RAPID ASSESSMENT OF SOIL CONDITION IN KENYA THROUGH DEVELOPMENT OF NEAR INFRARED SPECTRAL INDICATORS

BY

MUTHONI OLIVE
N50/12953/2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS OF THE DEGREE OF
MASTER OF ENVIRONMENTAL SCIENCE,
KENYATTA UNIVERSITY

DECEMBER 2009

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university or any other award.

Name: **Muthoni Olive** Signature:……………………….

Date:…………………………….

We confirm that the work reported in this thesis was carried out by the candidate under our supervision.

Prof.James Kung'u
Department of Environmental Sciences, Signature…………………………
Kenyatta University
Date……………………………..

Dr.Keith.D.Shepherd Signature…………………………
World Agroforestry Centre (ICRAF)
Date………………………………

DEDICATION

To my mother, who instilled strong values in me, and my brother and sisters who have been part of my academic journey

ACKNOWLEDGEMENTS

The work in this thesis was made possible through the support of a number of laudable people and institutions. I am grateful to my supervisors Prof. James Kung'u, the late Prof. Wellington Wamicha and Dr. Keith Shepherd who graciously guided me through the entire research period. Their invaluable suggestions and constructive criticism contributed significantly to this work.

I earnestly appreciate the tremendous support from the World Agroforestry Centre (ICRAF) staff for the entire course of my research study. Specifically, I thank Andrew Sila, Elvis Weullow, Dickens Ateku, and Caroline Mbogo who assisted me with library research, spectral and data analysis during the period. I am grateful to my family and in particular, my mother and sister Wanjugu who walked the path with me during my studies. Their support was invaluable.

To Joseph, thanks for the encouragement, continuous shove and support! I further extend my deepest appreciation to friends: Leah Mutanu, Mumbi Kamoche, Njoka, M. Ouko, Pauline Makutsa and Victor for their invaluable moral, academic and material support during my study period. My gratitude also goes out to Citibank staff; Kigira, Kioko, Vincent, Wachuka, Njihia, P.Marima and Caroline Wangui, who supported me, stood by me and gave me an opportunity to earn extra skills and income in trying moments of my study period. You are remarkable - God Bless you!

I acknowledge the Higher Education Loans Board of the Government of Kenya for tuition support and the Department of Environmental Sciences, Kenyatta University for academic support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

DEFINITION OF OPERATIONAL TERMS

Absorbance: The quantity of the ratio of absorbed light by a sample surface over incident light.

Calibration: A statistical model relating an independent laboratory test of a set of samples to the intensity of the light absorbance at different wavelengths.

Correlation: Correlation is a statistical technique, which can show whether and how strongly pairs of variables are related.

Indicators: Simple metrics to indicate general soil fertility level.

Infrared: Electromagnetic radiation of a wavelength longer than that of visible light, but shorter than that of radio waves. Infrared radiation spans three orders of magnitude and has wavelengths between approximately 750 nm and 1 mm.

Metrics: A system of parameters or ways of quantitative and periodic assessment of a process that is to be measured, along with the procedures to carry out such measurement and the procedures for the interpretation of the assessment.

Model Validation: The process of testing how well a calibration model predicts target values on a set of independent samples taken from the same population.

Multicollinearity: A statistical term for the existence of a high degree of linear correlation amongst two or more explanatory variables in a regression model. This makes it difficult to separate the effects of them on the dependant variable.

Peak: A peak means the wavelength of radiation where a sample absorbs.

Reflectance: The ratio of reflected light from a sample surface to the incident light on a surface

Reflectance spectroscopy: The study of light as a function of wavelength that has been reflected or scattered from a solid, liquid or gas.

Spectroscopy: The theoretical study of the interaction between electromagnetic radiation, or light and matter used for the understanding of the structure of matter and for qualitative and quantitative analyses.

Spectrum: The spectrum of an object is the frequency range of electromagnetic radiation that it emits, reflects, or transmits.

ACRONYMS AND ABBREVIATIONS

AGRA	Alliance for a Green Revolution in Africa
BNF	Biological Nitrogen Fixation
CEC	Cation Exchange Capacity
CR	Continuum Removal
ENVI	Environment for Visualizing Images
FWHM	Full Width at Half-Maximum
GIS	Geographic Information System
GPS	Global Positioning System
GDP	Gross Domestic Product
ICRAF	World Agroforestry Centre
IR	Infrared
Lab ID	Laboratory Identification Number
MDG	Millennium Development Goals
MIR	Mid-infrared
MPA	Bruker Multipurpose Analyzer
NEPAD	New Partnership for Africa's Development
NIR	Near infrared
NIRS	Near-Infrared Reflectance Spectroscopy
OPUS	Optical User Software
PCA	Principal Components Analysis
PLS	Partial Least Squares
PCR	Principal Component Regression
RMSEC	Root Mean Square Error of Correlation
RMSECV	Root Mean Square Error of Cross-validation
RMSEP	Root Mean Square Error of Prediction
RS	Remote Sensing
SOC	Soil Organic Carbon
SSA	Sub-Saharan Africa

ABSTRACT

Soil degradation in Kenya has environmental and economic impacts. Large area assessments are needed to quantify and diagnose problems of soil fertility and environmental degradation and target sustainable land management interventions, such as agroforestry, and to measure impacts of interventions. Near-infrared reflectance spectroscopy (NIRS) is a low cost, rapid and robust method for characterizing soil condition. The World Agroforestry Centre (ICRAF) has compiled an extensive spectral library of several thousand soil samples from Kenya. However, soil test data of properties determined using wet chemistry is available for only a subset of samples in the database because the methods are too expensive to measure on large numbers of samples. In addition it is tedious to develop separate individual calibrations for each soil property and region. Soil spectra integrate information on a number of soil physical and chemical components and it is against this background, that this study was carried out to derive integrated spectral indictors of soil condition for Kenya based on near infrared spectra. The specific objectives were to summarize the main variation in reflectance spectra in the soil samples into simple metrics, summarize the main variation in the soil chemical and physical properties and relate the spectral metrics to the individual soil chemical and physical properties and their combined principal components. Eight hundred and forty three soil samples (0-20 and 0-45cm depths) from different parts of Kenya that had complete physical and chemical data were randomly selected. The soil properties tested were organic carbon, pH, exchangeable Ca and Mg, extractable P and K, sand and clay content. Soil spectra were recorded on these samples using a Fourier Transform infrared spectrometer. The absorbance peak heights of the three principal spectral absorption features after baseline correction were proposed as the basis for the spectral condition index. The soil properties were related to the absorbance peak heights using Partial Least Squares Regression (PLS) regression. A second set of relationships was developed by relating the soil properties to the full spectrum using the Bruker Quant 2 method, which is also based on PLS but uses all the wavebands. The peak heights displayed higher correlation with soil properties after baseline correction - Continuum Removal (CR) as compared to before CR. The peak heights predicted exchangeable Ca which is a key soil fertility parameter moderately well, with calibration and cross-validated r^2=0.60. Sand, exchangeable K and extractable P however, had the poorest correlations with the spectral peak data. The full spectrum (use of Quant 2 method) provided effective predictions for the individual soil properties: ExCa (r^2 =0.86), ExMg (r^2 =0.74) and pH (r^2 =0.61). The method was also most robust in predicting the first principal component of the soil properties(r^2=0.76). Better models to predict soil condition from spectral metrics are obtained by using the full wavelength range, as opposed to restricting the models to the use of the three peak regions (using the PLS method), using basic peak height and width information related to the key absorption features. The study showed that the full spectrum method was most robust in developing

spectral indicators of soil condition in Kenya based on the first principal component of the soil conditions. In conclusion, the soils in Kenya are highly variable in distribution, types and condition and near infrared spectroscopy has an important role to play in assessment of soil condition by enabling more intensive soil sampling schemes by virtue of the simple, rapid and cheap nature of the analytical method. The technique can assist farmers to carry out frequent analysis to determine whether their farming activities are degrading or when giving recommendations on how to improve depleted soils and boost agricultural productivity. On-going research on the use of the technique in Kenya and its applications in environmental management should facilitate the transfer of this knowledge from the laboratory to the landowners, farmers, governments and other resource managers for more effective decision making. Based on the findings of this study, further research could examine ways of classifying spectra into groups and schemes based on full spectra as an alternative pathway for identifying simple soil fertility spectral indicators. This would help to capture the complexity in the shapes of the absorption features, while simplifying the complexity in the spectra.

CHAPTER ONE

INTRODUCTION

1.1 Background

Strategies to prevent soil degradation are critical for sustainable development, particularly in agriculture. Soil degradation has environmental and economic impacts at scales ranging from the field level to national levels where soil nutrient depletion, degraded soil structure and lost organic matter affect farm livelihoods and watershed effects level. Several studies have documented the significance of erosion in soil degradation throughout Sub-Saharan Africa (SSA) at these varied scales, and given the minimal use of soil amendments by rural farmers, which has profound implications on sustained regional agricultural production (Shepherd and Walsh, 2004). Though some local success in controlling and reversing soil degradation has been documented, the soil resource continues to decline alarmingly in the region (Cohen *et al.*, 2006).

Cohen *et al.* (2006) estimated that soil erosion losses in Kenya are equal in magnitude to national electricity production or agricultural exports (equivalent to $ 390 million annually or 3.8% of Gross Domestic Product (GDP).For example the densely populated Nyando District in Western Kenya, soil erosion costs were several fold higher than in neighboring districts (Kericho, Kisumu), indicating deeply marginalized rural farmers. The principal causes of erosion include devegetation (deforestation, clearance of woody vegetation, unprotected farmland), overgrazing and overuse of extensive areas of fragile lands on both hill slopes and plains.

Soil fertility decline involves the closely interlinked soil chemical, physical and biological degradation. For example, after converting forest to crop land, reduced litter inputs and soil tillage lead to reduced soil organic matter, which in turn leads to decreased soil nutrient supply and soil structural deterioration. Soil structural deterioration leads to increased run-off and erosion, which causes further loss of organic matter and exchange surfaces associated with finer soil particles. As soil nutrients levels decline, vegetation is reduced and nutrient loss through leaching is exacerbated.. Reduced vegetation cover also exposes the soil to increased erosion. This is a vicious cycle that can lead to sudden switches from fertile to degraded soil states. To reflect these combined closely interlinked processes, the term ‘soil condition’ is preferred to ‘soil fertility’ (Shepherd and Walsh, 2004).

One of the biggest challenges for managing soil degradation is the lack of empirical data on soil condition and trend of degradation (Shepherd and Walsh, 2007). Large area surveys are required to establish baselines of current soil condition to target interventions and assess their impact on the improvement of soil condition. However, conventional analysis of soil properties is time-consuming and expensive. This prohibits analysis of the large number of samples required in large area surveys and monitoring programmes. In addition, national programmes in sub-Saharan Africa are struggling to maintain functional soil and plant analytical laboratories and maintain the supply of required chemicals (Swift and Shepherd, 2007). Near infrared spectroscopy (NIRS) has shown promise as a rapid, low cost and reliable method for assessment of soil properties (Brown, *et al.*, 2006; Shepherd and Walsh, 2002; 2004). The method requires minimal sample preparation, is non-destructive, and requires no chemicals. NIR technology may provide a sustainable way for national programmes to provide reliable soil and plant analyses at low cost.

Infrared (IR) spectroscopy employs the infrared part of the electromagnetic spectrum to identify a compound and to investigate the composition of a sample. It provides precise information on soil conditions more cheaply and rapidly than the traditional analyses that uses costly chemicals. The IR technique uses just light, in a non-destructive analysis of soil and plant materials. Working from a digital scan, a ‘reflectance fingerprint’ that can be used to identify multiple soil properties is obtained.

In Kenya, large-area diagnostic surveillance of soil condition using IR is expected to increase the efficiency of scaling up soil and crop improvement practices by improving both the targeting of interventions as well as impact monitoring.

1.2 Statement of the Problem

The World Agroforestry Centre has compiled an extensive near infrared spectral library of several thousand soil samples from Kenya, including characterization of the entire Kenya Soil Survey archives(World Agroforestry Centre,2006) However, soil test data (soil properties determined using wet chemistry) is available for only a subset of samples from the spectral library.

The cost of acquiring further wet chemistry data to be able to establish calibrations between individual soil tests and infrared spectra for different regions or soil types is high, and costs $50 a sample as opposed to 50cents through the use of IR data. There is a further problem with standardization and benchmarking with different laboratories using different soil test methods and with quality control both within and among laboratories. In Africa, national soil testing laboratories are closing down at a time when they should be gearing up to meet development challenges (Shepherd and Walsh, 2002).There is however new opportunity to use IR as an intervention method for soil testing in these laboratories.

At the same time, the large projected increases in population levels, coupled with the increased economic growth associated with development, will result in massively increased demand on ecosystems in developing countries over the next several decades, threatening environmental sustainability (MDG Goal 7) (Shepherd and Walsh, 2007). Measures aiming at the enhancement of soil fertility therefore obviously need to be based on a more precise characterization of both the soil and the socio-economic conditions in the country concerned (Barrett *et al.*, 2002).

The current laboratory soil assessment situation relies largely on expert opinion to extrapolate information on soil constraints based on very few sampling locations (Shepherd and Walsh, 2002). There is need for large area assessment and monitoring of soil properties to understand human impacts on the soil resource base and reliably guide sustainable development policy. However, the high cost of soil analysis currently makes such large studies infeasible (Shepherd and Walsh, 2005).

Walsh (unpublished) has developed a soil condition indicator for Western Kenya using Lake Victoria sediment as a standard, and has shown preliminary promise for use in other regions. Further work is required to develop simple spectral indicators of soil fertility that could be generally applied in more widely in Kenya and in the tropics.

Rapid and cost effective characterization methods of soil condition are therefore needed for the screening and completion of soil surveys (Brown *et al*, 2006).Ideally, generalizable spectral indicators of soil fertility could be found that could be interpreted in terms of multivariate soil properties.

1.3 Justification of the study

Measurement of soil physical, chemical and biological properties, especially over large areas, is time consuming and expensive, and as a result the sampling strategies adopted are less than ideal. IR technology offers solution to this. The analysis of multiple constituents of soil condition can be made rapidly in a single operation using IR technology (400 samples can be processed in a day by a single operator) and with minimal sample preparation (Martens and Naes, 1989). IR is therefore low-cost (savings of at least 100 fold of normal laboratory costs are easily achievable). IR could be used as a basis for providing recommendations on how to improve depleted soils and boost agricultural productivity. IR analysis is non-destructive and produces no chemical wastes, and thus reduces environmental pollution from the soil science laboratory.

Another area in which IR can be used is in fertilizer recommendations. For example, current fertilizer recommendations in Kenya are broad based and were developed using old data (1980's) and need to be revised to reflect changes. This requires rapid, cheap and reliable method of analysis.

Until now, the absence of tools for rapid screening of soil and plant health has limited the completion of soil and plant nutrition surveys. The ability to analyze large numbers of soil samples from landscapes allows uncertainties in the estimates of soil condition to be expressed in quantitative terms. In Kenya, large-area diagnostic surveillance will increase the efficiency of scaling up soil and crop improvement practices by improving both the targeting of interventions as well as impact monitoring. This is because land use in developing countries is changing so rapidly, soil conditions are in constant flux, which greatly limits the value of one-off surveys.

The IR technology has the capacity to considerably expand national fertility mapping capabilities. Tests have shown that IR analysis can be effectively used in conjunction with GPS and satellite sensing to produce inexpensive maps pinpointing areas with soil nutritional problems. The combination of IR spectroscopy and geographic positioning systems provide one of the most powerful modern tools available for agricultural and environmental monitoring and analysis (Shepherd and Walsh, 2007). The method allows many georeferenced soil samples to be characterized from an area, and this permits soil constraints to be calibrated to satellite imagery and mapped out over large areas (Vagen *et al.*, 2006). The ability to analyze large numbers of geo-referenced samples also facilitates digital soil mapping by calibration of soil properties to satellite imagery. Also to greatly benefit from the technology are the development community, private sector and governments that are in interested in carrying out Environmental Impact Assessments of projects that could affect soil quality.

Generally applicable calibration libraries still need to be developed to relate standard soil properties and soil functions to NIR spectral data. Calibrations developed in one region may not hold up in another region and so there is still considerable investment required to build calibration libraries. Development of simpler, indicative spectral indicators of soil condition and fertility can be used in rapid diagnostic surveys and are a quantitative, evidence based approach to agriculture and environmental management in Kenya. This would bring about considerable savings in time and cost.

1.4 Research Questions

The study was guided by the following research questions:

1. Are there simple metrics that can summarize the main variation in reflectance spectra in the soil samples?
2. What variation exists in the chemical and physical properties of the soils in the soil library and are some properties closely inter-related?
3. Are there relationships of these metrics with the main variation in soil chemical and physical properties?

1.5 Objectives of the Study

The overall objective of the study was to develop a simple, generalizable indicator or set of indicators of soil condition for Kenya based on near infrared spectra. The specific objectives using an existing Kenya soil database were to:

1. Compute simple metrics summarizing the main variation in reflectance spectra in the soil samples based on height and width of absorption features and mean reflectance.
2. Summarize the main variation in soil chemical and physical properties in the soil library.
3. Relate the spectral metrics to the specific soil chemical and physical properties and their combined principal components.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

This chapter gives a general overview of previous work done in Sub-Saharan Africa in assessing soil condition using conventional methods and infrared spectroscopy, in relation to the problems of soil infertility, food insecurity and environmental degradation. It is divided into 6 sections as follows; the first two sections (2.1 and 2.2) address the nature of soils and agriculture and the resultant land and soil fertility degradation occasioned by population pressure, poverty and unsustainable farming practices. Section 2.3 deals with soil fertility evaluation in Sub-Saharan Africa. Section 2.4 spells out the physical and chemical analysis and infrared spectroscopy ways of assessing soil condition. The last two sections (2.5 and 2.6) show the application of Near Infrared Spectroscopy (NIRS) in soil condition diagnosis and how the application of the technique is important for better environmental and agricultural management.

2.1 Nature of soils and agriculture

A key measure of the long-term productive capacity of an agro ecosystem is the condition of its soil (Okeyo, 2006). Soils play a key role in sustaining many ecosystem services which include biomass production, carbon sequestration, habitat provision, water filtration, waste absorption and breakdown (Swift and Shepherd, 2007). Both, natural weather processes and human management practices can both affect soil quality. Sustaining soil productivity requires that soil-degrading pressures be balanced with soil-conserving practices (World Resource Institute, 2002).

Key soil nutrients in the soil have different importance in plant growth and food sustainability. Soil pH characterizes the chemical environment of the soil and may be used as a guide to suitability of soils for various pasture and crop species

(Hazelton and Murphy, 2007). Cation exchange capacity is the capacity of the soil to hold and exchange cations. The cations in the soil include exCa, exMg and exK.CEC is a major controlling agent of stability of soil structure, nutrient availability of plant growth, soil pH and the soil's reaction to fertilizers and other ameliorants (Brady and Weil, 1996). Phosphorus is an essential constituent of numerous substances involved in biochemical reactions including photosynthesis and respiration (Shepherd and Walsh, 2004).

Total nitrogen measures the total amount of nitrogen present in the soil, much of which is held in organic matter and is not immediately available to plants. A measure of carbon in the soil helps determine the rate of organic decay, the level of organic matter, and the amount of nitrogen available to plants (Brady and Weil, 1996).The size of mineral particles in soil including clay, sand and silt is critical to understanding soil behaviour and management (Dent and Young, 1981).

During the last decade, food security was not a global priority, but studies such as the 2020 Vision and the World Food Summit have shown that food security is a major global concern for this century. About 55% of Africa's land is unsuitable for agriculture and by 1990, soil degradation was estimated to have affected 500 million hectares, i.e. 17% of Africa's land (UNEP 1997). Only 11% of the continent, spread over many countries, has high quality soil that can be effectively managed to sustain more than double its current population (Eswaran *et al.*1997). Most of the remaining usable land is of medium or low potential, with at least one major constraint for agriculture. This is largely a result of continuing rapid population growth, the highest of any region in the world, and rapid land depletion (Sanchez *et al.*, 1997). While per capita food availability in the rest of the world has increased significantly over the past 45 years, the situation in sub-Saharan Africa (SSA) has improved only slightly (Eswaran *et al*., 1997). In contrast to sustained increases in other parts of the developing world, per capita food production continues to decrease in Africa (Sanchez *et al.*, 1997).

Poor soils and human poverty go hand-in-hand, especially in the tropics (Misiko, 2007). More than 70% of Africa's poor live in rural areas, a pattern that is expected to continue for many years (Shepherd and Walsh, 2007). Since the rural poor derive most of their livelihood from agriculture, increasing agricultural productivity is essential for significant poverty reduction (Swift, *et al.,* 2004*).* Massive increases (for example, 6% per annum) in agricultural productivity in developing countries are required if the world is to achieve the targets it has set itself towards meeting the Millennium development goals (MDGs) to eradicate extreme poverty and hunger (Goal 1) and to improve human health (Goals 4,5 and 6) (UN Millennium Project, 2005). The need for rapid agricultural development for poverty alleviation is particularly pronounced in Sub-Saharan Africa, where for example, there has been no increase in fertilizer use and crop yields per hectare or *per capita* Gross Domestic Product (GDP) over the past 30 years, whereas population has more than doubled over the same period and is expected to double again over the next 30 years (UN Millennium Project, 2005).

According to the World Bank Agriculture for Development Report (2008), 75 percent of the world's poor live in rural areas in developing countries, where only a mere 4 percent of official development assistance goes to agriculture. The report therefore calls for greater investment in agriculture in developing countries. The report warns that the sector must be placed at the center of the development agenda if the goals of halving extreme poverty and hunger by 2015 are to be realized. Further, agricultural and rural sectors have suffered from neglect and underinvestment over the past 20 years. The World Bank Group is advocating a new dynamic 'agriculture for development' agenda. According to the Report, GDP growth originating in agriculture is about four times more effective in reducing poverty for the poorest people, than GDP growth originating outside the sector. This can benefit the estimated 900 million rural people in the developing world who live on less than $1 a day, most of whom are engaged in agriculture.

African agriculture has performed dismally, in sharp contrast to Asia and Latin America, regions that benefited from the Green Revolution, (Sanchez, 2002).

Some African states and developed countries are now considering programmes that give high priority to restoring agricultural development in Africa. Paramount among them includes the Alliance for a Green Revolution in Africa (AGRA).

Agriculture remains the most important economic activity in Kenya, although less than 8% of the land is used for crop and feed production. Less than 20% of the land is suitable for cultivation, of which only 12% is classified as high potential (adequate rainfall) agricultural land and about 8% is medium potential land. The rest of the land is arid or semiarid (Britannica online, 2007). Several reasons have been advanced to explain the decline of food production in Kenya. They are natural disasters (flood, drought), a high incidence of pests and diseases and degradation of soil resource base characterized by a decline in soil organic matter and nutrient availability (Macharia, 2004). Agricultural production in low potential areas of Kenya is constrained by declining soil fertility and unpredictable and erratic rainfall. One of the challenges faced by those involved in rural development is improving soil fertility in low potential areas (Kinyanjui *et al*., 2000).

Stagnation of agricultural development in many sub-Saharan African regions is commonly attributed to the limited adoption of new and improved technologies (Misiko, 2007). By fundamental root cause, it means that no matter how effectively other conditions are remedied, per capita food production in Africa will continue to decrease unless soil-fertility depletion is effectively addressed. However, soil fertility in Africa has seldom been considered a critical issue by the development community, who until very recently has focused primarily on other biophysical constraints such as soil erosion, droughts, and the need for improved crop germplasm (Sanchez *et al*., 1997).

The agriculture and environment agendas are inseparable. Therefore managing the connections between agriculture, natural resource conservation, and the environment must become an integral part of using agriculture for development in order to achieve more sustainable agricultural production systems (World Bank Development Report, 2008).

To lift Africa out of poverty, large increases in the scale of agricultural research and extension and environmental assessment will be required, both in the private and public sector (Shepherd and Walsh, 2007). It is becoming increasingly recognized that developing countries need to draw on the latest scientific and technological advances available to accelerate their agricultural development, but these technologies need to be robust enough to be used under basic conditions and to be applied at scale (Shepherd and Walsh, 2002).

2.2 Land and soil fertility degradation

The continued threat to the world's land resources is exacerbated by the need to reduce poverty and unsustainable farming practices (Sanchez *et al.*, 1997). Little reliable data is available on the extent of land degradation in Africa. However, anyone who has traveled through the continent has observed that land degradation is widespread and serious. The presence of gullies and sand dunes, degraded forests and grazing lands are obvious, although the effects of sheet erosion and declining soil fertility are less noticeable.

The degradation of soil fertility, specifically the capacity of soil to support agricultural production, has been identified as one of the main causes of Africa's agricultural failure. Beyond this, African farmers face other degradation processes such as erosion, salinization and acidification (Swift *et al.,* 2004). Soil fertility depletion in smallholder farms is now recognized as the biophysical root cause of declining food security in this region (Sanchez and Leakey, 1997). Degradation of soil quality poses a serious threat to human welfare and the environment (Jones and Doran, 1996; Trimble and Crosson, 2000). A particular challenge is the high level of variation in the properties of African soils. This variation is often highest at the smallest scale, a farmer's field (Swift *et al.,* 2004*).* The magnitude of nutrient depletion in Africa's agricultural land is enormous. Sanchez *et al*. (1997)

quotes calculations based on Smaling's seminal works that indicate that an average of 660 kg N ha^{-1}, 75 kg P ha^{-1} and 450 kg K ha^{-1} has been lost during the last 30years from about 200 million ha of cultivated land in 37 African countries, excluding South Africa. These figures represent the balance between nutrient inputs as fertilizer, manure, atmospheric deposition, biological N_2 fixation (BNF),and sedimentation, and nutrient outputs as harvested products, crop residual removals, leaching, gaseous losses, surface runoff, and erosion(Cohen *et al*., 2006). These values are the aggregate of a wide variety of land-use systems, crops, and agro ecological zones in each country (Stoorvogel & Smaling, 1990).

Though some local success in controlling and reversing soil degradation has been documented, the soil resource continues to decline regionally, at alarming rates in some areas (Cohen *et al*., 2006). It has been recognized that the problem of soil fertility degradation is a microcosm of that of land degradation as a whole. Therefore, reversing soil-fertility depletion has been identified as a requirement for increasing per capita agricultural production. (Sanchez *et al.,* 1997). In scientific research, soil fertility is broadly recognized. Low soil fertility as a fundamental bottleneck for food security in smallholder farms in Africa (Misiko, 2007). For instance, the Soil Science Society of America (2001), in a special issue of the society's journal (No.58), introduces soil fertility as necessary to sustain agriculture on a continent threatened with high population growth. Swift *et al*. (2004) states:' the soil fertility problem remains intractable largely because of the failure to deal with the issue in a sufficiently holistic way. Soil fertility decline is not a simple problem. In ecological parlance it is a slow variable, which interacts over time with a wide range of other biological and socioeconomic constraints to sustainable agro ecosystem management. Tackling soil fertility issues thus requires a long-term perspective and a holistic approach that integrates biological and social elements.'

Increasing pressures on agricultural land have resulted in much higher nutrient outflows and the subsequent breakdown of many traditional soil-fertility maintenance strategies. Such strategies have not been replaced by an effective fertilizer supply and distribution system (Sanchez *et al.*, 1997). Soil fertility decline is associated with several simultaneous degradation processes feeding on each other to produce a downward spiral in productivity and environmental quality. For example, the combined effects of tillage and insufficient applications of nutrient and organic matter inevitably lead to a decline in soil organic matter (Swift *et al.*, 2004*).* This reduces the retention of essential plant nutrients, breaking down soil physical structure and in turn diminishing water infiltration and the water storage capacity of the soil.

High nutrient depletion is due to high outputs of nutrients in harvested products and erosion and also in the relatively high inherent fertility of the soils (Smaling *et al*., 1997). Soil-fertility depletion leads to decline in crop productivity, decreased food security, exacerbates several environmental problems at the national and global scales, increased government expenditure on famine relief and increased poverty (Sanchez *et al.*, 1997).

Several studies have documented the significance of erosion in soil functional degradation throughout SSA at varied scales, which, given minimal use of soil amendments by rural farmers, has profound implications on sustained regional agricultural production (Smaling *et al.*, 1997). In general, the factors that cause environmental degradation, such as soil erosion are well known. However, the magnitude and contribution of each of these factors, in specific situations is rarely known with any accuracy (Okeyo, 2006).

There is evidence of increased erosion occurring in some African rivers and lakes (Sanchez *et al.,* 1997), including Lake Victoria, where erosion from surrounding nutrient-depleted lands is widespread. Further, the loss of topsoil organic carbon associated with soil nutrient depletion results in additional CO_2 emissions to the atmosphere from decreasing soil and plant C stocks, (Sanchez, 1995). Therefore apart from low food production, inadequate soil management has serious consequences for other natural resources essential to African livelihoods and development (Swift *et al*., 2004).

In Kenya, like many other Sub-Saharan countries, soil fertility and hence productivity is declining at an alarming rate because areas of high agricultural potential are densely populated and in most cases farm holdings are less than one hectare (Okeyo, 2006). Declining soil fertility is a therefore a serious concern in Kenya. Since the soil resource has not kept its productive capability over time, farmers have witnessed low and declining yields. For example, a long-term trial in Kabete, Kenya, indicates that a previously fertile, red soil (Oxic Rhodudalf) lost about 1 ton ha^{-1} of soil organic N and 100kg ha^{-1} of soil organic P during 18 years of continuous maize-common bean rotation in the absence of nutrient inputs (Sanchez *et al.*, 1997). Further, for most cereal and legume crops grown in the country, nutrients removed in the harvested parts are not returned to the fields (Macharia, 2004). Many smallholder farmers in the country have not yet realized the value of using green manure and compost to maintain and improve soil fertility (KARI, 2001). The consequences of these actions for smallholder farms, particularly in the last two decades, have been a progressive decline in yields and diminishing soil fertility.

Sustaining soil productivity requires that soil-degrading pressures be balanced with soil-conserving practices (Okeyo, 2006). One example of a natural resource management approach deals with a problem that was invariably identified by farmers in characterization and diagnosis exercises throughout the sub-humid and semi-arid tropics of Africa (Sanchez, 1999). According to Shepherd and Walsh (2002), agricultural development in the tropics will require tens of thousands of field trials to develop recommendations for improved crop and soil management practices, supported by millions of soil analyses. This is not practical and methods for repeatable, rapid assessments for soil diagnosis are recommended.

2.3 Assessment of soil condition

Most agricultural and environmental plans require soil analysis, or at least should require analysis to better implement any change (Shepherd and Walsh, 2002). The purpose of a soil fertility evaluation laboratory is to provide rapid diagnostic services to farmers. In the tropics, few effective soil fertility evaluation methods exist in Africa and Southeast Asia (Sanchez, 1976). Typically, soil samples are analyzed, application rates calculated, and the fertility interpretations are made (Nanni and Demattê, 2006). The most widespread methods are based on soil testing, missing element techniques, simple fertilizer trials, and frequently a combination of these (Sanchez, 1976). Soil samples are normally composites consisting of 15 to 20 individual samples for an area of 12 to 20 ha.

Conventional assessments of soil quality rely on local calibration of soil performance to soil laboratory tests, but these analyses are expensive and large numbers of samples are required to adequately characterize spatial variability (Jones and Doran, 1996). Soil data on which the diagnosis is based are dangerously deficient. The combination of laborious methods and a shortage of scientific and technical expertise means that diagnostic analysis has been limited geographically and has rarely been repeated (Swift. *et al.,* 2004*).*

In Brazil, there are legislative regulations that limit the use of chemical products in the laboratory especially for methods that produce environmental pollutants (Nanni and Demattê, 2006). It is important to develop low cost, highly efficient, new methods of soil analyses that produce less environmental pollutants. Standard procedures for measuring soil properties are complex. According to (Nanni and Demattê, 2006), the costs of soil analyses with precision agriculture systems are very expensive when compared with more traditional methods.

Thus, more affordable alternatives are required and these can be achieved through the use of new technologies to estimate soil attributes. Furthermore, better practical methods that can rapidly estimate soil properties are needed to improve quantitative assessments of land management problems (Shepherd and Walsh, 2002). Many studies have been devoted to unraveling major factors constraining tropical soil fertility in order to achieve sustainable agriculture (Misiko, 2007). Sensing soil quality builds national capacity in soil health surveillance, which will directly contribute to priority areas in the Action Plan of The Environment

Initiative of the New Partnership for Africa's Development (NEPAD) and the UN Millennium Project (Shepherd and Walsh, 2005).

Soil fertility evaluation is the process by which nutritional problems are diagnosed and fertilizer recommendations made (Sanchez, 1976). Soil fertility is not a static feature. On the contrary, it changes constantly and its direction (accumulation or depletion) is determined by the interplay between physical, chemical, biological and anthropogenic processes (Smaling *et al.*, 1997). The detailed information necessary to conduct management and operational control tasks is thus rarely available to resource managers and hence they cannot determine the best way to address specific issues (Okeyo, 2006). Until now, the absence of tools for rapid screening of soil and plant health and the cost has limited the completion of soil and plant nutrition surveys (Shepherd and Walsh, 2002).

People depend on soil to provide a wide range of essential 'ecosystem services'. Therefore apart from low food production, inadequate soil management has serious consequences for other natural resources essential to African livelihoods and development (Swift and Shepherd, 2007). The evaluation approach therefore advocates for careful management of soil fertility aspects that optimize crop production potential through incorporation of a wide range of adaptable soil management principles, practices and options for productive and sustainable agro ecosystems (Okeyo, 2006).The global threat of environmental problems cannot be reliably assessed without methods for rapid assessment of soil quality (Shepherd and Walsh, 2001). Lack of measurable attributes that reflect the capacity of soil to perform specific production or environmental functions makes broad scale quantitative assessment difficult (Trimble and Crosson, 2000).

The Round Table of Experts, which met in Nairobi in February 2006 proposed an approach that aims at providing reliable data on the condition of the soil-resource base and degradation trends; spatially explicit early warning of emerging soil-related problems and a scenario analysis; and reliable ex post information on impacts of large-area soil management interventions (Shepherd and Walsh, 2007). This diagnostic approach calls for application of the latest scientific and technological advances, including remote sensing and GIS; infrared spectroscopy for rapid soil analysis; new multivariate statistical tools for handling hierarchical data; simulation and spatial modeling; and environmental accounting and economic valuation (Swift and Shepherd, 2007).

A rapid and economical soil analytical technique is therefore needed before farmers and land managers will be able to fully utilize soil testing as an aid in precision farming and both assessment and management of soil quality. Finally, while it is clear that soil analyses are necessary, they are still expensive, time-consuming, and often create undesirable environmental waste products (Nanni and Demattê, 2006).

2.4 Near Infrared spectroscopy and soil diagnosis

Rapid and cost effective characterization methods of soil condition are needed for the screening and completion of soil surveys (Brown *et al.*, 2006). Advances in infrared spectroscopy and chemometrics have created new possibilities for rapid non-destructive assessment of soil constituents, but their routine application is limited (Reeves and McCarty, 2000). The ability to analyze large numbers of soil samples from landscapes allows uncertainties in the estimates of soil condition to be expressed in quantitative terms (Sanchez *et al*., 1997).

The use of precision agriculture, which required rapid assessment of soil attributes, began in the 1980s and has become increasingly more popular in the 1990s (Nanni and Demattê, 2006). Since the mid-1980s, developments in instrument technology and chemometrics (the application of mathematical and statistical techniques to chemical data) have led to the increased use of spectroscopy in the laboratory and field and from space platforms, notably in geological studies (Shepherd and Walsh, 2002).

The composition of organic matter is ultimately reflected in the types of bonds between the atoms or groups of atoms that make up tissues (Shepherd and Walsh, 2004). The signatures result from electronic transitions of atoms and vibrational stretching and bending of structural groups of atoms that form molecules or crystals. Spectral signatures of materials are defined by their reflectance or absorbance, as a function of wavelength in the electromagnetic spectrum (Shepherd and Walsh, 2002), and they provide a rapid prediction of soil physical, chemical and biological properties (Okeyo, 2006).

Infrared (IR) spectroscopy meets the requirement of a rapid, reliable and low cost screening tool for many agricultural and environmental applications and could play a major role in enabling diagnostic surveillance approaches in developing countries,(Shepherd and Walsh, 2007).

Another great opportunity IR spectroscopy has provided is the ability to relate high density, georeferenced measurements of soil condition to remote sensing (RS) data (Vlek, 1995). In fact, Nanni and Demattê, 2006 emphasized the importance of RS to estimate CEC, moisture, and soil nutrients as the basis for site-specific management, even though satellite results were still not readily available.

Correlations between the different bands of the electromagnetic spectra and RS data have led to a better understanding of complex soil components. These RS studies are the basis for new paradigms of nondestructive methods to quantify soil attributes (Nanni and Demattê, 2006).Shepherd and Walsh (2002) developed a scheme to use a spectral library as a fast and nondestructive estimation of soil attributes based on analyses of diffusion reflectance spectroscopy to facilitate soil RS analysis. According to Nanni and Demattê (2006), "estimation of soil variables from spectral data used in soil mapping is an important potential application of multispectral remote sensing."

Although infrared (IR) spectroscopy has so far had limited use in poorer developing countries, it has potential to make a huge contribution in helping these countries accelerate agricultural development while safeguarding the environment (Shepherd and Walsh, 2007).

Near-infrared reflectance spectroscopy (NIRS) is a nondestructive analytical technique for studying interactions between incident light and a material's surface (Chang *et al,* 2001). Near infrared (NIR) spectroscopy, is one of the most cost-effective and reproducible analytical techniques available for the 21^{st} century (Shepherd and Walsh, 2007). A wide range of materials can be analyzed for multiple constituents in seconds with no sample preparation.

When a sample of organic matter is irradiated, the bonds continually vibrate, which causes stretching and bending. This in turn causes a type of wave motion within the bond at a frequency that is characteristic of the functional group. The frequencies of the incident light that match the frequencies of the vibrational waves are absorbed whereas other frequencies are reflected or transmitted (Stuart, 2004). According to Chang *et al,* 2001, near-infrared spectra are dominated by weak overtones and combinations of fundamental vibrational bands for H-C, H-N, and H-O bonds from the mid-infrared region. Near infrared radiation (750-2500 nm) is absorbed mainly by C-H, N-H and O-H bonds (Osborne *et al*., 1993), which are the primary constituents of all organic compounds. The chemical constituents of the material determine the nature and number of bonds present and therefore the wavelengths and amount of light that is absorbed. Therefore, the spectrum that is absorbed from the sample contains detail on the chemical composition of that material (Shenk *et al*., 1992; Shenk and Westerhaus, 1994).

In addition to the chemistry of a material, near-infrared spectra are also influenced by the physical structure of a material. The size and shape of the particles, the voids between particles, and the arrangement of particles affect the length of the light transmission passing through a sample and thereby influence reflectance (Chang *et al*, 2001).

Near Infrared Spectroscopy (NIRS) is widely used in industry due to its simplicity, rapidness, and the need for little or no sample preparation. Several studies have focused on measuring soil properties using NIRS. Soil parameters used by Stenberg *et al.* (1995) included clay content, cation exchange capacity, base saturation, and pH. According to the study, cost savings of 70% were achieved when the technique was compared to the most appropriate sampling strategy involving only wet chemistry analysis.

NIRS relies on applying statistical models that test the intensity of the relationship between a particular absorbance and an independent laboratory test of different material (Shenk *et al*., 1992; Shenk and Westerhaus, 1994). Recent research has demonstrated the ability of reflectance spectroscopy to provide nondestructive rapid prediction of soil physical, chemical, and biological properties in the laboratory (Shepherd and Walsh, 2002). Chang *et al.*2001 concluded that reflectance spectra are strongly affected by soil moisture content, organic matter content, and particle size. A potential application of NIRS for soil biologists is its usefulness in selecting soil samples from populations to maximize the variation in particular soil properties in a minimum subset of samples from the population (Stenberg *et al*., 1995). This is a powerful application of NIRS, as it allows one to select samples over the full range of variation in any one or combinations of parameters.

More recently, Shepherd and Walsh (2002) developed a scheme for the development and use of soil spectral libraries for rapid estimation of soil properties based on analysis with this technique, using a library of over 1000 archived topsoils from eastern and southern Africa. Using a multivariate regression approach they calibrated 10 different soil properties to soil reflectance. Its real power is in providing prevalence data on soil problems, a basis for quantifying risk factors and a baseline for change detection (Shepherd and Walsh, 2007). Near infrared analysis is thus an indirect or secondary method that estimates chemical composition by comparing spectra with samples of known composition (Shenk *et al.*, 1992; Shenk and Westerhaus, 1994). This procedure is known as calibration. This evidence-based approach to soil survey is a departure from conventional soil survey approaches where policies and management recommendations are made for whole soil mapping units on the basis of a limited number of observations of soil properties in each mapping unit (Shepherd and Walsh, 2007).

Most NIR spectroscopy work on soils has focused on the prediction of individual soil properties, whereas the ultimate objective is often to predict soil functional capacity, the ability of soil to perform functions (such as support plant growth, nutrient cycling, hydrological regulation and resistance to erosion) or some integrated measure of soil degradation (Shepherd and Walsh, 2007). Most NIRS studies related to prediction of soil properties were limited in one way or another due to small sample size, less diversity in soil types, and/or because only a few soil properties were tested.

2.5 Application of NIR spectroscopy

While visible (350-700 nm), near infrared (700-2500 nm) and mid-infrared (MIR) (2500-25,000 nm) have been used in soil studies, NIR technology is well commercially developed and less technically demanding than MIR (Shepherd and Walsh, 2004). Further, NIR is well suited to field portability, remote sensing, copes better with moist samples and can deal with larger bulk soil samples because of its more intense sources and sensitive detectors (Merry and Janik, 2001). This study used near infrared (NIR) in the range 1250-2500 nm (4000-8000 cm^{-1}).

NIR spectroscopy has the potential for making schemes for monitoring soil condition over large areas practical. There are several advantages of IR spectroscopy technology for agricultural and environmental laboratories in developing countries. First, a laboratory can meet multiple analytical needs with just one instrument, which is simple to operate and maintain, robust and requires no chemicals; this reduces the need for infrastructure and advanced technical capacity and minimizes the risk of breakdown. Second, the method is highly repeatable and also far more reproducible than wet chemistry methods. This is an important advantage as there is currently no standardization of the methods used among different wet chemistry laboratories, which diminishes the potential value of data collected in any one laboratory. Third, the low cost and high throughput capability of IR spectroscopy enables objective, evidence-based approaches to agricultural and environmental management to be employed, which can speed reliable learning. (Shepherd and Walsh, 2007). Further, the rapid, low cost and reproducible nature of NIR spectroscopy measurements will enable large samples of resources and materials to be analyzed. As a result, statistical information will be built up for key indicators.

NIRS has therefore shown promise as a rapid low cost and reliable method for rapid assessment of soil properties (Brown, *et al.*, 2006; Shepherd and Walsh, 2002; 2004). It provides precise information on soil conditions more cheaply and is cost effective in time and finances. Therefore, robust NIR-based soil condition indices have potential to provide valuable information on soil constraints that could be transmitted to farmers via extension agents and other farm service providers.

Emphasis in NIR spectroscopy studies will be on development of spectral indicators that are interpretable in terms of suites of functional properties and symptoms, with less emphasis on development of calibrations for individual constituents. Associations between spectral indicators and functions will be made using probabilistic approaches that capture uncertainties in the associations, learn the associations from the data, allow continuous updating of data, permit inference in any direction and allow entry of partial evidence (Shepherd and Walsh, 2007).

2.6 Literature gaps identified

The ability to acquire high density spatial data on the quality of resources and materials linked to case definitions for specific problems, will lead to an emphasis on *interpretation* of measurements with respect to important functions and problems and *interpolation* of these results.

There has been some success with reflectance spectroscopy for sensing of soil organic matter in the field, and for the discrimination of major soil types from satellite multi-spectral and aircraft hyper spectral data (Shepherd and Walsh, 2002). Despite these indications of the potential of the technique, there are few examples and therefore gaps in literature on the application of reflectance spectroscopy for non-destructive assessment of soils.

Although geological spectral libraries exist that include soil mineral spectra, there are few examples of soil spectral libraries that include a wide diversity of soils with information on physical, chemical, and biological properties (Chang *et al.*, 2001). In particular there are literature gaps in the development of soil spectral libraries for application to risk-based approaches to soil evaluation that explicitly consider predictions and interpretations of soil properties.

Further, there exists a gap in the application of NIRS in the broad quantification of soil variability and the interpretation of this with respect to important problems and interpolation of the results for agricultural and environmental management. This approach constitutes a departure from current approaches in agricultural and environmental management, which tend to focus on detailed measurements of few samples and rely heavily on the use of *expert opinion* and *extrapolation* of results.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Study design

World Agroforestry Centre (ICRAF)'s soil library has data for approximately 80,000 soil samples, including 1,109 data for samples from archives of the Kenya Soil Survey. The library soils consists of topsoil (0-20 or 0-45 cm depth) samples taken from multilocation experiments, on-farm trials, and soil surveys conducted in Eastern Africa during 1993 through 1999 for which the physical and chemical soil properties had been analyzed by the soil laboratory of ICRAF (Shepherd and Walsh, 2002). However, many of the samples do not have complete and even data on the soil chemical and physical properties. The approach in this study started with examination of variation in spectral shapes, particularly the characteristics of the principal absorption features, and then interpreting this information in terms of implications for the different soil properties. In this study, 843 samples from different locations in Kenya with complete reference data were selected from the ICRAF soil library and matched with their spectra.

3.2 Sample selection

A master set of soil samples from Kenya stored in the ICRAF soil laboratory was identified for this study. The samples had complete data on soil physical and chemical properties (Table 3.1). Random stratified sampling method was used to select 843 soil samples which had earlier been collected from; Siaya, Yala, Amboseli, Central Kenya, Vihiga, Shinyalu, Teso and other parts of Kenya to be used in the study. Samples were from ongoing projects and studies that ICRAF has been involved in. These include many multilocation experiments, on farm trials and soil surveys. ICRAF also considers these areas as hot spots in the country in terms of soil degradation.

Soils had been taken from a wide variety of landscape positions, parent materials, and land uses. Although the soils were not formally classified on-site, they were sampled from areas broadly mapped as soil taxonomy orders; vertisols, andisols, aridsols, alfisols, ultisols, histosols, nitisols.

Each soil sample was analyzed for physical and chemical properties as shown in Table 3.1. This raw data was then transformed to present an equally probable distribution of soil variables for analysis. The transformations used for each soil property are also indicated in the table.

Table 3.1: Soil physical and chemical properties analyzed in the study

Soil property	**Units**	**Transformations used**
pH water (1:2.5 soil: water ratio)		Ln
Exchangeable Ca	$\mathrm{cmol_c\ kg^{-1}}$	Sqrt
Exchangeable Mg	$\mathrm{cmol_c\ kg^{-1}}$	Ln
Exchangeable K	$\mathrm{cmol_c\ kg^{-1}}$	Ln
Extractable P	$\mathrm{mg\ kg^{-1}}$	Ln
Organic C	$\mathrm{g\ kg^{-1}}$	Ln
Clay	$\mathrm{g\ kg^{-1}}$	Sqrt
Sand	$\mathrm{g\ kg^{-1}}$	None

3.3 Sample preparation

The soils were spread on a plastic tray and placed in the solar dryer for about 1 to 2 days until air-dry. The maximum drying temperature was 40°C. Air-dried soils were used for convenience and to minimize effects of variation in soil moisture on reflectance (Ben-Dor *et al.*, 1999). Each soil sample was mixed thoroughly and spread on a plastic sheet to dry. Using a rolling pin, the sample was crushed and passed through a 2mm sieve. Plant material and pieces of gravel were removed. Anything that remained on the sieve was discarded.

The sieved soil samples were labeled with the researcher's identification number and stored at room temperature and stored at room temperature for further analysis. The label consisted of a unique and consistent number based on batch number and laboratory identification number and was logged on sample forms/sheets.

3.4 Soil physical and chemical properties

Each of these soil samples had complete data on the physical and chemical properties to be used in the study. All laboratory soil analyses were conducted by the ICRAF soil laboratory using standard wet-chemistry methods.

3.4.1 Methods for determining soil physical/chemical properties

Soil samples were analyzed using standard methods widely used for tropical soils (Shepherd and Walsh, 2002). Methods for each component include;

- **Sand and Clay content** – The particle size analysis of a soil estimates the percentage sand, silt and clay contents of the soil and is often reported as percentage by weight of oven-dry and organic matter-free soil (Okalebo *et al*., 2002). After dispersing the soil into the individual particles (sand 2.00 – 0.05 mm, silt 0.05-0.002 mm, clay<0.002 mm), the hydrometer method for silt and clay measurements was used which relies on the effects of particle size on the differential settling velocities within a water column. The settling velocity is proportional to the square of the radius of the particle. Percentage of sand was then deduced from the silt and clay amounts.
- **pH** - The pH of the soil solution controls the form and solubility of many plant nutrients. Soil pH was determined in water using a 1:2.5 soil/solution ratio whereby 50 ml deionised water was added to 20±0.1g soil. The pH of the soil suspension was measured. The electro conductivity of the supernatant liquid was measured to identify soils that are potentially saline.
- **Exchangeable Calcium, Potassium and Magnesium** - Soil samples were extracted with an excess of 1 *M* NH_4OAc solution such that the maximum

exchange occurs between the NH_4 and the cations originally occupying exchange sites on the soil surface. The amount of exchangeable K was determined by flame photometry, and analyzed by atomic absorption spectrometry for exchangeable Ca and Mg (Okalebo *et al*., 2002).

- **Extractable Phosphorus** - Samples were extracted with 0.5 *M* $NaHCO_3$ + 0.01 *M* EDTA (pH 8.5, modified Olsen) using a 1:10 soil/solution ratio and analyzed colorimetrically (molybdenum blue) for extractable P (ISFEIP, 1972; Yurimaguas Experiment Station Staff, 1989). The Olsen method is suitable for a wide range of soil types and pH values and by precipitation reactions; there is an increase in P concentration in the solution, which is then measured.
- **Organic Carbon** – Through the dichromate oxidation method, after complete oxidation from the heat of solution and external heating, the unused or residual aqueous potassium dichromate ($K_2Cr_2O_7$) (in oxidation) was titrated against ferrous ammonium sulphate. The used potassium dichromate, the difference between added and residual $K_2Cr_2O_7$, gave a measure of organic C content in the soil. An additional method is provided where the amount of chromic Cr^{3+} ions formed during the oxidation process was determined colorimetrically to give total amount of organic carbon present in soil samples. The method was however suitable for soils with higher carbon contents (e.g.>2%).

  - Another method that was used to measure the organic carbon is the dry combustion method. It involved conversion of inorganic carbon to dissolved CO_2, and purging this from the sample. The remaining (organic) carbon was then oxidized at a high temperature to CO_2 which can be detected by a sensor and directly correlated to total organic carbon (TOC) content.

3.4.2 Statistical analysis of soil chemical properties

All the 8 soil properties (Table 3.1) in each of the 843 soil samples were first examined for normality of frequency distribution and transformed where necessary to a normal distribution. Soil nutrient concentration data is often characterized by a high proportion of low values and a small proportion of very high values, leading to a frequency distribution that is skewed to the left, and requires a square-root or log transform to correct the skew and obtain a normal probability distribution.

The Unscrambler software (ref.) was used in the transformation where the square root or log functions were used to transform the raw data to a normal distribution. Determination of whether to use the square root or log functions was based on the extent of the left or right skewness. The function, which gave the smaller extent of skeweness in units, was used. Sand was however normally distributed and there was therefore no need for transformation of its raw data.

3.5 NIR spectral metrics

The approach in this study started with examination of variation in spectral shapes, particularly the characteristics of the principal absorption features, and then interpreting this information in terms of implications for different soil properties.

3.5.1 Scanning

Each sample was properly mixed with a spatula to create homogeneity before loading it on the dish. The standard size Duran glass petridishes were used. These samples were packed in 12 mm deep, 55-mm diameter Duran glass petridishes. The petridishes were half-filled with the 2-mm air-dried sieved soil samples. Then the soil was mixed well and then several spatula-fulls were taken to half-fill the dish ensuring a flat soil surface. This was in order to get a representative sample.

The supplied forms were used to enter the laboratory identification number (Lab ID) and each petridish number is matched to the Lab ID. After loading all the dishes, they were arranged sequentially then scanned using the Fourier-Transform near-infrared spectrometer (Plate 3.1) and spectra recorded for each soil sample. After scanning a batch of samples, the quality of spectra was checked by switching to Bruker OPUS software (ref.) and reading all the spectra that have just been run. If there were any suspect-looking spectra, these were repeated.

Plate 3.1: Bruker Multipurpose Analyzer (MPA) NIR Spectrometer

Spectra were recorded from 3,585 to 12,493 cm^{-1} at a spectral resolution of 4 cm^{-1}, gap- filled to 2 cm^{-1}. The spectra were then trimmed/ reduced to the range 4,000 to 8,000 cm^{-1} using the OPUS software for spectral processing. This range on the electromagnetic spectrum represented the area having the sharp peaks of importance in the study. The OPUS software assists in calibrating and evaluating a model. It also analyzes multivariate data.

3.5.2 Peak picking

Peak picking identifies the peaks associated with prominent absorption features in a spectrum, their position and the intensity of these peaks. This is important because characterization of the spectrum determines the properties of a soil sample.

To do peak picking, the spectra was read into OPUS software. This was achieved by using the data table, selecting the spectrum for peak picking, then from the file menu in the software Unscrambler, export as ascii file and use the spectrum name as the file name when prompted to give the file name to be exported. The process was repeated until all the spectra required for peak picking was exported. Once all the spectral files were saved as ascii files, they were read at once into OPUS ready for peak picking.

The following methods for absorption feature characterization were tested:

i. The peak picking routine in OPUS was used to identify peaks (absorption features) in the spectra and the peak characteristics measured (absorbance intensity, relative intensity, i.e. peak height, peak full width at half-maximum (FWHM), and wavelength at peak maximum) (Figure 3.1). This routine fits a local baseline to the absorption features. Peak picking was first done before continuum removal (Figure 3.2). The corresponding wavelengths at the peak maximum were classified into absorption wave regions, which were used for extracting absorption feature parameters in this study. These absorption wave regions contain peaks close to the same region of the electromagnetic spectrum. Upto six wave regions were identified.

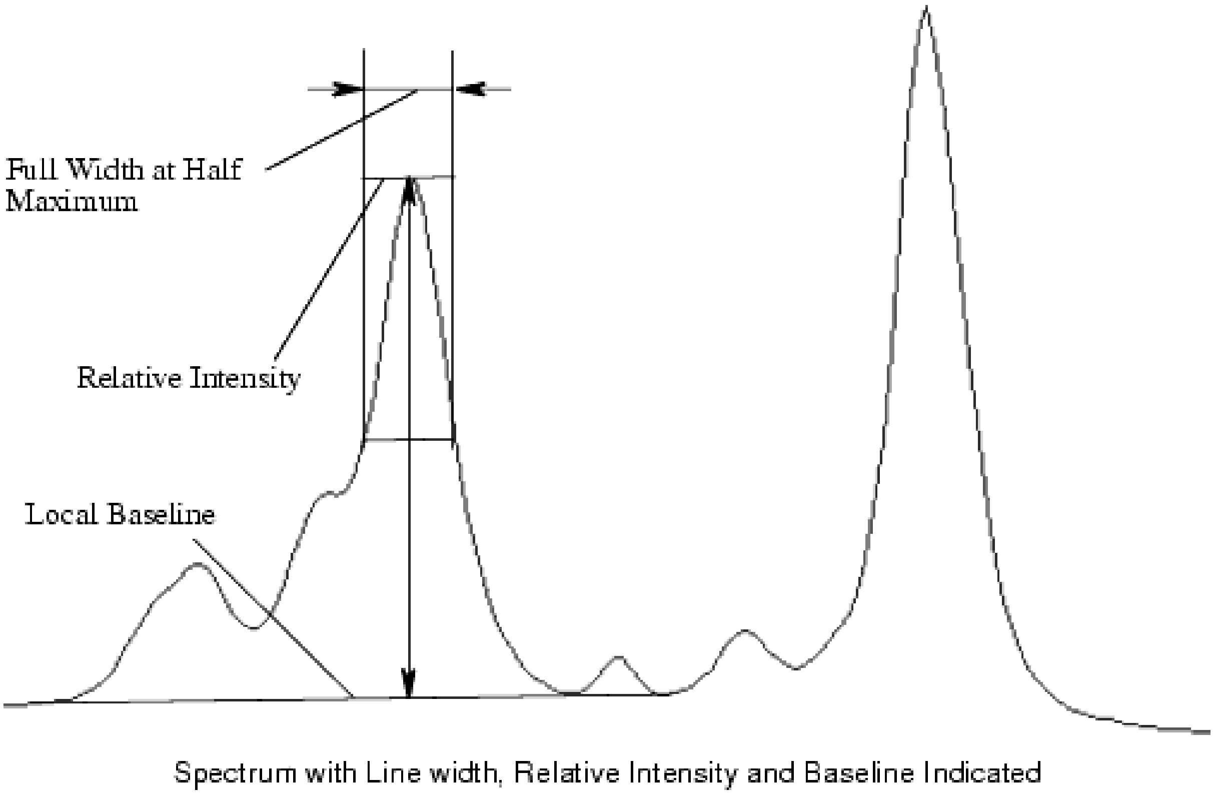

Figure 3.1: Spectrum peak and illustration of peak characteristics

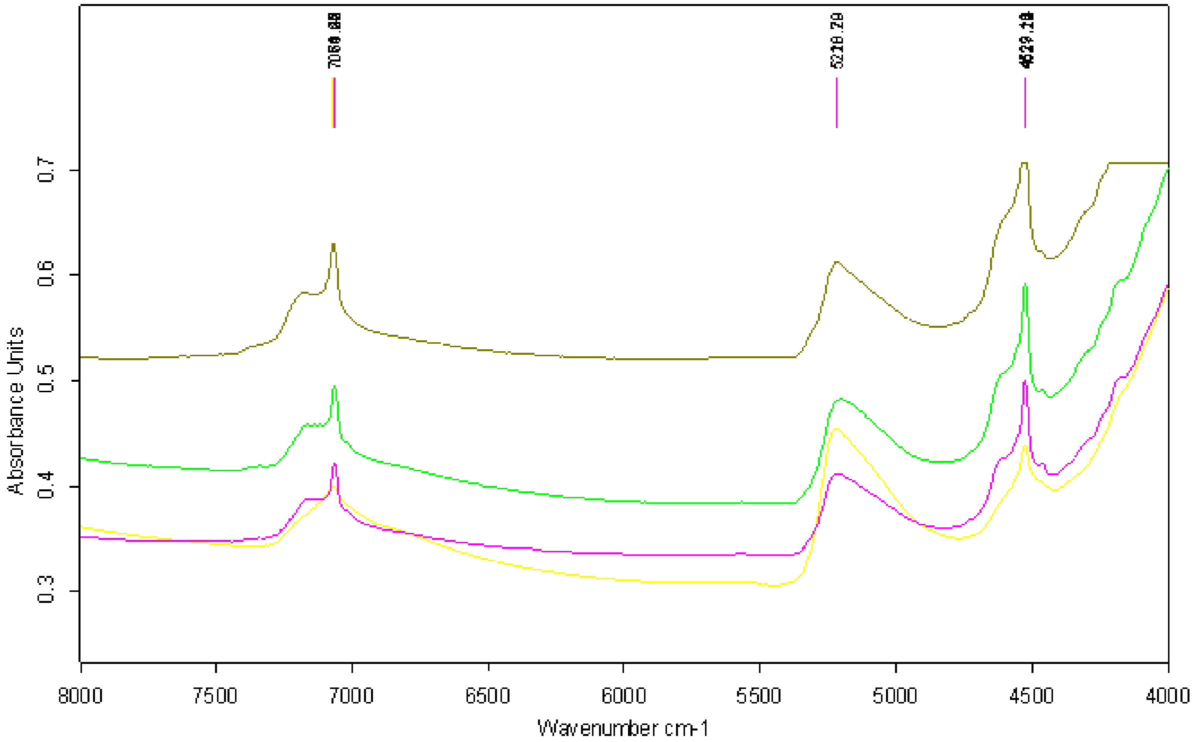

Figure 3.2: Spectra peaks identified for four different soil samples before Continuum Removal

ii. The peak picking routine was also applied after continuum removal, implemented in the Environment for Visualizing Images (ENVI) Version 4.3 software. Continuum removal is a means of normalizing reflectance spectra to allow comparison of individual absorption features from a common baseline (Figure 3.3). The continuum is a convex hull hit over the top of a spectrum utilizing straight-line segments that connect local spectra maxima. The absorbance spectra were first converted to reflectance spectra from continuum removal. The spectra are then exported as an ascii file that can be read by the software Unscrambler. After appending the spectrum names to the imported data table within Unscrambler, the continuum-removed reflectance spectra was transformed back into absorbance spectra. The data table was then transposed so that each column represented a spectrum. Up to seven absorption wave regions were identified.

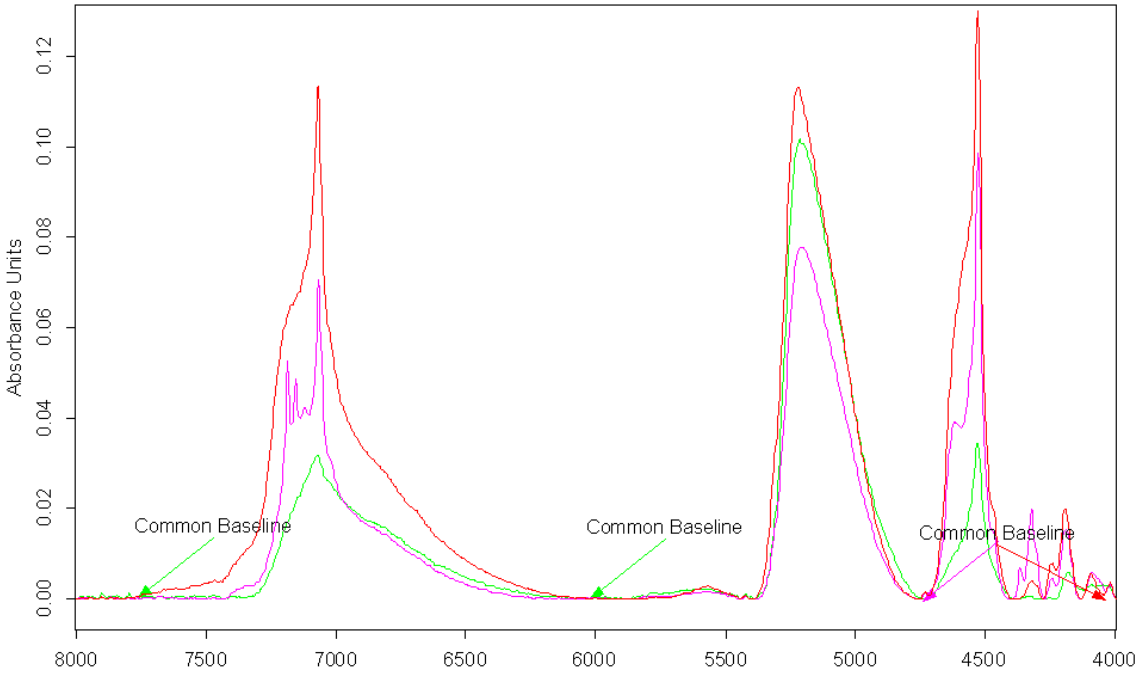

Figure 3.3: Spectra peaks identified for three different soil samples after Continuum Removal - Comparison of individual absorption features from a common baseline

3.5.3 Wave regions

The peaks on the spectral signatures of each soil sample are associated with the principal absorption features. They were classified into wave regions (Table 3.2) in which absorption feature parameters were extracted in this study.Waveregions are areas containing peaks that are close to each other along the electromagnetic spectrum.

Table 3.2: Spectral wave regions containing principal absorption features

Wave region	Wave number (cm^{-1})
1	0-4194
2	4195-4326
3	4327-4530
4	4531-5199
5	5200-5237
6	5238-7077
7	>7077

Wave regions containing spectral peaks that were present in each data set before and after continuum removal were identified as Wave regions 3, 5 and 6. The other peaks not found within these wave regions were excluded in the study. The samples containing the three wave regions were identified and the peaks corresponding to these wave regions were therefore the peaks containing the principal absorption features of importance in this study.

3.6 Relation between spectral metrics and soil properties

Soil properties are usually inter-correlated and can be reduced down to a smaller number of factors, to reduce the redundancy in the data, using principal component analysis. Principal components analysis (PCA) is a technique for simplifying a dataset, by reducing multidimensional datasets to lower dimensions for analysis while retaining those characteristics of the dataset that contribute most to its variance. It involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called *principal components (PC)*. The first principal component (PC01) accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. Principal component analysis is a linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (PC01), the second greatest variance on the second coordinate(PC02), and so on.

The scores of the fist two principal components of the soil properties usually account for more than 70% of the variation in soil properties.

3.6.1 Partial Least Squares regression method – peak region calibration and validation

Thus the relation between the first two principal components in turn and the spectral metrics were investigated using partial least squares (PLS) regression method which is a function of the Unscrambler software. These data sets were exported to the software and the spectral metrics were plotted against the two principal components. Transformed soil data was used in the regression.

PLS regressions were also developed for individual transformed soil variables per sample. The individual soil variables were calibrated against the absorbance intensity (spectral metrics) of each of the three wave regions per sample. The two soil data sets were used - before and after continuum removal. Partial least squares statistical method was used because of the high number of variables and the multicollinearity in the dataset. The spectral data was the independent factor while the physical/chemical soil data formed the predicted variables or the responses.

The results were models per individual soil variable and per PC used. Therefore, the calibration and validation determination coefficient (r^2), root mean squared error of calibration(RMSEC) and root mean squared error of prediction(RMSEP) were deduced from each model before and after continuum removal and laid out in a data set. In addition, the recommended optimal number of principal components explaining the distribution of each individual property was indicated against each property.

3.6.2 Quant 2 Method - Full spectrum calibration and validation

The relationship between the soil full spectrum was also investigated against individual soil properties. Using Quant 2 statistical method function of OPUS software), the full spectra of all the soil samples was correlated with the individual soil properties. 50% of the samples were randomly selected as the calibration data set while the other 50% was set aside as the validation set for the optimization process. The optimization process was carried out to specify the regions on the spectra that are important in the calibration and cross validation process.

After optimization, the whole soil data set was thereafter used as the calibration set, whereby the full spectra set was calibrated against each soil property. After calibration, this Quant 2 method was also able to validate the calibration models per soil property. Therefore, resultant were calibration models whose determination coefficient (r^2), Rank and root mean squared error of cross-validation (RMSECV) was recorded for the cross-validation process. The r^2 determined the percentage of variance present in the true component values, which is reported in the predicted concentration values. The r^2 approaches 100% as the predicted concentration values approach the true values. The r^2, Rank and RMSEC figures were also recorded for the calibration process.

The relation between the first two principal components in turn and the full spectra data was also investigated using this method. The full spectrum data was calibrated against soil PC 01 & 02 and the resulting models were validated per principal component. The models' determination coefficient (r^2), Rank, RMSEC and RMSECV were recorded for the cross-validation and calibration processes.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Soil physical and chemical properties

4.1.1 Sample characteristics

A total of 843 samples from various field locations in Kenya were sampled. Table 4.1 gives a summary of field locations and the distribution of soil samples per location. The number of samples represented the total samples that had complete soil physical and chemical reference data for each field location.

Table 4.1: Field locations and distribution of soil samples

Field Location	Number of Samples
Amboseli	19
Yala	42
Siaya	44
Meru South	32
Central Kenya	22
Vihiga	32
Shinyalu	155
Teso	25
Others	472
Total	**843**

There was high correlation between soil organic carbon (SOC) and non-acidified carbon as shown in Figure 4.1. Therefore, through inter-calibration these methods were transformed to one method to increase the size of the complete data set. The function $y=0.3759x + 0.0198$ was used to determine the SOC from the different methods used.

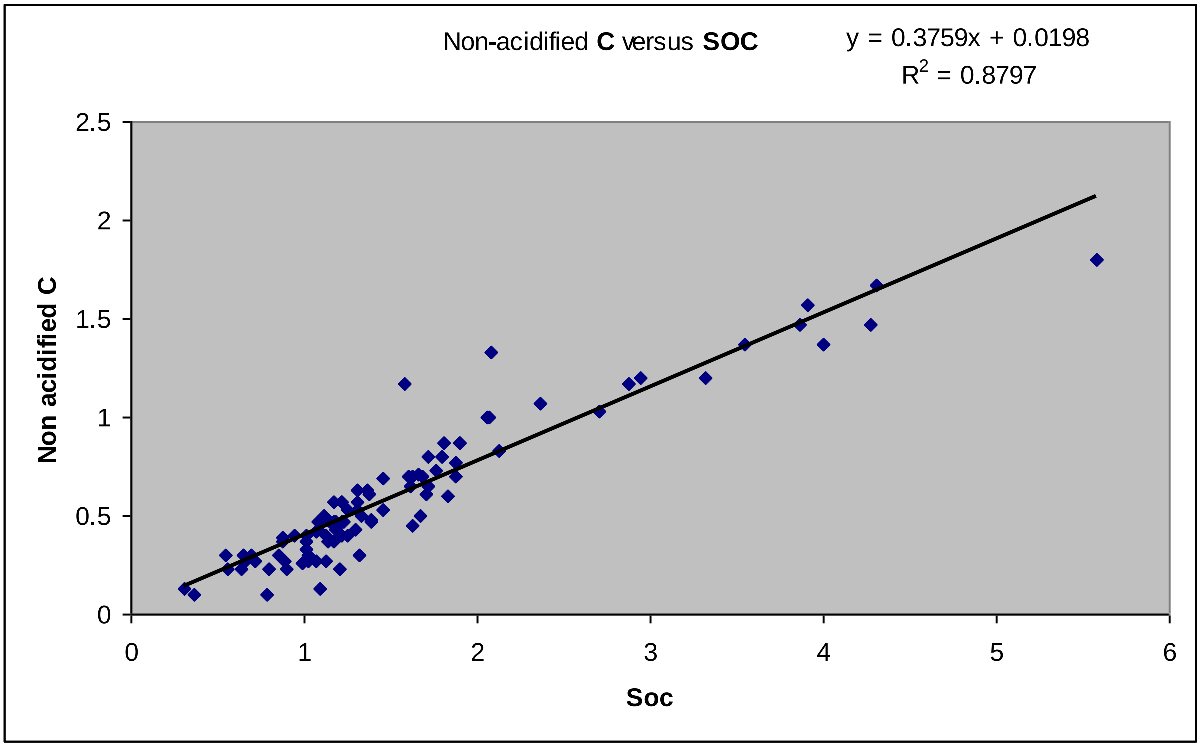

Figure 4.1: Non-acidified carbon versus soil organic carbon

4.1.2 Variation in soil physical and chemical properties data

The soil properties in the 843 soil samples did not have normal distribution and using the Unscrambler software, the square root or log functions were used to transform the raw data to a normal distribution. The function, which gave the smallest result, was used. Sand was however normally distributed and there was therefore no need for transformation of its raw data. The quantity of the transformed data was much smaller per property than the original soil property figures. The range was much lower for each property as compared to the untransformed data.

The variability of the data was explained using standard deviation (σ).Extractable phosphorus had the greatest change in standard deviation before and after transformation (24 before and 1.40 after). Table 4.2 shows the spread of distribution of soil physical and chemical data before and after transformation.

Table 4.2: Variation of soil data before and after transformation

Soil properties	Mean Before Transformation	Mean after Transformation	Standard Deviation Before Transformation	Standard Deviation After Transformation
Clay (g kg^{-1})	34.28	5.72	14.40	1.23
Sand (g kg^{-1})	42.58	42.58	18.44	18.44
pH	6.06	1.79	0.85	0.14
C (g kg^{-1})	4.24	0.24	6.91	1.46
Extractable P(mg kg^{-1})	9.97	1.16	24.00	1.40
Exchangeable K ($cmol_c$ kg^{-1})	0.49	-1.11	0.56	0.92
Exchangeable Ca ($cmol_c$ kg^{-1})	9.44	2.80	8.53	1.26
Exchangeable Mg ($cmol_c$ kg^{-1})	2.80	0.67	2.57	0.87

All the properties with the exception of exchangeable Potassium showed a lower measure of spread in distribution of the data set after transformation. The standard deviation in all the properties is small after normalizing the distribution because most of the values in the data set are close to the mean. Normalizing the distribution created a higher probability of using all the data in the variables for the analysis.

4.2 Near-infrared spectral metrics

Peak picking routine in OPUS was used to identify peaks (absorption features) in the spectra and the peak characteristics measured (absorbance intensity, relative intensity, i.e. peak height, peak full width at half-maximum (FWHM), and wavelength at peak maximum).

4.2.1 Peaks identified before Continuum Removal

Peak picking was first done before continuum removal (CR). Between one and three peaks were identified per soil spectrum. Figure 4.2 illustrates the number, nature of peaks and the corresponding wave numbers identified before CR.

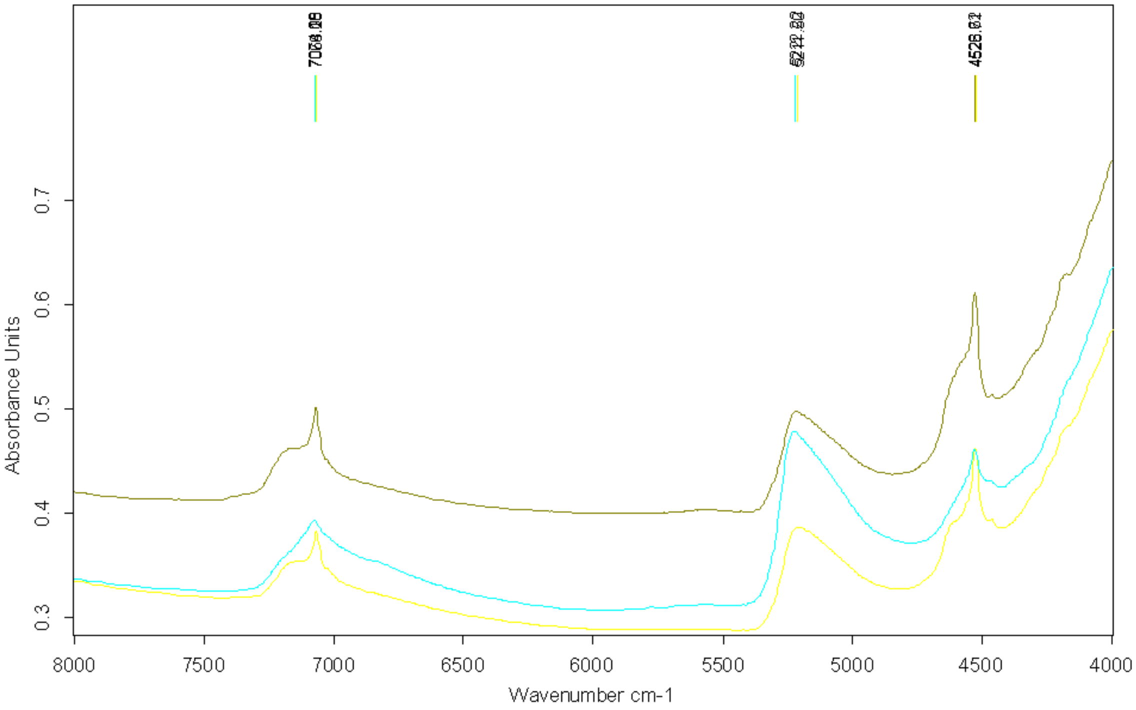

Figure 4.2: Near - infrared absorbance spectra and peaks identified for three soil samples before Continuum Removal

The corresponding wave numbers at the peak maximum were classified into absorption wave regions. All the peaks identified in the spectra were therefore included in these wave regions. Upto six wave regions were identified (Table 4.3).There were no peaks identified that fell within wave region 2

Table 4.3: Wave regions of soil spectra before Continuum Removal

Wave region	Wave number (cm^{-1})
1	0-4194
3	4327-4530
4	4531-5199
5	5200-5237
6	5238-7077
7	>7077

The peak characteristic parameters differed within the different wave numbers in each sample. They also differed from sample to sample and for each sample before and after Continuum Removal. For each of the wave numbers (peaks) identified the peak characteristics (the absorbance intensity, relative intensity and width of the peaks) were tabulated against each peak and sample. The characteristics of the peaks identified before continuum removal are shown in Table 4.4.

4.2.2 Peaks identified after Continuum Removal

After CR, the peak picking is done on the spectra from a common baseline unlike before Continuum Removal. This is illustrated in Figure 4.3 Hence, this technique enables in identifying more peaks that were not visible before CR. Resultantly, between one and five peaks were identified for most of the soil samples per spectrum. Table 4.4 is an illustration of the seven wave regions that all the peaks identified were grouped into.

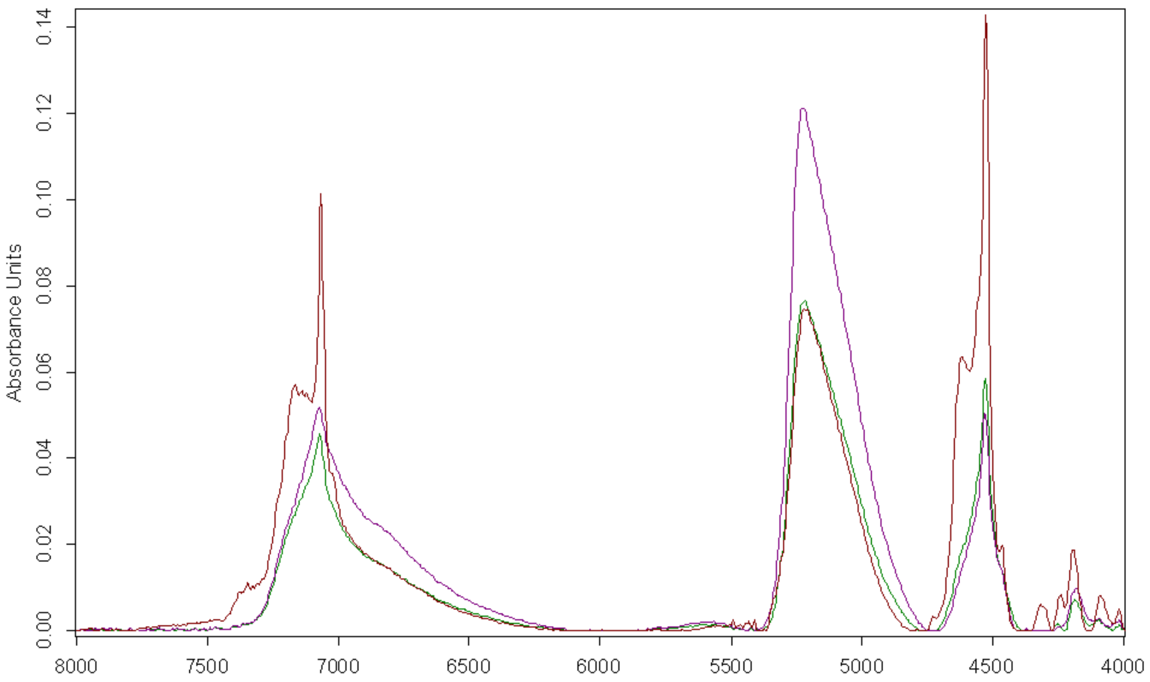

Figure 4.3: Near - infrared absorbance spectra and peaks identified for three soil samples after Continuum Removal

Table 4.4: Wave regions of soil spectra after Continuum Removal

Wave region	**Wave number (cm^{-1})**
1	0-4194
2	4195-4326
3	4327-4530
4	4531-5199
5	5200-5237
6	5238-7077
7	>7077

Table 4.5 represents the average and range values of the peak characteristics of the peaks identified before and after CR.

Table 4.5: Peak characteristics of the soil spectra before and after Continuum Removal

Peak characteristics	**Before CR**		**After CR**	
	Average	**Range(cm^{-1})**	**Average**	**Range**
Wave number(cm^{-1})	5586	2660	5581	2999
Absorbance Intensity	0.50	0.636	0.08	0.294
Relative Intensity	0.10	0.245	0.08	0.294
Width	164	223	175	567

The average values of the spectral characteristics are higher after CR in the case of the absorbance intensity and width. The average values of the wave numbers are almost equal before and after CR. However, the range of the peak wave numbers identified after CR increases (2999) as compared to before CR (2660).The ranges of the absorbance intensity and width of spectra also increased after CR. This NIR technique was able to pick peaks not visible before CR resulting to higher ranges in the absorbance intensity and width of spectra. The average and the range of the absorbance intensity and the relative intensity are similar respectively, after CR because the peaks are picked from a common baseline.

4.3 Relationship between the spectral metrics and soil properties

Relationships were developed between the soil spectra and the main variation in measured soil physical and chemical properties. This was done by calibrating the soil spectra against the soil physical and chemical properties resulting to calibration models. The robustness of the relationships in predicting soil properties was determined by validating the calibration models developed. This is in order to determine whether soil spectra can be a representation of the soil condition of an area.

There are many examples where several predictor variables used in combination can give dramatically better results than any of the individual predictors used alone. One rich source of such examples is NIR spectroscopy (Naes et *al.*, 2002). According to Change *et al* (2001), soil is a heterogeneous mixture of various chemical compounds, and thus a unique spectral response for soil properties is by no means certain. One method of evaluating the spectral response for a soil property is to study the relationship between wavelength, optical density, and the values of the soil property.

4.3.1 Relationship between the peak regions and the soil variables

The wave regions containing spectral peaks that were common in each data set in the study before and after CR were identified as Wave regions 3, 5 and 6.These were the peaks containing the principal absorption features of importance in the study. The other peaks not found within these wave regions were excluded in this section of the analysis. There were therefore two data sets, before and after CR. Resultantly, 210 samples were included in the analysis before CR and 766 samples included after CR. The relationships were investigated between the peak regions and the individual soil variables as well as the soil principal components.

4.3.1.1 Relationship between the peak regions and the individual soil variables

The spectral metrics were first plotted against the individual soil variables. These data sets were exported to Unscrambler software and the PLS regression method was used to calibrate the soil properties against the peak regions. The spectral metrics were calibrated against each soil variable before and after continuum removal. The results of the relationships that exist between the spectral metrics (peak regions) and individual soil variables are illustrated in Table 4.6.

4.3.1.2 Relationship between the peak regions and the soil principal components

The 843 samples in the study have been analyzed and constitute a large dataset of soil physical and chemical data. These soil variables are highly collinear with high correlations or other near linear relations among them. The explanatory variables were compressed onto a few linear combinations of the original variables before regression. Therefore using PCA, the data set of correlated soil variables was reduced to smaller uncorrelated variables called principal components. The first two principal components were used in this study whereby the first principal component explained 43% of the variability in the data set and the second principal component explained 25% of the variability.

The idea behind the partial least squares(PLS) and principal component regression (PCR) methods was to find a few linear combinations (components or factors) of the original *x*-values and to use only these linear combinations in the regression equation (Martens and Martens, 2001). In this way, irrelevant and unstable information was discarded and only the most relevant part of the *x*-variation is used for regression. A further idea behind the PCR method is to remove the factors with the least variability from the regression, to avoid instability of the predictions. The multicollinearity is therefore solved and more stable and reliable regression equations and predictions are obtained.

The spectral metrics were plotted against the soil PC 01 and 02 data. The results were correlation models per principal component, before and after continuum removal. The models in Figures 4.4(a-d) below are an illustration of relationships that existed between the spectral metrics (peak regions) and soil PCA 1 and 2 data each plotted before and after CR.

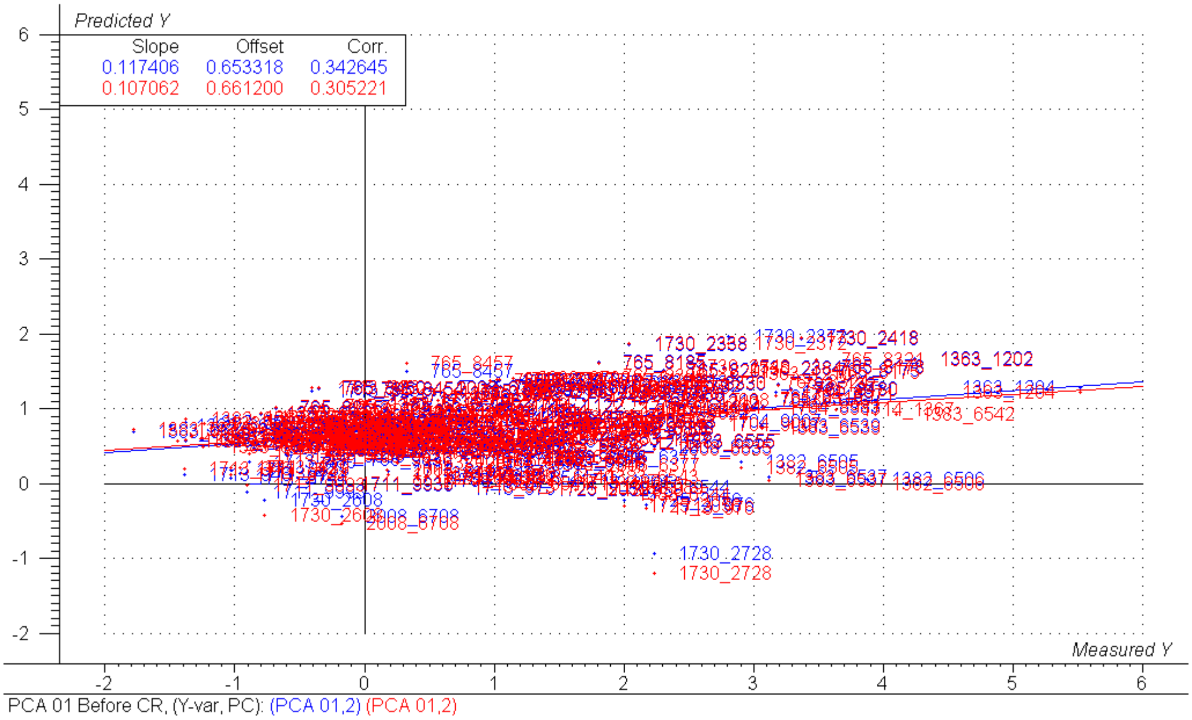

Figure: 4.4 (a) Calibration model of spectral peak regions against soil PCA 01 before CR

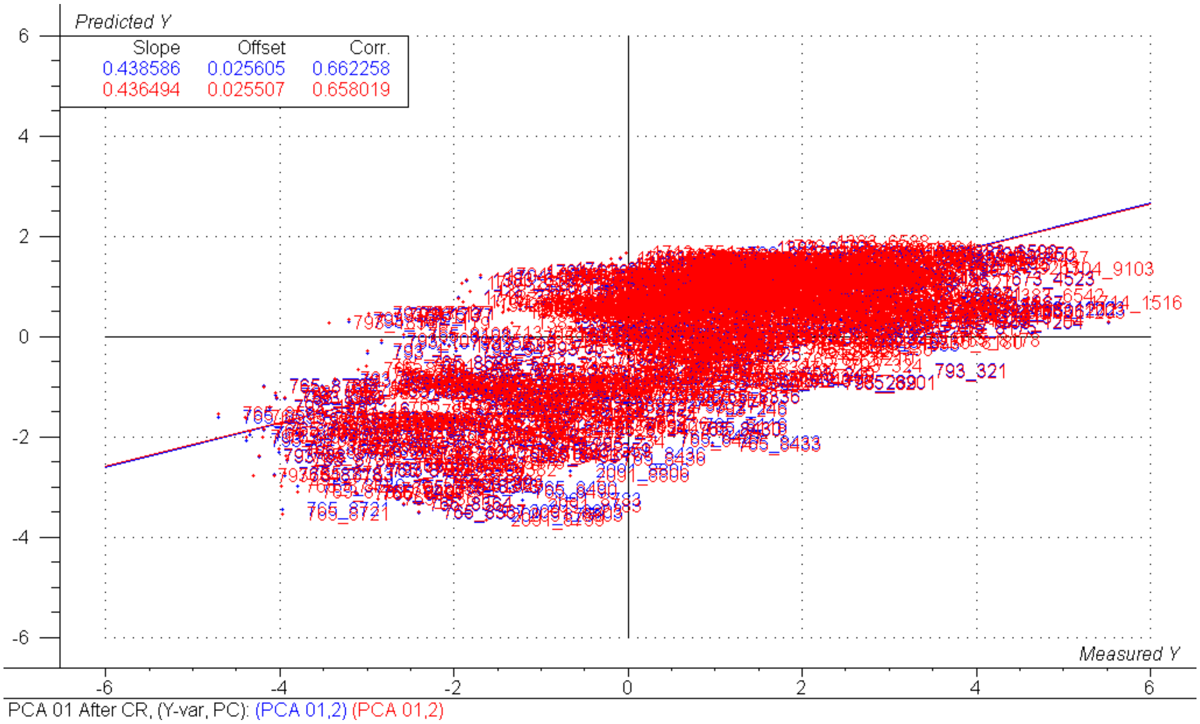

Figure: 4.4 (b) Calibration model of spectral peak regions against soil PCA 01 after CR

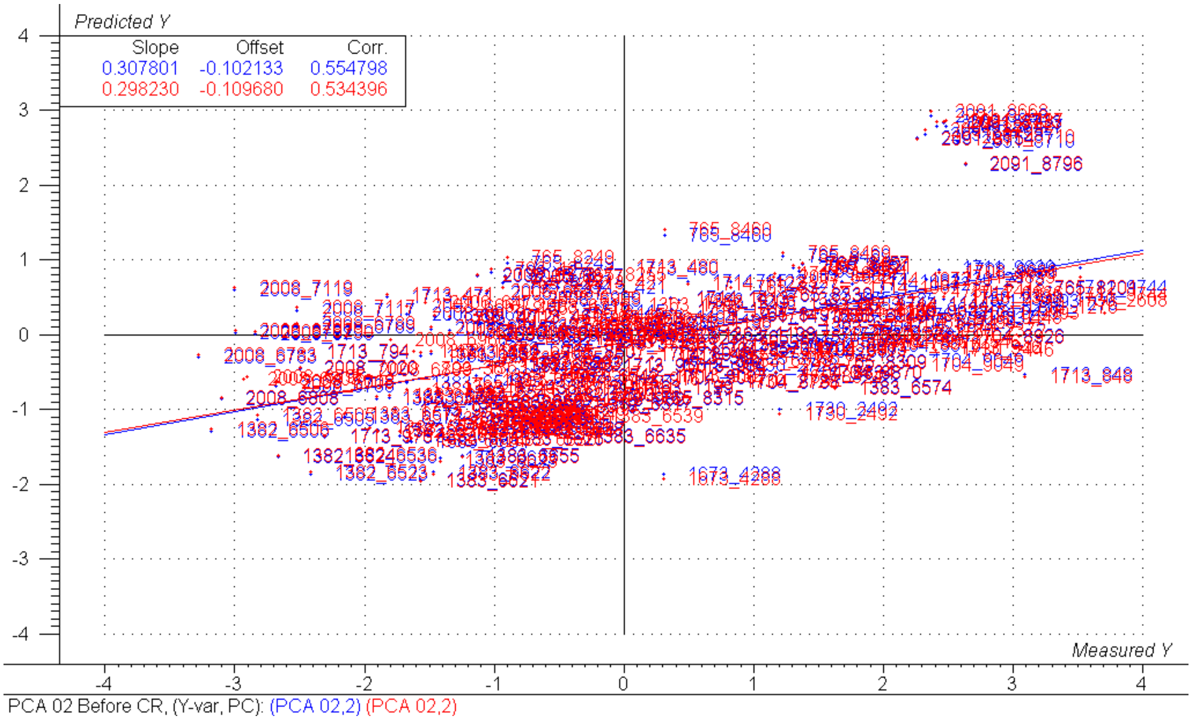

Figure: 4.4 (c) Calibration model of spectral peak regions against soil PCA 02 before CR

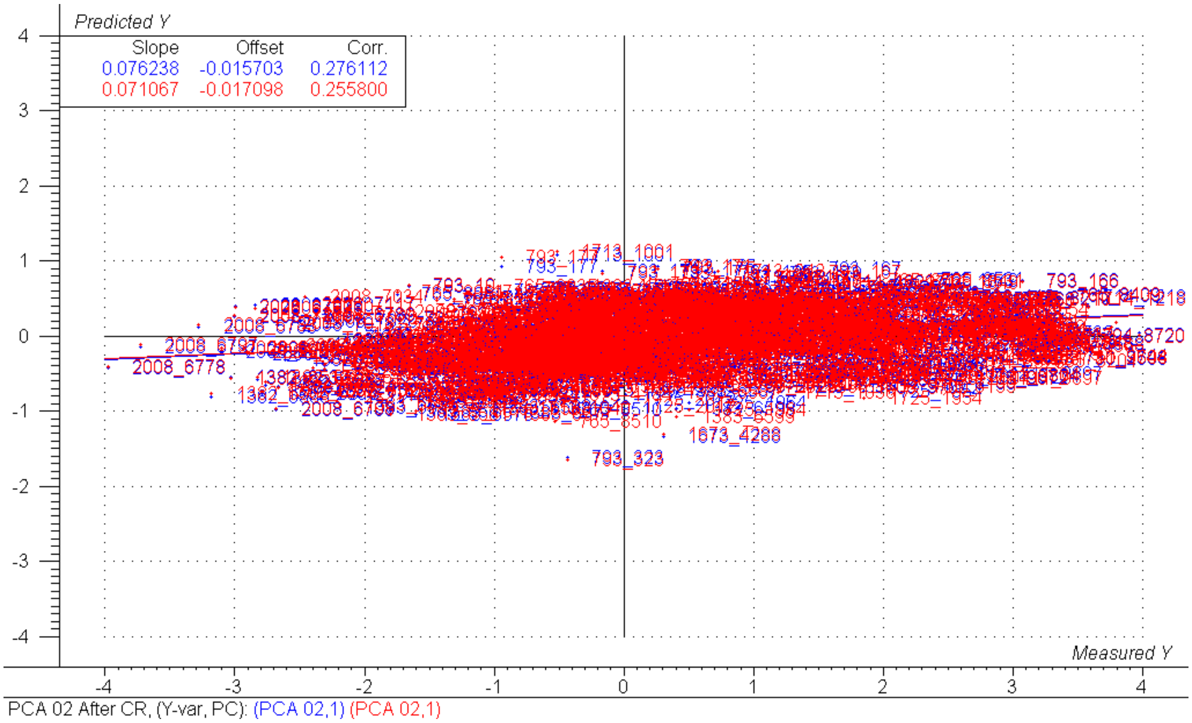

Figure: 4.4 (d) Calibration model of spectral peak regions against soil PCA 02 after CR

The calibration and validation r^2, root mean squared error of calibration(RMSEC) and root mean squared error of prediction(RMSEP) results were deduced from each model in the relationships with individual soil variables as well as soil principal component 01 and 02 data, before and after continuum removal. These were laid out in the Table 4.6. The calibration and validation r^2 explained the fit of the soil variables against spectral metrics as well as the robustness of these relationships in predicting soil properties. In addition, the recommended optimal number of principal components explaining the variability/distribution of each property was indicated against each property.

Table: 4.6 Results of the correlations of the individual soil properties and soil principal components with the peak regions of the spectrum

Property	AFTER CR					BEFORE CR				
	Calib. r^2	Valida. r^2	RMSEC	RMSEP	No. of spectra PC used	Calib. r^2	Valida. r^2	RMSEC	RMSEP	No. of spectra PC used
ExCa	0.601	0.598	0.798	0.801	2	0.094	0.073	0.636	0.644	1
pH	0.343	0.338	0.114	0.114	1	0.318	0.300	0.115	0.117	2
ExMg	0.299	0.292	0.716	0.72	2	0.111	0.076	0.69	0.704	2
C	0.262	0.225	1.278	1.311	2	0.167	0.141	1.369	1.391	2
Clay	0.240	0.234	1.078	1.082	1	0.269	0.233	0.941	0.965	2
ExK	0.137	0.132	0.87	0.873	2	0.091	0.069	0.942	0.955	1
ExP	0.049	0.042	1.357	1.362	2	0.234	0.206	1.12	1.14	2
Sand	0.097	0.093	17.473	17.517	1	0.283	0.260	15.213	15.453	2
PC 01	0.438	0.433	1.394	1.401	2	0.118	0.093	1.264	1.283	2
PC 02	0.076	0.066	1.371	1.379	2	0.308	0.285	1.341	1.363	2

As shown in the table 4.6 above, by using the individual soil properties, with the exception of Sand and ExP, all other soil properties displayed higher correlation and better fit with soil spectra after CR as compared to before CR. The effects of CR was to increase the number of absorption peaks identified on the spectrum, resulting to more spectral characteristics and information that was available to interpret the soil condition based on the soil property variation.

Therefore, the quality of relationships between spectrum and soil variation improved after CR. ExCa which is an important soil property had the highest change in correlation after CR (validation $r^2 = 0.073$ to $r^2 = 0.598$), and was the individual soil property having the best fit using this method. Sand, ExK and ExP however, had the poorest correlations with the spectral data. In previous studies, Janik *et al.* (1998) also observed poor prediction of bicarbonate-extractable K and P with mid-infrared analysis (validation $r^2 < 0.5$). There is evidence that the wet chemistry methods used to determine the parameters K and P have not been effective in extracting these parameters. Using the PCA data, PC 01 displayed highest correlation with the spectral data after Continuum Removal (validation $r^2 = 0.433$). PC 01 explained 43% of the total soil data variability and hence model had a better fit. It displayed a lower correlation before CR. As in the case of using individual soil properties, the model improved after CR of spectral data. PC 02 however displayed the inverse - a higher correlation before CR (validation $r^2 = 0.285$) as opposed to after CR.

Selecting the three peak regions omitted other spectral information that was vital in explaining the condition of the soil. Consequently, using the three peak regions as a method before or after CR did not provide robust relationships that could effectively provide means of assessing the soil condition due to the poor prediction coefficients (r^2).

4.3.2 Relationship between the full NIR spectrum method and the soil variables

Other workers (e.g., Palacios-Orueta and Ustin, 1998) have found that for the purpose of calibrating soil properties to spectral characteristics, it is preferable to use information over the entire spectrum, rather than attempting to interpret individual absorption features. Soil spectra result from overlapping absorption features of many organic and inorganic components, thus subtle differences in spectral shape may provide valuable information about soil properties.

This section describes the investigations of the relationships between the NIR full spectrum and the individual soil variables as well as the soil principal components.

4.3.2.1 Relationship between full NIR spectrum method and individual soil variables

Using Quant 2 statistical method of the OPUS software, the relationship between the soil full spectrum was investigated against individual soil properties. Each soil property was calibrated against the full spectra. After calibration, this Quant 2 method was also able to validate the models per individual soil property. Resultant were calibration models per soil property whose determination coefficient (r^2), number of principal components of the spectra and RMSEC was recorded for the calibration process. This method was able to detect the outliers and bring out the best models. The r^2, Rank and RMSECV figures were also recorded for the cross validation process. The r^2 determined the percentage of variance present in the true component values, which is reported in the predicted concentration values. r^2 approaches 100% as the predicted concentration values approach the true values.

4.3.2.2 Relationship between full NIR spectrum method and soil principal components

The full spectrum data was also plotted against the first two soil principal components in this study. Hence using PLS regression method, using the Unscramber software, PC 01 and 02 were calibrated against the soil full spectrum. The results were calibration models per principal component. The models in Figures 4.5(a-b) show the relationships that exist between the full spectrum and soil PC 01 and 02 data.

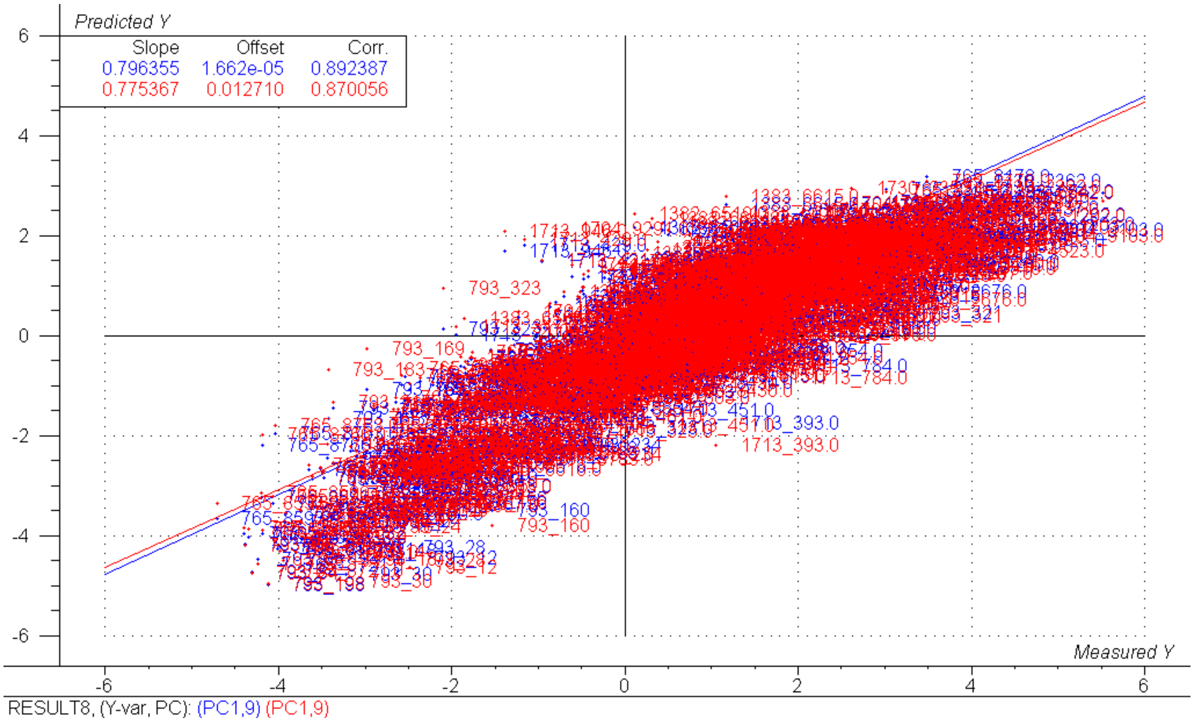

Figure 4.5: (a) Calibration model of soil PCA 01 data against full spectrum region

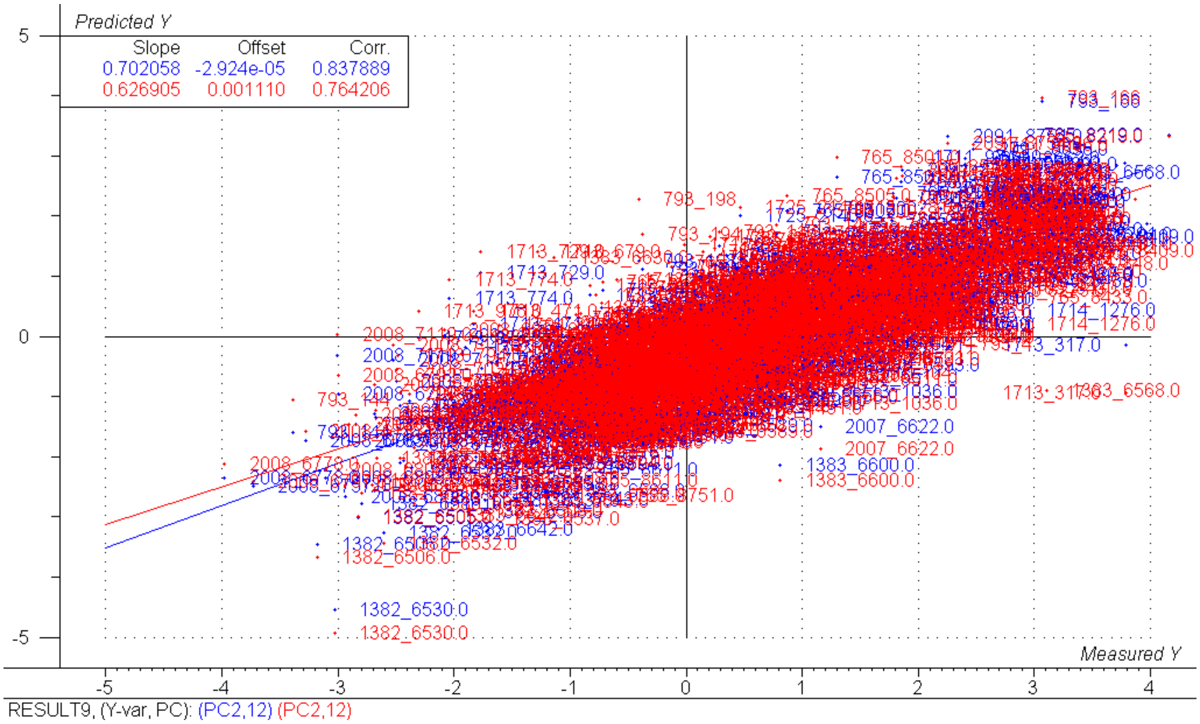

Figure 4.5: (b) Calibration model of soil PCA 02 data against full spectrum region

Table 4.7 illustrates the calibration and cross-validation results after calibrating individual soil properties against full spectrum as well as the soil principal components 01 and 02 against full spectrum. The calibration process is used to fit the models while the validation process is used to develop models for prediction of unknown values of the soil properties.r^2 is therefore higher in the calibration set as opposed to the validation set.

Table: 4.7 Results of the correlations of the full spectrum with the individual soil properties and soil principal components

Property	**Calibration r^2 (%)**	**Validation r^2 (%)**	**RMSEC**	**RMSECV**	**No. of spectra principal components**
ExCa	87.12	86.44	0.442	0.451	9
ExMg	75.1	73.69	0.403	0.412	9
pH	62.48	60.53	0.0809	0.0825	8
C	61.59	57.09	0.9	0.946	8
Sand	57.57	55.43	11.9	12.2	10
Clay	55.46	53.87	0.816	0.826	8
ExP	47.83	31.70	1.0072	1.160	N/A
ExK	45.96	43.58	0.647	0.657	9
PC 01	0.80	0.757	0.838	0.916	2
PC 02	0.702	0.584	0.772	0.916	2

The ability of NIRS to predict values of soil properties can be grouped into three categories; Category A includes properties with measured versus predicted r^2 values between 0.80 and 1.00 based on validation models. Category B includes soil properties with measured versus predicted r^2 values between 0.50 and 0.80. Category C consists of values $r^2 < 0.50$ (Chang *et al*, 2001). The authors believe that prediction of soil properties in Category B can be improved by using different calibration strategies, but properties in Category C may not be reliably predicted using NIRS.

As illustrated in the results in Table 4.7 above, using the relationship with individual soil data, root mean squared error of cross validation was larger for values of Sand (12.2) and ExP (1.16). Good calibrations (r^2 > 0.70), hence better model fits were obtained for soil exchangeable Ca and exchangeable Mg. Values of r^2 close to 100% indicate that NIR is almost as good as the laboratory technique against which it is calibrated. Sand (r^2=0.55), ExK (r^2=0.44) and ExP (r^2=0.32) however, had the poorest correlations with the spectral data. The same results were realized in the previous method of using the three wave regions which gave results for Sand (r^2=0.09), ExK (r^2=0.13) and ExP (r^2=0.04). As in the case with the peak region method, the NIR method was not good in predicting the soil ExP, which is the second most limiting nutrient of crop productivity in Sub Saharan Africa. The calibration models with the largest r^2 values for generalized cross-validation for a given attribute also resulted in the largest validation r^2 values, indicating that the cross-validation was effective in safeguarding against over-fitting.

It is apparent that the r^2 values improved after using the full spectrum method. For instance; ExCa - 0.86 with full spectrum and 0.60 with the peak regions; ExMg - 0.74 with full spectrum and 0.29 with the peak regions; Sand - 0.55 and 0.09 with the peak regions; pH – 0.61 and 0.34 with the peak regions. However, using the full spectrum method of calibrating the soil properties against full spectrum and validating the results as opposed to using the three wave regions did not effectively provide all the indicators of the soil condition, except when predicting ExCa(r^2 =0.86), ExMg(r^2 =0.74) and pH(r^2 =0.61).

Using the PCA data, PC 01 displayed a higher correlation with the full spectral data (validation r^2 = 0.757), as compared to PC 02(0.584). PC01 therefore had a better fit. Using all the methods illustrated, PC01 displayed the best model and fit by using the full spectrum method. Consequently, by relying on the inter-correlation among all the soil variables and fitting the soil principal components to the full spectra, the NIR analytical method was robust in developing effective indicators of soil condition.

These results can then be used as a basis for further research on ways of developing soil classification schemes based on soil spectra to enable diagnose large area soil nutrient health in a repeatable manner. The ability to rapidly and cheaply diagnose soil condition through NIRS will enable farmers and related institutions to monitor soil condition and develop appropriate farm based soil nutrient management programs.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

This chapter summarizes the main findings of the study in four sections as follows; Sections 5.1 summarizes the main findings of the study, while, Section 5.2 outlines some of the conclusions that can be drawn from the study. Recommendations of the study are stated in Section 5.3, whereas areas for further research work conclude the chapter in section 5.4

5.1 Main findings

In this study, spectral peaks to be used in the analysis were identified before and after CR. It was found that more peaks (upto five peaks) were identified after CR. The peaks were classified into seven wave regions. Spectra were therefore used for correlation both before and after CR.

The soil physical and chemical data set was large and highly collinear. The PCA method was able to reduce the dataset into fewer uncorrelated variables. A large part of the variability in the data set was explained by two principal components, which were used for the analysis. The first principal component accounted for 43% of the variability and the second component for 25% of the variability. This is reported in the method of calibration between the spectral metrics and the soil PC01 and PC02 data before and after continuum removal. Using PC01, validation r^2=0.433 after continuum removal, and using PC02, validation r^2=0.25 before continuum removal.

The effectiveness of NIRS technique in developing spectral indicators of multivariate soil properties in this study varied between the use of full spectrum and the use of the spectral peak regions. In the case of the PLS method (using spectral wave regions), better correlations were found after CR. The soil properties generally displayed higher correlations after CR. However in general, the wave region method gave poor correlations with the soil data and could not develop promising indicators of soil condition. ExCa was best predicted(r^2 = 0.598) by this method. All other properties were poorly predicted (validation $r^2 < 0.5$). Soil ExK, ExP and sand had the poorest correlations with the spectral data.

The use of Quant 2 method (using full spectrum), provided better predictions of individual soil properties as opposed to the use of the three peak regions, using the PLS method. Better models were developed using the Quant 2 method than the regression method and it gave higher calibration and validation r^2 values for all the models developed. E.g.; ExCa - 0.86 with full spectrum and 0.60 with the peak regions; Clay – 0.54 and 0.23

This signified the full spectrum method's higher prediction capability. However; this method was only able to predict effectively, ExCa(r^2 =0.86), ExMg(r^2 =0.74) and pH(r^2 =0.61), Clay (*r2* =0.54), Sand (*r2* =0.55) and C (*r2*=0.57) properties. As was observed in the PLS method, ExK(r^2=0.44) and ExP (r^2=0.32) had the poorest correlations with the spectral data. The full spectrum method further gave higher predictions when the soil properties were inter-correlated into PC 01 and 02. This method was effective in developing spectral indicators(r^2=0.76*)* for PC 01 soil data.

The peak height and width information appears not to have captured important variation in the shapes of the individual peaks, especially at the shoulders of the absorption features. The full spectrum Quant 2 method was able to employ this additional information and thus gave better predictions than using the peak information alone. The NIRS method of characterizing spectral data was effective in interpreting soil condition based on the soil property data.

5.2 Conclusions

For purposes of this study, PCA was effective in identifying the principal components that contained much of the soil property information required for analysis.

This study showed that when developing models to predict soil properties from infrared spectral data, the full spectrum method as opposed to restricting the models to using basic peak height and width information related to the key absorption features was more promising in diagnosing soil condition. By relying on the inter-correlation among all the soil variables and fitting the soil principal components to the full spectra, the study showed that the full spectrum method was most robust in developing spectral indicators of soil condition in Kenya based on the first principal component of the soil conditions.

Previous studies show that the soils in Kenya are highly variable in distribution, types and condition and soil science is becoming increasingly statistical in nature to help deal with this complexity. Near infrared spectroscopy has an important role to play in assessment of soil condition by enabling more intensive soil sampling schemes by virtue of the simple, rapid and cheap nature of the analytical method. It is important in monitoring the soil health and developing different soil management interventions.

This technique reduces the tedious, time consuming and costly process of soil analysis, especially for small-scale farmers. Frequent analysis can be carried out to determine whether their farming activities are degrading or helping in conserving the soil. In addition to the analysis, with further research, NIRS will give recommendations on how to improve depleted soils and boost agricultural productivity.

5.3 Recommendations

With the results from the study, the NIRS technique can be used for large area assessment of soil condition based on full spectrum method and the inter-correlation among all the soil variables. This would provide opportunities for development of management approaches in Kenya. It is therefore recommended that using Near-infrared spectroscopy, different calibrations are developed for the different soil types, properties, areas and studies.

On-going research on the use of the technique in Kenya and its applications in environmental management should take into consideration the low levels of awareness of the technique in the country and facilitate the transfer of this knowledge from the laboratory to the landowners, farmers, governments and other resource managers. This is for more effective decision making regarding monitoring land degradation, improving soil fertility and environmental management.

5.4 Further research areas

Based on the findings of this study, further research could examine ways of classifying spectra into groups and schemes based on whole spectra as an alternative pathway for identifying simple soil fertility spectral indicators. This would help to capture the complexity in the shapes of the absorption features (the mean reflectance, average absorption, width and peak height), while simplifying the complexity in the spectra, and provide a simple, first level screening and indication of soil fertility.

Opportunities for further research also exist in comparing the NIRS method with the use of other methods including the thermal method of assessing soil condition; hence assessing its robustness in diagnosing soil condition.

REFERENCES

Barrett CB, Place F, Aboud A and Brown DR (2002) The challenge of stimulating adoption of improved natural resource management practices in African agriculture. In: Barrett CB, Place F and Aboud A (eds). Natural Resources Management in African Agriculture. CAB International, Wallingford.

Ben-Dor, E, Irons JR, Epema GF (1999) Soil Reflectance. In: *Remote Sensing for the Earth Sciences: Manual of Remote Sensing*, Vol. 3; Rencz, N., Ed.; John Wiley & Sons: New York, NY, USA; 111–188.

Brady, N and Weil, R (1996) The Nature and Properties of Soils.Prentice-Hall, Inc.New Jersey, USA.

Britannica Online Encyclopedia (2007) Kenya: Soils. In: www.nationsencyclopedia.com/Africa/Kenya-AGRICULTURE.html

Brown D, Shepherd KD and Walsh MG (2006) Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma*, **132:** 273-290.

Camo Technologies (2003) Unscrambler Version 9.1 software package, Oslo.

Chang C, Laird DA, Mausbach MJ and Hurburgh CR Jr. (2001) Near-infrared reflectance spectroscopy-Principal components regression analysis of soil properties. *Soil Science Society of America Journal*, **65**:480–490.

Cohen MJ, Brown MT and Shepherd KD (2006) Estimating the environmental costs of soil erosion at multiple scales in Kenya using energy synthesis. *Agriculture, Ecosystems and Environment*, **114**: 249–269.

Dent A. and Young A. (1981) Soil survey and land evaluation. George Allen & Unwin Publ., Boston, MA.

Eswaran H, Almaraz R, van den Berg E and Reich P (1997) An assessment of the soil resources of Africa in relation to productivity. *Geoderma* **77**:1-18.

FAO (1988) Revised legend of the FAO-UNESCO soil map of the world. Food and Agriculture Organization of the United Nations, Rome.

Hazelton, P and Murphy, B (2007) Interpreting Soil Test Results: What do all the numbers mean? CSIRO Publishing, Oxford Street, Australia.

ISFEIP (1972) International Soil Fertility Evaluation and Improvement Program. Ann. Rep., North Carolina State Univ. Dept. of Soil Sci., Raleigh, NC.

Janik, L.J., Merry, R.H. and Skjemstad, J.O. (1998) *Australian Society of Agronomy Journal*, *Exp. Agric.* **38:**681-696.

Jones AJ and Doran JW (1996) Methods for Assessing Soil Quality. *Soil Science Society of America Journal*, Madison. Doran JW and Jones AJ (1996) Methods for Assessing Soil Quality. *Soil Science Society of America Journal*, Madison.

KARI (2001) Fertilizer Use Manual. Soil Fertility and Plant Nutrition Research Programme, Kenya Agricultural Research Institute, Nairobi.

Kinyanjui HCK, Obanyi SN, Onduru DD, Gachimbi LN, and Nandwa SM (2000) Stakeholder Perceptions of Agricultural Policies in Kenya. Managing Africa’s Soils No.18.International Institute for Environment and Development, London, UK.

Macharia P (2004) Gateway to Land and Water Information: Kenya national report. In: http://www.fao.org/ag/agL/swlwpnr/reports/ysf/z_ke/ke.html

Martens, H. and Martens, M. (2001) Multivariate Analysis of Quality. John Wiley & Sons, Chichester, UK.

Marten, H and Naes, T (1989) Multivariate Calibration. Wiley, Chichester, UK.

Merry RH and Janik LJ (2001) *Mid Infrared Spectroscopy for Rapid and Cheap Analysis of Soils. Proceedings of the 10th Australian Agronomy Conference.* Australian Society of Agronomy, CSIRO Land and Water, Urrbrae, South Australia.

Misiko M (2007) Fertile Ground? Soil Fertility Management and the African Smallholder.PhD thesis, Wageningen University, Wageningen, The Netherlands, 146pp.

Naes T, Isaksson T, Fearn T and Davis T (2002) A User-Friendly Guide to Multivariate Calibration and Classification. NIR Publications, Chichester, UK.

Nanni MR and Demattê JA (2006) Spectral Reflectance Methodology in Comparison to Traditional Soil Analysis. *Soil Science Society of America Journal,* **70**: 393 - 407.

Okalebo JR, Gathua KW and Woomer PL (2002) Laboratory Methods of Soil and Plant Analysis: A Working Manual.2nd Edition.TSBF and SACRED Africa, Nairobi, Kenya.

Okeyo JM (2006) Analysis of Spatial Variation of Soil Fertility Gradients in Vihiga and Siaya Districts of Western Kenya using Geostatistical Techniques. Master of Environmental Science thesis, School of Environmental Studies and Human Sciences, Kenyatta University, Kenya, 93pp.

OPUS Spectroscopic software, QUANT processing tool of Bruker Optik. Version 5.5

Osborne BG, Fearn T, Hindle PH (1993) Practical NIR spectroscopy with applications in food and beverage analysis. Longman, Singapore.

Palacios-Oreuta, A., and S.L. Ustin. 1998. Remote sensing of soil properties in the Santa Monica mountains. I. Spectral analysis. Remote Sens. Environ. **65**:170–183.

Reeves III JB, McCarty GW (2000) In: Davies AMC and Giangiacomo R (eds). *Near Infrared Spectroscopy: Proceedings of the 9th International Conerence*.NIR Publications, Chichester, UK, pp.587-591.

Research Systems (1999) ENVI User's Guide, ENVI Version 3.2.

Sanchez PA, Shepherd KD, Soule MJ, Place FM, Buresh RJ, Izac AN, Mokwunye AU, Kwesiga FR, Ndiritu CG and Woomer PL (1997) Soil fertility replenishment in Africa: an investment in natural resource capital.pp.1-46 In: Buresh RJ, Sanchez PA and Calhoun F (eds). Replenishing Soil Fertility in Africa, SSSA Special Publ.51. Soil Science Society of America and American Society of Agronomy. Madison, Wisconsin.

Sanchez PA (1995) Science in agroforestry. *Agroforestry Systems,* **30:**5-55.

Sanchez PA (1976) Properties and Management of Soils in the Tropics. John Wiley & Sons, Inc.New York.

Sanchez PA and Leakey RB (1997) Land use transformation in Africa: three determinants for balancing food security with natural resource utilization. *European Journal of Agronomy,* Vol.7.No 1:15-23.

Sanchez PA (1999) Natural resources management research: soil fertility replenishment in Kenya. *East African Agriculture and Forestry Journal,* Vol.65. No. 1-2.

Sanchez PA (2002) Soil fertility and hunger in Africa. *Science,* Vol. 295. No.5562: 2019-2020.

Shenk JS, Workman Jr JJ, Westerhaus MO (1992) Application of NIR spectroscopy to agricultural products. In: Burns DA, Ciurczak EW (eds) Handbook of near-infrared analysis. Marcel Dekker, New York, pp 383 – 431.

Shenk JS, Westerhaus MO (1994) The application of near infrared spectroscopy (NIRS) to forage analysis. In: Fahey GC Jr, Mossier LE, Mertens DR, Collins M (eds) National conference on forage quality evaluation and utilization. American Society of Agronomy, Madison, pp 406 -449.

Shepherd KD and Walsh MG (2002) Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal,* **66(3)**: 988-998.

Shepherd KD and Walsh MG (2007) Infrared spectroscopy - enabling an evidence based diagnostic surveillance approach to agricultural and environmental management in developing countries. *Journal of Near Infrared Spectroscopy, 15:1*-20.

Shepherd K (1995) Soil sampling and sample preparation- Technician training course series. ICRAF, Nairobi, Kenya.

Shepherd KD and Walsh MG (2001) Light reflectance provides rapid assessment of soil quality. Natural Resource Problems, Priorities and Policies Programme Working Paper 2001-1.International Centre for Research in Agroforestry, Nairobi, Kenya.

Shepherd KD and Walsh MG (2005) New Tools for Large Area Assessment of Soil Quality: Applications of Visible-Near-Infrared Reflectance Spectroscopy.

Shepherd KD and Walsh MG (2004). Diffuse reflectance spectroscopy for rapid soil analysis. In: Encyclopedia of Soil Science, edited By Rattan Lal. Online Published by Marcel Dekker.

Smaling EMA, Nandwa SM and Janseen BH (1997) Soil Fertility in Africa is at stake. pp.47-62 In: Buresh RJ, Sanchez PA and Calhoun F (eds). Replenishing soil fertility in Africa, SSSA Special Publ.51. *Soil Science Society of America*. Madison, Wisconsin.

Stenberg B, Nordkvist E, Salomonsson L (1995) Use of near infrared reflectance spectra of soils for objective selection of samples. *Soil Sci* **159**:109 – 114.

Stoorvogel, JJ and Smaling EMA (1990) Assessment of soil nutrient depletion in sub-Saharan Africa, 1983-2000.Rep.28.DLO Winand Staring Ctr.for Integrated Land, Soil a0nd Water Res., Wageningen, The Netherlands.

Stuart (2004) Infrared Spectroscopy: Fundamentals and Applications. John Wiley & Sons, Ltd.

Swift M.J, Stroud A, Shepherd K, Albrecht A, Batiano A, Mafongoya P, Place F, Tomich TP, Vanlauwe B, Verchot L and Walsh M (2004) Confronting land degradation in Africa: challenges for the next decade. In: Garrity DP, Okono A, Grayson M and Parrott S, (eds.). World Agroforestry into the Future. World Agroforestry Centre (ICRAF), Nairobi, pp 43-51.

Swift MJ (ed.) and Shepherd KD (2007). Saving Africa's soils: Science and Technology for improved soil management in Africa. ICRAF, Nairobi, Kenya, 18p.

Trimble SW and Crosson P (2000). *Science Magazine*, **Vol. 289,** No.5477:248-250.

United Nations Environment Programme (UNEP) (1997) World Atlas of Desertification.2nd ed. Arnold, London.

United Nations Millennium Project (2005).In: http://www.unmillenniumproject.org/index.htm

Vågen TG, Shepherd KD, Walsh MG (2006) Sensing landscape level change in soil quality following deforestation and conversion in the highlands of Madagascar using Vis-NIR spectroscopy. *Geoderma* 133(2006) p.281-294.

Verchot LV, Place F, Shepherd KD and Jama B (2007) Science and technological innovations for improving soil fertility and management in Africa: a report for the NEPAD Science and Technology forum. World Agroforestry Centre (ICRAF), Nairobi, Kenya: ICRAF Working Paper no. 31.

Vlek PLG (1995) The soil and its artisans in sub-Saharan African agriculture. *Geoderma*, Vol. **67**:165-170.

Woomer PL, Karanja NK and Okalebo JR (1999) Opportunities for improving integrated nutrient management by smallholder farmers in the central highlands of Kenya. *African Crop Science Journal,* Vol.**7**.No.4: 441 – 454.

World Agroforestry Centre (2006).Sensing Soil Condition: Infrared Diagnostics for Agriculture and the Environment. In: (http://www.worldagroforestry.org/sensingsoil/)

World Development Report 2008:Agriculture.In: www.worldbank.org/wdr2008

World Resource Institute, United Nations Environment Programme and the World Bank (2002). World Resources: 2000 – 2001.People and Ecosystems. The Fraying Web of Life. World Resource Institute, Washington DC.

Yurimaguas Experiment Station Staff. 1989. Soil and plant methods used at the tropical soil laboratory, Yurimaguas. pp. 163–175. *In* B. Hintze (ed.) First training workshop on acid tropical soil management and land development practices. International Board for Soil Research and Management technical notes 2. International Board for Soil Research and Management, Bangkok, Thailand.

APPENDICES

APPENDIX A1: SOIL DATA CHARACTERIZATION

Field Characteristics							Soil properties(units)							
Sample	**Study**	**Batch**	**LabID**	**Site**	**Plot Code**	**Depth**	**Clay**	**Ca**	**K**	**Mg**	**P**	**pH**	**Sand**	**C**

APPENDIX A2: SOIL LOGIN SHEET

Batch			
Batch Title			

Series			
Sample Type			
Title			
Login Date			
Responsible			
No. of samples			
Scientist			
Location			
Site			
Sample Descriptor	**Lab ID**	**Sample ID**	**Sampling Date**

APPENDIX A3: PLS REGRESSION DATASET

Principal wave	Soil chemical data(units)

Regions(Absorbance)											
Sample	Wave Region 3	Wave Region 5	Wave Region 6	sqrtClay	sqrtExCa	lnExK	lnExMg	lnExP	lnpH	Sand	lnC

APPENDIX A4: QUANT 2 CALIBRATION DATASET

Calibration data set	Soil chemical data(units)

Sample	Spectral absorbencies	sqrtClay	sqrtExCa	lnExK	lnExMg	lnExP	lnpH	Sand	lnC