuring these limitations be met, then only will the technique be effective. Subsequently, grass cuttings can also be applied to the surface. These residues serve a dual purpose in that they also act as a source of organic material for decomposition and soil organisms and thus, SOM.

Residents can also partake in planting natural vegetation (e.g. grasses, shrubs, trees, etc.) or cover crops (e.g. grains, legumes, and oil crops) to act as a soil cover, especially during the rainy and windy periods. This will protect the soil's surface and prevent further loss of topsoil and SOM as well as reduce the rate of erosion. It can also assist in the restoration of the soil's functioning by increasing microbial activity and biodiversity within the soil. This, in turn, will aid the SOM formation process.

In addition, the prolonged use of soil by the residents, especially for agricultural purposes, puts a strain on the soil and its functioning, health, and productivity. Therefore, farmers and residents should allow the soil to rest so that it can restore itself before being used again. By allowing the soil to rest for a specific period of time without any disturbances from human activities, the natural productivity of the soil can be regenerated. It can also allow the accumulation and storage of SOM. However, during this resting period, it is also important that the soil's surface be protected in order to prevent further deterioration of the SOM content.

Lastly, an awareness needs to be created amongst the residents of the Emakhosaneni area in terms of the deterioration of the SOM content present within the soils. This can be achieved through focus group discussions, workshops, pamphlets, posters, etc., where residents are provided with information on SOM and its formation, maintenance, and accumulation. Furthermore, they will provide information on SOM depletion and the many factors that

contribute to and exacerbate the process of SOM depletion and how this can impact the soil health, quality, and productivity. Subsequently, residents need to be given the appropriate management techniques and strategies that are specific to their area and surrounding environment in order to assist in alleviating and rebuilding the deteriorating SOM content within the Emakhosaneni area. Such awareness can also assist future generations as the information gathered from such tools and management strategies can be passed on overtime. This would assist both the soil and its surrounding environment within the Emakhosaneni area greatly.

REFERENCES

Adams, C.R., Bamford, K.M. and Early, M.P. (2011). Soil organic matter. *Principles of horticulture*: 184-194.

Aggarwal, S. (2004). Principles of remote sensing. *Satellite remote sensing and GIS applications in agricultural meteorology*: 23-38.

Agrawal, R.V. and Deshmukh, R.R. (2018). Develop Nondestructive Database of Seasonal Fruits and Reviews on spectral preprocessing techniques. *International Journal of Emerging Trends & Technology in Computer Science, 7*(2): 90-94.

Ait-Amir, B., Pougnet, P. and El Hami, A. (2015). Meta-Model Development. *Embedded Mechatronic Systems 2*: 151-179.

Amin, I., Fikrat, F., Mammadov, E. and Babayev, M. (2020). Soil Organic Carbon Prediction by Vis-NIR Spectroscopy: Case Study the Kur-Aras Plain, Azerbaijan. *Communications in Soil Science and Plant Analysis, 51*(6): 726-734.

Angelopoulou, T., Balafoutis, A., Zalidis, G. and Bochtis, D. (2020). From Laboratory to Proximal Sensing Spectroscopy for Soil Organic Carbon Estimation—A Review. *Sustainability, 12*(2): 443.

Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G. and Bochtis, D. (2019). Remote sensing techniques for soil organic carbon estimation: A review. *Remote Sensing, 11*(6): 676.

Aoyama, M. (2015). Functional Roles of Soil Organic Matter. *Humic Substances Research, 12*: 21-28.

Arriaga, F.J., Guzman, J. and Lowery, B. (2017). Conventional Agricultural Production Systems and Soil Functions. *Soil Health and Intensification of Agroecosytems*: 109-125.

Askari, M.S., Alamdari, P., Chahardoli, S. and Afshari, A. (2020). Quantification of heavy metal pollution for environmental assessment of soil condition. *Environmental Monitoring and Assessment, 192*(3): 162.

Astera, M. (2010). Soil CEC explained: understanding, measuring and using cation exchange capacity for nutritious crops. *Acres USA, 40*(3): 25-28.

Balasubramanian, A. (2017). Soil Forming Processes. University of Mysore: 1-10.

Bangelesa, F., Adam, E., Knight, J., Dhau, I., Ramudzuli, M. and Mokotjomela, T.M. (2020). Predicting Soil Organic Carbon Content Using Hyperspectral Remote Sensing in a Degraded Mountain Landscape in Lesotho. *Applied and Environmental Soil Science*: 1-11.

Bartholomeus, H.M., Schaepman, M.E., Kooistra, L., Stevens, A., Hoogmoed, W.B. and Spaargaren, O.S.P. (2008). Spectral reflectance based indices for soil organic carbon quantification. *Geoderma, 145*(1-2): 28-36.

Beaumont, R. (2012). An Introduction to Principal Component Analysis & Factor Analysis. *Using SPSS, 19*: 1-24.

Behera, B. K. and Prasad, R. (2020). Enhance organic matter. Chapter 4 Strategies for soil management. *Environmental Technology and Sustainability: Physical, Chemical and Biological Technologies for Clean Environmental Management.* Elsevier: 153.

Ben-Dor, E. and Banin, A. (1995). Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Science Society of America Journal*, *59*(2): 364-372.

Ben-Dor, E., Irons, J.R. and Epema, G.F. (1999). Soil reflectance. *Manual of Remote Sensing: Remote Sensing for Earth Science*: 111-187.

Ben-Dor, E., Taylor, R.G., Hill, J., Demattê, J.A.M., Whiting, M.L., Chabrillat, S. and Sommer, S. (2008). Imaging spectrometry for soil applications. *Advances in Agronomy*, *97*: 321-392.

Bhattacharyya, T. and Pal, D. (2015). The Soil: A Natural Resource. *Soil science-an introduction*, *1*: 1-19.

Bianchi, S.R., Miyazawa, M., Oliveira, E.L.D. and Pavan, M.A. (2008). Relationship between the mass of organic matter and carbon in soil. *Brazilian Archives of Biology and Technology*, *51*(2): 263-269.

Bilgili, A.V., Van Es, H.M., Akbas, F., Durak, A. and Hively, W.D. (2010). Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *Journal of Arid Environments*, *74*(2): 229-238.

Biology discussion. (n.d). Soil: Definition, Components and Role of Soil Organisms. Available at: http://www.biologydiscussion.com/soil/soil-definition-components-and-role-of-soil-organisms-with-diagram/7155. [Accessed 31 May 2019].

Birkeland, P.W. (1999). Soils and geomorphology. *Oxford University Press*: 430.

Biswas, J.C. and Naher, U.A. (2019). Soil nutrient stress and rice production in Bangladesh. *Advances in Rice Research for Abiotic Stress Tolerance*: 431-445.

Bloomenthal, A. (2020). Coefficient of Determination. Available at: https://www.investopedia.com/terms/c/coefficient-of-determination.asp [Accessed 06 October 2020].

Blume, H.P., Brümmer, G.W., Fleige, H., Horn, R., Kandeler, E., Kögel-Knabner, I., Kretzschmar, R., Stahr, K. and Wilke, B.M. (2016). Soil organic matter. *Scheffer/SchachtschabelSoil Science*: 55-86.

Bot, A. and Benites, J. (2005). The importance of soil organic matter: Key to drought-resistant soil and sustained food production. *Food and Agriculture Organisation*, *80*: 1-95.

Brady, N. and Weil, R. (2002). The Nature and Properties of Soils. *Upper Saddle River, New Jersey*.

Brady, N.C. and Weil, R.R. (2008). The nature and properties of soils. *Upper Saddle River, NJ: Prentice Hall*, *14*: 662-710.

Brevik, E.C. and Arnold, R.W. (2015). Is the traditional pedologic definition of soil meaningful in the modern context?. *Soil Horizons*, *56*(3): 1-8.

Buddenbaum, H. and Steffens, M. (2012). The effects of spectral pretreatments on chemometric analyses of soil profiles using laboratory imaging spectroscopy. *Applied and Environmental Soil Science*: 1-12.

Bullock, P. (2005). Climate change impacts. *Encyclopaedia of Soils in the Environment*: 254-262.

Campbell, J.B. and Wynne, R.H. (2011). Introduction to remote sensing. *Guilford Press*: 1-10.

Castaldi, F., Chabrillat, S., Don, A. and van Wesemael, B. (2019a). Soil organic carbon mapping using LUCAS topsoil database and Sentinel-2 data: An approach to reduce soil moisture and crop residue effects. *Remote Sensing*, *11*(18): 2121.

Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M. and van Wesemael, B. (2019b). Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. *ISPRS Journal of Photogrammetry and Remote Sensing*, *147*: 267-282.

Castaldi, F., Palombo, A., Santini, F., Pascucci, S., Pignatti, S. and Casa, R. (2016). Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. *Remote Sensing of Environment*, *179*: 54-65.

Castillo, E., Iglesias, A. and Ruiz-Cobo, R. (2005). Chapter 9- Functional Networks. *Functional equations in applied sciences*: 169-232.

Chai, T. and Draxler, R.R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development*, *7*(3): 1247-1250.

Chatterjee, A., Lal, R., Wielopolski, L., Martin, M.Z. and Ebinger, M.H. (2009). Evaluation of different soil carbon determination methods. *Critical Reviews in Plant Science*, *28*(3): 164-178.

Chenu, C., Rumpel, C., Lehmann, J. and Paul, E.A. (2015). Methods for studying soil organic matter: nature, dynamics, spatial accessibility, and interactions with minerals. *Soil microbiology, ecology and biochemistry*: 383-419.

Cherubin, M.R., Oliveira, D.M.D.S., Feigl, B.J., Pimentel, L.G., Lisboa, I.P., Gmach, M.R., Varanda, L.L., Morais, M.C., Satiro, L.S., Popin, G.V. and Paiva, S.R.D. (2018). Crop residue harvest for bioenergy production and its implications on soil functioning and plant growth: A review. *Scientia Agricola*, *75*(3): 255-272.

Cierniewski, J. and Kuśnierek, K. (2010). Influence of several size properties on soil surface reflectance. *Quaestiones Geographicae*, *29*(1): 13-25.

Dalgliesh, N.P. and Foale, M.A. (1998). Soil matters: monitoring soil water and nutrients in dryland farming.

Daniel, K., Tripathi, N.K., Honda, K. and Apisit, E. (2001). Analysis of spectral reflectance and absorption patterns of soil organic matter. *The proceedings of the 22nd Asian Conference on Remote Sensing*: 1-6.

Danner, M., Locherer, M., Hank, T. and Richter, K. (2015). Spectral sampling with the ASD FieldSpec 4–theory, measurement, problems, interpretation. *EnMAP Field Guides Technical Report, GFZ Data*: 1-25.

Danoedoro, P. and Zukhrufiyati, A. (2015). Integrating spectral indices and geostatistics based on Landsat-8 imagery for surface clay content mapping in Gunung Kidul area, Yogyakarta, Indonesia. *Proceeding of the 36th Asian Conference on Remote Sensing Fostering Resilient Growth*: 1-9.

Davis, R. and Luttrell, M. (2016). Bulk earthworks. Available at: https://www.mla.com.au/globalassets/mla-corporate/research-and-development/program-areas/feeding-finishing-and-nutrition/feedlot-design-manual/08-earthworks-2016_04_01.pdf [Accessed 28 May 2020].

DeGomez, T., Kolb, P. and Kleinman, S. (2015). Basic Soil Components. Available at: https://articles.extension.org/pages/54401/basic-soil-components [Accessed 31 May 2019].

Dobos, E., Carré, F., Hengl, T., Reuter, H.I. and Tóth, G. (2006). Digital soil mapping: as a support to production of functional maps. Office for Official Publication of the European Communities.

Donovan, P. (2013). Measuring soil carbon change, a flexible, practical local method. 1-61.

Dou, X., Wang, X., Liu, H., Zhang, X., Meng, L., Pan, Y., Yu, Z. and Cui, Y. (2019). Prediction of soil organic matter using multi-temporal satellite images in the Songnen Plain, China. *Geoderma*, *356*: 113896.

Doula, M.K. and Sarris, A. (2016). Soil Environment. *Environment and development*: 213-286.

Du Preez, C.C., Van Huyssteen, C.W. and Mnkeni, P.N. (2011). Land use and soil organic matter in South Africa 2: A review on the influence of arable crop production. *South African Journal of Science*, *107*(5-6): 35-42.

Duarte, R.M.B.O. and Duarte, A.C. (2019). Geochemistry Soil, Organic Components. *Encyclopedia of Analytical Science*, *1*(3): 329-339.

Durdevic, B., Jug, I., Jug, D., Bogunovic, I., Vukadinovic, V., Stipesevic, B. and Brozovic, B. (2019). Spatial variability of soil organic matter content in Eastern Croatia assessed using different interpolation methods. *International Agrophysics*, *33*(1): 31-39.

Earth eclipse. (2018). What is the earth?. Available at: https://www.eartheclipse.com/geography/4-different-spheres-of-earth.html. [Accessed 26 May 2019].

Encyclopaedia Britannica. (2019). Soils. Available at: https://www.britannica.com/place/South-Africa/Resources-and-power [Accessed 20 June 2019].

Environmental Protection Agency (EPA). (2020). Agricultural Pasture, Rangeland and Grazing. Available at: https://www.epa.gov/agriculture/agricultural-pasture-rangeland-and-grazing [Accessed 30 May 2020].

Erich, S. (2001). Soil, Physical Characteristics Of. Available at: https://www.encyclopedia.com/science/news-wires-white-papers-and-books/soil-physical-characteristics [Accessed 27 May 2019].

European Space Agency (ESA). (2015). Sentinel -2 User Handbook. Available at: https://earth.esa.int/documents/247904/685211/Sentinel-2_User_Handbook [Accessed 19 July 2020].

ESA. (2020a). Sentinel-2. Available at: https://sentinel.esa.int/web/sentinel/missions/sentinel-2 [Accessed 16 July 2020].

ESA. (2020b). Product types. Available at: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/product-types [Accessed 19 July 2020].

Escadafal, R. (1989). Remote sensing of arid soil surface color with Landsat thematic mapper. *Advances in space research*, *9*(1): 159-163.

Escribano, P., Schmid, T., Chabrillat, S., Rodríguez-Caballero, E. and Garcia, M. (2017). Optical remote sensing for soil mapping and monitoring. *Soil mapping and process modeling for sustainable land use management*: 87-125.

Fageria, N.K. (2012). Role of soil organic matter in maintaining sustainability of cropping systems. *Communications in Soil Science and Plant Analysis*, *43*(16): 2063-2113.

Fang, Q., Hong, H., Zhao, L., Kukolich, S., Yin, K. and Wang, C. (2018). Visible and near-infrared reflectance spectroscopy for investigating soil mineralogy: A review. *Journal of Spectroscopy*: 1-14.

Food and Agriculture Organization (FAO). (2015a). Revised World Soil Charter. *Rome: Food and Agriculture Organization of the United Nations*: 2.

FAO. (2015b). International Year of Soil Conference. Available at: http://www.fao.org/soils-2015/events/detail/en/c/338738/ [Accessed 28 March 2020].

FAO. (2017a). Soil organic carbon the hidden potential. *Food and Agriculture Organization of the United Nations Rome, Italy*: 1-90.

FAO. (2017b). GSP Guidelines for sharing national data/information to compile a Global Soil Organic Carbon (GSOC) map. *Food and Agriculture Organization of the United Nations*: 1-25.

FAO. (2019). Biological Properties. Available at: http://www.fao.org/soils-portal/soil-survey/soil-properties/biological-properties/en/ [Accessed 29 May 2019].

FAO. (2020). What is soil ?. Available at: http://www.fao.org/soils-portal/about/all-definitions/en/ [Accessed 16 March 2020].

Farifteh, J., Van der Meer, F., Atzberger, C. and Carranza, E.J.M. (2007). Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). *Remote Sensing of Environment*, *110*(1): 59-78.

Feng, Y. and Balkcom, K.S. (2017). Nutrient cycling and soil biology in row crop systems under intensive tillage. *Soil Health and Intensification of Agroecosytems*: 231-255.

Fey, M. (2010). Soils of South Africa. *Cambridge University Press*.

Finch, S., Samuel, A. and Lane, G.P. (2014). Soils and Soil management. *Lockhart and wiseman's crop husbandry including grassland, 9*: 37-62.

Foster, S., Schultz, B., McCuin, G., Neibling, N., Shewmaker, G. (2013). Soil Properties, Part 1 of 3: Physical Characteristics. *University of Nevada: Cooperative Extension*: 1-7.

Francioli, D., van Rijssel, S.Q., van Ruijven, J., Termorshuizen, A.J., Cotton, T.A., Dumbrell, A.J., Raaijmakers, J.M., Weigelt, A. and Mommer, L. (2020). Plant functional group drives the community structure of saprophytic fungi in a grassland biodiversity experiment. *Plant and Soil*: 1-15.

Frith, A. (2011). Okhahlamba. Available at: https://census2011.adrianfrith.com/place/571 [Accessed 23 June 2019].

Fu, W., Ma, J., Chen, P. and Chen, F. (2020). Remote Sensing Satellites for Digital Earth. *Manual of Digital Earth*: 55-123.

Gahlod, N.S.S., Jaryal, N., Roodagi, M., Dhale, S.A., Kumar, D. and Kulkarni, R. (2019). Soil organic carbon stocks assessment in Uttarakhand State using remote sensing and GIS technique. *International Journal of Current Microbiology and Applied Sciences*, *8*: 1646-1658.

Gallo, B.C., Demattê, J.A., Rizzo, R., Safanelli, J.L., Mendes, W.D.S., Lepsch, I.F., Sato, M.V., Romero, D.J. and Lacerda, M.P. (2018). Multi-temporal satellite images on topsoil attribute quantification and the relationship with soil classes and geology. *Remote Sensing*, *10*(10): 1571.

Gao, L., Zhu, X., Han, Z., Wang, L., Zhao, G. and Jiang, Y. (2017). Spectroscopy Based Estimation of Soil Organic Matter in Brown-Forest Areas of the Shandong Peninsula, China. *Pedosphere*: 1-12.

Geiger, B.C. and Kubin, G. (2012). Relative information loss in the PCA. *2012 IEEE Information Theory Workshop*: 562-566.

Gerba, C.P. and Choi, C.Y. (2006). Role of irrigation water in crop contamination by viruses. Viruses in foods: 257-263.

Gholizadeh, A., Borůvka, L., Saberioon, M. and Vašát, R. (2013). Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: State-of-the-art and key issues. *Applied spectroscopy*, *67*(12): 1349-1362.

Gholizadeh, A., Žižala, D., Saberioon, M. and Borůvka, L. (2018). Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sensing of Environment, 218*: 89-103.

Gia Pham, T., Kappas, M., Van Huynh, C. and Hoang Khanh Nguyen, L. (2019). Application of ordinary kriging and regression kriging method for soil properties mapping in hilly region of Central Vietnam. *ISPRS International Journal of Geo-Information*, *8*(3): 147.

Gitelson, A.A., Kaufman, Y.J. and Merzlyak, M.N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote sensing of Environment*, *58*(3): 289-298.

Glinski, J. (2011). Soil Phases. *Encyclopaedia of Agrophysics*: 264-267.

Gniazdowski, Z. (2017). New interpretation of principal components analysis. *Zeszyty Naukow, Warsaw School of Computer Science*, *16*(11): 43-65.

Gobin, A. (2000). Participatory and spatial-modelling methods for land resources analysis. Doctoral dissertation, Katholieke Universiteit Leuven.

Golicz, K., Hallett, S.H., Sakrabani, R. and Pan, G. (2019). The potential for using smartphones as portable soil nutrient analyzers on suburban farms in central East China. Scientific reports, 9(1): 1-10.

Gomez, C., Adeline, K., Bacha, S., Driessen, B., Gorretta, N., Lagacherie, P., Roger, J.M. and Briottet, X. (2018). Sensitivity of clay content prediction to spectral configuration of VNIR/SWIR imaging data, from multispectral to hyperspectral scenarios. *Remote Sensing of Environment*, *204*: 18-30.

Google Earth. (2020). Emakhosaneni. Available at: https://earth.google.com/web/@-28.79943144,29.23423194,1272.68696279a,3832.42087508d,35y,0h,0t,0r/data=ChUaEwoLL2cvMXRkM2NmYnQYAiABKAI [Accessed 12 May 2020].

Goovaerts, P. (1997). Geostatistics for Natural Resources Evaluation (Applied Geostatistics). Oxford University Press, New York: 496.

Guo, L., Luo, M., Zhangyang, C., Zeng, C., Wang, S. and Zhang, H. (2018). Spatial modelling of soil organic carbon stocks with combined principal component analysis and geographically weighted regression. *The Journal of Agricultural Science*, *156*(6): 774-784.

Hajar, A. (2020). Soil Building: an investment in your garden. Available at: https://cedarmillnews.com/article/soil-building-an-investment-in-your-garden/ [Accessed 11 June 2021].

Haynes, R.J., Dominy, C.S. and Graham, M.H. (2003). Effect of agricultural land use on soil organic matter status and the composition of earthworm communities in KwaZulu-Natal, South Africa. *Agriculture, Ecosystems & Environment*, *95*(2-3): 453-464.

He, T., Wang, J., Lin, Z. and Cheng, Y. (2009). Spectral features of soil organic matter. *Geo-spatial Information Science*, *12*(1): 33-40.

Hengl, T. (2009). A practical guide to geostatistical mapping. *Amsterdam: University of Amsterdam*, *52*: 1-293.

Hengl, T., Heuvelink, G.B. and Rossiter, D.G. (2007). About regression-kriging: From equations to case studies. *Computers & geosciences*, *33*(10): 1301-1315.

Hengl, T., Heuvelink, G.B. and Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, *120*(1-2): 75-93.

Hong, Y., Yu, L., Chen, Y., Liu, Y., Liu, Y., Liu, Y. and Cheng, H. (2018a). Prediction of soil organic matter by VIS–NIR spectroscopy using normalized soil moisture index as a proxy of soil moisture. *Remote Sensing*, *10*(1): 28.

Hong, Y., Chen, S., Zhang, Y., Chen, Y., Yu, L., Liu, Y., Liu, Y., Cheng, H. and Liu, Y. (2018b). Rapid identification of soil organic matter level via visible and near-infrared spectroscopy: Effects of two-dimensional correlation coefficient and extreme learning machine. *Science of the Total Environment*, *644*: 1232-1243.

Hoorman, J.J. and Islam, R. (2010). Understanding soil microbes and nutrient recycling. *Agriculture and natural resources*: 1-5.

Hou, C., Li, Y., Ma, J. and Zhang, X. (2019). Early-stage decomposition of maize litter at different positions in a semi-arid cropland. *Archives of Agronomy and Soil Science*: 1-11.

Hoyle, F. (2013). Managing soil organic matter: a practical guide. *Grains research and Development Corporation*: 1-110.

Huete, A. (1988). Huete, AR A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment. *Remote sensing of environment*, *25*: 295-309.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X. and Ferreira, L.G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of environment*, *83*(1-2): 195-213.

Huete, A.R. (2004). Remote sensing for environmental monitoring. *Environmental Monitoring and Characterization*: 183-206.

Hummel, J.W., Sudduth, K.A. and Hollinger, S.E. (2001). Soil moisture and organic matter prediction of surface and subsurface soils using an NIR soil sensor. *Computers and electronics in agriculture*, *32*(2): 149-165.

Idowu, J. and Angadi, S. (2013). Understanding and managing soil compaction in agricultural fields. *NM State University, Cooperative Extension Service*: 1-8.

Ismail-Meyer, K., Stolt, M.H. and Lindbo, D.L. (2018). Soil organic matter. *Interpretation of Micromorphological Features of Soils and Regoliths*: 471-512.

Ivezić, V., Kraljević, D., Lončarić, Z., Engler, M., Kerovec, D., Zebec, V. and Jović, J. (2016). Organic matter determined by loss on ignition and potassium dichromate method. *51st Croatian and 11th International Symposium on Agriculture, Opatija, Croatia*, *36*: 36-40.

Jacobson, M., Charlson, R.J., Rodhe, H. and Orians, G.H. (2000). Part Two - Properties of and Transfers between the Key Reservoirs. *International Geophysics*, *72*: 107-108.

Jacoby, R., Peukert, M., Succurro, A., Koprivova, A. and Kopriva, S. (2017). The role of soil microorganisms in plant mineral nutrition—current knowledge and future directions. *Frontiers in plant science*, *8*: 1617.

Jenny, H. (1994). Factors of soil formation: a system of quantitative pedology. Courier Corporation: 1-191.

Jobbágy, E.G. and Jackson, R.B. (2000). The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological applications*, *10*(2): 423-436.

Johnson, R.A. and Wichern, D.W. (2002). Applied multivariate statistical analysis. Upper Saddle River, NJ: Prentice hall: 1-581.

Jordan, C.F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, *50*(4): 663-666.

Kafle, G. and Balla, M.K. (2008). Effectiveness of Root System of Grasses Used in Soil Conservation in Paundi Khola Sub Watershed of Lamjung District, Nepal. *The Initiation*, *2*(1): 121-129.

Kalev, S.D. and Toor, G.S. (2018). The Composition of Soils and Sediments. *Green Chemistry*. Elsevier: 339-357.

Kamota, A., Muchaonyerwa, P. and Mnkeni, P.N.S. (2014). Decomposition of surface-applied and soil-incorporated Bt maize leaf litter and Cry1Ab protein during winter fallow in South Africa. *Pedosphere*, *24*(2): 251-257.

Karlen, D.L. and Rice, C.W. (2015). Soil degradation: Will humankind ever learn?. *Sustainability*, 7:12490-12501.

Kooistra, L., Wehrens, R., Leuven, R.S.E.W. and Buydens, L.M.C. (2001). Possibilities of visible–near-infrared spectroscopy for the assessment of soil contamination in river floodplains. *Analytica Chimica Acta*, *446*(1-2): 97-105.

Krishan, G., Saha, S.K. and Patel, S.K.N. (2014). Remote sensing in soil fertility evaluation and management. *Bioresources for sustainable plant nutrient management. New Delhi: Satish Serial Publishing House*: 509-533.

Krishnan, P., Alexander, J.D., Butler, B.J. and Hummel, J.W. (1980). Reflectance technique for predicting soil organic matter 1. *Soil Science Society of America Journal*, *44*(6): 1282-1285.

Kuang, B. and Mouazen, A.M. (2012). Influence of the number of samples on prediction error of visible and near infrared spectroscopy of selected soil properties at the farm scale. *European Journal of Soil Science*, *63*(3): 421-429.

Kuhn, L.W. and Edge, W.D. (2000). Controlling moles. Available at: https://www.oregon.gov/ODA/shared/Documents/Publications/PesticidesPARC/ControllingMolesOSU.pdf [Accessed 02 June 2020].

Kumar, A., Kuzyakov, Y. and Pausch, J. (2016). Maize rhizosphere priming: field estimates using 13 C natural abundance. *Plant and Soil*, *409*(1-2):87-97.

Kumar, N., Velmurugan, A., Hamm, N.A. and Dadhwal, V.K. (2018). Geospatial mapping of soil organic carbon using regression kriging and remote sensing. *Journal of the Indian Society of Remote Sensing*, *46*(5): 705-716.

Kumar, N.S., Anouncia, S.M. and Prabu, M. (2013). Application of Satellite Remote Sensing to find Soil Fertilization by using Soil Colour: study area Vellore District, Tamil Nadu, India. *International Journal of Online and Biomedical Engineering, 9(2)*: 44-49.

Laker, M.C. and Nortjé, G.P. (2019). Review of existing knowledge on soil crusting in South Africa. *Advances in Agronomy*: 189-242.

Lal, R. (2009). Use of crop residues in the production of biofuel. *Handbook of Waste Management and Co-Product Recovery in Food Processing*: 455-478.

Lal, R., Kimble, J.M., Follett, R.F. and Stewart, B.A. (1997). Pedospheric processes and the carbon cycle. *Soil processes and the carbon cycle, 11*: 1-8.

Le Roux, J. (2014). Soil erosion in South Africa - its nature and distribution. Available at: https://www.grainsa.co.za/soil-erosion-in-south-africa---its-nature-and-distribution [Accessed 30 May 2020].

Lefevre, C., Rekik, F., Alcantara, V. and Wiese, L. (2017). Soil organic carbon: the hidden potential. *Food and Agriculture Organization of the United Nations (FAO)*: 1-90.

Lefsky, M.A. and Cohen, W.B. (2003). Selection of remotely sensed data. *Remote sensing of forest environments*: 13-46.

Li, N., You, M.Y., Zhang, B., Han, X.Z., Panakoulia, S.K., Yuan, Y.R., Liu, K., Qiao, Y.F., Zou, W.X., Nikolaidis, N.P. and Banwart, S.A. (2017). Modeling soil aggregation at the early pedogenesis stage from the parent material of a Mollisol under different agricultural practices. *Advances in Agronomy*, *142*: 181-214.

Li, S., Shi, Z., Chen, S., Ji, W., Zhou, L., Yu, W. and Webster, R. (2015). In situ measurements of organic carbon in soil profiles using vis-NIR spectroscopy on the Qinghai–Tibet plateau. *Environmental Science & Technology*, *49*(8): 4980-4987.

Li, X., Ren, J., Zhao, K. and Liang, Z. (2019). Correlation between spectral characteristics and physicochemical parameters of soda-saline soils in different states. *Remote Sensing*, *11*(4): 388.

Li, Y. (2010). Can the spatial prediction of soil organic matter contents at various sampling scales be improved by using regression kriging with auxiliary information?. *Geoderma*, *159*(1-2): 63-75.

Liu, H., Zhang, Y. and Zhang, B. (2009). Novel hyperspectral reflectance models for estimating black-soil organic matter in Northeast China. *Environmental monitoring and assessment*, *154*(1-4): 147.

Liu, H., Zhao, C., Wang, J., Huang, W. and Zhang, X. (2011). Soil organic matter predicting with remote sensing image in typical blacksoil area of Northeast China. *Transactions of the Chinese Society of Agricultural Engineering*, *27*(8): 211-215.

Liu, H.J., Zhang, B., Zhang, Y.Z., Song, K.S., Wang, Z.M., Li, F. and Hu, M.G. (2008). Soil taxonomy on the basis of reflectance spectral characteristics. *Guang pu xue yu guang pu fen xi= Guang pu*, *28*(3): 624-628.

Liu, S., An, N., Yang, J., Dong, S., Wang, C. and Yin, Y. (2015). Prediction of soil organic matter variability associated with different land use types in mountainous landscape in southwestern Yunnan province, China. *Catena*, *133*: 137-144.

Lorenzo-Seva, U. (2013). How to report the percentage of explained common variance in exploratory factor analysis. *Unpublished manuscript*. Department of Psychology, Universitat Rovira i Virgili, Tarragona: 1-13.

Luce, M.S., Ziadi, N., Zebarth, B.J., Grant, C.A., Tremblay, G.F. and Gregorich, E.G. (2014). Rapid determination of soil organic matter quality indicators using visible near infrared reflectance spectroscopy. *Geoderma*, *232*: 449-458.

Mallick, J., Ahmed, M., Alqadhi, S.D., Falqi, I.I., Parayangat, M., Singh, C.K., Rahman, A. and Ijyas, T. (2020). Spatial stochastic model for predicting soil organic matter using remote sensing data. *Geocarto International*: 1-32.

Mandal, U.K. (2016). Spectral color indices based geospatial modeling of soil organic matter in chitwan district, Nepal. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *41*.

Mani, P.K. (2011). Manure For Vegetable Production. *Fundamentals of vegetable production*: 544.

Martens, H. and Martens, M. (2000). Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food quality and preference*, *11*(1-2): 5-16.

Martens, H. and Naes, T. (1989). Assessment, validation and choice of calibration method. *Multivariate calibration*: 237-266.

Matus, F.J., Escudey, M., Förster, J.E., Gutiérrez, M. and Chang, A.C. (2009). Is the Walkley–Black method suitable for organic carbon determination in Chilean volcanic soils?. *Communications in soil science and plant analysis*, *40*(11-12): 1862-1872.

Maynard, J.J., Dahlgren, R.A. and O'Geen, A.T. (2011). Soil carbon cycling and sequestration in a seasonally saturated wetland receiving agricultural runoff. *Biogeosciences*, *8*(11): 3391.

McCauley, A., Jones, Cl., Jacobsen, J. (2005). Basic soil properties. *Montana State University*: 1-12

McClaugherty, C and Berg, B. (2011). Soils and decomposition. *Encyclopaedia of life sciences*, Macmillan Publishers: 1-12.

McGranahan, D.A., Daigh, A.L., Veenstra, J.J., Engle, D.M., Miller, J.R. and Debinski, D.M. (2014). Connecting soil organic carbon and root biomass with land-use and vegetation in temperate grassland. *The Scientific World Journal*: 1-9.

McKenzie, R.H. (2010). Agricultural soil compaction: Causes and management. Available at: https://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/agdex13331/$file/510-1.pdf?OpenElement [Accessed 28 May 2020].

McNear Jr, D.H. (2013). The rhizosphere-roots, soil and everything in between. *Nature Education Knowledge*, *4*(3): 1.

Miller, C.E. (2001). Chemical principles of near-infrared technology. *Near-infrared technology in the agricultural and food industries*, *2*: 19-37.

Milton, E.J. (2004). Field spectroscopy. *Geoinformatics, Encyclopaedia of Life Support Systems (EOLSS)*.

Mirzaee, S., Ghorbani-Dashtaki, S., Mohammadi, J., Asadi, H. and Asadzadeh, F. (2016). Spatial variability of soil organic matter using remote sensing data. *Catena*, *145*: 118-127.

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

Mohamed, N. (2000). Greening land and agrarian reform: a case for sustainable agriculture. *At the crossroads: land and agrarian reform in South Africa into the 21st century*: 163-175.

Mondal, B.P. (2018) Hyper-spectral analysis of soil properties for soil management. *Advances in Agriculture for Sustainable Development*: 59-65.

Montgomery, D.R., Zabowski, D., Ugolini, F.C., Hallberg, R.O. and Spaltenstein, H. (2000). Soils, Watershed Processes, and Marine. *Earth System Science: From Biogeochemical Cycles to Global Changes*, *72*: 159-194.

Motaghian, H.R. and Mohammadi, J. (2011). Spatial estimation of saturated hydraulic conductivity from terrain attributes using regression, kriging, and artificial neural networks. *Pedosphere*, *21*(2): 170-177.

Motsara, M.R. and Roy, R.N. (2008). Guide to Laboratory establishment for plant nutrient analysis. *FAO Fertilizer and Plant nutrition bulletin. Food and Agriculture Organization, Rome:* 1-219.

Mouazen, A.M., Kuang, B., De Baerdemaeker, J. and Ramon, H. (2010). Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma*, *158*(1-2): 23-31.

Mujumdar, P.P. and Kumar, D.N. (2012). Remote sensing for hydrologic modelling. *Floods in a changing climate: hydrologic modeling*. Cambridge University Press: 90.

Mulder, V.L., De Bruin, S., Schaepman, M.E. and Mayr, T.R. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma*, *162*(1-2): 1-19.

Munna, N.M. (2017). Definition of Soil: The Composition of Soil. Available at: https://www.earthreview.org/soil-composition/ [Accessed 31 May 2019].

Mutanga, O., Dube, T. and Ahmed, F., (2016). Progress in remote sensing: vegetation monitoring in South Africa. *South African Geographical Journal*, *98*(3): 461-471.

Mutuo, P.K., Shepherd, K.D., Albrecht, A. and Cadisch, G. (2006). Prediction of carbon mineralization rates from different soil physical fractions using diffuse reflectance spectroscopy. *Soil Biology and Biochemistry*, *38*(7): 1658-1664.

Myers, J.S. and Miller, R.L. (2007). Optical airborne remote sensing. *Remote Sensing of Coastal Aquatic Environments*: 51-67.

Natural Resources Conservation Service. (2014). Soil physical and chemical properties. Available at: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/nj/home/?cid=nrcs141p2_018993 [Accessed 27 May 2019].

Navalgund, R.R., Jayaraman, V. and Roy, P.S. (2007). Remote sensing applications: An overview. *Current Science (00113891)*, *93*(12): 1747-1766.

Nawar, S., Buddenbaum, H., Hill, J., Kozak, J. and Mouazen, A.M. (2016). Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. *Soil and Tillage Research*, *155*: 510-522.

Nawaz, M.F., Bourrie, G. and Trolard, F. (2013). Soil compaction impact and modelling. A review. *Agronomy for sustainable development*, *33*(2): 291-309.

Neill, S.P. and Hashemi, M.R. (2018). Chapter 8-ocean modelling for resource characterization. *Fundamentals of Ocean Renewable Energy*: 193-235.

Nellis, M.D. and Briggs, J.M. (1992). Transformed vegetation index for measuring spatial variation in drought impacted biomass on Konza Prairie, Kansas. *Transactions of the Kansas Academy of Science (1903)*: 93-99.

Nimmo, J.R. (2005). Aggregation: physical aspects. *Encyclopaedia of soils in the environment. Elsevier*: 28-35.

Nortcliff, S., Hulpke, H., Bannick, C.G., Terytze, K., Knoop, G., Bredemeier, M. and Schulte-Bisping, H. (2000). Soil, 1. Definition, function, and utilization of soil. *Ullmann's Encyclopedia of Industrial Chemistry*, *30*: 399-420.

Nortcliff, S., Hulpke, H., Bannick, C.G., Terytze, K., Knoop, G., Bredemeier, M. and Schulte-Bisping, H. (2006). Soil, 1. Definition, function, and utilization of soil. *Ullmann's Encyclopedia of Industrial Chemistry*: 1-22.

Nowatzki, J. (2019). Air Seeders for Conservation Tillage Crop Production. *Handbook of Farm, Dairy and Food Machinery Engineering*: 133-148.

Oliveira, J.F., Brossard, M., Vendrame, P.R.S., Mayi III, S., Corazza, E.J., Marchão, R.L. and de Fátima Guimarães, M. (2013). Soil discrimination using diffuse reflectance Vis–NIR spectroscopy in a local toposequence. *Comptes Rendus Geoscience*, *345*(11-12): 446-453.

Ongley, E.D. (1996). Chapter 1: Introduction to agricultural water pollution. Control of water pollution from agriculture. Food and Agriculture Organization

Osman, K.T. (2012a). Physical Properties of Soil. *Springer, Dordrecht*: 49-65.

Osman, K.T. (2012b). Soil Organic Matter. *Soils*: 89-96.

Osman, K.T. (2012c). Forest soils. *Soils*: 229-251.

Osman, K.T. (2013). Soil as a part of the lithosphere. *Soils*: 9-16.

Osman, K.T. (2014a). Soil erosion by water. *Soil degradation, conservation and remediation*: 69-101.

Osman, K.T. (2014b). Wind erosion. *Soil degradation, conservation and remediation*: 103-123.

Parchami-Araghi, F., Mirlatifi, S.M., Dashtaki, S.G. and Mahdian, M.H. (2013). Point estimation of soil water infiltration process using Artificial Neural Networks for some calcareous soils. *Journal of hydrology*, *481*: 35-47.

Paterson, G., Turner, D., Wiese, L., Van Zijl, G., Clarke, C. and Van Tol, J. (2015). Spatial soil information in South Africa: Situational analysis, limitations and challenges. *South African Journal of Science, 111(5-6)*: 1-7.

Peddle, D.R., White, H.P., Soffer, R.J., Miller, J.R. and Ledrew, E.F. (2001). Reflectance processing of remote sensing spectroradiometer data. *Computers & geosciences*, *27*(2): 203-213.

Peng, J., Zhou, Q., Zhang, Y. and Xiang, H. (2013). Effect of soil organic matter on spectral characteristics of soil. *Acta Pedologica Sinica*, *50*(3): 517-524.

Phuong, T.T., Shrestha, R.P. and Chuong, H.V. (2017). Simulation of Soil Erosion Risk in the Upstream Area of Bo River Watershed. *Redefining Diversity & Dynamics of Natural Resources Management in Asia, 3*: 87-99.

Pietikäinen, J., Pettersson, M. and Bååth, E. (2005). Comparison of temperature effects on soil respiration and bacterial and fungal growth rates. *FEMS microbiology ecology*, *52*(1): 49-58.

Pimentel, D., Huang, X., Cordova, A. and Pimentel, M. (1997). Impact of population growth on food supplies and environment. *Population and Environment*, *19*(1): 9-14.

Pouget, M., Madeira, J., Le Floch, E. and Kamal, S. (1990). Caracteristiques spectrales des surfaces sableuses de la region cotiere nord-ouest de l'Egypte: application aux donnees satellitaires SPOT: 27-39.

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H. and Sorooshian, S. (1994a). A modified soil adjusted vegetation index. *Remote sensing of environment*, *48*(2): 119-126.

Qi, J., Kerr, Y. and Chehbouni, A. (1994b). External factor consideration in vegetation index development. *Proc. of Physical Measurements and Signatures in Remote Sensing, ISPRS, 723*: 730.

Qiao, X.X., Wang, C., Feng, M.C., Yang, W.D., Ding, G.W., Sun, H., Liang, Z.Y. and Shi, C.C. (2017). Hyperspectral estimation of soil organic matter based on different spectral preprocessing techniques. *Spectroscopy Letters*, *50*(3): 156-163.

Rai, P.K., Rai, A. and Singh, S. (2018). Change in soil microbial biomass along a rural-urban gradient in Varanasi (UP, India). *Geology, Ecology, and Landscapes*, *2*(1): 15-21.

Ramesh, T., Bolan, N.S., Kirkham, M.B., Wijesekara, H., Kanchikerimath, M., Rao, C.S., Sandeep, S., Rinklebe, J., Ok, Y.S., Choudhury, B.U. and Wang, H. (2019). Soil organic carbon dynamics: Impact of land use changes and management practices: A review. *Advances in Agronomy*, *156*: 1-107.

Ravisankar, T and Sreenivas, K. (2011). Soils and land degradation. *Remote Sensing Applications. National Remote Sensing Centre*: 81-109.

Reddy, P.P. (2016). Chapter 11: Soil Organic matter. *Sustainable Intensification of Crop Production*: 157-173.

Richardson, J. (2016). What is leaching?. Critical zone observations. Available at: https://criticalzone.org/national/blogs/post/what-is-leaching/ [Accessed 15 April 2020].

Roper, W.R., Robarge, W.P., Osmond, D.L. and Heitman, J.L. (2019). Comparing four methods of measuring soil organic matter in North Carolina soils. *Soil Science Society of America Journal*, *83*(2): 466-474.

Rossel, R.V. and Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma*, *158*(1-2): 46-54.

Rother, H.A., Hall, R. and London, L. (2008). Pesticide use among emerging farmers in South Africa: contributing factors and stakeholder perspectives. *Development Southern Africa*, *25*(4): 399-424.

Roujean, J.L. and Breon, F.M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote sensing of Environment*, *51*(3): 375-384.

Rouse, J.W., Haas, R.H., Schell, J.A. and Deering, D.W. (1974). Monitoring Vegetation Systems in the Great Plains with ERTS Proceeding. Third Earth Reserves Technology Satellite Symposium, Greenbelt: NASA SP-351, 30103017.

Roy, D.P., Li, J., Zhang, H.K. and Yan, L. (2016). Best practices for the reprojection and resampling of Sentinel-2 Multi Spectral Instrument Level 1C data. *Remote Sensing Letters*, *7*(11): 1023-1032.

Roy, P.S., Behera, M.D. and Srivastav, S.K. (2017). Satellite remote sensing: sensors, applications and techniques. *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, *87*: 465-472.

Salkind, N.J. (2010). Factor loadings. *Encyclopedia of research design* (Vol. 1). Sage.

Samuels, P. (2016). Advice on exploratory factor analysis. Centre for Academic Success, Birmingham City University: 1-8.

Savitzky, A. and Golay, M.J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, *36*(8): 1627-1639.

Schreier, H., Wiart, R. and Smith, S. (1988). Quantifying organic matter degradation in agricultural fields using PC-based image analysis. *Journal of soil and water conservation*, *43*(5): 421-424.

Schulte, E.E. and Hoskins, B. (1995). Chapter 8: Recommended soil organic matter tests. *Recommended Soil Testing Procedures for the Northeastern United States. Northeastern Regional Pub*, (493): 52-60.

Schulte, E.E., Walsh, L.M., Kelling, K.A., Bundy, L.G., Bland, W.L., Wolkowski, J.B. (2005). Management of Wisconsin Soils. *Madison, WI: Cooperative Extension Publishing, University of Wisconsin*.

Sehmi, N.S. and Kundzewicz, Z.W. (1997). Water, drought and desertification in Africa. *IAHS Publications-Series of Proceedings and Reports-Intern Assoc Hydrological Sciences*, *240*: 7-66.

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

Sharma, K.D. (1997). Assessing the impact of overgrazing on soil erosion in arid regions at a range of spatial scales. *IAHS Publication*, *245*: 119-123.

Shepherd, K.D. and Walsh, M.G. (2002). Development of reflectance spectral libraries for characterization of soil properties. *Soil science society of America journal*, *66*(3): 988-998.

Silleos, N.G., Alexandridis, T.K., Gitas, I.Z. and Perakis, K. (2006). Vegetation indices: advances made in biomass estimation and vegetation monitoring in the last 30 years. *Geocarto International*, *21*(4): 21-28.

Silva, N., Pereira, M.G., Fernandes, J.C.F. and Corrêa, N. (2016). Aggregate formation and soil organic matter under different vegetation types in Atlantic Forest from Southeastern Brazil. *Semina: Ciências Agrárias (Londrina)*, *37*(6): 3927-3939.

Simpson, R.T. (2010). Soil organic matter and aggregate dynamics in an Arctic ecosystem. [*Unpublished Doctoral thesis*], Colorado State University: 1-234

Sivakumar, M.V.K., Roy, P.S., Harmsen, K. and Saha, S.K. (2004). Satellite remote sensing and GIS applications in agricultural meteorology. *Proceedings of the Training Workshop in Dehradun, India. AGM-8, WMO/TD, 1182*: 1-423.

Smagin, A.V. and Scalenghe, R. (2006). Soil phases: the gaseous phase. Soils: Basic Concepts and Future Challenges. *Cambridge: Cambridge University Press*: 75–90.

Soil classification working group. (1991). Soil Classification: A Taxonomic System for South Africa. *Natural Agriculture. Resources for South Africa*: 15.

Soil Science Society of America. (2002). Why soil is important?. Available at: https://soils.org/files/science-policy/sssa-marketing-2013.pdf. [Accessed 03 June 2019].

Song, Y.Q., Yang, L.A., Li, B., Hu, Y.M., Wang, A.L., Zhou, W., Cui, X.S. and Liu, Y.L. (2017). Spatial prediction of soil organic matter using a hybrid geostatistical model of an extreme learning machine and ordinary kriging. *Sustainability*, *9*(5): 754.

South African National Space Agency (SANSA). (2015). Available at: http://atlas.sansa.org.za/atlas-geology.html [Accessed 12 May 2020].

South African Weather Service (SAWS). (2019). Historical rain maps. Available at: https://www.weathersa.co.za/home/historicalrain [Accessed 20 May 2020].

Srivastava, R., Sethi, M., Yadav, R.K., Bundela, D.S., Singh, M., Chattaraj, S., Singh, S.K., Nasre, R.A., Bishnoi, S.R., Dhale, S. and Mohekar, D.S. (2017). Visible-near infrared reflectance spectroscopy for rapid characterization of salt-affected soil in the Indo-Gangetic Plains of Haryana, India. *Journal of the Indian Society of Remote Sensing*, *45*(2): 307-315.

Statistics South Africa (SSA). (2011). Census 2011. Available at: http://www.statssa.gov.za/ [Accessed 20 June 2019].

Stenberg, B., Rossel, R.A.V., Mouazen, A.M. and Wetterlind, J. (2010). Visible and near infrared spectroscopy in soil science. *Advances in agronomy, 107*:163-215

Sullivan, D.G., Shaw, J.N. and Rickman, D. (2005). IKONOS imagery to estimate surface soil property variability in two Alabama physiographies. *Soil Science Society of America Journal*, *69*(6): 1789-1798.

Tabachnick, B.G. and Fidell, L.S. (2014). Using Multivariate Statistics, 6th edn. New International Edition. Harlow: Pearson Education Limited**.**

Targulian, V.O., Arnold, R.W., Miller, B. A. and Brevik, E.C. (2018). Pedosphere. Reference Module in Earth Systems and Environmental Sciences. *Encyclopaedia of Ecology*, *2*: 1-7.

The Unscrambler. (2014). The Unscrambler X v10.3 user manual. CAMO software: 1-1370.

Thompson, L.M. (1957). Soils and Soil Fertility. *McGraw-Hill Book Company Inc. New-York.*

Tomašic, M., Zgorelec, Ž., Jurišic, A. and Kisic, I. (2013). Cation exchange capacity of dominant soil types in the Republic of Croatia. *Journal of Central European Agriculture*, *14*(3): 937-951.

Tucker, C.J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, *8*(2): 127-150.

Uchida, R. (2000). Essential nutrients for plant growth: nutrient functions and deficiency symptoms. *Plant nutrient management in Hawaii's soils*: 31-55.

Vågen, T.G., Winowiecki, L.A., Tondoh, J.E., Desta, L.T. and Gumbricht, T. (2016). Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. *Geoderma*, *263*: 216-225.

Van Breemen, N. and Buurman, P. (2002). Soil formation. Springer Science & Business Media: 3-4.

van der Meer, F. (2018). Near-infrared laboratory spectroscopy of mineral chemistry: A review. *International journal of applied earth observation and geoinformation*, *65*: 71-78.

Van Der Merwe, G.M.E., Laker, M.C. and Bühmann, C. (2002). Factors that govern the formation of melanic soils in South Africa. *Geoderma*, *107*(3-4): 165-176.

Vasques, G.M., Grunwald, S.J.O.S. and Sickman, J.O. (2008). Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma*, *146*(1-2): 14-25.

Vibhute, A.D., Kale, K.V., Mehrotra, S.C., Dhumal, R.K. and Nagne, A.D. (2018). Determination of soil physicochemical attributes in farming sites through visible, near-infrared diffuse reflectance spectroscopy and PLSR modeling. *Ecological Processes*, *7*(1): 1-12.

Virgawati, S., Mawardi, M., Sutiarso, L., Shibusawa, S., Segah, H. and Kodaira, M. (2019). Explore the character of soil spectral reflectance relate to the soil organic matter content. *Jurnal Teknik Pertanian Lampung*, *8*(3): 214-223.

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

Wackernagel, H. (2003). Multivariate geostatistics: an introduction with applications. Springer Science & Business Media: 24.

Wadodkar, M.R., Ravisankar, T. and Joshi, A.K. (2014). Application of remote sensing techniques for soil fertility assessment: 111-119.

Walkley, A. and Black, I.A. (1934). An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil science*, *37*(1): 29-38.

Wang, J., Tiyip, T., Ding, J., Zhang, D., Liu, W. and Wang, F. (2017). Quantitative estimation of organic matter content in arid soil using Vis-NIR spectroscopy preprocessed by fractional derivative. *Journal of Spectroscopy*: 1-9.

Wang, X., Han, Z., Wang, W., Zhang, B., Wu, H., Nie, L., Zhou, J., Chi, Q., Xu, S., Liu, H. and Liu, D. (2019). Continental-scale geochemical survey of lead (Pb) in mainland China's pedosphere: Concentration, spatial distribution and influences. *Applied geochemistry*, *100*: 55-63.

Wang, X., Wang, J. and Zhang, J. (2012). Comparisons of three methods for organic and inorganic carbon in calcareous soils of northwestern China. *PlOS one*, *7*(8): e44334.

Weather Atlas. (2020). Monthly weather forecast and climate Bergville, South Africa. Available at: https://www.weather-atlas.com/en/south-africa/bergville-climate [Accessed 12 May 2020]

Webster, R. and Oliver, M.A. (2001). Geostatistics for environmental scientists. *John Wiley & Sons*: 183-184.

Wei, L., Yuan, Z., Wang, Z., Zhao, L., Zhang, Y., Lu, X. and Cao, L. (2020). Hyperspectral Inversion of Soil Organic Matter Content Based on a Combined Spectral Index Model. *Sensors*, *20*(10): 2777.

Wetterlind, J., Stenberg, B. and Rossel, R.A.V. (2013). Soil analysis using visible and near infrared spectroscopy. *Plant Mineral Nutrients*: 95-107.

Wetterstedt, M. (2010). Decomposition of soil organic matter. (Unpublished Doctoral thesis): 1-36.

Worsham, L., Markewitz, D., Nibbelink, N.P. and West, L.T. (2012). A comparison of three field sampling methods to estimate soil carbon content. *Forest Science*, *58*(5): 513-522.

Wuttichaikitcharoen, P. and Babel, M.S. (2014). Principal component and multiple regression analyses for the estimation of suspended sediment yield in ungauged basins of Northern Thailand. *Water*, *6*(8): 2412-2435.

Xue, J. and Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*: 1-18.

Yu, X., Liu, Q., Wang, Y., Liu, X. and Liu, X. (2016). Evaluation of MLSR and PLSR for estimating soil element contents using visible/near-infrared spectroscopy in apple orchards on the Jiaodong peninsula. *Catena*, *137*: 340-349.

Zhang, S., Huang, Y., Shen, C., Ye, H. and Du, Y. (2012). Spatial prediction of soil organic matter using terrain indices and categorical variables as auxiliary information. *Geoderma*, *171*: 35-43.

Zhang, T., Li, L. and Zheng, B. (2013). Estimation of agricultural soil properties with imaging and laboratory spectroscopy. *Journal of Applied Remote Sensing*, *7*(1): 073587.

Zhao, C., Long, J., Liao, H., Zheng, C., Li, J., Liu, L. and Zhang, M. (2019). Dynamics of soil microbial communities following vegetation succession in a karst mountain ecosystem, Southwest China. *Scientific reports*, *9*(1): 1-10.

Zheng, G., Dongryeol, R.Y.U., Caixia, J.I.A.O. and Changqiao, H.O.N.G. (2016). Estimation of organic matter content in coastal soil using reflectance spectroscopy. *Pedosphere*, *26*(1): 130-136.

Zheng, Q., Huang, W., Cui, X., Shi, Y. and Liu, L. (2018). New spectral index for detecting wheat yellow rust using Sentinel-2 multispectral imagery. *Sensors*, *18*(3): 868.

Zhou, M., Liu, C., Wang, J., Meng, Q., Yuan, Y., Ma, X., Liu, X., Zhu, Y., Ding, G., Zhang, J. and Zeng, X. (2020). Soil aggregates stability and storage of soil organic carbon respond to cropping systems on Black Soils of Northeast China. *Scientific Reports*, *10*(1): 1-13.

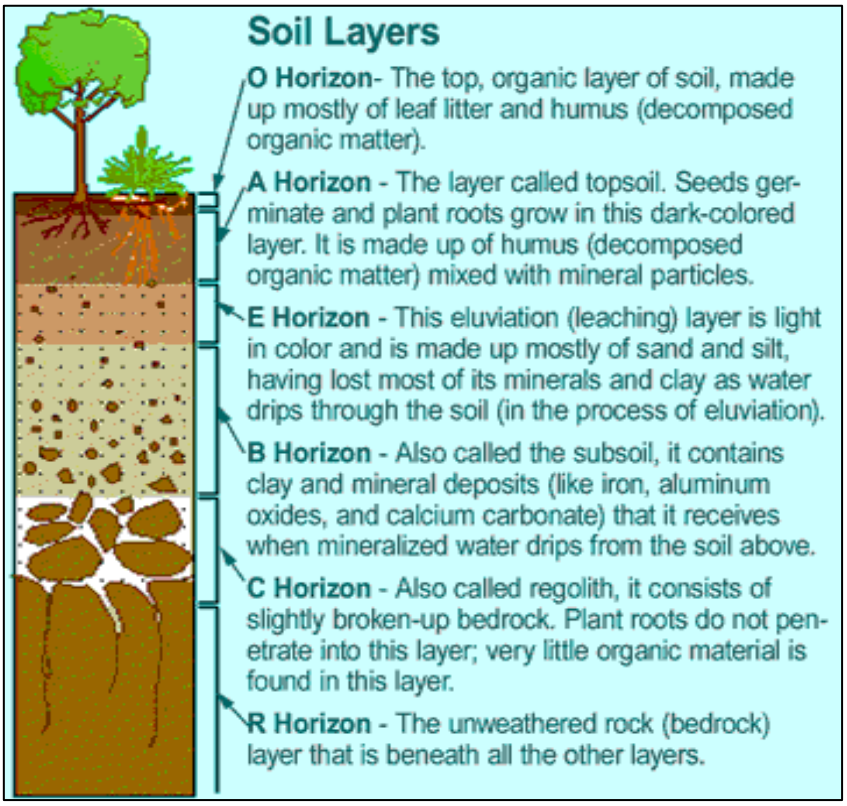
Zhu, L., Suomalainen, J., Liu, J., Hyyppä, J., Kaartinen, H. and Haggren, H. (2017). A review: remote sensing sensors. *Multi-purposeful application of geospatial data*: 19-42.

Zhuo, L., Liu, Y., Chen, J., Hu, C. and Wu, J. (2008). Quantitative retrieving of soil organic matter using field spectrometer and hyperspectral remote sensing. *International Conference on Earth Observation Data Processing and Analysis (ICEODPA)*, *7285*: 72850A.

Žížala, D., Minařík, R. and Zádorová, T. (2019). Soil Organic Carbon Mapping Using Multispectral Remote Sensing Data: Prediction Ability of Data with Different Spatial and Spectral Resolutions. *Remote Sensing*, *11*(24): 2947.

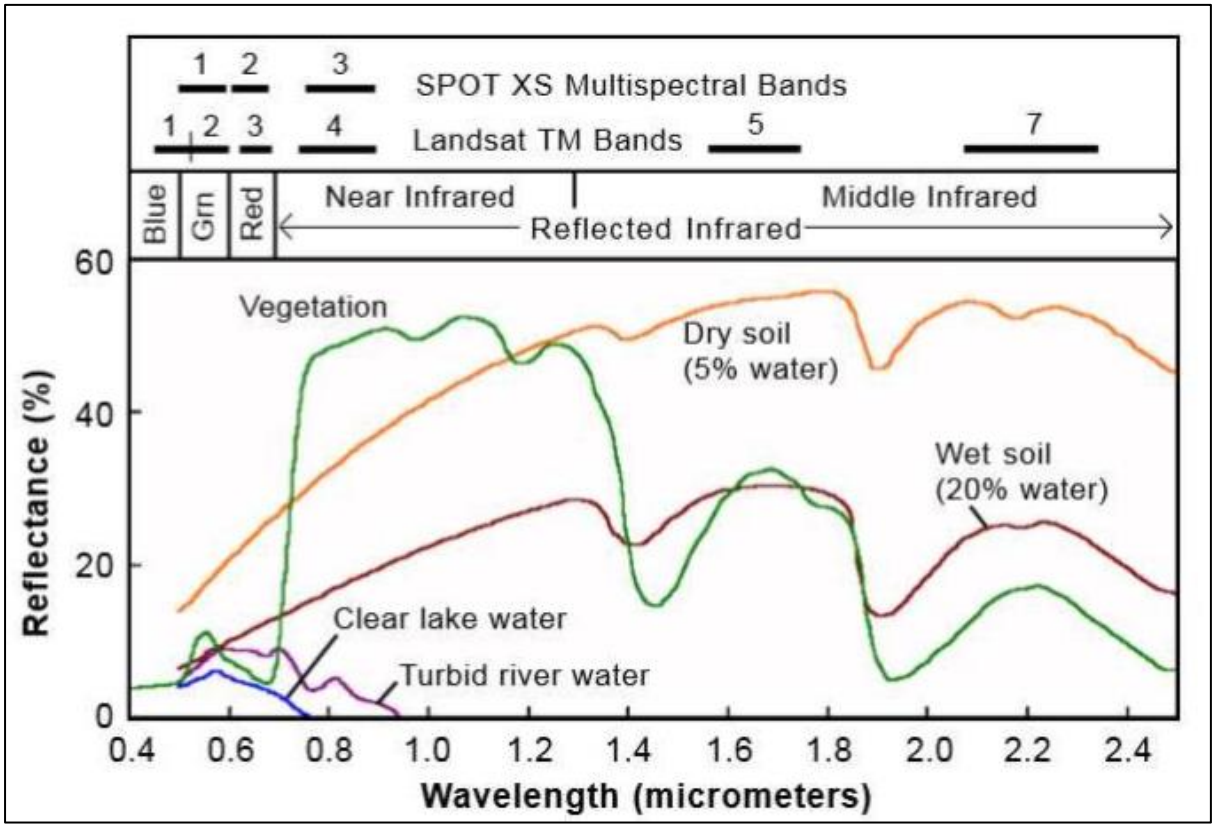
APPENDICES

Appendix A: Soil profile (Source: Hajar, 2020)



Appendix B: Soil samples covered with Parafilm M

Appendix C: Spectral reflectance curve of soil (dry and wet), water and vegetation (Source: Mondal, 2018)



Appendix D: Agricultural land use during the fallow period

Appendix E: Livestock grazing within the study area

Appendix F: Molehills located within the agriculture land use

Appendix G: Crop residues of maize scattered along the agriculture land use

Appendix H: The extensive rangeland land use

Appendix I: Compaction and removal of vegetation created by human pathways within the built-up land use

Appendix J: The extensive erosion within the study area

Appendix K: Portion of the raw Sentinel-2 image containing the study area