Diagnostic Screening and Prevalence Assessment of Soil Stability Related Problems Using Infrared Spectroscopy in Lake Victoria Basin, Kenya

Bernard Kariuki Waruru

A thesis submitted in partial fulfillment for the Degree of Doctor of Philosophy in Soil and Water Engineering in the Jomo Kenyatta University of Agriculture and Technology

2016

DECLARATION

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

Signature ---------- Date----------

Bernard Kariuki Waruru

This work has been submitted for examination with our approval as University supervisors.

Signature ---------- Date----------

Prof. George M. Ndegwa

JKUAT - Kenya

Signature ---------- Date----------

Dr. Keith D. Shepherd

World Agroforestry Centre-Kenya, ICRAF

Signature ---------- Date----------

Dr. Peter T. Kamoni

KALRO - Kenya

DEDICATION

I dedicate this work to my late parents Geoffrey Waruru and Esther Wangu and late siblings Peter Maina and Geoffrey Githinji.

ACKNOWLEDGEMENTS

A long sometimes bumpy journey has now reached a crucial milestone. On the outset I declare that this has taken the Hand of The Almighty. I do acknowledge with appreciation my employer KALRO for scholarship award through the Kenya Agricultural Productivity Project (KAPP) and a generous study leave extension. Additional financial support was kindly provided by the World Agroforestry Centre (ICRAF) Science Domain 4 – Land Health, courtesy of Dr. Keith Shepherd. The JKUAT, especially the Director, Board of Post-Graduate Studies (BPS), the Dean, School of Biosystems and Environmental Engineering (SOBEE), and Chair at the Soil Water and Environmental Engineering Department (SWEED) for guidance patience and trust that has allowed successful completion of this work. I acknowledge my university supervisors Professor G. M. Ndegwa, Dr K. D. Shepherd and Dr. P.T. Kamoni for their academic support and friendship. Professor Ndegwa for the trust and patience, understanding, and overall guidance during this work. Keith Shepherd has provided the mentorship, availed without restriction all the necessary resources (physical, technical, intellectual, valued time). Thank you Andrew for your patience while initiating me in multivariate data handling, storage and manipulation. Special thanks to the following persons who provided invaluable support during data collection: Joash Mango and ICRAF Kisumu field team during fieldwork; Paul Mutuo during aggregate wet-sieving analyses; Ms Akinyi at Materials Testing and Research Laboratory during soil mechanical analyses; KPC Rao at ICRAF with soil wet chemistry analyses, and; Elvis Weullow for soil processing and spectral measurements. My special gratitude is sent to the entire ICRAF community, especially colleagues at the Library and Information Services and at the Soil and Plant Spectral Diagnostics laboratory for "looking out for me". I thank my friends and colleagues at KALRO-Kabete for all manner of support (empathy, sympathy, encouragement). I am indebted to Ann Wamuongo who out of cheer goodness contacted me about the fast closing opportunity for scholarships at KALRO when I was away on official assignment. Without that call this study was not to be. Finally, heartfelt appreciation goes to my spouse Louise Kariuki and daughters Gina, Eva-Sheila, Sheila and Rosa and 'gacucu' Maya Mumbi for emotional support. Louise I knew you fasted and steadfastly prayed for us, especially during the most trying times. My pledge is that your love and devotion is not in vain.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF PLATES

LIST OF SYMBOLS, ACRONYMS AND ABBREVIATIONS

AfSIS - Africa Soil Information Service;

AS - aggregate stability (soil aggregation and aggregate stability);

CART - classification and regression tree;

CV - coefficient of variability;

DRIFTS - diffuse reflectance infrared Fourier transform spectroscopy;

eNa - exchangeable sodium content;

H - Mahalanobis distance (for detection/ definition of spectral outlier);

HB - Homa Bay sentinel site;

IR - infrared spectroscopy;

LDSF - land degradation surveillance framework;

LNY - Lower Nyando sentinel site;

looCV - leave-out-one cross validation;

LS - least square (model testing);

LVB - Lake Victoria Basin;

mame - stable macro aggregates from slaking plus mechanical shaking wet- sieving pretreatment; ;

masp - stable macro aggregates from slaking only wet-sieving pretreatment;

mechR - **R**atio of stable aggregate fractions from **mech**anical disruption (mame: mime);

mime - stable micro aggregates from slaking plus mechanical shaking;

MIR - mid-infrared spectral range (2500 - 25000 nm);

MIRwc - mid-infrared wavelet transform coefficients;

misp - stable micro aggregates from slaking only wet sieving pretreatment;

Na2.5 - sodium concentration (Na^{+}) in 1: 2.5 soil-water extract;

Na5	-	Na^+ in 1:5 soil-water extract;
NIR	-	near infrared spectral range (700 - 2500 nm);
NIRwc	-	near infrared wavelet transform coefficients;
PC	-	principle component;
PCA	-	principal component analysis;
pHw2.5	-	pH in 1:2.5 soil-water extract;
pHw5	-	pH in 1: 5 soil-water extract;
PLSR	-	partial least square regression (PLSR);
PTF	-	pedotransfer function;
r	-	Coefficient of correlation (pair-wise);
R	-	Pearsons' correlation coefficient
$\boldsymbol{R^2}$	-	coefficient of determination (r^2);
RE	-	relative error;
RMSE	-	root mean square error;
RMSECV	-	root mean square error of cross validation;
RMSEP	-	root mean square error of prediction;
RPD	-	ratio of prediction deviation (ratio of SD to RMSECV/ RMSEP);
SD	-	measurement standard deviation;
SE	-	measurement standard error;
SOC	-	soil organic carbon (OC, organic carbon);
sponR	-	**R**atio of stable aggregate fractions from **spon**taneous disruption (masp: misp);
TN	-	terminal node;
unme	-	unstable fraction (< 20 μm) from slaking plus mechanical shaking wet sieving pretreatment;
unsp	-	unstable fraction (< 20 μm) from slaking only wet-sieving pretreatment;

WSA	-	wet water stable aggregation
WDC	-	water dispersible (natural) clay

ABSTRACT

Application of infrared spectroscopy (IR) in soil studies is well established. However, there has been little focus on examining IR for soil stability pedotransfer purposes. This study aimed to evaluate the use of IR in diagnosing soil stability related problems and assessing their prevalence with a case study of Lake Victoria Basin (LVB) in Kenya. Specifically to develop alternative IR-based models for screening soil stability related properties and compare these with predictions using conventional soil properties, validate the IR-based models using independent datasets, assess prevalence of stability related problems in two sentinel sites, and assess indices of soil stability functional attributes most appropriate for screening stability problems using IR-based models. Two samples sets representing different soils were used for the study. A model calibration set (n = 136) was obtained following a conditioned Latin hypercube sampling, and a validation set (n = 120) using spatially stratified random sampling. Spectral measurements were obtained for air-dried (< 2 mm) and for ground (< 0.5 mm) soil sub-samples using multipurpose analyzer and Tensor 27 spectrometers for near infrared (NIR) and mid-infrared (MIR) ranges, respectively. Soil laboratory data were also obtained for wet aggregation indices (WSA): macro, micro and unstable fractions from two different wet-sieving pretreatments. Soil properties were screened for prediction of the WSA using Classification and Regression Tree analysis. The WSA were calibrated to the soil-based predictors and to smoothed first derivative spectra and to spectra wavelet transform variables using partial least squares regression (PLS). WSA threshold values developed using soil predictors and spectra were used to diagnose soil stability related problems and assess their prevalence in Lower Nyando (LNY) and Homa Bay (HB) sentinel sites. Key soil predictors were: soil organic carbon and pH water (macro), water dispersible clay (micro) and exchangeable sodium (unstable). Coefficient of determination (R^2) for full cross validation PLS and IR-methods was: 0.6 (macro and unstable); 0.4 (micro)

fractions. The R^2 for soil-based PLS was: 0.3 (micro); 0.5 (unstable). Independent testing of IR-methods gave R^2 and RPD (ratio of prediction deviation) (R^2, RPD) as follows: macro (0.81, 1.4); unstable (0.65, 1.1). The LNY and HB sites indicated 66 % low stability prevalence, and 80 % of the sites were at risk. The study showed that IR-based predictors are superior over soil properties for stability transfer purposes. That 70 to 80 % of the soils in the sites had low stability problems and the risk of stability related problems increased with soil depth. Soil wet stable macro aggregate at10 and 50 %, stable micro at 20 and 40 %, and unstable fraction at 70-65 and 20-40%, define low and high stability, respectively. The models developed can be used to diagnose and assess prevalence of soil stability related problems in the LVB in Kenya and other regions. Further efforts should, however, widely test similar soil property predictor sets, aggregation indices and IR to: (i) validate established soil-based predictors, (ii) counter variability from sample provenance for improved model geographic transferability, and (iii) assess suggested performance improvement with calibration spiking.

CHAPTER ONE
INTRODUCTION

1.1 Background

The requirements of agricultural and environmental sustainability have dramatically redefined soil quality or soil health from the traditional view of soil quality as measured by soil performance and productivity. Soil health is currently defined as the capacity of a soil to sustain ecosystem services, including provisioning, regulating, supporting and cultural services. The soil health paradigm (theory, methods and practice) in the temperate region was developed to establish standards in a bid to address air and water quality issues arising from large nutrient and energy inputs to agricultural lands. In the tropics the main concerns for soil health are food insecurity, rural poverty and ecosystem degradation including accelerated soil erosion, depletion of soil organic matter, soil nutrients, and the deterioration of soil structure. It has been estimated that the degraded area of the world's arable land increased from 10 % in the early 1970s to about 40 % in the early 1990s, and that globally more than 1.0 million ha of agricultural land is annually lost due to human induced soil degradation. This calls for remedial measures and that the greatest need for remediation is in the developing regions of the world, where the rate of loss of agricultural usable land has been estimated at 0.3 % per year (Farley, 2012; Bouma, 2010).

It has been hypothesized that in the new Millennium, a renaissance is taking place in soil science as a response to the current threat posed by global environmental degradation, climate change and world-food production. A renewed interest in agriculture (food, fibre, feed, and fuel), erosion, nutrient depletion and pollution, for example, has brought soils back onto the global research agenda and evoked an aggressive search for up-to-date soil information. A whole set of new techniques and methodologies are evolving particularly in developed economies. Soil resources information acquisition has been

accomplished nationally at large scale where, for example, Western Europe has coverage at 1: 100,000 or larger. In terms of capability and capacity Africa may not have joined the hypothesized soil renaissance where land resources inventory have hitherto followed conventional methods. Most countries in Africa have no country-wide soil information (Sanchez *et al*., 2009; Hartemink & McBratney, 2008; Swift & Shepherd, 2007).

In Kenya soil survey at country wide coverage is at a scale 1:1 million, 15 – 30 % coverage is at between 1:1,000,000 and 1:250,000 and 10-15 % at larger scale. Soils with stability related problems (including sodic and saline-sodic) are estimated to cover 30 - 40 % of the country's land area. However, there is currently no system for direct measurement and monitoring of soil functional capacity. Soil and water resources are especially significant in Kenya because agriculture remains the mainstay of the economy. Agricultural production has, however, been constrained by prevailing agro-ecological conditions where more than four-fifths of the total land surface is arid to semi-arid (ASAL). The ASAL constitute rangelands where over 25 % of the total human population, more than 50 % of livestock, and about 80 % of the total wildlife derive their livelihood. The lowlands within Lake Victoria Basin (LVB) of Kenya occur in the ASAL and are characterized by extremely deep, unconsolidated, stratified, sodic, fragile, eroding soils, rural poverty and low agricultural productivity (Waruru *et al*., 2003a; Swallow *et al*., 2001).

Sustainable management of land resources in the developing world particularly Africa has been hampered by the high cost of conventional research approaches and constrained budgets of National Agricultural Research Systems (NARS). Infrared spectroscopy (IR) has now been adopted for routine analysis of land and water resources in several laboratories in developed economies, but many of the developments have occurred in Africa (Shepherd & Walsh, 2007). The technology uses only light for rapid, non-destructive analyses of soil and plant materials in the laboratory and is therefore

cheap, fast and versatile. Combined with geographic positioning systems and satellite remote sensing the technology can provide rapid large area mapping of soil constraints.

Sustainable monitoring and protection of the Kenyan environmental health is envisioned in EMCA (Environmental Monitoring and Co-ordination Act) (EMCA, 1999). This is in line with also stated millennium development goals (MDG), specifically on goals 1 and 7 on eradication of extreme poverty and hunger and, ensuring environmental sustainability, respectively. The provisions under EMCA and MDGs support realization of Kenya's new long-term national planning strategy, Kenya Vision 2030 blueprint that aims at making the country a newly industrializing, 'middle income economy that provides high quality life for its citizens. The Social pillar of the vision seeks to build "a just and cohesive society with social equity in a clean and secure environment" (GoK, 2007).

1.2 Problem statement

Land resources inventory is recognized as basic prerequisite for development planning. Inadequacy of land resources inventory including soil health surveillance (Shephed and Walsh, 2007) is among the major challenges of national development planning and agricultural production intensification in the developing world, particularly in Africa where most countries have no adequate country-wide soil information. In Kenya country-wide soils information is only at scale 1: 1 million meaning that the current soil databases are of low intensity and land use planning recommendations are based on few observations and thus are less effective. Development planning is constrained when information on soil resources is inadequate, unreliable or unavailable. A challenge is that conventional soil information generation is slow, expensive, and not sustainable.

Conventional land capability assessment requires expensive soil data collection and sample analysis making broad-scale quantitative land suitability assessments difficult. The FAO Guidelines for Land Evaluation (FAO, 1986), for example, is resources intensive and not sustainable. Technological paucity in capacity in new and rapid methods to quantify soil properties, variability and prediction uncertainty have hitherto constrained capability for building appropriate databases for land capability assessment for temporal and spatial monitoring in Kenya. In a situation analysis on 'the changing face of irrigation in Kenya', it was concluded, among other things that the database was inadequate and of low quality and that efforts aimed at improving the quantity and quality is constrained by technology and hitherto low priority accorded to the building of databases (Sijali, 2001).

Efforts on research and development of appropriate technologies for management of stability related problems have been hampered by, among others, a lack of a clear definition of what constitutes stable and unstable soils. It is commonly accepted that sodic, highly dispersive, and /or soils of low wet water stable aggregates constitute unstable soils. No easily accessible studies have, however, demonstrated a clear correlation of sodicity, dispersivity, and/or low wet water stable aggregation. Sodicity has been applied as the yardstick for evaluation of soil stability in Kenya and ESP 15 % taken as the critical limit. No reported studies were available on the validity of this threshold. Some studies (Sumner, 2000) considered this rating rather severe as soil degradation can take place at lower ESP. No systematic effort has been made to delimitate stable from unstable soils using one standard definition. This has challenged effective diagnosis and management of soil stability related problems.

Application of IR for soil compositional analyses is now well established. However, there has been little focus on examining the application of soil reflectance in pedotransfer functions for prediction of soil functional attributes related to soil stability (Canasveras *et al*., 2010). Also, there are no studies comparing performance of alternative IR-based predictors: near infrared (NIR), mid-infrared (MIR), and their respective wavelet transform variables and soil-based predictors although the choice has both operational and performance implications.

Few found studies (Canasveras *et al*., 2010; Madari *et al*., 2006) have defined a limited number of water stable aggregation indices (mainly fraction > 250 µm and < 250 µm), whereas a wider range (macro, meso, micro, colloidal fractions) and from different breakdown mechanisms, could better mimic soil behavior under wetting field conditions. In Kenya little fundamental research, especially on soil function has been reported and more focus has been on adaptive research. No studies on, for example, soil aggregation and its stability to establish critical values for benchmarking soil condition were available although few (point) studies (Gachene *et al*., 2003) have been reported.

1.3 Study Justification

Re-building soil health is now being recognized as a prerequisite for African food security, development and environment and over the past decade a number of key science and technology developments have occurred in Africa that has potential to enable rapid progress in addressing the gaps and accelerating soil improvement at scale. These include advances in soil health surveillance concepts for diagnosing and monitoring soil health and associated risk factors, and the large area application of these concepts using new methods in remote sensing, geographic information systems and chemometrics. Putting these advances into large-scale action, however, implies a reorientation of conventional approaches to soil science, new skills and the re-tooling of soil laboratories with access to facilities for, among others, remote sensing and other

geographic information system (GIS) technologies, and new infrared spectroscopic techniques that allow rapid, reliable, low-cost soil analysis (Hartemink & McBratney, 2008; NEPAD OST, 2008).

The Africa Soil Information Service (AfSIS) uses spectral diagnostics – low cost, high throughput analytical techniques based on reflectance of electromagnetic radiation (infrared spectroscopy (IR), total x-ray fluorescence spectroscopy (TXRF), x-ray diffraction spectroscopy (XRD), and laser diffraction particle size analysis (LDPSA)) techniques to measure soil functional properties on tens of thousands of geo-referenced soil samples in a consistent way. The low cost high-throughput spectroscopy methods are being used both as a front line screening technique for development of pedotransfer functions and for the direct development of indicators of soil functional properties. The data generated by these high-throughput techniques is treated as spectra and used as input to pedotransfer functions for prediction of soil functional properties that are expensive or time-consuming to measure (McBratney *et al*., 2006). The potential for establishment and application of stability functional attributes based on wet-sieving for modeling against basic soil properties and infrared spectral data is anchored on the proof of concept using LDPSA data, and recently developed analytical protocols including a modern laboratory infrastructure at the World Agroforestry Centre (ICRAF) Soil-Plant Spectral Diagnostics Laboratory (Shepherd, 2010; www.africasoils.net). This now offers opportunity in research organizations for capacity building and mainstreaming this high technology but simple approach for rapid analysis of land resources and interpretation of these analyses into improved recommendations for integrated management and health improvement.

Characterization of key soil properties using IR is successful (Shepherd & Walsh, 2002), however, utility of reflectance spectroscopy for direct prediction of soil functional attributes for agricultural, environmental and engineering applications, and development

of operational schemes for its use in risk-based soil assessments remained largely unexplored. Also scanty information was available (Canasveras *et al*., 2010) on superiority of IR over basic soil properties for pedotransfer purposes, although Shepherd and Walsh (2002) argued that 'because soil reflectance provided an integrated measure of a number of fundamental properties, such calibrations could performan better, and would certainly be more rapid, than pedotransfer functions based on conventional measurement of soil properties'.

The selected spectral measurement region commonly near infrared (NIR) and mid-infrared (MIR) has shown variable performance for estimation of soil properties. Key question remained the veracity of reported superiority of MIR over NIR, where MIR is found superior for some properties and NIR superior for others (Canasveras *et al*., 2010; Madari *et al*., 2006). Assumed value added using spectral wavelet transform variables in place of Fourier transform spectral data is not also rigously tested. The choice on either NIR or MIR is informed by performance difference and intended use of the data considering also practicalities in data acquisition using MIR and NIR (Bellon-Maurel & McBratney, 2011).

Aggregate slaking or spontaneous disruption and slaking plus mechanical dispersion are two different aggregate breakdown mechanisms that better mimic soil behavior under wetting field conditions. Slaking common to soils subject to flooding or (basin, furrow) irrigation present hazard for dams through sub-terranean (piping, sink holes) (Bell, 2000), and gully development via retreat through slumping highly prevalent in lowland LVB of Kenya. Slaking plus mechanical disruption as common to heavy tropical storms (Barthes & Roose, 2002) or disturbances from tillage practices (Ashmana & Hallett, 2003), enhances surface runoff, nutrient depletion, sedimentation and eutrophication. No studies were found that developed and applied aggregation indices (stable macro, stable

micro, unstable/ colloidal and macro: micro ratio) from both slaking only and slaking plus mechanical dispersion wet-sieving pretreatments for air air-dried (≤ 2000 μm) soil.

Accelerated soil erosion is among key drivers of soil degradation in Kenya. The country continues to be highly dependent on its land and particularly soil resource for continued agricultural production. To avoid further depleting this resource, there is a need to identify soils and associated land management practices where there is a risk of accelerated soil degradation using new simple fast inexpensive yet accurate methods. Research has an obligation to feed-in and support sustainable land management and administration by providing efficient and effective evidence-based information on land health packages providing guidelines for best management practices for land managers and farmers alike that allow them to make sustainable soil management decisions. This will particularly be important for county governments as the focal points of development. The current study acknowledged the central role that soil stability plays in sustainable land management and proposed a framework for soil stability diagnosis using inexpensive data in support of assessment and monitoring of soil health status specifically for the Lake Victoria basin of Kenya and generally for similar environments.

1.4 Objectives of the study

1.4.1 Main objective

The main objective of this study was to evaluate the use of infrared spectroscopy (IR) in diagnosing soil stability related problems and assessing their prevalence with a case study of Lake Victoria Basin (LVB) in Kenya.

1.4.2 Specific objectives

Specific objectives were:

1. To develop alternative IR-based models for screening soil stability related properties across a range of sensitive soils and compare these with predictions using conventional physico-chemical properties.
2. To further validate the IR-based models using an independent set of soil samples set from two sentinel sites.
3. To assess the prevalence of soil stability related problems in two sentinel sites using IR-based models.
4. To assess indices of soil stability functional attributes most appropriate for screening stability problems using IR-based models.

1.4.3 Research questions

1. Do IR-based predictors provide superior models over soil physic-chemical properties for estimation of stability related properties?
2. Do alternative IR-methods provide robust models for estimation of stability related soil properties?
3. What is the prevalence of stability related problems in LVB of Kenya?
4. Which soil aggregation indices are more appropriate for screening stability related problems using IR in LVB of Kenya?

1.5 Study Conceptual Framework

A scheme based on development of soil spectral libraries and the double sampling approach was proposed for effective diagnostic screening and prevalence assessment of stability related problems using IR-based models in LVB of Kenya (Figure 1.1**)**. In the scheme two sets of soil samples are collected from LVB of Kenya. Set 1 is obtained from sites selected to represent a wide range of stability sensitive soils from across LVB. The sites are selected based on available legacy data (several soil profiles with property analytical data), following a strategy that allow accurate recovery of the original variation in the larger property dataset with fewer samples/ sampling sites **(**Stenberg *et*

al., 1995). Similar strategy has successfully been applied by other workers (Genot *et al*., 2011**;** Viscarra Rossel *et al*., 2008). Set 1 is used for model calibration. Set 2 is selected to represent a larger set that is representative of the major landform, soils and landuse and land cover types in two sentinel sites from LVB. Soil reflectance spectra are generated for Set 1 and for the larger set. A representative subset (Set 2) is obtained using the spectra of the larger set (Naes et al., 2002), and used for calibration validation. Reference data is obtained for a suite of selected soil basic properties commonly deployed in land capability/ suitability assessments for Set 1 and Set 2. Also obtained for Set 1 and Set 2 is reference data for selected stability functional attributes (SFA). The soil basic properties are screened for prediction of the SFA and key few soil predictor properties that are strongly correlated to spectra selected. Regression models for selected soil properties (developed using the more robust of Set 1 and Set 2) are used to predict these properties in the larger set, using the latters' spectra data. The predicted values are used to predict the SFA in the larger set from the two sentinel sites. The predicted SFA indices are used to define cutoffs for low/ high stability category for each sample of the larger set. The low stability tally is used for prevalence assessment of stability problems within the larger set. The potential for development of pedotransfer functions and inference systems for more effective estimation of functional attributes from spectra and basic soil properties has been demonstrated by other workers (Minasny & McBratney, 2008; Tranter *et al*., 2008). The double sampling approach and development and deployment of spectral libraries for rapid diagnostic screening for land health surveillance was illustrated by Shepherd and Walsh (2002; 2007).

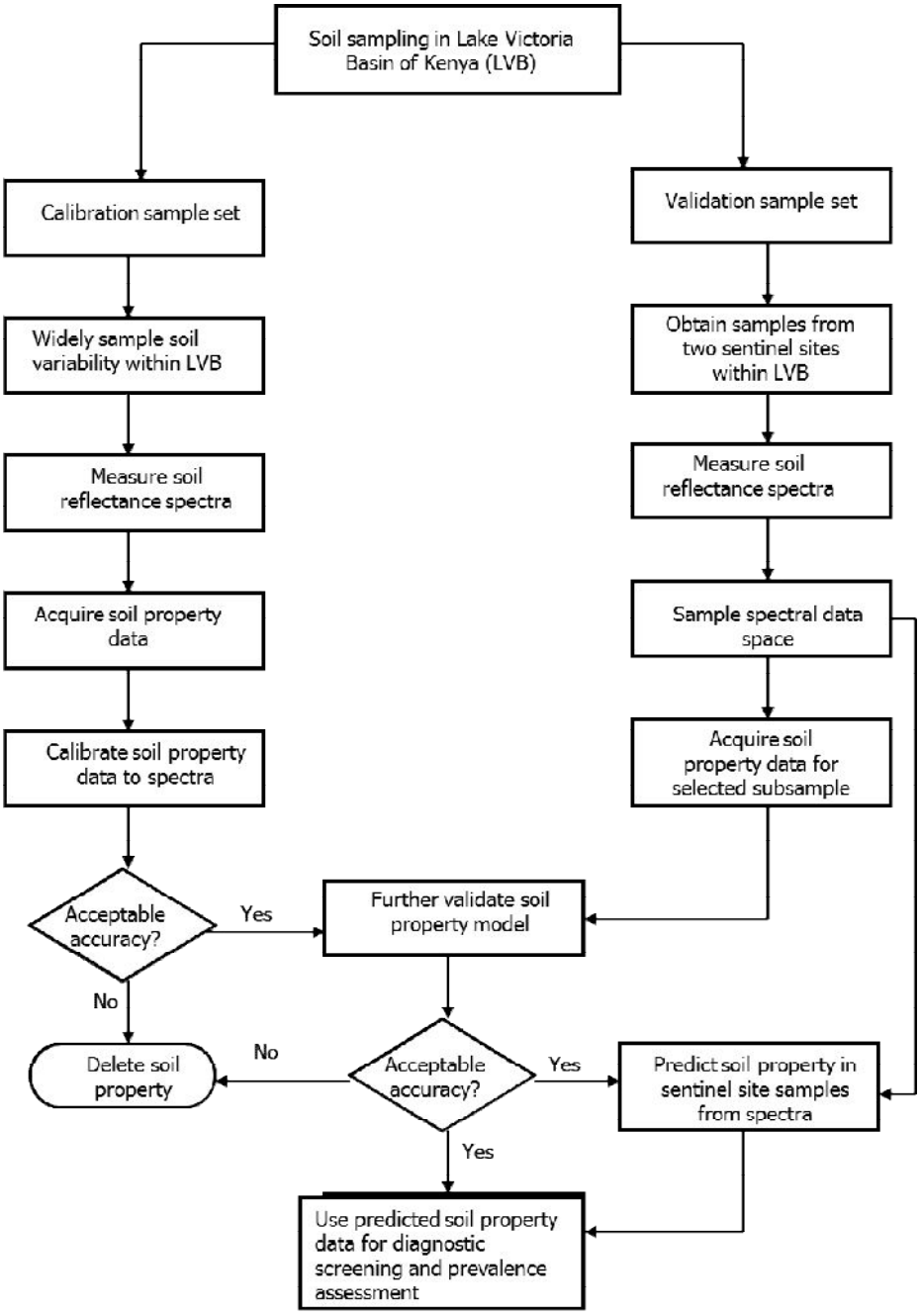


Figure 0.1.1: Scheme for development of IR-based models for diagnostic screening and prevalence assessment of stability related problems in LVB of Kenya.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

Soil health can be viewed as the capacity of a soil to sustain ecosystem services, including provisioning, regulating, supporting and cultural services. The role of soil aggregation and aggregate stability in assessment and monitoring of soil health is well documented (Osuji & Onweremadu, 2007; Wei *et al*., 2006). A renewed interest in soil wet stable aggregation (WSA) is probably borne of the need for development of fast cheap yet accurate soil health indicators. Conventional laboratory methods of soil analyses have challenges including: low precision, the use of analytical methods which generate chemical waste and are time-consuming (Nocita *et a*l., 2015; Lyons *et al*., 2011). Wet-sieving and aggregate fractionation, for example, is laborious and involves also sample peroxidation and use of surfactants (Marquez *et al*., 2004) whose effluents pose environmental hazard. Differences in aggregation and aggregate stability (AS) results, attributable to differences in critical aspects (sampled depth and sample distribution; sample pretreatment and treatment, selection of fractionation sizes and units of expression), all make it difficult to compare different studies and this has constrained inclusion of AS in minimum datasets (MDS) for soil health assessment (Merrington, 2006).

Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) (commonly IR) is known to be a physical, non-destructive, rapid, reproducible and low cost method for analyses of several constituents simultaneously in samples from very diverse materials (Shepherd & Walsh, 2007). The selected spectral measurement has shown variable performance for estimation of several soil properties. Key question from among comparative studies remained the veracity of reported performance superiority of

mid-infrared (MIR) over near-infrared (NIR) spectral ranges, and value added using spectra wavelet transform variables.

Key challenge with application of chemometrics and spectroscopy (soft modeling) in soil health studies remained model robusticity. The opportunity now is to harness technological advancement for more efficacious assessment and monitoring of soil health. This could be made feasible by deployment of approaches that combine among others, pedotransfer functions and spectral analysis to derive relationships that link the basic soil properties to functional soil properties that are more difficult to measure. The approaches promise to help in solving also the "model transferability" problem (Linker, 2012).

The deployment of data mining or machine learning techniques could help to improve on efficiency in selecting key basic variables from a complex multivariate data set of potential predictors without significant loss of information. The successful use of spectra for creation of diagnostic screening tests, especially for soil properties that were otherwise moderately calibrated to spectra (Shepherd & Walsh, 2002) has allowed satisfactory spectra classification for benchmarking soil condition and developing quality indices.

2.2 Soil aggregation and aggregate stability (AS)

Soil aggregation is the arrangement of primary soil particles into compound elements, which are separated from adjoining structural elements by surfaces of weakness. For good plant growth and soil health soil aggregation must be stable to disruption on wetting with optimal pores for water storage, water transmission, and for roots to grow. Soil aggregation and its stability (AS) is widely recognized as a key indicator of soil quality particularly soil physical health. This is attributed to the sensitivity of AS to both natural and/or cyclic factors, to anthropogenic (human-induced) soil and land

management practices (Wei *et al*., 2006), and also due to observed relationship of AS with basic soil properties (Canasveras *et al*., 2010; Li *et al*., 2005).

2.2.1 AS relationship with soil basic properties

Soil aggregation and aggregate stability (AS) exhibit variable relationship with soil basic physico-chemical properties including pH and electrical conductivity (EC), carbonates, soluble and exchangeable bases, soil moisture content (mc), texture, soil organic matter (SOM) (Canasveras *et al*., 2010**;** Igwe & Nwokocha, 2005; Igwe & Stahr, 2004).

AS relationship with pH: Westerhof *et al*. (1999) found that soil pH was positively correlated with the amount of clay dispersed after 3 h of shaking in water, suggesting that management practices that increases soil pH has deleterious effects on aggregation (Idowu, 2003), in support of earlier works (Auerswald, 1995) that established an inverse relationship of pH with percolation stability.

AS relationship with particle-size: The clay mineralogy of the soil influences aggregate slaking (Boucher, 2010), and soils with higher clay content have lower aggregate stability due to presence of expandable clay (Le Bissonnais *et al*., 2002). However, Levy *et al*. (2003) found that generally soils with < 25 % clay content were of inherently lower stability while > 35 % clay had inherently higher aggregate stability. Boix-Fayos *et al*. (2001) found that aggregates 1000 - 105 µm in diameter were positively correlated to medium, fine, very fine sand and silt fractions; aggregates < 105 µm were positively correlated to clay content, and; water stability of micro-aggregates was positively correlated with clay content. The later findings were affirmed by other workers (Zhang *et al*., 2008; Li *et al*., 2005). These studies suggested that textural separates could be potential proxies of water stable micro aggregates and unstable aggregate fraction.

AS relationship with SOM: Soil organic matter (SOM) contributes significantly to soil structural formation, however, contribution of SOM to maintenance of AS is controversial. Some studies reported significant relationship between AS and SOM, others found weak to no relationship, whereas others emphasize conditional relationship of AS and SOM including particle size, clay mineralogy, and characteristic of soil organic carbon (SOC) fraction (Wei *et al*., 2006). Significant relationships was found, for example, on the number of drop impacts and SOM (Cantón *et al*., 2009); AS against slaking was correlated to SOC (R^2 = 0.71), and water drop penetration time increased with carbon contents (Chenu *et al*., 2000). Chaudhary *et al*. (2009) found that SOM had no direct contribution to surface or subsurface stability in semi-arid shrublands. Zhang and Peng (2006) found significant (albeit weak) linear correlation (r^2 = 0.34 - 0.46, p < 0.05) of AS and SOC pools. Ashmana and Hallett (2003) suggested that the linkage between AS and SOM was an artefact of the fractionation procedure. This suggested that there was still much uncertainty in the dynamics of SOM in macro- and micro-aggregation. Important however, is that AS plays key role in SOC storage for carbon sequestration (Lal, 2009; Denef *et al*., 2004).

AS relationship with exchangeable bases and CEC: Significant relationship has been reported of AS with exchangeable Ca (eCa) and exchangeable Na (eNa) (Osuji & Onweremadu, 2007), and with effective CEC (Zhang & Horn, 2001). The clay activity (ratio CEC: % clay) values > 0.55 (Irvine & Reid, 2001) or > 0.*8* (Dawes & Goonetilleke, 2006), is associated with a predominance of 2:1 lattice clays mainly smectites and a higher predisposition to instability. Several workers (Ward & Carter, 2004; Sumner, 2000) consider sodicity (reflected by measures of eNa, exchangeable sodium percent-ESP and sodium adsorption ratio- SAR) a measure of soil AS by enhancing dispersion.

An increase of SAR has been associated, for example, with a decrease in water stable aggregates and an increase in amount of water dispersible clay (WDC) (Tajik *et al*.,

2003). The ratio of EC of a 1: 5 water extract (EC_5) to eNa and to ESP has been proposed Electrochemical Stability Indices (ESI) (ESI 1 and ESI 2, respectively), to predict dispersion (Hulugalle & Finlay, 2003). The 0.05 ESI thresh hold is proposed structural stability diagnostic tool (Dawes & Goonetilleke, 2006). Studies on soil sodicity amelioration using gypsum (Levy *et al*., 2006; Choudhary *et al*., 2004) indicated improved soil physical conditions, suggesting that soil gypsum requirement (GR) presented potential indicator of soil structural instability. These studies affirm that basic soil chemical properties provide potential pedotransfer functions for estimation of AS.

AS relationship with sodicity: Weak correlation of AS with sodicity (Levy *et al*., 2003; Levy & Mamedov, 2002), was partly attributed to lack of control/isolation of slaking (by capillary wetting pretreatment, for example), before wet sieving. Levy *et al*. (2003) argued that the effect of sodicity on AS was too low and required high sensitivity measurement method like the High Energy Moisture Characteristic (HEMC) (Mamedov *et al*., 2007), to quantify. Most studies (Sumner, 2000) associated sodicity (probably erroneously) with aggregate slaking and dispersion, although Yoder (1936) demonstrated that the stresses created upon rapid wetting of soil were caused by entrapped air and differential swelling of clay particles resulting in aggregate slaking, suggesting that slaking is more a physical than a chemical process. Arguably, however, sodicity might have significant contribution in enhancing differential swelling due to the high hydration power of Na^+ and this might be among the reasons sodicity confounded relationship of AS with other soil properties (Ward & Carter, 2004; Barthes' & Roose, 2002).

AS relationship with mechanical properties: Shrink-swell is a physical process that has profound effects on soil stability. Commonly used shrink-swell indicators COLE (coefficient of linear extensibility) and associated VS (volumetric shrinkage), have

shown close relationship with basic soil properties including elemental concentration, CEC, mc, liquid limit (LL), plastic limit (PL), plasticity index (PI), linear shrinkage (LS), and activity (A) (Fratta *et al*., 2007; Thomas *et al*., 2000). The LL combined with mc at 0.03 MPa provided a useful tool to evaluate the structural stability of a calcareous saline-sodic soil under reclamation, and; soils having a PI < 0.1 kg kg^{-1}, LL < 0.3 kg kg^{-1} have been categorized as dispersive and erodible (Lebron *et al*., 1994). Mechanical soil properties could, therefore, be treated as soil basic properties and included as potential predictors of stability functional attributes. These properties are otherwise commonly deployed as 'standard soil tests' for Materials Testing in engineering applications.

2.2.2 Aggregation and aggregate stability in soil studies

The sensitivity of aggregation and aggregate stability (AS) is manifest in its response to the effects of a wide range of natural and human-induced soil conditions including action of plants and microorganism (Chaudhary *et al*., 2009); rangeland conditions (Bird *et al*., 2007)*;* soil erodibility (Cantón *et al*., 2009); vegetation, land use, slope, and climatological gradients (Osuji & Onweremadu, 2007; Boix-Fayos *et al*., 2001)*,* allowing effective assessment and monitoring of soil health. Quantifying effects of biological communities (flora and fauna) on soil stability inform, for example, the relative amounts of resources that erosion control practitioners should devote to promoting these communities (Chaudhary *et al*., 2009). The AS has effectively been used to assess, for example, so-called natural ecosystem or earth engineers' contribution to improvement and maintenance of soil health and this has led to ways to enhance inputs for sustainable agricultural systems (Chaudhary *et al*., 2009). Other efforts (Krasilnikov *et al*., 2008) focused on stability spatial mapping.

Sensitivity to soil erodibility: AS is probably the single most important soil property governing soil erodibility. The susceptibility of soil to particularly crusting and erosion is often inferred from measurements of AS (Cantón *et al*., 2009). Stability of topsoil

aggregates is a valuable indicator of field-assessed runoff and inter-rill erosion (Cantón *et al*., 2009), and the development of the seal and soil infiltration capacity are related to AS (Le Bissonnais & Arrouays, 2005).

Response to influence of cyclic/seasonal conditions: The AS has effectively been used to quantify the influence of cyclic/ seasonal conditions on soil health including climatological gradients (Boix-Fayos *et al*., 2001); vegetation, land use and slope gradients (Cantón *et al*., 2009; Osuji & Onweremadu, 2007), and in differential soil wetting (water repellency) (Roper, 2005). Different aggregate sizes along a climatological transect provide good indicator of soil degradation. Large aggregates (>10,000, 10000 - 5000, 5000 - 2000 μm) present in highest proportions in the most arid of the studied areas were associated with high values of water-stable micro-aggregates and were related to high bulk density (BD) and low water retention, whereas small aggregate sizes (1000 - 105 and < 105 μm) had a positive influence on soil water retention (Boix-Fayos *et al*., 2001).

Response to human-induced management interventions: Measurements of AS provide effective indicator of the status and changes of soil health as a result of different human-induced management interventions (Wei *et al*., 2006). This includes: assessing the response to organic and inorganic additions (Li & Zhang, 2007; Haynes & Swift, 2006); response to land cover and land use intensities (Zhang *et al*., 2008); sodic and saline-sodic amelioration (Levy *et al*., 2006), and; prognostic monitoring post-restorative soil quality (Zhang & Peng, 2006).

Reported AS in the order: forestland > orchard > cropland > bareland (Zhang *et al*., 2008), affirmed now generally expected effect of land use intensity on soil health. The negative influence of conventional tillage on soil AS has demonstrated the gains of conservation tillage (Williams *et al*., 2005). Soils which have been cropped for long periods of time have unstable aggregates and on these soils, the type of crop grown has a

greater benefit on AS than the type of tillage or method of managing crop residues. This is in contrast to a soil with stable aggregates, where the type of tillage has a greater influence on AS than crop residue management or crop rotation.

These studies indicate the central role given to AS, as a sensitive indicator of effect of land management practices on soil health. Importantly that conventional cultivation decreased the quantity and quality of SOM and caused a loss in AS resulting in an overall decline in soil quality (Williams *et al*., 2005).

2.2.3 AS measurement methods and indices

Common approaches for evaluating the state of soil aggregation have included determination of: the aggregate-size distribution following dry or wet sieving where aggregation is expressed as mean weight diameter (MWD) (Canasveras *et al*., 2010) or mean geometric diameter (MGD) (Madari *et al*., 2006). The pore size distribution following changes in saturated and unsaturated infiltration rates and saturated hydraulic conductivity has been used as indices of stability (Vahyala, 2009). The ratio of the intrinsic permeability of the soil to air (k_a) and to water (k_w) (k_a/k_w) has been proposed also as structural stability index (Waruru & Wanjogu, 2002; Whelan *et al*., 1995), with unity representing fundamental stability. Determination of wet water stable aggregates and aggregate size distribution mainly from wet sieving-based method (Kemper & Rosenau, 1986) was the most widely reported.

AS from wet-sieving: Wet-sieving method has involved procedures that applied aggregate slaking (rapid wetting) (Angers *et al*., 2008), capillary wetting to control or enhance slaking (Levy *et al*., 2003), slaking plus mechanical disruption (Boucher, 2010), or combined three tests in one: slaked, capillary-wetted and slaked plus dispersed pretreatments (Marquez *et al*., 2004). The fraction of the original aggregates that has some degree of resistance to slaking is referred to as water stable aggregates and is used as an index for aggregate stability (Angers *et al*., 2008).

Aggregate slaking plus dispersion: Aggregate dispersion is either 'spontaneous' or 'mechanically- induced' (Rengasamy & Olsson, 1991) and several tests have been proposed for identification of dispersive soils including the water-drop test, crumb or Emerson test, pinhole test (PHT), turbidity ratio test, amount of total soluble salts in soil water, double hydrometer test (Boucher, 2010; Bell, 2000).

Based on the double hydrometer test (USDA- NRCS, 1996), different attributes have been proposed as aggregation indices including: water-dispersible clay (WDC) content, the clay (< 5 μm) dispersion ratio (CDR), soil (silt+clay) dispersion ratio (DR), clay flocculation index (CFI), ratio total silt : total clay, ratio WDC: WDSi, aggregated clay index (ACI) (WDC: tClay ratio), and aggregated silt and clay (ASC) (Canasveras *et al*., 2010; Oguike & Mbagwu, 2009). The soil material < 20 μm (wt %) define dispersion index (DI) [100 × (< 20 μm from water) / (< 20 μm total)] (Hulugalle *et al*., 1999). The material < 20 μm from wet sieving has been expressed also as 'instability index' (IS) = [% < 20 / (WSA_{ma} + WSA_{mi})], where WSA_{ma} and WSA_{mi} is the fraction > 250 μm and < 250 μm, respectively (Fortun *et al*., 2006; Hulugalle & Finlay, 2003). Dispersion groups (DR: < 30, 30 - 65, and > 65%) were used to effectively differentiate soils and the categories to establish association with basic soil properties (Ward & Carter, 2004).

The proposed aggregation indices from double hydrometer test could be efficacious for soil (in) stability assessment, however, some of the indices (CDR, CFI and FI; DR, DI and ASC) might present redundant information, and also the treatment of the indices as functional attributes could be subjective considering rules of pedotransfer functions (Minasny, 2007). Some of the dispersion tests (crumb, PHT, water-drop,) have also (reliability) limitations despite/inspite of their simplicity (Farres & Cousen, 2006).

AS universal measurement method and index: Differences in AS results, attributable to differences in critical aspects including: the sampled depth and sample distribution; sample pretreatment and treatment, including initial size of test sample, handling of primary sand; selection of fractionation sizes and units of expression, all make it difficult to compare different studies and to make meaningful interpretation of AS data. Also efforts towards developing universal AS measurement method and index have largely been uncoordinated. This has constrained inclusion of AS in minimum datasets (MDS) for soil health assessment (Merrington, 2006).

Sampled depth and sample distribution: Most AS studies (Le Bissonnais *et al*., 2007) are based on one depth (particularly top 0-20 cm) samples, and from a limited number of soil types. Reasons include that the surface horizon is most sensitive to disturbance from tillage, rainfall impact and runoff, and management practices. Some workers (Merrington, 2006; Yoder, 1936) argued that aggregate stability studies for subsoil horizons were not relevant and this could partly explain why relatively few studies (Ward & Carter, 2004; Tajik *et al*., 2003) have incorporated soils from subsurface, particularly deep subsoil (> 50 cm) horizons. Arguably, surface horizon are less susceptible to dispersion as a result of higher organic matter, lower accumulation of exchangeable cations, low bulk density and thus present lower AS hazard (piping, sinkholes, gully collapse and retreat), suggesting that subsurface horizons might be more critical for stability diagnostic studies. Aggregate breakdown and dispersion and particularly the influence of sodicity is more a subsurface than surface phenomena (Ward & Carter, 2004). Other workers (Bouajila & Gallali, 2008; Tajik *et al*., 2003) have demonstrated significant differences in AS between soil types and different soil pedon horizons.

Optimal aggregates size: A desirable range of pore sizes for a tilled layer occurs when most of the clay fraction is flocculated into micro-aggregates and the micro-aggregates are bound together into macro-aggregates (Angers *et al*., 2008), however, there were no set standards for soil size fraction for analysis. There was also no optimal stable macro- or micro- aggregate size and different fractionation sizes have been proposed: macro-aggregates (> 200, > 250, > 212 µm); micro-aggregates (< 200, < 250, and 50 - 250 µm); macroscopic and colloidal clay (< 20 µm) (Oguike & Mbagwu, 2009; Marquez *et al*., 2004). On the one hand, aggregates ranging from 1000 - 5000 µm size are considered valuable for agriculture and weighting is proposed for the various size ranges for stability measurements. The proportion of aggregates ≥ 2000 µm are considered suitable indicator of the influence of tillage systems on aggregation, and*;* the largest (> 2000 µm) and the smallest (< 250 µm) aggregates were preferentially enriched in carbon and nitrogen relative to the 2000 - 250 µm range aggregates (Angers *et al*., 2008; Wei *et al*., 2006). On the other hand, aggregates ranging from 200 to < 20 µm size are valuable for soil stability and erodibility studies (Le Bissonnais *et al*., 2007; Marquez *et al*., 2004).

Accounting for sand in stable aggregates: Sand plays a passive role in the formation and stabilization of aggregates, however, correction for sand is needed in AS determination to put on similar footing soil samples with different amounts of total sand and to remove undesirable effects introduced by sand content (Marquez *et al*., 2004). There were no set standards for accounting for the presence of various sand fractions [(coarse: 2000 - 500 µm; medium: 500 - 250 µm; fine: 250 - 50 µm)] (Zobeck, 2004). Some studies corrected for sand, for example, using the same sieve-size as for the stable aggregates fraction (Angers *et al*., 2008). Other workers (Fortun *et al*., 2006; Barthes & Roose, 2002) refer to correction for coarse sand but used the 200 µm sieve effectively correcting for coarse and medium sand content, and inardvertently not accounting for sand in < 200 µm size range. Other studies (Marquez *et al*., 2004) refer to correction for

total sand and used the 53 µm sieve. Notably, soil fraction 20 - 53 µm may comprise of fine sand and/or of stable micro-aggregate fraction and correction of sand using 53 µm sieve, might underestimate the stable micro-aggregate fraction.

Expression of results, measurement precision, and critical values: Expression of results from different AS methods and their interpretation might conflict (Le Bissonnais *et al*., 2007), or complicate comparison where for instance stable aggregates are expressed in seemingly unrelated units, for example, water stable aggregates as g kg^{-1} or g 100^{-1}g soil (Shulka *et al*., 2004; Barthes & Roose, 2002). There was paucity of information also on data reproducibility of the procedures (Le Bissonnais *et al*., 2007). Few found studies reported different results on AS for the same soil samples using different methods (Levy *et al*., 2003). Exception was the High Energy Moisture Moisture Characteristic (HEMC) method that demonstrated reproducible data on weakly aggregated (loessial) soils from the USA and Israel (Levy *et al*., 2003).

Critical (threshhold) values are important for soil health and also economy, however, scanty information was found on, for example, AS cost-benefit analysis *(*for instance in crop yields, soil loss, loss/gain in monetory terms*)*. Pringle (1975) found that the critical stability index values had been reached when the crop yield got lower than the farm average. Shulka *et al*. (2004) described water stable aggregate in terms of a sustainability index where value greater than 700 (g kg^{-1}) was described as highly sustainable and value less than 150 g kg^{-1} as unsustainable.

Advances towards universal AS method and index: There were no systematic advances towards development of a universal AS measurement method. Several workers (Cantón *et al*., 2009; Li & Zhang, 2007) focused, however, on comparing (correspondence/ agreement) and merit of indices from different fraction sizes and measurement methods, and validating the three tests method presented by Le Bissonnias

(1996). Considerable efforts have been put also in improving on the wet-sieving method of Yoder (1936) in methodology and definition of stability indices (Kemper & Rosenau, 1986; Le Bissonnais, 1996; Marquez *et al*., 2004; Angers *et al*., 2008) to better relate AS and other important soil properties.

Mbagwu (1992) found that for the soil micro aggregate indices: fraction < 250 µm, aggregated silt and clay (ASC), and aggregated clay index (ACI), the fraction < 250 µm reflected differences in organic cabon levels more than the ASC and the ACI. The ASC correlated positively whereas the fraction < 250 µm correlated negatively with the macro aggregate indices (fraction > 250 µm), suggesting that these two fractions measure different dimensions of AS.

Ashmana and Hallett (2003) found that in the slaked treatment, micro-aggregates (< 250 µm) contained 17 % more soil organic carbon (SOC) and had 30 % faster rates of respiration, while for slaked and shaken treatment the aggregates contained 12 % more SOC and had 14 % faster rates of respiration, suggesting that the linkage between soil aggregate size class, soil organic matter (SOM), and respiration rate were artefacts of the fractionation procedure. The authors concluded that relationships between aggregate size and other properties must be interpreted in terms of the disruptive mechanisms used to fractionate aggregated soil. This suggested that the two wet sieving pretreatment procedures (slaking only and slaking plus mechanical shaking), provided different dimensions on the relationship between AS and basic soil properties.

Le Bissonnais (1996) three treatments in one method allowed better correlation of stability with soil erodibility under field conditions. Amezketa (1999) emphasized attention to the conditions of sample collection in the field and sample preparation and treatments in the laboratory to address difficulties in obtaining consistent correlation between AS and other important soil properties such as soil erodibility, crusting

potential, and SOM enrichment. Marquez *et al*. (2004) proposed a framework involving three pretreatments: slaked; capillary-wetted, and; subsequent slaking of aggregates > 250 μm in size that permitted both accurate determination of aggregate-size stability distribution, and distinction between amounts of stable and unstable macro-aggregates (> 250 μm) based on their resistance to slaking. Angers *et al*. (2008) defined both macro and micro water stable aggregates and accounted for both the total sand and air-dried moisture content (mc), since initial or antecedent mc controls soil behavior including water stability of aggregates.

A challenge with these advances is in methodological rigor. The efforts were also largely uncoordinated and a universal method remained unlikely. What may be feasible is probably to rigorously test and 'standardize' field and laboratory methods for specific land uses, soil or land management practices. A broad criterion for choice of laboratory and/ or field method includes that the method is simple, fast, accurate, and sensitive for intended use. While this might reduce the number of test methods, a limitation would still be the criteria for choice of particular test method and acceptability. Angers *et al*. (2008) proposed detailing the particular procedure used from sample collection, transportation, pretreatment, treatment, data measurements, indices and expression of results. Arguably, the most appropriate AS measurement method will consider the local circumstances and purpose. For large area diagnostic screening, for example, key consideration could be simplicity (in terms of field data collection, sample preparation, wet-sieving pretreatment and aggregate fractionation), and measurement precision.

2.3 Diffuse reflectance infrared spectroscopy

Diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) or commonly infrared spectroscopy (IR) is the study of the interaction of infrared light with matter and an infrared spectrum is the fundamental measurement obtained in infrared spectroscopy. IR is defined to include near infrared (NIR: 0.75 μm to 2.5 μm) and mid-infrared (MIR:

2.5 µm to 25 µm) spectral ranges. As an analytical technique IR has shown great potential for solving analytical problems of samples found in numerous fields including agriculture, geology, pharmaceuticals, and medicine (Shepherd & Wash, 2007; McClure, 2003). The potential of IR for non-destructive analyses of soils have been also intensively investigated (Linker, 2012; Viscarra Rossel *et al*., 2011; Stenberg *et al*., 2010). Great efforts and advances have been made also in the use of IR for soil type differentiation and pedogenetic classification (Dematte *et al*., 2012; Ben-Dor *et al*., 2008), mapping of the spatial and temporal variability of soil properties in extensive areas (Deng *et al*., 2013), with implication for improved efficiency in soil mapping for soil survey**,** including digital soil mapping (Sanchez *et al*., 2009; Hartemink & McBratney, 2008). Efforts are being directed also to application of spectral libraries and spectroscopy as a diagnostic screening tool that can aid the development of a reliable spectral case definition to characterize soil health for agricultural and environmental management at the farm and landscape level (Shepherd & Walsh, 2007). Less focus has, however, been in the use of IR for soil aggregation (Canasveras *et al*., 2010) although soil functional attributes may be estimated with higher precision and sometimes accuracy from IR than from conventional laboratory measurements (Shepherd *et al*., 2005).

2.3.1 Soil spectra acquisition and interpretation

To generate a soil spectrum, radiation containing all relevant frequencies in the particular range is directed to the sample. Depending on the constituents present in the soil the radiation will cause individual molecular bonds (typically H - C, C – O, carbonyl groups, C – N bonds) to vibrate, either by bending or stretching, and they will absorb light, to various degrees, with a specific energy quantum corresponding to the difference between two energy levels. As the energy quantum is directly related to frequency (and inversely related to wavelength), the resulting absorption spectrum produces a

characteristic shape that can be used for analytical purposes (Viscarra Rossel *et al*., 2011).

Qualitative spectra interpretation: Reflectance increases, for example, as particle or aggregate size decreases (Atzberger, 2002), effectively lowering apparent absorbance. The reflectance curve shape changes due to also the strong moisture absorption bands near 1400 and 1900 nm (Stenberg *et al*., 2010) and the shoulder of the 1900 nm and 2200 nm absorption features is highly diagnostic in classifying aggregates (macro, meso, and micro fractions, and organic matter) across different sites, reflecting difference in OH vibrations in adsorbed water and lattice water, respectively (Mutuo *et al*., 2006). Verchot *et al*. (2011) found characteristic absorption band patterns in the frequency range of 800 - 4000 cm^{-1} with prominent diagnostic features for three different aggregate size fractions (macro, meso, and micro). Spectra for micro-aggregates were very similar across the two different sites and showed increased average absorbance values for all functional groups with depth, whereas there were no apparent trends with depth for the meso- and macro-aggregates (Verchot *et al*., 2011). However, soil spectra are complex attributed to soil compositional complexity making qualitative interpretation less effective.

Quantitative spectra interpretation: The complex soil absorption patterns are mathematically extracted from the spectra and correlated with soil properties using multivariate calibrations (Viscarra Rossel *et al*., 2006). Calibration performance is influenced, however, by several factors including samples characteristics (provenance, composition), soil property characteristics (spectral response, data quality), spectra measurement region, and calibration method (Reeves, 2010; Stenberg *et al*., 2010).

2.3.2 Factors influencing IR calibration performance

Influence of samples characteristics: Differences in calibration performance could be explained by the sample size and sampling strategy, sample provenance (location/ sites, sampled depth and soil type), and sample presentation (Reeves, 2010; Stenberg *et al*., 2010). A limited number of samples or samples from a limited number of sites of similar soil types can limit, for example, the calibration to a certain type of soil or only be applicable to the landscape studied (Islam *et al*., 2003). Shepherd and Walsh (2002) found a gradual decrease in predictive performance with decreasing sample size at large sample sizes, but a rapid decrease as sample size decreased below about 100 to 200 samples.

Sampling strategy: Important for a representative sample set is that the (re) sampling strategy recovers the original variation in the larger set with fewer samples. Genot *et al*. (2011) found that the data range (separate test set) obtained based on a "conditioned Latin hypercube sampling – cLHS" strategy (Minasny & McBratney, 2006) was within the range of values for calibration set, a prerequisite for robust modeling. Stenberg (2010) obtained a representative subsample from a large dataset of Swedish agricultural soils to cover clay and organic matter contents without co-variation using cLHS. Very poor prediction power for soil constituents in a large set of archived samples was highly improved using datasets from a representative sub sample obtained following cLHS (Viscarra Rossel *et al*., 2008).

Sampled depth: topsoil models for some properties (pH, OC, CEC, eCa and eMg) were superior over counterpart subsoil models, whereas the converse was true for other properties (eNa, EC, eK, ESP) (Dunn *et al*., 2002), attributed to also differences in sample composition that influence analyte content and data distribution. Similarly, better silt in subsurface than surface, and better clay activity in subsurface than surface data sets (Nanni & Dematte, 2006), presumably was attributable to variation in analyte concentration. This suggests the importance of developing also separate depth models

from interval depth sample set datasets although Merry and Janik (2001) argued that mixed depth models are more effective than separate depth models since they take account of both surface and subsurface soil variability.

Influence of soil property characteristics: Calibration performance is influenced by property characteristics including: spectral response (active or inactive); quality of reference data, the form of analyte (directly measured or derived), data range and data distribution (Linker, 2012; Reeves, 2010). Calibration is successful for wide range in reference analytical data and for test set data that is in the range of calibration set (Stenberg *et al*., 2010).

Spectral response: calibration of spectrally active (primary) soil properties including: mc, SOM, SOC, clay mineralogy, and Fe oxides (soil chromophores) is direct, whereas calibration for non-responsive (secondary) properties (including EC, pH, soluble and exchangeable bases, micro elements, and derived properties) is through co-variation (auto-correlation) with soil chromophores (Stenberg *et al*., 2010). Reeves (2010) refer to calibration of secondary properties as surrogate calibration. Calibration for primary properties is strong and robust (model geographic transferability is more successful). Secondary calibrations are: (i) highly variable and not-so-robust due to lack of direct relationships between the spectra and these properties, and (ii) work best over geographical areas with a homogeneous geology and anthropogenic history, therefore, are local and largely specific to a particular data set (Stenberg *et al*., 2010).

The mechanism responsible for calibration of aggregation indices remained controversial (Stenberg *et al*., 2010). Soil structure is assumed to have primary response to spectra, attributed to unique effect of soil surface physical properties (such as size and shape of soil aggregates) on influence of the light scattering and light path lengths (Atzberger, 2002; Chang *et al*., 2001). Efforts at establishing absorption features

associated with aggregation indices (Verchot *et al*., 2011; Mutuo *et al*., 2006) were, however, not conclusive. Canasveras *et al*. (2010); attributed calibration of aggregation indices (water stable aggregates, mean weight diameter, and water-dispersible clay) to also their individual association with spectrally responsive soil basic properties. High variability is also found for calibration of water stable aggregates on IR from different studies (R^2 = 0.46, 0.56, 0.60, 0.92) (Canasveras *et al*., 2010; Madari *et al*., 2006; Chang *et al*., 2001; Ben-Dor & Banin, 1995).

Quality of reference data: calibration performance depends on the quality of reference input data (Reeves, 2010), yet when performed by different laboratories routine soil analyses can yield very different results, even when they use the same methodologies (Nocita *et al*., 2015; Lyons *et al*., 2011; Cantarella *et al*., 2006). Different results are reported on AS for the same soil samples using different methods (Le Bissonnais *et al*., 2007). This is contrary to spectral measurements that are highly precise (Shepherd *et al*., 2005). Shepherd and Walsh *(*2003) showed how IR may be used to improve the accuracy of the reference wet chemistry method.

Properties from direct measurements perform better than derived, especially ratios of soil constituents may be predicted less accurately than the individual constituents (Stenberg *et al*., 2010), because of an error propagation of the individual error terms from reference measurements (Ludwig *et al*., 2008), and the level of interaction among the constituents (Nanni & Dematte, 2006). The silt: clay ratio (R^2 = 0.78) was better than for silt (R^2 = 0.27) but lower than clay (R^2 = 0.92) (Nanni & Dematte, 2006), suggesting more influence by clay and no apparent interaction between clay and silt. Sensitivity of the measurement may lower detection limit presenting (higher quality) reference data better correlated to spectra (Viscarra Rossel *et al*., 2006). Exceptionally high prediction of otherwise spectrally non-responsive soluble and exchangeable Na (R^2 > 0.9) in

surface samples (Chodak *et al*., 2004), was probably partly due to the high sensitivity measurement (mg g^{-1} and μg g^{-1}).

Influence of spectral measurement region: The selected spectral measurement region commonly: UV-VIS-NIR (250-2500 nm), VIS (400-700 nm), VIS-NIR (400-2500 nm), NIR (700 – 2500 nm), and MIR (2500 – 25000 nm) has shown variable performance for estimation of soil properties (Canasveras *et al*., 2010). Key question from among comparative studies is the veracity of reported superiority of MIR over NIR (NIR region to include: UV-VIS-NIR, VIS-NIR, and NIR) (Madari *et al*., 2006; Viscarra Rossel *et al*., 2006) for estimation of primary and secondary soil properties, especially for independent samples datasets.

Performance of MIR and NIR: Reported superiority of MIR over NIR for estimation of several soil properties is predicted on the relative strength of absorbance spectra. Absorption peaks associated with key soil chromophores (SOM and clay minerals) are more pronounced in MIR and weaker (masked due to overlapping bands) in NIR. The NIR is insensitive to quartz, a main soil constituent, whereas silica shows strong absorption features in the region 1200-900 cm^{-1} (Viscarra Rossel *et al*., 2006) although Terhoeven-Urselmans *et al*. (2010) showed that calibration of total sand in MIR is surrogate. The superiority of MIR over NIR is not, however, adequately interrogated. Especially the difference in performance might be cancelled out by the practical advantages using NIR (Bellon-Maurel & McBratney, 2011).

Superiority of MIR might be conditional: test sample characteristics and analyte characteristics. The MIR and NIR, for example, indicated comparable performance for estimation of mean-weight diameter (MWD) for soils from southern Spain (Canasveras *et al*., 2010), however, MIR was clearly superior for estimation of MWD in soils from Brazil (Madari *et al*., 2006). The NIR was superior for total clay for soils from Spain

(Canasveras *et al*., 2010), whereas performance was comparable for soils from Brazil (Madari *et al*., 2006). Canasveras *et al*. (2010) concluded that the wavelength range providing the highest model fit (R^2 values) and accuracy is dependent on particular soil property. Important also is that for the same dataset, IR has shown superiority over NIR for some properties, the methods are comparable for others and NIR is superior for others (Canasveras *et al*., 2010).

MIR superior over NIR: MIR performed better for MWD (R^2 = 0.79 vs 0.66) and MGD (R^2 = 0.79 vs 0.67), for interval depth samples from deeply weathered soils from Brazil (Madari *et al*., 2006). MIR was superior for estimation of pHw1:5 (R^2 = 0.71 vs 0.65), OC (R^2 = 0.85 vs 0.76), tClay (R^2 = 0.72 vs 0.61) in interval depth samples of soils from NSW and Queensland Australia (Pirie *et al*., 2005).

NIR/MIR comparable performance: NIR and MIR were comparable for pHw (R^2 = 0.69 vs 0.68), $CaCO_3$ (R^2 = 0.91 vs 0.95), Fe_{CBD} (R^2 = 0.79 vs 0.81), and MWD (R^2 = 0.52 vs 0.59) (Canasveras *et al*., 2010). The methods compared well for TOC (R^2 = 0.93 vs 0.90), TN (R^2 = 0.99 vs 0.97), tSa (R^2 = 0.99), and tClay (R^2 = 0.96 vs 0.94) (Madari *et al*., 2006), and for eNa (R^2 = 0.20 vs 0.18) (Pirie *et al*., 2005).

NIR superior over MIR: NIR was superior for prediction of tClay (R^2 = 0.84 vs 0.77, water stable aggregates (fraction > 250 µm) (R^2 = 0.60 vs 0.56), and WDC (R^2 = 0.66 vs 0.30) (Canasveras *et al*., 2010). NIR was superior for CEC (R^2 = 0.83 vs 0.56) (van Groenigen *et al*., 2003), and for water stable aggregates (fraction <.250 µm (R^2 = 0.92 vs 0.80) (Madari *et al*., 2006).

Properties based on ratios of soil constituents (ESP, C: N ratio, clay activity, silt: clay ratio) may be predicted less accurately than the single constituents for NIR analysis (Stenberg *et al*., 2010). The opposite may be true, however, for MIR analysis. The ESP

was better calibrated to MIR than eNa (R^2 = 0.68 vs 0.43) (Viscarra Rossel *et al*., 2008), however, the ESP model involved more rigorous data pretreatment and was therefore less efficient.

Comparative studies suggested that the NIR might perform as well or sometimes better than MIR for soil properties whose determination is related with moisture content and for secondary soil properties. Important from these findings is that the performance difference should be interpreted in terms of intended use of the data considering also practicalities in data acquisition using MIR and NIR (Bellon-Maurel & McBratney, 2011). Further, superiority of MIR for estimation of especially secondary properties is conditional and depends on sample provenance. The MIR models were more complex (more PLS factors) for physical properties total sand (11 vs 3), silt (11 vs 2), clay (11 vs 3), mean-weight diameter (5 vs 3), mean geometric diameter (5 vs 3) (Madari *et al*., 2006). The NIR models were more complex for chemical properties (pH, EC, OC, eCa, eK, CEC) (Viscarra Rossel *et al*., 2006). Model complexity has implications for robusticity of calibration.

MIR – NIR trade off: MIR is superior over NIR for majority of primary soil properties, however, the difference is not profound and NIR is better commercially supported for routine soil analysis (Bellon Maurel & McBratney, 2011). NIR provide higher throughput and great potential for on-site and on-the-go field measurements (Stenberg *et al*., 2010), whereas MIR measurements under uncontrolled condition are rare (Merry & Janik, 2001).

Available information is not conclusive, however, and more comparative studies are needed that, especially address the basic factors that influence calibration performance (sample provenance, quality of reference data) and some standards on sample presentation and spectral data preprocessing.

Spectra data pretreatment and handling of outliers: Spectra data pretreatment including four main procedures: Detrend-D, SNV, MSC, derivatives and combinations[1] have been applied, in attempt to linearly align spectra absorbance to reference values, to optimize calibration (Genot *et al*., 2011*)*.

Characteristics of most of the numerous soil studies that applied rigorous spectra data pretreatment (Knadel *et al*., 2013; Muhati *et al*., 2011; Madari *et al*., 2006; Chodak *et al*., 2004; Moron & Cozzolino, 2003) include: exceptionally high determination coefficient (R^2) and/ or low error rates for some soil properties for calibration that involved less validation rigor, suggesting high risk of over-fit models. Also, none of the studies provide information on calibration performance without the spectra pretreatment (except Genot *et al*., 2011). Genot *et al*. (2011) reported R^2 for no-pretreatment and for the best of 15 different techniques (in parenthesis) as follows: 0.72 (0.79), 0.39 (0.58), 0.32 (0.36), and 0.51 (0.67) for total carbn, total nitrogen (TN), clay, and CEC, respectively, suggesting that the rigor was to advantage for only TN and CEC.

Optimal spectral pretreatment: Spectra data pretreatment is, however, resources intensive; oftentimes results in modest to slight improvement, and; optimal pretreatment is specific to the soil property (Sorensen & Dalsgaard, 2005). The utility of IR as a viable alternative to wet chemistry is anchored on simplicity, rapidity and precision on

[1] *(i) Detrend (D) removes the linear and quadratic curvature of each spectrum; (ii) standard normal variate (SNV) reduces the light scattering caused by particle size effects. Each corrected value is the original absorbance from which the mean of the whole spectrum is subtracted and divided by the SD of the spectrum, (iii) multiplicative scatter correction (MSC) eliminates or reduces the difference in light scatter between samples, (iv) first (1_4_4) or second (2_8_6) derivatives of the data is used to remove part of the particle size influence. The first number represents the number of the derivative (1st or 2nd); second number represent the gap between wavelength over which derivative is calculated, and third number represent the smoothing of the points).*

measurements. Testing the main four mathematical treatments of spectra Sorensen and Dalsgaard (2005) found the first derivative spectra without any scatter correction was optimal for estimation of SOC, clay, sand, and silt. The first derivative spectra and smoothing using a Savitzky-Golay filter as the only spectral pretreatment have provided optimal models for several soil properties (Stenberg, 2010; Viscarra Rossel *et al*., 2008).

Handling of outliers: Samples coming from outside the population covered by the calibration set may not fit the calibration. Successful calibration requires detection of both spectra (*X*- outliers) and samples falling outside the content range (*Y*- outlier) of the calibration set. The X-outliers are detected before calibration, whereas Y- outliers are checked after model development (Sorensen & Dalsgaard, 2005).

Spectral outlier: commonly X-outliers (or *H*- statistic) are identified using the principle of Mahalanobis distance *(H)* applied on principal component analysis (PCA) reduced data[2] (Deng *et al*., 2013). The aim is to calculate the Mahalanobis distance *(H),* which illustrates the way the spectra deviates from the average spectrum. When the *H* value exceeds the warning limit, there is a risk of increased uncertainty of the IR result and the risk increases with increasing *H* value (Sorensen & Dalsgaard, 2005). The setting of *H* value is, however, not standard and values: $H > 3$ (Genot *et al*., 2011); $H > 3.5$ (Sorensen & Dalsgaard, 2005), and; $H > 12$ (Terhoeven-Urselmans *et al*., 2010) have been applied.

[2] *Spectra PCA scores are used to calculate a standardized H distance (Mahalanobis distance) for each sample. Relative H distances, calculated for all soil samples in the population are used to define and reduce the population. The algorithm ranks the samples according to their H distance from the average scores calculated for the population. The H statistics is associated with spectral characteristics of the sample and represent the distance from a sample to the population.*

Reference values outlier: The Y-outlier (or *T*- statistic) is a sample whose residual (difference between measured and predicted value) is > 3 × RMSECV or RMSEP (Sorensen &Dalsgaard, 2005). Moron and Cozzolino (2003) defined reference outlier as sample with residual greater than 2.5 * RMSECV. A sample that fails the *H* and or *T* – statistics is recommended for re-scanning and re- analysis by the reference method, and those that persist are excluded from analysis (Sorensen & Dalsgaard, 2005).

Influential outlier: Influential outlier is characterized by high leverage and high residual and is identified via a plot of residuals vs leverage. Influential outlier presents also spurious prediction (extremely high or negative values) (CAMO ASA Inc., 1998).

Few studies (Genot *et al*., 2011; Sorensen & Dalsgaard, 2005) were explicit on handing (*H*- and *T*- statistic) outliers, although their removal might have implication on sample size (especially for modest sets, $n < 200$) and model robusticity. Influential outliers are excluded from analyses (CAMO ASA Inc., 1998). For robust calibration validation, the validation dataset should be within the range of calibration set for the majority of soil attributes and this is why reference values outliers are commonly identified for the validation set (Sorensen & Dalsgaard).

A challenge with soil chemometric and spectroscopic modeling is on how to extract predictive information, which consists of largely localized features of the spectrum, from noisy and strongly correlated data (Viscarra Rossel & Lark, 2009). Also, linear modeling techniques are not optimal for data with non-linear behavior, especially at very high analyte concentration (Linker, 2012). Spectra wavelet transform ensures reduction or elimination of both irrelevant and noisy features, and optimizes calibration by avoiding the *a priori* assumption of linearity (Linker, 2012).

Spectra Wavelet transformation: A wavelet (commonly a wave-like oscillation with amplitude that starts out at zero, increases, and then decreases back to zero) is a compact analyzing kernel (mathematical function) that can be moved over a sequence of data to measure variation locally. A wavelet transform is the representation of a signal/ spectrum by wavelets. Wavelets have special properties valuable for analysis of soil diffuse reflectance spectra including that: (i) its basis functions (mother wavelet) represent distinct scales of variation, (ii) a single basis function is localized (i.e., the coefficients correspond to features in a delimited part of the spectrum), and (iii) for many wavelets, adjacent coefficients at a particular scale are typically uncorrelated (Viscarra Rossel & Lark, 2009).

Compressing a large data set with the wavelet transform and then performing regression analysis on some of the wavelet coefficients is fast comparing to calculating the PLS model on the original full spectrum data set, and this can drastically increase the model's predictive ability and its stability (Gributs & Burns, 2006). Viscarra Rossel and Lark (2009) found that transforming soil spectra into the wavelet domain and producing a smaller representation of the data improved the efficiency of the calibrations; the models were computed with reduced, parsimonious data sets using simpler regressions, and; further improvement in modeling efficiency and accuracy was found for some attributes.

Discrete wavelet transform (DWT): There are a large number of wavelet transforms (Torrence & Compo, 1998; Walczak & Massart, 1997). The discrete wavelet transform (DWT) is most common in infrared spectroscopic applications (Viscarra Rossel & Lark, 2009; Trygg & Wold, 1998). The DWT analyzes the scale-position information of a spectrum, resulting in a set of wavelet coefficients associated with a range of scales and positions. Each coefficient is directly related to the amount of energy in the spectrum at a particular position and scale. Characteristic of DWT is the use of orthogonal basis (orthogonal wavelets) that operates over a specific subset of scale and translation values

or representation grid. The DWT effectively compress the spectral data into discrete wavelet coefficients in a procedure referred to as multi-resolution analysis (MRA) using the pyramid algorithm. The DWT and MRA apply only to a spectrum of discrete length 2^n, where n is an integer, and a spectrum will need to be padded to the nearest 2^n prior to the decomposition (Viscarra Rossel & Lark, 2009). This explains why the same number of wavelet coefficients could be extracted for both NIR and MIR despite the fact that these represent different spectral ranges and wavebands (data points).

Choice of wavelet and their deployment: There are numerous different mother wavelets (for instance, Daubechies and Coiflet), and appropriate choice of the specific (child) wavelet will depend on several factors including: the structure of the data, the goals of the analysis, and computational considerations (Koger *et al*., 2003; Bradshaw & Spies, 1992). The Haar analyzing wavelet of the Daubechies' family (Daubechies, 1992), is among the most popular (Viscarra Rossel & Lark, 2009), due to their simplicity and ease of computation.

Wavelets have been deployed to extract information from different kinds of data and in diverse fields of science including: analytical chemistry, geophysics, medicine, agriculture, forestry, and soil science (Linker, 2012). Wavelets have been effective for spectral data compression and denoising for feature extraction (Trygg & Wold, 1998), multivariate calibration (Viscarra Rossel & Lark, 2009), image processing (Bradshaw & Spies, 1992), and in the classification of infrared spectra (Koger *et al*., 2003). Wavelet analysis has potential also in conventional and digital soil mapping provided sufficient data can be obtained from survey, including exploration of wavelet bases in established libraries (Sanchez *et al*., 2009). Lark (2007) presented a case study in which a decomposition of sensor data on soil electrical conductivity resulted in better predictive models for other soil properties.

Potential of wavelet-based modeling in soil studies: A challenge with wavelet transform for soil spectroscopy studies remained, to establish whether reported advantage over alternatives like Fourier Transform (FT) is reflected in improved calibration and classification for all soil properties. The one available study by Viscarra Rossel and Lark (2009) found that predictions of soil organic carbon (SOC) and total clay content (tClay) using all significant wavelet coefficients were not significantly different to those using the original spectra, and only slight improvement in prediction of SOC and degradation in tClay was found using only relevant coefficients. Classification based on spectra wavelet transform models were only slightly more robust than models based on original spectra and on spectra PCs (for instance, accuracies of 87, 83 and 81 %, comparing wavelets, raw spectra and PCs, respectively) (Koger *et al*., 2003). It remained subjective also whether wavelet coefficients in modeling should routinely be treated as spectral or non-spectral data and; the requisite data preprocessing (mean-centered and with or without standardization). Scaling of the compressed wavelet coefficient data matrix sometimes presented a more parsimonious regression model; however, this could lead to perturbation of the systematic information, and hence, model instability (Trygg & Wold, 1998).

2.4 Exploring relationship in multivariate data

The complexity and large amount and variance of environmental data sets with a multiplicity of variables limit the use of univariate statistical methods for applications including assessment and monitoring of environmental health. Paradoxically, working with many variables expands the size of the input space, complicating establishment of relationships, technically referred to as the "curse of dimensionality" (Dwinnell, 1998). However, environmental data sets including soil spectra and also reference property data tend to have some degree of correlation among their variables, and this reduces the effective dimensionality of the input space. Correlated inputs imply that the data will cluster in a subset of the entire space and paring down the number of inputs often results

in models that are more accurate and more robust. Multivariate statistical methods including exploratory data analysis (EDA), principal component analysis (PCA), principal component regression (PCR) and partial least square regression (PLSR) (commonly PLS) are among soft modeling (unsupervised) methods commonly deployed to explore multivariate empirical data sets and establish relationships (Maindonald & Braun, 2003).

2.4.1 Development of quality attributes and parametric modelling

Principal Component Analysis (PCA): PCA is a method of data reduction that applied to a data matrix of samples by variables (for instance soil spectra and/or reference data), it constructs new variables (latent variables, hidden variables, linear combinations, or principal components*)* such that: the new variables are linear combinations of the original ones; are uncorrelated (orthogonal); the first new variable captures as much as possible of the (co-) variability in all the original variables; each successive new variable accounts for as much of the remaining variability as possible (Naes *et al*., 2002). PCA makes it possible to describe a very large proportion of the variability in highly multivariate (and collinear) data with a modest number of principle components (PCs). The analysis places the PCs in order, using as an ordering (importance) criterion the amounts that the individual PCs contribute to the sum of the variances of the original variables. Usually, only the first few PCs contain genuine information, while the later PCs most likely describe noise. Therefore it is useful to study the first few PCs only instead of the whole raw data table. This is less complex and also ensures that noise is not mistaken for information (CAMO ASA Inc., 1998). The output of PCA has been interpreted using among others, the retained PCs through graphical inspection (scores plots, residual/explained variance plot) and numerical loadings (Martens & Martens, 2001).

A potential challenge is that with the different methods of extraction of important PCs (varimax rotation with Kaiser Normalization criteria of eigen-value > 1, the scree test and the broken- stick eigen-value criteria) (Maindonald & Braun, 2003), the same data set might present different interpretations even for the same objective. Maindonald and Braun (2003) caution that although PCA is tauted the 'mother' of all multivariate analysis methods, appropriate data set, and requisite preprocessing and indeed relevant optimal PC extraction criterion is necessary for a valid and meaningful PCA construct.

Appropriate dataset and requisite preprocessing: most of the variables should indicate moderate to high correlation. PCA exploits the noted correlations to produce a smaller set of new variables in such a way that most of the variability between the samples measured on the original variables is maintained (Maindonald & Braun, 2003). Preprocessing aims to remove or reduce irrelevant sources of variation or 'noise'. Pretreatment might include a monotonic transformation[3] (for instance, log, inverse, square, or square root- SQRT) to reduce data heterogeneity, and column mean - centering (variable means are subtracted from each element in their respective columns). Logarithmic (log) transformation, for example, makes the distribution of skewed variables more symmetrical, and is used also when measurement error on a variable increases proportionately with the level of that variable (ensures a relative uniform precision over the whole range of variation or variable stabilization). In the case of only slight asymmetry, SQRT can serve same purpose as log (CAMO ASA Inc., 1998).

[a] *A (positive) monotonic transformation is a way of transforming one set of numbers into another set of numbers so the the rank order of the original set of numbers is preserved. It is thus a function, f, mapping real numbers into real numbers, which satisfies the property that, if x > y, then f(x) > f(y).*

Where input variables come in very different measurement scales (for instance unitary, %, cmol (+) kg^{-1}, mS cm^{-1}, mg kg^{-1}), and/or are suspected to have very different error levels (measured by variance and SD), the variables are weighted, commonly by standardization or scaling. Standardization (1/SD) allows comparison of the variation of different variables by expressing the variable variance in the same standard unit, thus giving all variables the same chance to influence the estimation of the components (CAMO ASA Inc., 1998). Martens and Martens (2001) refer to scaling as 'balancing the error levels in the input variables'. Variables of the same type and measured on the same or similar scale for instance spectral data, and those based on linear measurements, however, should be analyzed in their original form (CAMO ASA Inc., 1998). Importantly, variable data preprocessing for PCA is standard procedure as demonstrated by some workers (Muhati *et al*., 2011), and should be explicit to avoid pitfalls of presenting highly redacted analyses. A potential drawback with scaling is that it might increase the relative influence of unreliable / noisy properties (Martens & Martens, 2001).

Multivariate calibration: The partial least square regression (PLSR) and principal component regression (PCR) are probably the two most common linear methods applied in soil chemometrics and spectroscopic studies (Linker, 2012; Stenberg *et al*., 2010). PLS and PCR are data projection methods whose algorithm reduces large numbers of correlated data into a limited number of orthogonal components, which are used as independent variables in a multiple linear regression (MLR) on dependent variable(s). Non-linear models (McBratney *et al*., 2003)[4] have, however, outperformed linear methods for estimation of some physico-chemical soil properties (Viscarra Rossel &

[4]*Non-linear or data mining methods including: multivariate adaptive regression splines (MARS), classification and regression tree (CART), artificial neural networks (ANN), TreeNet, boosted regression trees (BRT), random forests, and fuzzy systems.*

Behrens, 2010; Brown *et al*., 2006). Compared to PCR, PLS only selects components that are relevant to the requested model i.e., strong components that are not correlated to the soil variable in question, are not included in the final model. PLS results, therefore, in more parsimonious models with similar or better predictive capabilities (Naes *et al*., 2002). Numerous studies (Knadel *et al*., 2013; Genot *et al*., 2011; Stenberg *et al*., 2010; Viscarra Rossel *et al*., 2006) are based on PLS.

Calibration spectrometry: Three metrics: coefficient of determination or model fit (R^2), accuracy or root mean square error (RMSE) and model reliability (RPD) are commonly used to assess IR PLS performance for estimation of soil properties. Important is that model interpretation might be subjective when either of the statistics (R^2 and RMSE) is used independently (Bellon-Maurel & McBratney, 2011).

Coefficient of determination: The R^2 value is an expression of the percentage of variance present in the measured property values, which is reproduced in the prediction and approaches 100 % as the predicted values approach the true values. The R^2 is dependent on the measurement range, however, and when used alone could be a subjective indicator of prediction performance. Reported R^2 = 0.13, 0.07, and 0.04 for estimation of CEC, exchangeable calcium - eCa and electrical conductivity - EC, for example, might suggest no predictive relationship between spectra and these properties. Fair correspondence was found, however, in the data range (min – max) for observed and predicted (in parenthesis) values: [(CEC: 24.0 - 72.0: (29.7 - 56.0) mmol (+) kg^{-1}; eCa: 11.0 - 46.0: (13.2 - 37.5 mmol (+) kg^{-1}; EC: 0.02 - 0.07 (0.04 - 0.06) dS m^{-1})] (Viscarra Rossel *et al*., 2006), probably sufficient to allow quick management decision. This suggested that model performance should be interpreted based also on the purpose as suggested by Sorensen and Dalsgaard (2005). For cases with genuinely poor model fit, spectral screening using critical values of the analytes (Shepherd & Walsh, 2002; 2007), might be more informative than quantitative calibration.

Prediction accuracy/error: The RMSE present a more efficient measure of the uncertainty on future prediction *(Yp)* (*Yp* = ± 2SE×RMSE), whose measure is valid provided that the new samples are similar to the ones used for calibration (otherwise, the predicted error might be much higher) (Bellon-Maurel & McBratney, 2011). However, different measurement scales for the same soil property could constraint comparing error rates from different studies. It is not straight forward comparing RMSE and corresponding data range values, for example, for texture parameters for reference values expressed in g kg^{-1} (Araujo *et al.*, 2013), kg kg^{-1} (Chang *et al.* , 2001), mg g^{-1} (Couillard *et al*., 1997), or (the uncommon) dag kg^{-1} (Viscarra Rossel *et al*., 2006). Inter-conversion of the measurement scale might not be trivia. The use of R^2 and RMSE could present challenges also when comparing calibrations with different ranges for a constituent. The RPD has been proposed to boost R^2 and RMSE since it provides a more 'universal' statistic allowing comparison across soil properties and diverse databases (Chang *et al*., 2001).

RPD statistic: The RPD (ratio of prediction deviation) attempts to standardize the value of the RMSE, with respect to the natural dispersion of the samples (RPD = SD ×$RMSE^{-1}$). Larger values of RPD indicate better fitting models. The RPD has been widely adopted by the soil community (Knadel *et al*., 2013; Canasveras *et al*., 2010; Terhoeven-Urselmans *et al*., 2010; Pirie *et al*., 2005) since Chang *et al*. (2001) who set up RPD thresholds to gauge the performance of the prediction, although Bellon-Maurel *et al*. (2010) question the relevance of this parameter since it depends on the range and distribution of the population.

Calibration performances thresh holds: No universal standards have been established although largely comparable ratings for R^2 and RPD have been proposed. No thresholds for RMSE were available in the literature. The values of $0.5 < R^2 < 0.65$ were fixed to

indicate the possibility to differ high to low concentration of a soil attribute in a NIR model, and; values of R^2 0.66 - 0.81, 0.82 - 0.90, and > 0.90 were proposed to indicate, acceptable, good and excellent quantitative models, respectively (Araujo *et al*. , 2013). The R^2 values > 0.71 are satisfactory and values < 0.5 weak (Williams, 2001). Other workers (Saeys *et al*., 2005) proposed R^2 as follows: > 0.8 (very good), 0.8 – 0.7 (good), 0.6 – 0.5 (approximate), 0.5 – 0.3 (very approximate) and < 0.3 (not predictable). Notably, spectra of organic material (for instance Hog manure used by Sayes *et al*. (2005)) and cereals by Williams (2001)) is less complex compared with soil spectra and models from organic materials are better performing, suggesting that the thresh holds might be rather conservative for soil studies. Terhoeven-Urselmans *et al*. (2010) considered $R^2 > 0.75$; $0.65 < R^2 < 0.75$, and; $R^2 < 0.65$ to indicate good, satisfactory, and poor performance, respectively for MIR analysis.

RPD threshold: three categories of RPD, > 2.0; 2.0-1.4; < 1.4 and corresponding R^2: 1.0 - 0.8, 0.8 - 0.5, and < 0.5, respectively, indicated decreasing reliability of prediction of soil properties using NIR (Chang *et al*., 2001). Prediction of soil properties in the middle RPD category could be improved by using calibration optimization strategies including improving quality of reference data (Reeves, 2010), whereas properties in the lower category may not be reliably predicted by NIR (Chang *et al*., 2001). The RPD categories: ≥ 2.0; 1.5 < RPD < 2.0, and; < 1.5 are associated with high, moderate and poor reliability of MIR PLS analysis, respectively (Viscarra Rossel *et al*., 2008).

Calibration validation requirement: Soil chemometrics and IR modeling requires both calibration and validation stages; that the validation set data range is within the range of calibration set, and; that the calibration models are appropriately validated *(*Brown *et al.,* 2005*)*. Validation strategies include: leverage correction, cross validation, separate test set validation, and independent set validation *(*Naes *et al.,* 2002*)*. Leverage correction is effective for rapid selection of a representative set from a large sample set using, for

example, PC scores of spectra, however, the strategy can sometimes be highly over-optimistic and is not appropriate for actual analysis *(*Naes *et al.,* 2002*)*.

Cross validation: The leave-out-one cross validation (looCV) strategy provides a good estimate of the uncertainty of the model when the calibration samples are properly selected using for instance, a conditioned Latin hypercube sampling *(*Stenberg, 2010*)*. The potential risk is that looCV may underestimate the ruggedness of the calibration and the predicted uncertainty because validation samples are taken from the pool of samples used for calibration (Sorensen & Dalsgaard, 2005). Independent set is different from the calibration set both spatially and temporary *(*Sorensen & Dalsgaard, 2005*)*.

Independent validation: Few studies (Terhoeven-Urselmans *et al.,* 2010; McCarty *et al.,* 2002*)* demonstrated completely independent validation although the veracity of chemometric- based models could only be established using test sets, preferably collected after model calibration *(*Dardenne *et al., 2000)*. Testing calibration models of properties in bulk soil samples (BS) on spectra and reference values in aggregate size fractions (ASF) of the BS *(*Madari *et al*., *2006)* or interchanging surface and subsurface datasets as calibration and validation sets *(*Nanni & Dematte, 2006*)* are strategies, for example, that run the risk of pseudo-independent validation *(*Brown *et al*., 2005*)*, since the calibration and validation samples might be pseudo-replicates *(*Terhoeven-Urselmans *et al*., 2010*)*. The use of samples sets from the relatively more homogeneous surface horizon and a restricted clay range (< 26 %) for independent prediction of soil organic carbon and total clay *(*Sorensen & Dalsgaard, 2005*)*, might be also suspect for independent validation.

The requirement for calibration independent validation is especially critical/ basic for spectrally non-responsive soil properties characterized by high variation in (surrogate) calibration performance *(*Reeves, 2010*)*. A drop in performance for independent

datasets is to be expected, even for spectrally active soil constituents, attributed to variable sample characteristics from different provenance *(*Reeves, 2010; Minasny *et al*., 2009).

2.4.2 Deployment of soil and spectra for pedotransfer purposes

The role of aggregation and aggregate stability (AS) in soil health is well elucidated (Wei *et al*., 2006). The AS indicates significant association with other soil properties (Igwe & Nwokocha, 2005). Available works (Canasveras *et al*., 2010; Idowu, 2003) have developed aggregation proxies from multivariate calibration of a limited number of soil properties although multiplicity of properties could potentially influence soil function (Minasny, 2007). This is attributed partly to limiting/ stringent data quality requirement for successful calibration. Few studies (Canasveras *et al*., 2010; Madari *et al*., 2006) were available also on development of spectra-based pedotransfer functions for estimation of soil aggregation indices. There were scanty information (Canasveras *et al*., 2010) comparing spectra and soil-based properties for pedotransfer purposes.

Reported performance for estimation of aggregation indices by alternative spectral predictors (NIR and MIR) is variable and inconclusive: R^2 = 0.92 - 0.80 for stable micro aggregates (fraction < 250 µm) (Madari *et al*., 2006); R^2 = 0.60 - 0.31 for stable macro aggregates (fraction >250 µm) and R^2 = 0.66 - 0.28 for aggregated clay (< 20 µm) (Canasveras *et al*., 2010), and; R^2 = 0.60 - 0.46 for aggregation indices: > 2000, 1000, 500, 250 µm, and > 250 µm fractions *(*Chang *et al.,* 2001*)*. Performance for estimation of aggregation indices by soil-based predictors is also variable. The R^2 values 0.23, 0.43 and 0.49 were found for prediction of aggregation indices wet stable aggregates, mean-weight diameter, and water-dispersible (WDC), respectively, using six soil-based predictors. Corresponding R^2 0.60, 0.62, and 0.66, respectively, found using spectral predictors indicated better performance. A combined dataset of spectra and the six soil predictors was also better than the soil-based predictors (R^2 = 0.53, 0.60, and 0.47,

respectively), for estimation of the indices (Canasveras *et al*., 2010). This suggested potential of developing effective spectral and /or in combination with soil-based pedotransfer functions for estimation of aggregation indices. The inclusion of WDC as aggregation index by Canasveras *et al*. (2010) is insightful; however, this might be subjective considering rules of pedotransfer functions (Minasny, 2007). The release of WDC defines, however, stability for soil types (for instance Oxisols) whose aggregation is not supported by the hierarchical model (Azevedo & Schulze, 2007). Soil compositional complexity would suggest also that soil aggregation interacts with many more of the basic soil properties than the six found by Canasveras *et al*. (2010).

Prediction via pedotransfer functions and inference systems: The current challenges in direct calibration of soil properties to spectra has led to ' emerging applications' that improve on prediction performance and help in solving also the "model transferability" problem (Linker, 2012). In the new approaches estimates obtained by (MIR) spectroscopy are used in an inference system that uses pedotransfer functions for estimation of (functional) soil properties with no associated distinct spectral absorbance bands. Once regression models based on (MIR) spectra have been calibrated for properties with predictive relationship (such as clay, mc, and OC), these models are used to estimate the same properties of a new sample, together with their uncertainties. These values are then used to predict the functional attribute(s) (McBratney *et al*., 2006). An important feature of this approach is that the determination errors are propagated through the various estimation stages so that a confidence interval can be associated with each prediction (McBratney *et al*., 2006). This approach improved considerably the estimation of soil water retention, which was very poorly estimated directly from the MIR spectra (Tranter *et al*., 2008). More parsimonious models and greater accuracy was reported when using an inference system, than when using straightforward PLS regression on the MIR spectra (Minasny & McBratney, 2008).

Analogously, selected soil-based predictors of wet stable aggregation (WSA) indices would be calibrated to MIR spectra; the models are used to predict the values in samples from a new (larger) set using their MIR spectra, and; the predicted values are then used to estimate the WSA indices in the large set. This approach appeared most appropriate for diagnostic screening and prevalence assessment of stability related problems for large area soil health assessment and monitoring using spectral libraries and the double sampling approach (Shepherd & Walsh, 2007).

The appropriateness of selected aggregation indices (response) and the (predictive) relationship with established soil predictor variables and with IR-based predictors might influence opportunity of response to be estimated by selected regression methods (Canasveras *et al*., 2010; Idowu *et al*., 2008). This has attendant influence on efficacy of developed soil health status.

2.4.3 Relationship of basic properties and functional attributes: CART regression

CART an acronym for Classification and Regression Trees (Breiman *et al*., 1984) is among adaptive (data mining or "machine learning") modeling techniques that are based on highly automated search procedures to discover patterns and relationships in data that may be used to make valid predictions. CART can be used to advantage as a stand-alone package or as a pre-processing complement to other modeling packages.

CART operation and quality features: As a data preprocessor, CART can extract the most important variables from a very large list of potential predictors. CART algorithm bypasses "noise" and irrelevant variables, quickly and effectively selecting the best variables for input. This is attributed to inherent quality features including that CART is non-parametric and can handle numerical data that are highly skewed. Operationally, CART is not affected by outliers, missing values, multi-collinearities, heteroskedasticity or distributional error structures that affect parametric procedures. This permits minimal

preprocessing of the data. CART performs better when there are numerous independent variables, but the final decision tree will only include the independent variables that were predictive of the dependent variable (Steinberg & Golovnya, 2006; Steinberg & Colla, 2001).

The cutoff points CART uses to generate splits on continuous variables indicate where these variables may be split into categorical variables (Steinberg & Colla, 2001), allowing establishment of critical (thresh hold) values effective for diagnostic screening (Tittonell et al., 2008), for instance delimitation of low/high stability categories. A key feature of CART analysis is it's seemingly simplicity with semi-automation (in selection of variables, internal tests, and optimal decision trees), and ease with data exploration. However, some CART outputs could be counter-intuition and rationalization of the outputs is needed to allow making logical interpretation and valid conclusions (Steinberg & Colla, 2001).

CART application studies: The CART approach to multivariate data analysis has been used increasingly in diverse fields of research including medical (Lemon *et al*., 2003); network security systems (Srilatha *et al*., 2005); forensic science (Phelps & Merkle, 2008); food security and socio-economics (Yohannes & Webb, 1999), and; agriculture and soil science (Tittonell *et al*., 2008; Yang *et al*., 2004). In soil science, CART has gainfully been applied in developing screening tests for environmental health monitoring including soil physical degradation (Omuto & Shrestha, 2007); assessing organic quality attributes (Shepherd *et al*., 2003); soil fertility constraints (Shepherd & Walsh, 2002; 2007), and; establishing crop-yield controlling factors (Tittonell *et al*., 2008).

Paucity in CART regression studies: Characteristic of majority of the reported studies is that they are based on CART classification tree (McBratney *et al*., 2003) and only few (Tittonell *et al*., 2008) applied CART regression tree analysis. This was probably

attributed to paucity of information on veracity of predicted values from CART regression for defining critical values to enable assigning cases into different categories. Tittonell *et al*. (2008) suggested and demonstrated an approach that provides insight for establishment of robust thresholds for diagnostic screening from CART regression. The authors noted the need for making some subjective decisions, such as defining cut-off values for dividing variables into discrete classes, and defining the acceptable error in the final model, and that the decisions should be made explicit. Among great potential of CART regression is the efficiency in selecting key variables from a complex multivariate data set of potential predictors without significant loss of information (Tittonel *et al*., 2008).

2.4.4 Soil spectral libraries and diagnostic screening

The spectral libraries approach provides a tool for generalizing results of soil assessments that are conducted at a limited number of sites, and thereby increasing the efficiency of expensive and time- consuming soil-related studies (Shepherd & Walsh, 2002; 2007)). The approach is critical as a tool for advice on land management, such as the classification of new samples into basic soil quality classes, particularly in data-sparse situations (Terhoeven-Urselmans *et al*., 2010).

Spectral libraries for prediction and screening: The spectral libraries approach has been deployed to great effect for estimation of several soil properties (Terhoeven-Urselmans *et al*., 2010; Viscarra Rossel *et al*., 2008; Brown *et al*., 2006). Satisfactory performance was found, for example, for calibration of several soil chemical properties in a large set (n = 582) of topsoil samples collected from forest-cropland chronosequence age classes (Awiti *et al*., 2009). Good predictions for fertility valuable soil properties were reported for a globally distributed (n = 971) set (Terhoeven-Urselmans *et al*., 2010). Successful spectral screening tests developed for soil properties that were otherwise moderately calibrated to spectra (Shepherd & Walsh, 2002),

suggested that spectral screening would allow satisfactory classification of soils into basic quality classes. Following this lead other workers (Viscarra Rossel *et al*., 2010; deGraffenried Jr. & Shepherd, 2009; Vågen *et al*., 2006) have developed spectral screening tests for benchmarking soil condition and developing quality indices. Notable was paucity on deployment of reflectance spectra for screening stability problems although several workers (Reeves, 2010), related soil structure with reflectance spectroscopy.

Land degradation surveillance framework (LDSF): the LDSF (Walsh & Vagen, 2006), uses the spectral libraries approach for assessing large area land degradation and risk. Rigorous case definitions for target soil problem are established to define 'affected' and 'non-affected' states. Spectral screening tests are then developed to be able to rapidly dichotomize test observations to 'affected' or 'non-affected' categories (Shepherd & Walsh, 2007). Important is that the case definition itself may be defined in relation to the screening test where an arbitrary cut-off value of the screening test is used as a decision threshold for intervention (Tittonell *et al*., 2008). The framework applied to a 10*10 km block (sentinel site), uses a novel field sampling protocol (Vågen *et al*., 2013), whose geo-referenced sampling plots data can directly be linked to satellite imagery data. This provides opportunity to further develop scale flexible digital attribute maps that are amenable to temporal monitoring (UNEP, 2012; Sanchez *et al*., 2009). Important also is that the approach provides prevalence data on the target soil problem, a basis for quantifying risk factors and a baseline for change detection (Shepherd & Walsh, 2007). Vågen *et al*., (2006) demonstrated, for example, a dominant effect of historic land use on the spectroscopy-derived soil condition index, with areas originally under forest having higher soil fertility than areas originally under grassland, whereas current land use had relatively little effect on soil condition.

2.5 IR soil-based studies in Kenya

Several workers have demonstrated the potential of IR in soil studies in Kenya. These include works relating spectra to aggregation and aggregate size fractions (Verchot *et al*., 2011; Mutuo *et al*., 2006); assessing and monitoring changes in soil quality attributes (Shepherd *et al*., 2003; Awiti *et al*., 2009); soil erosion condition and condition benchmarking (deGraffenried Jr. & Shepherd, 2009); assessing soil physical degradation (Vågen, 2009; Omuto, 2008); assessing soil chemical and microbial properties within riparian wetland systems (Cohen *et al*., 2005); in soil nutrient diagnostics (Muhati *et al*., 2011), and; soil classification to swelling potential based on spectrally derived clay mineralogy differences (Kariuki *et al*. , 2003ab). Reflectance spectra is being used also as a soil monitoring tool in Kenya, among other African countries in the Africa Soil Information Service (AfSIS) project (Vågen *et al*., 2013; Shepherd, 2010; *www.AfricaSoils.net;)* to assess the impact of soil management interventions as part of a land degradation assessment framework applied to development projects over large areas.

2.6 Review concluding remarks; summary and research gaps

Soil aggregation and aggregate stability (AS): The behavior of a soil regarding the dispersion and aggregation of its particles is important for the development of environmental and agricultural soil functions. Soil aggregation and its stability (AS) play a critical role in both the physico-chemical and biological spheres of soil health. The relationship of AS with other soil properties; soil response to natural and /or cyclic conditions; land management practices, and; interaction with soil organic matter (SOM), is all attributed to sensitivity of the indices. Interaction of AS with SOC is complex; especially the uncertainty in the dynamics of SOM in macro and micro aggregate stability, important for C-sequestration, and more research is needed.

Sodicity is a major risk factor for dispersion, however, not all sodic soils are dispersed and certainly not all dispersive soils are sodic since several other physicochemical and mechanical factors also contribute to soil dispersion. Slaking is more a physical than a chemical process and contribution of exchangeable sodium (eNa) to aggregate slaking is not well supported, whereas influence of eNa could be significant with capillary wetting (controlled slaking) of aggregates. Oxisols and sodic soils indicate similar behavior regarding their aggregate breakdown and dispersion in water: Oxisols exhibit direct (mechanically-induced) dispersion of aggregates without first slaking; Sodic soils exhibit direct (spontaneous) dispersion, and; in both the hierarchical model concept for aggregation is not supported.

Various studies applied active selection (or exclusion) of study sites and/or data from tested parameters in analyses so as to establish significant relationships between AS and soil physicochemical properties and avoid confounding the results. Effective advances in better understanding and modeling relationship of AS with other properties would require, however, unbiased selection of study sites following, for example, randomized systematic sampling schemes.

A universal method for AS and indices might not be feasible and rigors might not be worthwhile. Angers *et al*. (2008) argument appears definitive: provide a detailed account on sampling, transport, storage, pretreatment and treatment methods, measurement indices for the specific purpose and method used. This could ensure effective communication and comparison. More important for wet-sieving approaches, is that the choice procedure is fast, simple, and ruthlessly reproducible and the indices highly sensitive for intended purposes.

Establishing relationships in multivariate data: The utility of principal component analysis (PCA) is anchored on the principle that information can be assimilated in

variation. Extracting information from a data table means finding out what makes one sample different or similar to another. PCA can be a very useful method in itself and can form the basis for several classification (soft independently modeling class analogy, SIMCA) and regression (PCR, PLS, MLR) methods.

Moderate to strong association of majority of soil properties in a multivariate non-designed (empirical) dataset is requisite for successful PCA construct. Requisite is also adequate and appropriate preprocessing (including summary statistics for quality control). These need to be explicit to avoid pitfalls of presenting highly redacted information.

Diffuse reflectance infrared spectroscopy (IR): The IR used in the laboratory can provide a framework for rapid and simultaneous prediction of several key soil physical and chemical properties, either directly or through secondary (surrogate) calibration. Performance of IR for estimation of aggregation indices and textural attributes is unreliable and at best modest, and IR-based predictor variables are superior over basic soil properties for prediction of soil aggregation indices. Aquaphotomics (water-light interactions) attributed to Stenberg (2010) could provide insights on mechanisms for prediction (direct or indirect), for especially soil properties whose determination is related to soil moisture content.

Factors influencing IR calibration performance: IR soil calibrations are influenced by several factors key among which are: sample characteristics especially provenance, property characteristics including spectral response and quality of reference data, the spectra measurement region, and the calibration method. Greater variation in calibration performance is to be expected for different samples datasets, especially for spectrally non-responsive soil constituents (including pH, EC, extractable, soluble and exchangeable bases and derived attributes). The rigor in spectral data pretreatment for

modest calibration improvement is not justified, considering that the first derivative spectra and smoothing provides optimal calibration performance for several soil properties, and; the utility of IR as alternative to wet chemistry methods is premised on simplicity, rapidity and precision. Accounting for outliers has implication for calibration performance; however, no studies were found that assessed the effect on prediction performance with and without accounting for both spectra and reference values outliers in both calibration and validation datasets.

Most appropriate spectral region: Reported MIR (2500 – 25000 nm) superiority over NIR (NIR to include: UV-VIS-NIR, VIS-NIR, and NIR) is not adequately supported, especially for spectrally non-responsive soil properties, and NIR could be superior for properties related to soil moisture content, including aggregation indices. The higher equipment maintenance (liquid nitrogen) for MIR, higher throughput with NIR and the demonstrated potential of NIR for on-site and on-the-go field measurements, could give NIR spectrometry practical advantage over MIR. Calibration performance for derived, especially ratio-based soil properties is lower than for directly measured individual properties for NIR PLS analyses, whereas the contrary might be the case for MIR PLS analyses.

Calibration geographic transferability: model transferability remains a key challenge for wide use of IR in soil studies, particularly for routine soil analysis. Chemometric and spectra-based PLS calibrations are inherently local and specific to applied datasets, and appropriate calibration validation strategy must be explicit to avoid pitfalls in (presumed) calibration robusticity prevalent in contemporary studies, for especially secondary properties. Independent set is different from the calibration set both spatially and temporary. Model transferability is more successful for primary than for secondary properties.

Spectra wavelet transform: There may be theoretical advantages in using wavelets over alternative data compression methods, however, the final proof will lie in whether they provide consistently better predictive performance on independent validation data sets than alternatives. Optimizing spectral transforms for each calibration data set and variable can just lead to over fitting. Scanty information on performance of spectra wavelet transform vis a vis Fourier transform for calibration of soil properties is inconclusive.

Relationship of basic properties and functional attributes: Screening of multiple soil properties for selection of a few key predictor variables (through CART regression), avoids potential pitfalls associated with inclusion of poor quality (noisy) predictor variables.

The CART methodology and the results from CART regression analysis suggest that this technique offers considerable potential for assisting in the analysis of large multivariate and complex soil datasets. Seemingly simplistic, CART regression decision tree efficiently uncover the predictive structure of the problem, unveiling the variables or interaction of variables that are responsible for a given phenomenon and that best determine one outcome rather than another. Such predictor variables or regression rules enable predicting the mean or median value, of future observation (s) from the profile characteristics submitted for analysis.

CART can effectively be used to improve conventional (parametric) models where CART can assist in variable selection (to pare a large number of potential predictors to few key variables), and; to suggest breakpoints for converting continuous to categorical variables. Few soil studies (Tittonell *et al*., 2008) have used CART regression trees for final analysis, whereas most CART-based studies are classification problems. Studies in other fields (medical, insurance, banking, security) present insights that suggest potential

success of CART regression in soil studies. CART is generic (as opposed to dedicated) software and can produce wide range of outputs (some spurious) for presented input data. For effective application including interpretation of CART regression outputs, there is need for contextual knowledge on the problem, adequate analytical skills and prognostic intuition.

Spectral libraries for calibration and diagnostic screening: There is paucity in information on the use of soil reflectance in pedotransfer functions for estimation of soil functional attributes. Soil spectral libraries approach provide a framework for effective development of calibration models and spectral screening tests for diagnostic screening and prevalence assessment of soil stability related properties to an accuracy level that is acceptable for large-area applications.

New modeling approaches has demonstrated the use of MIR spectroscopy in an inference system that uses pedotransfer functions for robust estimation of (functional) soil properties that otherwise indicate poor direct correlation to MIR. In an analogous way selected soil-based predictors of wet stable aggregation (WSA) indices would be calibrated to MIR spectra; the models are used to predict the values in samples from a new (larger) set using their MIR spectra, and; the predicted values are then used to estimate the WSA indices in the large set. This approach was considered most appropriate for 'diagnostic screening and prevalence assessment of stability related problems using diffuse reflectance infrared spectra' reported in this study.

CHAPTER THREE
MATERIALS AND METHODS

3.1 Study area and sample collection.

The study area falls within the Lake Victoria Basin (LVB) in the western part of the Republic of Kenya in an area covering approximately 46,400 km^2 and bound by latitudes $0^0$7' 48"N and $0^0$24' 36"S and longitudes $34^0$51'E and 35^0 43' 12"E (Figure 3.1). The area is within lowland LVB (1400 m asl and below contour) of Kenya and comprises of a wide-range of stability sensitive soils (Solonetz, Planosols, Vertisols, Luvisols), developed predominantly on Recent undifferentiated unconsolidated deposits from various sources. The area was within pilot districts of the Kenya Agricultural Productivity Project (KAPP), where the problem of sodium-affected soils has been identified among major constraints to smallholder agricultural productivity (Waruru *et a*l., 2003a). Two sets of soil samples were collected: a calibration set and a validation set.

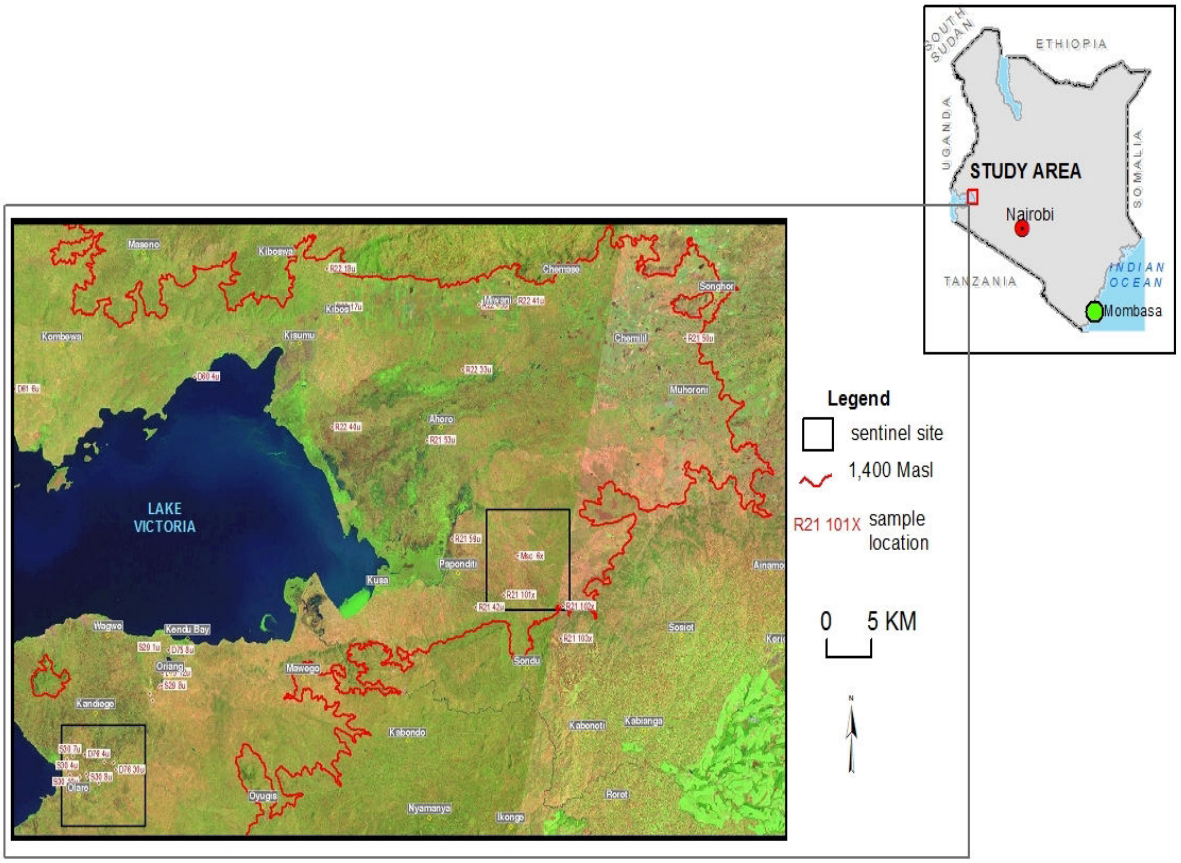


Figure 3.1: Location of study area and distribution of calibration set sampling sites within Lake Victoria basin of Kenya.

3.1.1 Calibration and validation sets

Calibration set: The calibration set (*n* =136) was obtained from a total of 46 representative sites. Selection of the sites followed a simplified version of a conditioned Latin hypercube sampling (Minasny & McBratney, 2006). The distribution of calibration sampling sites within lowland (< 1400 m asl contour) LVB is illustrated in Figure 3.1 (also indicating two sentinel sites used for obtaining validation samples). At each site in the field soil samples were collected at three depth intervals (0-20, 20-50, 50-100 cm) using a Dutch soil auger.

Validation set: The validation set (n =120) was obtained from a larger sample set (n = 417) from two different and spatially separated sentinel sites (10×10 km blocks) of Homa Bay (HB) (n = 207) and Lower Nyando (LNY) (n = 210) within LVB (see inset in Figure 3.1), one year after collection of the calibration set. Location of HB and LNY sentinel sites was selected to represent the major soil types, landforms, land cover and land use within the study area ((Waruru *et al*., 2003b). The sampling design and sample collection in the field followed the land degradation surveillance framework (LDSF) protocol (UNEP, 2012; Walsh & Vågen, 2006). At each site a 10×10 km block was demarcated and spatially stratified into 16 sub-blocks each 2.5×2.5 km. Within each sub-block a one km^2 cluster was randomly located. Within each cluster, ten (10) sampling plots were randomly selected from a possible 1000 plots" . Figures 3.2 and 3.3 illustrates layout and distribution of sampling plots within clusters in LNY and HB, respectively. In the field, samples were collected at three depth intervals (0 - 20, 20 - 50, 50 - 100 cm) using a Dutch soil auger.

Prior to analyses, bulk soil samples were air-dried in a room at 40 ^{o}C for two weeks followed by gently crushing to pass a 2-mm sieve. Each air-dried bulk sample was thoroughly homogenized and a representative sub sample taken using coning and quartering (Zobeck, 2004). Sub-samples were then used for various analyses.

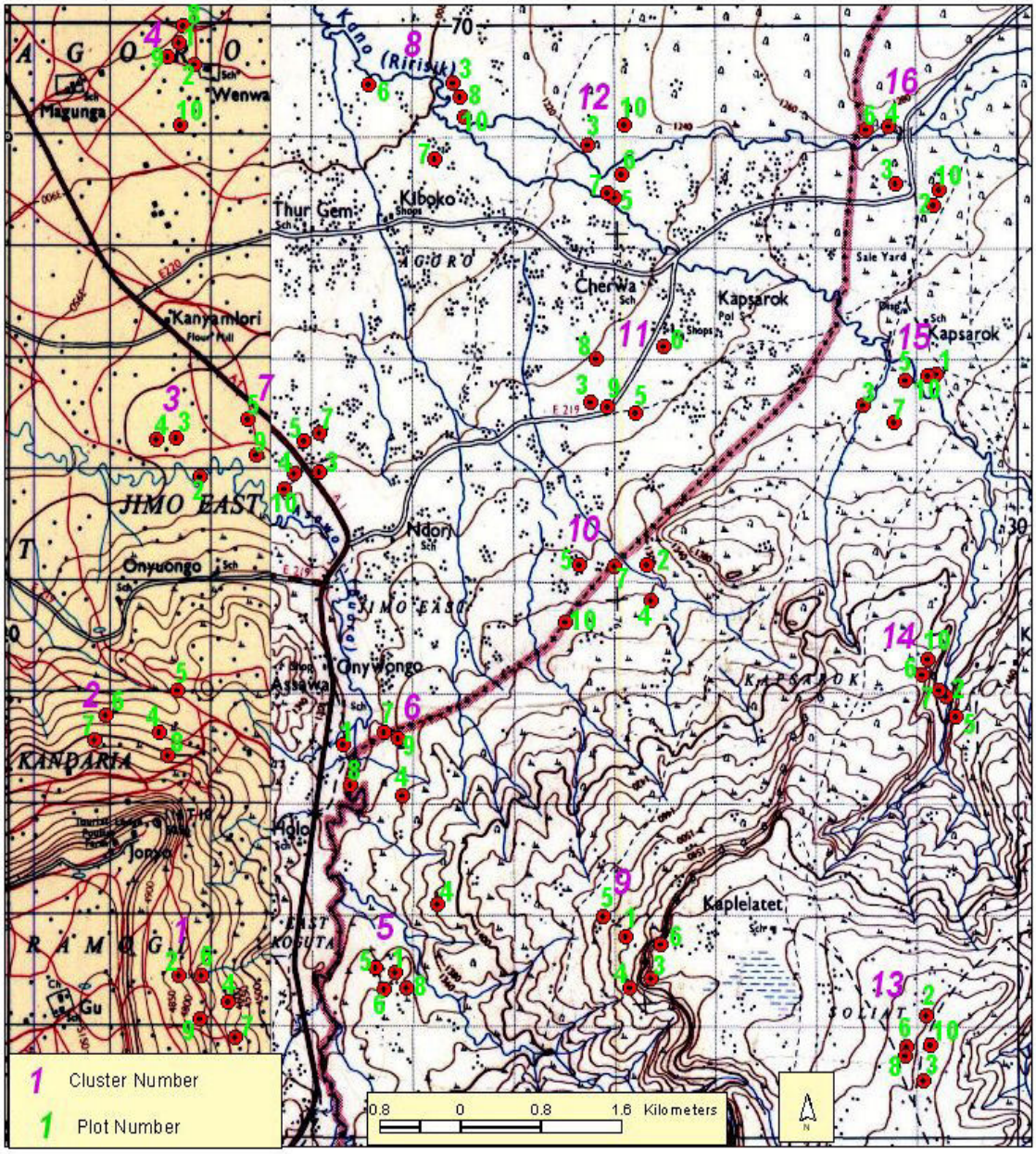


Figure 3.2: Distribution of clusters and sampling plots in Lower Nyando (LNY) sentinel site (plots marked in red were priority locations).

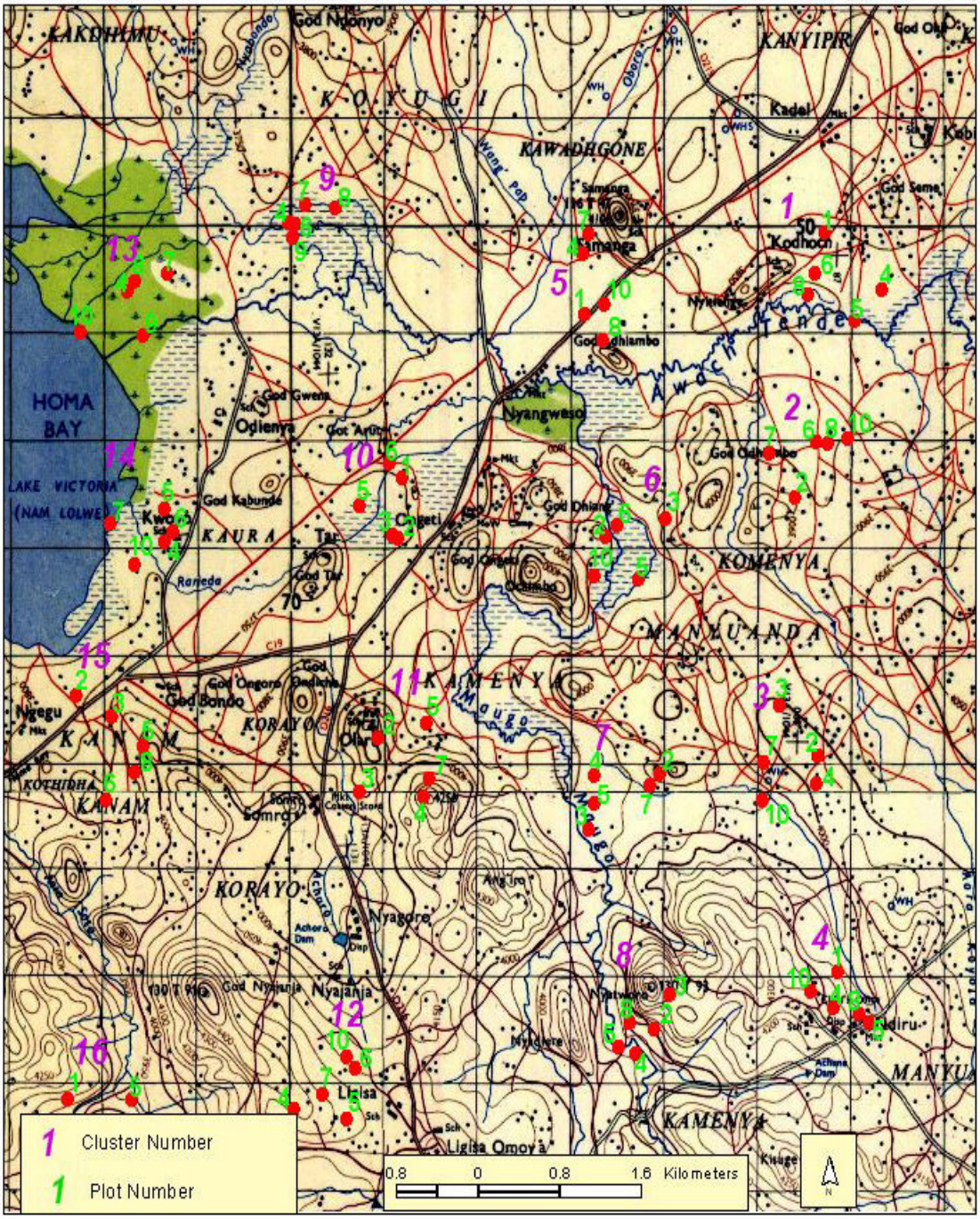


Figure 3.3: Distribution of clusters and sampling plots in Homa Bay (HB) sentinel site (plots marked in red were priority locations).

3.2 Development of soil spectra and property reference data

Spectral measurements (NIR and MIR) were conducted for calibration samples (n = 136) and for all samples from LNY and HB (n = 417). Reference data was generated for calibration set and for selected representative subsample from the larger set for validation. NIR and MIR spectra for calibration and validation set were also transformed into corresponding wavelet coefficients. Repeatability test was conducted for the two aggregate wet-sieving pretreatments.

3.2.1 Soil spectral data

MIR spectral data: Spectral measurements for calibration set (n = 135) and for the larger set from HB and LNY (n = 417) were conducted using a High Throughput Screening device (HTS–XT) attached to a TENSOR 27 spectrometer (Bruker Optics, Germany) customized for MIR spectral region (4,000 - 400 cm^{-1}) (Plate 3.1).

The measurement protocol elaborated by Terhoeven-Urselmans *et al*. (2010) was followed. About 5.0 g sub sample of the air-dry < 2 mm soil was ground to < 0.5 mm using natural stone pestle and mortar. The ground sample was thoroughly mixed and homogenized and approximately 0.5 g loaded into labeled wells in four replicates in aluminum micro-titre plates consisting of 96 wells. The first two wells of the plate were used for the standard and blank respectively and the first sample was placed in the third well. Scanning was done sequentially for each well. The average reflectance of 32 scans per sample was transformed to absorbance and recorded using the Optics Users Software (OPUS) (Bruker Optics, Germany).

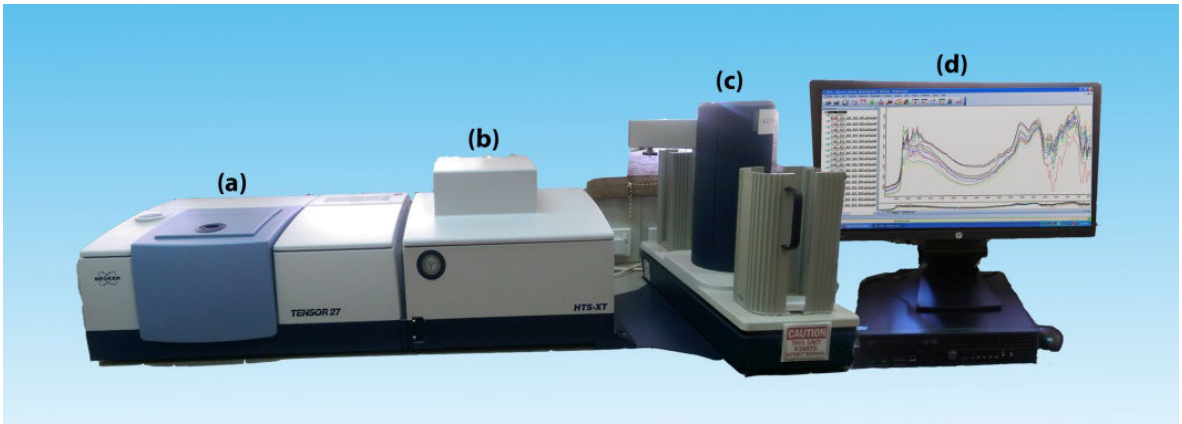


Plate 3.1: MIR diffuse reflectance measurement of soil samples ((a) Tensor 27 Fourier Transform Infrared spectrometer, (b) high throughput screening device (HTS – XT), (c) robotic arm for holding and conveying samples into HTS-XT and (d) measured soil absorbance spectra).

MIR spectral variables: MIR absorbance spectra were modified by first derivative transformation and smoothed using Savitzky - Golay filters (a nine point re-sampling algorithm) (Viscarra Rossel *et al.*, 2008). The spectral range 4,000 - 600 cm^{-1} (2500 - 25000 nm) were selected on the transformed spectra to eliminate wavebands with low signal to noise ratio. This provided a total of 1755 wavebands (data points), that were used as independent variables. Preprocessing was conducted in The Unscrambler version 9.02 (CAMO technologies, Inc, Woodbridge, NJ). The 1755 MIR wavebands were transformed also into 128 wavelet coefficients (MIRwc) using R software version 2.15.1 (R-Development Core Team, 2012).

NIR spectral data: Spectral measurements for calibration set (n = 136) and for the larger set from HB and LNY (n = 417) were conducted using a Fourier-transform diffuse reflectance spectrometer (Multi-Purpose Analyser (MPA), Bruker Optics, Germany) customised for the NIR region (12,500 - 3,600 cm^{-1}). Measurements for air-dried < 2

mm sub sample were made using resolution of 8 cm^{-1}, taking an average of 32 scans. Reflectance spectra were transformed to absorbance spectra and recorded using OPUS (Bruker Optics, Germany) (Shepherd, 2010).

NIR spectral variables: NIR absorbance spectra were transformed for first derivative and smoothed using Savitzky - Golay filter (a nine point smoothing function) (Shepherd & Walsh, 2002). The range (8,000 - 4,000 cm^{-1}) (700 - 2500 nm) was selected to eliminate noisy spectral bands (bands with low signal to noise ratio). This resulted in 1038 data points (NIR wavebands) that were used as predictors (independent variables). Preprocessing was conducted in The Unscrambler version 9.02. The 1038 NIR wavebands were transformed also into 128 wavelet coefficients (NIRwc) in R software.

Notably, the discrete wavelet transformation (DWT) and multi-resolution analysis (MRA) applies to a spectrum of discrete length 2^n, where n is an integer (Viscarra Rossel & Lark, 2009). Each NIR and MIR spectrum was padded to the same length of 2^n, (where $n = 7$), prior to DWT decomposition. For this reason the same number (128) of wavelet coefficients was extracted for NIR and MIR spectra despite the fact that they had different wavebands (data points).

Selection of validation sample set: Principal component analysis (PCA) of the MIR spectra for the combined (LNY and HB, $n = 417$) set was used to select representative 120 (~ 25 %) samples. The scores plot for the first two principal components (PC1 vs PC2) that together accounted for 68.1 % of the total variance was used to establish the distribution pattern of the soils. An average of 30 representative samples was selected from each of the 4 quadrants based on the Euclidean distance of the PC space (Naes *et al.*, 2002). Samples with extreme scores in different quadrants were selected also to broaden the spectral diversity and corresponding reference validation data range. The selected samples were used to provide reference and spectral values for validation.

3.2.2 Soil property reference data

Soil physicochemical properties: Reference data was developed for calibration (*n* = 136) (Set 1) and validation (*n* = 120) (Set 2) samples sets. Soil physical and chemical analyses were conducted using standard laboratory methods as reported by Shepherd and Walsh (2002) and Hinga *et al*. (1980). Air-dried soil moisture content (mc %) was determined by gravimetric method. Particle-size distribution was by hydrometer method. One sub sample set was with aided dispersion for total textural separates (tClay, tSi, tSa) (%) and other without for water-dispersed separates (WDC, WDSi, WDSa) (%). Flocculation index (FI, %) was computed as [100× (tClay - WDC) × $tClay^{-1}$)] ; clay dispersion ratio (CDR, %) as [100 × (WDC×$tClay^{-1}$)], and; soil dispersion ratio (DR, %) as [100 × (WDC + WDSi) × (tClay + tSi)$^{-1}$]. Median geometric particle diameter (dg, μm) and its standard deviation (δg, μm) were computed from total textural data (Shirazi & Boersma, 1984). Soil pH water (pHw, unit) and electrical conductivity (EC, dS/m) were read in 1:2.5 and 1:5 soil-water extracts using respective electrode meters. Na^+ in1:2.5 and 1:5 soil-water extracts (Na2.5, Na5 (mg kg^{-1}) were read on Na-ion electrode meter (Irvine & Reid, 2001). The pH-KCL (unit) was read on a 1:2.5 soil-to- 1.0 N KCL solution extract. Soluble Ca, Mg, K and Na were read in the 1:5 soil-water extract: Ca (sCa) and Mg (sMg) (mmol (+) L^{-1}) by atomic absorption using spectrophotometer (AAS) and Na (sNa) and K (sK) (mmol (+) L^{-1}) by atomic emission on flame photometer (AES). Sodium adsorption ratio (SAR, unit) was computed as [sNa× (sCa + $sMg)^2)^{-1}$] . Total C (totC, %) and total N (totN, %) were by dry combustion (Madari *et al*., 2006). One sub sample set was acidified for organic C (OC, %). The C: N was calculated as [totC × $totN^{-1}$]. Inorganic carbon (inC, %) was the numeric difference of totC and OC. Exchangeable Ca (eCa) and Mg (eMg) (cmol (+) kg^{-1}) were measured by AAS and Na (eNa, cmol (+) kg^{-1}) by AES in a 1:10 soil - to- 1.0 N KCL solution extract. Exchangeable K (eK), extractable P and micro-elements (Fe^{2+}, Mn^{2+}, Cu^{2+} and Zn^{2+}) were read in 1:10 soil-to- modified Olsen extracting solution extract. The eK (cmol (+) kg^{-1}) from flame photometer, P (mg kg^{-1}) from spectrophotometer using P color reagent

and the micro-elements (mg kg^{-1}) using AAS. Extractable boron (B, %) was by hot water extraction. Soil gypsum requirement (GR, cmol (+) kg^{-1}) followed US Salinity Laboratory Staff (1954) method 22d. Effective CEC (CEC1, cmol (+) kg^{-1}) was the numeric sum of exchangeable bases (eCa, eMg, eNa, and eK); clay activity (CEC2, cmol (+) kg^{-1}) as (CEC1× $tClay^{-1}$); and CEC3 (cmol (+) kg^{-1}) from CEC2 after accounting for contribution of SOC to CEC. The percent eNa (ESP), eK (ePP), eCa (eCaP), and eMg (eMgP) (%) were computed as the ratio of the respective bases to CEC1. The eNa ratio (eNaR, unit) as [eNa × (eCa + eMg)$^{-1}$)] *(*US Salinity Laboratory Staff, 1954*)*. Electrochemical stability indices (ESI) were computed as: (EC1:5×eNa^{-1}) (ESI 1, unit) and (EC1:5×ESP^{-1}) (ESI 2, unit) *(*Hulugalle & Finlay, 2003*)*. Determination of Atterberg limits (liquid limit (LL) and plastic limit (PL) and linear shrinkage (LS) (%) followed British Standards Institution (BSI: 1377, 1975). Plasticity index (PI %) was computed as the numerical difference between LL and PL. The activity number (*A*, unit) was obtained as the ratio of PI to tClay (Fratta et al., 2007). The coefficient of linear extensibility (COLE, unit) and soil volumetric shrinkage (VS, %) were computed using LS (%) data (COLE = (Lm-Ld) /Ld]; where: Lm = length of moist soil in the brass trough (= 140 mm), Ld = length of dry soil (140 mm – LS mm) and VS = [(COLE + 1)^3 - 1] ×100) as reported by Igwe (2003).

Data quality control: variable values suspected to be due to operator error (measurement or transcription) and/or samples with inadequate sample material for a property were marked 'missing'. Zero measurement value for a sample was replaced with a value obtained by multiplying the minimum (non-zero) value for the variable by 0.1 (Rule of Thumb). Variable values for samples that were non-plastic (recorded as 'NP') for Atterberg limits and LS determination were replaced with zero (0). The zero value was entered also for all other related variables (PL, PI, A, COLE, VS) for the particular sample. Established (total 59) soil physico-chemical and mechanical properties were screened for prediction of WSA indices.

Wet stable aggregation (WSA) indices: A simplified and rapid version of wet sieving and aggregate fractionation (Kemper & Rosenau, 1986) was used. One set of soil subsamples (50 g of < 2 mm soil weighed in triplicate) was subjected to slaking (rapid wetting) and another set to slaking plus mechanical shaking during 1 h. Manual sieving under tap water was used for stable macro aggregates (2000 - 212 μm) and the leachate secured in a plastic basin. The leachate was similarly sieved and fractionated for stable micro aggregates (212 - 20 μm). In order to correct for the sand (particles > 53 μm) present in the two aggregate fractions, a 50 mg L^{-1} Na hexametaphosphate-dispersed sample was sieved under the same operating conditions. The average sand free stable macro aggregates were computed and expressed in percentage [% macro = 100 × (wt of uncorrected macro - wt of sand in macro) × (wt of soil sample - wt of total sand)$^{-1}$], and stable micro aggregates [% micro = 100 × (wt of uncorrected micro - wt of sand in micro) × (wt of soil sample - wt of total sand)$^{-1}$], where wt of total sand = sand in macro fraction + sand in micro fraction (Canasveras *et al*., 2010; Madari *et al*., 2006). The unstable aggregate (fraction < 20 μm) (%) was obtained by subtraction [100 - (stable macro + stable micro)]. For each wet sieving pretreatment four (4) WSA indices were defined and assigned codes, respectively for macro, micro, unstable and macro: micro ratio as follows: slaking only (masp, misp, unsp and sponR); slaking plus mechanical disruption (mame, mime, unme and mechR).

Repeatability tests were conducted for Na-ion meter readings, for slaking only and for slaking plus mechanical disruption wet-sieving pretreatments. Repeatability was assessed using the coefficient of variability (CV), expressed in percentage [% CV = (SD/mean) ×100], where SD is the standard deviation of the measurements.

3.3 Exploratory data analyses (EDA)

Principal component analysis (PCA) and summary statistics were conducted for NIR and MIR spectra and property reference values for calibration (Set 1) and validation (Set 2) sets. The suite of (59) soil properties was screened for prediction of the WSA indices using CART regression tree analyses.

3.3.1 Soil spectral data

Plots for raw NIR and MIR absorbance spectra were used to illustrate characteristic spectral patterns and absorption features. PCA of spectra was used to assess association of calibration and validation soils and complexity of the data structure. The plots were made using R software.

3.3.2 Soil reference property data

Data distribution: One-way statistics (min, max, mean, SD, and percentiles/quartiles) was used to summarize soil property data for the suite of (59) properties and for wet stable aggregation (WSA) indices. Density plots and numerical skewness were used to determine appropriate linearization transformation (none, SQRT or log). Descriptive statistics and transformation were done using The Unscrambler software.

Data structure: PCA for combined soil properties and WSA indices was used to establish data structure and complexity. Prior to running, PCA data for each variable was mean centred then standardized by scaling (1/ SD). A full model with 20 principal components (PCs) and full cross validation for model testing was also specified. PCA was done using The Unscrambler software.

Association of soil properties: Pair-wise correlation coefficient was used to assess association of the (59) soil properties and the (8) WSA indices. Pearsons' correlation coefficient was used to establish association of soil properties with individual WSA

indices. Correlations were developed in R. Correlation was considered strong for $r \geq 0.71$; and weak for $r \leq 0.51$ (Mead *et al*., 2002).

3.3.3 Screening soil properties for prediction of WSA indices

The suite of 59 soil properties was screened for prediction of the eight (8) wet stable aggregation (WSA) indices. Prior to development of CART regression decision tree model**s**, soil property (predictor) and WSA indices (target) data matrix was preprocessed in Microsoft MS Excel (Steinberg & Golovnya, 2006). Variable code name with a mixture of character and numeric codes (for instance pH2.5, Na5) were revised to have complete character names (pHtwfv, Nafv). Missing variable data (N/A) was replaced with a blank; no data cell was allowed to have input of a character value, and; commas were not part of the data values. In CART one WSA indices was sequentially selected as the target and the (59) soil properties retained as potential predictors. Other model specifications were: 10-fold cross validation for model testing; least- square (LS) splitting rule method, and; the best tree (optimal decision tree) was set at one standard error (1SE) from the (maximal) minimum cost tree. Regression tree analysis was conducted using CART software version 6.0 (Salford Systems Inc., 2008). Output for each target included: the primary splitter, the (residual) relative error (RE), the primary splitter split point, and the five top primary split competitor and surrogate variables (Steinberg & Golovnya, 2006). The primary splitter was designated soil predictor for the target. Good decision tree model was indicated by low RE. Where predictions were unacceptable (high RE), CART returned a verdict that no decision tree could be grown, in which case suggested predictor variables that were qualitatively associated with the particular response were used to evaluate performance. The screened soil properties were used for development of partial least-square (PLS) models against predicted WSA indices. The principle and rationale of CART regression tree analysis is summarized in Appendix 1.

3.4 Calibration of WSA indices on IR- and soil-based predictors

Calibration set[5] comprising alternative IR-based predictors (NIR, NIRwc, MIR and MIRwc) and soil-based predictors from CART screening was used for developing calibration models for WSA indices (target). Prior to calibration development transformed reference data for WSA indices (target) and predictor soil property was mean-centred and then scaled (1/SD). Preliminary tests were conducted to optimize NIRwc and MIRwc input data for PLS modeling in The Unscrambler.

3.4.1 Calibration of WSA indices on IR-based predictors

The preprocessed WSA indices data was calibrated against each of the IR-based predictor variables: NIR (1038), MIR (1755), NIRwc (128), and MIRwc (128). Calibrations were developed using partial least-square regression (PLS 1) using The Unscrambler software. Calibrations were evaluated using leave-out-one cross-validation (looCV). In The Unscrambler, NIRwc and MIRwc were used as non-spectral data **(**based on optimization test runs**)**. A full model with a maximum 20 principle components (PCs) was set; however, the optimal number of PCs to be used for each property was determined using residual variances. The number of PLS factors (PLS PCs) chosen in each case was that resulting in the most parsimonious model with the lowest RMSECV (Stenberg, 2010). A plot of predicted vs measured values was used to compute the coefficient of determination (R^2). The R^2 and RMSECV and ratio of prediction deviation (RPD) (ratio of SD of the reference values to RMSECV) were used to evaluate the predictive ability of the models. The RMSECV was computed as follows:

$$\text{RMSECV} = [\sum(y - x)^2 / (n - 1)]^{1/2} \qquad \text{(Equation 1)}$$

[5] *The initial samples for calibration (n = 136) were sorted to correspond/ match with those available for laser diffraction particle size distribution (LDPSD) analyses (n = 128). Models based on LDPSD spectroscopy could, however, not be included in this report.*

Where; y is the predicted value by NIR, MIR, NIRwc and MIRwc and PLS technique, x is the observed or measured value, and *n* is the total number of samples used.

Optimization of wavelet input data for PLS modeling: Prior to use of NIRwc and MIRwc data two sequential tests were conducted to optimize wavelet coefficients (wc) PLS model performance in The Unscrambler. In Test 1 four different pre-processing procedures for the wc were tested for prediction of WSA indices: (i) the wc were entered as spectral data without standardization, (ii) the wc were entered as spectral data with standardization, (iii) the wc were entered as non-spectra data without standardization, and (iv) the wc were entered as non-spectra with standardization. The R^2 and RMSECV were used to assess the models. In Test 2, the wc were entered based on the optimal preprocessing established in test 1. PLS regression model computation time for selected diverse soil properties in calibration set (pH2.5, totC, eNa, tSi, Cu, tClay, and CEC1) against each of the IR-based methods (NIR, MIR, NIRwc and MIRwc) was recorded by noting the model start- and end- time in s. Identified influential outliers (samples with high residuals and high leverage) were removed and the model recalculated. Model computation time reduction factor by wc for both NIR and MIR was calculated as follows: Reduction factor by NIRwc = (NIR model time/ NIRwc model time) and by MIRwc = (MIR model time/ MIRwc model time). Input data reduction rate with wc was calculated as follows: Data reduction using NIRwc % = (NIR data points - NIRwc data points/ NIR data points) × 100 %, and using MIRwc % = (MIR data points - MIRwc data points/ MIR data points) × 100 %. Other workers (Viscarra Rossel & Lark, 2009) used similar approach to assess model performance with spectra wavelet transform.

Calibration optimization: The PLS looCV predictions were optimized following four different trials as follows: (i) prediction without removal of outliers, (ii) prediction with removal of identified spectral outliers, (iii) prediction with removal of reference values

outliers, and (iv) prediction with removal of both spectral and reference value outliers. Spectral outliers were checked for NIR, MIR, NIRwc and MIRwc using Mahalanobis distance *(H)* (Genot *et al*., 2011) implemented in R. Reference values outliers were defined as samples whose difference between predicted and measured value for a property was greater than 3× RMSECV (Sorensen & Dalsgaard, 2005).

3.4.2 Calibration of WSA indices on soil-based predictors

The established soil predictors were used for development of PLS looCV models against predicted WSA indices (target) in The Unscrambler. Prior to calibration, data for target and predictors were mean centred and standardized by scaling. In The Unscrambler the predictor and target variables were entered as non-spectra data. The R^2 and RMSECV and RPD were used to evaluate the predictive ability of the models. The RMSECV was computed using Equation (1), where; y is the predicted WSA indices value by the soil-based predictors and PLS technique; x is the observed or measured WSA indices value. The PLS looCV predictions were optimized following available two different trials as follows: (i) prediction without removal of outliers and (ii) prediction with removal of reference values outliers. Performance of the best IR- based method was compared with corresponding soil-based model using model fit (R^2) prediction error (RMSECV) and model reliability (RPD).

3.5 Further independent validation of IR-based WSA indices models

WSA indices and their CART selected soil-based predictors (WSA indices) in the calibration set (Set 1)[6] were used for developing PLS 1 looCV models against alternative IR-based methods. The looCV predicted values were further tested against

[6]*The initial samples for calibration (n = 136) were sorted to correspond/ match with those available for laser diffraction particle size distribution (LDPSD) (n = 128). Models based on LDPSD spectroscopy could, however, not be included in this report.*

corresponding values generated from independent validation set (Set 2)[7]. Calibration and validation models were developed in The Unscrambler version 9.02. Four different trials were conducted to assess effect of outliers on calibration validation as follows: (i) prediction without removal of outliers in the validation set, (ii) prediction with removal of reference values outliers in validation set, (iii) prediction with removal of spectral outliers in the validation set, and (iv) prediction with removal of both reference values and spectra outliers. For optimization criterion, trial (i) was compared with trials (ii), (iii) and (iv). A Rule of Thumb change in model R^2 during optimization trials was scored to assess the effect of removal of outliers in trials (ii), (iii) and (iv). Three categories as 'no effect', 'significant effect', and 'profound effect' were used to summarize the effects. The trials were then summarized with indication of maximal (trial with highest R^2) and optimal (trial with highest R^2 relative to trial (i)) model for each WSA indices and for each IR-method. The optimal model with the highest R^2 across the IR-methods was selected as the most appropriate validation and IR-method for each WSA indices.

3.5.1 Calibration validation without removal of outliers

Calibration of WSA indices on IR-based predictors: Preprocessed WSA indices data (target) was calibrated against preprocessed IR-based predictors (NIR, MIR, NIRwc and MIRwc) from the calibration set (n = 128) (Set 1). Calibrations were developed using PLS 1 in The Unscrambler and assessed using looCV. The R^2 and RMSECV and the RPD were used to evaluate the predictive ability of the models. The RMSECV was computed using Equation *(5)*. Influential outliers that met the set (Rule of Thumb) criterion (R^2 change with removal ≥ 10.0 %), were removed and the model recalculated.

[7]*The initial samples for independent validation (n = 120) were sorted to correspond/ match with those available for LDPSD (n = 79). Models based on LDPSD spectroscopy could, however, not be included in this report.*

The WSA indices looCV models with R^2 > 0.3 (Saeys *et al*., 2005) were further validated using independent sample set (n = 79) (Set 2).

Validation without removal of outliers (optimization trial (i)): Prediction of WSA indices were made using the looCV models and corresponding spectral-based data for the validation set (Set 2). Predictions were made in The Unscrambler. A plot of predicted vs measured/ observed values in Set 2 was used to compute the R^2. A check was made to confirm that both predicted and measured data were on the same measurement scale. Where the scales were different, the measured values were transformed to conform with the scale of predicted. The predictive ability of the looCV models was evaluated by the R^2, the root mean square error of prediction (RMSEP) and the RPD (ratio of SD to RMSEP for validation sample set). The RMSEP was computed using Equation (1), where; y is the predicted WSA indices value by the looCV predictors and PLS technique; x is the observed or measured WSA indices value in the validation set.

Prediction performance was compared for WSA indices from the same aggregate fraction but different wet sieving pretreatment (mame vs masp, mime vs misp, unme vs unsp, and mechR vs sponR).

3.5.2 Assessing effect of outliers on calibration validation

Reference values outliers (optimization trial (ii)): Reference values outliers were defined as samples in the validation set with data range (min-max) for particular indices that were out-of-range of the calibration set. Prediction of WSA indices were first made without removal of outliers. Identified outlier samples for particular indices were then removed and validation model (R^2, RMSEP and RPD) recalculated. The change (%) in validation R^2 and RMSEP after removal of outliers relative to the values before removal

was used to assess the effect of removal of the outlier(s) on validation. The change in R^2 and RMSEP respectively was calculated as follows:

R^2 (%) change = [100*(R^2 after- R^2 before) × (R^2 before)$^{-1}$] (Equation 2)

RMSEP (%) change = [100*(RMSEP after - RMSEP before) × (RMSEP before)$^{-1}$] (Equation 3)

Spectral outliers: optimization trial (iii): Spectral outliers in validation set were identified using Robust Mahalanobis distance measures *(H)* and implemented in R. A sample was defined outlier when $H > 2$. Outliers were identified for each of the IR-based predictor sets (NIR, NIRwc, MIR, and MIRwc). Prediction for WSA indices was first made without removal of outliers. Identified outlier sample(s) were then deleted and the prediction statistics (R^2, RMSEP and RPD) recalculated. The change (%) in R^2 and RMSEP with trial (iii) relative to trial (i) was used to assess the effect of removal of the outlier(s) using Equations (2) and (3), respectively.

Reference values and spectral outliers: optimization trial (iv): Prediction of WSA indices in the validation set was first made without removal of outliers. Identified reference values and spectral outlier samples for the particular indices were then removed and validation model (R^2, RMSEP and RPD) recalculated. The change (%) in R^2 and RMSEP with trial (iv) relative to trial (i) was used to assess the effect of removal of the outliers using Equations (2) and (3), respectively.

The change in model R^2 (%) after removal of reference values, spectral and both outlier types in the validation set relative to validation without removal of the outliers was summarized into three broad categories: (i) ± 0 – 10, as "no effect"; (ii) ± 10 – 50, as

"significant (positive/ negative) effect, and; (iii) ± > 50, as "profound (positive/ negative) effect".

Best validation and corresponding IR- method: Maximal validation for each WSA indices and for each IR- based method was the trial that yielded the highest R^2 comparing optimization trials (ii), (iii), and (iv) with trial (i). Optimal validation was the maximal validation that additionally presented $R^2 \geq 10$ % (Rule of Thumb) relative to trial (i), otherwise trial (i) was considered optimal validation. For each WSA indices the optimal validation with the highest R^2 across the IR-methods was selected the best (most appropriate) validation and IR-method.

3.6 Diagnostic screening and prevalence assessment of stability related problems using IR-based models in two sentinel sites

Calibration (Set 1, n = 128) and validation (Set 2, n = 79) sample sets were used to provide separate MIR PLS looCV models for the suite of 59 soil properties. Soil properties with modest to strong correlation to MIR ($R^2 \geq 0.6$) from each set were selected and screened for prediction of WSA indices using CART regression. MIR PLS looCV models for the screened properties from Set 1 were independently validated using Set 2 and key three properties selected. The more 'robust' of Set 1 and Set 2 based on CART regression screening was used for developing MIR PLS looCV models for each of the three selected properties. The models were used for prediction of the three properties in the larger set of samples from two sentinel sites of LNY and HB (n = 339) using their MIR spectra. The robust set was used also for developing grove files (calibration models in CART) for WSA indices with the selected three soil properties as the only predictors using CART regression. The predicted values of the properties in the larger samples set were used for predicting (scoring) each of the WSA indices in the set using the respective grove files. The predicted indices values were categorized into low or high stability based on observed predictor - response relationship and contextual

knowledge. A count (tally) of low and high stability categories for each sample across the WSA indices was taken for all samples of the larger set. A sample was designated as low stability when the tally for low stability category was higher than the tally for high stability. A sample was designated high stability when the tally for high stability was higher than for low stability. Prevalence of stability problems (low stability) for the two sentinel sites was calculated as the total 'low stability' divided by the total samples and expressed as percent. Similarly prevalence of stability problems in different clusters and sampled depth intervals of the sentinel sites was calculated as the total 'low stability' in a cluster/ depth interval divided by the total samples in that cluster/ depth interval. Analogous approaches for development of pedotransfer functions and inference systems for more effective estimation of functional attributes from spectra and basic soil properties has been demonstrated by other workers (Tranter et al., 2008; McBratney *et al.*, 2006).

3.6.1 Screening soil properties for prediction of WSA indices

Separate MIR PLS models for the suite of 59 soil properties were developed for calibration (Set 1, n = 128) and validation (Set 2, n = 79) sample sets. The models were developed in The Unscrambler and evaluated using looCV. Soil properties in Set 1 and Set 2 that were strongly correlated to MIR ($R^2 \geq 0.6$) were separately screened to establish key few predictors of WSA indices. Screening was conducted using CART Pro V6.0 software. In CART each of the WSA indices was entered as the target and the set of selected properties as predictors. Ten (10) fold cross validation was confirmed model testing method, the best CART regression tree was set at 1SE from the minimum cost tree and least square (LS) splitting rule as regression method. Output for each WSA indices included: the primary splitter(s), relative error (RE), split point for each splitter, and the variable importance score for the primary splitter(s). Complex decision tree models (more than two primary splitters) were rationalized by elimination of the predictor(s) with lower contribution. Simpler models were preferable (Steinberg &

Golovnya, 2006), however, an objective trade-off between model complexity and compromise on RE was observed. The screening provided information also on the more 'robust' of the two samples sets (Set 1 and Set 2).

Predictive relationship of predictor soil properties and predicted WSA indices was illustrated using scatter plot plots. Raw data for predictor variable(s) were plotted against raw data for corresponding predicted indices. Indication of split points on the plots was used to illustrate the efficacy of the predictive relationships.

MIR PLS looCV models for CART selected soil properties from Set1 were tested using MIR spectra and reference data in Set 2. The more strongly validated soil properties were selected and key three properties further selected. The selected (3) properties were used (as predictors) for development of PLS models for prediction of wet stable aggregation (WSA) indices in the larger set from lower Nyando (LNY) and Homa Bay (HB) sentinel sites.

3.6.2 Prediction of WSA indices in LNY and HB sentinel sites

Prediction of key three soil properties in LNY and HB: The more robust of the two sets (Set 1 and Set 2) from CART screening was used to develop MIR PLS looCV models of the selected key three soil properties. The models were used to predict the three properties in the larger sample set (n = 339) from LNY and HB using the MIR spectra of the larger set. Calibration and prediction was conducted in The Unscrambler. For prediction, the MIR spectra were entered as predictors and the looCV model for each of the properties as the response. Where the predicted values were on transformed scale the values were back-transformed to actual values. The predicted values were sorted in a data matrix that included sentinel site (LNY or HB), cluster ID (1-16), sampling plot number (between 1 and 10), and sampled depth interval (0-20, 20-50, 50-

100 cm). The data matrix provided the Data File used in CART to predict (score) WSA indices using the respective grove files.

Developing grove files for WSA indices: The more robust of Set 1 and Set 2 was used also for developing grove files (calibration models in CART) for each WSA indices. Grove files were developed using CART Pro V6.0. In CART, the selected three soil properties were marked as the predictors and one WSA indices as target. Other model set-up specifications included 10-fold cross validation for model testing, optimal tree at one standard error (1SE) from the minimum cost tree, and LS splitting rule. For each model run the RE, terminal nodes, splitters and split points and splitter importance score were recorded. The grove files were used for scoring (predicting) WSA indices in samples from the larger set (LNY and HB, $n = 339$).

Scoring WSA indices in two sentinel sites: Grove files for WSA indices and the Data File of predicted three soil properties in the larger sample set were used for scoring (predicting) each of the WSA indices in samples of the larger set. CART Pro V6.0 was used for scoring. In CART, the grove file was used to split (predict mean value for terminal node) for particular WSA indices for each sample in a procedure known as "dropping predictor data down a tree model" or "scoring data" (Steinberg & Golovnya, 2006).

Dropping data down CART tree model: during prediction, each sample was processed case by case (case = node content) beginning at the root node. The splitting criterion specific for each WSA indices was applied, and in response to each yes/ no question, the case moved left or right down the tree until it reached a terminal node. The main predictor variable for the particular WSA indices in the grove file was used for scoring the data. If the primary criteria could not be used because the case was missing data, a surrogate split criterion was used. The number of predicted (mean) values for the indices

was the same as the number of terminal nodes of the corresponding grove file. After all samples in the larger set (cases) were dropped down the particular indices tree model, the output (predicted score value (s)) was saved. The procedure was repeated for each WSA indices grove file and the predictor Data File for all cases and saved in separate files. The score values were used for prevalence assessment.

3.6.3 Assessing prevalence of stability related problems in LNY and HB sentinel sites

Prevalence of stability categories for individual WSA indices: The scored (predicted) value(s) for each of the WSA indices for each sample in the larger set were categorized into low or high stability. Stability category assignment was dependent on observed (scatter plot) relationship between predictor variable(s) and target indices, the number of score values, and contextual judgment. For WSA indices with two score values, for example, one of the values was assigned to low stability and the other to high stability category. For three splits the intermediate score value was considered transitional and assigned to moderate category. For more than three splits a rationalization was used to incorporate the intermediate splits into the main low and high categories (for instance, 'slightly low, and slightly high). For each WSA indices, a total count of each stability category for all sentinel samples was made. Prevalence of the category was calculated and expressed in percentage as follows:

Prevalence of stability category (%) = [100 × (count of category) × (total sentinel samples)$^{-1}$] (Equation 4)

Prevalence of stability problems in LNY and HB: The two main categories of stability (low and high) established for each sentinel sample and for each WSA indices were used to assess prevalence of stability problems (low stability) within LNY and HB. The categories were assigned to codes one (1) for low and zero (0) for high stability. The

total count of low and high category for each sample across WSA indices was taken. Stability sub-categories 'slightly low' and 'slightly high' were counted with the corresponding main category. Stability sub-category moderate was, however, uniquely categorized and assigned code (mod). A column each for the total count of low, moderate and high stability was used. Where the tally for low stability for a sample was higher than for high stability the sample was designated low stability. Where the total count for high stability was higher than low stability the sample was designated high stability. Where the tally for low plus moderate stability was higher than that for high stability the sample was assigned to low stability. Likewise where the tally for high plus moderate stability was higher than that for low stability the sample was assigned to high stability. Where low and high stability tally were equal, the sample was assigned to moderate stability. In this way all the sentinel samples were assigned to low, moderate or high stability. Prevalence of low stability in each of the two sentinel sites was computed as the number of samples of low stability divided by the total samples and expressed in percentage as follows:

$$\text{Prevalence of low stability (\%)} = [(\text{low stability count}) \times (\text{total samples})^{-1}] \times 100] \quad \text{(Equation 5)}$$

Prevalence of soils at risk of stability related problems was assessed as prevalence of low stability plus prevalence of moderate stability.

Prevalence of stability problems in different clusters: The assigned stability category for each sample in the two sentinel sites was used for prevalence assessment of low stability in different clusters of the sites. Clusters from LNY and HB were assessed separately. A count of low stability samples for each cluster was made and prevalence computed as follows:

Low stability prevalence in cluster (%) = [(total count of low stability in cluster) × (total samples in cluster)$^{-1}$] × 100] (Equation 6)

Cluster(s) that were scored as high stability for all or majority of WSA indices were considered stable, whereas clusters that were scored as low stability for all the indices were designated 'hotspot'.

Prevalence of low stability for different depth intervals: The data matrix with assigned stability categories for all sentinel samples and information on sentinel site name, cluster number, sampling plot number and, depth sampled was used for prevalence assessment of low stability for different depth intervals in LNY and HB. The samples were sorted on sampled depth interval as follows: 0 - 20 cm; 20 - 50 cm and 50 - 100 cm. A count of samples of low stability for each depth interval was made. Prevalence of low stability within a depth interval was computed as follows:

Low stability prevalence in depth interval (%) = [(total count of low stability in depth interval) × (total samples in the depth interval)$^{-1}$] × 100] (Equation 7)

In this way low stability prevalence for each sentinel site and for both sites were computed for the three depth intervals. Similarly prevalence of moderate and high stability within depth intervals was computed.

3.7 Assessing WSA indices most appropriate for screening stability problems using IR-based models in LVB of Kenya

The selected key soil-based predictors of WSA indices from calibration (Set 1, n = 128) and validation (Set 2, n = 79) sets (see section 3.6.1**)**, were sorted into four different subsets. Each subset comprised of three properties. The four subsets and MIR were

compared for prediction of WSA indices in Set 2 using PLS. CART regression was used also for screening each of the four soil-based sets for prediction of WSA indices. The two more robust sets were established from the PLS and CART regression. MIR PLS looCV models for each of the three properties in each of the selected two sets were developed and used for predicting these properties in the larger set (n = 339) from LNY and HB sentinel sites. The predicted values constituted the Data File, whereas the CART models for the two selected sets provided the grove files. The Data File and grove files from each of the two sets were used for scoring (predicting) WSA indices in samples of the larger set. The two sets were compared for their performance in assessing prevalence of stability related problems in the two sentinel sites. This involved assessing: (i) prevalence of stability categories for individual WSA indices, (ii) prevalence of low stability in the two sites, (iii) prevalence of low stability in clusters within each of the sites, and (iv) prevalence of stability problems for sampled depth intervals in the sites. The final determination of the most appropriate WSA indices for diagnostic screening stability related problems in LVB considered also: performance of the sets during CART regression screening; ease of data acquisition (laboratory determination) of the predictor properties; predictor property based on direct measurement was preferred over derived.

3.7.1 Selection of most appropriate soil predictor sets of WSA indices

The selected soil-based predictors of WSA indices in the two reference sample sets (Set 1, n = 128 and Set 2, n = 79) (see section 3.6.1) were sorted into four different subsets each comprised of three properties. The subsets sets were separately screened for prediction of WSA indices in Set 2 using CART regression. In CART each set of properties was marked as predictors and one WSA indices as the target. Model testing was confirmed as 10-fold cross validation, best decision tree set at 1SE from the minimum cost tree, and LS as splitting rule. Selected output was: primary splitter(s) and split points, RE, terminal nodes, and variable importance score for the primary splitter(s). The two predictor sets that had predictive relationship with most of the WSA

indices (considering also the RE) were proposed as most appropriate for predicting the indices.

Two sets of PLS looCV models were developed for WSA indices using Set 2. The first set of models calibrated WSA indices directly onto MIR. The second set calibrated WSA indices on each of the four soil-based predictor sets. The models were developed in The Unscrambler. Model R^2 and RMSECV were used to compare performance and affirm the two most appropriate soil-based sets for prediction of WSA indices. The CART models for the selected two sets were used as Grove files for scoring (prediction) of these indices in the larger set samples from LNY and HB sentinel sites.

3.7.2 Prediction (scoring) of WSA indices in LNY and HB sentinel sites

MIR PLS looCV models for the properties in the selected two soil-based sets were developed using Set 2. The models were used for prediction of the properties in the larger sample set (n = 337) from LNY and HB, using the MIR spectra of the larger set. Calibration and prediction models were developed in The Unscrambler. The Data File with predicted values of soil properties in the larger set together with the grove file for each WSA indices (see section 3.7.1) were used to score value(s) of the indices in samples from the larger set. The number of predicted values for particular indices was dependent on the number of terminal nodes of the corresponding grove file. The predicted values were used for assessing the prevalence of stability related problems in LNY and HB.

3.7.3 Assessing prevalence of stability related problems in LNY and HB sentinel sites

Prevalence of stability categories for individual WSA indices: The scored value(s) for each WSA indices and for each of the larger set samples was categorized into low, moderate or high stability. For each WSA indices a total count of each stability category for all sentinel samples set was made. Prevalence of the category was calculated and expressed in percentage using Equation (4). A comparison was made of prevalence of stability category for WSA indices from the same aggregate fraction but from the two different wet sieving pretreatments.

Prevalence of stability problems in LNY and HB: The total count of low, moderate (mod) and high stability categories for each sample across the WSA indices for LNY and HB was taken. Sub-categories 'slightly low' and 'slightly high' were counted with the corresponding main category. Where the tally for low was higher than for high the sample was designated low stability. Where the tally for high was higher than low, the sample was assigned to high stability. Where the tally for low plus moderate was higher than that for high, the sample was assigned to low stability. Likewise where the tally for high plus moderate stability was higher than that for low the sample was assigned to high stability. Where low and high stability tally were equal, the sample was assigned moderate stability. Prevalence of low stability in the two sentinel sites was computed using Equation (5). Prevalence of soils at risk of stability related problems was assessed as prevalence of low plus moderate stability.

Prevalence of stability problems in clusters within LNY and HB: The assigned stability category for samples in LNY and HB was used for prevalence assessment of low stability in the different clusters of the sites. Clusters from each site were assessed separately. A count of low stability samples for each cluster was made and prevalence calculated using Equation (6). Cluster(s) that were scored as high stability by all or

majority of WSA indices were considered stable. Clusters that were assigned low stability for all indices were designated 'hotspot'.

Prevalence of stability problems for different depth intervals: The data matrix with assigned stability categories for all sentinel samples and also information on depth sampled was used for prevalence assessment of low stability for different depth intervals (0 - 20, 20 - 50, and 50 - 100 cm). A count of samples of low stability for each depth interval was made. Prevalence of low stability for a depth interval was computed using Equation (7). Similarly prevalence of moderate and high stability classes was computed.

3.7.4 Determination of the most appropriate WSA indices for screening stability related problems in LNY and HB

Performance correspondence for prevalence assessment using models based on the two alternative sets of soil-based predictor properties was used to establish WSA indices and their thresh hold values most appropriate for diagnostic screening and prevalence assessment of stability related problems in LNY and HB. The final determination of the most appropriate indices considered also: performance of the soil-based predictor sets during CART regression screening; ease of data acquisition (laboratory determination) of the predictor soil properties; property based on direct measurement was preferred ahead of derived; residual value of the predictor property, including support in the literature of the property in soil stability related studies.

Performance correspondence of soil predictor sets: Performance of the models of selected two soil predictor sets in assessing stability related problems in LNY and HB was compared (for concurrence). This involved comparing: (i) prevalence of stability categories for individual WSA indices, (ii) prevalence of low stability in the two sites, (iii) prevalence of low stability in clusters within each of the sites, and (iv) prevalence of stability problems for sampled depth intervals in the sites.

Most appropriate WSA indices and thresholds: The most appropriate WSA indices for diagnostic screening stability related problems in LVB were determined by the correspondence of the two sets in the prevalence assessment. This included correspondence of prediction error (RE), the number of grove files terminal nodes (TN) and corresponding score (predicted) WSA indices values from CART analysis. The selection took into consideration also the ease of soil property data acquisition (laboratory determination), and potential residual value of the property.

Soil property based on direct measurement was preferred over derived. The WSA indices predicted by the two soil sets were designated most appropriate indices, and their score values used to define low/ high stability thresholds.

CHAPTER FOUR
RESULTS AND DISCUSSION

This chapter presented results for reference property data and soil spectra, performance of IR- and soil-based predictors for estimation of WSA indices, and on further testing of the IR-methods using dataset of independent samples. Presented is also observed prevalence of soil stability related problems in Lower Nyando (LNY) and Homa Bay (HB) sentinel sites, including the different clusters and sampled depth intervals. Finally the most appropriate WSA indices for screening soil stability problems using IR-based models in LVB of Kenya.

4.1 Soil properties and calibration of WSA indices on IR- and soil-based predictors

Results are presented for observed soil (NIR and MIR) absorbance spectra; spectra and soil basic property data structure (pattern and complexity); wet-sieving precision and WSA indices distribution; association of soil properties with WSA indices; the key soil-based predictors of WSA indices from CART screening; and; performance of spectra wavelet transform coefficients in PLS modeling. Performance of alternative IR-methods (NIR, MIR, NIRwc and MIRwc) for estimation of WSA indices is compared with performance using soil-based predictors.

4.1.1 Soil Properties

Soil spectra: Soil absorbance spectra pattern for calibration and validation samples sets were similar for NIR and for MIR, indication that the samples belonged to the same soil population (Figure 4.1 (a) and (b); spectra is after removal of the noisy parts (> 8,000 and < 4,000 cm^{-1} for NIR and < 600 cm^{-1} for MIR; spectra is shown in both wavenumbers (cm^{-1}) and wavelength (nm) for easy reference, where [(10,000,000 × $wavenumbers^{-1}$) = wavelength)]).

NIR absorbance spectra (8,000 - 4000 cm^{-1}) followed established basic shape (Canasveras *et al*., 2010), with subtle but characteristic peaks at about 1400, 1900 and 2200 nm, ascribed to moisture (vibrational absorbance due to the –OH in minerals, and to –OH and -CH, and NH organic functional groups in soil organic matter (Viscarra Rossel *et al*., 2006). The absolute magnitude and range of the MIR absorption (4000 - 600 cm^{-1}) are much greater than those found for the corresponding NIR spectra, and the MIR spectra indicate more distinctive spectral features than the NIR spectra (see also Canasveras *et al*., 2010). This suggested that MIR might be more resourceful than NIR. Few atypical spectra (wavebands between 2000 and 1000 cm^{-1} and at frequencies below 1100 cm^{-1} (Figure 4.1), suggested potential outliers.

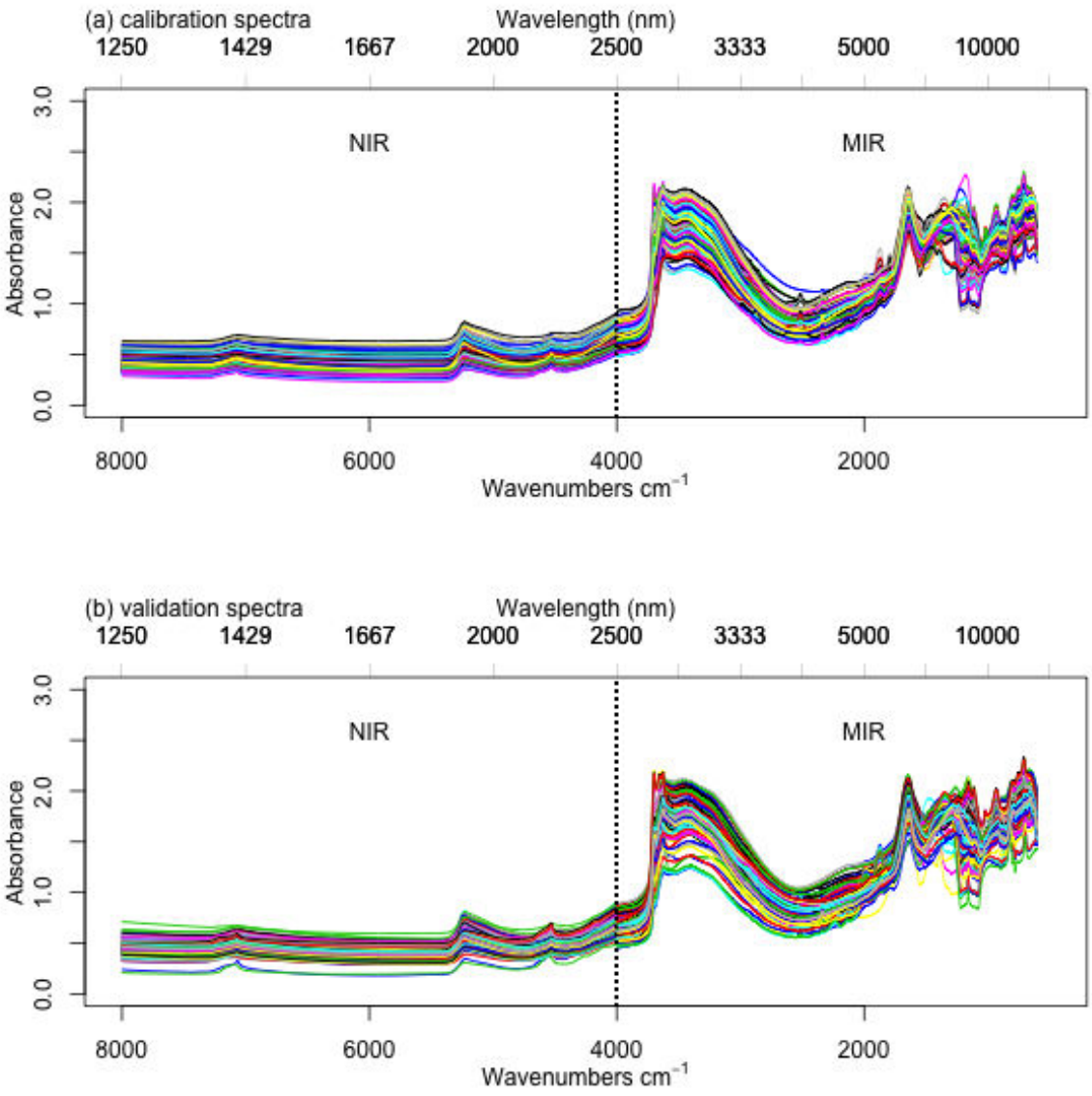


Figure 4.1: NIR and MIR soil absorbance spectra for (a) calibration (*n* = 136) and (b) validation (*n* = 120) samples sets.

The first principle component (PC1) explained 79 - 76 % and 51 - 45 %; and the first 5 PCs explained 98 and 88 % of the total variance based on NIR and MIR measurements, respectively (Figure 4.2). The first 5 PCs individually contributed 76.4, 8.7, 1.6, 1.0 and 0.6 % (NIR) and 44.7, 22.8, 9.7, 7.6 and 4.7 % (MIR) of the total explained variance for calibration set, suggesting spectra provides effective data structure. Notably fewer PCs were required to explain the limited extractable information from NIR spectra, demonstrated by the few NIR absorption features, whereas several absorption features from MIR (Figure 4.1), required more PCs to explain all the variance.

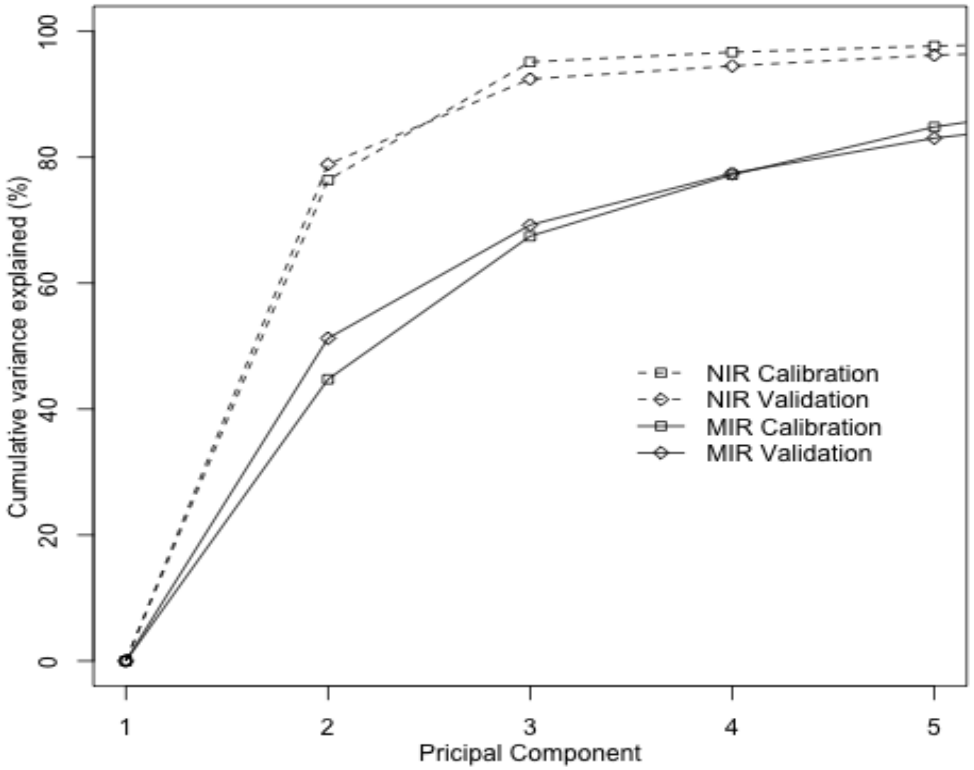


Figure 4.2: Cumulative variance explained by the principal components based on measurements with multipurpose analyzer (NIR range) and the Tensor 27 (MIR range).

Soil property reference data: Soil from three depth intervals (0 - 20, 20 - 50, 50 - 100 cm) were combined providing a wide data range for calibration (Set 1) and validation (Set 2) sets. The soils ranged, for example, from acid to alkaline (pHw 4.6 -10.5), very coarse to clayey (tSa 11 - 89 %; tClay 9 - 81 %) texture, low to high OM (OC 0.1 - 7.1 %), from non-sodic to severely sodic (eNa 0.0 - 64 cmol (+) kg^{-1}), and; from low to high dispersibility (WDC 3 - 69 %). Data for EC, C and N, soluble bases, exchangeable bases (eNa, eK) and derived variables (ESP, EPP, ESI 1, ESI 2, CEC2, and CEC3), micro elements (Fe, Mn, Zn) and P, particle-size based variables (tSa, tSi:tClay, WDSa, CDR, WDSi:WDC), and mechanical properties (PL, A) indicated high skewness in their distribution. Long tails of a majority of samples with low values and few samples with very high values characterized the skewed distributions. The natural logarithm (ln) transformation was applied to reduce skewness. For variables with slight skewness (sMg, eCa, eMg, CEC1, Cu, B, and VS), square root (SQRT) transformation was used. Data for pH, particle-size (tClay, tSi, WDC, WDSi, DR, FI), eMgP, mechanical properties (LL, PI, LS, COLE) and mc assumed a Gaussian distribution and no transformation was used. The suite of properties and data range for individual properties was typical, therefore, of soil property datasets commonly used for land capability/ suitability evaluation (Hazelton & Murphy, 2007; Peverill *et al*., 1999). Appendix 2 presents summary data of soil properties for the calibration and validation sets.

Soil property data structure: PCA for combined (59 soil properties and 8 WSA indices) dataset revealed that 16 PCs were needed to explain 83 % of the variance in calibration soils (Table 4.1). The first PC explained 26 %, the first four PCs explained 55 %, whereas the first 10 accounted for 75 % of the total variation. This is contrary to soil spectra where the first 5 PCs accounted for ≥ 88 % of the total variation. The last six PCs (PC 11 - PC16) contributed about 9 % of the variation. The PCs 9, 15, and 16, made higher contributions than preceding PCs, suggesting that the reference data was noisy

(complex). It is characteristic of complex soil data sets, however, that important variables may be lurking in otherwise lowly ranked PCs (CAMO ASA Inc., 1998).

Table 4.1: Variance explained by the principal components based on reference measurements of soil properties for calibration samples.

PC [a]	***% explained variance cumulative (calibration)***	***% explained variance cumulative (validation)***	***% explained variance each PC***
PC_01	28.6	25.7	25.7
PC_02	46.0	41.8	16.1
PC_03	54.4	48.3	6.5
PC_04	61.8	55.4	7.1
PC_05	66.9	58.3	2.9
PC_06	71.6	63.7	5.5
PC_07	75.5	67.9	4.2
PC_08	78.2	69.3	1.4
PC_09	80.8	72.8	3.5
PC_10	82.9	74.5	1.7
PC_11	84.7	75.7	1.2
PC_12	86.2	77.2	1.5
PC_13	87.6	78.0	0.8
PC_14	88.9	78.6	0.6
PC_15	90.2	81.2	2.6
PC_16	91.3	83.4	2.2
Total explained variance by first 16 PCs			83.4

Note: *PC[a], principle component;*

For representative data sets cumulative % variance explained at the validation stage closely follows corresponding values for the calibration (CAMO ASA Inc., 1998). The difference between explained calibration and validation variance increased from 3- 5 % for the three first PCs to an average of about 10 % for lower ranked PCs (Table 4.1), suggesting potential for less representative reference calibration data (CAMO ASA Inc., 1998).

The complexity in input data (contrasting spectra and wet chemistry data) has implications on performance of linear models. The reference properties indicated presence of non-linear structure with potential suboptimal performance for parametric analyses (Linker, 2012). Fewer PCs based on NIR and MIR measurements (Figure 4.2), suggested higher parsimony, therefore, more robust models, affirming Canasveras *et al*., (2010) who found superior usefulness of diffuse reflectance over basic soil properties for developing pedotransfer functions for estimation of soil aggregation indices.

Precision of aggregate wet-sieving pretreatments: Slaking plus mechanical shaking wet-sieving pretreatment was more precise than slaking only (% CV 28.7-2.8 vs 57.5-3.3). This was attributed to breakdown of unstable and metastable wet aggregates with shaking (Marquez *et al*., 2004) to the limiting stability analogous to percolation stability (Auerswald, 1995). Micro aggregates were more repeatable than macro aggregates, associated with difficult to (standardize) control manual wet sieving, especially for macro aggregate fraction, where sieving was assumed complete when tap water passing under the sieve screen was clear. There was a general increase in measurement precision with increasing eNa levels (Table 4.2).

Table 4.2: Repeatability (% CV) of slaking only and slaking plus mechanical shaking wet sieving pretreatments.

Sodicity unit	**Slaking only**			**Slaking + mechanical shaking**		
eNa	**masp**[a]	**misp**[a]	**unsp**[a]	**mame**[a]	**mime**[a]	**unme**[a]
0.25	25.2	10.3	4.3	13.8	3.9	0.9
0.5	57.5	4.8	6.5	28.7	7.2	2.1
14.3	17.2	4.6	1.7	6.7	3.4	0.8
19.3	18.4	5.1	2.5	10.1	3.7	0.7
45.3	6.8	3.3	0.1	7.2	2.8	0.1

Note: [a] *Wet aggregation indices: stable macro (masp/ mame); stable micro (misp/ mime); unstable (unsp/ unme); eNa, exchangeable sodium (cmol (+) kg^{-1});*

Lower precision (% CV up to 57) for low sodicity soils, especially for macro aggregate fraction (Table 4.2**)**, suggested unreliable reference data and potentially weak NIR PLS performance (Stenberg *et a*l., 2010). Low wet chemistry precision is contrary to reported high precision of soil spectral measurements (Shepherd *et al*., 2005).

Wet stable aggregation (WSA) indices: Three interval-depths (0-20, 20-50, 50-100 cm) were combined for both calibration and validation samples sets. This presented a wide range in WSA indices. The validation data were within the range of the calibration set for all indices (except for minimum values of mime and misp), requisite for successful calibration (Stenberg *et al*., 2010). PLS analysis is suboptimal for skewed data (Linker, 2012*)* and requisite linearization transformation was used (Table 4.3).

Table 4.3: Wet stable aggregation (WSA) indices for calibration samples set (the corresponding values for validation samples is shown in parenthesis).

Indices	min	25%	50%	75%	max	SD	transformation[a]
mame	0.1	2.7	4.7	9.1	63.9	10.2	
	(0.4)	(2.4)	(4.1)	(10.0)	(58.7)	(12.2)	ln (ln)
masp	0.8	6.0	9.5	17.4	68.5	12.4	
	(0.5)	(6.0)	(13.1)	(23.3)	(61.3)	(14.6)	ln (ln)
mime	9.6	22.2	28.6	34.2	48.9	9.1	
	(3.9)	(15.9)	(21.6)	(28.5)	(49.2)	(8.4)	ln (sqrt)
misp	11.8	30.61	38.8	46.6	66.9	11.1	
	(3.9)	(20.3)	(27.3)	(35.8)	(59.3)	(11.3)	none (sqrt)
unme	19.0	57.2	64.5	71.2	90.4	13.5	
	(18.6)	(61.0)	(72.4)	(80.0)	(95.7)	(16.5)	ln (none)
unsp	12.0	40.5	47.5	56.3	86.2	14.9	
	(13.5)	(40.9)	(54.7)	(68.4)	(100)	(19.6)	none (none)
mechR	0.01	0.10	0.17	0.31	4.15	0.51	
	(0.04)	(0.12)	(0.20)	(0.44)	(2.60)	(0.51)	ln (ln)
sponR	0.02	0.16	0.25	0.46	3.89	0.57	
	(0.03)	(0.23)	(0.41)	(0.85)	(2.77)	(0.63)	ln (ln)

Note: [a] *linearization transformation applied to WSA indices data: ln, logarithmic; sqrt, square root; none, no transformation; validation data for WSA indices are shown in parenthesis.*

Association of WSA indices: Very strong association of stable macro (mame and masp) and micro (mime and misp) aggregate indices and unstable fractions (unme and unsp) from slaking only and slaking plus mechanical disruption wet sieving pretreatments (pair-wise correlation coefficient r = 0.8), suggested potential for information redundancy among the two sets of indices (Table 4.4). Stable macro aggregates were uncorrelated with their stable micro aggregate counterparts (r < 0.2). Stable micro aggregates were moderately associated with the unstable aggregate fractions (r = 0.7 - 0.4). Notable was the very strong (positive) association of mechR (ratio mame: mime) with mame (r = 0.95), and of sponR (ratio masp: misp) with masp (r = 0.92), and; poor (negative) correlation of mechR with mime (r = 0.2) and sponR with misp (r = 0.4), respectively (Table 4.4). This suggested dominant influence of macro aggregates over micro aggregates for the studied soils.

Table 4.4: Correlation coefficients (*r-value*) (upper diagonal) of wet stable aggregation (WSA) indices for calibration sample set.

mame	masp	mime	misp	unme	unsp	mechR	sponR	
1.00	0.83	-0.02	-0.12	-0.74	-0.60	0.95	0.80	mame
	1.00	0.01	-0.19	-0.63	-0.69	0.78	0.92	masp
		1.00	0.77	-0.66	-0.58	-0.20	-0.18	mime
			1.00	-0.43	-0.58	-0.26	-0.38	misp
				1.00	0.84	-0.58	-0.48	unme
					1.00	-0.46	-0.49	unsp
						1.00	0.85	mechR
							1.00	sponR

Correlation of WSA indices with basic soil properties: Calibration on spectra of spectrally non-responsive soil attributes is through co-variation (auto-correlation) with soil chromophores (mc, texture, SOC, Fe-oxide) (Stenberg *et al*., 2010). Canasveras *et al*. (2010) ascribed attained prediction of water stable aggregates (fraction > 250 μm) (R^2 = 0.60 - 0.23) partly to correlation (Pearson's correlations coefficient, R = 0.45 - 0.23)

with basic soil properties (tSa, tClay, $CaCO_3$, pHw, OM, and Fe-oxides). Observed level of association of individual WSA indices with soil properties including mc, totC**,** OC, totN, WDC, WDSa, tClay, tSa, CEC, LL, PL, PI, pH2.5 and Fe (Table 4.5), was considered to influence calibration of the indices on spectral based predictors.

Table 4.5: Pearson's correlations coefficient (*R* - value) of wet stable aggregation (WSA) indices with soil properties.

Property	mame	masp	mime	misp	unme	unsp	mechR	sponR
mc	-0.07	-0.02	-0.29	-0.40	0.25	0.32	-0.02	0.04
pHw2.5	-0.39	-0.43	-0.11	-0.08	0.37	0.42	-0.33	-0.36
totC	0.49	0.47	-0.09	-0.33	-0.30	-0.14	0.52	0.57
totN	0.57	0.58	0.02	-0.19	-0.44	-0.34	0.58	0.65
OC	0.52	0.53	-0.08	-0.30	-0.34	-0.22	0.55	0.62
inC	0.15	0.10	-0.06	-0.21	-0.07	0.08	0.17	0.16
Fe	0.40	0.38	-0.01	-0.15	-0.29	-0.21	0.39	0.45
CEC	-0.30	-0.29	-0.41	-0.52	0.51	0.63	-0.21	-0.17
tclay	-0.21	-0.16	-0.36	-0.46	0.40	0.48	-0.13	-0.06
tSa	0.21	0.14	0.21	0.43	-0.30	-0.43	0.16	0.07
WDC	-0.39	-0.33	-0.47	-0.51	0.61	0.64	-0.29	-0.21
WDSa	0.37	0.31	0.24	0.42	-0.45	-0.57	0.30	0.20
WDSi: WDC	0.32	0.26	0.32	0.13	-0.46	-0.31	0.26	0.28
PL	0.37	0.41	-0.13	-0.33	-0.17	-0.05	0.35	0.43
LL	0.09	0.13	-0.29	-0.46	0.15	0.26	0.12	0.20
PI	-0.06	-0.03	-0.33	-0.45	0.28	0.38	-0.01	0.05

4.1.2 Soil predictors of WSA indices from CART screening

Modest cross-validated decision tree models (2 terminal nodes, TN) were grown for stable micro aggregate (mime, misp) and unstable (unme, unsp) fractions. No decision tree could be grown for macro aggregate (mame, masp*)* and ratio (mechR, sponR*)* indices ($RE \geq 1.0$).

CART regression decision tree typology and model statistics is illustrated using unme (Figure 4.3 screenshot**)**. The tree shows one split (2 TN) at the Root node (upper panel); the relative error (RE) curve showing the optimal tree (tree with 2 TN, RE = 0.714) at one standard error (1SE) from the minimum cost tree (tree with 10 TN, RE = 0.654) (lower panel), and; model statistics: potential (59) predictor variables, model selected important (6) predictors, 2 TN and minimum (62) samples in a TN (bottom right hand corner). Notably, the largest (maximal) tree (27 TN) is less accurate (higher RE) than the minimal cost tree (10 TN), and has similar accuracy as the optimal (2 TN) tree**.** The larger trees (more TN) are complex and overfit and less robust a factor in favour of the small tree, a key quality feature of CART regression analysis (Steinberg & Colla, 2001).

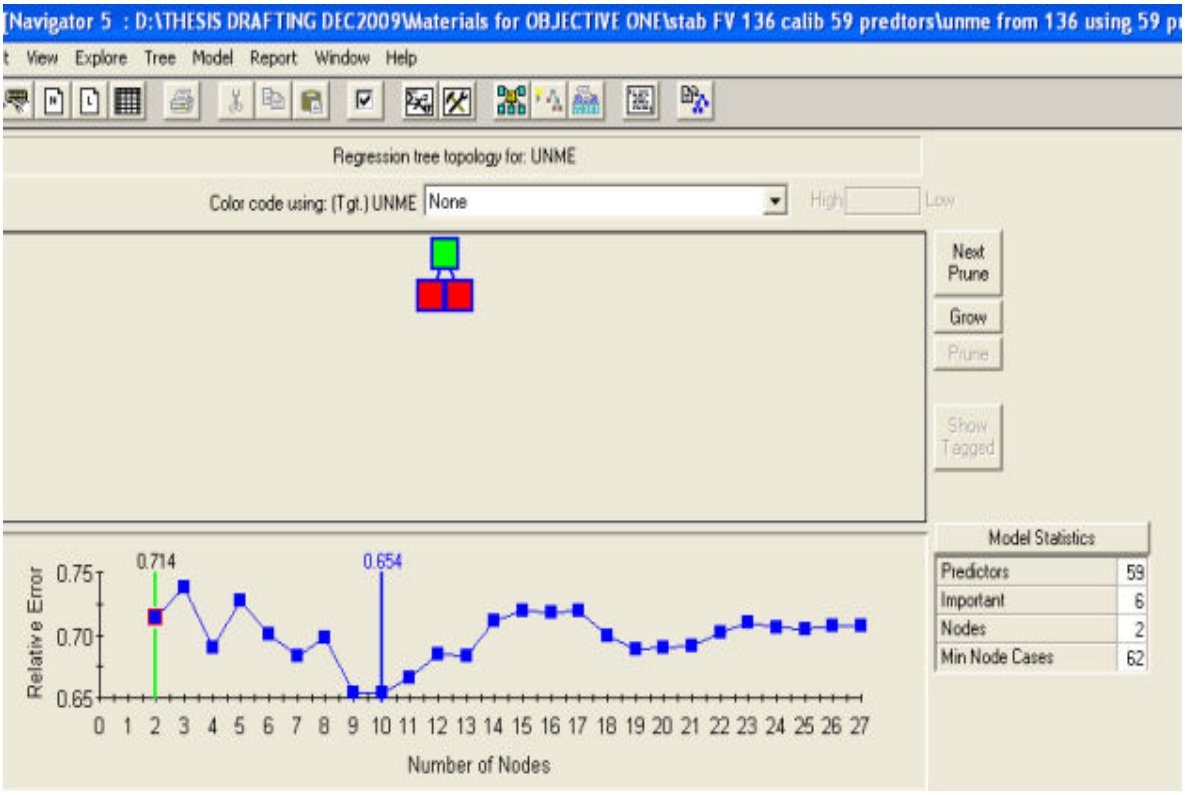


Figure 4.3: CART regression decision tree typology and model statistics for prediction of unme.

Six important variables (from total 59) could together explain variation in unme and unsp with accuracy of 28.6 % (RE 0.714) and 35.6 % (RE 0.644), respectively (see model statistics in Figure 4.3). The exchangeable sodium (eNa) at a split point 1.25 (unme) and 0.98 (unsp) cmol (+) kg^{-1} provided the greatest improvement in reducing heterogeneity in unme and unsp data (dichotomous separation), and was designated the primary splitter (soil predictor) of unme and unsp (Table 4.6)". Six important variables could together explain variation in mime and misp with accuracy of 18.0 % (RE 0.822) and 31.7 % (RE 0.683), respectively. The water dispersible clay (WDC) at a split point 34. 0 % (mime) and 32.5 % (misp) provided the greatest improvement in reducing heterogeneity in mime and misp data, and was designated the primary splitter (soil predictor) of mime and misp (Table 4.6).

Table 4.6: Soil predictors of wet stable aggregation indices and their competitor and surrogate variables.

WSA indices	RE [a]	primary split	split point	competitor [b]	surrogate [c]
unme	0.71	eNa	1.25	ESP, WDC, eNaR, ESI 1, sCa	ESP, eNaR, ESI 1, Na2.5, sNa
unsp	0.64	eNa	0.98	ESP, WDSa, Na2.5, WDC, eNaR	ESP, eNaR, Na2.5, Na5, ESI 1
mime	0.82	WDC	34	eNa, ESP, eNaR, tClay, CEC1	δg, tSa, WDSi:WDC, WDSa, tClay
misp	0.68	WDC	32.5	tClay, sMg, eNa, mc, ESI 1,	tClay, δg, WDSa, eMg, WDSi:WDC

Note: [a]*RE, relative (prediction) error;* [b]*Competitor variable compete for the primary split;* [c]*Surrogate is "back-up splitter" when the primary splitter is missing (for instances missing cases in the data set).*

Table 4.6 present also top five split competitors and surrogates of eNa and WDC. The competitor and surrogate variables are listed sequentially in order of importance. Notably, all eNa surrogates were Na-based, whereas WDC surrogates were based on particle size (except eMg for misp). This affirmed that stability could be assessed from

measurement of Na^+ concentration (Ward & Carter, 2004) and/or from particle-size distribution (Canasveras *et al*., 2010), and; superiority of eNa over ESP (Cook & Muller, 1997).

At 1.25 and 0.98 cmol (+) kg^{-1}, eNa provided the best dichotomous separation of calibration samples along the mean values of unme and unsp, respectively (Steinberg & Golovnya, 2006). *S*amples with eNa <= 1.25 (unme) or <= 0.98 (unsp) were allocated to the Left Child Node (TN 1), while samples with values > 1.25 (> 0.98) were allocated to the Right Child Node (TN 2) in the binary split **(**Figure 4.3**)**. Cases in the Left Child Node comprised soils with lower than mean levels of unme (or unsp) thus fell under relatively stable category. Soils in the Right Child Node presented higher than mean levels of unme (or unsp) and fell under less stable category.

Figure 4.4 (screenshot) illustrates box plots for the binary split and terminal node (TN) contents for unme. The blue box depicts the inter-quartile range, with the top of the box (upper hinge) marking the 75^{th} quartile and the bottom (lower hinge) marking the 25^{th} quartile. The horizontal (green line) denotes the node-specific median value, while the whiskers (upper and lower fences) extend to ± 1.5 times the inter-quartile range. Red crosses (+) represents values outside the fences, probably outliers (Steinberg & Golovnya, 2006). The eNa split resulted in clear separation of unme data into higher than average (~ 64 %) values (eNa > 1.25) represented by TN 2 (right hand box plot) and to lower than average values of unme (eNa ≤ 1.25) represented by TN 1(left hand box plot); TN 1 had median unme value of 58 % while TN 2 had median value 69.2 % presenting, therefore, considerable data overlap, a measure of residual impurity or unexplained variation. The tail end of TN 1 box plot indicates presence of samples with extremely low unme values (probably influential outliers).

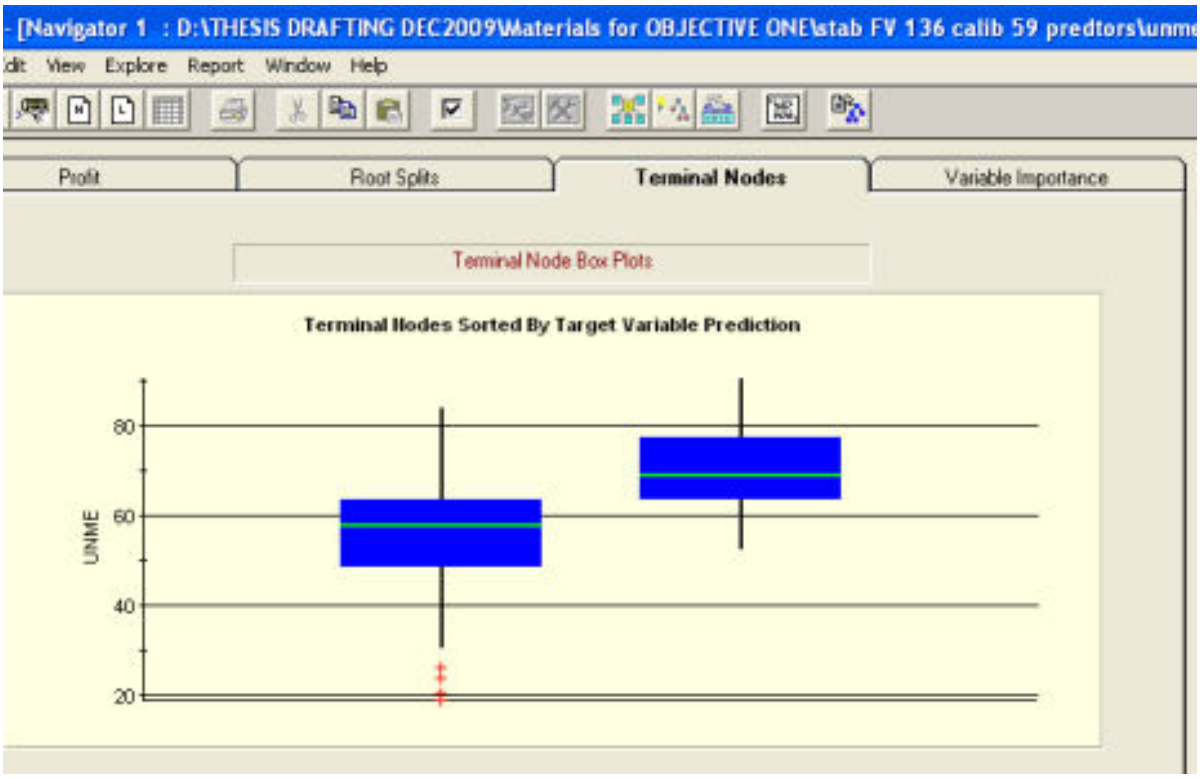


Figure 4.4: Box plots and terminal node (TN) allocation of unme by eNa split for calibration samples.

At 34 and 32.5 %, WDC provided the binary split and separation of the calibration samples along the mean values of mime and misp, respectively. Samples with WDC ≤ 34 % (mime) or ≤ 32.5 % (misp) were allocated to the Left Child Node (TN 1), while samples with WDC values > 34 (> 32.5) were allocated to the Right Child Node (TN 2) (similar to Figure 4.3).

Figure 4.5 **(**screenshot) illustrates box plots and TN allocation of misp by WDC split. All samples with WDC ≤ 32.5 % were allocated to the Right hand box plot and represented soils that were relatively stable (with higher than mean value of misp (~ 38 %). All samples with WDC > 32.5 were allocated to the Left hand box plot representing soils that were relatively less stable (lower than mean values of misp). The split resulted

in a clear separation of misp data. The TN 2 had a median misp value of 29 % while TN 1 had median value 44 % presenting considerable overlap, indication of relatively high unexplained variance; extreme values (tail end of TN 1) suggested potential influential outlier samples.

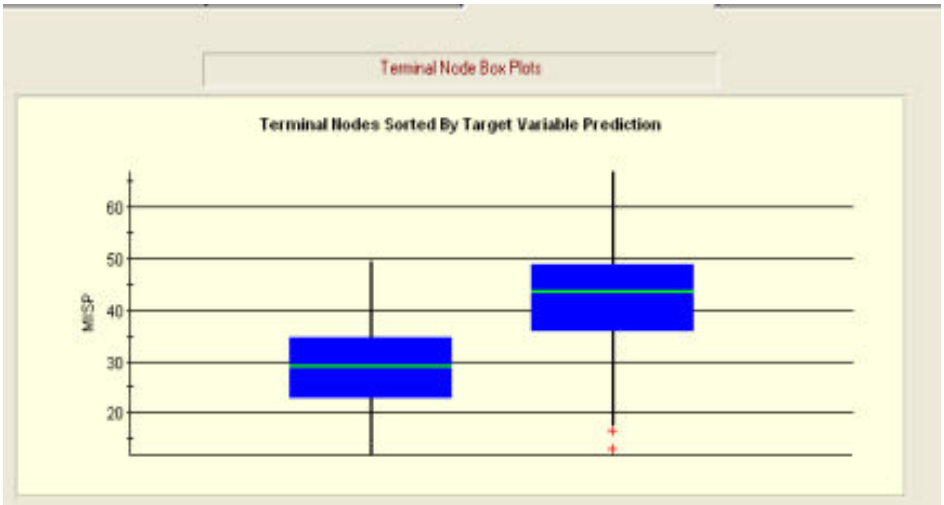


Figure 4.5: Box plots and terminal node (TN) allocation of misp by WDC split for calibration samples.

Prognostic intuition is key for proper interpretation of CART regression split terminal node (TN) allocation (Figure 4.3) and TN box plot contents (Figure 4.4 and 4.5) (Steinberg & Golovnya, 2006). Low values of unme (or unsp) are to be associated, for example, with low values of eNa, whereas low values of misp (or mime) are to be associated with high values of WDC.

The quality of data split reflects the efficacy of the predictive relationship between soil predictors and target WSA indices. For values of eNa > 1.25 (vertical line, Figure 4.6), for example, there were no values of unme lower than 50 %, and for values ≤ 1.25, only few samples hand unme larger than 70 %. At WDC > 32.5 % (vertical line, Figure 4.7),

majority of the soils indicated misp values < 40.0 %, whereas for values of WDC ≤ 32.5 %, only few samples had misp values lower than 30 %.

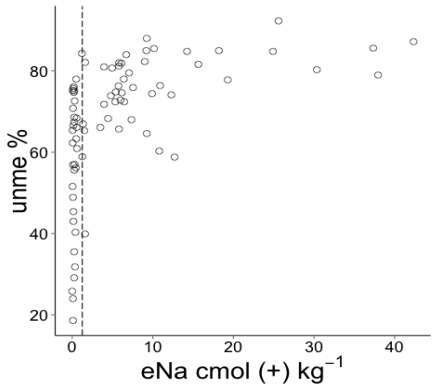


Figure 4.6: Relationship of eNa and unme in the calibration sample set (vertical line indicates unme (%) data dichotomous split by eNa at 1.25 cmol (+) kg^{-1}); raw data plot was for calibration set).

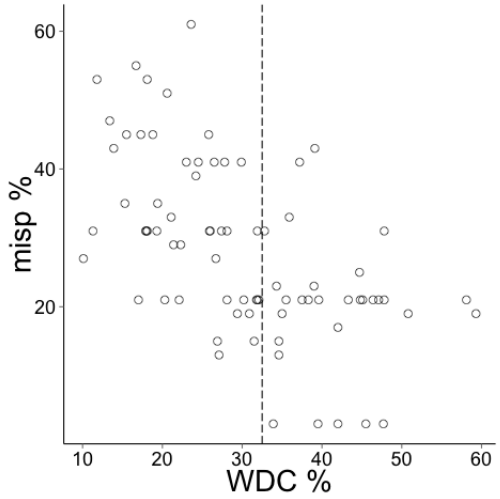


Figure 4.7: Relationship of WDC and misp in the calibration sample set (vertical line indicates misp (%) data dichotomous split by WDC at 32.5 %; raw data plot was for calibration set).

The soils were inherently unstable demonstrated by concentration of most of the samples at high level of unme, even at very low levels of eNa. The eNa split was edge cut where all samples with unme > 50 % had eNa > 1.25 (Figure 4.6), whereas WDC presented a more even split and separated majority of misp at 40 - 45 % (Figure 4.7).

The modest dichotomous separation of stable micro and unstable aggregate fractions (appreciable unexplained heterogeneity) (Figures 4.4 - 4.7), suggested that eNa and WDC individually provide weak indicators of soil (in) stability. This was associated partly also with observed modest correlation of eNa with unme and unsp (Pearson's correlations coefficient R-value 0.53-0.42) and WDC with mime and misp (R-value 0.51-0.47) (Table 4.5). Soil behavior especially aggregate breakdown and dispersion under wetting field conditions is a function of sodicity among other factors (types and amount of clay, electrolyte concentration, amount of SOM, and soil physical or mechanical disturbance (Ward & Carter, 2004).

The (59) soil properties had no predictive relationship with stable macro aggregates (mame, masp) and ratio indices (mechR and sponR) for the calibration set. However, CART suggested soil organic carbon (SOC) among potential predictors. This affirmed other works that link stability of macro aggregates to enmeshing by organic materials (mainly plant roots) (Wei *et al*., 2006). CART ranked eNa among zero - score importance variables for prediction of mame. The eNa indicated very poor ($r < 0.2$) (albeit negative) association with mame, similar to (negative) $r = 0.05$ for correlation of eNa with soil macro-aggregate (2000 - 250 μm) reported by Chang *et al*. (2001). The more abrasive slaking plus mechanical disruption wet sieving pretreatment might have obscured sensitivity of macro aggregates to eNa as suggested by Levy et al. (2003). Notably, ESP was among potential splitters for prediction of masp, affirming that sodicity is more important in slaking of macro aggregates (Levy *et al*.,2003) as shown also by the Crumb Test (Boucher, 2010). Successful development of CART decision

tree models is a function of, among others, the provenance of test soils. Using the validation sample set and a similar set of (59) soil properties, decision tree model was generated for mame (RE = 0.599) with soil predictors as pH2.5 in water and soil organic carbon (SOC). At pH2.5 values ≤ 5.85 few samples had mame values > 15 % (Figure 4.8 (a)). Few samples indicated mame values > 15 % for SOC values lower than split point at 2.2 % (Figure 4.8 (b)). This affirmed established relationship of aggregation with pH (Idowu, 2003; Auerswald, 1995), and with SOC (Wei *et al*., 2006). CART regression screening the suite of (59) soil properties suggested that eNa, pH2.5 water, SOC, and WDC were the key predictors of selected (8) WSA indices.

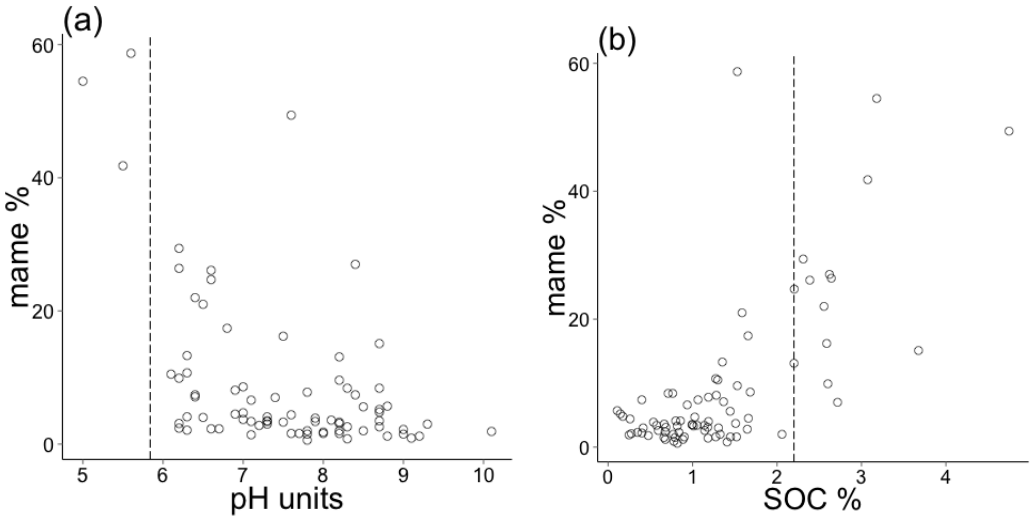


Figure 4.8: Relationship of (a) mame and pHw and (b) mame and SOC (dotted vertical lines indicates mame (%) data dichotomous split by pHw at 5.84 and by SOC at 2.2 %, respectively; raw data plots were for validation set).

Optimal wavelet data input for PLS modeling: There was no difference in performance (R^2 and RMSECV) for PLS prediction of the WSA indices for NIRwc and MIRwc entered in The Unscrambler as spectra or as non-spectral data. Models for WSA indices (except mime) were optimal with scaling (1/SD) of MIRwc. All NIRwc models

were optimal without scaling. The NIR model run time averaged 58.0 s, whereas corresponding NIRwc averaged 8.0 s, a reduction of 88 %. MIR model run time averaged 94 s, whereas corresponding MIRwc averaged 8.0 s, a reduction of > 90 %. NIRwc indicated degraded model performance (lower R^2 and higher RMSECV) compared to NIR. Improved model performance was observed with MIRwc relative to MIR (except for CEC 1) (Table 4.7).

Table 4.7: Model computation time and performance for NIR and MIR and corresponding wavelet coefficients (NIRwc and MIRwc) for selected soil properties.

	NIR			NIRwc			MIR			MIRwc		
Test	R^2	**RMSE**	**t, s**	R^2	**RMSE**	**t, s**	R^2	**RMSE**	**t, s**	R^2	**RMSE**	**t, s**
pH2.5	0.74	0.09	56	0.65	0.11	8	0.72	0.1	94	0.74	0.09	7
totC	0.79	0.18	61	0.56	0.26	9	0.89	0.13	94	0.90	0.12	8
eNa	0.55	1.81	54	0.50	1.89	8	0.74	1.39	100	0.75	1.34	9
tSi	0.23	8.92	62	0.14	9.18	8	0.33	8.06	87	0.37	7.85	8
Cu	0.46	0.12	66	0.27	0.14	8	0.61	0.1	98	0.63	0.10	10
tClay	0.46	11	55	0.38	11.72	8	0.60	9.43	89	0.65	8.84	8
CEC1	0.75	0.7	53	0.74	0.72	8	0.88	0.48	99	0.87	0.51	8

Note: *R^2, coefficient of determination; RMSE, root mean square error with cross-validation; t, model run time in second (s))*

Wavelet transform effectively reduced the MIR spectral data matrix to slightly over 7 % of its original size [(128/1755)×100], and NIR data matrix to slightly more than 12 % [(128/1030)×100] of its original size. Viscarra Rossel and Lark (2009) achieved reduction of soil vis-NIR (1076 wavebands) and mid-IR (933 wavebands) data to less than 7 % of their original size. Trygg and Wold (1998) reduced model computation time by almost 80 % and attained compression of vis-NIR spectral data to 3 % of its original size with almost no loss of information. There was up to 10 % improvement in R^2 and 6 % decline in prediction error by substituting MIR with MIRwc for the test runs (Table 4.7). Viscarra Rossel and Lark (2009) found also only a slight improvement using selected vis-NIR and MIR wavelet coefficients relative to vis-NIR and MIR PLS for

prediction of tClay (vis-NIR vs NIRwc: R^2 = 0.82, RMSE = 8.2 % vs R^2 = 0.86, RMSE = 7.1 %; MIRwc vs MIR: R^2 = 0.88, RMSE = 6.4 % vs R^2 = 0.90, RMSE = 5.8 %) and SOC (vis-NIR vs NIRwc: R^2 = 0.79, RMSE = 1.1 % vs R^2 = 0.86, RMSE = 0.9 %; MIRwc vs MIR: R^2 = 0.95, RMSE = 0.5 % vs R^2 = 0.94, RMSE = 0.5 %).

The preliminary test runs indicated that wavelet transform of NIR and MIR spectra drastically reduced model input data and speeded up model computation time. However, NIRwc caused degradation in model performance, whereas, there was only a slight gain in predictive ability by substituting MIR with MIRwc. The tests showed that for PLS modeling in The Unscrambler, it does not matter whether wavelet coefficients are entered as spectra or non-spectra data, however, MIRwc need to be standardized, whereas NIRwc should be used without standardization. The test runs suggested limited advantage including MIRwc and NIRwc as alternative predictors of WSA using PLS regression.

4.1.3 Calibration of WSA indices on IR-based predictors

Stable macro aggregate (mame/ masp): The looCV calibration for mame and masp was moderate to weak (R^2 = 0.58 - 0.42, RMSECV = 8.43 - 6.26 %, RPD = 1.64 - 1.47), across IR-methods (Table 4.8). This was attributed to variable effect of particle-size variation on the path of light and reflectance spectra for different samples (Stenberg *et al*., 2010). Observed performance was attributed to also moderate to weak correlation (R = 0.58 - 0.40) of mame/ masp with soil chromophores (totC, OC, totN, Fe, WDC) (Table 4.5) (Canasveras *et al*., 2010). The models were optimal without removal of outliers (optimization trial (i)) (Table 4.8). The MIR/MIRwc models were better than NIR/NIRwc, a result of MIR being energetic enough to excite molecular vibrations to higher energy levels than NIR (Bellon-Maurel & McBratney, 2011). Observed weaker performance for masp (R^2 = 0.54 - 0.42, RMSECV = 8.43 - 7.54 %, RPD = 1.64 - 1.48) was attributed to lower quality reference data for masp (Table 4.2**)** (Reeves, 2010).

Few comparative studies were available using interval- depth samples data sets. Performance for mame/ masp compared well with R^2 = 0.5, RMSE = 7.7 %, RPD = 1.4 (NIR) and R^2 = 0.56, RMSECV = 6.9 %, RPD = 1.5 (MIR) PLS prediction of WSA (> 250 μm) in surface soils from southern Spain (Canasveras *et al*., 2010). Performance compared well also with R^2 = 0.58 - 0.46, RMSECV = 5.67 - 4.28 %, RPD = 1.55 - 1.34 for individual macro aggregation indices (2000, 1000, 500, and 250 μm) using NIR-PCR in interval-depth samples (0-30 cm depth) of soils from Major Land Resource Areas (MLRAs) in the USA (Chang *et al*., 2001). However, prediction of macro aggregate fraction (0.25 to 2.0 mm) (R^2 = 0.60, RPD = 1.58) by Chang *et al*. (2001) indicated higher error than observed for this study (RMSECV = 14.01 vs 8.43 - 6.26 %), probably due to poor quality of their reference data attributed to error propagation in determination of the individual 4 aggregation indices.

Table 4.8: Performance of alternative IR-based methods for PLS looCV prediction of WSA indices and their soil-based predictor variables.

Indices	IR-method	PLS looCV model statistics					Optimization[3]
		PLS PCs	*n*	r^2	RMSE	RPD	
mame							
	NIR	3	128	0.56	6.52	1.56	(i)
	NIRwc	3	128	0.55	6.45	1.58	(i)
	MIR[b]	3	126	0.58	6.26	1.63	(i)
	MIRwc	2	127	0.56	6.32	1.61	(i)
masp							
	NIR	3	128	0.43	8.38	1.48	(i)
	NIRwc	3	128	0.42	8.43	1.47	(i)
	MIR[b]	3	127	0.54	7.54	1.64	(i)
	MIRwc	2	127	0.51	7.62	1.63	(i)
mime							
	NIR	14	127	0.40	7.2	1.26	(iii)
	NIRwc	1	127	0.11	8.45	1.08	(i)
	MIR	8	126	0.30	7.66	1.19	(iii)
	MIRwc[b]	9	125	0.41	6.9	1.32	(iii)
misp							
	NIR[b]	5	121	*0.39*	7.97	1.39	(ii)
	NIRwc	5	123	0.38	8.38	1.32	(ii)
	MIR	10	122	0.38	8.46	1.31	(ii)
	MIRwc	4	121	0.36	8.41	1.32	(ii)
unme							
	NIR[b]	13	126	*0.65*	7.37	1.83	(iii)
	NIRwc	5	127	0.45	9.25	1.46	(i)
	MIR	10	126	0.63	7.58	1.78	(iii)
	MIRwc	8	126	0.62	7.72	1.75	(iii)
unsp							
	NIR[b]	12	125	*0.62*	8.28	1.80	(iii)
	NIRwc	5	127	0.47	9.84	1.51	(iii)
	MIR	10	126	0.52	9.37	1.59	(iii)
	MIRwc	5	127	0.44	10.24	1.46	(i)
mechR							
	NIR[b]	3	128	*0.41*	0.38	1.34	(i)
	NIRwc	3	128	0.39	0.38	1.34	(i)
	MIR	3	127	0.39	0.38	1.34	(i)
	MIRwc	2	127	0.37	0.38	1.34	(i)

Table 4.8 Contd.

Indices	IR-method	PLS looCV model statistics					Optimization[3]
		PLS PCs	*n*	r^2	RMSE	RPD	
sponR							
	NIR	7	128	0.59	0.32	1.78	(i)
	NIRwc	6	128	0.45	0.35	1.63	(i)
	MIR[b]	3	127	0.45	0.36	1.58	(i)
	MIRwc	2	127	0.44	0.36	1.58	(i)
WDC							
	NIR[b]	18	118	0.75	5.82	2.08	(iii)
	NIRwc	6	117	0.64	6.78	1.78	(iii)
	MIR[b]	8	117	0.73	5.88	2.06	(iii)
	MIRwc	6	119	0.70	6.32	1.91	(iii)
eNa							
	NIR	5	127	0.44	25.45	0.50	(iii)
	NIRwc	5	126	0.37	10.64	1.18	(iii)
	MIR	11	127	0.11	38.18	0.33	(i)
	MIRwc[b]	9	124	0.73	5.15	2.45	(iii)

Note*: Optimization*[a]*: (i), prediction without removal of outliers; (ii), prediction with removal of identified spectral (Robust Mahalanobis) outliers; (iii), prediction with removal of reference values outliers;* [b] *the best IR-method for each indices is marked; NIRwc were used without scaling (1/SD) of the coefficients whereas MIRwc were with scaling).*

Stable micro aggregate (mime/ misp): Calibration for mime and misp was weak to poor across IR-methods (R^2 = 0.40 - 0.11, RMSECV = 8.5 - 6.9 %, RPD = 1.4 - 1.1). This is a reflection of the negative effect of particle-size variation on the path of light and reflectance spectra, including specular reflectance, especially for the < 425 µm particle-size fraction (Bellon-Maurel & McBratney, 2011). This was attributed to also weak correlation (R = 0.5 - 0.4) of the indices with mc, tClay, WDC, LL and PI and poor correlation ($R \leq 0.3$) with totN, totC, and OC (Table 4.5). Prediction for mime and misp was comparable across IR-methods (except mime using NIRwc). Notably, mime models were optimal with removal of reference values outliers (except NIRwc). The misp

models were optimized with removal of spectral outliers. The mime using NIRwc presented the poorest model (R^2 = 0.1, RMSECV = 8.5 %, RPD = 1.1) (Table 4.8**)**.

Observed poor prediction of misp/ mime was contrary to R^2 = 0.8, RMSECV = 3.83 %, RPD = 2.22 (MIR) and R^2 = 0.92, RMSECV = 2.41 %, RPD = 3.53 (NIR) for estimation of aggregate stability index (fraction < 250 µm) for mixed depth (0-30 cm depth) sample set reported by Madari *et al*. (2006). Madari *et al*. (2006) followed, however, very rigorous spectral data pretreatment (a total of 22 different spectra pretreatments were tested) prior to calibration development, compared with only first derivative spectra and smoothing used in this study. Rigorous spectra pretreatment could, however, over train calibration with implication on model robusticity (CAMO ASA Inc., 1998).

Unstable aggregate fraction (unme/unsp): Calibration for unme and unsp was satisfactory to moderate (R^2 = 0.7 - 0.5, RPD = 1.8 - 1.5) across IR-methods (Table 4.8**)**. This fairly good performance was attributed to observed moderate correlation (R = 0.6 - 0.5) with texture parameters tClay, WDC, WDSa (Table 4.5**)**, considered spectrally active (Viscarra Rossel *et al*., 2008). Notably performance was better than for stable macro and micro aggregate fractions. Presumably metastable macro- and micro-aggregate information lost through wet-sieving was added to the unstable fraction (Marquez *et al*., 2004). The weaker models for unsp (R^2 = 0.62 - 0.44, RMSECV = 10.24 - 8.28, RPD = 1.80 - 1.46) was attributed to lower quality reference data (Table 4.2**)**. The MIR/NIR models were superior (but more complex) than MIRwc/ NIRwc, indicating that suggested improved performance with spectra wavelet transform (Viscarra Rossel & Lark, 2009) is not supported for all soil properties. However, wavelet transform presented more parsimonious models as found by Viscarra Rossel and Lark (2009). The wavelet-based models were optimal with removal of reference values outliers. Exception was unme (NIRwc) and unsp (MIRwc) that were optimal without

removal of outliers (Table 4.8). There were no available studies in the literature to compare with these results.

Ratio macro: micro (mechR/sponR): Calibration of sponR (masp: misp) and mechR (mame: mime) was moderate to weak (R^2 = 0.6 - 0.4, RPD = 1.8 - 1.3) across IR-methods (Table 4.8**)**. This was attributed to observed moderate correlation (R = 0.65 - 0.40) of the indices with basic properties totC, OC, totN, Fe, and PL (Table 4.5**).** The performance suggested superior influence on behavior of the soils of stable macro aggregates (mame, masp) over micro aggregates (mime, misp). The models were optimal without removal of outliers. The NIR models were superior over MIR, probably due to microstructural information loss by sample grinding for MIR (Canasveras *et al*., 2010), and; enhanced performance in the NIR from effect of water-light interactions (Stenberg, 2010). The sponR models were more reliable (higher RPD) than mechR, mame, masp, mime and misp. Directly measured properties are better calibrated to spectra than derived properties (Stenberg *et al*., 2010). This was the case for mame and mechR; however this was not true for masp and sponR (Table 4.8). Also, observed lower prediction of mechR (R^2 = 0.4, RPD = 1.3) than sponR (R^2 = 0.6 - 0.5, RPD = 1.8 - 1.6) was unexpected since mame/mime presented better quality reference data than masp/misp (Table 4.2). No immediate adequate explanation was found. Probably relict non-linear data distribution even with logarithmic transformation (Table 4.3), where PLS analyses is suboptimal (Linker, 2012), had more (negative) influence on mechR model.

Exchangeable sodium content (eNa): Calibration for eNa ranged from satisfactory to very poor (R^2 = 0.7 - 0.1, RPD = 2.5 - 0.3) across IR-methods, a reflection of high variation in calibration performance of spectrally non-responsive (secondary) soil properties, whose calibration is based on strength of association with responsive soil properties (Stenberg *et al*., 2010). Observed poor performance for eNa (R^2 = 0.4 - 0.1,

RMSECV = 38.2 - 10.6 cmol (+) kg^{-1}, RPD = 1.2 - 0.3) (Table 4.8**)**, was attributed to poor association ($r < 0.3$) of eNa with totC, OC, tClay and WDC (Appendix 3); relic skewed data distribution (Appendix 2), where PLS analyses is suboptimal (Linker, 2012). Prediction was optimized with removal of reference values outliers (except MIR model). Prediction followed the order: MIRwc >> NIR > NIRwc >> MIR (Table 4.8).

Few available studies (Viscarra Rossel *et al*., 2008; Pirie *et al*., 2005; Islam *et al*., 2003) have reported on IR-based prediction of eNa using interval-depth samples sets data sets. Observed prediction using NIR (R^2 = 0.44) was comparable, however, with R^2 = 0.46 reported by Islam *et al.* (2003) and better than R^2 = 0.18 attained by Pirie *et al.* (2005) for prediction of eNa. Chang *et al*. (2001) found very poor NIR-PCR prediction of eNa (R^2 = 0.09, RPD = 0.92) and attributed this to among others, the very low and narrow range of eNa values (mean: 0.2, min: 0.1, max: 1.8, cmol (+) kg^{-1}). Observed performance using MIR (R^2 = 0.11, RPD = 0.33) was poorer than R^2 = 0.39 and RPD = 1.2 found by Viscarra Rossel *et al*. (2008) and R^2 = 0.2 and RPD = 1.1 reported by Pirie *et al*. (2005) for prediction of eNa. High performance (R^2 = 0.73, RMSECV = 5.2 cmol (+) kg^{-1}, RPD = 2.5) was observed for estimation of eNa using MIRwc.

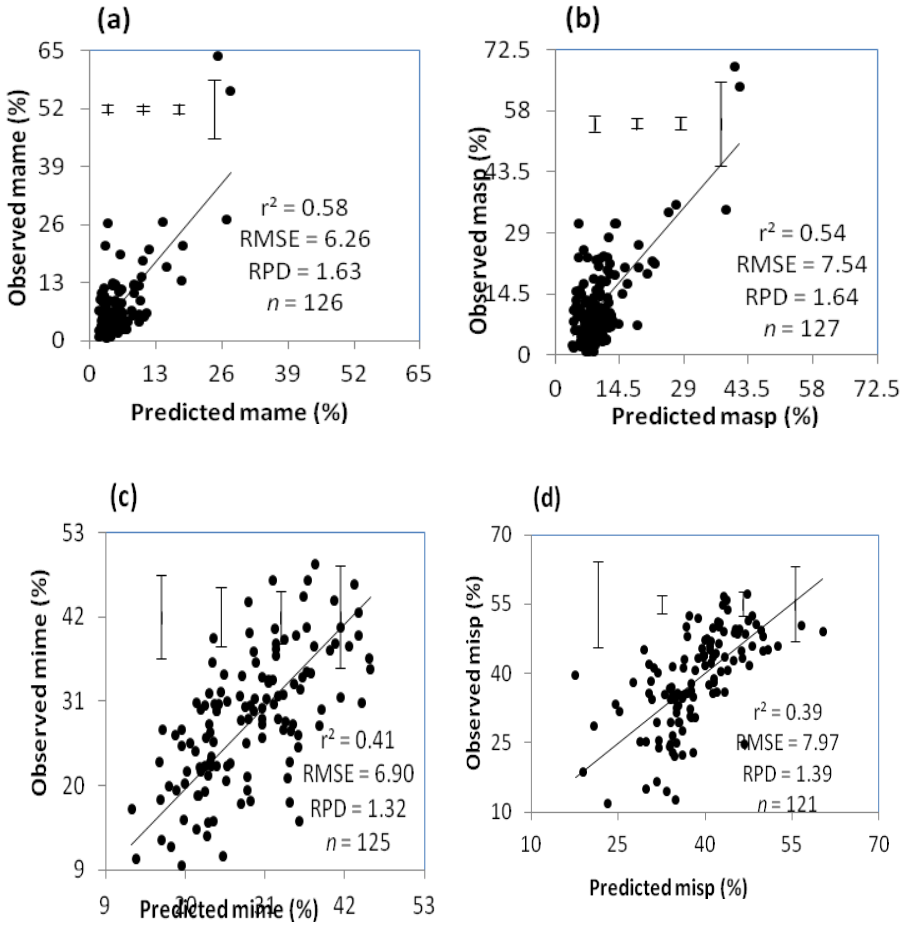
The selection of 128 wavelet coefficients from 1750 MIR spectral wavebands (together with exclusion of one influential outlier), effectively removed noise in the data resulting in better correlation of the coefficients with eNa reference data. However, Viscarra Rossel and Lark (2009) found only slight improvement using selected relevant MIR wavelet coefficients relative to MIR PLS for prediction of tClay (R^2 = 0.90 vs 0.88) and SOC (R^2 = 0.95 vs 0.94). Also, wavelet transform of spectral variables was not superior for prediction of WSA indices (exception mime) (Table 4.8). Observed great variation in MIRwc and MIR performance for prediction of eNa is insightful and requires further investigations.

Water dispersible clay content (WDC): The looCV estimation of WDC was satisfactory (R^2 = 0.75 - 0.64, RPD = 2.1 - 1.8) across IR-methods, attributed to the fact that soil texture exhibits a primary response to IR spectra (Viscarra Rossel *et al*., 2008; Chang *et al*., 2001). The WDC indicated also strong association with tClay (r = 0.78), WDSa (r = 0.80) and modest association (r = 0.67) with mc. There was insignificant difference in WDC prediction performance across IR-methods (except NIRwc that indicated weaker performance) (Table 4.8), affirming spectral activity across IR-methods. The models were optimal with removal of reference values outliers. The MIR/ NIR models were superior over MIRwc/ NIRwc, however, Viscarra Rossel and Lark (2009) found slight model improvement for estimation of tClay using MIRwc. Canasveras *et al*. (2010) found that UV-vis-NIR (R^2 = 0.66) and vis-NIR (R^2 = 0.55) performed better than MIR (R^2 = 0.30) for estimation of WDC.

The weaker performance of WDC relative to other spectrally active soil constituents like mc ($R^2 \geq 0.8$) for the same data set, could be a result of the negative effect of particle-size variation for different samples on the path of light and reflectance spectra, especially for the < 425 µm particle-size fraction (Bellon-Maurel & McBratney, 2011). However, WDC was better estimated than tClay (NIR: R^2 = 0.46, RMSECV = 11 %, RPD = 1.4 and MIR: R^2 = 0.6, RMSECV = 8.7 %, RPD = 1.7), for the same data set. This suggested that WDC correlates better with spectra than tClay. The WDC indicated also strong association (r = 0.7) with LL, whose strong looCV prediction (NIR R^2 = 0.8, RPD = 2.7; MIR: R^2 = 0.9, RPD = 2.8) was predicated on the influence of water-light interactions, suggesting aquaphotomics (Stenberg, 2010) enhance WDC performance. There were no available studies in the literature using mixed depth samples data sets for estimation of WDC. For surface horizon samples Canasveras *et al*. (2010) found weaker prediction of WDC using NIR (R^2 = 0.66 - 0.55, RMSECV = 2.9 - 3.3 %, RPD = 1.7 - 1.5) and MIR (R^2 = 0.30, RMSECV = 4.2 %, RPD = 1.2). Notably, Canasveras *et al*. (2010) WDC prediction indicated lower determination coefficients (R^2) and reliability

(lower RPD) but higher accuracy (lower prediction error) than reported study, suggesting higher measurement errors (SD) from their reference methods.

Plots of predicted vs measured values for the best IR-based PLS looCV prediction of WSA indices and their soil-based predictor variables are presented in Figure 4.9.



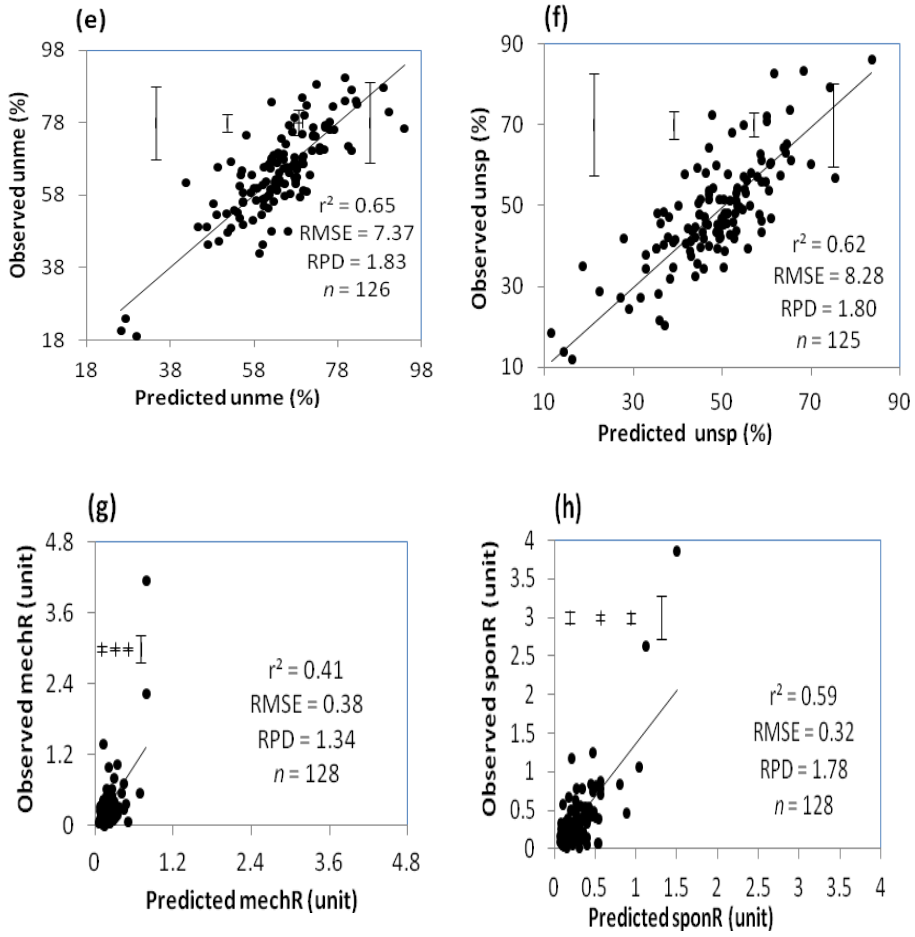


Figure 4.9: Scatterplot plots for predicted vs observed values for the best PLS looCV predictions of WSA indices ((a), mame using MIR; (b), masp using MIR; (c), mime using MIRwc; (d), misp using NIR; (e), unme using NIR; (f), unsp using NIR; (g), mechR using NIR; (h), sponR using NIR; (i), WDC using NIR, and; (j), eNa using MIRwc). Indicated on each plot also is R^2, coefficient of determination; RMSECV, root mean square error of cross validation; RPD, the ratio of prediction deviation (SD/RMSECV); n, number of samples used; 1:1 regression target line, and error bars for quartile data).

The concentration of data points at lower values of mame/ masp (Figure 4.9 (a) and (b)) and at higher values of unme (Figure 4.9 (e)) and less evident for unsp (Figure 4.8 (f)), affirms that the soils are largely unstable as observed in CART regression (Figures 4.4, 4.6 and 4.7)). Large standard errors for quartile data suggests that PLS analyses is less useful for extreme values of microaggregates (Figure 4.9 (c) and (d)), WDC (Figure 4.9 (i)), and is more useful for non-sodic soils (Figure 4.9 (j)). Data distribution for ratio indices mechR and sponR (Figure 4.9 (g) and (h)), illustrate the predominant influence of macro aggregate over micro fractions for the studied soils.

Other workers (Mouazen *et al*., 2005; Sorensen & Dalsgaard, 2005) found improved performance for fit-for-purpose models of soil properties by thresh holding data range to develop separate models. Such strategy could improve calibration for mame/ masp and unme.

4.1.4 Calibration of WSA indices on soil- based predictors

The WDC was the key predictor of stable micro aggregates mime and misp, whereas eNa predicted unstable fraction unme and unsp. There were no soil-based predictors for mame, masp, mechR and sponR (Table 4.9). Prediction without removal of outliers was superior for estimation of mime and unsp. Prediction of misp and unme was optimal with removal of reference values outliers**.**

Table 4.9: Performance of IR- and soil-based methods for estimation of WSA indices.

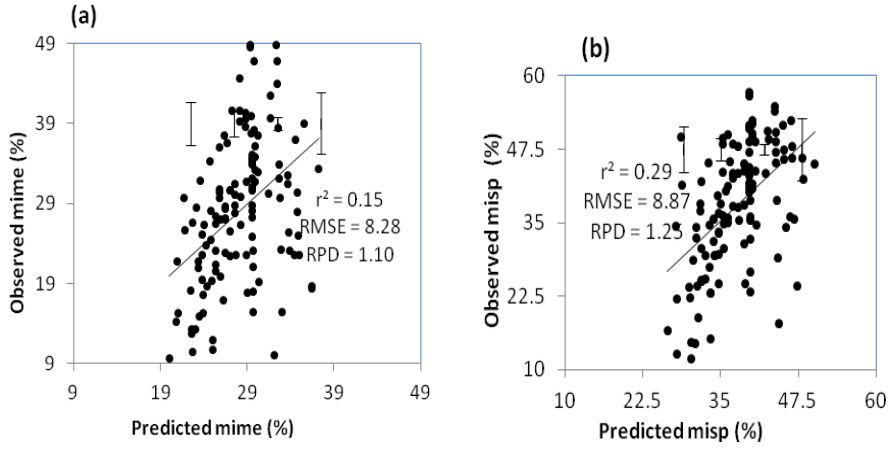
		PLS looCV prediction				PLS looCV prediction		
Indices	**IR-predictor**	R^2	**RMSECV**	**RPD**	**soil predictor**	R^2	**RMSECV**	**RPD**
mame	MIR	0.58	6.26	1.63	N/A	N/A	N/A	N/A
masp	MIR	0.54	7.54	1.64	N/A	N/A	N/A	N/A
mime	MIRwc	0.41	6.90	1.32	WDC	0.15	8.28	1.10
misp	NIR	*0.39*	7.97	1.39	WDC	0.29	8.87	1.25
unme	NIR	0.65	7.37	1.83	eNa	0.29	10.62	1.27
unsp	NIR	0.62	8.28	1.80	eNa	0.45	11.04	1.35
mechR	NIR	0.41	0.38	1.34	N/A	N/A	N/A	N/A
sponR	NIR	0.59	0.32	1.78	N/A	N/A	N/A	N/A

Note*: N/A (not applicable), means there were no soil-based predictors*.

Prediction of mime and misp was poor ($R^2 \leq 0.2$, RMSECV = 10.0 - 8.3 %, RPD = 1.1). Prediction of misp was slightly improved (R^2 = 0.29, RMSECV = 8.9 %, RPD = 1.3), however, with removal of three (3) outliers (Table 4.9). Calibration performance (albeit weak), reflects the (inverse) predictive relationship between WDC (< 20 μm) and stable micro aggregate fraction (212 - 20 μm) (Figure 4.7). The unsp could be modestly predicted (R^2 = 0.5, RMSECV = 11.0 %, RPD = 1.4), whereas, unme was poorly predicted (R^2 = 0.3, RMSECV = 10.6 %, RPD = 1.3) (Table 4.9). Better estimation of unsp than unme affirmed the stronger resonance of soil eNa content with spontaneous (slaking) aggregate disruption in agreement with Levy *et al*. (2003) and the Crumb Test (Boucher, 2010).. There was scant available information (Canasveras *et al*., 2010) on development of soil-based PLS models for prediction of aggregation indices from wet sieving. Canasveras *et al*. (2010) reported R^2 = 0.23, 0.43 and 0.49 for prediction of water stable aggregates (fraction > 250 μm), mean-weight diameter, and water dispersible clay (WDC), respectively, using six soil-based predictors (tSa, tClay, pH-water, $CaCO_3$, OM, and Fe). Notably, Canasveras *et al*. (2010) considered WDC among

functional attributes, whereas rules on pedotransfer functions (PTFs) (Minansny, 2007), would suggest inclusion of WDC among basic soil PTFs, as applied in this study.

Figure 4.10 presents scatterplot plots for predicted vs observed values for the soil-based PLS looCV predictions of WSA indices. The wide scatter of observation points from the target line suggested weak relationships. Scatter was particularly more for lower values for unme and unsp (see error bars, (c) and (d)). Narrow data range for predicted mime and misp ((a) and (b)), suggested insensitivity of the soil-based predictors to extreme values of the indices. Concentration of data points at higher values of unme ((c)), affirmed high aggregate instability (see also Figure 4.6). Slaking is mild and allows sensitivity to lower values of eNa (Figure 4.10 (d)).



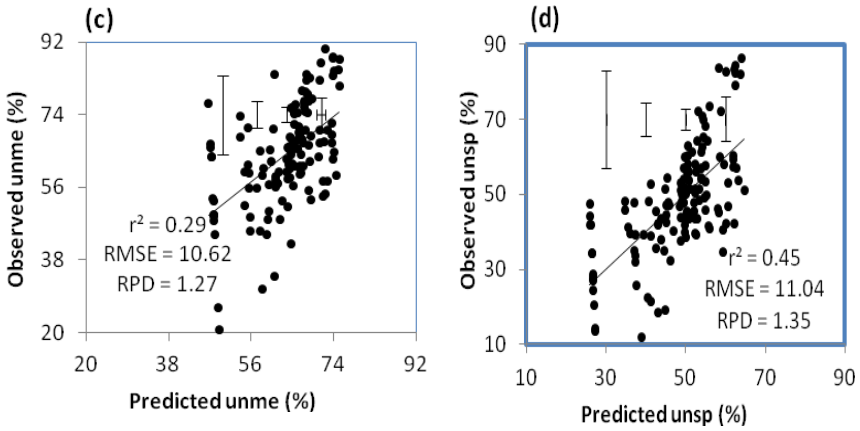


Figure 4.10: Scatterplot plots for predicted vs observed values for the soil-based PLS looCV predictions of WSA indices ((a) mime using WDC, (b) misp using WDC, (c) unme using eNa, and (d) unsp using eNa.

Performance of IR- and soil-based models: There were no soil-based predictors of mame, masp, mechR and sponR, whereas IR could predict these attributes with accuracy ranging from 41 to 59 % (Table 4.9). The IR predictors were superior over soil-based predictors, however, WDC could closely match NIR for prediction of misp and eNa could closely match NIR for prediction of unsp. Canasveras *et al*. (2010) found R^2 = 0.23 for prediction of water stable aggregates using six basic soil properties, whereas spectral predictors achieved R^2 = 0.6, affirming superiority of spectral predictors over soil-based predictors. Performance for indices from slaking plus mechanical disruption (mime and unme) was lower than from slaking only (misp and unsp) wet sieving pretreatment using soil-based predictors. This was contrary to IR-based methods where slaking plus mechanical disruption models indicated higher performance than slaking only (Table 4.9). No immediate adequate explanation was found. The soil is a non-ideal system; mechanisms of soil processes are only partially understood and the fundamental links between measured soil chemistry and particular soil attributes for specific data sets may be complex (Stenberg *et al*., 2010). Probably the abrasive wet-sieving pretreatment

obliterated all sensitivity for the weak predictive relationships of the WSA indices with soil- based predictors, whereas IR readily detects and quantifies subtle changes in absorption features (Shepherd & Walsh, 2002).

4.1.5 PLS and IR calibration efficiency and fitness for purpose

Overall performance: PLS looCV prediction of the WSA indices was modest to weak (R^2 = 0.65 - 0.39, RPD = 1.8 - 1.3) (Table 4.10). Prediction performance followed the general order: unme ~unsp > mame~masp ~sponR > mechR > mime~misp. Models for indices from slaking plus mechanical disruption wet sieving pretreatment were superior over those from slaking only pretreatment across IR-methods (except mechR). There was no clear trend on performance of alternative IR-methods for particular WSA indices**;** however, MIR indicated superiority for stable macro aggregates, whereas NIR was superior for stable micro aggregates, ratio-based indices and unstable fraction (Table 4.10**)**. Canasveras *et al*. (2010) concluded that the wavelength providing the highest R^2 and accuracy is specific to the soil property.

Table 4.10: Model statistics and associated IR-method for optimal PLS looCV estimation of WSA indices and their soil-based predictor variables.

Indices	IR-method	PLS PCs	R^2	RMSE	RPD	*n*	optimisation
mame	MIR	3	0.58	6.26	1.63	126	(i)
masp	MIR	3	0.54	7.54	1.64	127	(i)
mime	MIRwc	9	0.41	6.9	1.32	125	(iii)
misp	NIR	5	*0.39*	7.97	1.39	121	(ii)
unme	NIR	13	0.65	7.37	1.83	126	(iii)
unsp	NIR	12	0.62	8.28	1.80	125	(iii)
mechR	NIR	3	0.41	0.38	1.34	128	(i)
sponR	NIR	7	0.59	0.32	1.78	128	(i)
WDC	NIR	18	0.75	5.82	2.08	118	(iii)
eNa	MIRwc	9	0.73	5.15	2.45	124	(iii)

Note: *Optimization: (i), prediction without removal of outliers (exception influential outliers); (ii), prediction with removal of identified spectral (Robust Mahalanobis) outliers; (iii), prediction with removal of reference values outliers.*

The looCV models for mame, masp, mechR and sponR were optimal without removal of outliers across alternative IR-methods. Removal of reference values outliers optimized models for unme, unsp, WDC and eNa (Table 4.10). Spectral outliers had no influence on estimation of the WSA indices (except misp using NIR), affirming that there were no serious atypical samples from the calibration sample set (see Figure 4.1).

Highly reliable PLS looCV predictions were obtained for eNa (RPD = 2.5) and WDC (RPD = 2.1) using MIRwc and NIR, respectively. Moderately reliable predictions were obtained, in decreasing order (RPD, IR-method), for: unme (1.83, NIR), unsp (1.80, NIR), sponR (1.78, NIR), masp (1.64, MIR) and mame (1.63, MIR). Models for mime (MIRwc) and mechR (NIR) were unreliable (RPD < 1.4).

Soil-based PLS looCV predictions were unreliable with, RPD in decreasing order: unsp (1.35), unme (1.27), misp (1.25) and mime (1.10). Reliability for mame and masp using MIR (RPD = 1.64) and using NIR (RPD = 1.6 - 1.5) (Table 4.8) was better than RPD = 1.5 and RPD = 1.4 - 1.3 for looCV estimation of WSA using MIR and NIR, respectively found by Canasveras *et al*. (2010). Reliability for the WDC models was higher than RPD = 1.7 - 1.5 (using NIR ranges) and RPD = 1.2 (using MIR) reported by Canasveras *et al*. (2010). Observed performance was in the same range as RPD = 1.4 - 2.0 for NIR-PCR prediction of wet aggregation measures macro-, 1.0 and 0.5 mm and better than RPD < 1.4 for the 2 and 0.25 mm aggregates achieved by Chang *et al*. (2001).

The PLS looCV IR-based analyses were unreliable for assessment of mime and mechR (RPD = 1.3). Important is that model improvement strategies including improving quality of reference data (Reeves, 2010) could be used to improve reliability for prediction of misp (RPD = 1.4).

A statistical description (mean, SD and data range) of the observed WSA indices analyzed using conventional methods of analyses and their looCV predictions using IR-methods is shown in Table 4.11. In most cases there was a good correspondence between predicted and observed ranges. Even for poor calibrations (mime, misp), the predictions are in the same order of magnitude as observed values. However, some of the estimates were less accurate than those obtained by routine laboratory methods.

Table 4.11: Observed WSA indices analyzed using conventional methods of analyses and their looCV predictions using IR- methods.

	Observed		Predicted	
Indices	**mean ± SD**	**data range**	**mean ± SD**	**data range**
mame	8.0 ± 10.2	0.1 - 63.9	5.8 ± 4.4	1.8 - 27.8
masp	13.6 ± 12.4	0.8 - 68.5	10.5 ± 6.2	3.9 - 41.4
mime	28.4 ± 9.1	9.6 - 48.9	28.8 ± 7.5	12.5 - 45.5
misp	38.1 ± 11.1	11.8 - 66.9	38.4 ± 7.1	17.6 - 60.3
unme	63.6 ± 13.5	19.0 - 90.4	63.9 ± 10.8	26.2 - 93.9
unsp	48.4 ± 14.9	12.0 - 86.2	48.4 ± 11.8	11.5 - 83.7
mechR	0.3 ± 0.5	0.0 - 4.2	0.2 ± 0.1	0.1 - 0.8
sponR	0.4 ± 0.6	0.0 - 3.9	0.3 ± 0.2	0.1 - 1.5
WDC	26.9 ± 12	5.0 - 60.0	26.9 ± 11.0	5.2 - 47.9
eNa	6.6 ± 12.6	0.0 - 63.8	4.5 ± 9.6	0.0 - 65.9

Given the relative speed and cost of IR approach and the large local variation of soil properties, the ability to analyse large number of samples involving multiple variables at finer sampling intervals using the spectroscopic technique may in some circumstances outweigh the loss in analytical accuracy. The IR is also known to be more reproducible (precise) than the reference methods (Genot *et al*., 2011) and this could convey a distinct advantage. Ultimately, the key criterium for judging acceptable prediction accuracy and utility of soil IR-based PLS analysis is fitness for purpose, well illustrated by Sorensen and Dalsgaard (2005). For rapid preliminary site investigations, it might be sufficient to sort soils into stability classes (Canasveras *et al*., 2010), allowing to make management decision (for instance, site is stable or unstable).

Wide range in calibration performance: wide range in prediction performance was observed across alternative IR-methods (Table 4.12). This is in agreement with reported highly variable and sometimes weak soil chemometrics and spectra-based models. This is due to several reasons, including: complexity in the composition, spectral inactivity of several soil constituents; lack of correlation of secondary properties with primary

properties, poor quality reference data, and; confounding and distortion of weak spectral intensities, due to overlap with more intense absorption features (Stenberg *et al*., 2010). Weak and less reliable estimation could be due to also artifacts related to sample handling (size-sorting, redistribution); use of first derivatives of reflectance data may remove some of particle size influence (Chang *et al*., 2001), and; calibration model used may be less appropriate for analyzing some of the soil properties.

Table 4.12: Range in performance (R^2) of alternative IR-based methods for PLS looCV prediction of WSA indices and their soil-based predictor variables.

Indices	NIR	NIRwc	MIR	MIRwc	across IR-methods
mame	0.56-0.31	0.55-0.31	0.58-0.46	0.56-0.43	0.58-0.31
masp	0.43-0.19	0.42-0.18	0.54-0.34	0.51-0.33	0.54-0.18
mime	0.40-0.36	0.11-0.10	0.30-0.27	0.41-0.33	0.41-0.10
misp	0.39-0.38	0.38-0.37	0.38-0.37	0.36-0.34	0.38-0.34
unme	0.65-0.54	0.45-0.33	0.63-0.51	0.62-0.50	0.65-0.33
unsp	0.62-0.51	0.47-0.39	0.52-0.47	0.44-0.39	0.62-0.39
mechR	0.41-0.16	0.39-0.14	0.39-0.18	0.37-0.30	0.41-0.14
sponR	0.59-0.35	0.45-0.34	0.45-0.30	0.44-0.29	0.59-0.29
WDC	0.75-0.72	0.64-0.54	0.73-0.67	0.70-0.67	0.75-0.54
eNa	0.44-0.41	0.37-0.27	0.11-0.10	0.73	0.73-0.10

The PLS method is easy and straight forward and among the most commonly used methods for spectra and chemometrics soil multivariate calibration (Naes *et al*., 2002). PLS spectroscopy analysis is suboptimal, however, for property data with non-linear behavior. Spectra wavelet transform is among proposed spectra pretreatment prior to calibration development (Linker, 2012).

Influence of spectra wavelet transform: From the preliminary test run (see section 4.1.3), WSA indices models were superior without standardization (1/SD) of NIRwc, whereas the models were superior with scaling of MIRwc (except mime). The models were based on calibration set (n = 136) and without optimization (handling of outliers).

For optimized models (using reduced calibration sample size, n = 128), NIRwc models without standardization were clearly superior over NIRwc models with standardization for estimation of all WSA indices and for optimization trials (i) through (iv). Variation was observed, however, with MIRwc models as follows: without removal of outliers or with removal of spectral outliers, models without scaling were superior over those with scaling for mame, masp, unsp, mechR, and sponR; with removal of reference values outliers, models without scaling were superior for masp, unsp, and sponR, and; with removal of spectral and reference values outliers models without scaling were superior for mame, masp, unsp, and sponR (data not shown**)**. A number of observations can be made, that: (i) the sample size (and attendant samples composition) influence effect of wavelet coefficients in PLS MIR analyses; (ii) optimal preprocessing of MIRwc input data for PLS analyses is specific for WSA indices, suggesting low efficiency for studies involving large sample sizes and multiple variables; (iii) the NIRwc was not superior over alternative methods for estimation of any of the WSA indices, and; (iv) MIRwc was overall superior for only mime and eNa (Table 4.8). This suggested limiting advantage with spectra wavelet transformation relative to Fourier transform (DRIFT) in PLS multivariate calibration in soil aggregation studies. There were no available soil aggregation related comparative studies. Viscarra Rossel and Lark (2009) found only a slight improvement in prediction of tClay with MIR wavelet transform (R^2 0.90 vs 0.88 and RMSEP 5.77 vs 6.35 %).

Performance of MIR relative to NIR: Models from MIR were weaker than NIR for WSA indices (mime, misp, unme, unsp, mechR, and sponR) and WDC (Table 4.8). This is despite reported superiority of MIR over NIR for estimation of several soil physical and chemical properties (Viscarra Rossel *et al*., 2006). Notably, Madari *et al*. (2006) and Canasveras *et al*. (2010) did not find MIR superior over NIR for prediction of wet aggregation indices. The moisture related absorption features (bands around 1430, 1920, and 2200 nm) are prominent in NIR (Figure 4.1), and probably aggregation indices data

based on wet sieving find higher resonance with the moisture absorption features in NIR range. Presumably water-light interactions (aquaphotomics), recently introduced in soil spectroscopy studies by Stenberg (2010) provide additional insight in the mechanism responsible for higher calibration of WSA indices in the NIR region. Important, however, is the robusticity of the NIR looCV models *Vis a Vis* those for MIR, especially for independent predictions. Higher model complexity of looCV calibrations has negative implication also for independent validation (Naes *et al*., 2002). The NIR PLS looCV performed better than MIR for instance mime, misp, unme, unsp, mechR, sponR, and WDC, however, the NIR models indicated similar or higher complexity than MIR models (Table 4.8). This suggested potential for lower NIR robusticity when subjected to independent testing. The higher performing NIR PLS looCV model for prediction of aggregate stability index (fraction < 250 μm) found by Madari *et al*. (2006) was more complex (9 PLS factors) than the corresponding MIR model (5 PLS factors), however, the authors did not report on independent testing of the looCV models. Madari *et al*. (2006) found that the NIR-PLS looCV prediction of otherwise spectrally active soil carbon (10 PLS factors) degraded from R^2 = 0.90 to 0.58, whereas corresponding model using MIR (4 factors) degraded from R^2 = 0.93 to 0.74, with somewhat independent testing, suggesting higher robusticity of MIR analysis. Other workers (Viscarra Rossel *et al*., 2006) found MIR models more parsimonious than corresponding NIR models for several soil properties.

Implication for soil stability studies: The scanty available literature on spectroscopic calibration of measures of aggregation (Canasveras *et al*., 2010; Madari *et al*., 2006) has focused on few stability indices water stable aggregates and aggregate size distribution, although soil aggregation and aggregate stability is among key functional attributes that define the dynamism in soil behaviour. This study has demonstrated efficacy also of measures of aggregate "instability" and also the potential influence of mild/ spontaneous

(slaking only) and abrasive (slaking plus mechanical shaking) on aggregate breakdown dynamics.

Slaking common to soils subject to flooding or (basin, furrow) irrigation present hazard for dams through sub-terranean (piping, sink holes) (Bell, 2000), and gully development via retreat through slumping highly prevalent in lowland LVB of Kenya. Slaking plus mechanical disruption as common to heavy tropical storms (Barthes & Roose, 2002*)* or disturbances from tillage practices (Ashmana & Hallett, 2003), enhances surface runoff, nutrient depletion, sedimentation and eutrophication. The mame/ masp indicate distribution of macro pores for aeration and water movement, whereas mime/misp indicates distribution of water retention pores. Boix-Fayos *et al*. (2001) found small aggregate sizes (1.0 – 0.105 and < 0.105 mm) to have a positive influence on soil water retention and they seemed a good indicator of soil degradation. The mechR (ratio mame: mime) and sponR (masp: misp) provide good indicator of porosity (Mbagwu, 1992), and soil air- moisture balance.

The study affirmed eNa and WDC as "wet stable aggregation indices", as suggested by Canasveras *et al*. (2010) that is, WDC. The study affirmed also, stability threshold at 1-2 cmol (+) kg^{-1} for eNa and and 34 % for WDC, suggested in earlier works by US Salinity Laboratory Staff (1954) and Knodel (1991), respectively.

This study showed that spectra based predictors were superior over soil-based predictors for estimation of WSA indices, in agreement with Canasveras *et al*. (2010) who found that 4 spectra PCs were superior over 6 soil-based predictors (sand, clay, $CaCO_3$, pHw, OM, Fe) for estimation of water stable aggregates (R^2 = 0.42 vs 0.23), and that a combination of spectra and the soil predictors was even better (R^2 = 0.53). This suggested great potential of spectral variables and /or in combination with soil-based

variables for development of pedotransfer functions for estimation of soil aggregation and aggregate stability.

Canasveras *et al*. (2010) considered water dispersible clay (WDC) among functional attributes, whereas Rules on pedotransfer functions (PTFs) by Minansny (2007), would suggest inclusion of WDC among basic soil PTFs as applied by Igwe (2005) and also in this study. The WDC is, however, a measure of stability of soils (for instance Oxisols) whose aggregation is not supported by the hierarchical model concept (Azevedo & Schulze, 2007). Important is the renewed discovery of the great potential of WDC in aggregate stability studies and the need to mainstream its determination in routine soil survey and mapping and in soil capability assessments.

Further validation requirement: The looCV testing provides a good indicator of the robustness of a model, especially when calibration sample set is well selected using a conditioned Latin hypercube sampling strategy (Stenberg, 2010), also used in this study. This validation strategy could overestimate, however, the predictive performance of a model since samples from the same set are used for calibration and validation (Brown *et al*., 2005). The looCV models present practical application challenges, foremost being limitation for geographic transferability, especially for spectrally non-responsive soil properties. The efficacy of IR-based models could only be ascertained with independent testing using samples collected after calibration development (Sorensen & Dalsgaard, 2005). The robusticity of the looCV models of WSA indices and their soil-based predictors (eNa and WDC) were, therefore, subjected to further testing using data sets of similar soils from independent sites.

4.2 Further validation of IR-based models for estimation of WSA indices

Results are presented for looCV calibration and independent validation of WSA indices (including their soil-based predictor variables eNa and WDC) for alternative IR-based

methods. Models are for calibration (n = 128) and validation (n = 79) data sets. The influence of outliers on calibration performance is assessed for validation optimization trials as follows: Trial (i), prediction without removal of outliers in validation set; Trial (ii), prediction with removal of reference values outliers (samples in validation set with reference data out-of-range of calibration set); Trial (iii), prediction with removal of spectral outliers in validation set, and; Trial (iv), prediction with removal of both reference values and spectra outliers.

4.2.1 Calibration and independent validation of WSA indices for alternative IR-based methods

PLS looCV calibration of WSA indices and corresponding performance for independent set for validation without removal of outliers in the validation set is presented in Table 4.13.

Calibration of WSA indices: Stable macro aggregates: The looCV calibration for mame and masp was moderate (R^2 = 0.6 - 0.4, RPD = 1.6 - 1.5), across IR-methods (Table 4.13). This was attributed to the effect of particle-size, shape and aggregation on light transmission through the soil and hence its reflectance (Chang *et al*., 2001). Performance was attributed to also modest correlation (R = 0.58 - 0.40) of the indices with soil chromophores (totC, OC, totN, Fe, WDC (Table 4.5). The MIR/ MIRwc models were superior over NIR/ NIRwc). The MIR range is energetic enough to excite molecular vibrations to higher energy levels than NIR (Bellon-Maurel & McBratney, 2011). The weaker models for masp (R^2 = 0.5 - 0.4, RPD = 1.6 - 1.5) were attributed to lower quality reference data (Table 4.2) (Reeves, 2010). Performance for mame and masp compared well with R^2 = 0.5 and RPD = 1.4 (NIR) and R^2 = 0.56 and RPD = 1.5 for estimation of water stable aggregate (fraction > 250 μm) reported by Canasveras *et al*. (2010). Prediction of mame and masp using NIR compared well with R^2 = 0.58 - 0.46

and RPD = 1.55 - 1.34 reported by Chang *et al*. (2001) for prediction of macro aggregation indices (2000, 1000, 500, and 250 μm).

Table 4.13: Calibration and independent validation statistics for estimation of WSA indices for alternative IR-based methods without removal of outliers in the validation set.

		PLS looCV model statistics					Validation with Trial (i)		
Soil test	**IR-method**	**PLS PCs**	***n***	R^2	**RMSECV**	**RPD**	R^2	**RMSEP**	**RPD**
mame	MIR	3	126	0.58	6.26	1.63	0.70	7.42	1.37
	MIRwc	2	127	0.56	6.32	1.61	0.72	7.68	1.33
	NIR	3	128	0.55	6.52	1.56	0.81	7.40	1.38
	NIRwc	3	128	0.55	6.45	1.58	0.57	8.36	1.22
masp	MIR	3	127	0.54	7.54	1.64	0.26	13.53	0.92
	MIRwc	2	127	0.51	7.62	1.63	0.31	12.73	0.97
	NIR	3	128	0.43	8.38	1.48	0.42	11.44	1.08
	NIRwc	3	128	0.42	8.43	1.47	0.35	53.12	0.23
mime	MIRwc	9	127	0.33	7.68	1.18	0.13	9.48	0.96
	NIR	14	128	0.36	7.54	1.21	0.03	35.10	0.26
misp	MIR	10	127	0.37	8.64	1.28	0.14	15.03	0.74
	MIRwc	4	126	0.34	8.55	1.30	0.21	14.64	0.76
	NIR	5	126	0.38	8.12	1.37	0.27	15.62	0.71
	NIRwc	5	128	0.37	8.50	1.31	0.34	14.65	0.76
unme	MIR	10	127	0.60	8.03	1.68	0.44	13.56	1.00
	MIRwc	8	127	0.58	8.21	1.64	0.56	11.24	1.20
	NIR	13	127	0.61	7.83	1.72	0.56	12.50	1.08
	NIRwc	5	127	0.45	9.25	1.46	0.65	12.16	1.11
unsp	MIR	10	127	0.51	9.67	1.54	0.21	18.33	0.81
	MIRwc	5	127	0.44	10.24	1.46	0.37	17.37	0.86
	NIR	12	128	0.56	9.30	1.60	0.46	16.41	0.91
	NIRwc	5	128	0.45	10.28	1.45	0.48	16.04	0.93
mechR	MIR	3	127	0.39	0.38	1.34	0.59	0.37	1.38
	MIRwc	2	127	0.37	0.38	1.34	0.64	0.36	1.42
	NIR	3	128	0.41	0.38	1.34	0.50	0.41	1.24
	NIRwc	3	128	0.39	0.38	1.34	0.31	0.42	1.21
sponR	MIR	3	127	0.45	0.36	1.58	0.21	0.58	0.98
	MIRwc	2	127	0.44	0.36	1.58	0.26	0.55	1.04
	NIR	7	128	0.59	0.32	1.78	0.14	0.60	0.95
	NIRwc	6	128	0.45	0.35	1.63	0.32	0.51	1.12
eNa	MIR	11	126	0.49	9.34	1.35	0.30	15.68	0.80
	MIRwc	9	126	0.73	10.30	1.22	0.46	14.79	0.85

		PLS looCV model statistics					Validation with Trial (i)		
Soil test	**IR-method**	**PLS PCs**	***n***	***R^2***	**RMSECV**	**RPD**	***R^2***	**RMSEP**	**RPD**
	NIR	5	127	0.41	45.52	0.28	0.14	10.01	1.26
WDC	MIR	8	120	0.68	6.75	1.79	0.65	9.13	1.33
	MIRwc	6	120	0.68	6.75	1.79	0.71	8.25	1.47
	NIR	18	120	0.72	6.33	1.91	0.40	14.18	0.85
	NIRwc	6	121	0.55	8.08	1.50	0.39	11.50	1.05

Note: (*PLS-PCs, partial least-square regression factors; n, calibration samples without influential outliers (computed as N-n; where N = 128; except WDC where N = 121); R^2, coefficient of determination; RMSECV, root mean-square error of cross validation; RMSEP, root mean-square error of prediction; looCV models for sample set (n = 128) from across LVB of Kenya were further validated with independent set (n = 79) from selected two sentinel sites of LNY and HB from LVB; NIRwc data was used without standardization, whereas MIRwc was standardized; RMSECV/ RMSEP is shown for backtransformed values; looCV models with R^2 < 0.3 (mime: MIR and NIRwc; eNa: NIRwc) were not considered for further validation).*

Stable micro aggregates: Calibration for mime and misp was weak (R^2 = 0.4 - 0.3, RPD = 1.4 - 1.2) across IR-methods (Table 4.13), attributed to weak correlation (R = 0.5 - 0.4) with mc, tClay, WDC, LL and PI and poor correlation ($r \leq 0.3$) with totN, totC, and OC (Table 4.5). The mime models for MIR and NIRwc were very poor (R^2 = 0.27, RMSECV = 7.53 %, RPD = 1.2 and R^2 = 0.11, RMSECV = 8.44 %, RPD = 1.1, respectively). The NIR/ NIRwc methods were superior over MIR/ MIRwc for both mime and misp. The NIR (< 2 mm) is much more affected by the surface physical properties (such as size and shape of soil aggregates) because of their influence on the light scattering and light paths lengths (Chang *et al*., 2001), unlike MIR (< 0.1 mm) whose smoother surface results in higher specular reflectance. Superiority of NIR is presumably due to also loss of microstructure information with sample grinding for MIR as suggested by Canasveras *et al*. (2010). Probably also due to the effect of water-light interactions (aquaphotomics) (Stenberg, 2010) more pronounced in the NIR region. The

MIR/ NIR models were more complex than MIRwc/ NIRwc models (Table 4.13), due to the smoothing effect of spectra wavelet transform (Viscarra Rossel & Lark, 2009). Madari *et al*. (2006) found higher performance for prediction of micro aggregate indices (fraction < 250 μm) using both MIR (R^2 = 0.8; RPD = 2.2) and NIR (R^2 = 0.9; RPD = 3.5). Madari *et al*. (2006) followed very rigorous spectral data pretreatment prior to calibration development, whereas only first derivative spectra and smoothing was used prior to calibration of mime and misp.

Unstable aggregate fraction: The looCV calibration for unme and unsp was moderate (R^2 = 0.61 - 0.44, RPD = 1.72 - 1.45), across IR-methods (Table 4.13), attributed to the moderate correlation (R = 0.6 - 0.5) with texture parameters tClay, WDC, WDSa, and ratio WDSi: WDC (Table 4.5), considered spectrally active. The weaker models for unsp was attributed to lower quality reference data (with CV % 6.5 - 0.1 and 2.1 - 0.1 for unsp and unme, respectively). The performance for unstable fraction was much better than for micro aggregate fraction (mime, misp). Presumably metastable micro- aggregate information lost through wet-sieving was added to the unstable fraction (Marquez *et al*., 2004). The MIR/ NIR models performed better (but were more complex) than MIRwc/ NIRwc for both unme and unsp (Table 4.13), suggesting that improved performance with wavelet transform is specific to the soil property.

Ratio macro: micro: Estimation of sponR (ratio masp: misp) and mechR (ratio mame: mime) was moderate to weak (R^2 = 0.6 - 0.4, RPD = 1.8 - 1.3) across IR-methods (Table 4.13). This was attributed to moderate correlation (R = 0.65 - 0.40) of the indices with basic properties totC, OC, totN, Fe, and PL (Table 4.5). Estimation of sponR (R^2 = 0.6 - 0.5, RPD = 1.8 - 1.6) was better than mechR (R^2 = 0.4, RPD = 1.3). This was unexpected since mame and mime indicated higher quality reference data than masp and misp (Table 4.2), and; calibration of mame was superior over masp. Performance for the ratio indices was better than for individual indices for sponR and mechR (than mime). This

suggested that the widely held fact that directly measured properties presented better models than derived properties (Bellon-Maurel & McBratney, 2011), could not be supported. The NIR/ NIRwc methods were superior over MIR/ MIRwc for sponR and mechR (Table 4.13). This was partly ascribed to microstructural information loss with sample grinding for MIR (Canasveras *et al*., 2010), and; enhanced performance in the NIR from effect of water-light interactions (Stenberg, 2010). There were no available comparative studies in the literature.

eNa content: Calibration of eNa was satisfactory to very poor (R^2 = 0.7 - 0.3, RPD = 1.4 - 0.3) across IR-methods (Table 4.13), affirming high variability for secondary calibration (Stenberg *et al*., 2010) . The poor performance (R^2 = 0.5 - 0.4, RMSECV = 45.5 - 9.3 cmol (+) kg^{-1}, RPD = 1.4 - 0.3), was attributed to poor association of eNa (r < 0.3) with spectrally active totC, OC, tClay and WDC **(**Appendix 3**)**; and possible residual non-linear eNa data **(**see strongly skewed eNa in Appendix 2**)**, where PLS analyses is suboptimal (Linker, 2012). Observed high model fit using MIRwc (R^2 = 0.73), that Stenberg *et al*. (2010) refer to as 'occasional high performance' is characteristic of surrogate calibration (Reeves, 2010). The high performance was after excluding one identified influential outlier. The performance using NIR (R^2 = 0.41) was comparable with R^2 = 0.34 reported by Islam *et al.* (2003) and better than R^2 = 0.18 and 0.09 attained by Pirie *et al.* (2005) and Chang *et al*. (2001), respectively. Prediction of eNa using MIR (R^2 = 0.49, RPD = 1.35) was better than R^2 = 0.39 and RPD = 1.2 found by Viscarra Rossel *et al*. (2008) and R^2 = 0.2 and RPD = 1.1 reported by Pirie *et al*. (2005). Very poor looCV model for eNa (NIRwc: R^2 = 0.27, RMSECV = 50.53 cmol (+) kg^{-1}, RPD = 0.25) *(Saeys et al., 2005)* was not considered for further validation.

WDC content: The looCV calibration for WDC was satisfactory to moderate (R^2 = 0.7 - 0.6, RPD = 1.9 - 1.5) across IR-methods. This was attributed to strong association (r = 0.78) with tClay, affirming r = 0.76 of WDC with tClay by Canasveras *et al*. (2010).

There was insignificant difference in WDC prediction across IR-methods (except NIRwc) (Table 4.13), attributed to the fact that soil texture exhibits a primary response to IR spectra (Viscarra Rossel *et al*., 2008). The NIR presented the best estimation of WDC (R^2 = 0.72, RPD = 1.9) attributed to resonance to strong moisture absorption bands in NIR. The WDC indicated also strong association (r = 0.7) with LL and mc, suggesting that aquaphotomics (Stenberg, 2010) enhanced performance. The NIR model was, however, more complex (Table 4.13), with potential for lower robusticity. There was no available information on prediction of WDC for interval- depth samples data set. However, observed lower looCV prediction of tClay for the same data sets (NIR: R^2 = 0.46, RMSECV = 11 %, RPD = 1.4 and MIR: R^2 = 0.6, RMSECV = 8.7 %, RPD = 1.7) suggested that WDC correlated better with spectra than tClay. Canasveras *et al*. (2010) found weaker performance for prediction of WDC for surface horizon samples using NIR (R^2 = 0.66 – 0.55, RPD = 1.7 – 1.5) and MIR (R^2 = 0.30, RPD = 1.2).

Notably looCV models for WSA indices and eNa and WDC were similar as in previous study (see section 4.1.3; Table 4.8) since the same calibration sample set (n = 128) was used. The looCV models in section 4.1.3 were optimized with/ without removal of (reference values and spectral) outliers, to partly explain slightly better performance than corresponding looCV models. However, this inadvertently occasioned some level of overlap in observed results and discussion for respective WSA indices (Tables 4.8 and 4.13). This was deemed necessary to minimize cross-referencing.

The looCV testing provides a good indicator of the robustness of a model for a well selected calibration sample set for instance using a conditioned Latin hypercube sampling strategy (Viscarra Rossel *et al.* 2008). This testing could overestimate, however, the predictive performance of a model since samples from the same set are used for calibration and validation (Brown *et al*., 2005). The robusticity of the looCV

models of WSA indices and their soil-based predictors (eNa and WDC) were, therefore, further tested and optimized using data sets of similar soils from independent sites.

Validation without removal of outliers (optimization trial (i)): macro aggregates: The looCV model indicated excellent to moderate performance for independent estimation of mame across IR-methods (R^2 = 0.8 - 0.6, RPD = 1.4 - 1.2) (Table 4.13). This was attributed to robust response of particle-size and aggregation to spectra, and to presumed correlation of mame with soil chromophores (totC, OC, totN, Fe, WDC). The weak prediction of masp (R^2 = 0.4 - 0.3, RPD = 1.1 - 0.2) was attributed to low quality reference data from the reference method (Table 4.2) (Reeves, 2010). The NIR model was more robust for mame and masp (Table 4.13), affirming more robust effect of water-light interactions in the NIR region (Stenberg, 2010).

Micro aggregates: The weak mime and misp looCV models could not reproduce the indices in the independent set ($R^2 \leq 0.34$) (Table 4.13). This was attributed to poor correlation of the indices with mc, tClay, WDC, LL, PI, totC, and SOC (Table 4.5), also in the validation set. This was probably due to also confounding/distortion of their weak absorption features by strong features of SOC and quartz (Stenberg *et al*., 2010).

Unstable aggregate fraction: Satisfactory estimation of unme and unsp (R^2 = 0.7) was ascribed to robust association with texture parameters tClay, WDC and WDSa. The wide range in performance (R^2 = 0.7 - 0.2, RPD = 1.2 - 0.8) (Table 4.13), is a reflection of highly variable effect of particle-size variation on the path of light and reflectance spectra for the < 425 µm particle-size fraction, especially for samples from different provenance (Bellon-Maurel & McBratney, 2011). Similar high variability in performance for prediction of tClay (R^2: 0.88, 0.83, 0.72; RPD: 2.9, 2.4, 1.6) is reported in other studies (Terhoeven-Urselmans *et al*., 2010; Viscarra Rossel *et al*., 2008; Pirie *et al*., 2005). Lower performance for unsp (R^2 = 0.5 - 0.2, RPD = 0.9 - 0.8) was associated

with lower quality reference data (Table 4.2). Observed more robust NIR/NIRwc models than MIR/ MIRwc, suggested that aggregation indices from wet-sieving find higher resonance with the moisture absorption features in NIR range.

macro: micro ratio: The moderate to weak looCV mechR model was sustained (R^2 = 0.6 - 0.3, RPD = 1.4 - 1.2) across IR-methods (Table 4.13). This suggested robust (albeit moderate) correlation of mame and mime with basic properties totC, OC, totN, Fe, and PL for the independent set. Degradation of sponR (R^2 = 0.3 - 0.1, RPD = 1.0) could partly be associated with low quality of the reference data, that is, masp (Table 4.2), and higher looCV model complexity (NIR and NIRwc). The MIR/ MIRwc models were more robust than NIR/ NIRwc models for mechR, where the contrary was the case for looCV prediction (Table 4.13). No immediate explanation was found for this reversal.

eNa content: Prediction of eNa was weak to poor (R^2 = 0.5 - 0.1, RPD = 1.3 - 0.8) across IR-methods (Table 4.13), courtesy of surrogate calibration (Reeves, 2010), and residual skewness. The surrogate poor looCV performance (R^2 = 0.5 - 0.4, RPD = 1.4 - 0.3) was reliably sustained, however, for the independent set (R^2 = 0.5 - 0.1, RPD = 1.3 - 0.8), suggesting that surrogate relationship i.e. correlation of eNa with totC, OC, tClay, WDC (albeit weak) worked as well in independent as in the calibration set (Reeves, 2010). The calibration and validation soils came from the same soil population (Table 4.3). The drastic drop (collapse) of eNa MIRwc model (from R^2 = 0.73 to 0.46) (Table 4.13) with independent testing, is characteristic of surrogate calibrations.

WDC content: Estimation of WDC ranged from satisfactory to weak (R^2 = 0.7 - 0.4, RPD = 1.5) across IR-methods (Table 4.13), attributed to primary response of texture parameters to spectra. The wide range in performance is a reflection of highly variable effect of particle-size and shape variation on the path of light and reflectance spectra for different samples (Bellon-Maurel & McBratney, 2011). Models based on MIR/ MIRwc

were more robust (R^2 = 0.7, RPD = 1.3 - 1.5) than from NIR/ NIRwc (R^2 = 0.4, RPD = 0.9 - 1.1) (Table 4.13). This was a consequence of lower parsimony (higher complexity) of looCV NIR models. There were no available studies in the literature to compare with these results. However R^2 = 0.61, RMSEP = 13.3 %, RPD = 1.3 and R^2 = 0.52, RMSEP = 16.0 %, RPD = 1.1 was observed for independent testing of tClay looCV models of MIR and NIR respectively, for the same datasets. This suggested superiority of MIR over NIR for estimation of textural parameters.

Without removal of outliers in the validation set, looCV models for WSA indices based on slaking plus mechanical disruption (mame, unme, mechR) were more robust than those from slaking only wet-sieving pretreatment (masp, unsp, sponR). Notably models based on spectral variables (MIR/ NIR) were more complex, and less robust than corresponding models from wavelet transform variables (MIRwc/ NIRwc) (misp, unme, unsp, sponR). The MIRwc models were more robust, for example, than their MIR counterparts for all WSA indices, eNa and WDC (Table 4.13). The NIR models were more robust for mame, masp, mime, mechR, eNa and WDC, whereas NIRwc models were superior for estimation of misp, unme, unsp, and sponR (Table 4.13). Without removal of outliers, parsimonious wavelet transform-based models were more robust than their spectra-based counterparts. Viscarra Rossel and Lark (2009) found slight improvement for SOC and degradation for tClay using selected relevant wavelet coefficients.

4.2.2 Effect of outliers on independent estimation of WSA indices

Reference values outliers (optimization trial (ii)): The mame, masp, mechR, sponR and eNa did not present reference values outliers in the validation set (sample with data in validation set that was out-of-range in calibrations set). Outliers in other indices were as follows: 2 (mime), 3 (misp), 2 (unme), 4 (unsp), and 6 (WDC). Removal of the outliers degraded performance (-ve change in R^2) of the looCV models across IR-

methods (except for mime using NIR**)**. Effectively, removal of the outliers had a profound (positive) effect on mime NIR model. The removal had significant (negative) effect on models for misp (NIR, NIRwc), unme (MIR, MIRwc and NIR), and unsp (MIR, NIR). The removal had no impact on mime (MIRwc), misp (MIR, MIRwc), unme and unsp (NIRwc) and WDC (Table 4.14).

Table 4.14: Effect of removal of reference values outliers on independent estimation of WSA indices.

	Stable micro aggregate		unstable fraction		
IR-method	**mime**	**misp**	**unme**	**unsp**	**WDC**
MIR	N/A	(o)	(-)	(-)	(o)
MIRwc	(o)	(o)	(-)	(o)	(o)
NIR	(++)	(-)	(-)	(-)	(o)
NIRwc	N/A	(-)	(o)	(o)	(o)

Note: (*R^2 change (%) ± 0 – 10 was assigned zero (o) and designated "no effect"; ± 10 – 50 % was assigned (-/+) and designated "significant effect", and;* ± > 50 was *assigned (++/--) and designated "profound effect" comparing trial (i) with trial (ii); N/A (not applicable) for looCV model with R^2 < 0.3).*

Spectral outliers (optimization trial (iii)): Identified Robust Mahalanobis distance (*H*) (*H* > 2) outliers were: 7 (MIR), 4 (MIRwc), 2 (NIR), and 1 (NIRwc). Effectively, removal of MIR outlier samples had a positive impact on unsp and eNa, no impact on mame, masp misp, unme, mechR and sponR, and negatively impacted on WDC (Table 4.15). The MIRwc outliers had no effect on models for mame, masp, misp, unme, unsp, mechR, sponR and WDC, but had negative effect on mime and eNa. The NIR outliers had positive influence on estimation of misp, sponR and eNa and profound influence on mime. The removal had no effect on mame, unme, unsp and WDC, but negatively affected masp and mechR. Removal of NIRwc outlier sample had a positive impact on WDC and no effect on mame, masp, misp, unme, unsp, mechR, sponR (Table 4.15).

Table 4.15: Effect of removal of spectral outliers on independent estimation of WSA indices.

	Stable macro aggregate		stable micro aggregate		ratio macro:micro		unstable fraction			
IR-method	**mame**	**masp**	**mime**	**misp**	**mechR**	**sponR**	**unme**	**unsp**	**eNa**	**WDC**
MIR	(o)	(o)	N/A	(o)	(o)	(o)	(o)	(+)	(+)	(-)
MIRwc	(o)	(o)	(-)	(o)	(o)	(o)	(o)	(o)	(-)	(o)
NIR	(o)	(-)	(++)	(+)	(-)	(+)	(o)	(o)	(+)	(o)
NIRwc	(o)	(o)	N/A	(o)	(o)	(o)	(o)	(o)	N/A	(+)

Note: (*R^2 change (%) ± 0 – 10 was assigned zero (o) and designated "no effect"; ± 10 – 50 % was assigned (-/+) and designated "significant effect", and;* ± > 50 was *assigned (++/--) and designated "profound effect" comparing trial (i) with trial (iii); N/A (not applicable) for looCV model with R^2 < 0.3).*

Notable, prediction of WSA indices was more sensitive to NIR outliers than to other IR-methods. Removal had, for example, profound positive effect on prediction of mime, and significant effect on prediction of masp, misp, mechR, mechR and unme (Table 4.15). The models were less sensitive to MIRwc and NIRwc outliers, ascribed to smoothing effect of spectra wavelet transform that effectively bypasses outliers (Viscarra Rossel & Lark, 2009). The model for eNa was sensitive to NIR, MIR, and MIRwc outliers (Table 4.15), characteristic of surrogate calibrations (Reeves, 2010). Wavelet coefficients were treated as non-spectral data in PLS modeling in The Unscrambler and were probably not appropriate for spectral outlier detection.

As might be expected, the same outlier samples identified for MIR were obtained for MIRwc. Similarly the same outliers for NIR were identified for NIRwc. Outliers for the wavelets were fewer than those for the corresponding spectra (7 MIR vs 4 MIRwc and 2 NIR vs 1 NIRwc), since wavelet transform effects noise removal (Viscarra Rossel &

Lark, 2009). Also, identified outlier samples for MIR/ MIRwc were different from those for NIR/ NIRwc, since the effective spectral ranges are mutually exclusive. This suggested complementarity in NIR and MIR. The presentation of both NIR and MIR models has, therefore, the potential advantage of giving different dimensions of soil behavior, and probably, the term "alternative IR-methods" could be more appropriately substituted with ' comparable IR-methods'.

Reference value and spectral outliers (optimization trial (iv)): Performance degraded with sequential removal of the outliers. The exception was models for mime (NIR), unsp (MIR) and WDC (NIRwc) that indicated improvement. Noteworthy, removal of the outliers downgraded MIR model for unme but enhanced model for unsp by a similar margin (16 - 20 %). Models for mime (MIRwc), misp, unme, and unsp indicated further degradation compared with removal of only spectral outliers. The exception was WDC (MIR) that indicated more than 80 % performance improvement. Notably, there were no samples that were both reference values and spectral outliers for mime, misp, unme, and unsp, whereas, five samples (of the 7 MIR and 6 WDC outliers) were MIR and WDC outliers. Prediction for a majority of WSA indices got poorer compared to when only either spectral or reference value or none were removed.

Effectively, removal of both reference and spectral outliers had either no effect or negatively affected prediction of WSA indices with exception of mime (NIR), unsp (MIR) and WDC (NIRwc) (Table 4.16).

Table 4.16: Effect of removal of reference value and spectra outliers on independent estimation of WSA indices.

	Stable micro aggregate		unstable fraction		
IR-method	**mime**	**misp**	**unme**	**unsp**	**WDC**
MIR	N/A	(o)	(-)	(+)	(o)
MIRwc	(-)	(-)	(-)	(-)	(o)
NIR	(++)	(o)	(-)	(-)	(o)
NIRwc	N/A	(-)	(o)	(o)	(+)

Note: (*R^2 change (%) ± 0 – 10 was assigned zero (o) and designated "no effect"; ± 10 – 50 % was assigned (-/+) and designated "significant effect", and;* ± > 50 was *assigned (++/--) and designated "profound effect" comparing trial (i) with trial (iv); N/A (not applicable) for looCV model with R^2 < 0.3).*

4.2.3 PLS and IR performance and reliability for estimation of WSA indices

Maximal and optimal validation models: Validation without removal of outliers (Trial (i)), and with removal of spectral outliers (Trial (iii)) presented maximal[8] models for estimation of WSA indices across IR-methods. The maximal trial was also the optimal model[9] across IR-methods for majority of WSA indices (data not shown). Trial (i) was optimal for estimation of WSA indices. Exceptions were mime, misp and sponR (NIR) and unsp (MIR) where prediction with removal of spectral outliers (trial (iii)) was optimal. Trial (iii) was optimal for eNa (MIR and NIR) and WDC (NIRwc).

Models for majority of WSA indices for MIR and MIRwc were optimal following trial (i), whereas NIR and NIRwc indicated more indices optimized following trial (iii), suggesting practical advantage of MIR/MIRwc over NIR/NIRwc. Noteworthy, where Trial (i) had an edge over Trial (iii), the difference was not significant, suggesting a

[8] *Maximal model defined the trial with highest R^2 for estimation of WSA indices across trials (i), (ii), (iii), and (iv)) for each IR- method*.

[9] *Optimal validation trial was the maximal trial with ≥ 10.0 % change in R^2 (Rule of Thumb) relative to Trial (i), else Trial (i) was defined optimal for each IR- method.*

lurking influence of spectral outliers. However, where Trial (iii) maximized validation, there were large differences with Trial (i) particularly for NIR and sometimes NIRwc models (data not shown). This is reflection of somewhat weakness (lower stability) of NIR models, a consequence of higher sensitivity to (spectral) outliers.

Overall validation without removal of outliers (Trial (i)) presented the best validation[10] for all WSA indices except mime (Table 4.17).

Table 4.17: Validation statistics for the best model and associated IR-method for independent estimation of WSA indices.

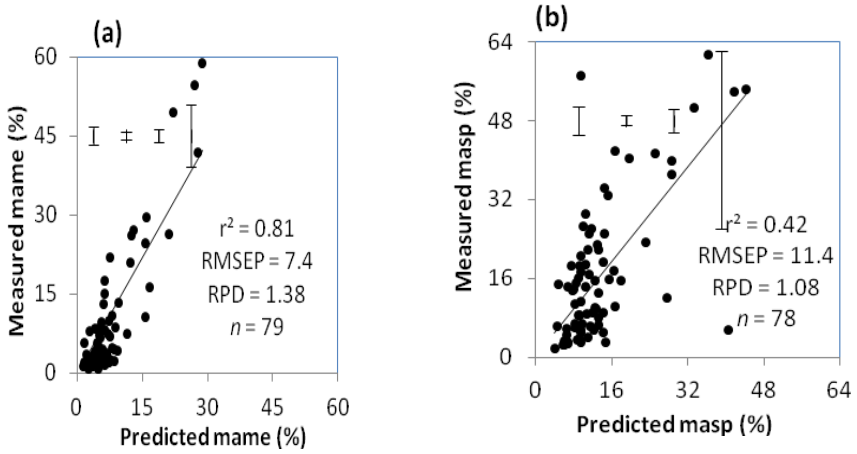
Soil test	IR-method	R^2	RMSEP	RPD	*n*	optimization trial
mame	NIR	0.81	7.40	1.38	79	(i)
masp	NIR	0.42	11.44	1.08	78	(i)
mime	NIR	0.30	12.80	0.71	77	(iii)
misp	NIRwc	0.34	14.65	0.76	78	(i)
unme	NIRwc	0.65	12.16	1.11	79	(i)
unsp	NIRwc	0.48	16.04	0.93	79	(i)
mechR	MIRwc	0.64	0.36	1.42	79	(i)
sponR	NIRwc	0.32	0.51	1.12	78	(i)
eNa	MIRwc	0.46	14.79	0.85	79	(i)
WDC	MIRwc	0.71	8.25	1.47	78	(i)

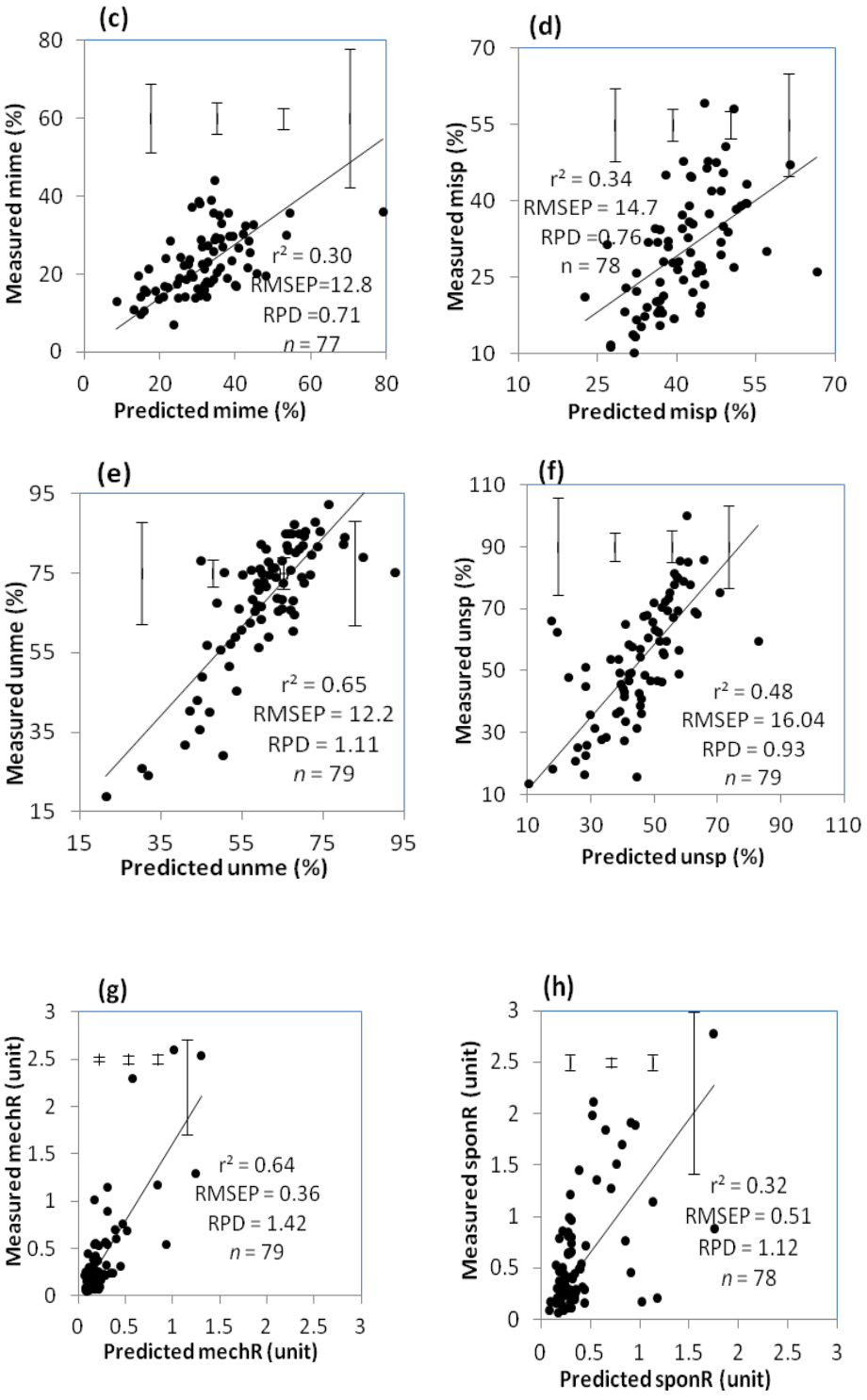
Note: *(trial (i), prediction without removal of outliers in the validation set; (iii), prediction with removal of spectral outliers in the validation set; n, validation samples).*

Figure 4.11 presents comparison of measured vs predicted values of WSA indices for the best validation and corresponding IR-based method (indicated also for each plot is 1:1 regression target line and error bars for quartile data). Excellent performance was attained for independent estimation of mame (R^2 = 0.8, RPD = 1.4; Figure 4.11 (a)), whereas estimation of masp was weak (R^2 = 0.4, RPD = 1.1; Figure 4.11 (b)). Stable

[10] *Optimal trial with the highest R^2 across the IR-methods was designated the best trial.*

micro aggregates (mime and misp) could not be reproduced in the validation set (R^2 = 0.3, RPD = 0.7), illustrated by the wide data scatter (Figure 4.11 (c) and (d)). Estimation of unstable fraction (unme and unsp) was satisfactory to moderate (R^2 = 0.7 - 0.5, RPD = 1.1 - 0.9; Figure 4.11 (e) and (f)). Prediction of mechR was satisfactory (R^2 = 0.64, RPD = 1.4) with data distribution similar to mame (Figure 4.11 (g)), whereas estimation of sponR was poor (R^2 = 0.3, RPD = 1.1; Figure 4.11 (h)). Performance for eNa was weak (R^2 = 0.5, RPD = 0.9; Figure 4.11 (i)), especially prediction error (RMSEP = 14.8 cmol (+) kg^{-1}) was very high (Table 4.17). Estimation of WDC was satisfactory (R^2 = 0.7, RPD = 1.5; Figure 4.11 (j)). Model for WDC was more robust than eNa (Table 4.17); like tClay, WDC is spectrally responsive, whereas calibration for eNa is surrogate. Notably estimation of WDC using MIRwc was much better than R^2 = 0.3 and RPD = 1.2 using MIR reported by Canasveras *et al*. (2010) using the less rigorous cross validation. Similarly, estimation of eNa was better than R^2 = 0.39 (Viscarra Rossel *et al*., 2008) and R^2 = 0.2 (Pirie *et al*., 2005) using MIR following separate test set validation strategy. This affirmed improved (and more robust) performance with MIRwc for the soil-based predictors (WDC and eNa).




(c)
Measured mime (%)
Predicted mime (%)
r² = 0.30
RMSEP=12.8
RPD =0.71
n = 77
(d)
Measured misp (%)
Predicted misp (%)
r² = 0.34
RMSEP = 14.7
RPD =0.76
n = 78
(e)
Measured unme (%)
Predicted unme (%)
r² = 0.65
RMSEP = 12.2
RPD = 1.11
n = 79
(f)
Measured unsp (%)
Predicted unsp (%)
r² = 0.48
RMSEP = 16.04
RPD = 0.93
n = 79
(g)
Measured mechR (unit)
Predicted mechR (unit)
r² = 0.64
RMSEP = 0.36
RPD = 1.42
n = 79
(h)
Measured sponR (unit)
Predicted sponR (unit)
r² = 0.32
RMSEP = 0.51
RPD = 1.12
n = 78

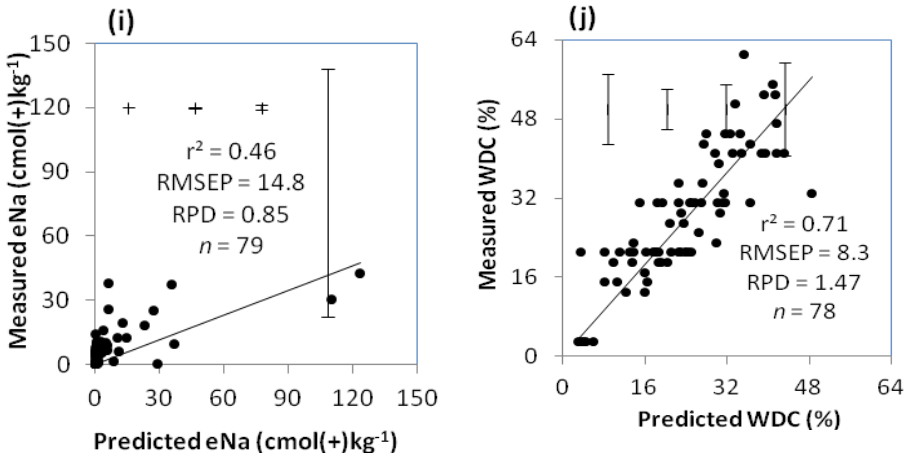


Figure 4.11: Scatterplot plots of measured vs predicted values for the best model and associated IR-method for independent estimation of WSA indices ((a) mame using NIR, (b) masp using NIR, (c) mime using NIR, (d) misp using NIRwc, (e) unme using NIRwc, (f) unsp NIRwc, (g) mechR using MIRwc, (h) sponR using NIRwc, (i) eNa using MIRwc, and (j) WDC using MIRwc).

Influence of wet-sieving pretreatment and IR-method: Models for indices from slaking plus mechanical disruption wet sieving pretreatment were superior over those from slaking only pretreatment (except mime) (Table 4.17). This was attributed to lower quality reference data for the slaking only-based indices. Performance followed the general order: mame > unme ~ mechR > unsp > masp > misp ~ sponR ~ mime. The performance has implications for utility of the indices for diagnostic screening and prevalence assessment of stability related problems in LVB. The poor performing sponR was strongly associated with mechR (r = 0.8) and with masp (r = 0.92). There was no clear trend on performance of alternative IR-methods. However, NIR and NIRwc were associated with the best validation for WSA indices (except mechR). The NIR was associated with indices from direct measurements; whereas NIRwc was associated with derived indices. The MIR was not associated with the best estimation of the indices (Table 4.17). The NIR (< 2 mm) is much more affected by the surface physical

properties (such as size and shape of soil aggregates) because of their influence on the light scattering and light paths lengths (Chang *et al*., 2001*)*, whereas MIR (< 0.1 mm) present more homogeneous/ smooth surface with higher reflectance (including specular reflectance) and consequently lower relative absorbance. Canasveras *et al*. (2010) attributed weaker MIR performance relative to NIR to loss of microstructure information by grinding. Presumably, NIR models were also more enhanced by the effect of water-light interactions (Stenberg, 2010) since very strong mc absorption bands are found in the NIR region (mainly around 1400, 1900 and 2200 nm). Remarkably, replacing NIR with NIRwc resulted in degradation of looCV models for estimation of all WSA indices (Tables 4.8 and 4.13)*,* whereas NIRwc presented best models for independent estimation of misp, unme, unsp, and sponR (Table 4.17). This affirmed the fact that robusticity (stability and reliability) of PLS IR-based analyses could only be ensured with independent testing (Brown *et al*., 2005).

Reliability of PLS analyses for estimation of WSA indices: Calibration spectrometry (model fit (R^2), prediction error (RMSE) and associated model reliability (RPD) is critical for assessing utility of PLS (Bellon-Maurel & McBratney, 2011), for also 'fit-for-purpose' soil analyses (Sorensen & Dalsgaard, 2005).

Model fit and reliability: PLS looCV calibration (reflected by coefficient of determination, R^2) could be reproduced in independent set with similar or enhanced level of efficiency for WSA indices (except sponR and eNa). This affirmed robusticity of WSA indices models, attributed to the fact that soil particle-size and soil structure exhibits primary (albeit weak) response to reflectance spectra (Chang *et al*., 2001). Observed lower performance for eNa affirmed inherent challenges with geographic transferability of surrogate calibration (Reeves, 2010) (Tables 4.13 and 4.17**)**.

Larger RPD values indicate better fitting models (Bellon-Maurel & McBratney, 2011). Based on established RPD categories (> 2.0, 2.0-1.4, and < 1.4 for high, moderate, and low reliability of PLS and IR models, respectively, moderately reliable looCV predictions were obtained for WDC (RPD = 1.8), mame and sponR (RPD = 1.6), masp, unme, and unsp (RPD = 1.5). Models for mime, misp, mechR and eNa were unreliable (RPD < 1.4). Importantly, moderately reliable independent prediction was achieved for WDC (RPD = 1.5) similar to Canasveras *et al*. (2010) following less rigorous cross validation. Also models for mame and mechR (RPD = 1.4) could be improved by improving quality of reference data (Stenberg *et al.*, 2010).

Reliability for 'fit-for-purpose' analyses: Observed acceptable looCV prediction of stable macro aggregate (mame and masp) and unstable fraction (unme and unsp) (R^2 = 0.6, RPD = 1.7-1.6), and for soil-based predictors eNa (R^2 = 0.7, RPD = 1.2), and WDC (R^2 = 0.7, RPD = 1.9), suggested that IR-based calibrations using air-dried (< 2 and < 0.5 mm) soil samples are feasible in principle (Reeves, 2010), for rapid characterization of the indices. Independent testing suggested strong potential for estimation of mame and WDC (R^2 = 0.8 - 0.7) and moderate potential for unme, unsp, mechR, and eNa ($R^2 \geq$ 0.5), however, reliability of the models was moderate to poor (RPD = 1.5 - 0.9). Estimation of sponR and stable micro aggregates (mime and misp) was unacceptable (R^2 = 0.3, RPD = 1.1 - 0.7). Importantly, models for WDC, mame and mechR could be improved (RPD ≥ 1.4) by improving quality of reference values (Chang *et al*., 2001). Table 4.18 presents a statistical description (mean, SD and data range) of the observed data of WSA indices analyzed using conventional methods and their independent predictions.

Table 4.18: Statistical description of the observed WSA indices analyzed using conventional methods of analyses and their independent predictions using IR-methods.

	Observed		predicted	
Soil test	**mean ± SD**	**data range**	**mean ± SD**	**data range**
mame	9.0 ± 12.0	0.6 - 58.7	7.4 ± 5.8	1.5 - 28.8
masp	16.8 ± 14.4	1.9 - 61.3	13.5 ± 8.5	4.0 - 44.3
mime	23.0 ± 8.2	7.0 - 44.1	31.8 ± 10.9	8.7 - 79.1
misp	30.2 ± 11.3	10.1 - 59.3	41.5 ± 8.0	22.7 - 66.6
unme	68.0 ± 16.4	18.6 - 92.3	60.7 ± 11.8	21.5 - 92.7
unsp	53.5 ± 19.3	13.5 - 100.0	45.4 ± 13.1	10.5 - 82.9
mechR	0.4 ± 0.5	0.0 - 2.6	0.2 ± 0.2	0.1 - 1.3
sponR	0.6 ± 0.6	0.1 - 2.8	0.4 ± 0.3	0.1 - 1.8
eNa	6.4 ± 9.2	0.0 - 42.4	6.5 ± 19.5	0.0 - 123.5
WDC	28.2 ± 12.9	3.0 - 61.0	23.6 ± 10.8	2.9 - 48.6

There was general correspondence of observed and predicted values for WSA indices and WDC. Even for poor calibrations (for instance misp, Table 4.17), the predictions are in the same order of magnitude as observed values (Table 4.18). Notably, the IR-methods indicated higher precision (lower SD) (except for mime and eNa) than conventional methods. However, the PLS analyses overestimated lower values of mame, masp, and misp (see also error bars Figure 4.11 (a), (b), and (d), respectively), and higher values of mime and eNa (Table 4.18; Figure 4.11 (c) and (i), respectively). The PLS analyses underestimated higher values of mame, masp, mechR, sponR and WDC (Table 4.18; Figure 4.11 (a), (b), (g), (h), and (j), respectively).

The quality of the quantitative independent models for mame, masp, unme, unsp, mechR, WDC, and eNa might be valuable for providing indices for sorting soils into broad stability classes. These findings support the few studies available in the literature on performance of IR-methods for prediction of soil aggregation indices. Canasveras *et al*. (2010) found, for example, R^2 = 0.60 - 0.67 and RPD = 1.7-1.5 for prediction of

aggregation indices (water stable aggregate, mean-weight diameter, and water dispersible clay), and concluded that the spectral predictors are valuable for sorting soils into stability categories. Madari *et al*. (2006) reported R^2 values 0.66 and 0.67 for NIR PLS prediction of soil mean-weight diameter and mean geometric diameter, respectively, and concluded that the performance provide a reasonable estimate of soil aggregation properties, and valuable information on soil structure, especially since the standard way of determination of these indices is the result of a long and laborious laboratory procedure that itself holds various opportunities for error inclusion.

The relatively high speed and lower cost of IR-methods combined with PLS calibration, and; the large local variation of soil properties suggest that the ability to analyze large number of samples involving multiple variables at finer sampling intervals using the spectroscopic technique may in some circumstances outweigh the loss in analytical precision. The IR-methods are known additionally to be more reproducible (precise) than reference methods (Shepherd *et al*., 2005) and this could convey a distinct advantage.

4.2.4 Factors influencing PLS performance for estimation of WSA indices

Overall, there was observed lack of impact of outlier removal where validation Trial (i) was adjudged optimal validation for WSA indices (except mime) (Table 4.17). This suggested that the rigors in optimization (Trials (ii), (iii) and (iv)) might not be worthwhile for studies involving large sample sizes and multiple variables. However, interpretation of performance of PLS and IR-methods has to consider a number of factors. These include outlier definition and choice of thresh hold values, calibration spectrometry (R^2, RMSEP, RPD), correspondence of IR-methods, and appropriateness of selected WSA indices including data quality.

Soil analyte characteristics (quality, distribution): The poorer predictive performance for WSA indices at high and/or extreme (quartile) values (Figure 4.11) was partly attributed to low quality (skewed distribution) reference values (Table 4.3), rather than genuine lack of predictive power of IR-methods. At very high concentration of eNa (Figure 4.11 (i)), for example, there is greater variability in amounts of ions extracted (therefore higher analytical error) resulting in weak correlation of reference values with spectra or soil chromophores as suggested by Shepherd and Walsh (2002). Other possible sources of error include changes in soil properties between the times of analytical and spectral measurement and variation among subsamples used for analytical and spectral measurements (Shepherd & Walsh, 2002). Notably, validation samples and datasets were obtained one year after calibration development. Reference data range thresh holding and development of separate models for the different (more linearly distributed) segments, is a strategy commonly applied for improved calibration performance (Stenberg, 2010). This was not feasible in this study due to the small validation quartile sample size ($n < 30$).

Outlier definition and (choice of) threshold values: The set Rule of Thumb threshold change in R^2 to assess effect of outliers [(0 - 10 %, no effect; 10 - 50 %, significant effect, and; > 50.0 %, profound effect)], might be too broad probably masking the effects. Robust Mahalanobis distance (*H) to* define spectral outlier is not also standardized. The set value ($H > 2$) might be restrictive considering that the higher the H value selected the broader the diversity of validation set. However, Terhoeven-Urselmans *et al*. (2010) found Mahalanobis H > 12 outliers to have no significant influence on independent estimation of fertility valuable soil properties.

Reference values outliers were defined as samples in validations set with data range outside the range of calibration set and were as follows: 2 (mime), 3 (misp), 2 (unme), 4 (unsp), and 6 (WDC). Defining reference values outlier as the sample whose prediction

residual is greater than 3× RMSE (Sorensen & Dalsgaard, 2005), the number of outliers for each WSA indices across the IR-methods was as follows: mame: 2 (MIR, MIRwc), 3 (NIR, NIRwc); masp: 2 (MIR, MIRwc, NIR); mime: 1 (MIRwc, NIR); misp: 1 (MIR); unsp: 1 (NIR, NIRwc); mechR: 2 (MIR), 3 (MIRwc, NIR, NIRwc); sponR: 2 (MIR, MIRwc, NIR), 1 (NIRwc); eNa: 3 (MIR, NIR, NIRwc); WDC: 1 (MIR, MIRwc, NIR, NIRwc). Notably, WSA indices mame, masp, mechR, sponR, and eNa presented also reference values outliers, whereas these indices presented no outliers based on data range between calibration and validation sets. This could have implications on calibration performance and effect of Trial (ii). The assessment could, however, not be accommodated in this study.

Correspondence of model goodness of fit and prediction error: The model goodness of fit (R^2) was used to assess the influence of outliers on calibration validation. Improvement (positive change) in R^2 with validation trial (ii), (iii), and (iv) relative to trial (i) was not always followed, however, by improvement (negative change) in RMSEP. Degraded R^2 was associated with improved RMSEP, for example, for unsp and WDC (trial (ii)) and for mime (MIRwc), mechR (NIR), eNa (MIRwc), and WDC (MIR) (trial (iii)) (data not shown). The conflict of R^2 and RMSEP could challenge practical interpretation of effect of outlier on estimation of the WSA indices. Also, the RMSEP was substantial for some indices (Table 4.17), with negative implications for model robusticity. The moderate R^2 = 0.5 for estimation of eNa was associated, for example, with high prediction error (RMSEP = 14.8 cmol (+) kg^{-1}) and corresponding low reliability (RPD = 0.9). Such performance might be unacceptable since they mask effect of sodicity in stability diagnostic screening, where threshold values for eNa are in the range 1- 2 cmol (+) kg^{-1} (Sumner, 2000). Some authors (Stenberg, 2010) have focused on prediction errors for assessing model performance since RMSE relates more directly with model reliability (RPD), also attested to in this study.

Correlation of WSA indices to IR-based predictor variables: Observed independent estimation of the WSA indices was strong to poor (R^2 = 0.8 - 0.3), especially model reliability was weak to poor (RPD = 1.4 - 0.7). This is not unusual in ecological studies, especially for chemometrics and spectra-based soil analyses, attributed to several factors including: sample characteristics (provenance, composition); soil property characteristics (spectral response, data quality); spectra measurement region; spectral data pretreatment, and; calibration method (Stenberg *et al*., 2010).The complexity in soil composition distorts (through masking for example) resonance of spectra absorbance and reference values of analyte of interest, for example, contrary to high resonance of spectra and organic (plant) constituents, ascribed to the simple organic structural composition that is genetically controlled (Shepherd & Walsh, 2007).

No absorption features have been directly associated with aggregation indices (Canasveras *et al*. 2010). However, robust (albeit weak) models for independent estimation of WSA indices, affirm that soil structure exhibits primary response to spectra as suggested by Chang *et al*. (2001). The role of several mechanistic processes responsible for calibration performance (enhancing or confounding), however, could only be speculative. These include effects of water-light interactions (aquaphotomics) (Stenberg, 2010); effect of particle-size and shape variation on spectra (Chang *et al.*, 2001); dynamics of microstructure and other artifacts (sample sorting) associated with test sample preparation and handling (Canasveras *et al*., 2010), and; the choice of aggregation indices. Further studies should focus on unraveling the mechanisms involved, effective indices, and the more appropriate IR-method. Stenberg *et al.* (2010) noted that rapid uptake of IR for routine soil analyses will depend on demonstrated understanding of the (theoretical) mechanisms involved in calibration of (especially secondary) soil properties.

IR-method sensitivity to outliers and model complexity: The NIR- based models were more sensitive to outliers than MIR- based models. The 2 NIR spectral outliers had significant effect, for example, on performance of NIR models than was the case for 7 MIR outliers on MIR models, probably due to confounding (distortion) in absorption features from the weak combinations and overlapping bands in NIR. Additionally, with removal of spectral outliers, improved R^2 was in most cases accompanied with higher accuracy (reduced RMSEP) for the NIR models.

Models from MIRwc and NIRwc were more robust than corresponding models from raw spectra. Removal of spectral outliers for wavelet-based models, for example, had no effect on estimation of WSA indices (Tables 4.15 and 4.17). The general order of sensitivity to outliers was as follows: NIR > MIR ~ NIRwc/ MIRwc. Wavelet coefficients were entered as 'non-spectral' data in PLS regression modeling in The Unscrambler, suggesting that the coefficients might not be appropriate for subjection to spectral outlier removal.

NIR WSA indices models indicated similar or higher complexity (more PLS PCs) than MIR models (except mime) (Table 4.13). Models from wavelet transform indicated higher parsimony than models from MIR and NIR, a finding reported also by Viscarra Rossel and Lark (2009). The MIRwc and NIRwc models were more robust than those of MIR and NIR for the soil-based predictors (eNa and WDC). The WDC models were more robust than eNa across IR-methods, attributed to surrogate calibration for eNa, whereas WDC like tClay was considered spectrally active.

Effect of wet-sieving pretreatment: Models of WSA indices based on slaking plus mechanical disruption wet sieving pretreatment (mame, mime, unme) were less sensitive to outliers than those from slaking only (misp, unsp) wet sieving pretreatment. Also, indices based on slaking plus mechanical disruption indicated superior models over

indices from slaking only through trials (i) to (iv). This was attributed to higher precision for the slaking plus mechanical shaking pretreatment (Table 4.2). The breakdown of unstable and metastable wet aggregates with shaking (Marquez *et al*., 2004) reduced aggregation to the limiting stability analogous to percolation stability demonstrated by Auerswald (1995). The quality of reference data greatly influences performance of IR-based models (Reeves, 2010).

Noteworthy, NIR/ NIRwc were better calibrated to wet stable micro- aggregates (mime/ misp) and unstable aggregate fractions (unme and unsp) than to macro aggregates (mame/ masp). Also NIR/NIRwc-based models were superior over their MIR/MIRwc counterparts across trials (i) through (iv) for WSA indices from slaking only wet-sieving pretreatment. This suggested that slaking only and NIR/NIRwc method was more appropriate for modeling wet stable micro aggregates and unstable aggregate fractions for the studied soils.

Stable micro aggregate indices from the two wet-sieving pretreatments (mime and misp) were poorly calibrated to alternative IR-methods. No immediate explanation could be advanced since micro aggregates presented higher quality reference data than the better performing macro aggregates (Table 4.2). The poor performance suggested less utility of IR for assessing soil micro aggregation.

4.2.5 Implication for soil health assessment

IR-method and calibration validation requirement: The NIR and NIRwc were the IR-methods associated with the best independent estimation of the WSA indices (Table 4.17). This might be insightful since several workers (Viscarra Rossel *et al*., 2006) found MIR superior over NIR for estimation of several soil properties. The validation strategies in the reported studies (including cross validation, separate test sets) are, however, not as rigorous as independent testing used in this study. Notably, none of the

IR-methods associated with the best independent estimation of the WSA indices presented also the best fitting looCV models (except eNa) (Table 4.19). The strong looCV models could not provide, for example, the optimal (or maximal) models for the WSA indices when subjected to independent testing. This adds to the subjectivity of studies based on cross validation and affirmed the need for independent testing of soil IR-based looCV models (Brown *et al*., 2005).

Table 4.19: Validation statistics and associated IR-methods comparing looCV and independent estimation of WSA indices.

	Cross validation model statistics					Independent model statistics				
Soil test	**IR-method**	**best trial**	**r^2**	**RMSECV**	**RPD**	**IR-method**	**best trial**	**r^2**	**RMSEP**	**RPD**
mame	MIR	(i)	0.58	6.26	1.63	NIR	(i)	0.81	7.4	1.38
masp	MIR	(i)	0.54	7.54	1.64	NIR	(i)	0.42	11.44	1.08
mime	MIRwc	(ii)	0.41	6.9	1.32	NIR	(iii)	0.30	12.80	0.71
misp	NIR	(iii)	0.39	7.97	1.39	NIRwc	(i)	0.34	14.65	0.76
unme	NIR	(ii)	0.65	7.37	1.83	NIRwc	(i)	0.65	12.16	1.11
unsp	NIR	(ii)	0.62	8.28	1.80	NIRwc	(i)	0.48	16.04	0.93
mechR	NIR	(i)	0.41	0.38	1.34	MIRwc	(i)	0.64	0.36	1.42
sponR	NIR	(i)	0.59	0.32	1.78	NIRwc	(i)	0.32	0.51	1.12
eNa	MIRwc	(ii)	0.73	5.15	2.45	MIRwc	(i)	0.46	14.79	0.85
WDC	NIR	(ii)	0.75	5.82	2.08	MIRwc	(i)	0.71	8.25	1.47

Note: (*the same procedure was followed for optimization of looCV models (Table 4.10) and independent (Table 4.17) models; Optimization trial: (i), prediction without removal of outliers; (ii), prediction with removal of reference values outliers; (iii), prediction with removal of spectral outliers).*

Choice/ selected wet stable aggregation indices: Scanty information was available (Canasveras *et al*., 2010) on spectroscopic assessment of measures of aggregation although soil aggregation and its stability is key for soil health (Idowu *et al.*, 2008ab*;* Wei *et al.,* 2006). Specifically no studies were found that further extends cross validation or separate test sets using independent sample datasets.

Superior NIR performance over MIR for independent estimation of aggregation indices from slaking plus mechanical disruption affirms and extends the work of Canasveras *et al.* (2010) using looCV and that of Madari *et al*. (2006) using datasets highly susceptible to pseudo- replication and attendant pseudo-independent validation. Superior NIR models for independent estimation suggested also better geographic transferability, with implications for robust on-site and on-the-go measurements. Notably, NIR-based methods were superior over MIR methods for independent estimation of aggregation indices from wet sieving, whereas, MIR-based methods were superior for soil-based indices eNa and WDC.

Performance for independent estimation of macro aggregates and unstable aggregate fraction in this study has demonstrated potential of IR for rapid assessment of soil physical quality. Also available studies have focused on few aggregation indices (fraction > 250 or < 250μm) and measures of aggregate-size distribution, whereas this study tested 8 indices including efficacy also of IR for estimation of measures of aggregate "instability" and also the potential influence of mild/ spontaneous (slaking only) and abrasive (slaking plus mechanical shaking) aggregate breakdown.

Slaking only and slaking + mechanical disruption of aggregates under controlled conditions mimic the effect on soil stability of low and high intensity rainfall, respectively. Slaking as is common to soils subject to flooding (basin, furrow) irrigation, and slaking plus mechanical disruption common to overhead (sprinkler) irrigation. The two pretreatments represent two different triggers of soil (in) stability and erosion (aggregate break down, detachment and transport). The selected WSA indices provide, therefore, good indicators/ proxies of stability. The mame/ masp indicate, for example, distribution of macro pores for aeration and water movement; mime/misp distribution of water retention pores; unme/ unsp the risk of permeability problems with pore clogging.

The mechR (ratio mame: mime) and sponR (masp: misp) provide good indicator of porosity (and soil moisture balance).

The WSA indices provide measures for wide range/ broad spectrum of stability sensitivity, effective for screening and prevalence assessment of stability related problems. The scatter for stable micro aggregates (mime), for example, is the converse/inverse of unme; the concentration of data points at high levels of unme and low levels of mime, respectively, suggested that the soils were largely unstable (see Figure 4.11 (c) and (e)).

Most appropriate IR-method for screening WSA indices in LVB: The NIR/ NIRwc were associated with the best independent estimation of WSA indices. The difference was not profound however, comparing MIR models from Trial (i) with the statistics for the best validation, especially for commonly deployed stability indices (mame, eNa, WDC) (Table 4.20). This suggested that the rigors in validation optimization including testing alternative IR-methods might not be worthwhile strategy for studies on diagnostic screening of stability related problems involving several hundred samples and multiple variables.

Table 4.20: Validation statistics comparing the best validation model and MIR (trial (i)) models for independent estimation of WSA indices.

	IR-method & optimal trial		validation model			MIR validation with Trial (i)		
Soil test	**IR-method**	**best trial**	**r^2**	**RMSEP**	**RPD**	**r^2**	**RMSEP**	**RPD**
Mame	NIR	(i)	0.8	7.4	1.4	0.7	7.4	1.4
Masp	NIR	(i)	0.4	11.4	1.1	0.3	13.5	0.9
Mime	NIR	(iii)	0.3	12.8	0.7	N/A	N/A	N/A
Misp	NIRwc	(i)	0.3	14.7	0.8	0.1	15.0	0.7
Unme	NIRwc	(i)	0.6	12.2	1.1	0.4	13.6	1.0
Unsp	NIRwc	(i)	0.5	16.0	0.9	0.2	18.3	0.8
mechR	MIRwc	(i)	0.6	0.4	1.4	0.6	0.4	1.4
sponR	NIRwc	(i)	0.3	0.5	1.1	0.2	0.6	1.0
eNa	MIRwc	(i)	0.5	14.8	0.9	0.3	15.7	0.8
WDC	MIRwc	(i)	0.7	8.3	1.5	0.7	9.1	1.3

Observed calibration performance improvement with spectra wavelet transform might not justify, for example, the additional analytical step. The MIR looCV models indicated similar or higher parsimony as NIR models, suggesting high robusticity. MIR was also less sensitive to outliers than MIRwc and NIRwc, whereas NIR was most susceptible. This relative immunity to the presence of outliers suggested MIR to have additional practical advantage over the other IR-methods. Local practicalities (MIR infrastructure, operator skills, soil response) suggested MIR as the more appropriate IR-method for screening stability related problems in LVB of Kenya.

4.3 Evaluation of soil stability related problems in Lower Nyando and Homa Bay sentinel sites

Results are presented for key soil-based predictors of WSA indices and the predicted (score) value (s) of the WSA indices in LNY and HB. The established stability categories for individual WSA indices are followed by prevalence of stability related problems in the sites, within the different clusters and for different depth intervals.

4.3.1 Key soil-based predictors of WSA indices

Twenty eight (28) soil properties presented strong MIR PLS models (looCV $R^2 \geq 0.6$) in sample Set 1 (n = 128) and Set 2 (n = 79) (performance is shown in Appendix 4). The properties were screened (separate run for Set 1 and Set 2) for prediction of each of the (8) WSA indices using CART regression. There were no soil predictors of masp, mechR and sponR ($RE \geq 0.9$). The mame (Set 1) and mime (Set 2) presented also no acceptable predictors (Table 4.21).

Stable macro aggregate: The mame could be predicted by pH in 1:2.5 soil-to-water extract (pH2.5) and soil organic carbon (OC) with 40 % accuracy. The pH2.5 was the main predictor with 100 % importance score while OC contributed about 56.0 %. Exclusion of OC as predictor (model rationalization) elevated total carbon (totC) to primary predictor with similar contribution as OC. This resulted, however, in a slight increase in prediction (relative) error (*RE* 0.60 to 0.68) (Table 4.21).

Stable micro aggregate: The mime could be predicted by water dispersible clay (WDC) with modest 20.0 % accuracy. The misp presented a rather complex model with WDC, plastic limit (PL), and exchangeable sodium (eNa) as the key primary predictors. The WDC was the main predictor (100 % importance score), whereas PL and eNa made a contribution of 41 and 24 %, respectively. The PL split misp data twice at PL 19.3 and 33.8 %. Removal of PL and eNa as predictors presented WDC as the sole predictor of misp but the *RE* increased from 0.50 to 0.61(Table 4.21).

Unstable fraction: The unme could be predicted with 25 % accuracy by eNa in Set 1. The Na+ in 1:2.5 soil-to-water extract (Na2.5), PL and totC could predict unme with 34 % accuracy in Set 2. The Na2.5 split unme data twice at 3.3 and 14.5 mg kg^{-1}. Exclusion of PL and totC presented Na2.5 and pH in 1:2.5 soil-to- 1.0 N KCL solution extract (pHKCL) as primary predictors, however, the prediction error increased (*RE* 0.66 to

0.79). The Na2.5 could predict unsp with 32 % accuracy in Set 2. In Set 1, eNa, WDC and PL could estimate unsp with 37 % accuracy. Exclusion of WDC and PL presented eNa and effective cation exchange capacity (CEC1) as main predictors and with marked improvement in predictive performance (RE 0.63 to 0.59) (Table 4.21).

Table 4.21: Performance of selected soil properties for prediction of WSA indices in two different reference sample sets.

Target	samples set	RE	predictor	split point	Importance score (%)
mame	Set 2	0.599	pH2.5	5.84	100
			OC	2.2	55.7
(-) OC		0.676	pH2.5	5.84	100
			totC	2.39	50.9
mime	Set 1	0.807	WDC	34	100
misp	Set 1	0.499	WDC	32.5	100
			PL	19.3	40.7
			PL	33.8	
			eNa	0.34	24.1
(-) PL & eNa		0.606	WDC	32.5	100
	Set 2	0.636	WDC	26	100
unme	Set 1	0.75	eNa	1.25	100
	Set 2	0.661	Na2.5	3.25	100
			Na2.5	14.5	
			PL	23.6	13.4
			totC	1.9	15.7
(-) PL & totC		0.792	Na2.5	3.25	100
			Na2.5	14.5	
			pHkcl	4.39	13.3
unsp	Set 1	0.632	eNa	0.95	100
			WDC	36	42.4
			PL	23.4	33.8
(-) WDC & PL		0.594	eNa	0.95	100
			CEC1	42.95	36.6
	Set 2	0.684	Na2.5	12.75	100
			Na2.5	61.25	

Note: *(Set 1, n = 128 from across LVB and Set 2, n = 79 representative of LNY and HB sentinel sites in LVB).*

`````

**Rationalized soil predictor models:** The pH2.5 and OC were retained as the main predictors of mame. The WDC was retained sole predictor of mime and misp. The relatively higher *RE* (for misp) with exclusion of PL and eNa could be justified by the easy determination of WDC relative to PL and eNa. The model with Na2.5, PL, and totC as primary predictors (Set 2) was used for prediction of unme since the seemingly less complex model after removal of PL and totC significantly compromised prediction accuracy. The improved model for unsp with eNa and CEC1 was used (Table 4.21). Presumably, WDC and PL were stronger competitors of eNa (Table 4.6) masking importance of CEC1 for prediction of unsp.

From 28 potential predictors, 8 soil properties (pH2.5, OC, WDC, eNa, Na2.5, PL, totC, and CEC1) indicated predictive relationship with one or more of the WSA indices**.** Notably, the list included properties shown to relate well with soil aggregation and stability. Kodešová *et al*. (2009) found, for example, that the strong correlation of WSA and aggregate vulnerability indices (akin to WSA indices from slaking and slaking plus mechanical shaking) were dependent on soil contents of OM, tClay, CEC, and pH. Six soil properties (tSa, tClay, pHw, $CaCO_3$, OM and free Fe) were key predictors of aggregation indices water stable aggregates, mean-weight diameter, and water dispersible clay (Canasveras *et al*., 2010). This suggested that the established 8 properties could provide effective surrogates of the predicted WSA indices. To establish the three (3) more robust soil predictors, MIR PLS looCV models for the 8 properties in Set 1 were independently tested in Set 2.
`````

Key three predictors of WSA indices: The 8 soil properties presented excellent to modest independent MIR models (R^2 = 0.9 - 0.5) (Table 4.22). Three properties (pH2.5, Na2.5 and WDC) had comparative advantage over the others including: prediction efficiency (*RE*) and variable importance score (Table 4.21). The pH2.5, Na2.5 and WDC were selected as the key soil-based predictors of WSA indices mame, mime, misp, unme and unsp.

Table 4.22: Calibration and independent MIR PLS models for selected 8 soil-based predictors of WSA indices.

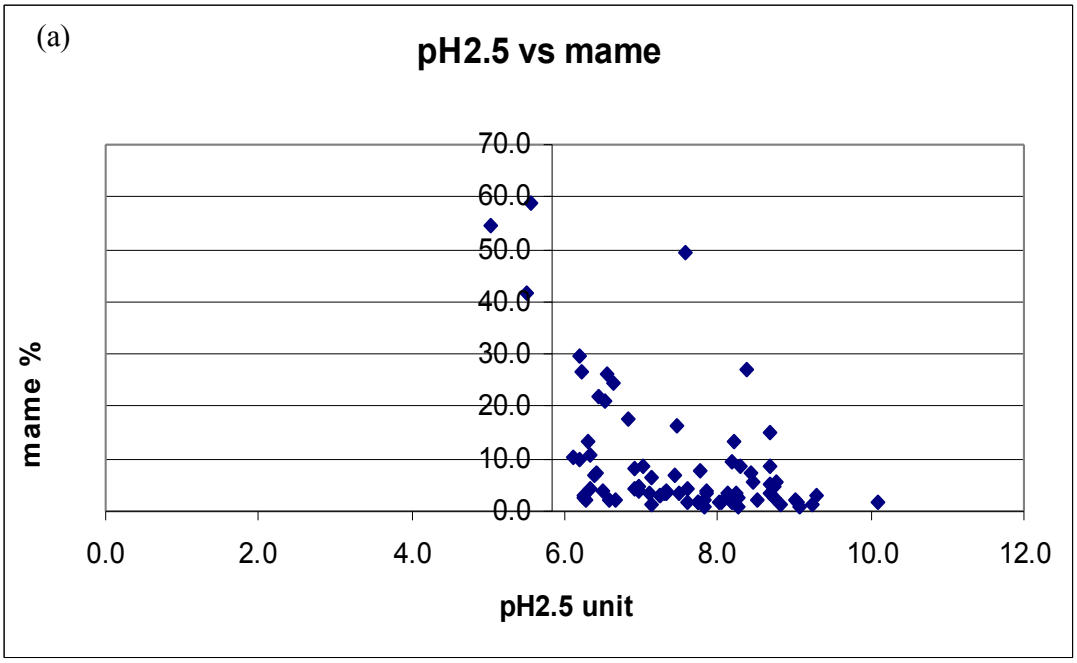
		Calibration with looCV (Set 1)		**independent validation (Set 2)**	
Soil test	**transform**	r^2	**RMSECV**	r^2	**RMSEP**
pH2.5	sqrt	0.67	0.13	0.64	0.13
Na2.5	ln	0.57	1.61	0.52	2.80
totC	ln	0.89	0.24	0.87	0.28
OC	sqrt	0.86	0.12	0.81	0.18
eNa	ln	0.73	1.38	0.53	2.03
CEC1	sqrt	0.86	0.52	0.89	0.63
WDC	sqrt	0.72	0.62	0.65	0.95
PL	ln	0.62	0.18	0.61	0.18

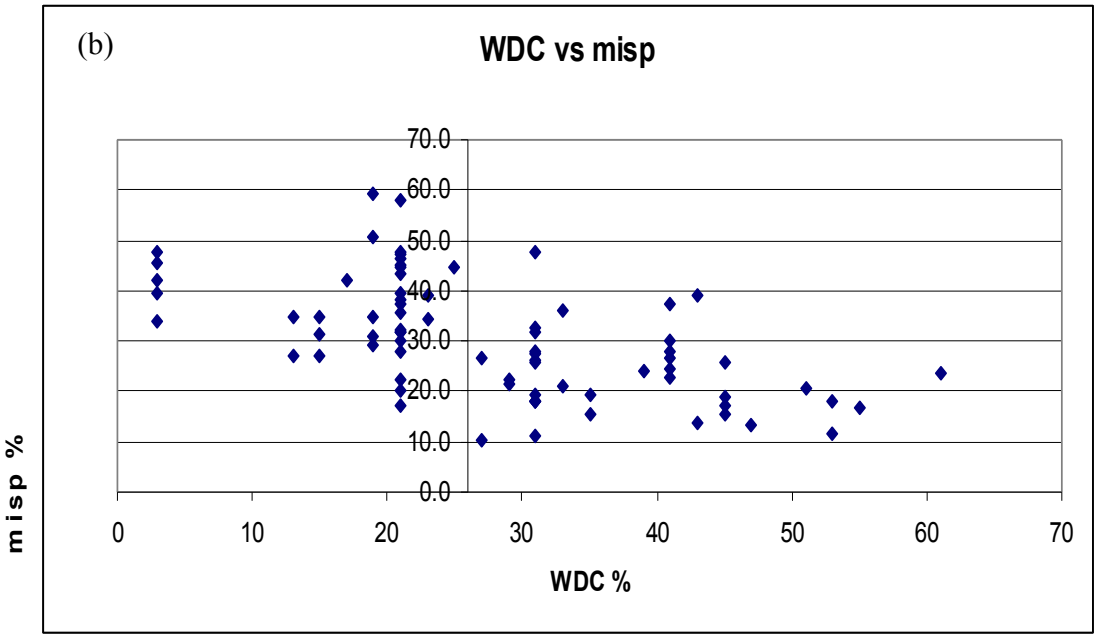
Note: *The RMSECV/ RMSEP are for transformed data*.

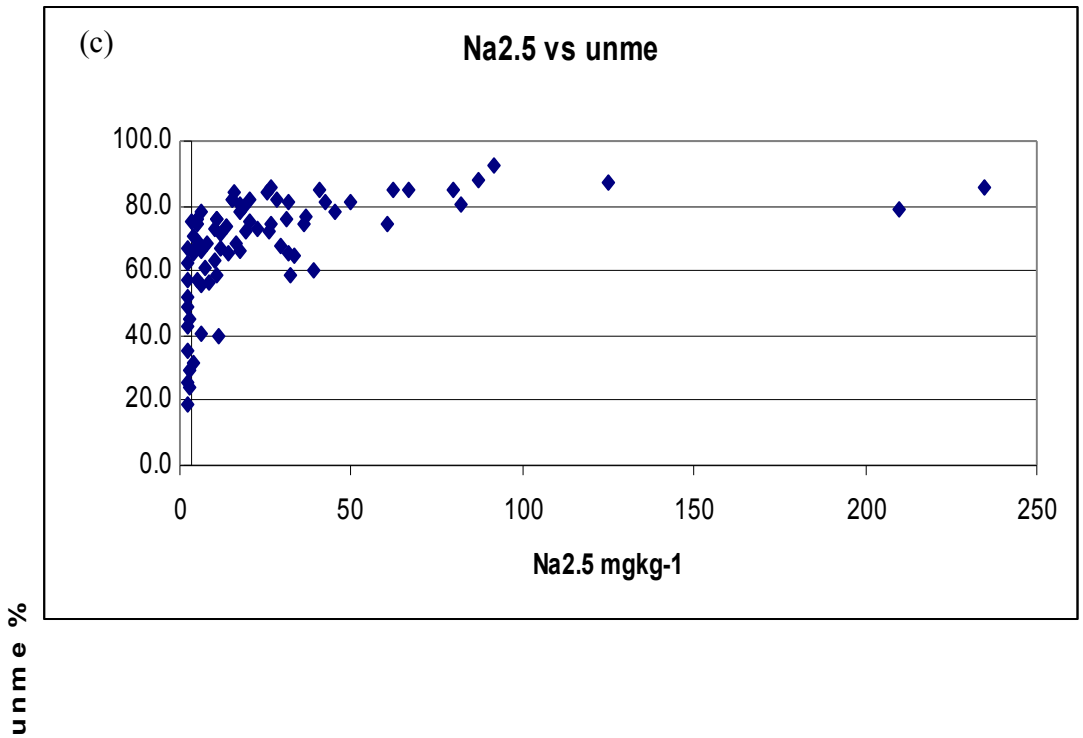
Selection of Na2.5 ahead of eNa and CEC1: The eNa and Na2.5 indicated similar independent prediction (R^2 = 0.52) **(**Table 4.22**)**, and the two were strongly associated (*r* = 0.9). The Na2.5 had comparative advantage and was selected ahead of eNa. The Na2.5 was easier to determine than eNa and could be read from the same soil-water suspension used for pH2.5 making measurements more efficient. The Na2.5 indicated slight superiority over eNa for prediction of unme and unsp (Set 2) (Table 4.21). The two data split points of Na2.5 at 3.3 and 14.5 (mg kg^{-1}) (unme) and 12.8 and 61.3 (mg kg^{-1}) (unsp), suggested potential cutoffs for more effective stability screening (Tittonell *et al*., 2008). The Na2.5 was also a strong competitor and/or surrogate of eNa for prediction of

unme and unsp (Table 4.6). The CEC1 was strongly validated (Table 4.22) and has great residual value, however, routine laboratory determination was considered too rigorous. Figure 4.12 illustrates the predictive relationships of data for soil predictor variables pH2.5, WDC and Na2.5 against predicted WSA indices mame, misp, and unme.

Relationship of WSA indices and predictor variables: The moderate to weak predictive relationships of WSA indices and the three soil properties pH2.5 water, Na2.5, and WDC (*RE* = 0.50 - 0.75) (Table 4.21) were largely non-linear (Figure 4.12).




(b)
WDC vs misp
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
0
10
20
30
40
50
60
70
misp %
WDC %


(c)
Na2.5 vs unme
100.0
80.0
60.0
40.0
20.0
0.0
0
50
100
150
200
250
unme %
Na2.5 mgkg-1

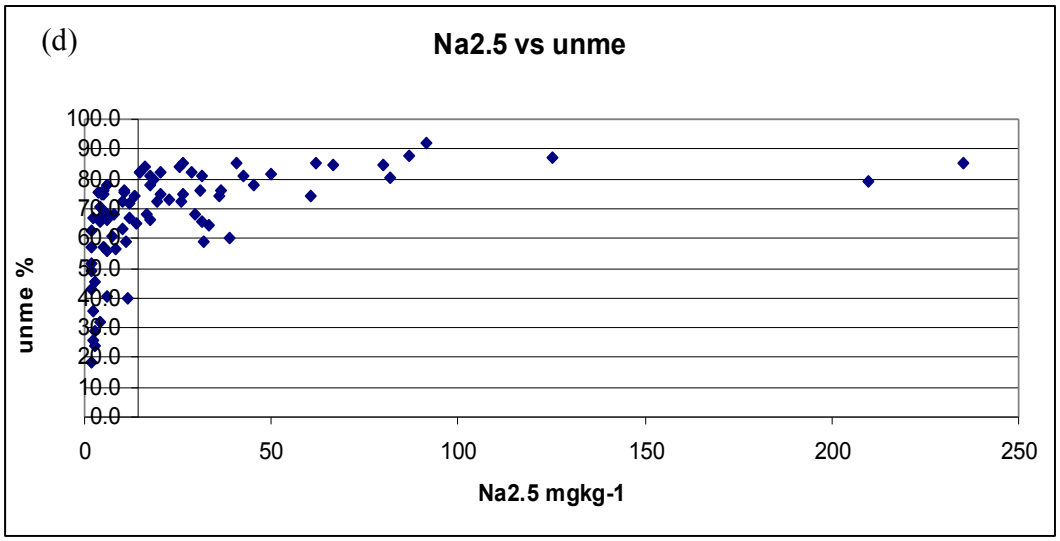


Figure 4.12: Scatterplot plots for predictive relationship of soil-based predictors and predicted WSA indices ((a) pH2.5 vs mame, (b) WDC vs misp, and (c) Na2.5 vs unme (for Na2.5 split at 3.3 mg kg^{-1}), and (d) Na2.5 vs unme (for Na2.5 split at14.5 mg kg^{-1})); the illustrations are based on sample Set 2, n = 79).

Split interpretation for macro and micro- aggregate fraction: the combination of pH2.5 and OC could predict mame with accuracy of 40 % (Table 4.21). At a split point of 5.84 (vertical line Figure 4.12 (a)), the pH2.5 could separate lower than average values from higher than average values of mame. The pH2.5 values ≤ 5.84 were associated with higher than average values of mame (relatively stable soils). High values of stable macro aggregates are commonly associated with soils of low pH (Westerhof *et al*., 1999; Auerswald, 1995). Notably, there were many samples with higher than average mame values for pH > 5.84 (Figure 4.12 (a)), a reflection of substantial unexplained variance (residual heterogeneity). The WDC could predict misp with 36 % efficiency (Table 4.21). The WDC dichotomized misp data at WDC 26 %. Higher than average values of misp had WDC values lower than the split point (Figure 4.12 (b)), suggesting that lower than split point values of WDC are associated with more samples with higher than mean values of stable micro-aggregates. This is in agreement with

Boix-Fayos *et al*. (2001) who found that aggregates < 105µm were positively correlated to clay content, and; water stability of micro-aggregates was positively correlated with clay content. The observation on WDC appeared contrary to interpretation for total clay where, soils with < 25 % clay were considered of inherently lower stability and those of > 35 % clay of inherently higher aggregate stability (Levy *et al*., 2003).

Split interpretation for unstable fraction: Values of unme higher than about 40 % were associated with values of Na2.5 higher than the first split point at 3.3 mg kg^{-1} (Figure 4.12 (c)). Values of unme higher than about 60 % were associated with values of Na2.5 higher than the second split point at 14.5 mg kg^{-1} (Figure 4.12 (d)). The splits indicated that soils with values of Na2.5 higher than the split point were relatively less stable than those with lower values (as expected) and also that there existed two identifiable levels of stability.

Majority of the studied soils were unstable, demonstrated by the occurrence of most of the soils at low levels of wet stable macro- and micro aggregate fractions (mame and misp), higher levels of unstable fractions (unme and unsp), and edge cut splits (Figure 4.12).

4.3.2 Predicted WSA indices in LNY and HB sentinel sites

Predicted key soil predictors in LNY and HB: predicted pH2.5, Na2.5 and WDC in LNY and HB (Data File) is summarized in Table 4.23.

Table 4.23: Predicted key three soil properties in LNY and HB sentinel sites.

Property	min	25th	50th	75th	max	SD
pH2.5	4.1	6.6	7.4	8.1	10.2	1.1
	(5.0)	(6.6)	(7.4)	(8.2)	(10.1)	(1.0)
Na2.5	0.4	5.7	14.0	32.5	755.2	49.8
	(2.0)	(5.0)	(14.5)	(30.6)	(235.0)	(38.3)
WDC	1.1	23.2	28.9	35.5	50.0	9.3
	(3.0)	(21.0)	(27.0)	(41.0)	(69.0)	(14.0)

Note: (*Shown are predicted values for samples in the larger set (n = 339)*; c*orresponding observed values for initially selected set 2 (n = 120) representative of the sentinel sites (in parenthesis) affirm that both sets belong to the same soil population).*

WSA indices grove files: Calibration models (CART grove files) for WSA indices mame, misp, unme, and unsp using Set 2 (n = 79) are summarized in Table 4.24. The mime had no predictive relationship with the set of three soil predictors (RE = 0.869). The unme could be predicted by Na2.5 and pH2.5 with accuracy of about 45.0 %, and each of the predictors split unme data twice. Removal of pH2.5 as predictor resulted in a slight improvement in accuracy. The Na2.5 could predict unsp with 38 % accuracy.

Scoring (prediction) of WSA indices in LNY and HB: The Data File and grove files were used to predict (score) WSA indices mame, misp, unme and unsp for each of the samples in LNY and HB. The main predictor variable in the grove file determined the splitting (mean value (s) for the target. When the Data File was 'dropped down' the tree model (Grove File) of mame, for example, it was pH2.5 splitting rule (Table 4.24) that was used for scoring the data. Similarly, WDC and Na2.5 splitting rules were used for

splitting misp and unme/ unsp data, respectively. The score (predicted) WSA indices value(s) corresponded with the number of terminal nodes (TN) in the grove file (Steinberg & Golovnya, 2006).

Table 4.24: Calibration models (CART grove files) for WSA indices using soil predictor set.

Target	RE	terminal nodes	predictor	split point	importance score
mame	0.749	2	pH2.5	5.84	100
misp	0.624	2	WDC	26	100
unme	0.552	5	Na2.5	3.25	100
			Na2.5	14.5	
			pH2.5	5.84	23.2
			pH2.5	6.64	
(-) pH2.5	0.549	3	Na2.5	3.25	100
			Na2.5	14.5	
unsp	0.615	2	Na2.5	12.75	100

Note: *(pH2.5, Na2.5, and WDC was the predictor set; Set 2 (n = 79) was used for developing grove files; split point: pH (unit), WDC (%), Na2.5 (mg kg^{-1})).*

Grove files of mame, misp, and unsp had two terminal nodes (Table 4.25). The corresponding prediction (score output) for mame was 7.3 % (TN 1) and 51.7 % (TN 2), misp 23.3 % (TN 1) and 37.5 % (TN 2), unsp 39.1 % (TN 1) and 66.3 (TN 2). The unme grove file had five terminal nodes and corresponding five score (split) values were 22.8 (TN 1), 48.8 (TN 2), 55.2 (TN 3), 69.7 (TN 4), and 77.8 (TN 8). With exclusion of pH2.5 as predictor, unme grove file presented 3 terminal nodes and corresponding 3 score (split) values as 42.3 % (TN 1), 64.9 (TN 2) and 77.8 (TN 3) (Table 4.25).

Table 4.25: Predicted (score) of WSA indices and stability class in LNY and HB sentinel sites.

WSA indices	grove TN	score (%)	stability class	score (%)	stability class	score (%)	stability class
mame	2	7.3	low (1)	51.7	high (0)	N/A	N/A
misp	2	23.3	low (1)	37.5	high (0)	N/A	N/A
unme	3	42.3	high (0)	64.5	mod	77.8	low (1)
unsp	2	39.1	high (0)	66.3	low (1)	N/A	N/A

Note: (*TN, terminal node; LNY, lower Nyando; HB, Homa Bay).*

Assigning WSA indices score values to stability classes: The low amounts of mame (limit at 7.3 %) were used to define low stability and the higher score (51.7 %) to define high stability. Similarly low amounts of misp (23.3 %) defined low stability and the higher score (37.5 %) high stability. There was no available literature providing stability thresh hold limits, however, the assignment was logical since the higher the amounts of wet stable aggregates from wet-sieving is an indication of high stability. The higher the amounts of water dispersed aggregates (100 – (stable macro + stable micro)) is an indication of low stability of the soil. For this reason, the low amounts of unme and unsp (22.8 and 39.1 %) were used to define relatively stable soils, while the highest amounts (77.8 and 66.1 %) were used to define low stability soils. Tittonell *et al*. (2008) alluded to efficacy of imputing split values from CART regression screening to effective categories in soil studies. In view of the five score values for unme data, a moderate category (55.2 %) was defined together with two sub-categories, slightly high and slightly low to define transitional categories towards low and high stability soils. These subcategories were used only in assessing the prevalence of the different categories of WSA indices and were rationalized to low and high stability in assessing prevalence of stability problems within the sentinel sites. Samples with ≤ 48.8 % unme were classified as high stability, those with ≥ 69.7 % were categorized as low stability, while those with unme value 55.2 % were considered transitional and categorized moderate stability.

4.3.3 Prevalence of stability related problems in LNY and HB sentinel sites

Prevalence of stability categories for individual WSA indices (mame, misp, unme and unsp) and prevalence of stability problems in LNY and HB sites, for different clusters within the sites and for sampled depth intervals is summarized in Tables 4.26 - 4.31.

Stable macro aggregates: more than 90.0 % of the soils in LNY and HB indicated low macro aggregate stability, associated with the low score (7.3 %) of mame (Table 4.26). This suggested high susceptibility of the soils to problems associated with poor aeration, low infiltration capacity and flooding hazard. Slaking plus mechanical disruption of wetted aggregates mimic effect of raindrop impact under field conditions (Cantón *et al*., 2009). This is characteristic of tropical environments with high intensity rainfall (Barthes & Roose, 2002). The rigorous mechanical disruption (shaking for 1 h in a reciprocal shaker) might cause, however, excessively higher aggregate breakdown and low prevalence of stable macro-aggregates.

Table 4.26: Prevalence of stability categories for WSA indices in LNY and HB sentinel sites.

WSA indices	Score (%)	count	stability category	Prev (%) =100×(count/339)
mame	7.3	310	low	91.4
	51.7	29	high	8.6
misp	23.3	219	low	64.6
	37.5	120	high	35.4
unme	22.8	14	high	4.1
	48.8	33	slightly high	9.7
	55.2	38	moderate	11.2
	69.7	88	slightly low	26.0
	77.8	166	low	49.0
unsp	39.1	158	high	46.6
	66.3	181	low	53.4

Stable micro-aggregates: based on misp, 65 and 35 % of the soils indicated low and high stable micro-aggregate stability, respectively (Table 4.26). The rather narrow separation between low and high stability (split of misp at 23.3 and 37.5 % respectively), could partly be explained by the very mild aggregate break down (slaking/ spontaneous disruption), characteristic of overland flow/ runoff, drip or flood irrigation. The soils presented higher prevalence of low macro- aggregate stability (90 %) than low micro-aggregate stability (65 %) (Table 4.26), suggesting potential for high aeration problem for these soils.

Unstable aggregate fraction: more than 95 % of the soils in the two sentinel sites were dispersible based on unme. Combining the subcategories (slightly high, slightly low) indicated that < 15 % of the soils could be considered stable, whereas 75 % were in low stability category (Table 4.26). For mild disruption (unsp) a more even split in prevalence between low and high stability was observed (Table 4.26). The five score values for unme suggested that the unstable fraction had inherent functional characteristics that could be segregated. It was not possible to establish definitive stability sub-categories, however, for the intermediate split points. The rigorous wet sieving pretreatment for unme might partly explain the lower split (score) value for high stability (23 %) compared to unsp (39 %) (Table 4.26).

The WSA indices indicated variation in prevalence for low stability: mame (90 %), misp (65 %), unsp (53 %), and unme (75 %), suggesting that the aggregation indices explained different dimensions of stability problems. The low amounts of mame raises hazard for low air and water movement. The low stable micro aggregates and high unstable fraction aggravates aeration problems and high water retention raising hazard of waterlogging. Between 50 and 65 % of the soils were highly susceptible to disruption as would be occasioned by light rainfall shower or flood irrigation. About 75 % of the soils

were highly susceptible to effect of high intensity rainfall (aggregate breakdown and surface transport). There were no comparable prevalence studies in the literature.

Prevalence of stability problems in LNY and HB: The soils within lower Nyando (LNY) and Homa Bay (HB) sentinel sites indicated 66 % low stability prevalence. Prevalence of low stability was higher in HB (73 %) than LNY (61 %). The sites indicated about 80 % risk (combining low and moderate categories) of stability problems (Table 4.27).

Table 4.27: Prevalence of stability related problems in LNY and HB sentinel sites.

Site	stability class	count	Prevalence (%) = (count/n)×100
LNY *(n = 172)*	low	104	60.5
	mod	33	19.2
	high	35	20.3
HB *(n = 167)*	low	122	73.1
	mod	9	5.4
	high	36	21.6
both *(n = 339)*	low	225	66.4
	mod	41	12.1
	high	71	20.9

Low stability in clusters within LNY and HB: Four clusters representing 25.0 % of lower Nyando (LNY) site indicated 100.0 % low stability prevalence. The clusters are flagged (*) and marked 'hotspot'. Additional 25.5 % (clusters 2, 3, 7, and 12) had ≥ 70.0 % low stability prevalence. Clusters with between 0.0 and 25.0 % low stability prevalence (clusters 1, 5, 13, and 14) (Table 4.28) were considered at low risk. Figure 3.2 illustrates the distribution of the different clusters in LNY.

Table 4.28: Low stability prevalence of different clusters in LNY sentinel site.

Cluster ID	low count	mod count	high count	total samples	low prev (%) = (100×(low count/ total))
1**	2	5	5	12	16.7
2	9	1	2	12	75.0
3	10	3	0	13	76.9
4*	12	0	0	12	100.0
5**	2	1	5	8	25.0
6	5	5	0	10	50.0
7	12	3	0	15	80.0
8*	15	0	0	15	100.0
9	3	1	7	11	27.3
10*	10	0	0	10	100.0
11*	11	0	0	11	100.0
12	5	1	1	7	71.4
13**	1	0	11	12	8.3
14**	0	6	1	7	0.0
15	3	5	1	9	33.3
16	4	2	2	8	50.0

Note: *(*, low stability count by all WSA indices and designated hotspots; * *- high stability count by all or majority WSA indices).*

Four clusters (1, 3, 5 and 9), representing about 25.0 % of Homa Bay (HB) site indicated 100.0 % low stability prevalence (hotspot). Additional six clusters (2, 6, 7, 8, 10 and 13) (Table 4.29) had ≥ 70.0 % low stability prevalence and were considered at risk. Clusters 12 and 16 indicated 0.0 % low stability prevalence and were considered at low risk. Figure 3.3 illustrates the distribution of the different clusters in HB.

Table 4.29: Low stability prevalence in different clusters in HB sentinel site.

Cluster ID	low count	mod count	high count	total samples	low prev (%) = (100×(low count/total))
1*	12	0	0	12	100.0
2	13	2	0	15	86.7
3*	11	0	0	11	100.0
4	5	0	5	10	50.0
5*	12	0	0	12	100.0
6	8	0	2	10	80.0
7	7	1	2	10	70.0
8	5	0	1	6	83.3
9*	14	0	0	14	100.0
10	8	0	1	9	88.9
11	3	2	5	10	30.0
12**	0	0	7	7	0.0
13	13	0	2	15	86.7
14	4	3	0	7	57.1
15	5	1	3	9	55.6
16**	0	0	7	7	0.0

Note: *(*, low stability count by all WSA indices and designated hotspot; * *- high stability count by all or majority WSA indices).*

Study field- data collection following the LDSF protocol included recording of also ancillary data in each of the 16 clusters for each sentinel site. The data included field soil texture, land use/ land cover including bare ground and vegetation cover, visible erosion, and basic infiltration rates. Analyses of this data together with predicted values for stability indicator soil properties (pH, Na2.5 and WDC) should enable future more effective characterization of the 'hotspot' and stable clusters within sentinel sites. This analyses could, however, not be realized in this study.

Low stability prevalence for each cluster was based on a maximum of 15 samples (5 plots sampled at three depth intervals) due to cost constraint of wet chemistry analysis). There were clusters also where far less samples were available where depth was limiting. This indicated low intensity data suggesting caution in prevalence

interpretation. The trends were evident, however, for rapid diagnostic screening purposes. Also the LDSF protocol is unbiased in cluster and sampling plot selection. The LDSF protocol allows determining the number of sampling plots within a cluster and future similar studies should consider using more plots for instance 10 to 15 (deGraffenried & Shepherd, 2009), for more effective cluster-based low stability prevalence assessment. Soil stability is a major risk factor of soils degradation in LVB of Kenya (Waruru *et al*., 2003ab).

Low stability for different depth intervals in LNY and HB: In LNY 50 - 57 % of the top (0-50 cm soil depth) and 71 % of the deeper (50 - 100 cm) subsoil were of low stability. In HB 60 % of the surface (0 - 20 cm), 80 % of immediate subsurface (20 – 50 cm) and > 90 % of deeper subsoil (50 - 100 cm) were of low stability. About 55 % of topsoil, 68 % subsoil and 80 % deeper subsoil in LNY and HB indicated low stability (Table 4.30).

Table 4.30: Low stability prevalence for different depth intervals in LNY and HB sentinel sites.

Site	depth (cm)	low count	total samples	prev (%) = [(low count/total)×100]
LNY	0 -20	29	58	50
	20 - 50	31	54	57.4
	50 - 100	32	45	71.1
HB	0 -20	35	59	59.3
	20 - 50	37	46	80.4
	50 - 100	32	35	91.4
Both	0 -20	64	117	54.7
	20 - 50	68	100	68
	50 - 100	64	80	80

In LNY 75 % for the top and 85 % for the deeper subsoil were at risk (low + moderate prevalence) of stability related problems. More than 66 % of the topsoil, 83 % of subsoil and 94 % of deeper subsoil in HB were at risk (low + moderate) of stability related problems. About 72 % of topsoil, 78 % of immediate subsoil and 89 % of deeper subsoil in LNY and HB were at risk (low + moderate prevalence) of stability related problems (Table 4.31).

Table 4.31: Risk of stability problems for different depth intervals in LNY and HB sentinel sites.

		Count				**prev (%) = 100×(count/total)**			
Site	**Depth (cm)**	**low**	**mod**	**high**	**total**	**low**	**mod**	**high**	**low + mod**
LNY	0 -20	29	16	13	58	50	27.6	22.4	77.6
	20 - 50	31	9	14	54	57.4	16.7	25.9	74.1
	50 - 100	32	6	7	45	71.1	13.3	15.6	84.4
HB	0 -20	35	4	20	59	59.3	6.8	33.9	66.1
	20 - 50	37	1	8	46	80.4	2.2	17.4	82.6
	50 - 100	32	1	2	35	91.4	2.9	5.7	94.3
Both	0 -20	64	20	33	117	54.7	17.1	28.2	71.8
	20 - 50	68	10	22	100	68	10	22	78
	50 - 100	64	7	9	80	80	8.8	11.3	88.8

Prevalence of stability problems for different depth-intervals informs potential risk for soil splash, sheet wash, rill and inter-rill and gully erosion. About 75 % of the surface and 85 % of the deeper subsoil in LNY were highly susceptible to erosion. In HB 66 % of the surface and 94 % of the deeper subsoil were highly susceptible. Removal of the topsoil through inappropriate land use practices has exposed the more vulnerable deeper soils contributing to severe gully development within LVB (Waruru & Wanjogu, 2002).

4.4 Wet stable aggregation (WSA) indices most appropriate for screening soil stability problems using IR-based models in LVB of Kenya

Performance of soil property sets: (i), pH2.5, Na2.5, WDC; (ii), pH2.5, Na5, WDC; (iii), pH2.5, eNa, WDC, and; (iv), pH2.5, ESP, WDC) for prediction of WSA indices from CART regression is presented in Tables 4.32 - 4.34. Tables 4.35 - 4.42 present performance of sets (i) and (iii) in assessing prevalence of stability related problems in lower Nyando (LNY) and Homa Bay (HB) sites. Table 43 presents the most appropriate WSA indices for screening stability problems in LVB.

4.4.1 Performance of soil property sets and MIR for estimation of WSA indices

Soil properties pH2.5, Na2.5, Na5, eNa, ESP and WDC were key for prediction of WSA indices (Table 4.22). Tables 4.32 and 4.33 present the performance of the four three property sets ((i), pH2.5, Na2.5, WDC; (ii), pH2.5, Na5, WDC; (iii), pH2.5, eNa, WDC, and; (iv), pH2.5, ESP, WDC) and MIR for PLS estimation of the WSA indices.

Soil-based estimation of WSA indices: The sponR could not be predicted by the four soil predictor sets (*RE* > 1.0) (Table 4.32). Sets (i) and (ii) could not predict masp and mime. Set (i) (with Na2.5) could not predict mechR, whereas set (ii) (with Na5) could predict mechR with 42 % accuracy. Set (i) presented an elaborate (albeit more complex) decision tree model for prediction of unme with double split points for both Na2.5 and pH2.5, whereas set (ii) presented weak model with < 10 % accuracy. Set (i) could predict unsp with higher efficiency than set (ii), whereas set (ii) could predict mame with higher accuracy. Set (i) was considered more appropriate for prediction of WSA indices ahead of set (ii), also because both pH2.5 and Na2.5 could be read from the same soil-water extract making data acquisition more efficient. Sets (iii) and (iv) indicated similar *RE* for prediction of mame, mime, misp, and mechR. Set (iv) (with ESP) could not predict masp, but could better predict unme and unsp than set (iii) (with eNa). Set (iii) models for unme and unsp were more complex than set (iv), a practical advantage for

screening purposes. Set (iii) was considered more appropriate ahead of set (iv) for prediction of WSA indices, also due to easier determination of eNa than ESP.

Notably, eNa was less sensitive to unme (eNa was for instance outdone by pH2.5 for prediction of unme), but was more sensitive to unsp in set (iii)**,** affirming Bell (2000) that eNa is key for assessing spontaneous aggregate disruption (slaking). The effect of sodicity on aggregate stability was low to quantify (Levy *et al*., 2003), and was probably further obscured by mechanical disruption. Effective assessment of sodicity on soil aggregation requires use of the more sensitive High Energy Moisture Characteristic (HEMC) method by Mamedov *et al*. (2007).

Table 4.32: Performance of four different soil predictor sets for estimation of WSA indices.

		mame	masp	mime	misp	unme	unsp	mechR
Set (i)	RE	0.75	0.97	0.9	0.62	0.55	0.62	0.79
	TN	2	N/A	N/A	2	5	2	N/A
	Splitter (s)	pH2.5	N/A	N/A	WDC	Na2.5[a]/pH2.5[b]	Na2.5	N/A
	Split (s)	5.84	N/A	N/A	26	3.3; 14.5 / 5.8; 6.4	12.8	N/A
	Score	100	N/A	N/A	100	100 / 23	100	N/A
Set (ii)	RE	0.52	0.85	0.9	0.62	0.9	0.7	0.58
	TN	2	N/A	N/A	2	2	2	2
	Splitter (s)	pH2.5	N/A	N/A	WDC	Na5	Na5	pH2.5
	Split (s)	5.84	N/A	N/A	26	14.5	14.5	5.84
	Score	100	N/A	N/A	100	100	100	100
Set (iii)	RE	0.52	0.71	0.89	0.62	0.89	0.72	0.58
	TN	2	2	2	2	2	2	2
	Splitter (s)	pH2.5	eNa	WDC	WDC	pH2.5	eNa	pH2.5
	Split (s)	5.84	0.47	28	26	6.23	3.72	5.84
	Score	100	100	100	100	100	100	100
Set (iv)	RE	0.52	0.88	0.84	0.62	0.61	0.61	0.58
	TN	2	N/A	2	2	4	5	2
	Splitter (s)	pH2.5	N/A	WDC	WDC	ESP/pH2.5/WDC	ESP[c]/pH2.5/WDC	pH2.5
	Split (s)	5.86	N/A	28	26	2.4/6.3/26	4.3 ; 27.8/6.3/26	5.84
	Score	100	N/A	100	100	100/44/19	100/6/22	100

Note: *(CART regression models were developed using Set 2, n = 79; predictor set (i): pH2.5, Na2.5, WDC; set (ii): pH2.5, Na5, WDC; set (iii): pH2.5, eNa, WDC, and; set (iv): pH2.5, ESP, WDC; Na2.5[a], Na2.5 split unme data twice, at 3.3 and 14.5 (mg kg^{-1}); pH2.5[b], the pH2.5 split unme data twice, at 5.8 and 6.4 (unit); ESP[c], the ESP split unme data twice at 4.3 and 27.8 (%); RE, regression relative error; TN, terminal node; N/A (not applicable), no regression tree model was grown (RE ≥ 0.9)).*

PLS calibration of WSA indices on soil predictors and MIR: The MIR was superior over soil-based predictor sets for estimation of mame, unme and mechR. The soil-based sets were superior over MIR for masp, misp, unsp, sponR and for mime. The sponR could not be estimated by MIR or soil property sets (R^2 < 0.24) (Table 4.33). Notably, the MIR was superior over the soil property sets for prediction of WSA indices from slaking plus mechanical disruption wet-sieving pretreatment, whereas the soil-based sets were superior over MIR for indices from slaking only pretreatment. This suggested combining to advantage MIR and soil-based sets for development of more effective pedotransfer functions for estimation of WSA indices from the two different wet-sieving pretreatments. Canasveras *et al*. (2010) found R^2 = 0.23 for prediction of water stable aggregates (fraction > 250 μm) using a set of six soil properties (tSa, tClay, pHw, $CaCO_3$, OM, and Fe), and R^2 = 0.53 combining spectra and the soil property set datasets.

Table 4.33: Performance of four different soil predictor sets and MIR for estimation of WSA indices.

	Set (i)		set (ii)		set (iii)		set (iv)		MIR	
Target	**r^2**	**RMSECV**	**r^2**	**RMSECV**	**r^2**	**RMSECV**	**r^2**	**RMSECV**	**r^2**	**RMSECV**
mame	0.44	0.78	0.42	0.79	0.41	0.80	0.45	0.78	0.68	0.60
masp	0.47	0.63	0.46	0.64	0.42	0.66	0.46	0.64	0.34	0.71
mime	0.31	0.70	0.28	0.71	0.26	0.72	0.28	0.71	0.19	0.76
misp	0.32	9.33	0.30	9.41	0.28	9.59	0.27	9.63	0.09	10.80
unme	0.44	12.17	0.41	12.48	0.41	12.55	0.45	12.12	0.65	9.69
unsp	0.58	12.49	0.55	12.83	0.54	13.09	0.56	12.73	0.38	15.13
mechR	0.28	0.78	0.28	0.78	0.28	0.78	0.30	0.77	0.58	0.59
sponR	0.23	0.76	0.23	0.76	0.21	0.77	0.24	0.75	0.18	0.79

Note: *(PLS models were developed using samples Set 2, n = 79; calibration is with looCV; RMSECV values for mame, masp, mechR and sponR are shown on logarithmic (ln) scale and mime on sqrt scale); (set (i): pH2.5, Na2.5, WDC; set (ii): pH2.5, Na5, WDC; set (iii): pH2.5, eNa, WDC, and; set (iv):pH2.5, ESP, WDC).*

The two most appropriate soil predictor sets: Set (i) (pH2.5, Na2.5, WDC) and set (iii) (pH2.5, eNa, WDC) were considered most appropriate for estimation of WSA indices. The PLS models using these predictor sets could match MIR for masp, mime, misp, unsp, and sponR (Table 4.33).

The four soil-based sets alternated sodicity indices (Na2.5, Na5, eNa and ESP) in the three-variable predictor sets. Sodium is a major risk factor in soil degradation, especially in relation to stability in LVB of Kenya (Waruru *et al*., 2003b; Waruru & Wanjogu, 2002). Using set (i) with Na2.5 unme data was split by Na2.5 and pH2.5 and at different data points, presenting better opportunity for separation of low and high stability categories for prevalence assessment, unlike set (ii) with Na5 that presented only one split (Table 4.32). The accuracy was much higher also using set (i) than using set (ii). Also Na2.5 could be read from the same soil-water extract used for pH2.5. Set (iv) with ESP appeared more appropriate proxy for WSA indices ahead of set (iii) with eNa from

CART screening, however, ESP is a derivative of eNa and its computation requires also determination of effective CEC making data acquisition more rigorous and expensive. Table 4.34 illustrates rationalization (based on RE) used to establish set (i) and set (iii) as the more appropriate soil-based sets for estimation of WSA indices.

Table 4.34: Efficiency of soil predictor sets for estimation of WSA indices.

Target	set (i)	set (ii)	set (iii)	set (iv)
mame	(o)	(-)	(-)	(-)
masp	N/A	N/A	**	N/A
mime	N/A	N/A	*	**
misp	(-)	(-)	(-)	(-)
unme	**	(o)	(-)	*
unsp	*	(-)	(-)	**
mechR	N/A	(-)	(-)	(-)

Note: [*(**), highest efficiency ;(*), second best; (o), lowest efficiency; (-), no difference between sets; N/A (not applicable), no predictive relationship; set (i), (ii), (iii) and (iv) as defined in Table 4.33*].

4.4.2 Predicted (score) values for WSA indices in LNY and HB

Tables 4.36 and 4.37 present grove files (calibration models) and predicted (score) values of WSA indices for soil predictor sets (i) (pH2.5, Na2.5, WDC) and (iii) (pH2.5, eNa, WDC).

Predicted (score) WSA indices in LNY and HB: Predicted (MIR) soil property data in LNY and HB (n = 339) (Data File) was in similar range as for observed values in the representative (n = 79) set. MIR overestimated, however, upper range values for Na2.5 and eNa (Table 4.35).

Table 4.35: Data range for predictor soil properties in LNY and HB sentinel sites.

Soil property	observed in representative set (n = 79)		predicted in LNY and HB (n = 339)	
	min	max	min	max
pH2.5	5.0	10.1	4.1	10.2
WDC	3.0	61.0	1.1	50.0
Na2.5	2.0	235.0	0.4	755.2
eNa	0.0	42.4	0.0	83.9

Grove Files for WSA indices: Set (i) and (iii) could not predict sponR ($RE > 1.0$). Set (i) models for also masp, mime, and mechR were not acceptable ($RE \geq 0.8$), whereas set (iii) could predict these indices with 12 to 42 % accuracy. Set (i) was superior over set (iii) for unme and unsp, whereas set (iii) was superior for mame and masp. Notably only mame, misp, unme and unsp were available for prevalence assessment using set (i) predictor properties (Table 4.36), similar to Table 4.21 using identical set of soil predictor properties.

Table 4.36: Grove files for WSA indices using selected two different soil predictor sets.

	Set (i)					set (iii)				
Target	**RE**	**TN**	**splitter (s)**	**split**	**score**	**RE**	**TN**	**splitter**	**split**	**score**
mame	0.749	2	pH2.5	5.84	100	0.522	2	pH2.5	5.84	100
masp	0.966	N/A	N/A	N/A	N/A	0.707	2	eNa	0.47	100
mime	0.896	N/A	N/A	N/A	N/A	0.888	2	WDC	28	100
misp	0.624	2	WDC	26	100	0.623	2	WDC	26	100
unme	0.552	5	Na2.5[a]/pH2.5[b]	3.25/ 5.84	100/ 23	0.893	2	pH2.5	6.23	100
unsp	0.615	2	Na2.5	12.8	100	0.718	2	eNa	3.72	100
mechR	0.785	N/A	N/A	N/A	N/A	0.581	2	pH2.5	5.84	100
sponR	1.126	N/A	N/A	N/A	N/A	1.018	N/A	N/A	N/A	N/A

Note: *(set (i): pH2.5, Na2.5, WDC; set (iii): pH2.5, eNa, WDC; grove files were developed using reference Set 2, n = 79 representative of LNY and HB; Na2.5[a], Na2.5 split unme data twice at 3.3 and 14.5 mg kg^{-1}; pH2.5[b], pH2.5 split unme data twice at 5.8 and 6.6; Na2.5 indicated 100 % importance score, whereas pH2.5 contributed 23 % to unme split; TN, terminal node; RE, relative error).*

Predicted (score) values for WSA indices in LNY and HB: Performance correspondence was observed for set (i) and set (iii) for scoring misp and unsp data (Table 4.37). Observed five score values for unme (18.3, 27.5, 55.2, 69.7, and 77.8 %) using set (i), suggested high sensitivity of unstable aggregate fraction to different levels of Na2.5 and pH2.5. The single score value for mame, unme, and mechR using set (iii) (for 2 TN), suggested less efficacy of set (iii) for prediction of these indices. CART is generic software, however, and some of its outputs could be counter-intuition (Steinberg & Colla, 2001).

Table 4.37: Score values of WSA indices for selected two soil predictor sets.

			WSA predicted (score) values	
Target	**predictor set**	**grove TN**	**TN1**	**TN2**
mame	set (i)	2	7.3	51.7
	set (iii)	2	7.3	
masp	set (iii)	2	10.7	28.4
mime	set (iii)	2	18.3	27.5
misp	set (i)	2	23.3	37.5
	set (iii)	2	23.3	37.5
unme	set (i)[a]	5	22.8	48.8
	set (iii)	2	71.1	
unsp	set (i)	2	39.1	66.3
	set (iii)	2	40.2	66.5
mechR	set (iii)	2	0.31	

Note: *(set (i): pH2.5, Na2.5, WDC; set (iii): pH2.5, eNa, WDC; TN, terminal node; set (i)[a], prediction (score) values for TN 3, 4 and 5 were 55.2, 69.7 and 77.8 (%), respectively); Data file of predicted soil predictors was dropped down (scored) grove file for each of the WSA indices; the score value (s) coincided with the TN of the particular grove file).*

The scored WSA indices were used to categories samples in the sites to (low/ high) stability. The tally for category score was used to assess prevalence for the category in LNY and HB. Separation of low and high stability categories for slaked only (and micro-aggregate) indices (masp, mime, misp, unsp) was evident but not so for mame, unme, and mechR (Table 4.37). This affirmed observed more sensitivity of WSA indices to eNa for slaked only (and for also micro-aggregates) fractions. The eNa was outclassed by Na2.5, however, for prediction of both unme and unsp (Table 4.36).

4.4.3 Prevalence of stability related problems in LNY and HB sentinel sites

Stability categories for indivindual WSA indice: Separation (albeit weak) of low/high prevalence based on predictor set (i) indicated that mame was more sensitive to pH2.5 with Na2.5 than to pH2.5 with eNa in set (iii). The five levels of prevalence from set (i) for unme, suggested better delimitation of soil dispersion potential in the sentinel sites. Set (iii) could better discern differences in prevalence of masp. Correspondence for score values for misp and unsp (set (i) and set (iii)) was followed by corresponding prevalence (misp that is). Single score value for mame, unme and mechR and corresponding 100 % low prevalence using set (iii), indicated less sensitivity of the predictor variables, suggesting more advantage using set (i) for these indices (Table 4.38).

Table 4.38: Prevalence of stability category for WSA indices for selected two soil predictor sets.

Target	predictor set	score	stability class	count	prev (%) = 100×(count/339)
mame	set (i)	7.3	low	310	91.4
		51.7	high	29	8.6
	set (iii)	7.3	low	339	100
masp	set (iii)	10.7	low	246	72.6
		28.4	high	93	27.4
mime	set (iii)	18.3	low	183	54
		27.5	high	156	46
misp	set (i)	23.3	low	219	64.6
		37.5	high	120	35.4
	set (iii)	23.3	low	219	64.6
		37.5	high	120	35.4
unme	set (i)	22.8	high	14	4.1
		48.8	slightly high	33	9.7
		55.2	moderate	38	11.2
		69.7	slightly low	88	26
		77.8	low	166	49
	set (iii)	71.1	low	339	100
unsp	set (i)	39.1	high	158	46.6
		66.3	low	181	53.4
	set (iii)	40.2	high	117	34.5
		66.5	low	222	65.5
mechR	set (iii)	0.31	low	339	100

Stability problems in LNY and HB: high stability prevalence was about 20 % in LNY and HB from both set (i) and set (iii). Set (i) indicated also 20 % moderate stability prevalence (Table 4.39). The low plus moderate prevalence using set (i) was similar to low prevalence using set (iii).

Table 4.39: Prevalence of stability problems in LNY and HB for selected two soil predictor sets.

	Stability class	count		prev (%) = 100×(count/*n*)	
Sentinel site		**set (i)**	**set (iii)**	**set (i)**	**set (iii)**
LNY (*n* = 172)	low	104	133	60.5	77.3
	mod	33	N/A	19.2	N/A
	high	35	39	20.3	22.7
HB (*n* = 167)	low	122	130	73.1	77.8
	mod	9	N/A	5.4	N/A
	high	36	37	21.6	22.2
both (*n* = 339)	low	225	263	66.4	77.6
	mod	41	N/A	12.1	N/A
	high	71	76	20.9	22.4

Set (i) was more sensitive to levels of WSA indices than set (iii), allowing prevalence assessment of moderate stability category. Combining moderate and low stability prevalence (set (i), provide better indication of stability risk for management intervention (Table 4.39).

Stability problems in different clusters from LNY: Set (i) could discern a moderate stability category for some clusters, whereas set (iii) presented only low and high categories. The low plus moderate prevalence using set (i) was similar to low prevalence using set (iii) for most of the clusters. Both sets similarly identified 'hotspot' (100 % low stability prevalence) clusters 4, 8, 10, and 11, however, stable cluster 14 using set (i) was identified as moderate using set (iii) (Table 4.40).

Table 4.40: Prevalence of stability problems in LNY clusters for selected two soil predictor sets.

	Count					prev (%) = 100× (count/total)			
	Set (i)		set (iii)			set (i)		set (iii)	
Cluster	**low**	**mod**	**low**	**mod**	**total**	**low**	**mod**	**low**	**mod**
1	2	5	8	N/A	12	16.7	41.7	66.7	N/A
2	9	1	10	N/A	12	75	8.3	83.3	N/A
3	10	3	12	N/A	13	76.9	23.1	92.3	N/A
4*	12	0	12	N/A	12	100	0	100	N/A
5	2	1	6	N/A	8	25	12.5	75	N/A
6	5	5	5	N/A	10	50	50	50	N/A
7	12	3	14	N/A	15	80	20	93.3	N/A
8*	15	0	15	N/A	15	100	0	100	N/A
9	3	1	5	N/A	11	27.3	9.1	45.5	N/A
10*	10	0	10	N/A	10	100	0	100	N/A
11*	11	0	11	N/A	11	100	0	100	N/A
12	5	1	6	N/A	7	71.4	14.3	85.7	N/A
13	1	0	9	N/A	12	8.3	0	75	N/A
14	0	6	1	N/A	7	0	85.7	14.3	N/A
15	3	5	7	N/A	9	33.3	55.6	77.8	N/A
16	4	2	6	N/A	8	50	25	75	N/A

Note: (*, *hotspot, clusters scored as low stability by all WSA indices; see Figure 3.2 for distribution of clusters in LNY)*.

Stability problems in different clusters from HB: Set (i) and set (iii) indicated perfect correspondence in identifying hotspot (100 % low stability prevalence) clusters 1, 3, 5, and 9. There was perfect correspondence also in low stability prevalence for clusters 6, 7, 10, 13, and stable cluster 16. Predictor set (i) identified moderate stability category (for instance cluster 2, 7, and 14), whereas set (iii) indicated only low/high categories. Low plus moderate prevalence using set (i) was similar; for example, to low prevalence using set (iii) for clusters 2, 11, and 15 (Table 4.41).

Table 4.41: Prevalence of stability problems in HB clusters for selected two soil predictor sets.

	Count					prev (%) = 100× (count/ total)			
	Set (i)		set (iii)			set (i)		set (iii)	
Cluster	low	mod	low	mod	total	low	mod	low	mod
1*	12	0	12	N/A	12	100	0	100	N/A
2	13	2	15	N/A	15	86.7	13.3	100	N/A
3*	11	0	11	N/A	11	100	0	100	N/A
4	5	0	7	N/A	10	50	0	70	N/A
5*	12	0	12	N/A	12	100	0	100	N/A
6	8	0	8	N/A	10	80	0	80	N/A
7	7	1	7	N/A	10	70	10	70	N/A
8	5	0	6	N/A	6	83.3	0	100	N/A
9*	14	0	14	N/A	14	100	0	100	N/A
10	8	0	8	N/A	9	88.9	0	88.9	N/A
11	3	2	5	N/A	10	30	20	50	N/A
12	0	0	1	N/A	7	0	0	14.3	N/A
13	13	0	13	N/A	15	86.7	0	86.7	N/A
14	4	3	3	N/A	7	57.1	42.9	42.9	N/A
15	5	1	6	N/A	9	55.6	11.1	66.7	N/A
16	0	0	0	N/A	7	0	0	0	N/A

Note: *(*, hotspot, clusters scored as low stability by all WSA indices; see Figure 3.3 for distribution of clusters in HB).*

Stability problems for different depth-intervals: Low stability prevalence was between 50 - 70 % (LNY) and 60 - 90 % (HB) across depth-intervals for set (i). For set (iii) low stability prevalence was 70 - 80 % (LNY) and 70 - 95 % (HB) across depth-intervals. Set (i) identified also moderate stability category for the different depth-intervals for both LNY and HB. The low plus moderate prevalence using set (i) was similar to low prevalence using set (iii) for the different depth-intervals from LNY and HB. Based on the two predictor sets between 25 and 30 % of topsoil, 20 - 25 % of subsoil, and 15 - 20 % of deep subsoil of LNY were of high stability (Table 4.42).

Table 4.42: Prevalence of stability problems in different depth intervals for selected two soil predictor sets.

		Count					prev (%) = 100× (count/total)			
		Set (i)		set (iii)			set (i)		set (iii)	
Site	depth (cm)	low	mod	low	mod	total	low	mod	low	mod
LNY	0 -20	29	16	40	N/A	58	50	27.6	69	N/A
	20 - 50	31	9	41	N/A	54	57.4	16.7	75.9	N/A
	50 - 100	32	6	37	N/A	45	71.1	13.3	82.2	N/A
HB	0 -20	35	4	40	N/A	59	59.3	6.8	67.8	N/A
	20 - 50	37	1	38	N/A	64	57.8	1.6	59.4	N/A
	50 - 100	32	1	33	N/A	35	91.4	2.9	94.3	N/A
both	0 -20	64	20	80	N/A	167	38.3	12	47.9	N/A
	20 - 50	68	10	79	N/A	118	57.6	8.5	66.9	N/A
	50 - 100	64	7	70	N/A	80	80	8.8	87.5	N/A

Stability decreased with depth for low (plus moderate) stability prevalence in LNY (73, 75, and 83 %) and low prevalence in HB (67, 59, and 94 %) for surface, subsoil, and deep subsoil respectively (Table 4.42). Erosion of topsoil has exposed the more vulnerable subsoil causing rills and gullies in lowland LVB (Waruru & Wanjogu, 2002).

4.4.4 Most appropriate WSA indices for screening stability problems in LVB

Performance correspondence of soil predictor sets: Soil predictor sets (i) (pH2.5, Na2.5, WDC) and set (iii) (pH2.5, eNa, WDC) indicated high correspondence in prevalence assessment of stability related problems.

The single score values for mame, unme and mechR were all categorized as low stability (Table 4.38), giving the tally for low stability already a high head start using set (iii). This might suggest that either: all the studied soils were unstable based on slaking plus mechanical shaking; that the shaking was too abrasive disrupting even stable aggregates and therefore was not realistic, or; that the predictor set (iii) with eNa was less appropriate for assessing stability problems for the studied soil.

Implicit in prevalence assessment was the possibility of subjectivity also in tallying low, moderate, and high stability categories across WSA indices. Where the tally for moderate count was equal to that for low or high stability count, for example, the sample was assigned to moderate category, and this inadvertently might have overestimated moderate prevalence and probably more important underestimated low stability prevalence. However, low plus moderate prevalence using set (i) was similar to low prevalence using set (iii), suggesting that the moderate category belong to low (more than high) stability in set (i). This suggested that low stability prevalence for set (iii) and combination of low plus moderate prevalence for set (i) provided better indicator of soils at risk of stability related problems in LVB.

Most appropriate WSA indices and thresh hold values: The mame, misp, unme, and unsp and their respective thresholds for low and high stability (10 and 50, 20 and 40, 70 and 20, 65 and 40 %, respectively), were the most appropriate indices for diagnostic screening and prevalence assessment of stability related problems using IR-based models in LVB of Kenya. This was informed by the correspondence and convergence in performance of the two most appropriate soil-based predictor sets for estimation of the WSA indices. The rounding up of the indices is for simplicity, taking into account also the moderate to low levels of accuracy achieved for predictions by the soil-based predictor sets (Table 4.32).

Table 4.43: Indices of WSA most appropriate for screening stability related problems using IR-based models.

WSA indices	low stability index (%)	high stability index (%)
mame	7.3 (10.0)	51.7 (50.0)
misp	23.3 (20.0)	37.5 (40.0)
unme	69.7 (70.0)	22.8 (20.0)
unsp	66.5 (65.0)	39.1 (40.0)

Note: *(values in parenthesis are approximations for easy reference).*

Implication for stability related studies in LVB: The most appropriate WSA indices (mame, misp, unme and unsp) suggested that soil aggregation and aggregate stability under wetting field conditions in LVB is influenced by disruptive forces that range from aggregate slaking to dispersion. The combined influence of the critical limits of indices from slaking only (misp, unsp) and slaking plus mechanical disruption (mame, unme) provides efficacious assessment of the prevalence of stability related problems in LVB.

Requisite for successful rapid large area (diagnostic) screening stability related problems using IR-based models is development of key few basic soil properties that are easy to determine and that have strong and robust IR-based models and that have predictive relationship with target stability functional attributes. The suite of 59 basic soil properties initially screened for prediction of WSA indices was reduced to three key soil properties (pH2.5, Na2.5 and WDC). The three property set in combination with MIR showed to be effective for developing pedo-transfer functions for diagnostic screening and prevalence assessment of stability related problems in LVB of Kenya. The three property set could estimate, for example, WSA indices with accuracy not uncommon in ecological studies: macro aggregates (mame) with 44 - 48 %), micro aggregates (misp) with up to 38 %), and unstable aggregate fraction unme (41 - 45 %) and unsp (54 - 58 %) (Table 4.33). This was better than Canasveras *et al*. (2010) who found 23 % accuracy for estimation of macro aggregates (fraction > 250 µm) from a set of six soil predictor properties (tSa, tClay, $CaCO_3$, pHw, OM, Fe).

Set (i) (pH2.5, Na2.5, WDC) and (iii) (pH2.5, eNa, WDC) presented comparable performance for prevalence assessment (stability categories for individual WSA indices, low stability in LNY and HB, low stability in different clusters, and low stability for depth-interval samples). However, set (i) indicated superior performance (higher sensitivity to stability) and practical advantage (ease of data acquisition) over set (iii). For example, Na2.5 could be read from the same extract used for pH2.5; Na-ion meter

readings were very precise for sodic soils (% CV 1.8 - 4.4), and; Na2.5 and eNa were very strongly associated (r = 0.9).

The study showed that > 77 % of soils in LNY and HB are of low stability. That > 90 % of the soils in LNY and HB are of low macro aggregate stability, ≥ 65 % of low micro aggregate stability and 50 to > 75 % are susceptible to aggregate slaking, break down and total dispersion. Low stability prevalence was 50 - 80 % and 60 – 95 % across depth-intervals in LNY and HB, respectively. Established critical values for low and high stability from macro aggregates (10 vs 50 %), micro aggregates (20 vs 40 %) and unstable aggregate fraction (65 vs 20 %) provide critical input data for modeling and monitoring soil health in LVB. No related information was available for LVB.

Notably, differences (albeit slight) in prevalence assessment was observed (comparing results under sections 4.3 and 4.4) with the same soil-based predictor set (pH2.5, Na2.5, WDC) and the same reference samples set (n = 79). This could not be adequately explained. The several (steps) iterations in selection of the three properties in section 4.3 could propagate errors, whereas the set in section 4.4 was selected directly. The use of a larger list of potential predictors (initially 28, then 8) in section 4.4 could result in better CART regression decision tree models for grove files. CART performs better when there are numerous independent variables (Steinberg & Colla, 2001) although more potential predictors could mask (lower importance score) for potentially more powerful primary predictors (Steinberg & Golovnya, 2006).

CHAPTER FIVE
SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary of the study

Application of infrared spectroscopy (IR) for soil compositional analyses is well established. However, there has been little focus on examining IR in pedotransfer functions for prediction of functional attributes related to soil stability. This is despite the fact that the potential for development of pedotransfer functions and use of inference systems for more efficient and effective estimation of functional attributes from spectra and basic soil attributes has been demonstrated. Soil physical health, especially related to stability problems is a major risk factor for degradation in the Lake Victoria basin (LVB) in Kenya. This study evaluated the use of infrared spectroscopy (IR) in diagnosing soil stability related problems and assessing their prevalence with a case study of LVB in Kenya. Specifically, the study developed alternative IR-based cross-validated (looCV) models for screening soil stability related properties across a range of sensitive soils from LVB of Kenya and compared these with predictions using conventional physico-chemical properties. The IR-based looCV models were further tested using datasets of similar soils from independent sites in LVB. The prevalence of soil stability related problems was assessed using IR-based models in two sentinel sites from LVB. Finally, indices of soil stability functional attributes most appropriate for screening stability problems using IR-based models in LVB of Kenya were developed.

A suite of 59 soil basic properties from direct measurements and derived were compared with spectral variables from four alternative IR-methods (MIR, MIRwc, NIR, NIRwc) for PLS leave-one-out cross validation (looCV) estimation of selected eight (8) water stable aggregation (WSA) indices from slaking only (masp, misp, unsp, sponR) and slaking plus mechanical disruption (mame, mime, unme, mechR) wet-sieving pretreatments. The IR-based looCV models were further independently validated following optimization trials that involved with and without removal of reference values

and spectral outliers in the validation set. Few key soil-based predictors that were strongly calibrated to MIR ($R^2 \geq 0.6$) were used to develop PLS calibration models that were in turn used to estimate these properties in the larger set from LNY and HB using the laters' MIR spectra. The soil-predictors were used to also develop calibration models (grove files in CART) for the predicted WSA indices. The grove files and the data matrix of predicted soil-based predictors in the larger set were used to predict (score) the WSA indices in LNY and HB samples in CART using a procedure called ' dropping data down a tree model' or scoring data. The predicted indices values were categorized into low or high stability and used to assess prevalence of stability related problems in LNY and HB. Two different sets of soil properties that were strongly correlated to MIR and that indicated predictive relationship with WSA indices were used to develop the most appropriate indices of WSA for diagnostic screening and prevalence assessment of stability related problems using IR-methods in LVB.

5.2 Conclusions

i. The infrared spectroscopy (IR) methods were superior over soil physicochemical predictors for PLS cross validation estimation of wet stable aggregation (WSA) indices. This affirmed that soil diffuse reflectance spectra is superior over soil properties for development of stability pedotransfer functions (PTFs). Further advantage of reflectance spectra over wet chemistry laboratory measurements includes rapidity, high throughput and high precision.

ii. Soil pH in water (pHw) and soil organic carbon (SOC) provided effective PTFs for assessment of soil wet stable macro aggregation; soil wet stable micro aggregation could be inferred from water-dispersible clay content (WDC), and; soil aggregate instability (aggregate break down and dispersion) from Na^+ concentration. This suggested exchangeable sodium (eNa), WDC, pHw and SOC as 'soil stability indices', effective for development of stability minimum

datasets. This affirms few found studies that established modest association of wet stable aggregates to soil basic properties (total sand, total clay, calcium carbonate, pHw, soil organic matter, and iron).

iii. Satisfactory independent estimation of wet stable macro aggregation (R^2 = 0.81) and colloidal clay (unstable aggregate fraction) (R^2 = 0.7 - 0.5) suggested that the accuracy of IR-PLS technique is sufficient to support field and pedon-scale analysis of water stable aggregation indices, and therefore reflectance spectroscopy has the potential to be used as a rapid soil testing technique for assessing soil physical quality.

iv. The NIR-based methods (NIR spectra and corresponding wavelet coefficients) were more robust than counterpart MIR-based methods for estimation of the WSA indices. The NIR spectra was associated with indices from direct measurements (stable macro and micro aggregates), whereas wavelet coefficients were effective for derived indices (ratio and unstable aggregate fraction). Wavelet transform counteracted errors (data quality) associated with derived properties, improving resonance of wavelet coefficients and reference data and model performance.

v. Superiority of NIR over MIR for estimation of aggregation indices is insightful for stability studies. The NIR is amenable to field-based (on-site or on-the-go) measurements, unlike MIR that is restricted to controlled (laboratory) conditions. This is important where utility of IR PLS is predicated on effective and efficient (simple, rapid, cheap, precise) soil analyses, especially for large sample sizes and multivariate data sets. Greater potential of chemometrics and spectroscopic soil studies is in landscape-level (for instance 10×10 km block) soil health suerveillance.

vi. The study showed that 70 to 80 % of the soils in Lower Nyando and Homa Bay sentinel sites had low stability problems and the risk of stability related problems

increased with soil depth. This affirmed that soil stability is a major risk factor for observed soil degradation and attendant rill and inter-rill erosion and gully development in LVB in Kenya.

vii. Stable macro (2000-212 µm) and unstable (< 20 µm) aggregate fraction from slaking plus mechanical disruption and stable micro (212-20 µm) and unstable fraction from slaking only wet-sieving pretreatments were the most appropriate indices for screening stability related problems and assessing their prevalence using IR in LVB of Kenya. Soil wet stable macro aggregate at10 and 50 %, stable micro at 20 and 40 %, and unstable fraction at 70-65 and 20-40%, define low and high stability, respectively.

viii. A combination of indices from slaking only and slaking plus mechanical disruption wet-sieving pretreatments was found most appropriate for screening stability problems in LVB of Kenya. This suggested that soil aggregation and aggregate stability under wetting field conditions in LVB is influenced by disruptive forces that range from aggregate slaking to complete dispersion.

5.3 Recommendations

i. This study illustrated and demonstrated a methodological framework for rapid large area soil health surveillance. The study benchmarks diagnostic screening and prevalence assessment of soil stability related problems using diffuse reflectance infrared spectra. The models developed can be used to diagnose and assess prevalence of soil stability related problems in the LVB of Kenya and other regions. Further research should, however, widely test similar soil property predictor sets, aggregation indices and IR methods to: (i) validate established soil-based predictors (eNa, WDC, pHw, SOC), (ii) counter variability from sample provenance for improved model geographic transferability, and (iii) assess suggested performance improvement with calibration spiking (addition of few of the new samples in the calibration set prior to model development), and

combining IR and soil properties. Basic soil properties water-dispersible clay content and sodium in soil solution (for instance from Na-ion meter readings), should be incorporated and mainstreamed in routine soil data collection to improve on development of effective minimum datasets for stability pedotransfer purposes.

ii. Soil stability is a major risk factor of soil degradation in LVB in Kenya. Cluster-based low stability prevalence was based on a maximum of 15 samples (constrained by cost of wet chemistry analysis), suggesting low intensity data for more efficacious prevalence interpretation. The land degradation surveillance framework (LDSF) protocol is otherwise flexible on the number of sampling plots within a cluster. Future studies should involve selection of more sampling plots and incorporation in analysis of also ancillary data (field soil texture, land use/ land cover including bare ground and vegetation cover, visible erosion, and basic infiltration rates) collected for each cluster in the LDSF protocol. This should enable more effective characterization of clusters within sentinel sites.

iii. This study has assessed (geo-referenced) point-based prevalence of stability related problems in two sentinel sites (10×10 km blocks) that were selected to represent the diversity of soil types, landforms, land cover and land use types in the LVB of Kenya. Further investigation should assess the spatial distribution of stability related problems in the two sites using geo-statistical tools and integrate the outputs with other remote sensing techniques in a geographic information system to allow for mapping of stability hazard for the entire LVB in Kenya.

iv. Soil- and spectra-based transfer functions developed in this study are efficacious for evaluation and monitoring soil health related problems for agriculture and environmental applications. This suggests high potential of chemometrics and

soil spectroscopy for civil engineering applications. Further studies should pursue, “diagnostic screening materials attributes using IR” in the context of civil engineering (roadway and building construction). Availability of accessible (more than 500) archived “Roads Soil Samples” from Kenya adds to feasibility of such a study.

v. A major advantage with spectroscopy techniques is that several soil properties can be determined from a single spectrum, which greatly reduces the cost of analysis compared to conventional laboratory methods. This study suggests feasibility and recommends building/development of dynamic soil spectral libraries and their use for assessing soil compositional and health (quality) attributes in Kenya at the national/ country, county and sub-county level. National soil spectral library would contribute also to the Global Spectral Libraries initiative by the International Union of Soil Science Societies. The New Partnerships for Africa Development Office of Science and Technology (NEPAD OST) and the international community could support increased training in ‘new’ soil science methods for improved soil health evaluation and monitoring, especially in data sparse case for Africa.

REFERENCES

Amezketa, E. (1999). Soil aggregate stability: A review. *Journal of Sustainable Agriculture, 14*(2-3), 83-151.

Angers, D. A., Bullock, M. S., & Mehuys, G. R. (2008). Aggregate stability to water. In M. R. Carter & E. G. Gregorich (Eds.), *Soil sampling and methods of analysis*, 2nd ed. (pp. 811-819), Canandian Society of Soil Science. Boca Raton, FL: CRC Press.

Araújo, S. R., Demattê, J. A. M., & Bellinaso, H. (2013). Analysing the effects of applying agricultural lime to soils by VNIR spectral sensing: a quantitative and quick method. *International Journal of Remote Sensing, 34*(13), 4570-4584.

Ashmana, M. R., & Hallett, C. P. D. (2003). Are the links between soil aggregate size class, soil organic matter and respiration rate artefacts of the fractionation procedure? *Soil Biology and Biochemistry, 35*, 435-444.

Atzberger, C. (2002). Soil Optical Properties: A review. *Remote Sensing Report, D-54286*. Tier: Germany.

Auerswald, K. (1995). Percolation stability of aggregates from arable topsoils. *Soil Science, 159*(2), 142-148.

Awiti, A. O., Walsh, M. G., Shepherd, K. D., & Kinyamario, J. (2008). Soil condition classification using infrared spectroscopy: A proposition for assessment of soil condition along a tropical forest-cropland chronosequence. *Geoderma, 143*(1-2), 73-84.

Azevedo, A. C., & Schulze, D. G. (2007). Aggregate distribution, stability and release of water dispersible clay for two subtropical oxisols. *Science and Agriculture (Piracicaba, Brazil), 64*(1), 36-43.

Barthes, B., & Roose. E. (2002). Aggregate stability as an indicator of soil susceptibility to runoff and erosion: validation at several levels. *Catena, 47(2)*, 133-149.

Bell, F. G. (2000). *Engineering Properties of Soils and Rocks,* 4th ed. Osney Mead, Oxford OX2 OEL: Blackwell Science Ltd.

Bellon-Maurel, V., & McBratney, A. B. (2011). Review: Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. *Soil Biology and Biochemistry, 43*(7), 1398-1410.

Bellon-Maurel, V., Fernandez, E., Palagos, B., Roger, J. M., & McBratney, A. B. (2010). Prediction of soil attributes by NIR/MIR Spectroscopy. Coming back to statistics fundamentals to improving prediction accuracy. *Trends in Analytical Chemistry, 29*(9), 1073-1081.

Ben-Dor, E., Heller, D., & Chudnovsky, A. (2008). A novel method of classifying soil profiles in the field using optical means. *Soil Science Society of America Journal, 72*(4), 1113-1123.

Ben-Dor, E., & Banin, A. (1995). Near infrared analysis as a method to simultaneously evaluate spectral featureless constituents in soils. *Soil Science, 159*(2), 259 - 270.

Bird, S. B., Herrick, J. E., Wander, M. M., & Murray, L. (2007). Multi-scale variability in soil aggregate stability: Implications for understanding and predicting semi-arid grassland degradation. *Geoderma, 140*(1), 106-118.

Boix-Fayos, C., Calvo-Cases, A., Imeson, A. C., & Soriano-Soto, M. D. (2001). Influence of soil properties on the aggregation of some Mediterranean soils and the use of aggregate size and stability as land degradation indicators. *Catena, 44*(1), 47-67.

Bouajila, A., & Gallali, T. (2008). Soil organic carbon fractions and aggregate stability in carbonated and no carbonated soils in Tunisia. *Journal of Agronomy, 7*, 127-137.

Boucher, S. C. (2010). Aggregate slaking and clay Dispersion. Retrieved from http://www.dpi.vic.gov.au/dpi/vro/vrosite.nsf/pages/soil_mgmt_slaking/.

Bouma, J. (2010). Implications of the knowledge paradox for soil science. *Advances in Agronomy, 106*, 143-71.

Bradshaw, G. A., & Spies, T. A. (1992). Characterizing canopy gap structure in forests using wavelet analysis. *Journal of Ecology, 80*(2), 205-215.

Breiman, L. J., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees*. Boca Raton, FL: CRC Press.

British Standard Institution (BSI). (1975). *Methods of Test for Soils for Civil Engineering Purposes*. 2 Park Street, Landon: British Standard Institute.

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

Brown, D. J., Shepherd, K. D., Walsh, M. G., Mays, M. D., & Reinsch, T. G. (2006). Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma, 132*(2-4), 273-290.

CAMO ASA Inc. (1998). *The Unscrambler User Manual*. Corvallis, OR: CAMO Inc.

Canasveras, J. C., Barron, V., Campillo, M. C., Torrent J., & Gomez, J. A. (2010). Estimation of aggregate stability indices in Mediterranean soils by diffuse reflectance spectroscopy. *Geoderma, 158*(1), 78-84.

Cantarella, H., Quaggio, J. A., van Raij, B., & Abreu, M. F. (2006). Variability of soil analysis of commercial laboratories: Implications for lime and fertilizer recommendations. *Communications in Soil Science and Plant Analysis, 37*(15-20), 2213-2225.

Cantón, Y., Solé-Benet, A., Asensio, C., Chamizo, S., & Fábregas, P. J. (2009). Aggregate stability in range sandy loam soils relationships with runoff and erosion. *Catena, 77*(3), 192-199.

Chang, C.W., Laird, D. A., Mausbach, M. J., Hurburgh, C. R. (2001). Near-infrared reflectance spectroscopy-principal components regression of soil properties. *Soil Science Society of America Journal, 65*(2), 480-490.

Chaudhary, B. V., Matthew, B. A., O'Dell, T. E., Grace, J. B., Redman, A. E., Rillig, M. C., & Johnson, N. C. (2009). Untangling the biological contributions to soil stability in semi-arid shrublands. *Ecological Applications, 19*(1), 110-122.

Chenu, C., Le Bissonnais, Y., & Arrouays, D. (2000). Organic matter influence on clay wettability and soil aggregate stability. *Soil Science Society of America Journal, 64(4)*, 1479-1486.

Chodak, M., Partap, K., Balazs, H., & Friedrich, B. (2004). Near infrared spectroscopy for determination of total and exchangeable cations in geologically heterogeneous forest soils. *Journal of Near Infrared Spectroscopy, 12*(5), 315-324.

Choudhary, O. P., Josan, A. S., Bajwa, M. S., & Kapur, M. L. (2004). Effect of sustained sodic and saline-sodic irrigation and application of gypsum and farmyard manure on yield and quality of sugarcane under semi-arid conditions. *Field Crop Research, 87*(2), 103-116.

Cohen, M. J., Prenger, J. P., & Debusk, W. K. (2005). Visible near infrared reflectance spectroscopy for rapid, non-destructive assessment of wetland soil quality. *Journal of Environmental Quality, 34*(4), 1422-1434.

Cook, G. D., & Muller, W. J. (1997). Sodium content a better index of soil sodicity than exchangeable sodium percentage? A reassessment of published data. *Soil Science, 162*(5), 343-349.

Couillard, A., Turgeon, A. J., Shenk, J. S., & Westerhaus, M. O. (1997). Near infrared reflectance spectroscopy for analysis of Turf soil profiles. *Crop Science, 37*(5), 1554-1559.

Dardenne, P., Sinnaeve, G., & Baeten, V. (2000). Multivariate calibration and chemometrics for near infrared spectroscopy: which method? *Journal of Near Infrared Spectroscopy, 8*(4), 229-237.

Daubechies, I. (1992). Ten Lectures on Wavelets. Philadelphia, PA: SIAM.

Dawes, L., & Goonetilleke, A. (2006). Using multivariate analysis to predict the behaviour of soils under effluent irrigation. *Water Air Soil Pollution, 172*(1-4), 109-127.

de Graffenried Jr., J. B., & Shepherd, K. D. (2009). Rapid erosion modeling in a western Kenya watershed using visible near infrared reflectance, classification tree analysis and ^{137}Cesium. *Geoderma, 154*(1), 93-100.

Demattê , J. A. M., da Silva, T. F., & Quartaroli, C. F. (2012). Spectral behavior of some modal soil profiles from São Paulo State, Brazil. *Bragantia Campinas, 71*(3), 413-423.

Denef, K., Six, J., Merckx, R., & Paustian, K. (2004). Carbon sequestration in micro aggregates of no-tillage soils with different clay mineralogy. *Soil Science Society of America Journal, 68*(6), 1935-1944.

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

Dunn, B. W., Beecher, H. G., Batten, G. D., & Ciavarella, S. (2002). The potential of near infrared reflectance spectroscopy for soil analysis: a case study from the riverine plain of south-eastern Australia. *Australian Journal of Experimental Agriculture, 42*(5), 607-614.

Dwinnell, W. (1998). Modeling Methodology 2: Model Input Selection (pp23-26). In *Principal Component Analysis (PCAI), 12*.

EMCA. (1999). *The Environmental Management and Coordination ACT (EMCA)*. Kenya Gazette Supplement 3 (Acts No. 1). Nairobi: Government Printer.

FAO. (1986). *Guidelines: Land Evaluation for Rainfed Agriculture, Soils Bulletin 52*. Rome: FAO.

Farley, J. (2012). Ecosystem services: The economics debate. *Ecosystem Services, 1*(1), 40-49.

Farres, P. J., & Cousen, S. M. (2006). An improved method of aggregate stability measurement. *Earth Surface Processes and Landforms, 10*(4), 321-329.

Fortun, A., Fortun, C., & Ortega, C. (2006). Effect of farmyard manure and its humic fractions on the aggregate stability of a sandy–loam soil. *European Journal of Soil Science, 40*(2), 293-298.

Fratta, D., Aquettant, J., & Roussel-Smith, L. (2007). *Introduction to Soil Mechanics Laboratory Testing*. Boca Raton: CRC Press.

Gachene, C. K. K., Ngetich, F. K., & Anyika, F. (2003). Soil aggregate stability as influenced by different residue management practices (pp10-11). *Legume Research Network Newsletter, 9*. KARI: Nairobi.

Genot, V., Colinet, G., Bock, L., Vanvyve, D., Reusen, Y., & Dardenne, P. (2011). Near infrared reflectance spectroscopy for estimating soil characteristics valuable in the diagnosis of soil fertility. *Journal of Near Infrared Spectroscopy, 19*(2), 117-138.

Government of Kenya (GoK). (2007). *Kenya Vision 2030: A Competitive and Prosperous Nation*. Government of the Republic of Kenya. Nairobi: Government Printer.

Gributs, C. E. W., & Burns, D. H. (2006). Parsimonious calibration models for near-infrared spectroscopy using wavelets and scaling functions. *Chemometrics and Intelligent Laboratory Systems, 83*, 44-53.

Hartemink, A. E., & McBratney, A. B. (2008). A soil science renaissance. *Geoderma, 148*, 123-129.

Haynes, R. J., & Swift, R. S. (2006). Stability of soil aggregates in relation to organic constituents and soil water content. *European Journal of Soil Science, 41*, 73-83.

Hazelton, P., & Murphy, B. (2007). *Interpreting Soil Test Results: What Do All The Numbers Mean?* Collingwood, Australia: CSIRO Publishing.

Hinga, G., Muchena, F. N., & Njihia, C. M. (1980). *Physical and Chemical Methods of Soil Analyses*. Nairobi: National Agricultural Research Laboratories.

Hulugalle, N. R., & Finlay, L. A. (2003). $EC_{1:5}$ / exchangeable Na, a sodicity index for cotton farming systems in irrigated and rainfed Vertosols. *Australian Journal of Soil Research, 41*(4), 761-769.

Hulugalle, N. R., Entwistle, P. C., & Mensah, R. K. (1999). Can Lucerne (*Medicago sativa L*) strips improve soil quality in irrigated cotton (*Gossypium hirsutum L.)* fields? *Applied Soil Ecology, 12*(1), 81-92.

Idowu, O. J., Moebius, B. N., van Es, H. M., Schindelbeck, R. R., Abawi, G. S., Wolfe, D. W., Thies, J., Gugino, B., & Clune, D. (2008a). Soil health assessment and management: measurements and results. *What's cropping up: Cornell Soil Health PWT*. Retrieved from http://soilhealth.cals.cornell.edu/ .

Idowu, O. J. (2003). Relationships between aggregate stability and selected soil properties in humid tropical environment. *Communications in Soil Science and Plant Analysis, 34*(5-6), 695-708.

Idowu, O. J., van Es, H. M., Abawi, G. S., Wolfe, D.W., Ball, J. I., Gugino, B. K., Moebius, B. N., Schindelbeck, R. R., & Bilgili, A.V. (2008b). Farmer-oriented assessment of soil quality using field, laboratory, and VNIR spectroscopy methods. *Plant and Soil, 307*, 243-253.

Igwe, C. A. (2003). Shrink – swell potential of floodplain soils in Nigeria in relation to moisture content and mineralogy. *International Agrophysics, 17*(2), 47-55.

Igwe, C. A. (2005). Erodibility in relation to water-dispersible clay for some soils of eastern Nigeria. *Land Degradation and Development, 16*(1), 87-96.

Igwe, C. A., & Nwokocha, D. (2005). Influence of soil properties on the aggregate stability of a highly degraded tropical soil in Eastern Nigeria. *International Agrophysics, 19*(2), 131-139.

Igwe, C. A., & Stahr, K. (2004). Water-stable aggregates of flooded Inceptisols from southeastern Nigeria in relation to mineralogy and chemical properties. *Australian Journal of Soil Research, 42*(2), 171-179.

Irvine, S. A., & Reid, D. J. (2001). Field prediction of sodicity in dryland agriculture in Central Queensland. *Australian Journal of Soil Research, 39*(6), 1349-1357.

Islam, K., Singh, B., & McBratney, A. B. (2003). Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared spectroscopy. *Australian Journal of Soil Research, 41*(6), 1101-1114.

Kariuki, P. C., van der Meer, F. D., & Siderius, W. (2003a). Classification of soils based on engineering indices and spectral data. *International Journal of Remote Sensing, 24*(12), 2567-2574.

Kariuki, P. C., Woldai, T., & van der Meer, F. D. (2003b). Effectiveness of spectroscopy in identification of swelling indicator clay minerals. *International Journal of Remote Sensing, 25*(2), 455-469.

Kemper, W. D., & Rosenau, R. C. (1986). Aggregate stability and size distribution. In A. Klute (Ed.), *Methods of Soil Analysis, Part I,* 2nd ed. (pp. 425-442). Madison, WI: American Society of Agronomy.

Knadel, M., Stenberg, B., Deng, F., Thomsen, A., & Greve, M. H. (2013). Comparing predictive abilities of three visible –near infrared spectrophotometers for soil organic carbon and clay determination. *Journal of Near Infrared Spectroscopy, 21*(1), 67-80.

Knodel, P. C. (1991). *Characteristics and problems of dispersive clay soils*. Materials Engineering Branch, Research and Laboratory Services Division. Denver Office, Denver, Co: USDI Bureau of Reclamation.

Kodešová, R., Rohošková, M., & Žigová, A. (2009). Comparison of aggregate stability within six soil profiles under conventional tillage using various laboratory tests. *Biologia, 64*(3), 550-554.

Koger, C. H., Bruce, L. M., Shaw, D. R., & Reddy, K. N. (2003). Wavelet analysis of hyperspectral reflectance data for detecting pitted morning glory *(Ipomoea lacunosa*) in soybean (*Glycine max*). *Remote Sensing for Environment, 86*(1), 108-119.

Krasilnikov, P., Carré, F., & Montanarella, L. (2008). *Soil geography and geostatistics: Concepts and Applications*. Joint Research Centre, Scientific and Technical Reports: Institute for Environment and Sustainability. ISSN: 1018-5593. EUR 23290: European Commission.

Lal, R. (2009). Sequestering carbon in soils of arid ecosystems. *Land Degradation and Development, 20*(4), 441-454.

Lark, R. M. (2007). Decomposing digital soil information by spatial Scale. In P. Lagacherie, A.B. McBratney, & M. Voltz (Eds.), *Digital soil mapping: An introductory perspective* (pp. 301-326). Amsterdam: Elsevier.

Le Bissonnais, Y. (1996). Aggregate stability and assessment of soil crustability and erodibility: I. Theory and methodology. *European Journal of Soil Science, 47*(4), 425-437.

Le Bissonnais, Y., & Arrouays, D. (2005). Aggregate stability and assessment of soil crustability and erodibility: II. Application to humic loamy soils with various organic carbon contents. *European Journal of Soil Science, 48*(1), 39-48.

Le Bissonnais, Y., Blavet, D., De Noni, G., Laurent, J.Y., Asseline, J., & Chenu, C. (2007). Erodibility of Mediterranean vineyard soils: relevant aggregate stability methods and significant soil variables. *European Journal of Soil Science, 58*(1), 188-195.

Le Bissonnais, Y., Cros-Cayot, S., & Gascuel-Odoux, C. (2002). Topographic dependence of aggregate stability, overland flow and sediment transport. *Agronomie, 22*(5), 489-501.

Lebron, I., Suarez, D. L., & Alberto, F. (1994). Stability of a calcareous saline - sodic soil during reclamation. *Proceedings of Soil Science Society of America Journal, 58*(6), 1753-1762.

Lemon, S. C., Roy, J., Clark, M. A., Friedmann, P. D., & Rakowski, W. (2003). Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. *Annals of Behavioral Medicine, 26*(3), 172-181.

Levy, G. J., & Mamedov, A. I. (2002). Aggregate stability and seal formation. *Soil Science Society of America Journal, 66*(5), 1603-1609.

Levy, G. J., Mamedov, A. I., & Goldstein, D. (2003). Sodicity and water quality effects on slaking of aggregates from semi-arid soils. *Soil Science, 168*(8), 552-562.

Levy, G. J., Tang, Z., Yu, J., Shainberg, A., Mamedov, I., & Ben-Hur, M. (2006). Runoff and interill erosion in sodic soils treated with dry PAM and Phosphogypsum. *Soil Science Society of America Journal, 70*(2), 679-690.

Li, J. T., & Zhang, B. (2007). Paddy soil stability and mechanical properties as affected by long-term application of chemical fertilizer and animal manure in subtropical China. *Pedosphere, 17*(5), 568-579.

Li, Z. X., Cai, C. F., Shi, Z. H., & Wang, T.W. (2005). Aggregate stability and its relationship with some chemical properties of red soils in subtropical China. *Pedosphere, 15*(1), 129-136.

Linker, R. (2012). Application of FTIR Spectroscopy to Agricultural Soils Analysis. In *Fourier Transforms – New Analytical Approaches and FTIR Strategies* (pp. 385-404).

Ludwig, B., Nitschke, R., Terhoeven-Urselmans, T., Michel, K., & Flessa, H. (2008). Use of mid-infrared spectroscopy in the diffuse-reflectance mode for the prediction of the composition of organic matter in soil and litter. *Journal of Plant Nutrition and Soil Science, 171*(3), 384-391.

Lyons, D., Rayment, G., Hill, R., Daly, B., Marsh, J., & Ingram, C. (2011). *Aspac soil proficiency testing program report 2007-08*. Technical Report, ASPAC. Melbourne: Victoria. Retrieved from http://www.aspac-australasia.com/index.php/documents/.

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

Maindonald, J. H., & Braun, J. (2003). *Data Analysis and Graphics Using R: An Example –Based Approach*, *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press.

Mamedov, A. I., Huang, C. H., Skidmore, E., & Levy, G. J. (2007). HEMC: a sensitive aggregate stability method for soil quality evaluation. *ASA – CSSA - SSSA International Annual meetings Nov 4 – 8 (2007), New Orleans, Louisiana*.

Marquez, C. O., Garcia, V. J., Cambardella, C. A., Schulz, R. C., & Isenhart, T. M. (2004). Aggregate-size distribution and soil stability. *Soil Science Society of America Journal, 68*(3), 725-735.

Martens, H., & Martens, M. (2001). *Introduction to Multivariate Data Analysis for Understanding Quality*. Chichester, UK: John Wiley & Sons.

Mbagwu, J. S. C. (1992). A comparison of three micro-aggregation indices with other tests of structural stability. *International Agrophysics, 6*(1-2), 27-32.

McBratney, A. B., Mendonca, S. M. I., & Minasny B. (2003). On digital soil mapping. *Geoderma, 117*(1), 3-52.

McBratney**,** A. B., Minasny, B., & Viscarra Rossel, R. A. (2006). Spectral soil analysis and inference systems: a powerful combination for solving the soil data crisis. *Geoderma, 136*(1), 272-278.

McCarty, G. W., Reeves III, J. B., Reeves, V. B., Follet, R. F., & Kimble, J. M. (2002). Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil

carbon measurements. *Soil Science Society of America Journal, 66*(2), 640-646.

McClure, W. F. (2003). Review: 204 years of near infrared technology: 1800-2003. *Journal of Near Infrared Spectroscopy, 11*(6), 487-518.

Mead, R., Curnow, R. N., & Hasted, A. M. (2002). *Statistical Methods in Agriculture and Experimental Biology, 3rd edition*. UK: Chapman & Hall/ CRC.

Merrington, G. (2006). *The development and use of soil quality indicators for assessing the role of soil in environmental interactions*. Science Report: SC030265. Almondsbury, Bristol: Environmental Agency.

Merry, R. H., & Janik, L. J. (2001). Mid infrared spectroscopy for rapid and cheap analysis of soils, In *Proceedings of the 10th Australian agronomy conference, Hobart, 28th Jan. to 1st Feb. 2001, Urrbrae, South Australia*. CSIRO land and Water.

Minasny, B. (2007). Predicting soil properties. *Journal of Ilmu Tanah dan Lingkungan, 7*(1), 54-67.

Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences, 32*(9), 1378-1388.

Minasny, B., & McBratney, A. B. (2008). Regression rules as a tool for predicting soil properties from infrared spectroscopy. *Chemometrics and Intelligent Laboratory Systems, 94*(1), 72-79.

Minasny, B., Tranter, G., McBratney, A. B., Brough, D. M., & Murphy, B.W. (2009). Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. *Geoderma, 153*(1), 155-162.

Moron, A., & Cozzolino, D. (2003). Exploring the use of near infrared reflectance spectroscopy to study physical properties and microelements in soils. *Journal of Near Infrared Spectroscopy, 11*(2), 145-154.

Mouazen, A. M., Karoui, R., De Baerdemaeker, J., & Ramon, H. (2005). Classification of soil texture classes by using soil visual near infrared spectroscopy and factorial discriminant analysis techniques. *Journal of Near Infrared Spectroscopy, 13*(4), 231-240.

Muhati, S. I., Shepherd, K. D., Gachene, C. K., Mburu, M. W., Jones, R., Kironchi, G. O., & Sila, A. (2011). Diagnosis of soil nutrient constraints in small-scale groundnut (*Arachis hyopaea L.*) production systems of western Kenya using infrared spectroscopy. *Journal of Agricultural Science and Technology, A 1*, 111-127.

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

Næs, T., Isaksson, T., Fern, T., & Davies, T. (2002). *A User friendly Guide to Multivariate Calibration and Classification*. Chichester, UK: NIR Publications.

Nanni, M. F., & Demattè, J. A. (2006). Spectral reflectance methodology in comparison to traditional soil analysis. *Soil Science Society of America Journal, 70*(2), 393-407.

NEPAD OST. (2008). *New Partnerships for Africa's Development, Office of Science and Technology*. Johannesburg and FAO, Rome.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthes, B., Dor, E. B., Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Dematte, J.A., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J.M., Shepherd, K.D., Stenberg, B., Towett, E.K., Vargas, R., &Wetterlind, J. (2015). Soil spectroscopy: An alternative to wet chemistry for soil monitoring. *Advances in Agronomy, 132*, 139-159.

Oguike, P. C., & Mbagwu, J. S. C. (2009). Variations in some physical properties and organic matter content of soils of coastal plain sand under different land use types. *World Journal of Agricultural Science, 5*(1), 63-69.

Omuto, C. T. (2008). Assessment of soil physical degradation in eastern Kenya by use of a sequential soil testing protocol. *Agriculture Ecosystem and Environment, 128*(21), 199-211.

Omuto, C. T., & Shrestha, D. (2007). Remote sensing techniques for rapid detection of soil physical degradation. *International Journal of Remote Sensing, 28*, 4785-4805.

Osuji, G. E., & Onweremadu, E. U. (2007). Structural stability of dystric Nitisol in relation to some edaphic properties under selected land uses. *Nature Science, 5*(4), 7-13.

Peverill, K. I., Sparrow, L. A., & Reuter, D. J. (1999). *Soil Analysis: An Interpretation Manual*. ISBN 0 643 063765. Victoria, Australia: CSIRO Publishing.

Phelps, M. C., & Merkle, E. (2008). Classification and regression trees as alternative to regression (pp.77-78). *Proceedings of the 4th Annual GRASP Symposium, Wichita State University (2008)*.

Pirie, A., Singh, B. & Islam, K. (2005). Ultra-violet, visible, near-infrared and mid-infrared diffuse reflectance spectroscopic techniques to predict several soil properties. *Australian Journal of Soil Research, 43*(6), 713-721.

Pringle, J. (1975). Soil stability measurement. Newcastle University, Cyclostyled.

R-Development Core Team. (2012). *R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing*. ISBN 3-900051-07-0. Vienna, Austria.

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

Rengasamy, P., & Olsson, K. A. (1991). Sodicity and soil structure. *Australian Journal of Soil Research, 29*(6), 935-952.

Roper, M. M. (2005). Managing soils to enhance the potential for bioremediation of water repellency. *Australian Journal of Soil Research, 43*(7), 803-810.

Saeys, W., Xing, J., De Baerdemaeker, J., & Ramon, H. (2005). Comparison of transflectance and reflectance to analyse hog manures. *Journal of Near Infrared Spectroscopy, 13*(2), 99-108.

Salford Systems Inc. (2008). *CART1 V 6.0*. San Diego: Salford Systems Inc.

Sanchez, P. A., Ahamed, S., Carré, F., Hartemink, A. E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A. B., McKenzie, N. J., & Mendonça-Santos, M. (2009). Digital soil map of the world. *Science, 325*(5941), 680-681.

Shepherd, K. D. & Walsh, M. G. (2007). Review: Infrared spectroscopy - enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries. *Journal of Near Infrared Spectroscopy, 15*(1), 1-19.

Shepherd, K. D. (2010). Soil spectral diagnostics – infrared, x-ray and laser diffraction spectroscopy for rapid soil characterization in the Africa Soil Information Service. *19th World Congress of Soil Science, Soil Solutions for a Changing World, 1-6 August 2010, Brisbane, Australia*. Published on CDROM.

Shepherd, K. D., & Walsh, M. G. (2002). Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal, 66*(3), 988-998.

Shepherd, K. D., & Walsh, M. G. (2003). Improving accuracy and quality of routine soil analyses using diffuse reflectance spectroscopy. *Paper presented at ASA-CSSA-SSSA Annual Meetings, 2-6 November 2003, Denver, Colorado, USA*.

Shepherd, K. D., Vanlauwe, B., Gachengo, C. N., & Palm, C. A. (2005). Decomposition and mineralization of organic residues predicted using near infrared spectroscopy. *Plant Soil, 277*(1-2), 315-333.

Shepherd, K. D., Walsh, M. G., & Awiti, A. (2003). Use of soil spectral indicators for assessing and monitoring soil quality. *Paper presented at ASA-CSSA-SSSA Annual Meetings, 2-6 November 2003, Denver, Colorado, USA*.

Shirazi, M. A., & Boersma, L. (1984). A unifying quantitative analysis of soil texture. *Soil Science Society of America Journal, 48*(1), 142-147.

Shulka, M. K., Lal, R., & Ebinger, M. (2004). Soil quality indicators for the north Apalachian experimental watersheds in Coshocton, Ohio. *Soil Science, 169*(3), 195-205.

Sijali, I. V. (2001). Drip Irrigation: options for smallholder farmers in eastern and southern Africa. *Technical Handbook No. 24, Regional Land Management Unit, RELMA/ Sida*. ICRAF House, Gigiri, Nairobi.

Sorensen, L. K., & Dalsgaard, S. (2005). Determination of clay and other soil properties by near infrared spectroscopy. *Soil Science Society of America Journal, 69*(1), 159-167.

Srilatha, C., Abraham, A., & Thomas, J. P. (2004). Feature deduction and ensemble design of intrusion detection systems. *Computers & Security, 5*, 1-14.

Steinberg, D., & Colla, P. (2001). *CART: Tree-Structured Non-Parametric Data Analysis*. San Diego, CA: Salford Systems Inc.

Steinberg, D., & Golovnya, M. (2006). *CART 6.0 User's Guide Manual*. San Diego, CA: Salford Systems Inc.

Stenberg, B. (2010). Effects of soil sample pretreatments and standardized rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. *Geoderma, 158*(1), 15-22.

Stenberg, B., Nordkvist, E., & Salomonsson, L. (1995). Use of Near infrared reflectance spectra of soils for objective selection of samples. *Soil Science, 159* (2), 109-114.

Stenberg, B., Viscarra Rossel, R. A., Mounem, A., & Wetterlind, J. (2010). Visible and NIR spectroscopy in soil science: A Review. *Advances in Agronomy, 107*, 163-215.

Sumner, M. E. (2000). *Hand Book of Soil Science*. Taylor and Francis, CRC Press.

Swallow, B. M. (2001). *Improved land management in the Lake Victoria Basin*. *Annual Technical Report July 2000 to June 2001*. Natural Resources Problems, Priorities and Policies Programme, Working Paper 2001 - 4. Nairobi: ICRAF.

Swift, M. J., & Shepherd, K. D. (2007). *Saving Africa's Soils: Science and Technology for Improved Soil Management in Africa.* ISBN: 92 9059 2109. Nairobi: World Agroforestry Centre (ICRAF).

Tajik, F., Rahimi, H., & Pazira, E. (2003). Effects of electrical conductivity and sodium adsorption ratio of water on aggregate stability in soils with different organic matter content. *Journal of Agricultural Science and Technology, 5*, 67-75.

Terhoeven-Urselmans, T., Vagen, Tor-G., Spaargaren, O., & Shepherd, K. D. (2010). Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library. *Soil Science Society of America Journal, 74*(5), 1792-1799.

Thomas, P. J., Baker, J. C., & Zelazny, L.W. (2000). An expansive soil index for predicting shrink–swell potential. *Soil Science Society of America Journal, 64*(1), 268-274.

Tittonell, P., Shepherd, K. D., Vanlauwe, B., & Giller, K. (2008). Unravelling the effects of soil and crop management on maize productivity in smallholder agricultural systems of western Kenya – an application of classification and regression tree analysis. *Agriculture Ecosystem and Environment, 123*(1), 137-150.

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of American Meteorological Society, 79*(1), 61-78.

Tranter, G., Minasny, B., McBratney, A. B., Viscarra Rossel, R. A., & Murphy, B.W. (2008). Comparing spectral soil inference systems and mid-infrared spectroscopic predictions of soil moisture retention. *Soil Science Society of America Journal, 72*(5), 1394-1400.

Trygg, J., & Wold, S. (1998). PLS regression on wavelet compressed NIR spectra. *Chemometrics and Intelligent Laboratory Systems, 42*(1), 209-220.

UNEP. (2012). *Land health surveillance: An evidence-based approach to land ecosystem management. Illustrated with a case study in the West Africa Sahel*. Nairobi: United Nations Environment Programme.

US Salinity Laboratory Staff. (1954). Diagnosis and improvement of saline and alkali soils. *Agricultural Handbook 60*. Washington, DC: USDA Superintendent of Documents.

USDA-NRCS. (1996). Indicators of soil quality evaluation: Soil Quality Information Sheet. *USDA Natural Resources Conservation Service*. Washington, DC: USDA Superintendent of Documents.

Vågen, Tor-G. (2009). *Assessment of land degradation in the Sasumua watershed. Baseline Report*. Nairobi: World Agroforestry Centre (ICRAF).

Vågen, Tor-G., Shepherd, K. D., & Walsh, M. G. (2006). Sensing landscape level change in soil fertility following deforestation and conversion in the highlands of Madagascar using Vis-NIR spectroscopy. *Geoderma, 133*(3), 281-294.

Vågen, Tor-G., Winowiecki, L. A., Tondoh, J. E., & Desta, L. T. (2013). *Africa Soil Information Service (AfSIS), Soil Health Mapping*. http://hdl.handle.net/1902.1/19793 V2 [Version].

Vahyala, I. E. (2009). Soil structure control in Corn and Soybean residue management system. *International Annual Meeting of ASA – CSSA – SSSA Nov, 1-5, 2009, Theme: Footprints in the Landscape: Sustainability through Plant and Soil Science*. Pittsburgh: PA.

van Groenigen, J. W., Mutters, C. S., Horwath, W. R., & van Kessel, C. (2003). NIR and DRIFT - MIR spectrometry of soils for predicting soil and crop parameters in a flooded field. *Plant Soil, 250*(1), 155-165.

Verchot, L. V., Dutaur, L., Shepherd, K. D., & Albrecht, A. (2011). Organic matter stabilization in soil aggregates: Understanding the biogeochemical mechanisms that determine the fate of carbon inputs in soils. *Geoderma, 161*(3), 182-193.

Viscarra Rossel, R. A., Jeon, Y. S., Odeh, I. O. A., & McBratney, A. B. (2008). Using a legacy soil sample to develop a mid-IR spectral library. *Australian Journal of Soil Research, 46*(1), 1-16.

Viscarra Rossel, R. A., & Lark R. M. (2009). Improved analysis and modeling of soil diffuse reflectance spectra using wavelets. *European Journal of Soil Science, 60*(3), 453-464.

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

Viscarra Rossel, R. A., Adamchuk, V. I., Sudduth, K. A., McKenzie, N. J., & Lobsey, C. (2011). Proximal soil sensing: An effective approach for soil measurements in space. *Advances in Agronomy, 113*(1), 243-291.

Viscarra Rossel, R. A., Rizzo, R., Demattе, J. A. M., & Behrens, T. (2010). Spatial modeling of a soil fertility index using vis-NIR spectra and terrain attributes. *Soil Science Society of America Journal, 74*(4), 1293-1300.

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

Walczak, B., & Massart, D. L. (1997). Wavelets – something for analytical chemistry? *Trends in Analytical Chemistry, 16*(8), 451-463.

Walsh, M. G., & Vagen, Tor-G. (2006). *Land Degradation Surveillance Framework (LDSF): Guide to Field Sampling and Measurement Procedures*. Nairobi: World Agroforestry Centre (ICRAF).

Ward, P. A., & Carter, B. J. (2004). Dispersion of saline and non-saline Natric Mollisols and Alfisols. *Soil Science, 169*(8), 554-566.

Waruru, B. K., Njoroge, C. R. K., & Wanjogu, S. N. (2003a**)**. *Biophysical baseline information for the Nyando catchment area, The soils of the Nyando catchment area*. *Reconnaissance Soil Survey Report No. R21*. Nairobi: Kenya Soil Survey.

Waruru, B. K., & Wanjogu, S. N. (2002). Intrinsic permeability and the air-to-water permeability ratio as a measure of the structural stability of some highly eroded soils in the lowlands of the Lake Victoria Basin, Kenya. *Proceedings of the 20th Soil Science Society of East Africa, Mbale, Uganda* (pp. 282-288). Nairobi: Soil Science Society of East Africa.

Waruru, B. K., Wanjogu, S. N., Njoroge, C. R. K., & Wagate, P. N. (2003b**)**. Erosion Hazard within the River Nyando catchment Lake Victoria basin of Kenya. *Proceedings of the 21st Soil Science Society of East Africa Conference 1st – 5th December, 2003, Eldoret, Kenya* (pp.47-55). Nairobi: Soil Science Society of East Africa.

Wei, C., Gao, M., Shao, J., Xie, D., & Pan, G. (2006). Review: Soil aggregate and its response to land management practices. *China Particuology, 4*(05), 211-219.

Westerhof, R., Buurman, P., van Griethuysen, C., Ayarza, M., Vilela, L., & Zech, W. (1999). Aggregation studied by laser diffraction in relation to plowing and liming in the Cerrado region in Brazil. *Geoderma, 90*(3), 277-290.

Whelan, B. M., Koppi, A. J., McBratney, A. B., & Dougherty, W. J. (1995). An instrument for the in situ characterization of soil structural stability based on the relative intrinsic permeabilities to air and water. *Geoderma, 65*(3), 209-222.

Williams, A., Xing, B., & Veneman, P. (2005). Effect of cultivation on soil organic matter and aggregate stability. *Pedosphere, 15*(2), 255-262.

Williams, P. C., & Norris, K. (2001). Variables affecting near-infrared spectroscopic analysis. In P. Williams and K. Norris (Eds.), *Near – Infrared Technology in the Agricultural and Food Industries* (pp. 171- 185). St. Paul, Minnesota: American Association of Cereal Chemists.

www.AfricaSoils.net/. Africa Soil Information Service.

Yang, C. C., Prasher, S. O., & Goel, P. K. (2004). Differentiation of crops and weeds by decision – tree analysis of multispectral data. *Transactions of American Society of Agricultural Engineers, 47*(3), 873-879.

Yoder, R. E. (1936). A direct method of aggregate analysis of soils and a study of the physical nature of erosion losses. *Journal of American Society of Agronomy, 28*(5), 337-351.

Yohannes, Y., & Webb, P. (1999). *Classification and Regression Trees. A User Manual for Identifying Indicators of Vulnerability to Famine and Chronic Food Insecurity*. ISBN 0-89629-337- 8. 2033 K Street, N.W. Washington DC: International Food Policy Research Institute.

Zhang, B., & Horn, R. (2001). Mechanisms of aggregate stabilization in Ultisols from subtropical China. *Geoderma, 99*(1), 123-145.

Zhang, B., & Peng, X. (2006). Organic matter enrichment and aggregate stabilization in a severely degraded Ultisol after reforestation. *Pedosphere, 16*(6), 699-706.

Zhang, Z., Wei, C., Xie, D., Gao, M., & Zeng, X. (2008). Effects of land use patterns on soil aggregate stability in Sichuan Basin. *China Particuology, 6*(3), 157-166.

Zobeck, T. M., 2004. Rapid soil particle size analyses using Laser diffraction. *Applied Engineering in Agriculture, American Society of Agricultural Engineers, 20*(5), 633-639.

\

APPENDICES

Appendix 1: Building CART regression decision tree.

In building CART regression decision tree model, all the observations are initially placed in the root node. This node is impure (heterogeneous) because it contains (target) observations of mixed variances. The goal is to apply a rule that will break up these observations and create binary nodes that are internally more homogenous than the root node. Partitioning of the root node into binary nodes is based upon a two-format question as follows: Is $X \leq d$, where, X is a continuous variable and d is a constant within the range of X values. The 'yes' answers are sent to the left child node and the 'no' answers to the right child node. To come up with candidate splitting rule, CART uses a computer-intensive algorithm that searches for the best split at all possible split points for each variable. The rule is to choose that split that result in the maximum reduction in the impurity of the parent node. The procedure separates high (mean) values of the dependent variable from its low values and results in left (high) and right (low) values nodes that are now internally more homogeneous than the parent node. The methodology that CART uses for growing trees is technically called binary recursive partitioning (Steinberg & Golovnya, 2006).

CART regression tree construction centers on the definition of three major elements: (1) the sample-splitting rule, (2) the goodness-of-split criteria, and (3) the criteria for choosing an optimal tree. The CART default splitting rule (impurity measure function) is the Least Squares (LS) criterion, where node impurity is measured by within-node sum of squares, SS (t), which is defined as:

$SS(t) = \Sigma (yi(t) - ybar(t))^2$, for i = 1, 2, …., Nt,

Where yi(t) = individual values of the dependent variable at node t; ybar(t) = the mean of the dependent variable at node t, and; Nt = number of cases (samples) in node t. Given the impurity function SS(t), and a split 's' that sends cases to left (tL) and right (tR) child nodes, the goodness of a split is measured by the function

$Ø(s, t) = SS(t) - SS(tR) - SS(tL)$

Where, SS (tR) is the sum of squares of the right child node, and SS (tL) is the sum of squares of the left child node. The best split is that split for which Ø (s, t) is the highest (Yohannes & Webb, 1999).

CART grows an overly large (complex) tree, some of whose branches are sequentially pruned from the bottom up, using validation (default is 10-fold cross-validation). CART then rank-orders the sequence of pruned trees, and once the minimal-cost tree (tree with lowest mean squared error-MSE) is identified, the tree at one standard error (1SE) from the minimum-cost tree is picked as the optimal tree. CART avoids over-fit decision tree since pruning and validation is automated, and where validation fails, CART reports that no decision tree could be grown (Steinberg & Golovnya, 2006). After choosing an optimal tree, CART computes summary statistics (min, Q1, mean/ median, Q3, max, N-cases) for each terminal node (TN). For LS, CART computes the mean and SD of the target. The mean of the TN become the predicted value of the target for cases in that TN (Steinberg & Golovnya, 2006).

A goal of CART is to develop a simple tree structure for data; therefore, relatively few variables may appear explicitly in the splitting criteria. However, unlike a linear regression model, a variable in CART regression can be considered highly important even if it never appears as a primary node splitter (Steinberg & Golovnya, 2006). Variables earn credit towards their importance in two ways: as primary splitters and as surrogate splitters. To calculate a variable importance score, CART looks at the improvement measure attributable to each variable in its role as a primary splitter and as surrogate to the primary split at every node in which the variable appears. The values of these improvements are summed over each node and totaled, and are scaled relative to the best performing variable. The variable with the highest sum of improvements is scored 100, and all other variables will have lower scores ranging downwards towards zero (Steinberg & Golovnya, 2006). Eliminating the primary splitter (variable with 100 % importance score) allows the top surrogate to effectively split the data, illustrating the

phenomenon of one variable hiding (masking) the significance of another (Steinberg & Golovnya, 2006).

Appendix 2: Soil properties in the calibration sample set (corresponding values in the validationset are shown in parenthesis).

Soil property	minimum	25th percentile	50th percentile	75th percentile	maximum	SD
pHw2.5	4.6 (5.0)	6.2 (6.6)	7.0 (7.4)	8.2 (8.2)	10.5 (10.1)	1.3 (1.0)
pHw5	5.0 (3.0)	6.6 (6.9)	7.2 (7.4)	7.6 (7.9)	9.7 (9.9)	0.9 (1.1)
pHKCL	3.6 (3.6)	5.0 (4.8)	5.8 (5.5)	6.5 (6.4)	10.2 (8.0)	1.2 (1.0)
EC2.5	0.00 (0.01)	0.03 (0.04)	0.05 (0.07)	0.11 (0.12)	2.35 (2.11)	0.36 (0.33)
EC5	0.00 (0.01)	0.06 (0.05)	0.11 (0.08)	0.24 (0.18)	2.52 (3.35)	0.42 (0.42)
totC	0.17 (0.18)	0.77 (0.99)	1.29 (1.47)	1.78 (2.18)	8.18 (6.45)	1.03 (1.07)
OC	0.12 (0.11)	0.61 (0.80)	0.96 (1.23)	1.44 (1.66)	7.05 (4.75)	0.83 (0.82)
inC	0.01 (0.02)	0.11 (0.10)	0.20 (0.20)	0.31 (0.41)	3.42 (1.80)	0.47 (0.40)
totN	0.01 (0.00)	0.06 (0.08)	0.09 (1.13)	0.14 (0.18)	0.70 (0.56)	0.08 (0.09)
C:N	4.40 (2.67)	11.03 (10.06)	12.60 (11.67)	15.10 (13.62)	70.52 (164.0)	8.13 (17.11)
sCa	0.04 (0.1)	0.3 (3.7)	0.6 (6.0)	0.8 (11.3)	6.0 (84.0)	0.6 (10.0)
sMg	0.03 (0.00)	0.1 (1.2)	0.2 (2.3)	0.4 (4.1)	0.9 (25.5)	0.2 (4.6)
sK	0.00(0.01)	0.08(0.04)	0.19 (0.07)	0.32 (0.27)	1.75 (0.05)	0.25 (0.05)
sNa	0.00(0.00)	0.1 (0.0)	0.4 (1.5)	1.0 (4.0)	15.0 (45.0)	2.6 (7.2)
Na2.5	0.1 (2.0)	0.0 (5.0)	2.6 (15.5)	8.5 (30.6)	355.0 (235.0)	55.3 (38.3)
Na5	0.1 (2.0)	1.9 (5.0)	5.5 (17.0)	17.8 (33.8)	455.0 (270.0)	75.8 (46.6)
SAR	0.01(0.00)	0.7 (0.0)	1.9 (0.7)	4.8 (2.9)	80.7 (30.5)	12.2 (4.6)
eCa	0.2 (2.4)	9.7 (11.0)	15.0 (18.7)	22.2 (27.0)	39.6 (44.5)	9.0 (9.9)
eMg	0.01 (0.30)	3.1 (2.6)	4.5 (4.5)	6.2 (6.9)	10.5 (10.4)	2.5 (2.6)
eK	0.1 (0.0)	0.2 (0.1)	0.4 (0.2)	0.6 (0.4)	4.5 (1.3)	0.6 (0.2)
eNa	0.0 (0.0)	0.2 (0.3)	1.5 (3.9)	4.8 (7.6)	63.8 (45.3)	12.6 (9.3)
eCaP	0.4 (5.1)	55.7 (59.5)	68.6 (65.5)	73.5 (72.2)	90.1 (95.2)	21.3 (14.4)
eMgP	0.03 (1.1)	13.9 (12.5)	18.8 (18.0)	24.5 (21.9)	36.2 (41.6)	8.0 (7.8)
ePP	0.2 (0.0)	0.9 (0.3)	1.4 (0.8)	2.6 (1.7)	16.3 (6.1)	2.3 (1.3)
ESP	0.0 (0.1)	1.5 (2.3)	5.4 (13.0)	16.3 (21.6)	99.2 (92.7)	25.3 (17.0)
eNaR	0.00 (0.00)	0.02 (0.02)	0.06 (0.15)	0.20 (0.28)	166.9 (13.4)	22.2 (1.3)

Appendix 2: Contd.

Soil property	minimum	25th percentile	50th percentile	75th percentile	maximum	SD
CEC1	3.4 (4.8)	16.6 (16.8)	26.4 (30.3)	36.6 (43.8)	67.0 (76.4)	14.0 (16.6)
CEC2	14.0 (6.7)	50.6 (52.3)	65.6 (65.0)	89.2 (81.8)	322.6 (172.4)	45.3 (28.8)
CEC3	6.1 (1.8)	46.0 (47.0)	60.9 (61.5)	83.2 (77.7)	322.1 (167.4)	45.9 (28.7)
GR	0.02 (0.00)	0.2 (0.6)	0.4 (1.2)	1.0 (1.9)	3.7 (7.5)	0.8 (1.2)
ESI 1	0.01 (0.00)	0.03 (0.02	0.06 (0.03)	0.33 (0.18)	36.7 (2.0)	5.4 (0.3)
ESI 2	0.00 (0.00)	0.01 (0.01)	0.02 (0.01)	0.05 (0.03)	5.43 (0.28)	0.80 (0.06)
Fe	0.2 (0.0)	0.7 (7.8)	1.4 (18.6)	3.4 (40.0)	36.6 (279.0)	5.0 (44.0)
Mn	0.2 (2.7)	1.0 (13.4)	2.1 (24.6)	4.9 (45.3)	77.3 (364.7)	7.4 (44.1)
Cu	0.01 (0.01)	0.1 (2.1)	0.2 (3.7)	0.3 (4.9)	2.4 (8.9)	0.2 (1.8)
Zn	0.01(0.4)	0.01(1.4)	0.1 (2.3)	0.1 (3.8)	2.2 (19.5)	0.2 (2.9)
B	0.00 (0.01)	0.37 (0.43)	0.59 (0.95)	1.09 (2.14)	4.64 (5.35)	0.76 (1.20)
P	0.4 (0.1)	2.9 (1.5)	6.8 (3.0)	18.8 (7.4)	147.9 (100.7)	20.9 (17.8)
tSa	11.0 (11.0)	19.0 (19.0)	31.0 (33.0)	51.0 (49.5)	88.0 (89.0)	19.0 (19.2)
tSi	3.0 (2.0)	20.0 (14.0)	26.0 (20.0)	32.0 (24.0)	56.0 (40.0)	9.8 (7.9)
tClay	9.0 (7.0)	25.0 (30.5)	41.0 (47.0)	51.0 (61.0)	61.0 (81.0)	14.8 (17.0)
Dg	0.01 (0.01)	0.01 (0.02)	0.02 (0.03)	0.07 (0.04)	0.49 (0.06)	0.08 (0.01)
Δg	7.8 (3.9)	13.4 (13.5)	15.7 (17.9)	18.5 (21.5)	25.1 (44.7)	3.7 (7.0)
tSi : tClay	0.20 (0.07)	0.44 (0.32)	0.62 (0.39)	0.96 (0.61)	2.75 (1.90)	0.46 (0.31)
WDSa	19 (17)	29 (25)	39 (39)	59 (55)	91 (91)	18 (19)
WDSi	3 (2)	22 (22)	30 (30)	36 (36)	52 (60)	11 (11)
WDC	5 (3)	20 (21)	25 (27)	35 (41)	60 (69)	12 (14)
WDSi:WDC	0.23 (0.13)	0.84 (0.79)	1.09 (1.03)	1.60 (1.60)	4.86 (3.33)	0.67 (0.61)
CDR	23.8 (27.3)	56.7 (53.9)	69.5 (67.2)	84.0 (75.6)	100 (100)	19.5 (15.8)
DR	29.0 (73.9)	84.2 (90.1)	88.2 (92.9)	91.8 (95.1)	133.3 (97.6)	12.4 (4.9)
FI	0.4 (0.4)	16.0 (24.1)	30.5 (32.8)	43.3 (45.7)	76.2 (72.7)	19.4 (15.9)

Appendix 2: Contd.

Soil property	minimum	25th percentile	50th percentile	75th percentile	maximum	SD
PL	10.8 (11.4)	17.1 (18.1)	21.2 (22.4)	25.8 (26.8)	39.3 (45.3)	6.9 (8.0)
LL	21.8 (22.2)	40.5 (44.3)	52.7 (60.8)	66.0 (73.5)	90.7 (96.7)	18.7 (22.2)
PI	5.5 (8.4)	21.8 (22.8)	30.1 (35.6)	41.1 (47.0)	62.8 (66.1)	14.0 (16.1)
LS	2.9 (3.6)	9.9 (11.4)	12.1 (14.3)	14.3 (15.7)	21.2 (20.0)	4.2 (4.7)
COLE	0.00 (0.04)	0.10 (0.13)	0.10 (0.17)	0.17 (0.19)	0.27 (0.25)	0.1 (0.1)
VS	9.1 (11.5)	36.6 (43.9)	47.5 (58.9)	58.8 (67.3)	104.3 (95.3)	20.6 (22.7)
A	0.4 (0.3)	0.7 (0.6)	0.8 (0.7)	1.0 (0.9)	1.8 (1.9)	0.3 (0.3)
Mc	0.7 (0.7)	4.4 (4.2)	6.5 (6.8)	8.3 (9.0)	15.9 (13.4)	3.1 (3.3)

Note: *pHw2.5, pH in 1:2.5 soil-water suspension (unit); pHw5, pH in 1: 5 soil-water suspension (unit); pHKCl, pH in 1: 2.5 soil-1.0 N KCL (unit); EC2.5, electrical conductivity (EC) in 1:2.5 soil-water suspension (dS/m); EC5, EC in 1:5 soil-water suspension (dS/m); totC, soil total carbon (%); OC, organic carbon (%); inC, inorganic/carbonate-carbon (%); totN, total nitrogen (%); C : N, totC/totN ratio (unit); soluble calcium (sCa), magnesium (sMg), potassium (sK), sodium (sNa) (cmol(+) L^{-1}); Na2.5, sodium concentration (Na^+) in 1: 2.5 soil-water suspension (mg kg^{-1}); Na5, Na^+ in 1:5 soil-water suspension ((mg kg^{-1}); SAR, sodium adsorption ratio (unit); exchangeable calcium (eCa), magnesium (eMg), potassium (eK), sodium (eNa) (cmol (+) kg^{-1});exchangeable calcium percent (eCaP), magnesium (eMgP), potassium (ePP), sodium (ESP)(%); eNaR, exchangeable sodium ratio (unit); CEC1, effective cation exchange capacity (cmol(+) kg^{-1}; CEC 2, clay activity (100*CEC1/tClay) (cmol(+) kg^{-1}; CEC 3, clay activity accounting for contribution of SOC (cmol(+) $kg^{-1)}$; GR, soil gypsum requirement ((cmol(+) $kg^{-1)}$; ESI, electrochemical stability index (ESI): ESI 1 (EC/eNa), ESI 2 (EC/ESP) (unit); extractable/ available iron (Fe), manganese (Mn), zinc (Zn), copper (Cu), phosphorus (mg kg^{-1}), boron (B) (%); total sand (tSa), silt (tSi), clay (tClay) (%); tSi : tClay, total silt/ total clay ratio (unit); dg, median particle diameter (µm); δg, standard deviation for dg measurements ((µm); water-dispersible sand (WDSa), silt (WDSi), clay (WDC) (%); WDSi : WDC, ratio WDSi/ WDC (unit); CDR, clay dispersion ratio (%); DR, soil dispersion ratio (%); FI, flocculation index (%);PL, plastic limit, LL,*

liquid limit, PI, plasticity index (%); LS, linear shrinkage (%); VS, volumetric shrinkage (%), COLE, coefficient of linear extensibility (unit); A, activity number (PI/tClay) (unit); mc, air-dried moisture content (%).

Appendix 3: Correlation coefficient (*r-value*) (upper diagonal) for selected soil properties in the calibration samples set.

pHKCl	OC	totC	eCa	eMg	eNa	Na5	ESI 1	CEC1	tclay	WDC	WDSa	dg	tSi:tclay	FI	mame	sponR	LL	pHw2.5	pHw5	
1	-0.4	-0.2	-0.13	-0.36	0.63	0.61	-0.03	0.42	-0.33	-0.15	0.22	0.22	0.25	-0.24	-0.25	-0.27	-0.22	0.90	0.86	**pHKCl**
	1	0.9	0.16	0.28	-0.2	-0.17	0.28	-0.02	0.19	0.03	-0.15	-0.2	-0.06	0.24	0.52	0.62	0.37	-0.45	-0.34	**OC**
		1	0.12	0.2	-0.1	-0.04	0.24	0.04	0.12	0.01	-0.1	-0.1	0	0.16	0.49	0.57	0.34	-0.27	-0.17	**totC**
			1	0.61	-0.29	-0.25	-0.14	0.48	0.5	0.42	-0.53	-0.5	-0.2	0.02	-0.19	-0.1	0.67	-0.02	0.01	**eCa**
				1	-0.16	-0.17	-0.13	0.42	0.68	0.61	-0.65	-0.6	-0.26	0.03	-0.17	-0.03	0.69	-0.20	-0.17	**eMg**
					1	0.87	-0.15	0.69	0.08	0.29	-0.17	-0.1	0.05	-0.37	-0.19	-0.13	0.05	0.60	0.61	**eNa**
						1	-0.11	0.59	0.06	0.14	-0.12	-0	0.04	-0.19	-0.17	-0.14	0.07	0.52	0.53	**Na5**
							1	-0.24	-0.15	-0.27	0.32	0.09	-0.04	0.21	0.39	0.33	-0.09	-0.15	-0.15	**ESI 1**
								1	0.51	0.64	-0.61	-0.5	-0.12	-0.32	-0.3	-0.17	0.62	0.49	0.53	**CEC1**
									1	0.78	-0.8	-0.7	-0.6	0.17	-0.21	-0.06	0.79	-0.19	-0.12	**tclay**
										1	-0.8	-0.6	-0.34	-0.44	-0.39	-0.21	0.69	0.01	0.04	**WDC**
											1	0.77	0.15	0.17	0.37	0.2	-0.67	0.07	0.02	**WDSa**
												1	0.15	-0.13	0.11	-0.02	-0.72	0.12	0.02	**dg**
													1	-0.41	0.14	0.03	-0.3	0.16	0.17	**Si:clay**
														1	0.32	0.28	0.04	-0.30	-0.27	**FI**
															1	0.8	0.09	-0.39	-0.30	**mame**
																1	0.2	-0.36	-0.25	**sponR**
																	1	-0.15	-0.03	**LL**
																		1	0.86	**pHw 2.5**
																			1	**pHw5**

Appendix 4: Soil properties strongly correlated to MIR in two different reference sample sets.

		LooCV models Set 1 (n = 128)		looCV models Set 2 (n = 79)	
Soil test	**transform**	**r^2**	**RMSECV**	**r^2**	**RMSECV**
mc	N/A	0.91	0.96	0.87	1.16
pH2.5	sqrt	0.67	0.13	0.81	0.08
pH5	N/A	0.63	0.51	0.58	0.74
pHKCL	ln	0.86	0.07	0.85	0.07
Na2.5	ln	0.57	1.61	0.71	0.64
Na5	ln	0.62	1.53	0.65	0.75
totC	ln	0.89	0.24	0.90	0.21
totN	sqrt	0.86	0.04	0.60	0.08
OC	sqrt	0.86	0.12	0.91	0.11
CEC2	ln	0.63	0.33	0.59	0.27
eCa	N/A	0.83	3.79	0.71	5.01
eMg	N/A	0.80	1.10	0.71	1.39
eNa	ln	0.73	1.38	0.66	1.21
CEC1	sqrt	0.86	0.52	0.86	0.59
ESP	ln	0.63	1.36	0.74	0.86
ESI 1	ln	0.64	1.16	0.58	1.04
tClay	N/A	0.62	9.29	0.66	9.41
tSa	ln	0.78	0.25	0.73	0.27
WDC	sqrt	0.72	0.62	0.77	0.64
WDSa	ln	0.80	0.18	0.78	0.21
PL	ln	0.62	0.18	0.63	0.17
LL	N/A	0.78	8.75	0.86	8.18
PI	N/A	0.65	8.25	0.78	7.41
LS	N/A	0.65	2.51	0.87	1.79
COLE	N/A	0.67	0.03	0.85	0.02
VS	N/A	0.63	12.53	0.83	9.44
δg	N/A	0.62	0.05	0.57	0.01
eNaR	ln	0.66	1.64	0.69	1.04

Note: *(Set 1, n = 128 from across LVB and Set 2, n = 79 representative of LNY and HB sentinel sites in LVB)*