Digital Soil Mapping and GIS-based Land Evaluation for Rice Suitability in Kilombero Valley, Tanzania

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Boniface Hussein John Massawe, M.S.
Graduate Program in Environment and Natural Resources

The Ohio State University
2015

Dissertation Committee:
Dr. Brian Slater, Advisor
Dr. Warren Dick
Dr. Laura Lindsey

Abstract

A GIS-based multi-criteria land evaluation was performed in Kilombero Valley, Tanzania in order to avail decision makers and farmers with evidence based decision support tool for improved and sustainable rice production in this important region for agricultural investment. Five most important criteria for rice production in the area were identified through literature search and discussion with local agronomists and farmers. The identified criteria were 1) soil properties, 2) surface water resources, 3) accessibility, 4) distance to markets, and 5) topography. Spatial information for these criteria for the study area was generated.

To generate spatial soil information, field survey and lab analysis was conducted using base map generated from a legacy soil map and 30 m Digital Elevation Model (DEM). OSACA, a *k*-means based clustering application was used to perform distance metric numerical classification of soil profiles. The profiles were classified into 13 clusters. The clusters were demonstrated to be different from each other through comparison of modeled continuous vertical variability of selected attributes of modal soil cluster centroids by using equal area spline functions. Two decision tree based algorithms, J48 and Random Forest (RF) were applied to construct models to spatially predict the soil clusters using environmental correlates derived from 30 m DEM, 5 m RapidEye satellite

image and legacy soil map using the *scorpan* digital soil mapping framework. The RF predicted soil cluster map was picked for land evaluation because the algorithm demonstrated superiority by having comparatively higher predictive rate and pixel contiguity. Topsoil attributes values of predicted soil clusters were used to produce soil physical and chemical properties maps.

On-screen digitization, reclassifications and overlays in ArcMap and Whitebox GIS software were used to create spatial layers of the other identified criteria. Rivers were digitized from the satellite image and topographic map of the study area to create surface water resources map. Roads were digitized to create accessibility map and market centers' coordinate points were digitized to create distance to market map. Slope gradient derivative from DEM was used to create topography map.

Analytical hierarchy process (AHP) method was used to score the criteria by the local extension staff and lead farmers on a scale of 0.0 – 1.0. Surface water resource scored the highest weight (0.462) followed by soil chemical properties (0.234). Other criteria and their weight in paranthesis are soil physical properties (0.19), topography (0.052), accessibility (0.036), and distance to market (0.025).

The multi-criteria land evaluation results showed that about 8% of the study area was classified as having low suitability for rice production while only 2% was highly suitable. The majority of the area (about 89%) was classified as having medium suitability for rice production. Since the suitability decision was dominated by the surface water resource criterion, the rice suitability in the study area can be greatly improved by improving the water resources management.

Dedicated to my children

Acknowledgments

I first and foremost thank God, the Almighty for His unconditional love, guidance, protection and blessings. Words cannot express my gratitude for all He has done to me.

This work was made possible by the scholarship I was offered through a USAID funded project - innovative Agricultural Research Initiative (*i*AGRI). I greatly appreciate kind and timely support I got from the iAGRI Project Management Unit in Tanzania under Dr. David Kraybill when I was in my research work and from the Ohio State University’s International Programs in Agriculture (IPA) office under Dr. Mark Erbaugh when I was doing my course work and dissertation writing. All staffs have been very supportive, but I would like to especially mention Pat Rigby for being like a mother to me, Wendi Howell for being like a good sister and Dave Hansen for his encouraging confidence on me.

I also appreciate Norman E. Borlaug Leadership Enhancement in Agriculture Program (Borlaug LEAP) for offering me a grant which facilitated additional field trips for my dissertation research. The trips facilitated early completion of this work. I heartily appreciate Dr. Leigh Winowiecki of Center for Tropical Agriculture (CIAT), Nairobi for her contributions and readiness to work with me as a co-mentor in this grant.

I sincerely thank my Dissertation Committee: Dr. Brian Slater, Dr. Warren Dick, and Dr. Laura Lindsey for their constructive criticism which helped me to shape this dissertation. I especially thank my Advisor, Dr. Brian Slater for his professional guidance and criticism throughout this work. His expertise in digital soil mapping and GIS, and the comments thereof have made me a better scientist. I also sincerely thank Dr. Abel Kaaya of Sokoine University of Agriculture (SUA) for agreeing to serve as my Tanzania based advisor as per iAGRI scholarship format. His field work participation and logistic support are deeply appreciated.

I appreciate support offered by Dr. Namik Sonmez in the Remote Sensing part of my work, and Dr. Sakthi Subburayalu in the Machine Learning work. Your support and advice were invaluable.

I am grateful to my employer, Sokoine University of Agriculture for granting me the study leave which gave me time to concentrate on my course work and research.

I would like to acknowledge dedicated and unlimited support from Kilombero District extension staff for facilitating and participating in the field works, and lab technicians from soil lab of the Department of Soil Science, Sokoine University of Agriculture for sometimes accepting my request to prioritize working on my samples. You all made my work smoother and less stressful.

Special thanks are due to all my family members and true friends who stood with me through good and challenging times. Resilience and understanding shown by my

children were a true motivation for me to finish this work. Mr. Ignas Chilewa and his wife Neema deserve a mention for being true and generous friends in Columbus. Thanks for being there when I needed time-out from academic stresses. The live NBA games, Super Bowl events, match madness, Thanksgivings, and some outings to enjoy real American life you facilitated will make me feel indebted throughout mylife.

It is not possible to mention everybody who helped me in one way or another to successfully finish this work. I thank you all.

Vita

1990 ....Leganga Primary School, Arusha, Tanzania

1994 ....Moshi Secondary School (Ordinary level), Kilimanjaro, Tanzania

1997 ....Moshi Secondary School (Advanced level), Kilimanjaro, Tanzania

2002 ....B.Sc. Agronomy, Sokoine University of Agriculture, Tanzania

2003 - 2004 ....Company Agronomist, Balton (Tanzania) Limited

2004 - 2008 ....Project Officer, Producer Empowerment and Market Linkage, FARM Africa - Tanzania

2008 - 2010 ....Tutorial Assistant, Department of Soil Science, Sokoine University of Agriculture, Tanzania

2010 - 2014 ....Assistant Lecturer, Department of Soil Science, Sokoine University of Agriculture, Tanzania

2011 ....M.Sc. Land Use Planning and Management, Sokoine University of Agriculture, Tanzania

2014 to present ....Lecturer, Department of Soil Science, Sokoine University of Agriculture, Tanzania

Fields of Study

Major Field: Environment and Natural Resources

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Background Information and Rationale

1.1.1 Land Evaluation and Land Use Planning

Land evaluation is the process of assessing land potentials for specified land use type (FAO, 1985). The land use type can be recreational, agricultural, industrial, settlement, fishing, grazing, and nature reserves, among others. The evaluation for a specific land use provides a rational basis for taking decisions pertaining its management and optimization of benefits related to the use (FAO, 1985; Deckers et al., 2002). Hence, if resources and technology are available, any land use decision would ideally first require land evaluation.

In agriculture, evaluating land for a crop performance is referred to as crop suitability analysis. It is a prerequisite to achieving optimum utilization of the available land resources for sustainable agricultural production (Perveen et al., 2008). One of the most burning needs in Tanzania is to improve agricultural land management and to impart suitable cropping patterns in order to increase the agricultural production with efficient use of land resources.

To be able to efficiently accomplish the land use planning exercises, dynamic land evaluation systems are required. These systems should be able to collect, store, and process the inputs needed for land evaluation and display the results. One of the most important inputs for land evaluation processes is spatial soil information. Others include climatic conditions, topography, accessibility, and socio-economic factors such as availability of financial and health services (Kuria et al., 2011). Choice for inputs for a land evaluation process depends on the land use type requirements.

Geographical information system (GIS) provides the much needed analysis and means of results communication for land evaluation processes (Mendas and Delali, 2012; Pan and Pan, 2012). GIS has improved the efficiency of data processing, facilitates data integration and supports spatial analysis. Land use planning is a multicriteria decision, that is, we need more than one criterion to decide suitability of the land for a given land use type. GIS supports algorithms which can be used to make multicriteria decision analyses (Marinoni and Hoppe, 2006). In multicriteria analyses, the criteria used for making decision have different weights. For example, if soil fertility and accessibility are considered as criteria for a land use type decision, planners may give more weight to one of the criteria over the other depending on their relative importance for that land use type. Weighting tools for multicriteria decision making processes have been developed, and some of them such as analytical hierarchy analysis – AHP (Saaty, 1980) have been incorporated in GIS application software such as in the System for Automated Geoscientific Analyses – SAGA (Cimmery, 2007).

1.1.2 Meeting Soil Spatial Information Challenges

Inadequacy of spatial information including soil data is common in most developing countries including Tanzania (Msanya et al., 2002; Cook et al., 2008). According to McBratney et al. (2003), the main reason for the inadequacy of soil spatial information is related to the slowness and expense of the conventional methods used for its collection. The inadequacy is recorded more in developing countries like Tanzania where little has been done to collect and harmonise spatial soil information.

Recent develoments in computational and information technology have improved methods in which soil information can be collected, shared, communicated and updated (Scull et al., 2003; McBratney et al., 2003; Malone, 2013). Adding to the increasing power of GIS, other tools such as GPS, and remote and proximal sensors are making spatial data acquisition and interpolation easier (McBratney et al., 2003). Soil prediction models such as *scorpan* which are capable of predicting continuous soil classes and soil attributes and other recent advances such as use of machine learning which have been used in digital soil mapping elsewhere (Subburayalu and Slater, 2013), can be useful in predicting soil spatial data required for land evaluation where it is lacking, such as in Tanzania. These prediction procedures, which have never been applied in Kilombero Valley, are applied in this study to generate the spatial soil data which was used in the multi-criteria land evalution for rice suitability analysis in the valley. Such an outcome will enable farmers and the policy makers to identify the most suitable areas and management options that could develop and promote sustainable rice farming.

1.1.3 General Background Information on Rice

Rice (*Oryza sativa*) is one of the most important cereal crops in the world. It provides 20% of the world's dietary energy supply, as compared to other important staples such as wheat (supplies 19%) and maize (supplies 5%) (FAO, 2004).

Rice is a grass (monocot) plant with round culms, flat leaves and terminal panicles (De Data, 1989). As a crop, it is normally grown as an annual plant, but in favourable conditions such as in tropical areas, it can survive as a perennial and produce up to 20 ratoon crops (Haifa, 2008). It is a plant that produces tillers (branches). The primary tillers grow from the lowermost nodes of the transplanted/germinated seedlings, further giving rise to secondary and tertiary tillers. The plant height can reach 1.8 m depending on the variety (there are more than 40,000 varieties) and favourable growth conditions such as soil fertility (Haifa, 2008). The plant takes 70 to 160 days to mature, depending on the type of variety and growth conditions. The edible part of the plant is a grain produced at the panicle of each tiller. The panicle may contain about 80-120 grains depending on crop performance and variety (De Data, 1989). Different from many other crops which would suffocate if grown in wetlands because of oxygen deficiency, rice have special anatomy (aerenchyma) which enable it to transport oxygen from other parts of the plant down to the root tissues to make respiration possible (Sahrawat, 2012).

Historians believe that rice was domesticated in years earlier than 5000 B.C., but the production was first documented in China in year 2800 B.C. From China it spread to India, Iraq, Iran, Egypt and Japan by the year 500 B.C. (Haifa, 2008). In East Africa, the crop is believed to have been introduced by traders from India and Sri Lanka sailing to and trading in the coastal towns of Somalia, Zanzibar and Kilwa about 2000 years ago (Carpenter, 1978). In the mainland of Tanzania, the crop was introduced by Arab traders and settlers along the slave trade routes around year 1852 (Meertens et al., 1999).

Rice is grown in four ecosystems: dryland, rainfed wetland, deep water and mangrove swamps, and irrigated wetlands (Balasubramanian et al., 2007). Worldwide, around 79 million ha of rice is grown under irrigated conditions. While this accounts for about 55% of the total rice area, it accounts for about 75% of the world's annual rice production (Dobermann and Fairhurst, 2000). The rainfed wetlands ecosystem accounts for 35% of total rice area and is characterized by lack of water control leading to potential problems of floods and droughts (Haifa, 2008).

Most important problems facing rice production include drought, flood, extreme temperatures, and salinity. Others include weeds; diseases such as rice blast, Rice yellow mottle virus (RYMV), and African rice gall midge (AfRGM); and soil fertility (Balasubramanian et al., 2007). According to De Data (1989), modern high-yielding varieties producing around 5 t/ha of grain, can remove from the soil about 110 kg N, 34 kg P_2O_5, 156 kg K_2O, 23 kg MgO, 20 kg CaO, 5 kg S, 2 kg Fe, 2 kg Mn, 200 g Zn, 150 g Cu, 150 g B, 250 kg Si and 25 kg Cl per ha. Replenishing nutrients through application of fertilizers and crop residues incorporation are thus, important practices for sustainable rice production.

1.1.4 Rice Production in Tanzania

Rice is the second most important cereal crop in Tanzania after maize and the majority of rice farmers depend on it both for food and cash (Bucheyeki et al., 2011). Rice farming in Tanzania is characterised by many small holder farmers, cultivating small farms ranging from 0.5 to 10 acres (Massawe and Amuri, 2012; TIC, 2013). Only a small portion of rice produced in the country is from irrigated land. It is estimated that rain fed agriculture accounts for over 71% of rice production in Tanzania (RLDC, 2011). The average rice yield per unit area under small scale farms is 1.0 to 1.5 t ha^{-1} (Bucheyeki et al., 2011) comparing to yields of up to 15 t ha^{-1} in some Asian countries (Ly et al., 2012). This low rice production is hardly keeping pace with the rapidly rising demands from neighbouring countries and steeply rising demand from local population who have started preferring rice from the traditional maize, sorghum, cassava and banana as staple food (TIC, 2013).

There have been efforts by the government of Tanzania and other stakeholders including America’s Feed the Future programme to increase the production of rice in the country. These efforts have resulted to an increase of annual rice production from 511 tons in 2000 to 990 tons in 2013 (Indexmundi, 2014).

The Kilombero Valley with a total length of 250 km and width of up to 65 km, covering an area of about 11,600 km^2 (Kato, 2007) presents great potential for expansion and intensification of rice production. The government of Tanzania projects the production of rice to quadruple by 2018. This can only be possible if the land resources which are able to support the production are identified, evaluated, and mapped. This will enable making of informed decisions about optimization and sustainability of the production.

1.1.5 Objectives of the Study

This study used digital soil mapping techniques to generate spatial soil information, which was combined with other biophysical and socio-economic factors to perform GIS-based multi-criteria land evaluation for rice cultivation suitability in a portion of Kilombero Valley, Tanzania. It is anticipated that this information will provide farmers, policy and decision makers with evidence based, more flexible and superior tool to aid in improving rice production and curb the food insecurity problem.

1.2 Description of the Study Area

1.2.1 Geographical Location

In UTM coordinate system, the study site is located in zone 37 south, occupying the area lying between 9064697 and 9089031 m northings and 175422 to 197033 m eastings. It covers three wards of Kilombero district, namely Mngeta, Mchombe and parts of Mbingu. This area occupies a land of about 300 km^2 on the west side of the Kilombero Valley. The study area location is shown on Figure 1.1.

The Kilombero Valley, which is part of Rufiji Basin, covers Kilombero and Ulanga districts in Morogoro region. On the west and north-west this valley is bounded by an escarpment which rises abruptly from around 250 m above sea level to about 1400 m (Kato, 2007). The valley borders the Mahenge Mountain on the south east. The north part of the valley connects to the rest of the Rufiji Basin (Figure 1.2).

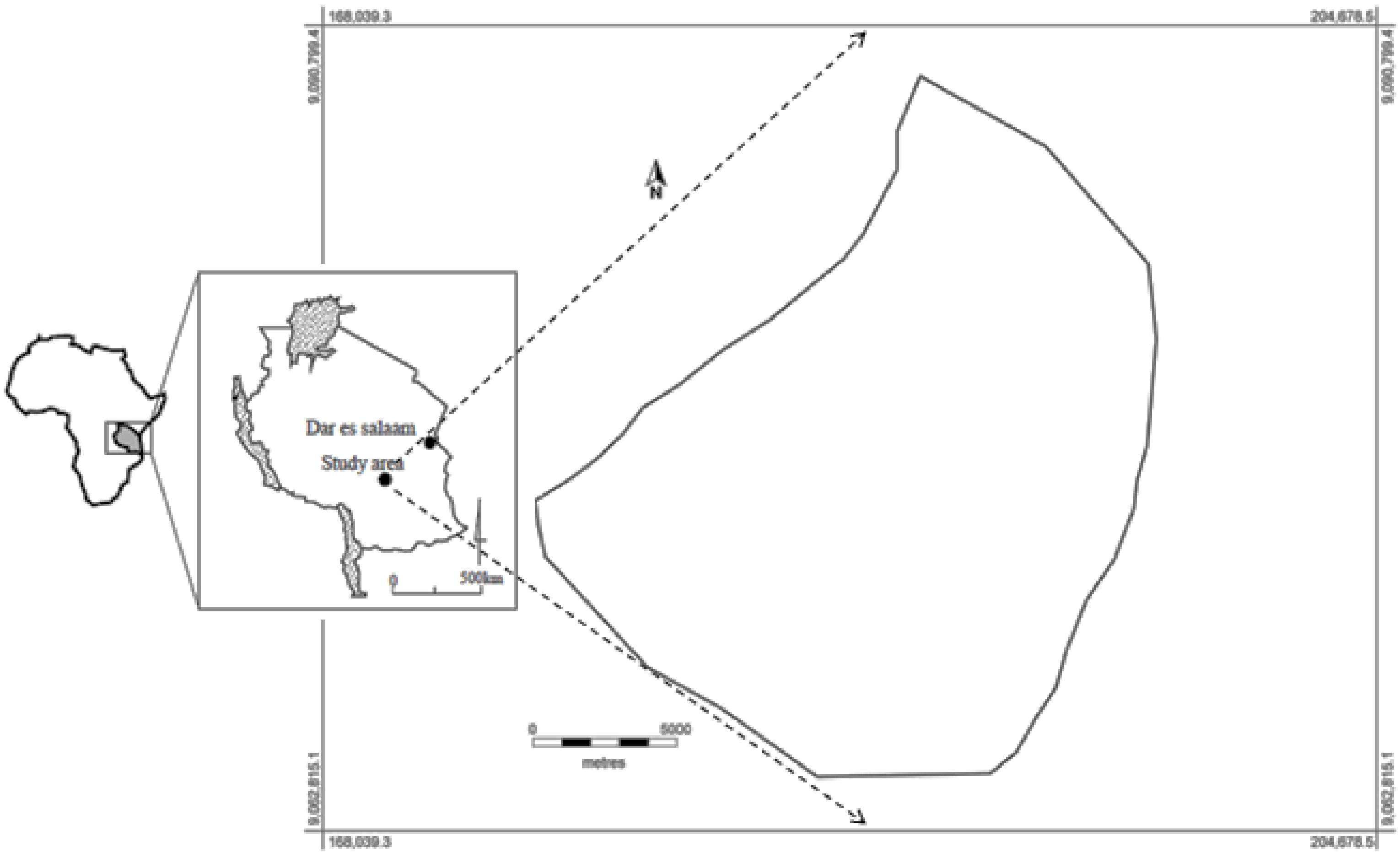


Figure 1.1. Study area location (Modified from Kato, 2007)

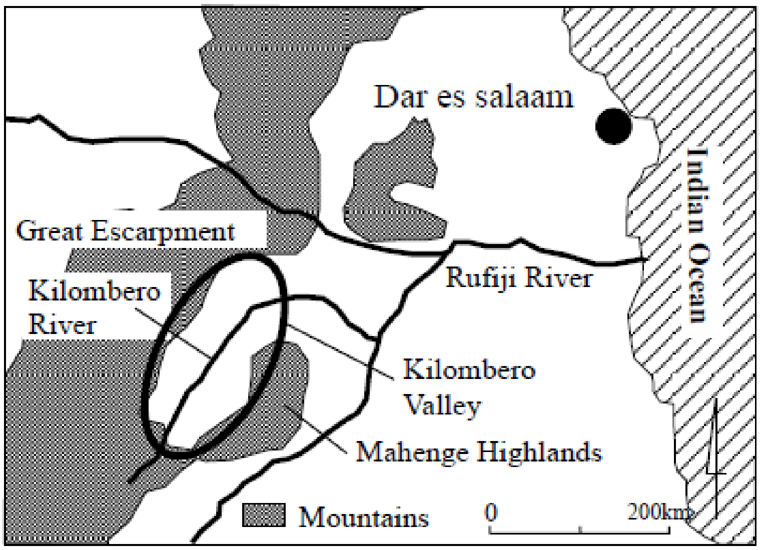


Figure 1.2. The Kilombero Valley as part of Rufiji Basin which stretches to the Indian Ocean (Adopted from Kato, 2007).

1.2.2 River Networks

The River Kilombero begins at the south-west side where Mpanga, Mnyera and Ruhuji rivers enter the valley. This end can be perceived as the beginning of the Kilombero Valley. Numerous permanent and seasonal rivers contribute to the Kilombero River. Important tributaries on the western bank of the Kilombero are Kihansi, Mngeta, Ruipa, Lumemo and Msolwa while Furua flow from the eastern bank. River Mngeta passes through the study area, while Kihansi is located a few kilometers south of the site. The rivers tend to build up alluvial fans because of lessening of the gradient. River networks drawn by Bonarius (1975) are shown on Figure 1.3.

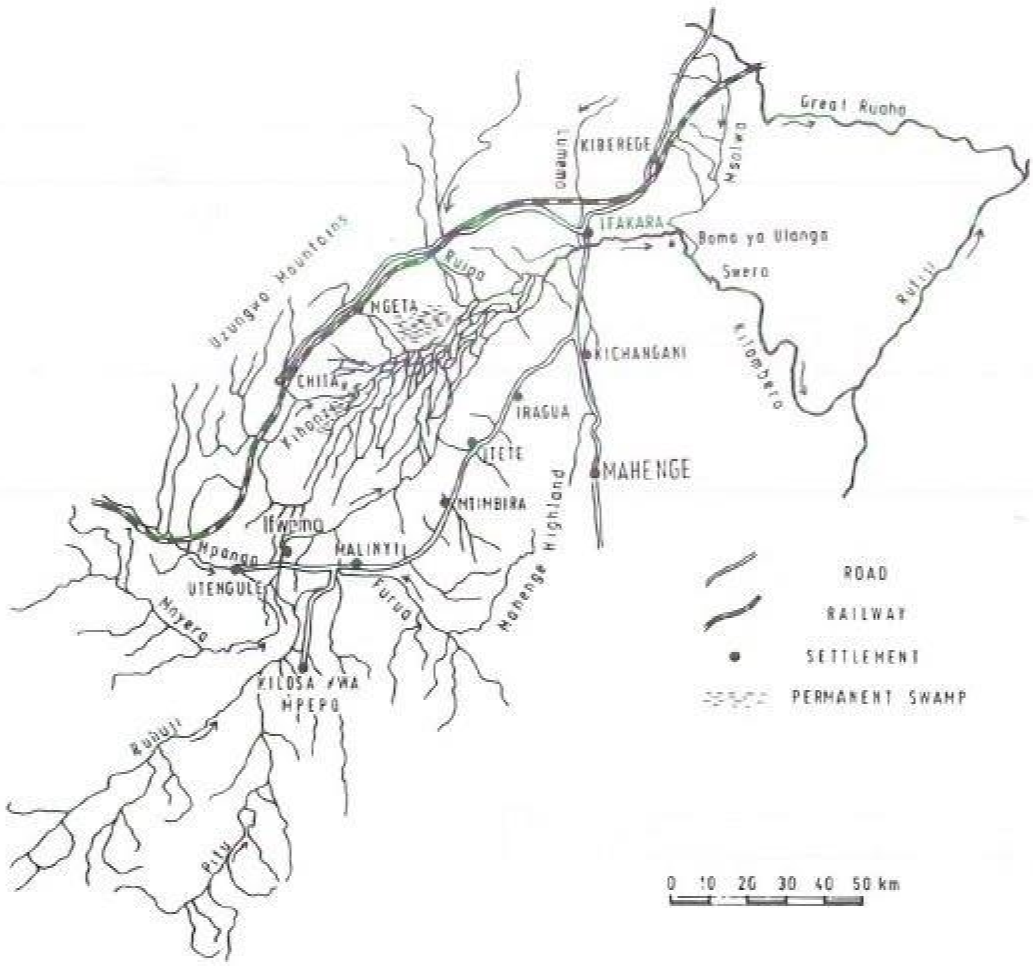


Figure 1.3. Major rivers of Kilombero Valley (Adopted from Bonarius, 1975). Arrows on the sides of the major rivers indicate directions of water flow.

1.2.3 Geology

The Kilombero Valley lies largely within the Usagaran System of the Basement Complex. In this region the Usagaran System consists chiefly of migmatitic biotite gneiss and acid granulites. There are occasional dyke rocks such as dolorites and mica pegmatites. The Usagaran rocks underlie the alluvium of the valley floor for a major portion of the valley. Generally, the rocks of the Kilombero Valley have been covered with Pliocene and Pleistocene deposits and alluvial materials of recent age (Beck, 1964). Soils of the valley are generally from the alluvial material.

1.2.4 Climate

The rainfall in the Kilombero Valley is higher closer to the escarpment and the Mahenge Mountains and decrease with distance to the mid of the valley (Kato, 2007). The effect is more pronounced where the valley is the widest. In the center of the valley, the annual rainfall is about 1000 mm while close to the escarpment the annual rainfall is about 1200 mm and 1800 mm in areas close to Mahenge Mountains. The rain falls mostly between December and April. Generally, the mean daily maximum and minimum temperature varies from 22° to 28°C, while the relative humidity is generally between 70% and 90%.

1.2.5 Vegetation

Major part of the study area is used for agriculture. Settlements are concentrated along the main road close to the escarpment. Some settlements are sparsely scattered within the valley on the fans with some grazing taking place around them or in the fields after crop removal.

Natural vegetation is dominantly tall grasses, mainly elephant grass (*Penisetum purpureum*), guinea grass (*Panicum maximum*), *Hyparrhenia spp*. and reed (*Phragmites mauritianus*) which cover protected areas close to the centre of the valley. Some trees such as *Borassus* palm, *Ficus spp*. and sausage tree (*Kigelia africana*) appear on alluvial fans. Teak trees (*Tectona grandis*), cocoa (*Theobrama cocoa*) and banana (*Musa spp*.) are among non-traditional crops which are now grown in drier parts of the valley. The miombo woodlands, mainly consisting of *Brachystegia spp*. are common in the south-west part of the valley and protected pieces of land close to the escarpment.

1.3 Dissertation Organization

This dissertation has seven chapters, each dedicated for a topic which builds up to the final objective: to develop suitability map for rice production in the study area.

The first chapter gives a general background and rationale of the study. It also describes the study area and also highlights on the organization of the dissertation.

The second chapter is dedicated on soil data collection. In this chapter, field work and lab work are reported. Also characterization and classification of soil pedons using USDA Soil Taxonomy (Soil Survey Staff, 2014) are reported.

Chapter 3 starts to introduce the concept of digital soil mapping (DSM). In this chapter, distance metrics are used to perform numerical classification to cluster the soil profiles and horizons. The work in this chapter uses information generated previously in chapter 2.

Chapter 4 reports on the predictive soil mapping. The soil clusters numerically classified in chapter 3, are spatially predicted and mapped using machine learning. Two decision tree algorithms: J48 and Random Forest are used to train the data derived from two digital elevation models (DEMs), satellite image and a legacy soil map. Predicted soil cluster maps from the two learning algorithms and the two DEMs are evaluated and compared to reveal a better soil map which is later used as a layer in multi-criteria land evaluation for rice suitability analysis.

Chapter 5 describes the process of identification of the criteria used in the land evaluation. The chapter also reports on the Analytical Hierarchy Process (AHP) which was used to assign weights and rank the criteria. It was also important to note how lead farmers and extension staff groups who were used in the ranking, differed in prioritizing the identified criteria.

In the sixth chapter, GIS is used to develop spatial data layers for identified land evaluation criteria. Overlays and reclassifications are also done. Using the AHP scores, multicriteria analysis was done in GIS using the identified criteria and the best predicted soil map. Suitability for rice production in the study area was developed in this chapter.

The last chapter summarizes major conclusions made from this study.

1.4 Reference

Balasubramanian, V., M. Sie, and R.J. Hijmans, K. Otsuka. 2007. Increasing rice production in Sub-Saharan Africa: challenges and opportunities. Advances in Agronomy 94:55-133.

Beck, A.D. 1964. The Kilombero Valley of south-central Tanganyika. East Africa Geography Review 2:37-43.

Bonarius, H. 1975. Physical properties of soils in the Kiolmbero Valley, Tanzania. German Agency for Technical Cooperation (GTZ), Germany.

Bucheyeki, T.L., E. Shennkalwa, D. Kadadi, and J. Lobulu. 2011. Assessment of rice production constraints and farmers preferences in Nzega and Igunga Districts. Journal of Advances in Developmental Research 2(1):30-37.

Carpenter, A.J. 1978. The history of rice in Africa. In: I.W. Buddenhagen, and G.J. Persley, editors, Rice in Africa. Academic Press, New York. p. 3–10.

Cimmery, V. 2007. User guide for SAGA (version 2.0). Vern Cimmery, Hamburg.

Cook, S.E., A. Jarvis and J.P. Gonzalez. 2008. A new global demand for digital soil information. In: A.A. Hartemink, A. McBratney, and M.L. Mendonca-Santos, editors, Digital soil mapping with limited data. Springer. p. 31-41.

De Data, S.K. 1989. Rice. In: D.L. Plucknett, and H.B. Sprague, editors, Detecting mineral nutrient deficiencies in tropical and temperate crops. Westview Press Inc. Boulder, CO. p. 41-51.

Deckers, J., O. Spaargaren, and S. Dondeyne. 2002. Soil survey as a basis for land evaluation. Encyclopedia of Life Support Systems.

Dobermann, A., and T. Fairhurst. 2000. Rice: nutrient disorders & nutrient management. Oxford Graphics Printers Pte Ltd. Singapore.

FAO 1985. Guidelines: Land evaluation for irrigated agriculture. FAO Soils Bulletin 55. Food and Agriculture Organization of the United Nations, Rome.

FAO. 2004. International year of rice 2004: rice is life. FAO, Rome

Haifa. 2008. Nutritional recommendations for rice. Haifa, Israel.

Indexmundi. 2014. United Republic of Tanzania, milled rice production by year. http://www.indexmundi.com/agriculture/?country=tz&commodity=milled-rice&graph=production. Last visited May 2014.

Kato, F. 2007. Development of a major rice cultivation area in the Kilombero Valley, Tanzania. African Study Monographs, Supplement 36: 3-18.

Kuria, D., D. Ngari, and E. Waithaka. 2011. Using geographic information systems (GIS) to determine land suitability for rice crop growing in the Tana delta. Journal of Geography and Regional Planning 4(9):525-532.

Ly, P., L.S. Jensen, T.B. Bruun, D. Rutz, and A. de Neergaard. 2012. The system of rice intensification: Adapted practices, reported outcomes and their relevance in Cambodia. Agricultural Systems 113:16–27.

Malone, B. 2013. Use of R for digital soil mapping. Soil Security Laboratory, University of Sydney, Australia.

Marinoni, O., and A. Hoppe. 2006. Using the analytical hierarchy process to support sustainable use of geo-resources in metropolitan areas. Journal of Systems Science and Systems Engineering 15(2):154-164.

Massawe, B.H.J., and N.A. Amury. 2012. Report on agronomic practices and soil fertility analysis for improved rice production in Kilombero and Wami Valleys in Tanzania. USAID Feed the Future NAFAKA staples value chain activity project. Morogoro, Tanzania.

McBratney, A.B., M.L. Mendonca-Santos, and B. Minasny 2003. On digital soil mapping. Geoderma 117:3–52.

Meertens, H.C.C., L.J. Ndege, and P.M. Lupeja. 1999. The cultivation of rainfed, lowland rice in Sukumaland, Tanzania. Agriculture, Ecosystems and Environment 76:31–45.

Mendas, A., and A. Delali. 2012. Integration of MultiCriteria Decision Analysis in GIS to develop land suitability for agriculture: application to durum wheat cultivation in the region of Mleta in Algeria. Computers and Electronics in Agriculture 83:117–126.

Moore, I.D., P.E. Gessler, G.A. Nielsen, and G.A. Peterson. 1993. Soil attributes prediction using terrain analysis. Soil Science Society of America Journal 57: 443–452.

Msanya, B.M., J.P. Magoggo, and H. Otsuka. 2002. Development of soil surveys in Tanzania. Pedologist 46(1):79-88.

Pan, G. and J. Pan. 2012. Research in crop land suitability analysis based on GIS. Computer and Computing Technologies in Agriculture 365:314–325.

Perveen, F., R. Nagasawa, A.O.C. Ahmed, I. Uddin, and R. Kimura. 2008. Integrating biophysical and socio-economic data using GIS for land evaluation of wheat cultivation: A case study in north-west Bangladesh. Journal of Food, Agriculture and Environment 6(2):432-437.

Rural Livelihood Development Company-RLDC. 2011. Rice sector strategy. Dodoma, Tanzania.

Saaty, T.L. 1980. The Analytic Hierarchy Process, McGraw-Hill, New York.

Sahrawat, K.L. 2012. Soil fertility in flooded and non-flooded irrigated rice systems. Archives of Agronomy and Soil Science 58(4):423-436.

Soil Survey Staff (2014), Keys to soil taxonomy. 12th Edition. USDA, Natural Resource Conservation Service, USA.

Subburayalu, S.K. and B.K. Slater. 2013. Soil series mapping by knowledge discovery from an Ohio county soil map. Soil Science Society of America Journal 77:1254–1268.

TIC. 2013. Tanzania investment guide 2013-14. Tanzania Investment Center, Dar es Salaam.

Chapter 2: Characterization and Classification of Soils of Kilombero Valley

2.1 Introduction

2.1.1 Soil Characterization and Classification

The importance of soil as part of landscape lead to the quest of understanding its nature, properties, dynamics and functions in relation to the support of life and and ecosystems in general (FAO, 2006). To be able to achieve this, reliable information on soil morphological, chemical, mineralogical, and physical characteristics are needed. These properties are obtained through examination and description of the soil in the field and laboratory analysis of representative samples.

One of the biggest users of soil information is agriculture. Apart from being used in areas where agriculture is already intensified, soil information is needed to facilitate the intensification of agriculture in underdeveloped and underutilized lands. The information enables analysis of the potential and constraints of major soils, thus facilitating designing of the most appropriate soil management systems (Mueller et al., 2010; Kihoro et al., 2013). The entry point for this is characterization and classification of the soils of the intended area.

Different systems of soil classification to group soils of similar properties in one class have been developed. The widely used are two: the USDA Soil Taxonomy (Soil Survey Staff, 2014) and the FAO World Reference Base for soil resources (IUSS Working Group WRB, 2014). Soil profiles, which are composed of a series of horizons, are crucial to soil characterization and classification of soils. Recognizing, naming, recording morphological properties and sampling for further analysis of physical and chemical attributes of the soil profile horizons are the activities which when completed give data for soil classification. Individual soils have one or more soil horizons, but very young soils may not yet have started the soil horizonization process (Schaetzl and Anderson, 2010; Brady and Weil, 2010).

2.1.2 Importance of Alluvial Soils

The soils of Kilombero valley are developed on an alluvial basin (FAO, 1961; Bonarius, 1975) and can generally be categorized as alluvial soils. Alluvial soils are normally supplied with excessive water from different sources such as rainfall and flooding from streams and rivers. Because of poor drainage, they normally show redoximorphic features in parts or throughout the soil profile (Schaetzl and Anderson, 2010). Their texture may range from sands to clays and their profiles may range from underdeveloped to strongly developed (Egbuchua, 2011).

The distribution of soils in the floodplain environment is highly complex and the extreme spatial variability of alluvial soils, in both the vertical and lateral dimensions, creates challenges to adequately capture the range in physical and chemical properties values that are important to characterize the soil suitability for a particular land use type (Iqbal et al., 2005; Alves and Ross, 2010). This is particularly due to constant input of new sediment leading to poor conformity of the alluvial soils to the standard ABC model of profile development (Alves and Ross, 2010). In the USDA Soil taxonomy (Soil Survey Staff, 2014), most alluvial soils have been classified as Entisols, Inceptisols, Alfisols, Ultisols or Vertisols (Egbuchua, 2011).

Despite their complex nature, alluvial soils are among the most important soils especially in arid and semi-arid areas. This is because they are associated with water and are able to support agriculture, wildlife, and aquatic life to feed the people and the economy. Many of the earlier civilizations were developed upon alluvial soils of arid or semiarid regions. Holmes and Hearn (1942) write that much of the wealth of Babylonia was derived from the alluvial soils of the Tigris and Euphrates Rivers. The authors add that, "the ancient people of Babylon appreciated the value of their alluvial soils, expending much of the wealth derived from them in building a vast system of irrigation which won for the country the title of 'granary of the world'". Another example is the Egypty. The ancient Egyptian civilization and present economy were and are still largely supported by the alluvial soils along the Nile (Gill, 2003). Likewise, the alluvial soils of China associated with Yellow River, Hwai Ho, and Yangtze Kiang, constitute the most important agricultural soils in China; supporting a significant part of China's population (Lin, 1997).

2.1.3 Justification and Objectives

Given the importance of alluvial soils to the past and present societies, there is need to investigate these soils in Tanzania for the purpose of tapping its potential. The Kilombero Valley, which is a part of Rufiji Basin in Tanzania, covering an area of about 11,600 km^2 (Kato, 2007), presents great potential for expansion and intensification of agricultural production. Apart from the reconnaissance work done to map the valley in 1959 for irrigation suitability (FAO, 1961) and characterization of a portion of Chita National Service rice farm by Msanya and Meliyo (personal communication), no any documented characterization of the valley has come to our attention. Massawe and Amuri (2012) used composite top soil samples to assess nutrient availability status for rice production in major part of the eastern Kilombero, but did not use soil profiles, and hence were unable to study the subsoils and classify the soils.

The objective of this work is to characterize and classify soils of Mngeta, Mchombe, and Njage areas of Kilombero valley in order to provide soil information which will be used to assess suitability analysis of the study area for rice production using GIS and predictive soil mapping techniques in later stages of this dissertation.

2.2 Methods

2.2.1 Study Site

The study was conducted in Kilombero Valley, Tanzania. The valley is about 300 km east of Indian Ocean covering about 11000 km^2. The study site covered about 300 km^2 within the valley. In UTM coordinate system, the study site is located in zone 37 south, occupying the area lying between 9064697 and 9089031 m northings and 175422 to 197033 m eastings. More description of the study area is given in the introductory chapter of this document.

2.2.2 Pre-field Works

A legacy soil map developed by Anderson in the later 1950s was used as the base map (FAO, 1961). The soil units of this map were delineated using aerial photo interpretation.

The map was extracted from a FAO (1961) report and georeferenced in QGIS software (QGIS Development Team, 2014). The study site soil polygons were extracted by clipping it with a previously prepared polygon covering the boundaries of the study area. The soil units were then digitized using on-screen digitization tool in ArcMap 10.1 software (ESRI, 2010). The digitized soil Anderson's legacy map is shown on Figure 2.1.

Anderson grouped the Kilombero valley soils into four major categories: alluvial soils, non-alluvial soils, soils from composite profiles in which a shallow layer of alluvium overlies alluvium of a different texture or else non alluvial material, and lastly the soils which did not fit exactly into the above categories. The last group includes soil groups which Anderson adopted from an earlier work done by R. F. Loxton in 1953 - 1954 (FAO, 1961). Anderson mapped a total of 33 soil units in Kilombero Valley from the four categories.

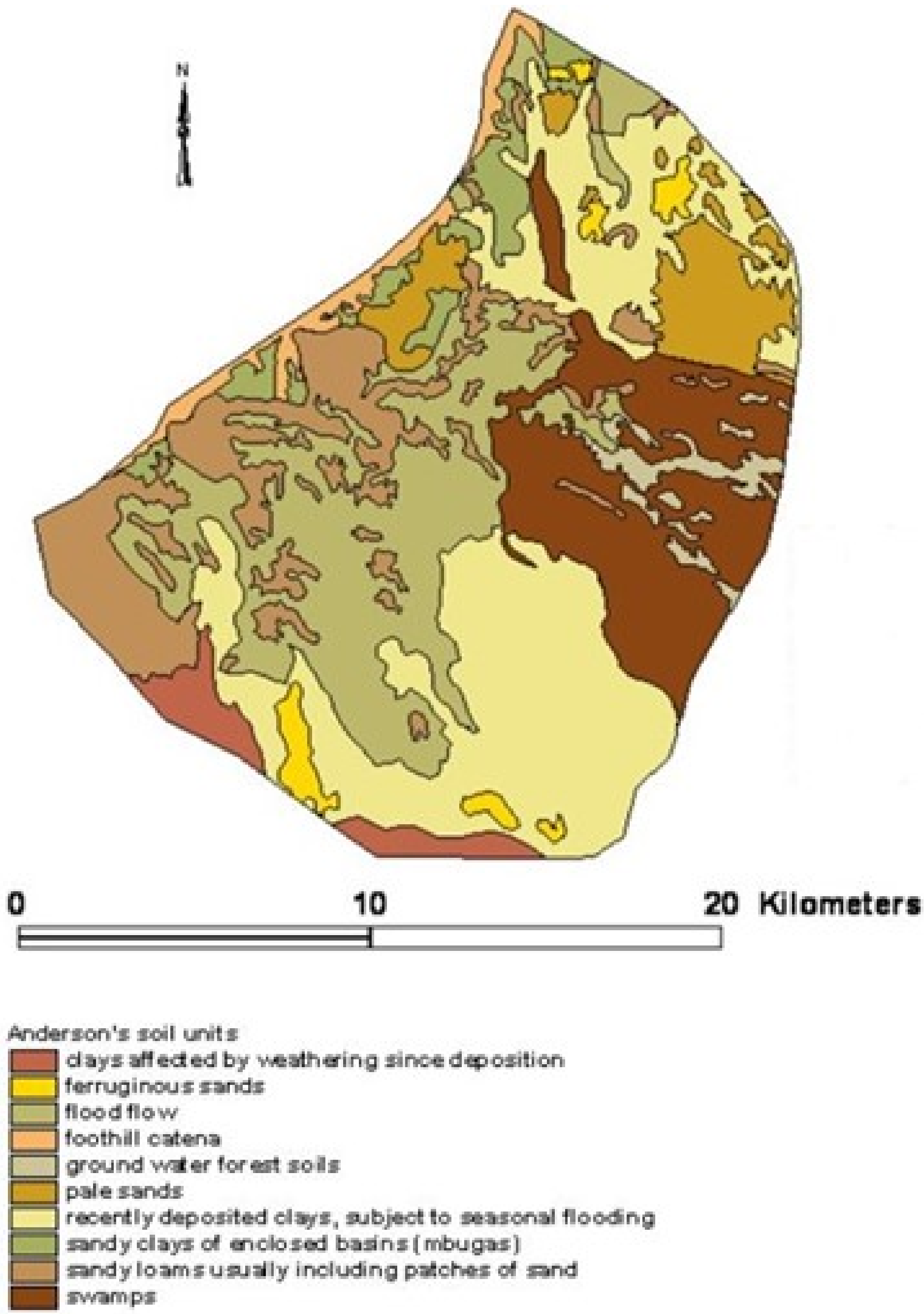


Figure 2.1. Anderson's reconnaissance soil units of the study area. The soil units were digitized from FAO report (FAO, 1961)

The study site has nine soil groups from three of the above four categories.

From Anderson's alluvial soils group the study site has the following units:

i) recently deposited clays, subject to seasonal flooding (K),

ii) sandy loams, usually including patches of sand (A), and

iii) clays affected by weathering since deposition (C).

From non-alluvial groups the units are:

i) pale sands without permanent water table and not developed from underlying sediment (P), and

ii) sandy clays of enclosed basins (D).

From the units adopted from Loxton's work, the study area contained:

i) ferruginous sands (U),

ii) flood flow soils (F), and

iii) ground-water forest soils (W).

A significant portion of the study area was mapped as 'swamp' and no soil description was given in the legacy map. During the field visit it turned out that this area has been converted into agriculture and can no longer be categorized as swamp. This area was therefore treated as an independent unit on the base map.

In addition to the base map, the 1 arc (30m resolution) Shuttle Radar Topography Mission (SRTM) terrain model (USGS, 2000) was used to visualize differences in elevation in the study area (Figure 2.2) in ArcView 3.2a (ESRI, 2002). The visualization aided in deciding where else to add soil observations so as to be able to capture the soil variability related to landscape positions. The elevations are generally observed to be decreasing from the escarpment towards the center of the valley. The decrease towards the center of the valley is also observed from the other side of the river, whereas, the landscape is generally sloping down from the Mahenge highlands. Figure 2.3 shows a typical West – East profile of the valley from both sides of river.

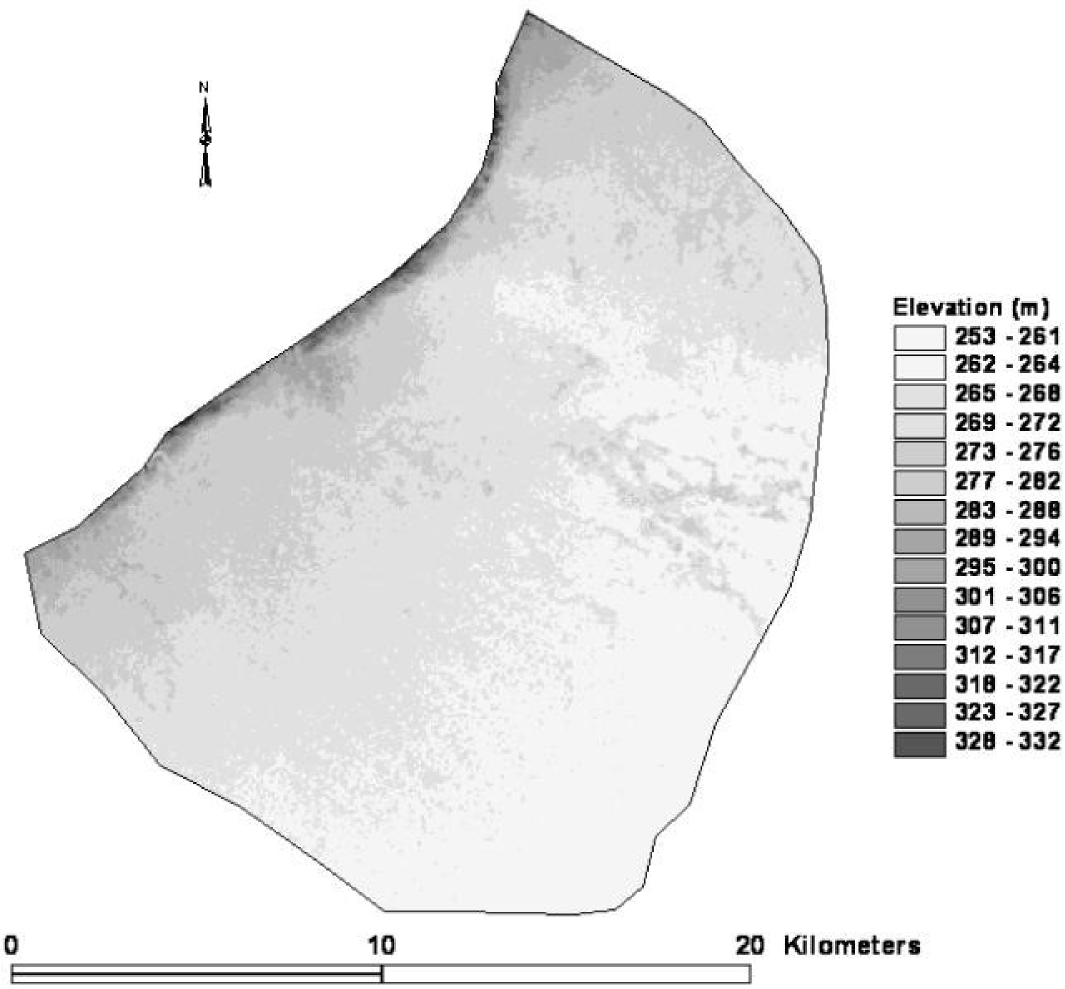


Figure 2.2. Change in elevation within the study area

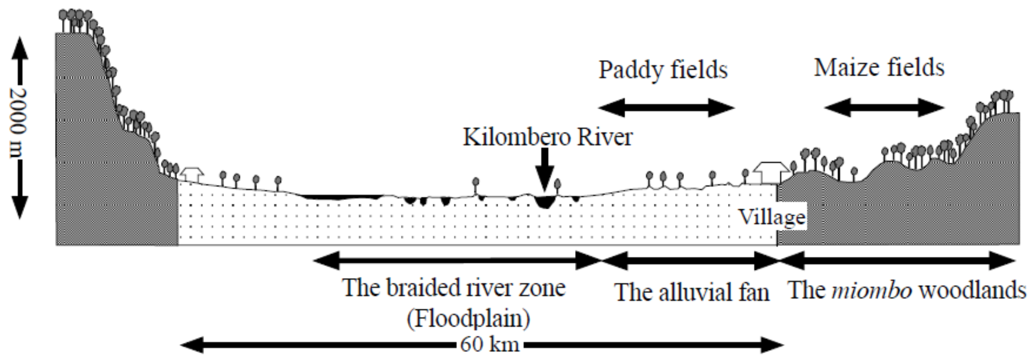


Figure 2.3. Typical West – East profile of the Kilombero Valley (adopted from Kato, 2007).

2.2.3 Field Description and Soil Sampling

A total of 33 soil profiles were excavated from the study site. Additional nine soil profiles were excavated in other parts of the valley during the reconnaissance phase of the study. The number of soil profiles in the study site exceeded the number of soil groups in the legacy soil map because it was necessary to describe and sample on areas which appeared not to be well represented with the base map. There has been some land cover/ land use change since the preparation of the legacy map in the late 1950s such as deforestation and draining of swampy areas for agriculture, grazing and settlements. Figures 2.4 and 2.5 show the locations of observations and sampling points on Anderson's reconnaissance soil units and the SRTM DEM determined elevation ranges.

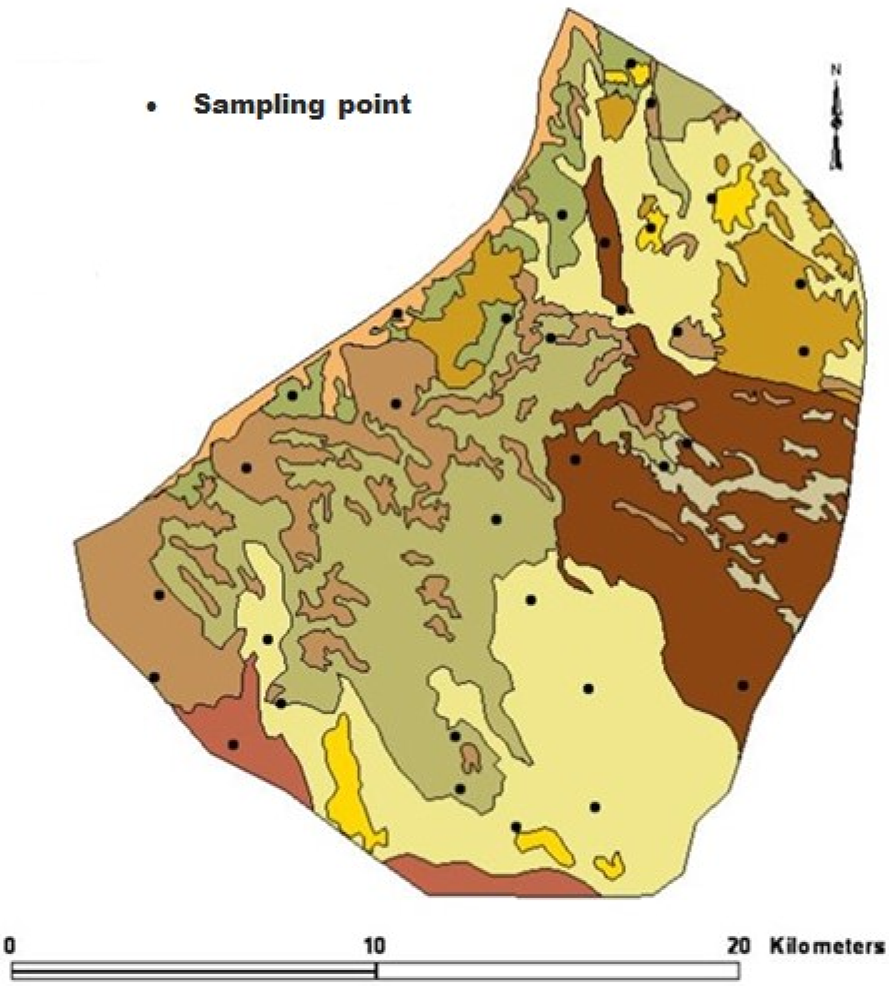


Figure 2.4. Locations of observations and sampling points on Anderson's reconnaissance soil map

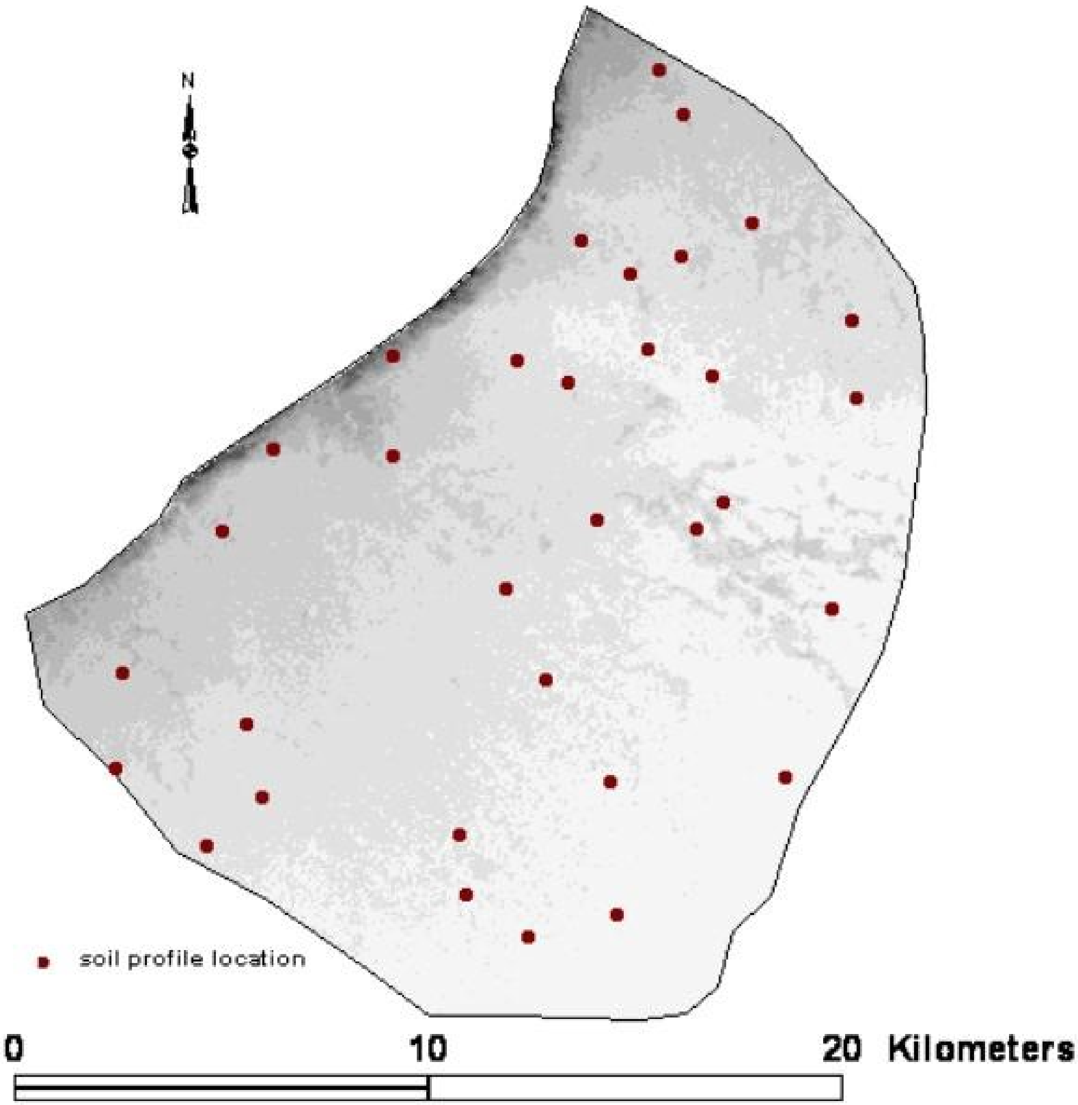


Figure 2.5. Locations of observations and sampling points with respect to elevation ranges

Every soil profile site was georeferenced. Site characteristics including land use, elevation, vegetation, slope characteristics, and flooding were recorded in a prepared field data sheet. Profile pits were dug to a depth of 120 cm or to a limiting layer. In most cases it was not possible to dig below 80 cm because of the high water table. The soil profile pits were studied and described according to FAO Guidelines for Soil Description (FAO, 2006). Soil profile morphological characteristics studied in the field included soil colour, texture, consistence, structure, porosity, effective depth, presence or absence of cutans, mottles, concretions and type of primary minerals and rock fragment. Soil colour was determined by Munsell soil colour charts (Munsell Colour Co., 1992) whereas soil texture was estimated by feel method (Thien, 1979).

In each soil profile pit, bulk soil samples were taken from designated natural horizon for physical and chemical analysis in the laboratory.

2.2.4 Laboratory Analysis

Soil attributes which were analyzed in the laboratory are soil pH, electrical conductivity (EC), soil texture, total nitrogen (TN), organic carbon (OC), available phosphorus (P), exchangeable potassium (K), exchangeable sodium (Na), exchangeable magnesium (Mg)and exchangeable calcium (Ca). Others included cation exchange capacity (CEC), extractable manganese (Mn), extractable iron (Fe), extractable copper Cu), and extractable zinc (Zn).

The soil pH was determined using pH meter in water and in $CaCl_2$ at the ratio of 1:2.5 soil-water and soil-$CaCl_2$ as described by McLean (1982) while electrical conductivity (EC) was determined by conductivity meter in a 1:2.5 soil:water suspension following a method by Rhoades (1982). Soil texture was determined by hydrometer method using calgon (5%) as a dispersing agent (Gee and Bauder, 1986). Organic carbon was determined by the Walkley and Black wet oxidation method as outlined by Nelson and Sommers (1982). The total nitrogen in the soil samples was determined by Kjeldahl method (Bremner and Mulvaney, 1982). Available phosphorus was extracted by Bray and Kurtz-1 method (Bray and Kurtz, 1945) for soils with pH_{water} less than 7 and Olsen method for soils with pH_{water} above 7 (Watanabe and Olsen, 1965). Cation exchange capacity of the soil (CEC) and exchangeable bases were determined by saturating soil with neutral 1M NH_4OAc (ammonium acetate) and the adsorbed NH_4^+ were displaced using 1M KCl. The exchangeable bases (Ca^{2+}, Mg^{2+}, Na^+, K^+) were determined by atomic absorption spectrophotometer (Thomas, 1982) while CEC was determined by Kjeldahl distillation method Schollenberger and Simon (1945). Diethylenetriaminepenta-acetic acid (DTPA) method (Lindsay and Norvell, 1978) was used to extract four micronutrients: iron, manganese, copper and zinc.

CEC:Clay ratio for each sample was determined as the ratio of CEC to clay, while percent base saturation (PBS) was calculated by multiplying the total exchangeable bases (Ca^{2+}+ Mg^{2+}+ Na^{+}+ K^{+}) by 100 and then divided by CEC (Landon, 1991).

2.2.5 Statistical Analysis

Microsoft Excel and Minitab (Minitab, 2004) packages were used for qualitative assessment and descriptive statistical analyses. Anderson-Darling statistic test normal plot graphs were used to determine whether the data met the assumption of normality. To attain normal distribution, log-transformation was done on CEC, Clay, K, Mg, Mn, Na, OC, P, PBS, TN, and Zn data, while Ca, Cu, and soil depth were transformed using Square-root function. Root one-third was used for Fe and silt content transformation, while soil pH, and sand content were not transformed because they were normally distributed.

Spearman correlation analysis was conducted to identify relationships among the studied soil properties. Significance level $p < 0.05$ was used to assess the significance of the Spearman correlations.

Equal area spline functions (Bishop et al, 1999) were employed to model vertical variation of selected soil attributes with soil depth at an interval of 1 cm.

Principal component analysis was conducted to assess variables which are highly correlated with the principal components which accounts for 73% of the data variability.

2.2.6 Soil Classification

Using field and laboratory data, the soils were classified to subgroup level of the Soil Taxonomy (Soil Survey Staff, 2014). The sand, silt and clay fractions were used to group the soil into textural classes using USDA textural triangle (Schoeneberger et al., 2012).

2.3 Results and Discussion

2.3.1 Segmentation of the Study Area

The terrain analysis of the study area revealed that elevations were decreasing somewhat parallel from the northwestern side towards the center of the valley where the Kilombero River is located (Figure 2.2). Using this trend, six landscape levels were identified with the first one occupying the highest altitude range and the sixth occupying the lowest (Figure 2.6). The study area lies in an elevation between 253 and 388 m asl, giving a range of 135 m from the lowest to the highest point. These broad landscape levels did not solely guide sampling locations, initially, but they finally appear to be a good reference in discussing the characterization results. The landscape levels and number of observations falling on each are shown in Table 2.1.

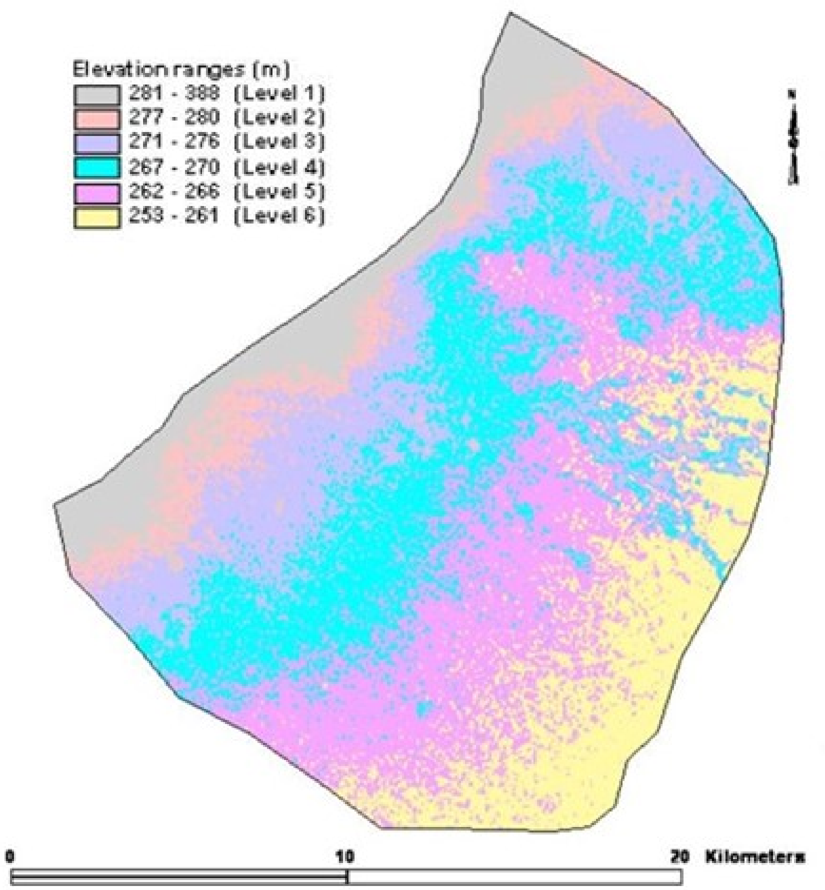


Figure 2.6. The six landscape levels demarcated on the study area based on elevation ranges

Table 2.1. Number of soil observations in each dermacated landscape level of the study area

Landscape level	Number of observations (soil profiles)	Elevation range (m)
1	3	281 - 388
2	4	277 - 280
3	5	271 - 276
4	6	267 - 270
5	12	262 - 266
6	3	253 - 261

Landscape level 1 occupied areas in the study site which have elevation ranges between 281 and 388 m asl. The altitudes of the lowest and highest points of this landscape differ by 107 m. This elevation range accounts for about 79% of the difference in altitude between the highest and lowest point in the entire study area. Despite its high contribution in elevation range, this landscape level is estimated to occupy only about 10% of the study area. It is therefore, the steepest part of the study area. Three soil sampling and observation points fall in this landscape level (Table 2.1).

Four soil observation and sampling points fall in the second landscape level. The lowest point in this landscape level is 277 m asl while the highest is 280 m, giving an altitude range of 4 m. It is a relatively flat elongated area immediately following the steepest landscape of the study site (Figure 2.6). It covers the smallest area compared to other landscape levels and it is an immediate transition zone between the steepest part and the flatter part of the study area.

The third Landscape level occupies an area with altitude between 271 and 276 m asl. Five soil observation and sampling points fall in this level. The altitude range between the highest and the lowest point is 5 m, but it occupies a wider area than Landscape level 1 which has an altitude range of 107 m. It can be deduced from these observations that this landscape level has less slope gradient compared to the first two levels.

Five soil sampling points are located in level 4 landscape. The altitude range in this level is 4 m, with 267 and 270 m asl being its lowest and highest points respectively. This level has lower slope gradient compared to the three above, and at points dissects the lower landscape levels (levels 5 and 6). It also occupies a larger area compared to each of the higher landscape levels (Figure 2.6).

Landscape level 5 occupies the area with elevations between 262 and 266 m asl. Twelve soil sampling sites fall into this level. With only a range of 4 m between the highest and the lowest points, this landscape level is estimated to be the largest and its flatness is only exceeded by landscape level 6. With its broad surface area and low slope gradients, deposition of finer materials by fluvial activities is expected to dominate especially with increased distance from side channels and streams.

The lowest positions in the study area are occupied by Landscape level 6. This level is occupying altitudes between 253 and 261 m asl, with 9 m being the difference between the highest and lowest points. Three soil sampling and observations done during the study fall in this area. This level is also closest to the center of the valley, and borders the conserved area were agriculture is not allowed.

2.3.2 Soil Texture

Soil Texture Variation in the Valley

USDA soil textural triangle groups soil separates into twelve classes (Schoeneberger et al., 2012). The Kilombero Valley soils were classified into seven of these classes. These were clay (C), clay loam (CL), loamy sand (LS), sand (S), sandy clay (SC), sandy clay loam (SCL), and sandy loam (SL) as shown in Table 2.2. The dominant soil textural classes are sandy clay loams and sandy loams. Of 152 studied and analysed soil horizons from 42 soil profiles, 38% of them was classified as sandy loam and 28% as sandy clay loams. The two classes accounted for 66% of the studied horizons. Clays accounted for 13% and sands accounted for only 3%. These soils are therefore dominated by medium to coarse soil separates.

Due to dominance of medium to coarse soil fractions, higher infiltration rates and hydraulic conductivities, as well as leaching of base cations are expected. Bonarius (1975) reported that the infiltration rates of Kilombero Valley soils vary widely but can generally be classified as rapid (10.0 – 25.0 cm/hr) to very rapid (> 25.0 cm/hr) using infiltration rates classification by Kohnke (1968). According to his study, soils with clayey textures had significantly lower infiltration rates. Positive correlation between infiltration rates and hydraulic conductivities were also observed in this valley in the same study. The workability of these soils is also expected to be generally good compared to finer textural class soils.

Table 2.2. Soil textural classes of the studied soil profiles horizons in Kilombero Valley

Horizon soil textural class	Percent of total number of studied horizons (n = 152)
Clay (C)	13
Clay loam (CL)	2
Loamy sand (LS)	9
Sand (S)	3
Sandy clay (SC)	7
Sandy clay loam (SCL)	28
Sandy loam (SL)	38

Soil Texture Variation in the Study Area

Textural classes for each landscape level are displayed on Table 2.3. As expected, soil textures appeared to generally become finer down the gradient, which is typical for alluvial fans. Landscape level 1 contains 85% sandy loams and loamy sands, with the finest being sandy clay loams accounting for 15% of the observations. In landscape level 2, sandy loams and loamy sands account for 68% of observation while clay loams and sandy clay loams accounts for 29% without any horizon with the clay (C) textural class. We start observing clayey horizons in landscape level 3 which accounts for 14% of the observations, increasing to 23% in landscape level 4 and 36% in the lowest landscape level (landscape level 6). This could be explained by decrease in flow rates in the rivers and side channels. High water flow rates deposit coarser materials while slow moving water deposits finer materials (Sweeney et al., 2013). The flow rate of water in the rivers and channels decreases with decrease in slope gradient and increase in river width (Gervais-Beaulac et al., 2013), as is the case in Kilombero Valley.

Table 2.3. Soil textural classes for each of the six demarcated landscape levels

Landscape Level	Number of studied horizons	Soil textural class	Percent of total number of studied horizons
1	13	Loamy sand (LS)	23
		Sandy clay loam (SCL)	15
		Sandy Loam (SL)	62
2	21	Clay loam (CL)	5
		Loamy sand (LS)	14
		Sand (S)	5
		Sandy clay loam (SCL)	24
		Sandy loam (SL)	52
3	21	Clay	14
		Loamy sand (LS)	10
		Sandy clay loam (SCL)	14
		Sandy loam (SL)	62
4	26	Clay C	23
		Clay loam (CL)	4
		Loamy sand (LS)	4
		Sandy clay (SC)	8
		Sandy clay loam (SCL)	35
		Sandy loam (SL)	27
5	57	Clay C	9
		Clay loam (CL)	2
		Loamy sand (LS)	7
		Sand (S)	7
		Sandy clay (SC)	12
		Sandy clay loam (SCL)	32
		sandy loam (SL)	32
6	14	Clay C	36
		Sandy clay (SC)	21
		Sandy clay loam (SCL)	43

The coarse material observed in the lower landscape levels despite dominance of finer materials, could be explained by the presence of side streams. The soils closer to the side streams are coarser than those away from it (Kemker, 2014). Another explanation could be the braided shape of the channels at the center of the valley. Braided channel shapes consist of bar deposits. Bar deposit happen when a stream deposit consists of sand or gravel deposited in the center of the channel. Bars form when the stream's velocity or discharge decreases and its bedload is dropped. Braided rivers deposits coarser materials at the bars and finer materials where there are no bars. This could also explain a mixture of soil textural classes near the center of the valley.

An analysis of variation of clays with depths in studied pedons for all landscape levels is depicted in Figure 2.7. The variations in clay contents in landscape levels and vertically down the soil profiles can be explained by factors which affect dynamics of sediments. Sediment refers soil-based, mineral matter (e.g. clay, silt and sand), and decomposing organic substances and inorganic biogenic material that can be carried out with water or other moving agent like ice and wind (Kemker, 2014). Materials carried by water can be divided into suspended load and suspended sediments. The two seem to overlap, but according to Southard (2006) suspended sediment are any particles found in the water column, whether the water is flowing or not, while the suspended load is the amount of sediment carried downstream within the water column by the water flow. The size of the particles that can be carried as suspended load is dependent on the flow rate. Suspension of larger particles requires faster flow rates.

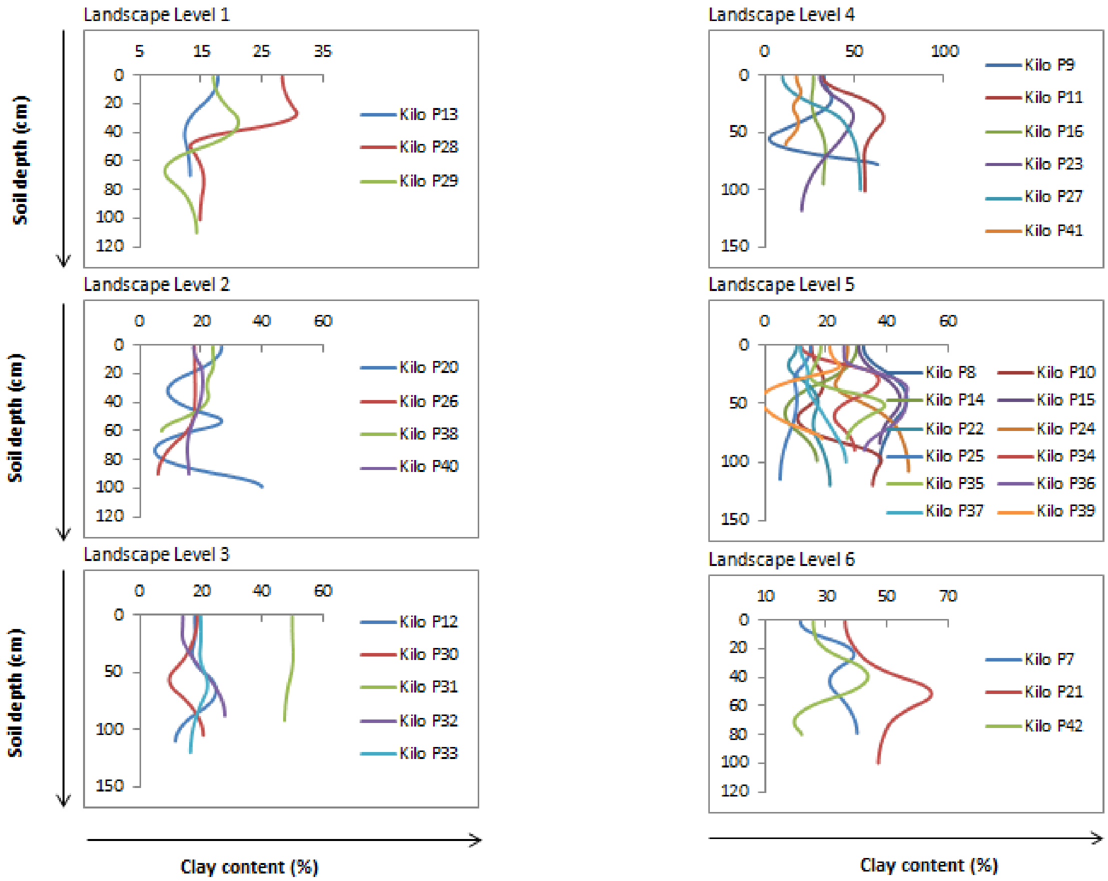


Figure 2.7. Clay contents variations with soil depth for each dermacated landscape level.

The keys (Kilo Pxx) represent codes of soil profiles studied in each level

In landscape level 1 soils, topsoils have more clays than subsoils, particularly below 35 cm. This could be explained by change of rivers and streams depths with time, thus affecting flow rates and size of material suspended and deposited. Normally, over time, streams or rivers with fast flow rates will erode their beds. This reduces the gradient and flow rates of the channel resulting to finer materials being suspended and deposited over the coarser materials which were deposited when the streams flow rates were faster.

In landscape level 2, there appear to be low variability of clay content with soil depth except for pedon Kilo P20. This pedon could be located in an area which has been receiving cyclic deposition of materials having variable levels of clay content. From level 3 landscape, the trends of clay content start to increase with depth. This is not very obvious in Level 3 most probably because the rate of clay deposition on the topsoil is lower due to relatively faster stream flow rate compared to the lower levels (level 4, 5, and 6). Some pedons show repeated patterns of increase and decrease suggesting that the higher levels of clay contents in the subsoils could be a result of both clay illuviation and cyclic deposition of materials carrying varying amounts of clays with them.

2.3.3 Soil pH

Soil pH in the study area varies widely spatially and vertically (Figure 2.8). It ranges from 4.7 to 7.2, with topsoil values predominantly being around 5.8. Following soil pH classifications by Soil Survey Staff (1993) the Kilombero valley soils can be generally classified as moderate acidic (pH ranges between 5.6 and 6.0). These pH levels are not limiting to many crops and soil microbial performances. Most upland crops perform normally in pH ranges between 5.6 and 7.2 (Hoskins, 1997), while lowland crops like flooded rice grown in Kilombero can perform normally even in pH levels lower than the range given above. Flooding rice soils have been documented to moderate the pH towards a neutral pH condition (Haifa, 2008).

The soil pH does not show strong correlation with any of the other determined parameters (Table 2.4). Weak negative correlations are observed with organic carbon (Pearson correlation, r = -0.41) and total nitrogen (r = -0.47). A weak positive correlation (r = 0.42) between pH and base saturation is also observed.

In the Landscape level 1, topsoils pH values vary from 5.6 to 6.9, but converge in the subsoil to around 6.3 (Figure 2.8). In landscape levels 3, 4, 5, and 6, pH values generally tended to increase with soil depth. The increase could be attributed to accumulation of leached base cations from the topsoils. Soil pH tend to increase with increase in base cations (Brady and Weil, 2010)

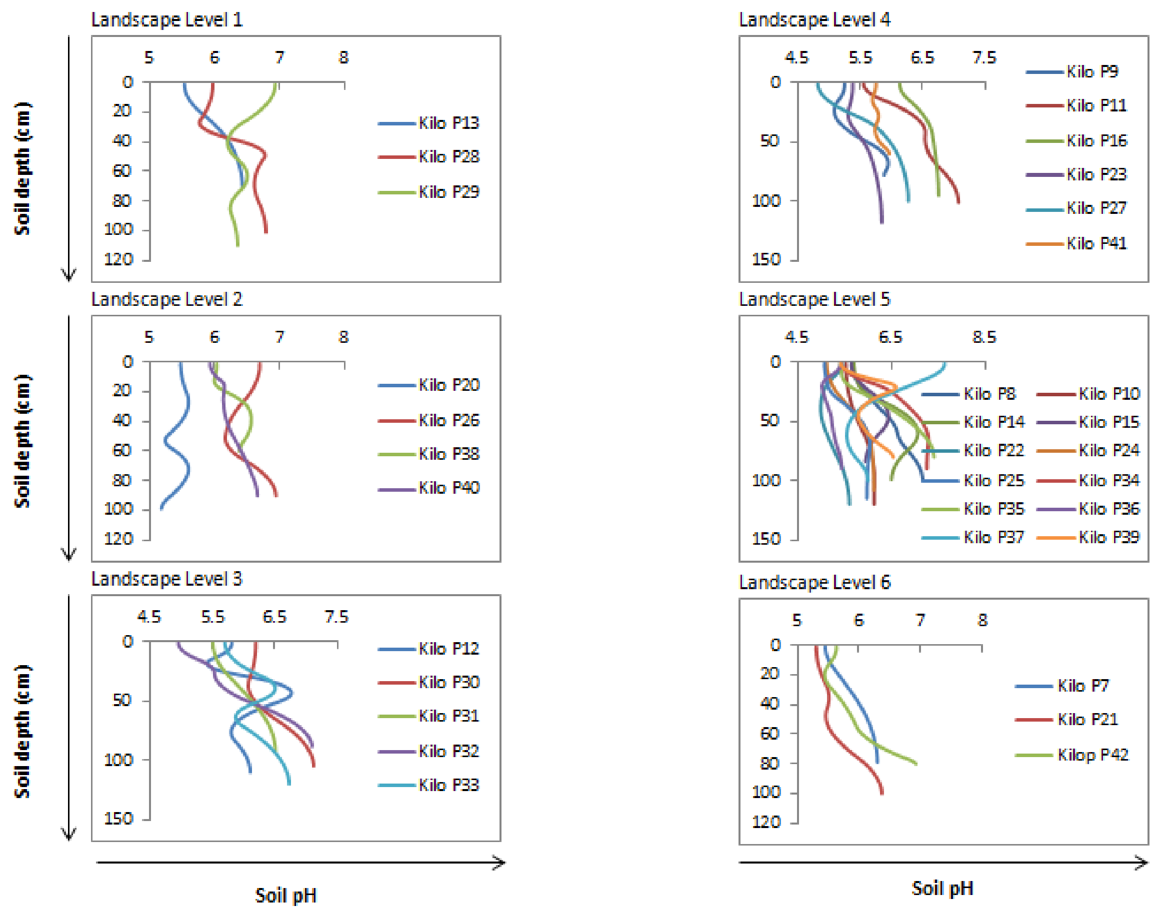


Figure 2.8. Variability of soil pH with soil depth for profiles studied in each landscape level. The keys (Kilo Pxx) represent codes of soil profiles studied in each level

Table 2.4. Pearson's correlations among studied soil properties.

	Depth	pH	TN	P	K	Ca	Mg	Na	OC	CEC	PBS	Cu	Mn	Fe	Zn	Sand	Clay
pH	0.395																
TN	-0.632	-0.473															
P	-0.389	-0.147	0.307														
K	-0.36	-0.175	0.307	0.308													
Ca	-0.195	0.119	0.109	0.152	0.208												
Mg	-0.051	0.376	-0.034	-0.22	-0.085	0.636											
Na	0.208	0.153	0.082	-0.123	-0.08	-0.48	-0.298										
OC	-0.659	-0.41	0.851	0.363	0.404	0.296	0.053	0.053									
CEC	-0.405	-0.249	0.632	0.307	0.318	0.281	0.065	0.06	0.711								
PBS	0.094	0.426	-0.275	-0.21	-0.109	0.674	0.891	-0.392	-0.179	-0.274							
Cu	-0.497	-0.44	0.599	0.327	0.112	0.403	0.293	-0.209	0.659	0.478	0.132						
Mn	-0.237	-0.208	0.284	0.037	0.214	0.296	0.296	-0.157	0.332	0.232	0.164	0.439					
Fe	-0.59	-0.49	0.559	0.344	0.183	0.342	0.101	-0.276	0.625	0.353	0.009	0.756	0.337				
Zn	-0.393	-0.28	0.499	0.175	0.38	0.258	0.215	0.03	0.493	0.357	0.064	0.475	0.586	0.473			
Sand	0.027	0.1	-0.473	0.164	-0.075	-0.13	-0.229	-0.433	-0.48	-0.529	0.03	-0.327	-0.228	-0.148	-0.321		
Clay	0.05	-0.012	0.342	-0.252	0.046	0.045	0.223	0.493	0.355	0.369	0.001	0.2	0.221	0.031	0.294	-0.926	
Silt	-0.353	-0.168	0.64	0.065	0.161	0.411	0.346	0.008	0.671	0.578	0.138	0.578	0.215	0.455	0.324	-0.672	0.479

2.3.4 Soil Organic Carbon

The soil organic carbon in the study area is generally low. The dominant levels are below 4% which correlate with dominance of coarse soil textures in the valley. The three upper landscape levels (1,2, and 3) have lower soil organic carbon compared to landscape levels 4, 5, and 6 which are lower in elevation. The difference could be attributed to the relatively coarser soil textures of the upper landscape levels. Coarse textured soils have been documented to have low organic matter contents (Lal and Shukla, 2004). The lower three landscape levels are characterized by a mix of pedons, some with more organic carbon, some with less. This also can be explained by presence of all ranges of soil textural classes in these landscape levels as discussed earlier.

Vertical variability of organic carbon in pedons of all landscape levels is shown in Figure 2.9. Generally, the organic carbon levels decrease with soil depth indicating accumulation in the surface horizons. A further comparison study revealed negative correlation ($r = -0.66$) between soil depth and soil organic carbon (Table 2.4). Some pedons showed irregular decrease/distributions of organic carbon with depth. These irregular changes occurred due to the presence of buried horizons containing more organic carbon or reflected changes in texture related to changes in depositional environments. The finer textured soils occur in slow stream flow rates environments that are conducive to the accumulation of organic materials mainly in the lower landscape levels and farther from the streams compared to the fast stream flow rates environments in the upper landscape levels and closer to the streams where the coarser soils are located.

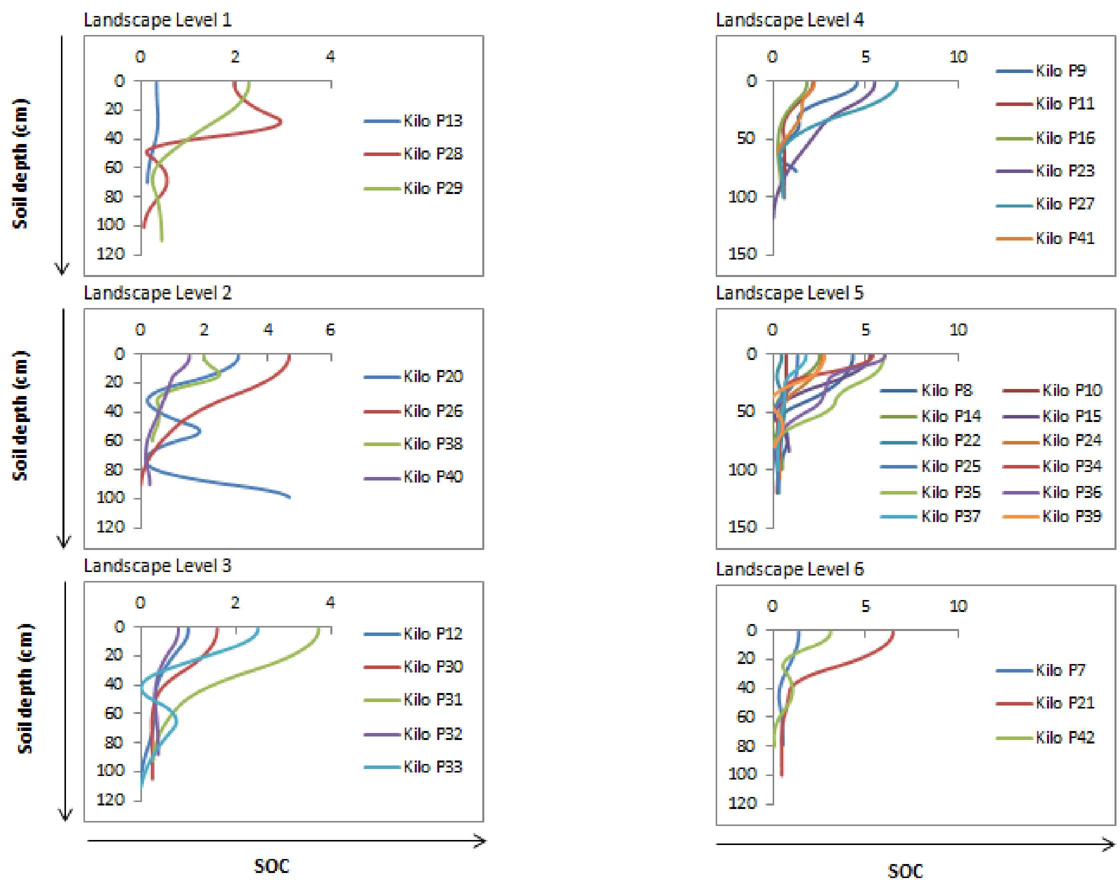


Figure 2.9. Variability of soil organic carbon with soil depth for profiles studied in each landscape level. The keys (Kilo Pxx) represent codes of soil profiles studied in each level.

Positive correlations were observed between OC and total nitrogen (r=0.85), CEC (r = 0.71), and silt content (r = 0.67) as shown in Table 2.4. These positive correlations were expected. Decomposing soil organic matter which is a proxy to soil organic carbon releases nitrogen (Bruland and Richardson, 2004), hence their positive correlation. Soil organic matter has been documented as an important contributor of soil CEC together with finer soil fraction (Lal and Shukla, 2004). Positive correlations were also observed with two micronutrients; Cu and Fe (r = 0.66 and 0.63, respectively).

2.3.5 Soil CEC

The soil CEC values are variable (Figure 2.10). They are generally low with many pedons having values below 20 cmol(+)/kg in the surface horizons, and below 15 cmol(+)/kg in their subsurface horizons. The vertical variability show that the CEC decreases with soil depth, generally in all landscape levels. These trends generally follow those of organic carbon (Figure 2.9), and not of clays as expected (Figure 2.8). Correlation studies also show that CEC in this valley is highly correlated to OC (r = 0.71) than clay (r = 0.37). Also CEC in this study area are more correlated to silt fraction (r = 0.58) than the clay fraction. As expected, there is a negative correlation (r = -0.53) between CEC and sand fraction of these soils. Weak correlation between CEC and clay suggests a relatively strong contribution of other sources of CEC in the soil, such as soil organic matter.

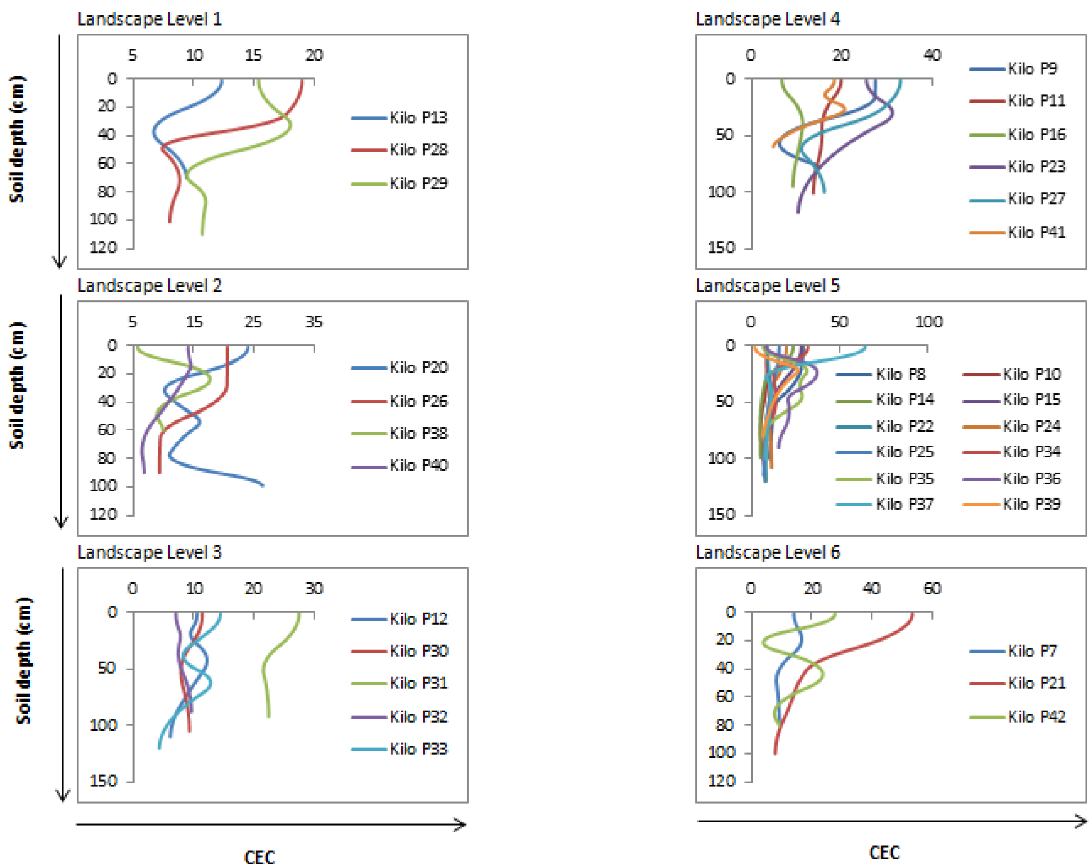


Figure 2.10. Variability of CEC with soil depth for profiles studied in each landscape level.

The keys (Kilo Pxx) represent names codes of soil profiles studied in each level.

Calculation of CEC:Clay ratio for the study area data suggest that the soils are dominated by Kaolinte and mixed clay mineralogy (Figure 2.11; Table 2.5). Kaolinites are CEC - poor non expanding (1:1) clays compared to the CEC - rich expanding (2:1) clays such as Smectite (Lal and Shukla, 2004). Figure 2.11 also shows some soil samples had CEC:Clay ratio > 0.95. According to Shaw et al. (1998), soils with CEC:Clay ratio greater than 0.95 may have CEC from Smectitic clays and feldspars or CEC from other than the clay fraction, such as the organic matter (Table 2.5). A study dedicated to clay mineralogy of Kilombero Valley would give more insights about this.

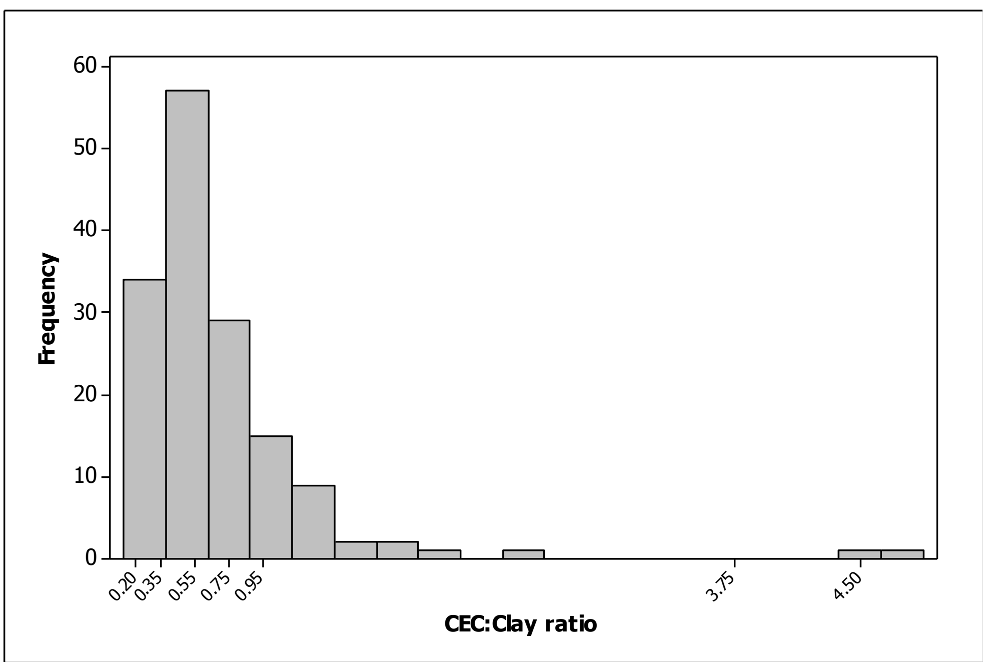


Figure 2.11. Histogram of CEC:Clay ratios of the studied soils

Table 2.5. Relationship between the CEC:Clay ratio and the clay mineral composition of a soil (Adopted from Shaw et al., 1998)

Clay:CEC ratio	Clay mineralogy
< 0.02	Kaolinite
0.2 – 0.35	Illite and kaolinite
0.35 – 0.55	Mixed clay mineralogies
0.55 – 0.75	Mixed clay mineralogies with a higher proportion of smectite
0.75 – 0.95	Dominantly smectite with the possibility of feldspars
> 0.95	Smectite, plus feldspars or CEC from other than the clay fraction (e.g. organic matter)

2.3.6 Gleying Properties

Gleying properties were observed in all landscape levels. This is because all landscape levels are saturated with water for durations long enough to allow reduction processes to take place. Flooding followed by lowering of water tables during dry seasons results to redoximorphic features which indicate gleying properties. In the process, reduction of iron (Fe^{2+}) depletes pigmentation and reveals background grey colors of soil minerals. Remaining or transported oxidized iron (Fe^{3+}) yields bright (reddish brown) colors (Schaetzl and Anderson, 2010).

Out of 33 studied pedons, 23 (approximately 82%) of them showed gleying properties. Mostly, the upper most gleyed horizons were observed to be Bg designated soil horizons (Table 2.6 and Figure 2.12). This was observed in 13 pedons, accounting for about 48% of studied gleyed pedons. Btg horizons accounted for 22% while Cg and ABg horizons accounted for 15 and 7% respectively. Bssg and BCg horizons had the lowest share.

The Bg horizons were generally closer to the surface (mean depth = 34.4 cm) than Btg horizons (mean depth = 43.8 cm), while pedons with Cg as the upper most gleyed horizon were found deeper (mean depth = 56 cm).

The variation in depths where gleying is observed partly indicates variations in water table levels with location within the landscape. Factors which would influence this include soil texture, amount of water and distance to water sources. The Bg horizons which in this study represent the shortest distance from soil surface to gleyed horizons are widely distributed in the study area. However, more than half are found in landscape level 5 correlating to deposition of finer materials. Most of the Btg horizons which were ranked second for having gleying closer to the soil surface are predominantly located in landscape level 6. The two landscape levels are generally the wettest part of the study area.

Gleying properties were not conspicuous in the topsoils. This could be because of tillage, which disturbs and mixes the top soil, thus restricting formation of the redoximorphic features.

Table 2.6. Descriptive statistics of depths from soil surface to the upper gleyed soil horizons

Upper horizon containing gleying features	Number of pedons with this horizon type	Mean depth from soil surface to gleying (cm)	SE Mean	StDev	Minimum	Maximum
Bg	13	34.38	5.13	18.49	14	82
Btg	6	43.83	3.73	9.13	29	54
Cg	4	56	13	26	31	84
ABg	2	25	5	7.07	20	30
Bssg	1	21	-	-	21	21
BCg	1	54	-	-	54	54
No gleying features to 100 cm depth	6	100	0	0	100	100

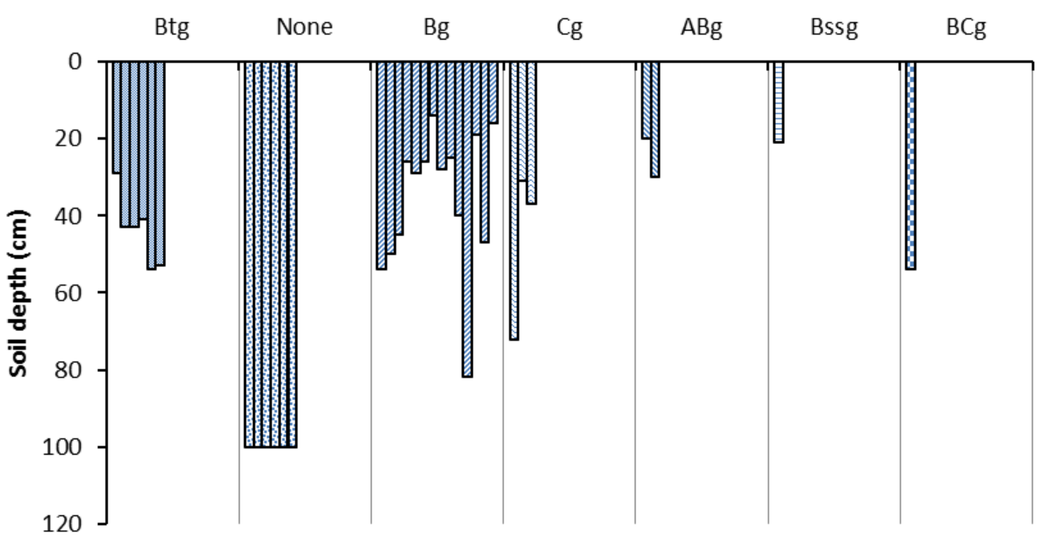


Figure 2.12. Depiction of depths from soil surface to soil horizons (Btg, Bg, Cg, ABg, Bssg, BCg) showing gleying properties. Horizons labeled ‘None’ did not show gleying properties to the depth of 100 cm.

2.3.7 Principal Component Analysis of the Soil Attributes

Principal component analysis (PCA) was done in order to be able to explain which soil attributes explain more variations in the study site. In the PCA, the first principal component (PC1) is the linear combination of variables that has maximum variance among all linear combinations, so it accounts for as much variation in the data as possible. The second principal component (PC2) is the linear combination of variables that accounts for as much of the remaining variation as possible. Likewise, all subsequent principal components are linear combinations that account for as much of the remaining variation as possible. The principal components are not correlated with each other.

In this study, the first principal component (PC1) explains about 35% of the variation. Furthermore, the first four principal components explain 73%, while the first six principal components explain 83% of the variation (Table 2.7). Results show that PC1 is not strongly correlated with any of variable since there is no any variable with Pearson correlation value > 0.5 (Table 2.7). However, using $r = 0.3$ as a cutoff point, this component is negatively correlated with total soil organic carbon, copper, silt, and nitrogen. All these 4 attributes vary together, that is, a decrease of one correlates to a decrease of the others, and vice versa. This is consistent with what we have seen in sections above that for example OC is positively correlated with Cu, silt, and TN.

Table 2.7. Principal components (PCs) of the studied soil properties. The first six PCs explained about 83% of the variations in the soil properties.

Variable	PC1	PC2	PC3	PC4	PC5	PC6
pH	0.19	-0.26	0.17	0.44	-0.16	0.38
Sand	0.23	0.00	-0.48	-0.01	-0.05	0.08
OC	-0.36	0.10	0.00	0.14	0.05	-0.01
P	-0.14	0.14	-0.31	0.41	-0.08	0.53
Cu	-0.32	-0.09	-0.10	-0.17	0.24	0.24
Mn	-0.20	-0.14	-0.02	-0.45	-0.42	0.12
Fe	-0.30	-0.01	-0.22	-0.20	0.21	0.16
Clay	-0.17	-0.01	0.51	-0.06	-0.06	-0.03
Silt	-0.30	-0.10	0.18	0.17	0.27	-0.15
Ca	-0.17	-0.42	-0.11	0.23	-0.02	-0.12
Mg	-0.09	-0.52	0.10	0.02	0.02	0.06
Na	0.03	0.27	0.42	0.06	-0.13	0.42
CEC	-0.29	0.10	0.09	0.31	0.02	-0.12
PBS	0.01	-0.55	-0.01	0.01	0.00	0.05
TN	-0.34	0.17	0.02	0.03	0.10	-0.01
Zn	-0.27	-0.04	0.00	-0.27	-0.45	0.22
K	-0.16	0.08	-0.13	0.30	-0.61	-0.43
Depth	0.27	-0.07	0.25	-0.07	-0.03	0.01
Eigenvalue	6.22	3.08	2.67	1.17	1.14	0.73
Proportion	0.35	0.17	0.15	0.07	0.06	0.04
Cumulative	0.35	0.52	0.67	0.73	0.79	0.83

PC2 decreases with calcium, magnesium and base saturation. Pearson correlation values for magnesium and base saturation are higher, suggesting the dominance of Mg compared to other bases in determination of base saturation. The correlations of attributes in PC1 and PC2 are also depicted in a loading plot in Figure 2.13. From the plot, soil pH and sand content vary positively with soil depth at varying correlation strengths. The two are the only attributes which increase with depth when 50% of the variability is explained. Mg, Ca, silt, Mn, Cu, Fe, clay, OC, CEC, K, TN, and P attributes are generally decreasing with increase in soil depth when considering the first two principal components.

Soil sand content and available P decrease with principal component three, while clay and sodium vary positively with the component. At the same time, principal component four increases with soil pH, available P and CEC and decreases with decrease in manganese.

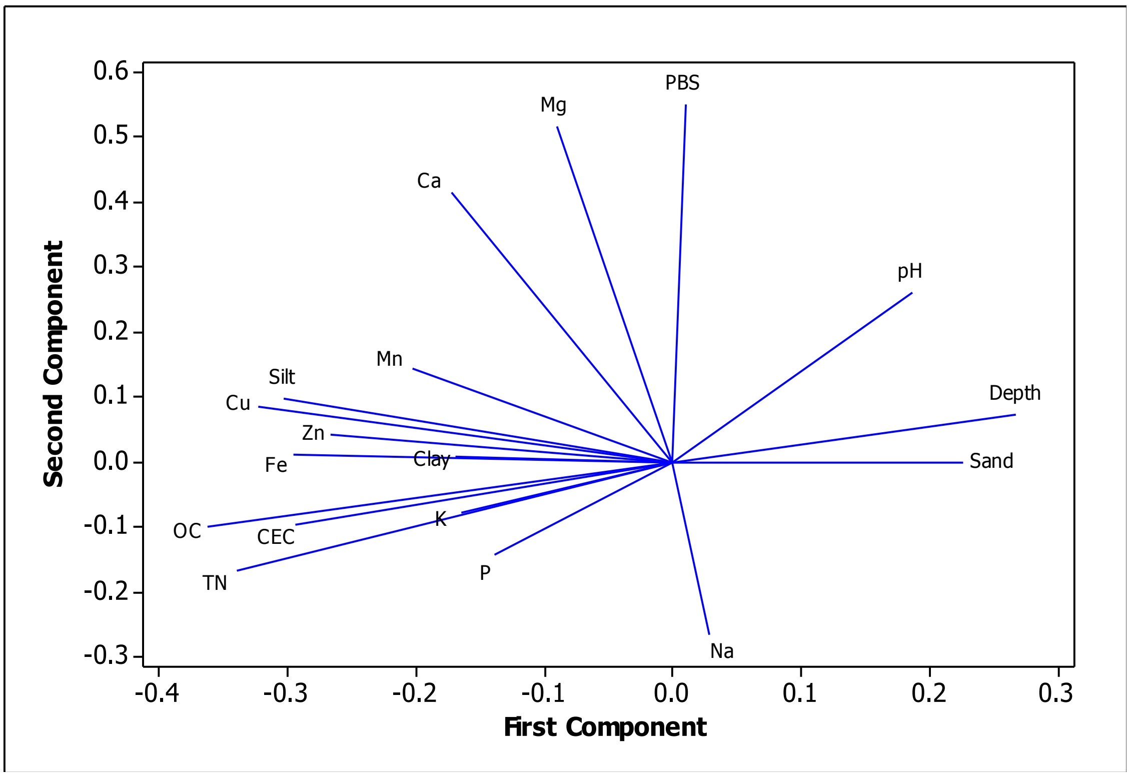


Figure2.13. Loading plot depicting correlation between the first two principal components (PC1 and PC2)

2.3.8 Soil Classification

The soils of Kilombero valley fall into 5 USDA Soil Taxonomy orders: Inceptisols, Entisols, Alfisols, Mollisols and Vertisols (Table 2.8). Inceptisols and Entisols are the dominant soil orders, represented by 31 out of 42 classified pedons. Five pedons were classified as Alfisols, four as Mollisols, and two as Vertisols. Formative elements 'aquic', 'udic', 'fluvic', connotative for wet soil moisture regimes and fluvial activities dominated the classification categories. Others dominant formative elements included 'endo' connotative for non-perched saturation, 'dystr' and 'eutr' for low and high base saturation respectively, 'aeric' and 'oxyaquic' for seasonal saturation, and 'hum' and 'mollic' for dark soil colors.

Table 2.8. Classification of soils of Kilombero valley to Sub group level of USDA Soil Taxonomy

Order (n=42)	Sub order	Great group	Sub group
Inceptisols (23)	Aquepts (17)	Dystrudepts (5)	Aeric Humaquepts (4)
	Udepts (6)	Endoaquepts (5)	Aquic Dystrudepts (1)
		Eutrudepts (1)	Aquic Humic Dystrudepts (1)
		Humaquets (12)	Fluvaquentic Dystrudepts (2)
			Fluvaquentic Endoaquepts (2)
			Fluvaquentic Humaquepts (6)
			Fluventic Endoaquepts (2)
			Oxyaquic Dystrudepts (1)
			Oxyaquic Eutrudepts (1)
			Typic Endoaquepts (1)
			Typic Humaquepts (2)
Entisols (8)	Aquents (3)	Endoaquents (1)	Aeric Endoaquents (1)
	Fluvents (4)	Fluvaquents (2)	Aquic Udifluvents (1)
	Orthents (1)	Udifluvents (4)	Humaqueptic Fluvaquents (1)
		Udorthents (1)	Mollic Fluvaquents (1)
			Mollic Udifluvents (1)
			Typic Udifluvents (2)
			Typic Udorthents (1)
Alfisols (5)	Aqualfs (2)	Endoaqualfs (2)	Aeric Umbric Endoaqualfs (1)
	Udalfs (3)	Hapludalfs (3)	Aquic Hapludalfs (1)
			Inceptic Hapludalfs (2)
			Umbric Endoaqualfs (1)
Mollisols (4)	Aquolls (2)	Argiudolls (1)	Aquic Argiudolls (1)
	Udolls (2)	Endoaquolls (2)	Fluvaquentic Endoaquolls (1)
		Hapludolls (1)	Fluvaquentic Hapludolls (1)
			Typic Endoaquolls (1)
Vertisols (2)	Aquerts (2)	Endoaquerts (2)	Aeric Endoaquerts (1)
			Typic Endoaquerts (1)

Distibution of soil orders in the study area is shown on Figure 2.14 and the subgroups for each landscape level are shown in Table 2.9.

Soils of landscape level 1 are generally very young, classifying to Fluvaquentic Endoaquepts, Fluvaquentic Humaquepts, and Mollic Udifluvents, while those of Landscape level 2 classified to Typic Endoaquolls, Fluvaquentic Hapludolls, Oxyaquic Dystrudepts, and Typic Udorthents. Two Mollisols were observed in this level indicating presence of thick, dark, high base surface horizons in areas where cyclic depositions are no longer active. These areas are cultivated with upland crops and some settlements are established.

The soils of landscape level 3 are mostly Inceptisols with pockets of Entisols and Mollisols. They classify to Aeric Humaquepts, Typic Humaquepts, Aquic Humic Dystrudepts, Humaqueptic Fluvaquents, and Fluvaquentic Endoaquolls. The soils of Landscape level 4 are also dominantly Inceptisols. Some of soils in this landscape level have developed argillic (Bt) horizons due to clay illuviation. Such soils are classified as Alfisols. Soil subgroups in this landscape level are Fluventic Endoaquepts, Typic Humaquepts, Aquic Dystrudepts, Inceptic Hapludalfs, Aeric Umbric Endoaqualfs, and Typic Udifluvents

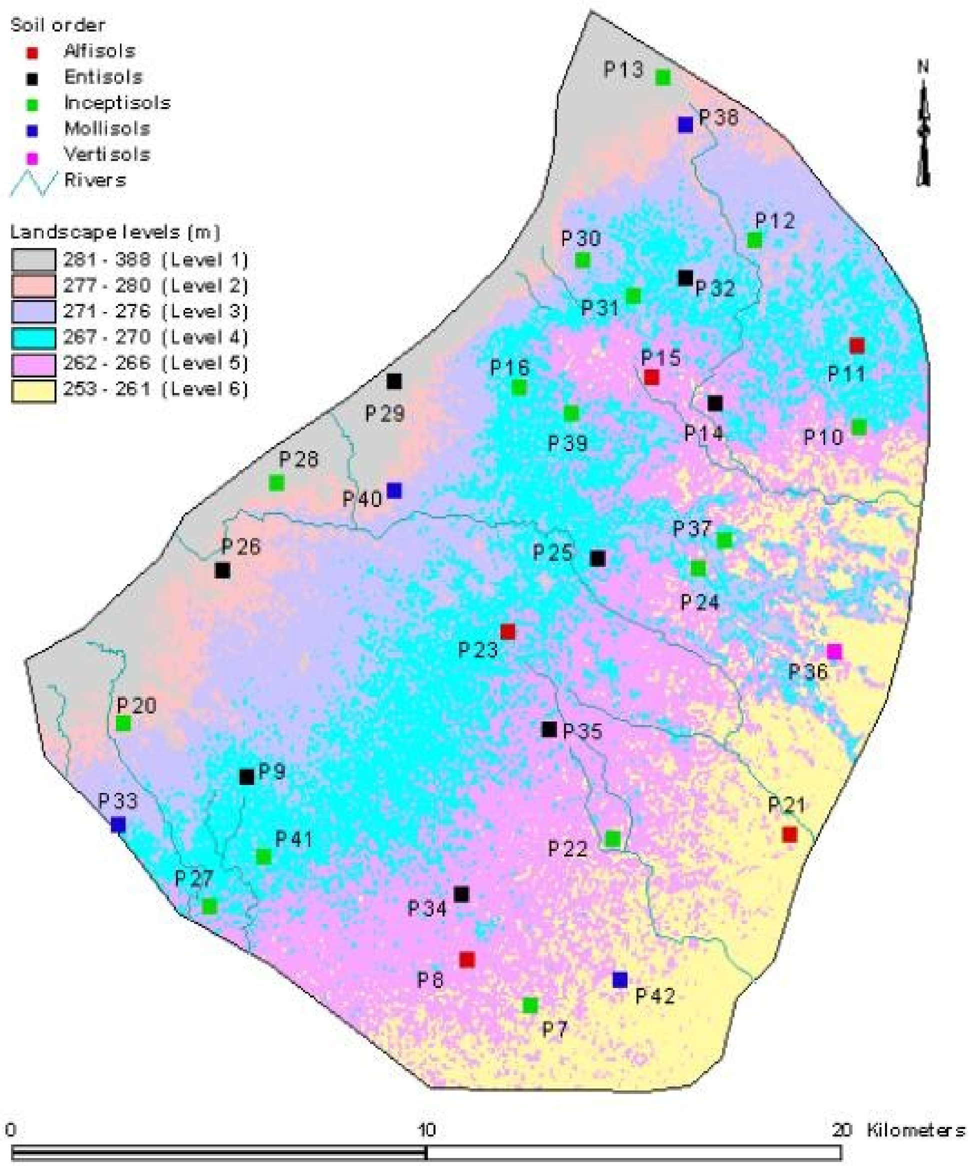


Figure 2.14. Distribution of soil orders in the study area as depicted by locations of soil profiles in the demarcated landscape levels. The P's in the map indicates soil profile identifications.

Table 2.9. Distribution of soil sub groups in the different landscape levels

Landscape level	Sub group	Landscape level	Sub group
1	Fluvaquentic Endoaquepts Fluvaquentic Humaquepts Mollic Udifluvents	5	Typic Endoaquepts Aeric Humaquepts Oxyaquic Eutrudepts Fluventic Endoaquepts Fluvaquentic Humaquepts Aquic Udifluvents Mollic Fluvaquents Aeric Endoaquents Typic Udifluvents Umbric Endoaqualfs Inceptic Hapludalfs Typic Endoaquerts
2	Typic Endoaquolls Fluvaquentic Hapludolls Oxyaquic Dystrudepts Typic Udorthents		
3	Aeric Humaquepts Typic Humaquepts Aquic Humic Dystrudepts Humaqueptic Fluvaquents Fluvaquentic Endoaquolls		
4	Fluventic Endoaquepts Typic Humaquepts Aquic Dystrudepts Inceptic Hapludalfs Aeric Umbric Endoaqualfs Typic Udifluvents	6	Fluvaquentic Humaquepts Aquic Hapludalfs Aquic Argiudolls

Landscape level 5 is generally a low flow rate stream area of the study site, and actively receiving new material from seasonal flooding. As a result of this, most of the soils in this level classified into less developed soil order of Inceptisols and Entisols. Some Alfisols were observed in relatively stable areas which allowed illuviation and accumulations of clays in their subsoils. The subgroups in this landscape level are Typic Endoaquepts, Aeric Humaquepts, Oxyaquic Eutrudepts, Fluventic Endoaquepts, and Fluvaquentic Humaquepts. Others are Aquic Udifluvents, Mollic Fluvaquents, Aeric Endoaquents, Typic Udifluvents, Umbric Endoaqualfs, Inceptic Hapludalfs, and Typic Endoaquerts.

The soils of Landscape level 6 appear to be more developed compared to the nearby Landscape level 5 despite being closer to the braided zone of the Kilombero River. The natural vegetation of this area is grass, particularly the elephant grasses. Most of these areas have just recently been opened for agriculture. Kato (2007) gives an account of how farmers first opened agricultural lands closer to their settlements in the upper slopes, and slowly moved down slope due to decreased soil fertility, infestation of rice fields by grassy weeds and increased human population. The stable, undisturbed environment and high organic matter turnover by grassy vegetation could explain presence of clay illuviations in subsurface horizons and development of dark, thick surface horizons, characteristic of Alfisols and Mollisols. The soils of this landscape level are classified as Aquic Hapludalfs, Aquic Argiudolls, and Fluvaquentic Humaquepts.

2.4 Conclusion

This chapter gave insights on important properties of the soils of the study area including soil texture, pH, organic carbon, CEC and water table fluctuations as indicated by gleying properties. The chapter also classified the soil using USDA Soil Taxonomy System to sub group level.

The dominant soil textural classes were found to be sandy clay loams and sandy loams, accounting for about 66% of the studied soil horizons. As expected, soil textures appeared to generally become finer down the gradient due to influence of water sediment deposition dynamics where coarser materials are deposited upstream due to higher stream flow rates. Some coarse materials were observed in the lower landscape levels, closer to the streams despite dominance of finer materials.

Soil pH values in the study area were found to vary widely spatially and vertically. It ranges from 4.7 to 7.2, with topsoil values predominantly being around 5.8. The soil pH did not show strong correlation with any of the other determined parameters. Weak negative correlations were observed with organic carbon and total nitrogen while weak positive correlation between pH and base saturation was also observed.

The soil organic carbon in the study area was found to be generally low. The levels are generally below 4% correlating with dominance of coarse soil textures in the valley. Generally, the organic carbon levels were found to decrease with soil depth indicating accumulation in the surface horizons. Some pedons showed irregular decrease with depth. The organic carbon irregular decrease with soil depth is a property of alluvial soils.

The soil CEC values are also variable. They are generally low with many pedons having values below 20 cmol(+)/kg in the surface horizons, and below 15 cmol(+)/kg in their subsurface horizons. The vertical variability shows that the CEC generally decreases with soil depth, following the trends displayed by organic carbon.

Gleying properties were observed in most part of the study area supporting records of repeated flooding and drying of the landscape. Mostly, the upper most gleyed horizons were observed to be Bg horizons followed by Btg, Cg, ABg, Bssg, and BCg in that order. Gleying properties were not conspicuous in the topsoils. This could be because of tillage, which disturbs and mixes the top soil, thus restricting formation of the redoximorphic features which are evidence of gleying.

The soils of Kilombero valley fall into 5 USDA Soil Taxonomy orders: Inceptisols, Entisols, Alfisols, Mollisols and Verisols. Inceptisols and Entsols are the dominant soil orders, represented by 31 out of 42 classified pedons.

Generally, the soils of Kilombero valley are young and demonstrated properties common for alluvial soils worldwide.

2.5 References

Alves, C., and D. Ross. 2010. Examining phosphorus contributions from alluvial soils – a comparison of three Vermont, U.S.A. River corridors. In Proceedings of 19th World Congress of Soil Science, Soil Solutions for a Changing World 1 – 6 August, Brisbane, Australia. 228 – 231.

Bishop, T.F.A., McBratney, A.B., and G.M. Laslett. 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91:27–45.

Bonarius, H. 1975. Physical properties of soils in the Kilombero Valley, Tanzania. GTZ, Stuttgart, Germany.

Brady, N.C., and R.R. Weil. 2010. Elements of the Nature and Properties of Soils. Pearson Hall, NY.

Bray, R.H., and L.T. Kurtz. 1945. Determination of total, organic and available forms of phosphorus in soils. Soil Science 59:39-45.

Bremner, J.M., and C.S. Mulvaney. 1982. Total nitrogen. In: A.L. Page, R.H.Miller, and D.R. Keeney, editors, Methods of soil analysis, Part 2; Chemical and mineralogical properties, 2nd edition. American Society of Agronomy, Madison, Wisconsin. p. 595-624.

Bruland, G.L., and C.J. Richardson. 2004. Hydrologic gradient and topsoil additions affect soil properties of Virginia Created wetland. Soil Science Society of America Journal 68:2069 – 2077.

Egbuchua, C.N. 2011. A pedological characterization of soils of the River Niger alluvial complex of Nigeria. Nigerian Journal of Soil and Environmental Research 9:65-70.

ESRI. 2002. What's New in ArcView 3.1, 3.2, and 3.3. Environmental Systems Research Institute, Inc., Redlands, California.

ESRI. 2010. ArcGIS - A Complete Integrated System. Environmental Systems Research Institute, Inc., Redlands, California.

FAO. 1961. The Rufiji Basin Tanganyika. FAO Exp. Techn. Ass. Progr. No. 1269 Vol. 7. Rome.

FAO. 2006. Guidelines for soil description. FAO, Rome, Italy.

Gee, G. W., and J.W. Bauder. 1986. Particle-size analysis. In: A. Klute, editor, Methods of soil analysis: Part 1—physical and mineralogical methods. SSSA Book Series 5.1 Soil Science Society of America, American Society of Agronomy. p. 383–411.

Gervais-Beaulac, V., D. Saint-Laurent, J.S. Berthelot, and M. Mesfioui. 2013. Organic carbon distribution in alluvial soils according to different flood risk zones. Journal of Soil Science and Environmental Management 4(8):169-177.

Gill, A. 2003. Ancient Egyptians: The Kingdom of the Pharaohs brought to Life. Harper Collins Entertainment, London, United Kingdom.

Haifa. 2008. Nutritional recommendations for rice. Haifa, Israel.

Holmes, R.S. and W.E. Hearn. 1942. Chemical and physical properties of some of the important alluvial soils of the Mississippi drainage basin. Technical bulletin No, 833. United States Department of Agriculture, Washington, D. C.

Hoskins, B.R. 1997. Soil testing handbook for professionals in agriculture, horticulture, nutrient and residuals management. Third edition, University of Maine, Maine, USA.

Iqbal, J., J.A. Thomasson, J.N. Jenkins, P.R. Owens, and F.D. Whisler. 2005. Spatial variability analysis of soil physical properties of alluvial soils. Soil Science Society of America Journal 69:1-14

IUSS Working Group WRB. 2014. World reference base for soil resources 2006. 2nd edition. World Soil Resources Reports No. 103. FAO, Rome.

Kato, F. 2007. Development of a major rice cultivation area in the Kilombero Valley, Tanzania. African Study Monographs, Supplement 36: 3-18.

Kemker, C. 2014. Sediment transport and deposition: Fundamentals of environmental measurements. Fondriest Environmental, Inc. http://www.fondriest.com/environmental-measurements/parameters/hydrology/sediment-transport-deposition/. Last visited November, 2014.

Kihoro, J., N.J. Bosco, and H. Murage. 2013. Suitability analysis for rice growing sites using a multicriteria evaluation and GIS approach in Great Mwea region, Kenya. SpringerPlus 2013, 2:265. http://www.springerplus.com/content/2/1/265

Kohnke, H. 1968. Soil Physiscs. McGraw Hill Book Company, New York.

Lal, R. and M.K. Shukla 2004, Principles of Soil Physics. Marcel Dekker, New York.

Landon, J.R. (1991). Booker tropical soil manual. A handbook for soil survey and agricultural land evaluation in the tropics and subtropics. Longman Scientific & Technical Publishers. Essex.

Lin, G.C.S. 1997. Red capitalism in South China: Growth and development of the Pearl River Delta. UBC Press, Vancouver, Canada.

Lindsay, W.L,. and W.A. Norvell. 1978. Development of a DTPA soil test for zinc, iron, manganese, and copper. Soil Science Society of America Journal 42:421-428.

Loxton, R.P. 1953. Kilombero land use survey - Block A. Unpublished report to the Tanganyika Government.

Loxton, R.P. 1954. Kilombero land use survey- Blocks B and C. Unpublished report to the Tanganyika Government.

Massawe, B.H.J. and N.A. Amury. 2012. Report on agronomic practices and soil fertility analysis for improved rice production in Kilombero and Wami Valleys in Tanzania. USAID Feed the Future NAFAKA staples value chain activity project.

McLean, E.O. 1986. Soil pH and lime requirement. In: A.L. Page, R.H.Miller, and D.R. Keeney, editors, Methods of soil analysis, part 2; chemical and mineralogical properties, 2nd edition. American Society of Agronomy, Madison. p. 199-223.

Minitab. 2004. Meet Minitab: Minitab user guide. Minitab Inc., USA.

Mueller, L., U. Schindler, W. Mirschel, T.G. Shepherd, B.C. Ball, K. Helming, J. Rogasik, F. Eulenstein, and H. Wiggering. 2010. Assessing the productivity function of soils. A review. Agronomy for Sustainable Development 30:601–614.

Munsell Color Company. 1992. Munsell soil color charts. Munsell Color Co. Inc. Baltimore.

Nelson, D.W., and L.E. Sommers. 1982. Total carbon, organic carbon and organic matter. In: A.L. Page, R.H.Miller, and D.R. Keeney, editors, Methods of soil analysis, part 2; chemical and mineralogical properties, 2nd edition. American Society of Agronomy, Wisconsin. p. 539-579.

QGIS Development Team. 2014. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org.

Rhoades, J.D. 1982. Soluble salts. In: A.L. Page, R.H.Miller, and D.R. Keeney, editors, Methods of soil analysis. Part 2, chemical and mineralogical properties, 2nd edition. American Society of Agronomy. Wisconsin. p. 149-158; 167-180.

Schaetzl, R., and S. Anderson. 2010. Soil genesis and geomorphology. Cambridge University Press, Cambridge, UK.

Schoeneberger, P.J., D.A. Wysocki, E.C. Benham, and Soil Survey Staff. 2012. Field book for describing and sampling soils, Version 3.0. Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE

Schollenberger, C.J., and R.H. Simon. 1945. Determination of exchange capacity and exchangeable bases in soils-ammonium acetate method. Soil Science 59:13-24

Shaw, R.J., K.J. Coughlan, and L.C. Bell. 1998. Root zone sodicity. In; Eds M.E. Sumner and R. Naidu. Sodic soils: Distribution, properties, management, and environmental consequences. Oxford University Press, New York. p. 95–106.

Soil Survey Staff. 2014. Keys to soil taxonomy. 12th Edition. USDA, Natural Resource Conservation Service, USA.

Southard, J. 2006. Introduction to fluid motions, sediment transport, and current-generated sedimentary structures. MIT Open Courseware: Massachusetts Institute of Technology. USA.

Sweeney, M.R., E.V. McDonald, and C.E. Markley. 2013. Alluvial sediment or playas: what is the dominant source of sand and silt in desert soil vesicular A horizons, southwest USA. Journal of Geophysical Research: Earth Surface 118:1–19.

Thien, S.J. 1979. A flow diagram for teaching texture by feel analysis. Journal of Agronomic Education 8:54-55.

Thomas, G.W. 1982. Exchangeable cations. In: A.L. Page, R.H.Miller, and D.R. Keeney, editors, Methods of soil analysis. Part 2, chemical and mineralogical properties, 2nd edition. American Society of Agronomy. Wisconsin. p. 595-624.

USGS. 2000. Shuttle Radar Topography Mission, 1 Arc Second scene SRTM. http://earthexplorer.usgs.gov/ page visited 24 October, 2014.

Watanabe, F.S., and S.R. Olsen. 1965. Test of ascorbic acid method for determining phosphorus in water and NaHCO3 extracts from soil. Soil Science Society of America 29:677-678.

Chapter 3: Numeric Classification of Soils of Kilombero Valley

3.1 Introduction

3.1.1 Soil Description

Mapping soils, being it classes or individual attributes, relies to a large extent on information generated from physical soil sampling. Soil profile descriptions and sampling appear to be indispensable processes in identifying and mapping soils in many soil classification systems (Minasny and McBratney, 2007). Soil profile, a vertical section of soil from the ground surface to the parent rock, enables enables soil scientists to understand the processes that influenced formation of the soil in question, and it is the means of determining the types of soil that occur in a given landscape. Soil profiles are composed of soil horizons which are layers generally parallel to the soil surface and whose physical characteristics such as colour, texture, structure, root density, and porosity differ from the overlying and underlying layers (Buol et al., 2002; Schaetzl and Anderson, 2010). In order to facilitate cross-references and comparison between soils and soil classification systems, and to enhance standardization and uniformity of soil profile descriptions; standard soil profile description guidelines such as FAO Guidelines for Soil Profile Description (FAO, 2006) and Field Book for Describing and Sampling Soils (Schoeneberger et al., 2012) are used.

3.1.2 Numerical Soil Classification

Soil properties, and hence soil types, vary spatially on the landscape (Odgers et al., 2012). Soil properties also vary with soil depth (Bishop et al., 1999, Malone et al., 2009). A soil at a point '*x*' could vary slightly or significantly from a soil at a far or a nearby point '*y*'. Traditional methods have been used to classify soil properties based on similarities and differences as depicted by sampled and described soil profiles (Muir, 1970). These methods are qualitative because they do not consider numerical relationships in deciding whether two profiles are similar or not. Numerical classification of profiles allows application of numerical procedures to determine similarity between soil profiles and subsequently identify the most representative groups for the dataset (Rizzo et al, 2014). Studies show superiority of numerical over traditional methods (Muir, 1970; Carré and Jacobson, 2009; Rizzo et al, 2014) and because of this, there is a movement in soil science away from qualitative soil classification toward a quantitative approach (Jacobson and Carré, 2006).

According to McBratney et al. (2010), the idea of numerical classification comes from the 1770s. It was not until the 1950's when it was put in real use in soil classification. Early works employing numerical soil classification methods includes those of Hole and Hironaka (1960), Bidwell and Hole (1964a; 1964b), Bidwell et al. (1964) and Rayner (1966).

The evolution of numerical classification is based on the idea that classification should be based on many characters, each with equal weight. In early works ordination methods were employed for numerical classification where similarity matrices were calculated between pedons and transformed to distances (Hole and Hironaka, 1960; Muir et al., 1970). The shorter the calculated distance between the pedons, the similar they are. Ordination is a statistical technique in which data from a large number of sites or populations are represented as points in a two- or three-dimensional coordinate frame.

One of the early limitations encountered in the field of soil mapping was the classification of soils as discrete – non continuous objects with abrupt boundaries between the neighboring soil types (McBrtaney and Odeh, 1997). Although sharp transitions may be observed in some cases where very dissimilar soils may border each others such as when they have different lithology, the transitions are generally gradual (Fridland, 1974).

Use of fuzzy set theory has enabled classifying soils as continuous objects. In fuzzy classifications, continuous class membership values are used instead of hard classes. Individuals that exactly match strictly defined classes are assigned a membership value of 1, while the rest are assigned membership values depending on their degree of closeness to the strictly defined classes or the class centroids (or class means). The approaches include use of fuzzy *k*-means (Diday, 1971) to cluster individuals into fuzzy classes. A detailed review of these methods can be obtained in McBratney and Odeh (1997).

Numerical methods have been used to generate horizon classes, soil classes, and to define taxonomic distance between the classes for existing soil classification systems in different studies. Carré and Jacobson (2009) used numerical classification techniques to predict soil taxonomy and available water capacity for a set of soil profiles in New South Wales, Australia. Odgers et al (2011a) used numerical methods to create classes of soil profiles whose horizons consist of common sequences of membership to the soil layer classes they named “soil series classes”. The soil series classes were found to be optimal for the set of soil profiles and the taxonomic distance to the modal profiles of the soil series classes was mapped using regression-kriging (Odgers et al., 2011b). Minasny and McBratney (2007) used taxonomic distances as a criterion in supervised classification for prediction of soil classes. In their study they used decision trees, but stressed that the concept is universal and can be used with other data – mining and prediction tools and to other soil classification systems. In another study, Rizo et al (2014) used numerical classification to correctly distinguish Argissolos from Latossolos and Nitossolos in the Brazilian soil classification system. The inputs for calculating taxonomic distances between classes and predict soil classes have been morphological, physical, and chemical properties derived from soil profiles (Minasny and McBratney, 2007).

The relevance and applicability of numerical soil classification techniques have been suggested and tested by many authors (Rayner, 1966; Muir, 1970, Carré and McBratney, 2005; Carre and Jacobson, 2009; McBratney et al, 2010; Odgers et al 2011a; Odgers et al 2011b; Rizo et al., 2014). These methods have not been widely used in Africa, where soil mapping is still very fragmented, coarse, and in great demand (Cook et al., 2008).

3.1.3 OSACA

OSACA (Outil Statistique d'Aide à la Cartogénèse Automatique), a Java Web Application (Jacobson and Carre, 2006) was used in this study to cluster and classify soil horizons and soil profiles using distance metrics. Its algorithm is based on the *k*-means clustering process (Diday, 1971). The quality of both soil horizon and solum clustering is described by the ratio D_{intra}/D_{inter}. D_{intra} is the mean of the distances of each observed horizon/solum to its nearest cluster centre, and D_{inter} is the mean of the distances between each cluster centre. Initial cluster centers in OSACA are not chosen at random as in many *k*-means clustering applications, but uses an iterative procedure to have consistent results.

The OSACA version used in this study provides two options for horizon distance metrics; Euclidean and Manhattan. Euclidean distance is calculated using Equation 3.1 and Manhattan distance using Equation 3.2 where D_{ij} is distance between two horizons v_i and v_j

$$D_{ij} = \sqrt{\sum_{k=1}^{n}\left(v_{ik} - v_{jk}\right)^2} \qquad \text{Eq. 3.1}$$

$$D_{ij} = \sum_{k=1}^{n}\left|v_{ik} - v_{jk}\right| \qquad \text{Eq. 3.2}$$

The distance *Dij* is used in calculation of soil profile (solum) distance metrics.

The soil profile metrics which can be performed in OSACA include Utilitarian, Joint, and Pedological distances. The utilitarian distance takes into account the depths of the horizons, but during the process two very similar soils, differing only by the depths of their horizons could have a large utilitarian distance due to overstating importance of depth weighting (Carre and Jacobson, 2009). The joint distance between two profiles involves scaling which has an effect of reducing depth weighting which takes better care of depth sensitivity than Utilitarian metrics. Pedological distance emphasizes succession of soil horizons types without taking into account their depths and has been shown to give better results than Joint and Utilitarian distances (Carre and Jacobson, 2009).

Figure 3.1 shows how pedological distance between two soil profiles with different number of horizons is calculated. The mean distance between the horizons in each profile is taken in sequence whereas the lowest horizon in the profile with fewer horizons is used repeatedly. Pedological distance (D_{ped}) is calculated using Equation 3.3, where *Sa* and *Sb* are profiles *A* and *B* respectively; *hi,j* is horizon *j* of profile *i; Dh* (*ha,j, hb,j*) can be either Euclidian or Manhattan distance, whichever used in horizon distance metric, between the *j* horizons of profiles *A* and *B; Ma* and *Mb* are the number of horizons of profile *A* and *B* respectively. In this case profile *A* has less number of horizons than profile *B*.

$$D_{ped}(s_a, s_b) = \frac{\sum_{j=1}^{M_a} D_{ij}(h_{a,j}, h_{b,j}) + \sum_{j=M_a+1}^{M_b} D_h(h_{a,M_a}, h_{b,j})}{M_b} \quad \text{Eq. 3.3}$$

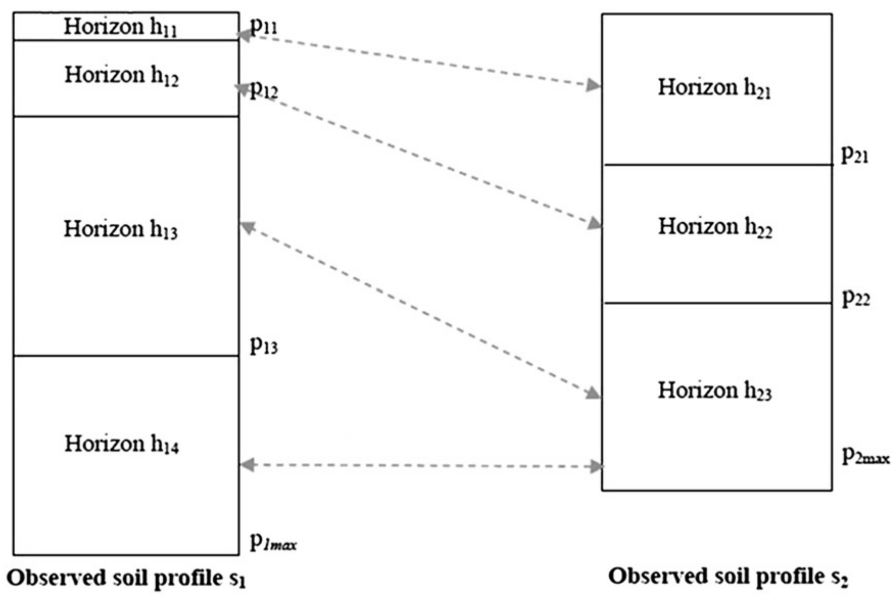


Figure 3.1. Pedological distance determination taking into account the sequence of horizons (adopted from Carre and Jacobson, 2009).

3.1.4 Objectives of the study

This study used numerical soil classification techniques to cluster soils in a selected site of an alluvial basin in Kilombero Valley Tanzania to assess its applicability and create soil classes for subsequent digital soil mapping of the area.

3.2 Methods

3.2.1 Study Site

The study was conducted in Kilombero Valley, Tanzania. The valley is about 300 km east of Indian Ocean covering about 11000 km^2. The study site covered about 300 km^2 within the valley. The study site is located in zone 37 south, occupying the area lying between 9064697 and 9089031 m northings and 175422 to 197033 m eastings. More description of the study area is given in the introductory chapter of this document.

3.2.2 Soil Information Assemblage

A total of 152 soil horizons were sampled from 33 soil profiles in the study area. The information determined and recorded from each horizon included morphological characteristics such as soil colour, depth, structure, cutans, mottles, concretions and redoximorphic features. Laboratory analyses were used to generate information about chemical and some physical properties for each horizon. The properties analysed were: soil pH, electrical conductivity (EC), soil texture, total nitrogen (TN), organic carbon (OC), available phosphorus (P), exchangeable potassium (K), exchangeable sodium (Na), exchangeable magnesium (Mg)and exchangeable calcium (Ca). Others included cation exchange capacity (CEC), extractable manganese (Mn), extractable iron (Fe), extractable copper Cu), and extractable zinc (Zn).

Field and laboratory methods for the soil data collected are described in detail in chapter 2 of this dissertation.

3.2.3 Data Pre-processing

The above data was tabulated and preprocessed before inputting it in OSACA for horizon and solum clustering and classification. The Munsell color notations recorded in the field (hue, value and chroma) were converted to red green and blue (RGB) numerical counts (0 – 255) using mColorConverter (He, 2013). The raw data were standardized because data measured at different scales may not contribute equally in the analysis. The equation below was used:

$$V_s = (V_o - \mu)/\sigma \qquad \text{Eq. 3.4}$$

where V_s is the standardized value of the attribute, V_o is its observed value, μ is sample mean and σ is the sample standard deviation.

3.2.4 Clustering

In the OSACA process options, the Euclidean distance metric was chosen for horizon clustering. A range of optimal clusters suggested was between 10 and 12. Pedological distance metric was chosen for soil profile clustering while the optimal number of clusters was suggested to be between 9 and 13. This range was chosen because the Anderson's basemap showed that the study area had 9 soil groupings.

From the outputs we obtained the following data:

1. Horizon class distance table: this gave distances between optimized clusters centres and was used to assess the similarities or differences between the horizon clusters. The shorter the distance between the clusters, the similar the clusters being compared.
2. Horizon cluster descriptions: this provided centroid values of parameters for each modal horizon. Back-transformation was done since the output was based on the standardized data.
3. Solum distance table: this provided distance metrics for each of the actual (observed) solum from each of the modal solums. From this table, it was possible to assign observed solums to the modal solums to which they are most similar.
4. Solum class distance table: this gave information which aided in assessing how close the modal profiles are to each other. The output shows the metric distance of a modal profile from each of the other modal profiles.
5. Solum cluster descriptions: this gave description of all horizons comprising each modal solum. Again, back-transformation was done to be able to interpret the attributes values for each solum and horizon.

3.2.5 Assessment of Differences between the Soil Clusters

To be able to visually assess the differences between generated clusters, equal area spline functions (Bishop et al, 1999) were employed to model vertical variation of selected soil attributes with depth. A smoothing value (λ) of 0.1 and the soil depth intervals of 0-5, 5-15, 15-30, 30-60, 60-100, and 100-200 cm in the Soil profile generator v2.0 were used.

3.3 Results and Discussion

3.3.1 Horizon Clusters

Properties of Predicted Modal Horizons

The OSACA output gave 11 horizon clusters as the optimum number. The centroids for each horizon cluster were also generated. These centroids are predicted values of the attributes which constitute the modal horizons. Modal horizons for the 11 clusters are referred to as H1 to H11 in this document. The attributes of the modal horizons are presented in Table 3.1.

In order to assess the reliability of the clustering results, an analysis to see if the interrelationships of attributes in the predicted modal horizons are consistent to the common knowledge in pedology was done. Selected attributes for the test were soil pH against base saturation, sand against clay content, CEC against clay content, and organic carbon against soil color.

As expected, positive correlation (r = 0.74) was observed between bases and soil pH (Figure 3.2). Soil pH tends to increase with increase in basic cations (Brady and Weil, 2002).

Table 3.1. Attributes (centroids) of modal horizons generated during numeric clustering of soil horizons. Modal horizons are named H1, H2, …H11.

Modal horizon	pH (water)	TN	OC	Sand	Clay	Silt	Texture	P	Cu	Zn	Mn	Fe	Colour name	Ca	Mg	K	Na	CEC
				(%)						(mg/kg)						(Cmol(+)/kg)		
H1	6.2	0.04	0.3	73	20	7	SCL	6.4	1.2	0.2	8.3	21.5	olive	1.7	6.8	0.07	0.26	8.7
H2	6.1	0.08	0.5	46	46	8	SC	2.3	1.1	0.4	22.4	17.4	olive	1.8	2.5	0.09	0.42	13.3
H3	5.9	0.08	0.8	74	18	7	SL	18.2	2.3	0.3	8.9	71.4	dark olive brown	2	4	0.08	0.24	12.1
H4	6.5	0.13	2.2	71	17	11	SL	168.6	1.6	0.8	5.5	43	very dark brown	7.2	1	0.31	0.2	16.5
H5	7.5	0.13	1.6	80	12	8	SL	38.4	1.5	0.9	11.2	26.4	very dark greyish brown	6.7	43.1	0.41	0.07	57.4
H6	6.8	0.06	0.8	78	14	8	SL	8.9	2.4	3.5	5.9	48.9	very dark greyish brown	4.1	29	0.54	0.07	15
H7	5.2	0.13	2.3	63	26	11	SCL	25.3	2.9	1.4	34.6	18	black	2.4	1.8	0.54	0.21	18
H8	5.5	0.29	3.6	50	30	20	SCL	24.9	3.8	0.7	14.2	120.3	black	2.2	3.5	0.13	0.33	25.2
H9	5.6	0.1	1.1	71	22	7	SCL	23.8	2.2	2.4	13.1	109.1	dark brown	1.5	1.5	0.05	0.23	8.8
H10	6.9	0.07	0.8	49	40	11	SC	11.5	1.8	0.4	3.7	30.3	olive grey	2.1	13.8	0.05	0.86	13.7
H11	5.7	0.16	2.2	54	33	14	SCL	11.5	5	1.2	40.3	257.1	black	6.3	20	0.14	0.18	18.1

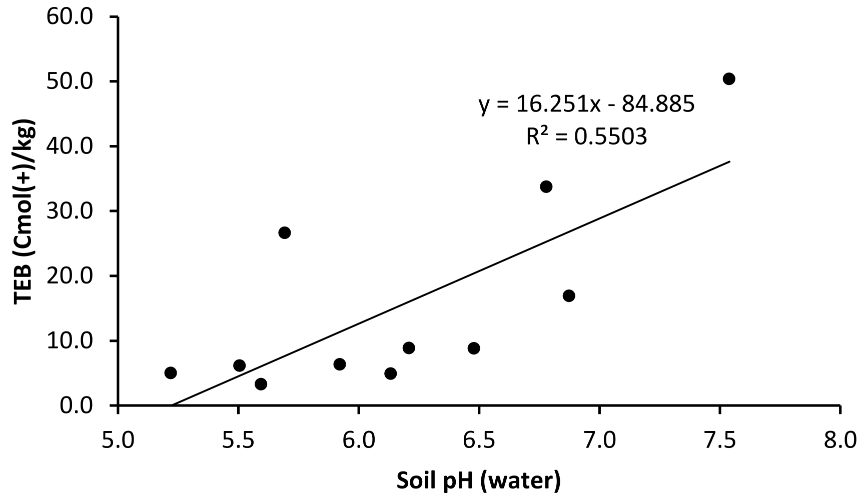


Figure 3.2. Relationship between exchangeable bases and soil pH of centroids of modal soil horizons generated during numeric clustering

A strong negative correlation (r = 0.96) between sand and clay contents within the horizons was also observed (Figure 3.3). This is consistent to observations that an increase in one of the two contents will lead to the decrease of the other. Except for modal horizons H8 with silt content 20% and H11 (14% silt), the rest of the horizons have silt fraction content of less than 11% (Table 3.1). These results suggest that the modal soil horizons are dominated by either sand or clay. This can be possible when we have lithologic discontinuity or cyclic deposition of soil materials in the landscape. Such situations are common in river valleys, where also the distance from the depositing river and flow rates largely dictate the texture of soil material to be deposited.

Characterization and classification of the sampled profiles discussed in detail in chapter 2, shows some profiles showing evidence of cyclic deposition of soil materials such as irregular decrease in soil organic carbon with soil depth. Description of the soils by Anderson (FAO, 1961) also showed dominance of cyclic sediment deposition as the soil forming process in the valley. These horizon clustering results are therefore consistent with soils in the study area.

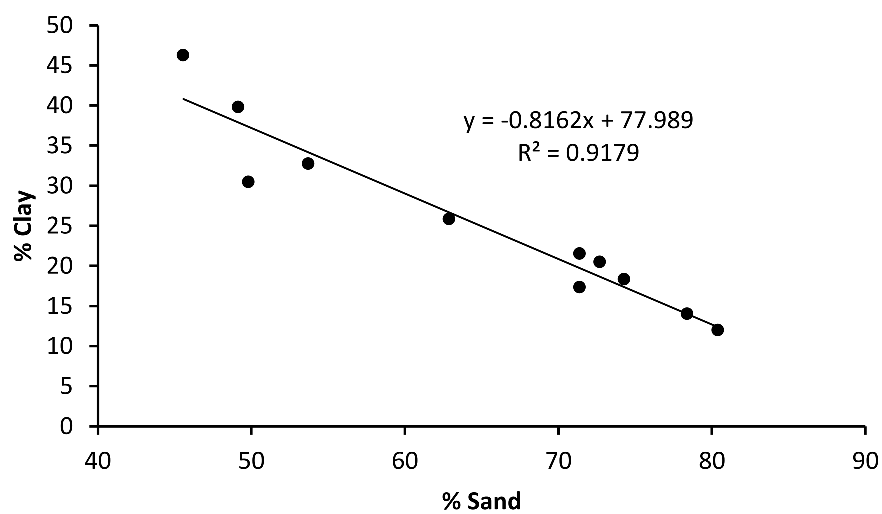


Figure 3.3. Relationships between sand and clay contents in the modal horizons generated by numeric clustering of soil horizons

The correlation between cation exchange capacity (CEC) and clay content is shown in Figure 3.4. There is weak correlation between the two attributes. This is also consistent with the raw data observations where CEC and clay were weakly correlated (Chapter 2). Calculation of CEC:Clay ratio shows that the area has mixed clay mineralogy including kaolinite, illite, smectite, and feldspars (Table 2.5 and Figure 2.13 in Chapter 2). The results also suggest that some of the CEC is coming from sources other than clay mineralogy. This is shown by having CEC:Clay ratios of greater than 0.95. Some materials are showing CEC clay ratio as high as 4.5 strongly suggesting CEC sources other than clays. As discussed earlier, a study of clay mineralogy in the study area could give more insights to this. The results suggest soil organic matter could be an important source of CEC in these soils. However, the horizon clustering results show that soil organic carbon (proxy to soil organic matter) values for the modal horizons are relatively low, ranging from 0.3 to 3.6% with mean value of 1.5% (Table 3.1).

With regard to soil color, the modal horizons with higher soil organic carbon were darker than those with less soil organic carbon (Fig. 3.5). This is consistent to the widely documented findings that the higher the soil organic matter the darker the soil (Jamer et al, 2010). When the soil organic carbon values for the modal horizons were plotted against their respective sum of converted soil Munsell colour to numerical red, green and blue (RGB) values, it showed that high organic carbon values correlated with low RGB values (Figure 3.6). The smaller the RGB numerical sum the darker the colour and vice versa.

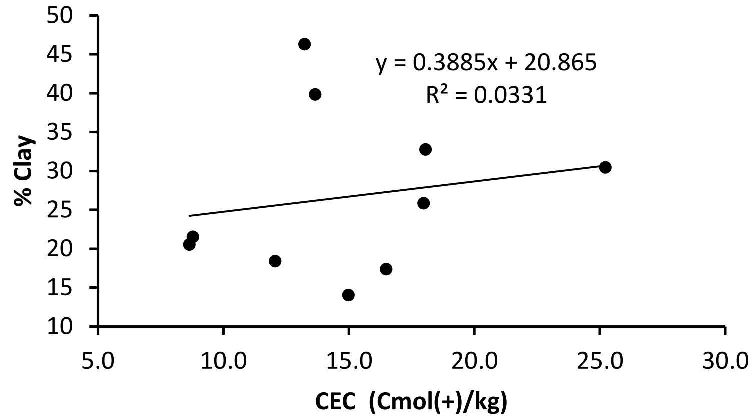


Figure 3.4. Relationship between CEC and clay content in the modal horizons generated in the numeric clustering of soil horizons

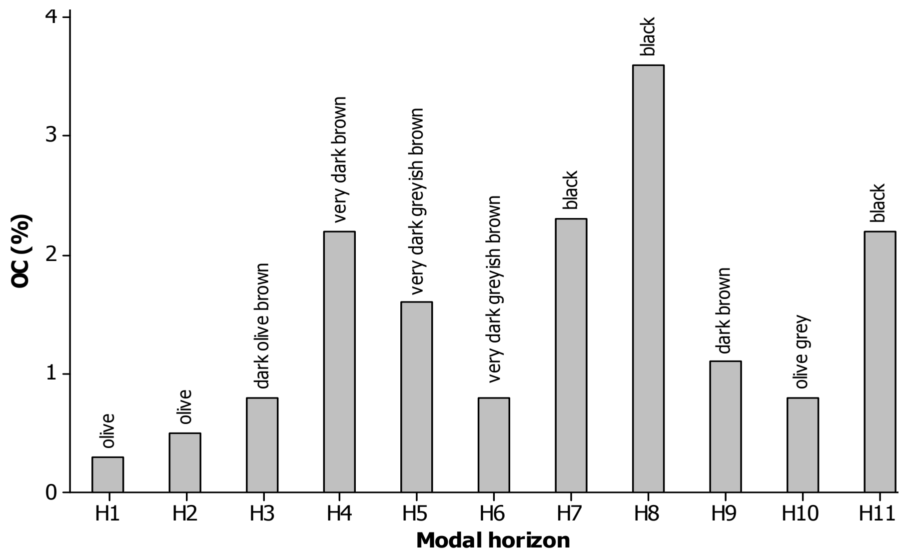


Figure 3.5. Relationship between soil organic carbon and soil colour in the modal horizons generated during clustering. H1 to H11 represent modal horizons

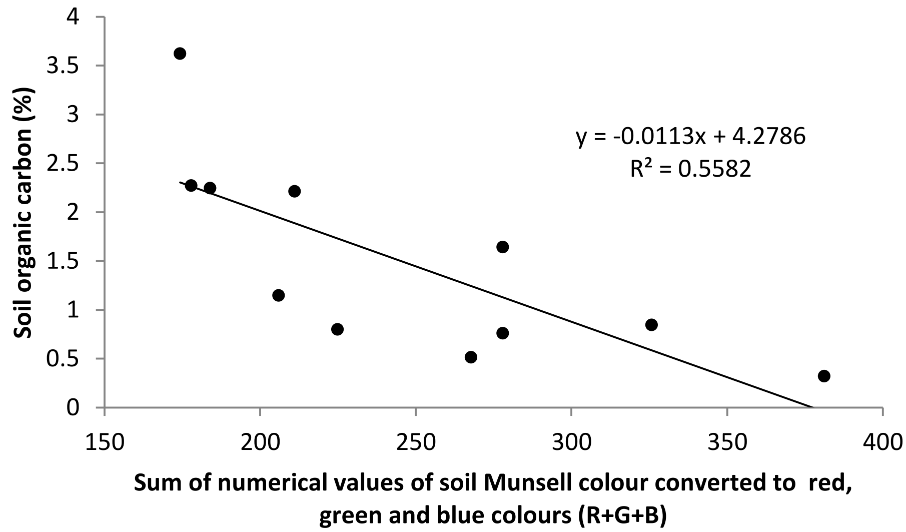


Figure 3.6. Relationship between modal horizons' soil organic carbon and the sum of red, blue and green (RGB) colours numerical values. The numerical values were obtained by converting the Munsell soil colours to numerical RGB colour values

Comparisons between Modal Horizons

One of the outputs of the horizon clustering is the table of distance metrics between one modal horizon and the rest. This output was used to assess the similarity between the clusters, since the shorter the calculated distance between the clusters the similar they are. In addition to distance metrics, basal depths of the modal horizons were used.

The distance metrics between modal horizons are shown in a matrix in Table 3.2. The smallest metric distance is 3.033 (comparing H3 against H2). This implies in all the 11 horizons clusters, horizon cluster H3 and horizon cluster H2 are the closest in similarity. The largest distance metric is 9.993 which is a result of comparison between H9 and H5. Horizon clusters H9 and H5 are therefore the most dissimilar clusters. The average metric distance for the comparisons is 6.9. The most similar clusters would have a metric distance comparison value of 0.

Some of the horizons which were shown to be relatively similar by smaller metric distances between them also occupied soil depths near to each other as depicted by their respective basal depths. For example, modal horizons H1 with basal depth at 79 cm and H2 with basal depth at 83 cm have a small distance metric difference of 3.43. Modal horizon H1 is also close to H10 with basal depth at 75 cm with distance metric of 3.85. This small difference between horizons could be indicative of gradual change of soil properties with depth in the landscape. Although these horizons are different enough to be assigned into different clusters by OSACA, they are virtually the same in terms of values of some attributes. For example, H1 and H2 have close values for pH, iron, organic carbon, copper and soil colour (Table 3.1).

Consistent with this, modal horizons with large distance metrics difference between them were shown to occupy relatively different locations in soil depths as shown with their respective basal depths. For example, modal horizon H5 (basal depth = 13) and H2 (basal depth = 83) are different by 9.28; modal horizon H5 (basal depth = 13) and H1 (basal depth = 79) are different by 9.24; modal horizon H7 (basal depth = 29) and H1 (basal depth = 79) are different by 9.34; while modal horizon H6 (basal depth = 28) and H1 (basal depth = 79) are different by 8.44.

Some modal horizons, though occupied basal depths that are relatively close to each other, were shown to be different by the metric distances comparisons. Modal horizon H7 (basal depth = 29) and H5 (basal depth = 13) were different by 9.44 metric distance. This could be an indication of buried horizons and/or lithological discontinuities.

Table 3.2. Matrix of distances metrics between modal horizons with respect to their basal depths. H1 to H11 represent modal soil horizons

Modal horizon	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	Basal depth (cm)
H1	0											79
H2	3.43	0										83
H3	3.19	3.033	0									47
H4	7.646	7.303	6.218	0								41
H5	9.244	9.28	8.865	8.366	0							13
H6	8.445	8.571	8.118	8.34	7.654	0						28
H7	9.335	8.551	8.065	8.066	9.438	7.647	0					29
H8	6.597	5.445	4.517	6.556	8.862	8.817	7.49	0				29
H9	5.404	4.887	3.853	7.001	9.993	6.891	8.093	5.223	0			35
H10	3.85	3.568	4.428	7.715	8.903	8.729	9.248	6.223	6.125	0		75
H11	6.845	6.117	5.256	7.064	8.261	7.535	7.104	4.462	5.383	6.636	0	34

3.3.2 Solum Clusters

Predicted Solum Clusters

OSACA generated 13 solum clusters from the specified range of 9 – 13 clusters. This range was picked based on the soil units on the legacy map which guided soil sampling. The legacy map (FAO, 1961) divided the study area in 9 soil units.

The 13 solum clusters with their respective soil profile memberships are shown in Table 3.3. The Table also shows Soil Taxonomy Sub group name of the representative pedon, that is the pedon with the shortest distance metric to the modal pedon of the cluster. Different clusters are represented by different soil groups. This suggests that clustering process was able to distinguish different soil groups in the study area. The results show that some clusters were comprised of as many as 5 solums while some were comprised only of 1 solum. The solum cluster which comprised the highest number of pedons is S58, while S63, S66, S67 and S68 are each comprised only of 1 pedon.

Table 3.3. Solum clusters and their representative solums. S56 to S68 represent generated solum clusters, Kilo_Ps stand for soil profiles (solums) names

Solum cluster	pedons in cluster	representative solum	classification (sub group)
S56	4	Kilo_P13	Fluvaquentic Endoaquepts
S57	2	Kilo_P10	Fluvaquentic Humaquepts
S58	5	Kilo_P41	Aquic Dystrudepts
S59	4	Kilo_P23	Aeric Umbric Endoaqualfs
S60	3	Kilo_P8	Umbric Endoaqualfs
S61	3	Kilo_P28	Fluvaquentic Humaquepts
S62	3	Kilo_P40	Fluvaquentic Hapludolls
S63	1	Kilo_P34	Mollic Fluvaquents
S64	1	Kilo_P22	Typic Endoaquepts
S65	4	Kilo_P25	Aquic Udifluvents
S66	1	Kilo_P36	Typic Endoaquerts
S67	1	Kilo_P37	Oxyaquic Eutrudepts
S68	1	Kilo_P35	Aeric Endoaquents

The spread within the solum clusters is depicted in Figure 3.7. Apart from profiles in clusters S63, S66, S67 and S68 which are zero distance from the modal soil profile (they are actually the modal soil profiles), the rest of the solum clusters have their constituent soil profiles at varying distances from their modal profiles. Solum cluster S56 appears to have most of its constituent soil profiles close to the modal profile followed by S62 and S65. This implies the similarities within S56 profile clusters is higher, followed by S62 and S65 in that order. The dissimilarities within solum clusters S58 and S61 are relatively higher compared to other clusters.

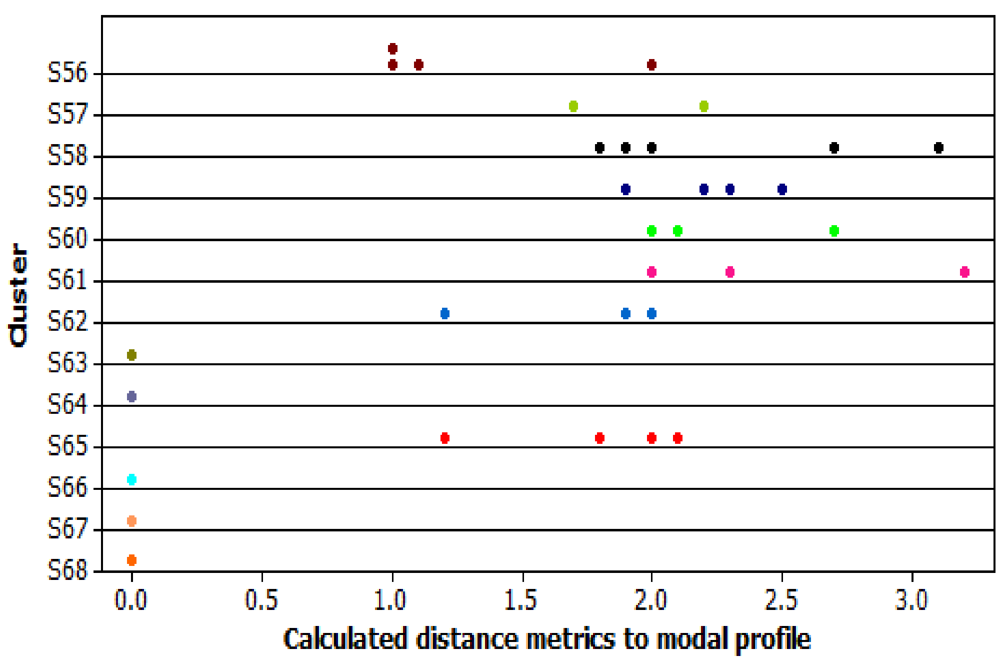


Figure 3.7. Distance metrics between constituent solum clusters' pedons and the clusters' modal pedons. The same colour dots represent constututients solums in a solum cluster. S56 to S58 represent solum clusters

Properties of Predicted Solum Clusters

One of the outputs of the solum clustering in OSACA is modal solum. The modal solums are populated with attribute data generated as centroids in this *k*-means based clustering. To determine cluster membership, the algorithm evaluated the distance between a point and the cluster centroids. Table 3.4 shows the centroid attribute values for the 13 solum clusters. The continuous vertical variability centroids of some attributes for each of the solum clusters are shown in Figures 3.8 – 3.13 and are discussed below.

Table 3.4. Attribute values of the generated clusters' modal solums. S56 to S68 represent soil clusters, TN = Total nitrogen, OC = organic carbon, P = available phosphorus, Cu = copper, Zn = zinc, Mn = manganese, Fe = iron, Ca = calcium, Mg = magnesium, K = potassium, Na = sodium, CEC = cation exchange capacity

Soil Cluster	Horizon	Basal depth	pH	TN	OC	P	Cu	Zn	Mn	Fe	Sand	Clay	Silt	Text. Class	Ca	Mg	K	Na	CEC
		(cm)	(water)	(%)		(mg/kg)					(%)				(Cmol(+)/kg)				
S56	1	21	5.6	0.09	0.82	8.3	2.4	0.8	26.0	169.3	76	16	8	SL	1.4	1.0	0.07	0.3	9.8
	2	39	5.9	0.06	0.47	3.8	1.5	0.1	5.0	35.8	81	14	5	SL	1.2	0.7	0.05	0.2	7.8
	3	63	6.2	0.04	0.26	2.8	0.7	0.1	4.1	13.7	82	14	5	SL	1.2	0.8	0.05	0.3	8.1
	4	88	6.7	0.04	0.26	3.1	0.5	0.1	7.3	6.3	76	19	5	SL	1.1	1.2	0.06	0.3	9.0
S57	1	20	5.4	0.12	1.45	20.2	2.9	1.1	28.6	111.6	72	22	6	SCL	1.0	1.1	0.23	0.5	13.8
	2	42	5.7	0.07	0.81	11.8	3.0	0.3	15.3	66.8	63	26	11	SCL	1.0	1.2	0.12	0.4	11.9
	3	68	6.0	0.05	0.35	1.2	1.0	0.3	7.1	16.3	66	28	7	SCL	0.6	1.3	0.09	0.4	8.4
	4	96	6.2	0.05	0.38	1.2	0.6	0.3	4.6	13.7	53	40	7	SC	1.0	1.6	0.10	0.4	8.7
	5	102	6.2	0.05	0.36	0.4	0.3	0.3	3.2	2.4	52	40	9	SC	1.1	2.2	0.10	0.5	9.3
S58	1	19	6.1	0.19	2.75	94.2	2.3	1.0	13.3	101.0	68	20	13	SCL	5.8	3.6	0.31	0.2	16.6
	2	35	6.1	0.08	1.39	60.9	2.7	0.2	4.4	80.4	72	18	9	SL	3.5	2.2	0.15	0.2	17.8
	3	53	6.0	0.07	0.70	53.0	2.1	0.1	5.5	54.9	80	12	8	SL	2.4	1.9	0.09	0.2	14.4
	4	72	6.0	0.06	0.77	40.3	2.6	0.2	5.9	53.0	78	14	8	SL	2.4	2.6	0.08	0.2	11.4
	5	83	6.2	0.04	0.27	23.3	1.5	0.2	4.4	34.4	82	12	6	SL	2.3	4.3	0.06	0.2	8.9
	6	86	6.2	0.04	0.40	26.1	1.8	0.2	4.8	40.9	80	13	7	SL	2.4	4.3	0.06	0.2	8.9
	7	88	6.2	0.09	1.08	26.4	2.7	0.3	8.6	74.5	70	18	12	SL	2.8	4.4	0.08	0.2	11.4
S59	1	21	5.2	0.39	5.17	17.7	2.3	0.7	7.5	119.0	49	30	21	SCL	1.4	0.3	0.17	0.4	33.2
	2	38	5.4	0.22	2.35	9.0	2.6	0.2	15.5	68.2	42	43	16	C	1.0	0.4	0.07	0.3	26.0
	3	59	5.6	0.12	1.21	9.8	2.2	0.2	14.9	60.4	48	44	8	SC	1.1	0.7	0.05	0.4	15.7
	4	91	5.9	0.08	0.47	10.3	1.4	0.2	15.3	18.5	57	36	7	SC	1.6	1.1	0.06	0.4	11.9
	5	99	6.0	0.07	0.66	8.3	1.8	0.5	17.9	36.8	44	47	9	C	1.5	1.3	0.07	0.5	12.8

Continued

Table 3.4 continued

Soil Cluster	Horizon	Basal depth	pH	TN	OC	P	Cu	Zn	Mn	Fe	Sand	Clay	Silt	Text. Class	Ca	Mg	K	Na	CEC
		(cm)	(water)	(%)		(mg/kg)					(%)				(Cmol(+)/kg)				
S60	1	20	5.7	0.25	3.54	31.2	5.2	1.9	30.0	107.5	47	34	19	SCL	2.3	2.5	0.14	0.5	24.8
	2	48	6.3	0.14	1.45	23.5	2.5	0.4	9.3	32.5	40	49	10	C	1.0	3.2	0.10	0.6	18.4
	3	72	6.4	0.17	0.61	6.8	1.2	0.6	21.5	10.5	42	47	10	C	1.2	5.0	0.07	0.6	13.3
	4	86	6.6	0.09	0.67	1.7	1.3	0.5	21.7	11.6	45	45	10	C	1.8	5.8	0.07	0.9	12.3
	5	89	6.7	0.13	0.69	17.7	1.7	0.5	21.6	10.8	46	45	9	SC	1.5	5.6	0.07	0.8	12.1
	6	95	6.7	0.11	0.60	0.9	1.0	0.4	21.3	9.9	49	44	7	SC	1.3	5.4	0.07	0.7	11.9
S61	1	21	5.7	0.17	2.78	23.5	5.1	1.1	17.4	308.8	50	35	15	SC	5.4	6.0	0.22	0.3	23.2
	2	39	5.8	0.12	1.64	6.5	4.5	0.6	16.3	208.2	48	36	15	SC	4.3	4.5	0.10	0.2	14.3
	3	65	6.3	0.06	0.60	5.4	1.9	0.3	35.2	40.8	55	35	10	SC	4.3	4.9	0.08	0.3	17.3
	4	82	6.4	0.05	0.37	6.9	1.8	0.3	34.7	31.8	61	28	10	SCL	3.9	6.5	0.08	0.3	13.5
	5	91	6.7	0.04	0.22	5.8	1.3	0.4	34.8	24.3	61	28	10	SCL	3.4	9.4	0.08	0.3	13.2
S62	1	12	5.9	0.13	1.94	20.6	3.7	1.2	55.0	275.3	68	21	12	SCL	6.2	21.0	0.16	0.1	11.7
	2	21	6.1	0.10	1.55	8.1	3.9	0.9	56.3	230.3	68	21	11	SCL	5.9	21.0	0.07	0.1	13.7
	3	44	6.4	0.04	0.52	3.2	2.5	0.2	22.4	36.2	70	21	9	SCL	4.8	20.4	0.06	0.1	12.6
	4	67	6.3	0.05	0.48	13.1	3.5	0.3	20.9	70.6	69	19	12	SL	3.9	19.7	0.04	0.1	9.9
	5	90	6.5	0.04	0.28	4.5	1.4	0.2	11.1	16.5	80	14	6	SL	3.4	18.3	0.04	0.1	7.7
S63	1	16	5.6	0.36	4.60	2.8	5.4	0.2	10.2	108.7	64	16	20	SL	3.0	6.3	0.04	0.1	28.6
	2	25	6.4	0.13	1.22	8.2	2.5	0.0	1.5	18.6	48	34	18	SCL	3.0	9.4	0.06	0.3	13.8
	3	43	6.8	0.07	0.72	7.4	3.4	0.2	0.7	16.3	48	36	16	SC	3.5	24.8	0.03	0.6	13.2
	4	68	7.2	0.04	0.30	3.5	1.6	0.1	1.2	8.4	64	24	12	SCL	3.5	30.0	0.03	0.6	13.2
	5	90	7.3	0.03	0.34	0.9	1.3	0.1	3.2	6.2	64	28	8	SCL	1.9	33.4	0.02	0.6	8.6

Continued

Table 3.4 continued

Soil Cluster	Horizon	Basal depth (cm)	pH (water)	TN (%)	OC (%)	P (mg/kg)	Cu (mg/kg)	Zn (mg/kg)	Mn (mg/kg)	Fe (mg/kg)	Sand (%)	Clay (%)	Silt (%)	Text. Class	Ca (Cmol(+)/kg)	Mg (Cmol(+)/kg)	K (Cmol(+)/kg)	Na (Cmol(+)/kg)	CEC (Cmol(+)/kg)
S64	1	10	5.4	0.06	0.43	23.5	0.8	0.2	3.0	34.6	87	10	3	LS	1.5	0.3	0.12	0.2	9.2
	2	26	5.1	0.04	0.22	9.7	0.5	0.1	1.5	13.7	89	8	3	LS	1.1	0.2	0.12	0.2	9.8
	3	55	5.0	0.06	0.57	7.8	1.7	2.4	3.2	58.2	77	16	7	SL	1.2	0.2	0.06	0.2	10.4
	4	79	5.2	0.05	0.49	8.9	1.4	0.3	2.4	45.9	79	16	5	SL	1.1	0.2	0.06	0.2	8.6
	5	120	5.5	0.04	0.35	1.3	0.8	0.1	3.0	19.6	75	20	5	SCL	0.7	0.3	0.06	0.2	8.2
S65	1	16	5.7	0.15	1.55	40.3	2.3	0.9	11.2	96.2	70	22	8	SCL	0.7	1.1	0.10	0.3	13.5
	2	33	5.8	0.09	0.86	19.2	2.1	0.4	5.2	44.5	74	19	6	SL	0.9	1.0	0.10	0.3	11.0
	3	67	6.5	0.05	0.20	8.4	0.5	0.2	3.5	8.5	77	17	6	SL	0.7	1.0	0.08	0.3	9.6
	4	98	6.3	0.04	0.30	10.7	0.4	0.2	7.7	11.4	75	19	5	SL	0.6	1.3	0.05	0.3	7.9
	5	105	6.4	0.03	0.25	10.5	0.7	0.2	7.3	13.8	79	17	4	SL	0.5	1.3	0.04	0.3	7.2
S66	1	10	5.4	0.50	5.75	4.8	6.9	1.4	13.5	311.8	56	26	18	SCL	5.1	18.4	0.31	0.1	11.2
	2	21	5.1	0.49	3.42	3.5	5.7	1.7	10.6	261.3	36	28	36	CL	3.5	12.0	0.26	0.2	33.2
	3	38	5.1	0.24	2.85	17.5	4.2	1.3	13.2	264.0	36	44	20	C	8.3	17.5	0.03	0.2	35.0
	4	48	5.2	0.12	2.40	1.5	5.6	0.9	18.5	275.3	36	46	18	C	10.5	32.4	0.02	0.1	21.6
	5	74	5.3	0.10	0.76	0.3	7.0	1.4	34.4	230.3	42	44	14	C	5.1	36.4	0.02	0.1	21.0
	6	90	5.4	0.03	0.19	1.9	4.1	1.5	67.8	7.3	60	34	6	SCL	5.7	50.3	0.06	0.2	16.2
S67	1	13	7.5	0.13	1.64	38.4	1.5	0.9	11.2	26.4	80	12	8	SL	6.7	43.1	0.41	0.1	57.4
	2	28	6.8	0.06	0.76	8.9	2.4	3.5	5.9	48.9	78	14	8	SL	4.1	29.0	0.54	0.1	15.0
	3	43	5.9	0.05	0.65	12.7	2.5	0.6	6.8	75.9	80	14	6	SL	1.9	20.1	0.35	0.1	10.8
	4	82	5.6	0.05	0.34	14.2	2.1	0.2	4.0	45.5	76	20	4	SCL	1.9	21.9	0.21	0.1	10.2
	5	100	6.0	0.03	0.27	2.6	2.3	0.2	7.6	26.4	68	26	6	SCL	4.1	26.6	0.11	0.1	8.0
S68	1	11	5.4	0.41	5.94	1.9	3.8	0.6	16.9	177.0	58	18	24	SL	4.1	13.0	0.22	0.2	9.8
	2	24	5.6	0.37	5.33	3.4	4.1	0.4	11.0	193.8	62	16	22	SL	3.0	7.5	0.07	0.2	29.8
	3	40	6.2	0.32	3.73	10.4	5.0	0.1	3.5	109.6	58	18	24	SL	2.5	7.1	0.04	0.4	27.8
	4	55	6.8	0.07	2.97	5.9	2.1	0.2	4.3	23.0	46	38	16	SC	3.0	14.4	0.03	0.9	27.6
	5	80	7.3	0.05	0.57	6.3	1.3	0.2	5.1	151.7	54	30	16	SCL	2.5	16.7	0.02	1.1	11.6

Available Phosphorus (P)

The differences in trends for available phosphorus (P) for the 13 soil clusters were obvious (Figure 3.8). Except for S63 and S68 which initially showed some increase in available P with depth to around 20 cm, the rest of the solums showed decreasing P down the soil. The low P at the surface compared to the subsurface could be attributed to deposition of new materials on the surface which are relatively richer in P. This can also be explained by higher soil organic matter in the topsoils. The solum cluster members most similar to the modal cluster for S63 and S68 were both classified as Aquents and located on the same landscape level where surface deposition of new material was active. The levels of available P and trends down the solum are almost peculiar for each cluster, indicating the differences between the clusters. Soil cluster S58 appears to present soil profiles which have abnormally high available P comparing to the other profiles, especially on the upper horizon. The reason could be inherent high P in the soils from parent material. Three out of four of the soil profiles allocated in this cluster are located close to the escarpment where soils are generally developed from pale, non alluvial ferruginous materials sometimes interspersed or overlain by thin cover of alluvium (Geological Survey of Tanganyika, 1963). Fertilization was ruled out because records do not show high application of fertlizers where the profiles are located and the relatively high available P values were observed also down the slope.

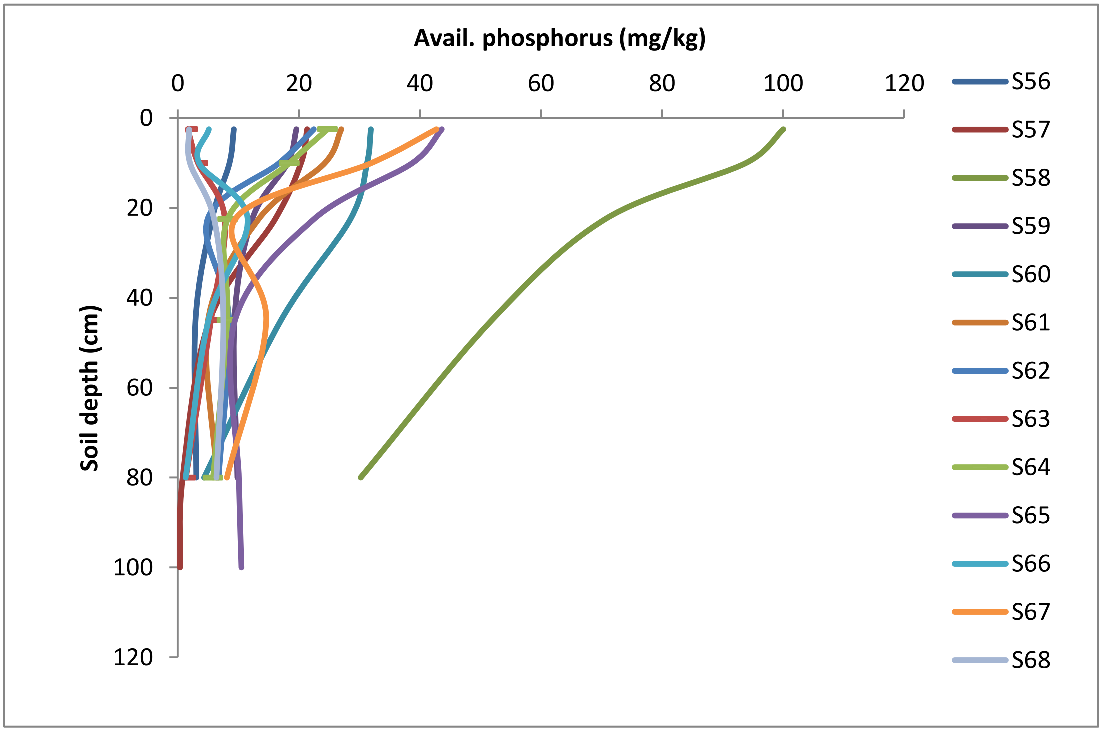


Figure 3.8. Variation of available phosphorus with depth in modal solums. S56 to S68 represent soil clusters

Percent Base Saturation

Percent base saturation for clusters S64, S65, S57 and S56 do not differ much in amount and they show somewhat similar trend with increase in soil depth (Figure 3.9). These clusters also show low base saturation (between 15 and 25%) throughout their studied profile depths, and as expected they also have low pH (Figure 3.13).

Solum cluster S63 showed an abrupt increase in base saturation with soil depth while also S59, S61 and S67 showed increase in base saturation with depth. Increase in base saturation with depth is normally a result of leaching process where the bases are washed down the soil profile with infiltrating water. The abrupt increase in S63 can be explained by abrupt change in soil texture between the top and the subsoil. Topsoil is sandy clay loam while the subsoil is clay (Table 3.4). The bases, therefore leaches easily from the topsoil due to relatively low CEC of coarse textured soils and are held in the finer soil down the solum since finer soils generally have higher CEC.

Trends showing repeated increase-decrease and vice versa with depth have been displayed by clusters S61, S58, S66 and S68. These are indications of lithologic discontinuity, whereby bordering soil layers have materials coming from different parent materials or deposited cyclically one over the other. For example, the field representative of S58 was found to have a buried B horizon while the representative of S66 showed evidence of lithologic discontinuity by having a profile sequence of Ap-2Bssg1-2Bssg2-2Bssg3-3Bg.

Solum cluster S62 maintained 100% base saturation throughout implying that its exchange sites are saturated by basic cations throughout the studied depth. The three pedons which fall under this cluster are all classified as Mollisols (Fluvaquentic Endoaquolls, Typic Endoaquolls, and Fluvaquentic Hapludolls). Mollisols are high base soils.

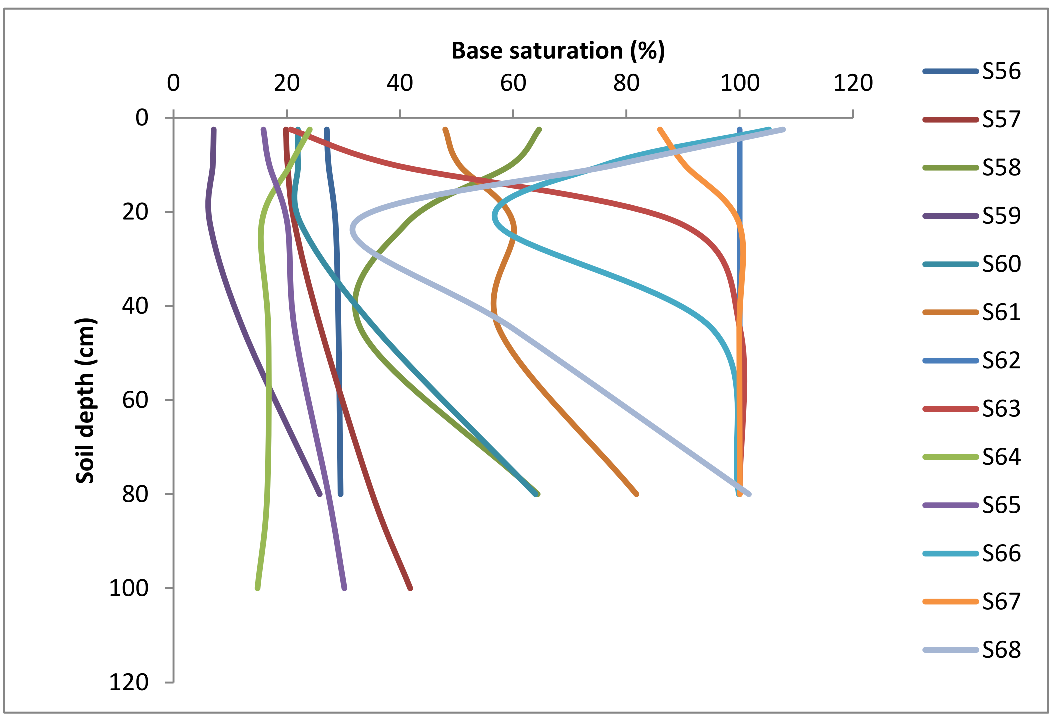


Figure 3.9. Variation of base saturation with depth in modal solums. S56 to S68 represent solum clusters

Cation Exchange Capacity (CEC)

The topsoil CEC values for clusters S56, S64, S65, S62, and S57 are relatively low, varying between 9 and 15 Cmol(+)/kg (Figure 3.10). These clusters and cluster S58 also show that their CEC values do not differ much with soil depth. The most probable cause for the low CEC and the observed trend could be the soil texture and soil organic matter contents in these clusters. Low clay contents are observed to be associated with these solum clusters. For example, S56 has clay content between 12 and 17%, S64 has between 8 and 20%, S65 between 15 and 22%, and S62 between 15 and 21% throughout the profile depth (Figure 3.11). Soil organic matter values, another significant contributor to CEC values in soils, are very low in these soil clusters. Figure 3.12 shows organic carbon contents (proxy to soil organic matter) for the above clusters are generally below 1% throughout the soil depths of the clusters.

Clusters S61, S63, S59, S60 and S67 CEC values are higher in the top soils and keep decreasing with increase in soil depths (Figure 3.10). The CEC trends in S61 and S67 are consistent to decreases in both clay content and soil organic matter (Figures 3.11 and 3.12 respectively). The CEC trends for clusters S63, S59, and S60 cannot be explained by clay contents variations in the clusters since clay trends were following an opposite direction (increasing with depth while CEC was decreasing with depth). However, the CEC trend can, at least in part be explained by the soil organic matter trends. Figure 3.12 shows high organic contents in top soils which kept decreasing with solum depths.

Solum cluster S63 has 5% organic carbon which decreased sharply to 2% at around 20 cm soil depth. S59 topsoil had 5.5% organic carbon which decreased to 2.5% at 30 cm soil depth, meanwhile S60's organic carbon decreased to 2.5% at 30 cm soil depth from 4% of the topsoil. These organic carbon trends are reflected by the trends of CEC.

Two solum clusters; S68 and S66 CEC values increased with depth to a certain point and then started to decrease. S68 and S66 CEC values respectively, increased with depth from around 10 Cmol(+)/kg in the top soil to 30 and 35 Cmol(+)/kg respectively at 30 cm soil depth. S68 CEC values then decreased to around 10 Cmol(+)/kg at 80 cm while S66 values decreased to around 18 Cmol(+)/kg at 50 cm soil depth (Figure 3.10). These trends are consistent to trends of clay contents in the respective clusters (Figure 3.11).

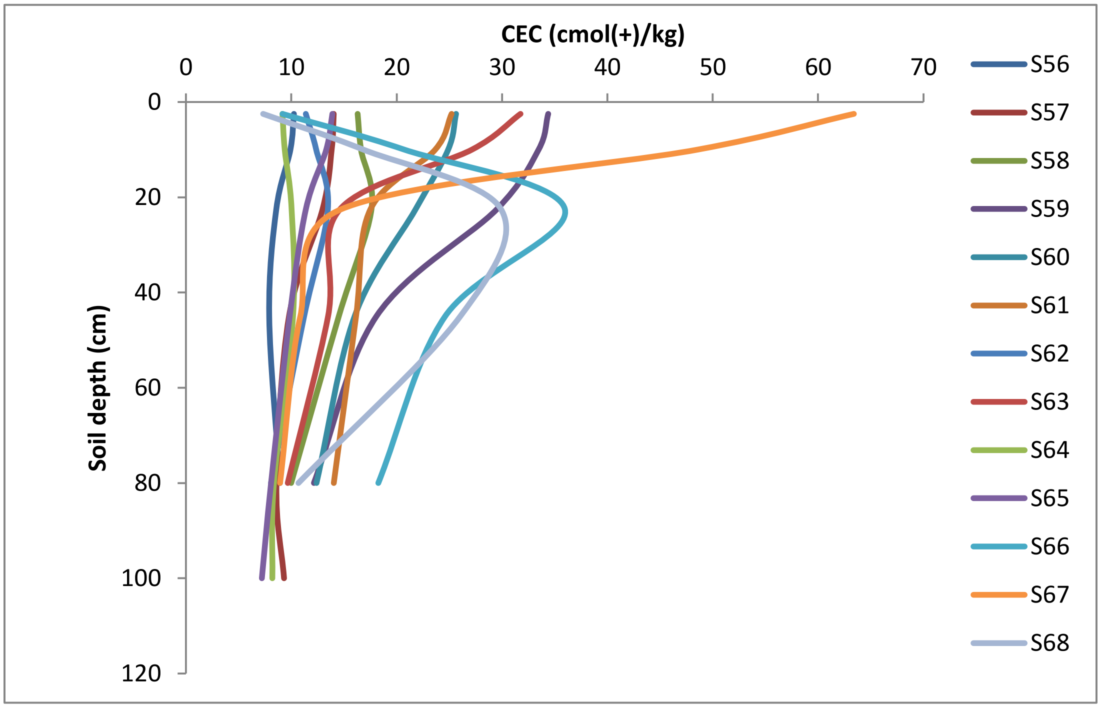


Figure 3.10. Variation of cation exchange capacity (CEC) with depth in modal solums.

S56 to S68 represent solum clusters

Clay Content

The continuous vertical variability of soil clay contents for the 13 solum clusters is shown on Figure 3.11. Three trends can be observed; (1) Clusters in which clay contents vary very slightly with soil depth, (2) clusters in which clay content increase with soil depth to certain point in the profile and then decreasing forming a bulge shape in the figure, and (3) clusters which show decrease in clay content from the topsoil and those forming repeated patterns of decrease and increase down the soil profile.

Clusters in which clay contents vary slightly with soil depths are represented by clusters S58 and S62. The pattern suggests that there has been very slight or no clay migration in these soil profiles. Such types of soils are normally located in areas where water infiltration and pedoturbation activities are limited. It can also be a property of very young soils which have not been exposed to pedogenic activities. In the study site which is an alluvial basin, these soils can be located in deep recent depositions. This is evidenced by morphological description of the soil profiles which constitute these clusters. Kilo P39, kilo P41, kilo P29 and kilo P20 all are having slightly altered cambic (Bw) subsurface horizons. Sequence of horizons for Kilo P20 is Ap-Bw1-Bw2-Ab-2Bw1-2Bw2-3Ab, while that of Kilo P29 is Ap-A-Bw1-Bw2-Bw3.

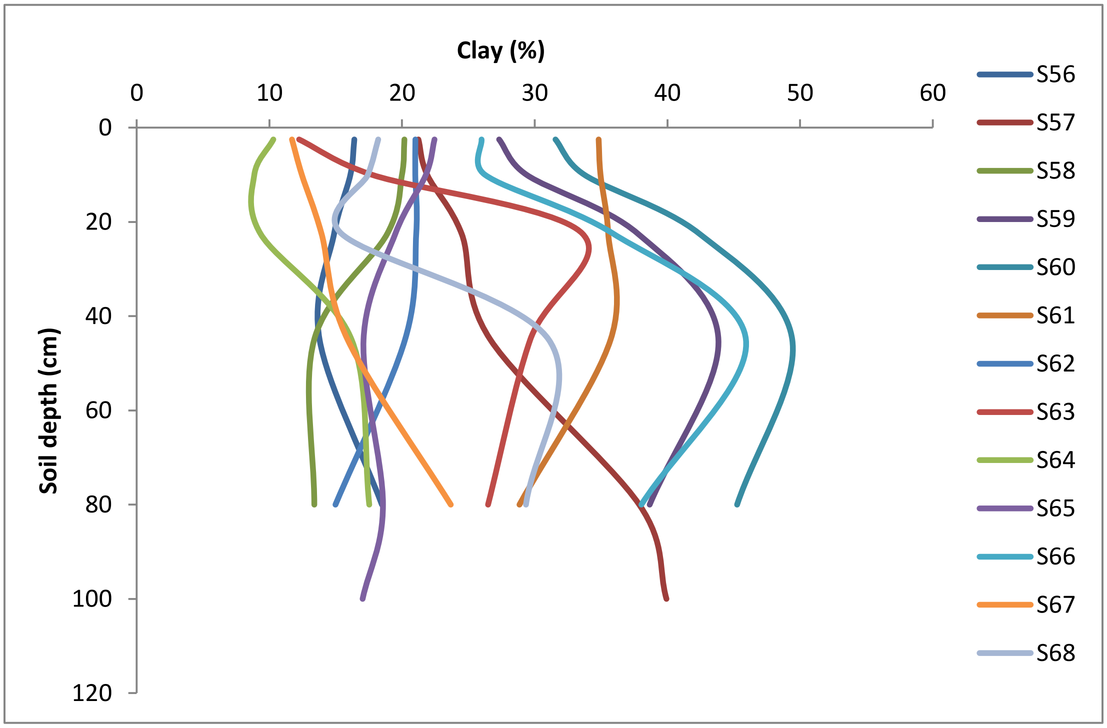


Figure 3.11. Variation of clay content with depth in modal solums. S56 to S68 represent solum clusters

The second group is depicted by solum clusters S63, S57, S66, S59, and S60. The soils which exhibit increase in clay contents in the subsoils are relatively old. The clays are normally translocated from higher horizons through illuviation process resulting to Bt horizons. This happens in relatively stable landforms where no new deposits are done or top soils are not removed by erosive processes. In alluvial basins like Kilombero, this can likely happen in fans which are no longer receiving deposition due to change of channel courses or reduced flooding levels and frequencies. For example, pedons Kilo P10 and Kilo P24 were both located on fans. Deposition contrasts could be another explanation as evidenced by pedon Kilo P36 of cluster S66.

The third group is presented by solum clusters S64, S56, S68, and S65. These solums show repeated patterns of increase and decrease of clay contents with increase in soil depth. This pattern could be caused by cyclic deposition of sediments which may have different clay contents. Higher clay contents on the topsoil could be due to inherently higher clay contents in the sediments or due to overlying a horizon which already had some clay elluviation. This third group is common in alluvial basins which still have active seasonal flooding. The representative soil profile for clusters S56 clasifies to Fluvaquentic Endoaquepts and cluster S65 classifies to Aquic Udifluvents, indicative of locations where soils are formed by cyclic deposition of materials.

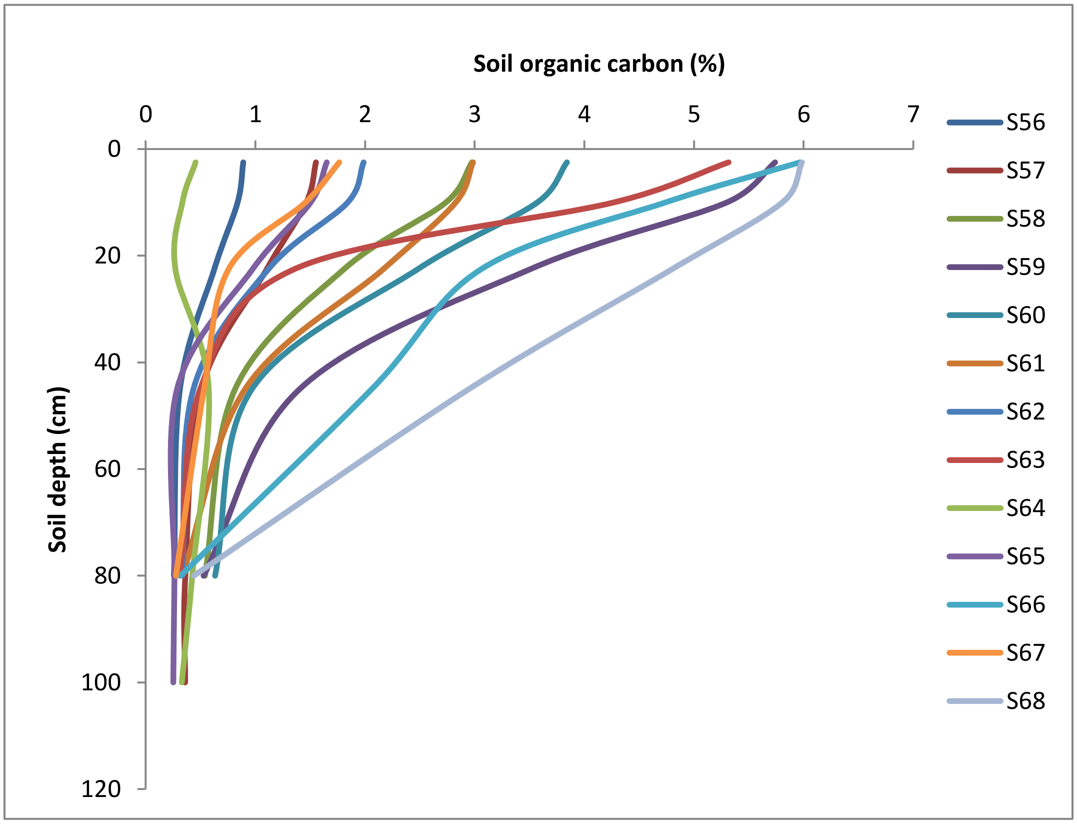


Figure 3.12. Variation of soil organic carbon with depth in modal solums. S56 to S68 represent solum clusters

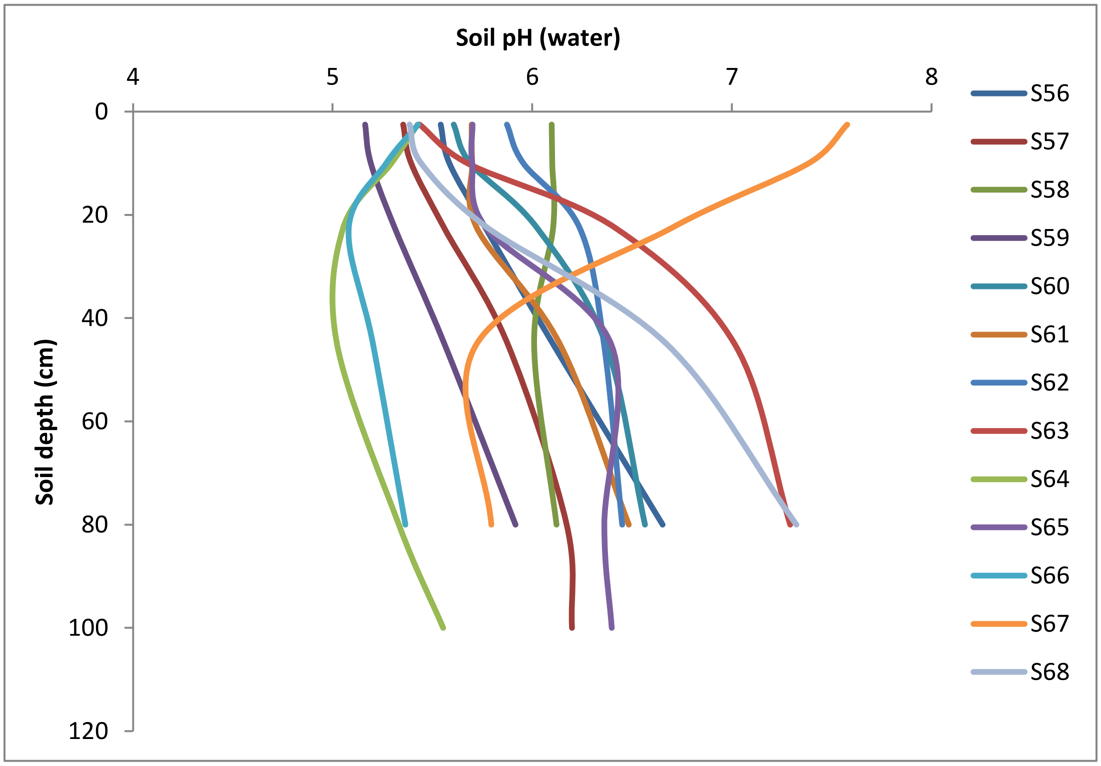


Figure 3.13. Variation of soil pH with depth in modal solums. S56 to S68 represent solum clusters

3.4 Conclusion

Numerical classification using distance metrics on a web based java application called OSACA was employed to cluster described and sampled pedons. In the first step, the soil horizons were clustered into 11 classes. The centroids for each horizon cluster were also generated to give modal soil horizons.

The reliability of the clustering was validated by studying interrelationships between modal horizon attributes. As expected, a positive correlation ($r = 0.74$) was observed between base saturation and soil pH. A strong negative correlation ($r = 0.96$) between sand and clay contents within the horizons was also observed. It was also observed that the modal horizons with higher soil organic carbon were darker than those with less soil organic carbon. These relationships are consistent with pedological knowledge, suggesting the reliability of the clustering results.

Further, OSACA was used to cluster the soil profiles. A total of 13 solum clusters with their modal pedons attributes were generated. The results show that some clusters were comprised of as many as 5 solums while some clusters were comprised only of single solum. Similarities within the clusters were also variable as depicted by distance metrics between the cluster member pedons and the clusters' modal pedons.

Selected modal solum attributes were used to assess continuous variability with soil depth to see how different the soil clusters are. These included CEC, Organic carbon, base saturation, clay content, soil pH and available phosphorus. Differences in selected attributes values and variability of trends with depth between the clusters confirmed that the clustering process was able to separate different soil types present in the study area. Different soil forming processes were also picked. For example cyclic deposition of sediments was shown by irregular decrease in soil organic carbon with depth. Likewise, clay illuviation to form Bt horizons was shown by increased clay content in the subsoil.

3.5 References

Bidwell, O.W., and F.D. Hole. 1964a. An experiment in the numerical classification of some Kansas soils. Soil Science Society of America Proceedings 28:263-268.

Bidwell, O.W., and F.D. Hole. 1964b. Numerical taxonomy and soil classification. Soil Science 97: 58-62.

Bidwell, O.W., L.F. Marcus, and P.K. Sarkar. 1964. Numerical classification of soils by electronic computer. Transactions of International Congress of Soil Science 5:933-941.

Bishop, T.F.A., McBratney, A.B., and G.M. Laslett. 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91:27–45.

Brady, N.C. and R. Weil. 2002. The Nature and Properties of Soils. Thirteenth Edition, Prentice Hall, Upper Saddle River, NJ.

Buol, S.W., R.J. Southard, R.C. Graham, and P.A. McDaniel. 2002. Soil Genesis and classification. 5th Edition. Iowa State University Press, Ames, IA.

Carré, F., and A.B. McBratney. 2005. Digital terron mapping. Geoderma 128:340-353.

Carré, F., and M. Jacobson. 2009. Numerical classification of soil profile data using distance metrics. Geoderma 148:336–345.

Cook, S.E., A. Jarvis and J.P. Gonzalez. 2008. A new global demand for digital soil information. In: A.A. Hartemink, A. McBratney, and M.L. Mendonca-Santos, editors, Digital soil mapping with limited data. Springer. p. 31-41.

Diday, E. 1971. Une nouvelle méthode en classification automatique et reconnaissance des formes: la méthode des nuées dynamiques. Revue de Statistique Applique. 19:283-300. In French.

FAO. 1961. The Rufiji Basin Tanganyika. FAO Exp. Techn. Ass. Progr. No. 1269 Vol. 7. Rome.

FAO. 2006. Guidelines for Soil Description. FAO, Rome.

Fridland, V.M. 1974. Structure of the soil mantle. Geoderma 12:35-41.

Geological Survey of Tanganyika. 1963. Chombe: Quarter degree sheet No. 234. Geological Survey Division, Dodoma.

He, Y. 2013. mColorConverter. Apple Inc. CA, USA.

Hole, F.D., and M. Hironaka. 1960. An experiment in ordination of some soil profiles. Soil Science Society of America Proceedings 24: 309-312.

Jacobson, M., and F. Carré. 2006. OSACA version 1.0. Land Management and Natural Hazards Unit, Institute for Environment and Sustainability. European Commission, Italy.

Jamer, T., J. Hill, H. Lavée, and P. Sarah. 2010. Mapping topsoil organic carbon in non agricultural semi-arid and arid ecosystems of Israel. Photogrammetric Engineering and Remote Sensing 76(1): 85-94.

Lal, R., and M.K. Shukla. 2004. Principles of soil physics. Marcel Dekker, New York, NY.

Malone, B.P., McBratney, A.B., Minasny, B., and G.M. Laslett. 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154:138–152.

McBratney, A., B. Minasny, and R.V. Rossel. 2010. Numerical soil classification: A missed, but not a lost, opportunity. 19th World Congress of Soil Science, Soil Solutions for a Changing World, 1 – 6 August, Brisbane, Australia.

McBratney, A.B., and I.O.A. Odeh. 1997. Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma 77:85-113.

Minasny, B., and A.B. McBratney. 2007. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. Geoderma 142:285–293.

Muir, J.W., H.G.M. Hardie,R.H.E. Inkson, and A.J.B. Anderson. 1970. The classification of soil profiles by traditional and numerical methods. Geordema 4: 81 – 90.

Odgers, N.P., A.B. McBratney, and B. Minasny. 2011a. Bottom-up digital soil mapping I. Soil layer classes. Geoderma 163:38–44.

Odgers, N.P., A.B. McBratney, and B. Minasny. 2011b. Bottom-up digital soil mapping II. Soil series classes. Geoderma 163:30–37.

Odgers, N.P., Z. Libohova, and J.A. Thompson. 2012. Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. Geoderma 189–190:153–163.

Rayner, J.H. 1966. Classification of soils by numerical methods. European Journal of Soil Science 17:79-92.

Rizzo, R., J.A. M. Demattê, and F.S. Terra. 2014. Using numerical classification of profiles based on VIS-NIR spectra to distinguish soils from the Piracicaba region, Brazil. Revista Brasileira de Ciência do Solo 38:372-385.

Schaetzl, R.J., and S. Anderson. 2010. Soils: Genesis and Geomorphology Cambridge University Press.

Schoeneberger, P.J., D.A. Wysocki, E.C. Benham, and Soil Survey Staff. 2012. Field book for describing and sampling soils, Version 3.0. Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE.

Chapter 4: Predictive Soil Cluster Mapping of Soils of Kilombero

4.1 Introduction

4.1.1 Soil Information in Tanzania

Collection and application of soil information for agricultural land use in many parts of the world is as old as agriculture itself (Deckers et al., 2002). This is not different in Tanzania. Under the traditional land use planning and management system, land users used their local knowledge and experience to identify and earmark land for different uses. Soil types were distinguished by vernacular names such as "*Kikungu*" for red soils, "*Ipwisi*" for pale brown light soils, and "*Ibushi*" for dark brown soils in parts of Tabora region (Acres, 1983). Identification of suitability of these soils for major land uses types such as crops growing, grazing, fuelwood, and residential purposes was also done (Kauzeni et al., 1993).

The first documented soil survey in Tanzania was done by Milne (1936). His classification put Tanzanian soils into 9 groups: saline soils, volcanic soils, podzolized soils, plain soils, laterized red earths, non-laterized red earths, loose sands, plateau soils, and desert soils. The usefulness of his classification was limited by it being based on a very general reconnaissance and limited analytical data. Another work was in 1954 by Calton who distinguished eluvial from illuvial soil types to produce a 1:4000000 soil map of Tanganyika (Calton, 1954). A soil map of the same scale, but covering the whole of East Africa was produced in 1963 by Scott, and a year later D'Hoore (1964) produced soil map for Africa at a scale of 1:5000000. These scales are too coarse for practical planning purposes at national and regional level. Baker (1970) classified the soil of Tanzania into 31 units by studying only 40 soil profile pits country wide, throwing doubts on its representativeness. Generally, the above works lacked reliability as they covered only the accessible areas and had very coarse scales (Wickama, 1997).

In 1977, Samki classified the soils of Tanzania in 19 groups based on soil ecological zones. He later used the legend of FAO-UNESCO to classy these soils and produced a 1:2000000 soil map (Samki, 1982). The map was used to prepare the district by district fertilizer recommendations (Samki and Harrop, 1984). In 1983, a compilation of different land evaluation and land capability analyses were used to produce a Soil Atlas of Tanzania at a scale of 1:2000000 (Hathout, 1983). No systematic soil mapping has been carried out to cover the entire country for a soil map of a scale finer than 1:1000000 since 1970 by Baker. De Pauw (1984) produced soil, physiography and agroecological zones map of the country at a scale of 1: 2000000 and in 1998, FAO through SOTER project produced a soil and terrain database for Tanzania at a scale of 1:2000000 (Eschweiler, 1998).

This brief review shows that Tanzania has a long history of collecting soils information. To date, a number of inventories at different scales have been made for various land use planning objectives. However, handling of basic soils and other land resources data are still to a great extent done manually and only a modest computerization has been done. Modern soil survey methods such as digital soil mapping techniques need to be applied in order to improve and harness the information from the fragmented land resource works currently available in the country. This will provide evidence based decisions for rational utilization of the country's soils and other land resources.

4.1.2 Concepts that Underlie the Recent Development in Soil Mapping

Recent advances in information and computational technology have created a vast potential through which soil information can be collected, mapped and communicated. Remote sensing, proximal sensors, geographic position systems (GPS), photogrammetry, geographic information systems (GIS), and decreasing costs of data storage allow efficient characterization, analysis, storage and display of vast amounts of data (Scull et al, 2003; McBratney et al, 2003)

Advances in Geostatistics have impacted soil mapping in a positive way. Geostatistical related methods such as trend surfaces, ordinary kriging, co-kriging, and regression kriging have been used to produce soil maps (Kiss et al, 1988; Heuvelink and Webster, 2001; Odeh et al, 1995, Kempen et al., 2009; Kerry et al., 2012). These methods are used to spatially interpolate soil property values at unmeasured sites from field-collected data.

Statistical methods such as linear regressions have been used in mapping works to exploit the relationship between quantifiable landscape indices and soil properties to create predictive soil maps (Scull et al, 2003). The methods include generalized additive models, generalized linear models, generalized least squares, principal component regression, and partial least squares. Scientists used these methods for mapping include Zheng et al. (1996), McBratney et al. (2000), Pachepsky et al. (2001), Park and Vlek (2002), and Kempen et al. (2009).

Decision tree analysis (DTA) is gaining popularity in predictive soil mapping. Its advantages over standard statistical techniques are its interpretability and its ability to deal with missing data and outliers. DTA also has advantage of being able to deal with nonadditive and nonlinear behavior, and it is not restricted by assumptions about data distribution (Scull et al, 2003; Odgers et al., 2014). Studies which employed decision tree analyses in soil mapping include Hansen et al. (2009), Haring et al. (2012) and Subburayalu et al. (2014).

DSM also has been developed to utilize expert knowledge. Soil surveyors accumulate information, knowledge and experience when working in the field and soil laboratories.

Expert systems make use of of the expertise and knowledge gained by the surveyor and integrate them into the predictive model (Scull et al., 2003). The expert systems are composed of information on spatial environmental variables, rules and facts related to soil variation, and an inference which combines the information and the rules to make logical conclusions (Scull et al., 2003). Expert systems differ from conventional models in that, they store and manipulate qualitative information and it allows for information to be easily updated.

These advances present favourable conditions to support the already popular shift from qualitative to quantitative soil mapping techniques (Lagacherie et al., 2007; Boettinger et al., 2010). The qualitative methods are referred to as conventional soil mapping (CSM) techniques, while the quantitative approaches are known as digital soil mapping (DSM) techniques.

Both DSM and CSM use a soil-landscape model to predict soil at unobserved locations. However, CSM soil-landscape model relies on expert knowledge, while the DSM soil-landscape model is based in numerical computations. Challenges in reproducing the maps, representation of continuous intergrading soil bodies as discrete homogeneous entities, and the lack of quantified measures of uncertainty are among the criticisms of CSM compared to DSM.

Despite their superiority to CSM, DSM methods have some limitations. CSM can better deal with the complex soil forming processes during mapping than by quantification of environmental explanatory variables used in DSM. Another limitation of DSM is the challenge on transferability of prediction models. This is because the DSM generally lack standaerdization as compared to CSM. Lastly, DSM depends on the availability of up-to-date soil data and environmental data layers. In areas with limited data, ability of DSM methods to produce reliable maps is reduced. Despite these challenges, DSM offers more flexibility and cheaper ways of producing the soil spatial information which is currently in high demand, especially in developing countries where it is lacking (Cook et al., 2008).

4.1.3 Soil Forming Factor and Digital Soil Mapping

DSM begins with the development of a numerical or statistical model of the relationship among environmental variables and soil properties. These are then applied to a geographic data base to create a predictive map. The environmental variables are derived from the factors which are perceived to have influence on soil formation process, or those which influence a property of soil in study. These are presented in Jenny's equation (Jenny, 1941);

$S = f(c,o,r,p,t)$

Where; '*S'* stands for soil, '*c'* for climate, '*o'* for organisms including vegetation and humans, '*r'* for reliefs, '*p'* for parent material, and '*t'* representing time. Innumerable soil scientsist have used this equation to understand and map the soils. These include qualitative works, mainly before the advent and wide application of sophisticated numerical methods, and later on, quantitative methods (McBratney et al., 2003).

4.1.4 The *scorpan* Model

Of late, a Jenny-like formulation called *scorpan* has been published for empirical quantitative descriptions of relationships between soil and other spatially referenced factors with a view to using these as soil spatial prediction functions (McBratney et al., 2003). This formulation known as *scorpan* model has seven factors compared to five of Jenny's formulation and is written as

Sc = f(s,c,o,r,p,a,n) or Sa = f(s,c,o,r,p,a,n)

Where;

Sc: stands for soil class

Sa: represents soil attribute

s: standing for soil, other properties of the soil at a point;

c: representing climate, climatic properties of the environment at a point;

o: for organisms, vegetation or fauna or human activity;

r: for topography, landscape attributes;

p: representing parent material, lithology;

a: standing for age, the time factor; and

n: representing space, spatial position.

It can be noted that *s* and *n* factors were not in Jenny's original formulation. *S* has been added as a factor following findings that soil can be predicted from its properties, or soil properties from its class or other properties. *n* has been added because soil or soil property can be predicted from variables depending on how near or far they are, such as in using geostatistical methods.

Prediction of soil class or soil attribute does not necessarily require population of all the seven factors in the model. As such, most of the predictions have employed less than 4 of the scorpan factors (McBratney et al., 2003).

4.1.5 Sources of *scorpan* Data for DSM

s

The soil information *s,* can be obtained from an existing soil map, from expert knowledge or from remote or proximal sensing. Conventionally produced soil maps have provides a valuable source of soil information which is used in DSM (Baxter and Crawford, 2008; Yang et al., 2011). Expert knowledge system was used by Lilburne et al. (1998) to predict topsoil carbon. Zhu and Band (1994) and Zhu et al. (2001) wrote on a knowledge-based approach to data integration for soil mapping in combination with GIS and fuzzy logic. With respect to remote sensing, Boettinger et al. (2008) demonstrated that Landsat remotely sensed spectra data represent useful environmental covariates for DSM while Viscarra-Rossel and McBratney (2008) discussed application of soil diffuse reflectance spectroscopy (visible near infra-red and mid infra-red) for cheap soil analysis and their subsequent use in DSM.

c

Important climate variables which can be used as predictors in soil mapping are temperature, precipitation, and potential evapotranspiration (McBratney et al., 2003). The sources of these variables are weather stations records. However, weathers stations in some parts of the world, especially in the developing world, are sparsely located to be able to have sound estimation of the local climatic conditions for areas where is no weather station. Mitchell and Jones (2005) and Hijmans et al (2005) have developed global surface rasters for temperature and precipitation using spatial geostatistic tools at 5 degrees and 1 km resolutions respectively. These modeled global climate information may be an important source of climate information where they are not locally available.

o

Land use and land cover have been identified and used for prediction of soil classes and properties in several studies representing the biological (living organism) soil formation factor. Remote sensing has been used as a source of this information for soil mapping where normalized difference vegetation index (NDVI) has been the most important input. Turreta et al. (2008) combined land cover maps from Landsat TM images and secondary data from soil survey reports in soil properties-landscape modelling in Rio de Janeiro, Brazil. In another study, Bell et al. (2000) used vegetation as input in a soil–terrain model for estimating patterns of soil organic carbon. Other living organisms which have influence on soil development such as humans and soil organisms (earthworms, termites, etc) have not been employed to a larger extent in soil mapping compared to vegetation.

r

Terrain and its derivatives are currently mainly derived from digital elevation models. Other sources could be contours from topographic maps, point measurements from traditional survey works using GPS receivers, and remotely sensed elevation data. Several other attributes important for soil classes and attributes prediction can be derived from terrain data. These include altitude, slope gradient, slope shape, aspect, wetness index and topographic ruggedness index (Wilson and Gallant, 2000). Among the scorpan model factors, *r* has the highest explanatory power for short scale variability of soil (Schaetzl and Anderson, 2010). The power is attributed to it providing potential and kinetic energy on water movement and thus conditioning the redistribution of energy and matter. The catena concept, originally defined by Milne (1936), also stresses on the influence of relief on the sequence of soils between high altitude areas and the adjacent low lands.

p

Geological maps are currently the main source of soil parent material information. Radiometrics can also be used to give some information about composition of the parent materials (McBratney et al., 2003). Studies which have used parent material information to predict soil maps include Carre´ and Girard (2002).

a

This represents the length of time pedogenesis of a given soil has been occurring. The information can be estimated from the age of dateable materials in the soil such as carbon dating of fossils and artefacts. Age of soil can also be estimated using the age of the material in which soil has developed. These sources of information for age of soil pose a limitation for scorpan prediction efficiency when dealing with polycyclic soils (McBratney et al., 2003). As a result of uncertainties and high technology needed to estimate the age of soils, this soil forming factor has not been used widely in digital soil mapping.

n

With invent and continued use of GIS, spatial coordinates are important information required for prediction of both soil classes and soil attributes. Spatial factor used in scorpan model are the x,y (longitude, latitude or easting, northing) coordinates. In field, this information is obtained using GPS receiver sets. In data mining works, the coordinates can be read from georeferenced maps.

4.1.6 Machine Learning for DSM

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to automatically learn programs from multi-source data sets and make predictions (Witten et al., 2011). Machine learning is used in many computer based applications such as web search, spam filters, credit scoring, and drug design, to mention a few, and it is predicted to be the driver of the next big wave of innovation (Manyika et al., 2011). Recently, machine learning techniques have found their way in soil mapping as well. A review by Brungard et al. (2015) show that the technique has been used to model soil depth classes, biological soil crust classes, soil drainage classes, presence/absence of diagnostic soil horizons, and soil class-landscape relationships.

Numerous machine learning algorithms are available including clustering algorithms, discriminant algorithms,multinomial logistic regression algorithms, neural networks algorithms, tree based algorithms, and support vector machine classifiers algorithms (Brungard et al., 2015)). The selection of suitable algorithms is determined by multiple factors, including the nature of data source, appropriate knowledge representation, and desired accuracy. The decision tree algorithms are however, favorable because of their interpretability and ability to deal with nonlinearity and both nominal and continuous data types.

WEKA is one of the applications that can be used to run decision tree analyses. It is an open-source Java application produced by the University of Waikato in New Zealand. DSM works done using WEKA application include those by Ramesh and Ramar (2011), Gholap et al. (2012), Subburayalu and Slater (2013) and Subburayalu et al (2014).

4.1.7 Objective of the Study

The objective of this study was to apply the use of decision tree based J48 and Random Forest algorithms in predicting the numerically classified soil clusters using two training data sets derived from 30 m resolution ASTER and SRTM DEMs. The better predicted soil cluster map will be used as the basis for soil properties derivatives inputs in subsequent multi-criteria land evaluation for rice suitability analysis.

4.2 Methods

4.2.1 Study Area Description

The study was conducted in Kilombero Valley, Tanzania. The valley is about 300 km east of Indian Ocean covering about 11000 km^2. The study site is located in zone 37 south, occupying the area lying between 9064697 and 9089031 m northings and 175422 to 197033 m eastings. More description of the study area is given in the introductory chapter of this document.

4.2.2 Attributes Selection

The attributes for prediction of soil classes were chosen based on *scorpan* formulation (McBratney., 2003).

s - was derived from a 1959 legacy soil map of the area by Andeson (FAO, 1961) and soil survey conducted specifically for this study.

c - was not used

o - a 5 m resolution RapidEye satellite image was used to extract vegetation information

r – 30 m resolution SRTM and ASTER digital elevation models were used to derive topography attributes

p – was not used. The soils are generally formed from cyclic deposition of alluvial materials from other parts. The study area bedrock has little influence on the pedogenesis of these soils

a – not used

n- spatial location was recorded for each attribute used in the prediction of soil classes

4.2.3 Generation of Attributes

s

Two sources were used to provide soil related information (*s*):

- a legacy soil map of the study area, and
- field collected and clustered soil information

Legacy soil map

A reconnaissance map developed by Anderson (FAO, 1961) was used to extract soil information of the study area. The map was extracted from the FAO report and georeferenced in QGIS software (QGIS Development Team, 2014). The study site was extracted by clipping it with a previously prepared polygon covering the boundaries of the study area. The soil map was then screen-digitized using ArcMap 10.1 software (ESRI, 2010). Figure 4.1 shows the digitized Anderson's reconnaissance soil map and the soil groups.

Soil Clusters from Field Description and Sampling

A total of 13 soil clusters were generated from 33 field-described and sampled soil profiles using OSACA application (Jacobson and Carré, 2006). OSACA clusters soil pedons uses a *k*-means based algorithm in which the sum of distance of points from the cluster center is used as objective function (Carré and Jacobson, 2009). More description about the soil information used and the clustering process is given in Chapters 2 and 3 of this dissertation.

o

A 5 m resolution RapidEye satellite image (Blackbridge LLC) was used to derive vegetation related attributes of the study area. The RapidEye satellite sensor specifications are listed on Table 4.1. ERDAS IMAGINE 2014 software was used for processing the image and the derivatives using unsupervised classification. The indices were calculated and analyzed using Spatial-Modeler module in the software. The attributes derived are landuse/landcover classes, normalized difference vegetation index (NDVI), optimized soil adjusted vegetation index (OSAVI) and soil enhancement ratio (SER).

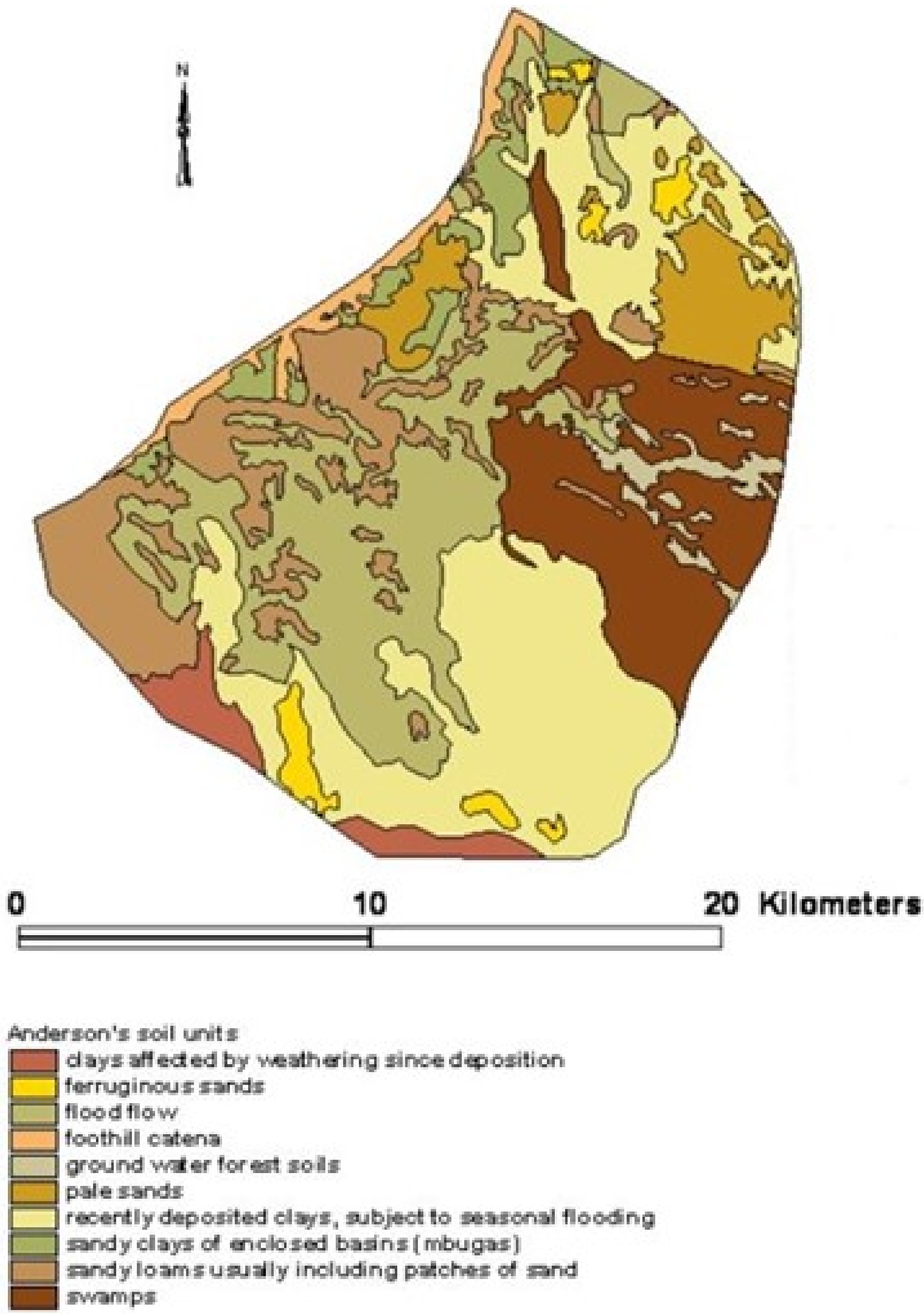


Figure 4.1: Anderson's reconnaissance soil units of the study area. (Digitized from FAO report (FAO, 1961))

Table 4.1. RapidEye satellite sensor specifications

Number of Satellites	5
Spacecraft Lifetime	7 years
Orbit Altitude	630 km in Sun-synchronous orbit
Equator Crossing Time	11:00 am local time (approximately)
Ground sampling distance (nadir)	6.5 m
Pixel size (orthorectified)	5 m
Swath Width	77 km
On board data storage	1500 km of image data per orbit
Revisit time	Daily (off-nadir) / 5.5 days (at nadir)
Image capture capacity	4 million sq km/day
Dynamic Range	12 bit
Sensor Type	Multi-spectral push broom imager
Spectral bands:	
Type	Wavelength (nm)
Blue	440 - 510
Green	520 - 590
Red	630 - 685
Red Edge	690 - 730
NIR	760 - 850

Land Use Land Cover Classification

Classification involves labeling the pixels as belonging to a particular spectral and categorizing the pixels in an image into land cover/ land use classes. An unsupervised classification was performed on the RapidEye image using the Iterative Self Organizing Data Analysis Technique (ISODATA) into 36 classes (Figure 4.2). Since the intention was to get different pixel values for prediction of soil clusters, efforts were not done to further give the names of the vegetation classes based on dominant vegetation types.

Normalized Difference Vegetation Index (NDVI)

The NDVI (Gitelson et al., 1999) is a numerical indicator that uses the visible and near-infrared bands of the electromagnetic spectrum. It is developed for estimating vegetation cover from the reflective bands of satellite data (Sahebjalal and Dashtekian, 2013). The NDVI layer is calculated as:

$$NDVI = \frac{(NIR - R)}{(NIR + R)}$$

Where NIR is the spectral reflectance in near infrared band and R is the red band.

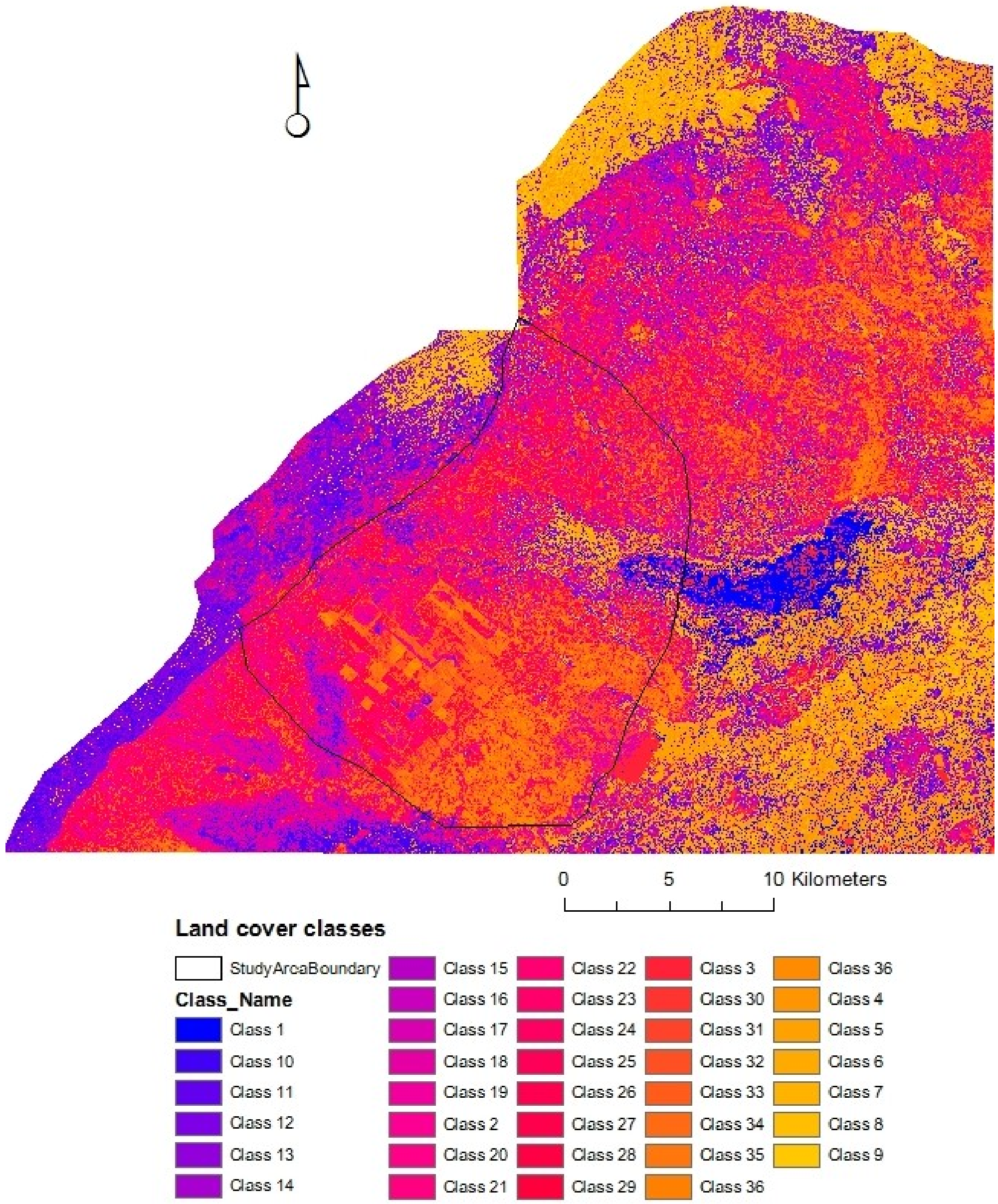


Figure 4.2. Land use/ land cover classes derived from RapidEye satellite image

Optimized Soil Adjusted Vegetation Index (OSAVI)

According to Rodeaux et al. (1996), the optimized soil adjusted vegetation index (OSAVI) was developed with an intention to minimize soil and atmospheric perturbation when calculating vegetation index. It is calculated as:

$$OSAVI = \frac{(NIR - R)}{(NIR + R + L)}$$

Where, L = 0.16 is an optimal value to minimize the variation in soil background.

Soil Enhancement Ratio (SER)

Soil enhancemement ratios are calculated as b3/b2, b3/b7, and b5/b7. Since RapidEye image has bands 1 – 5, it was only possible to calculate b3/b2. The calculated SER is shown on Figure 4.3.

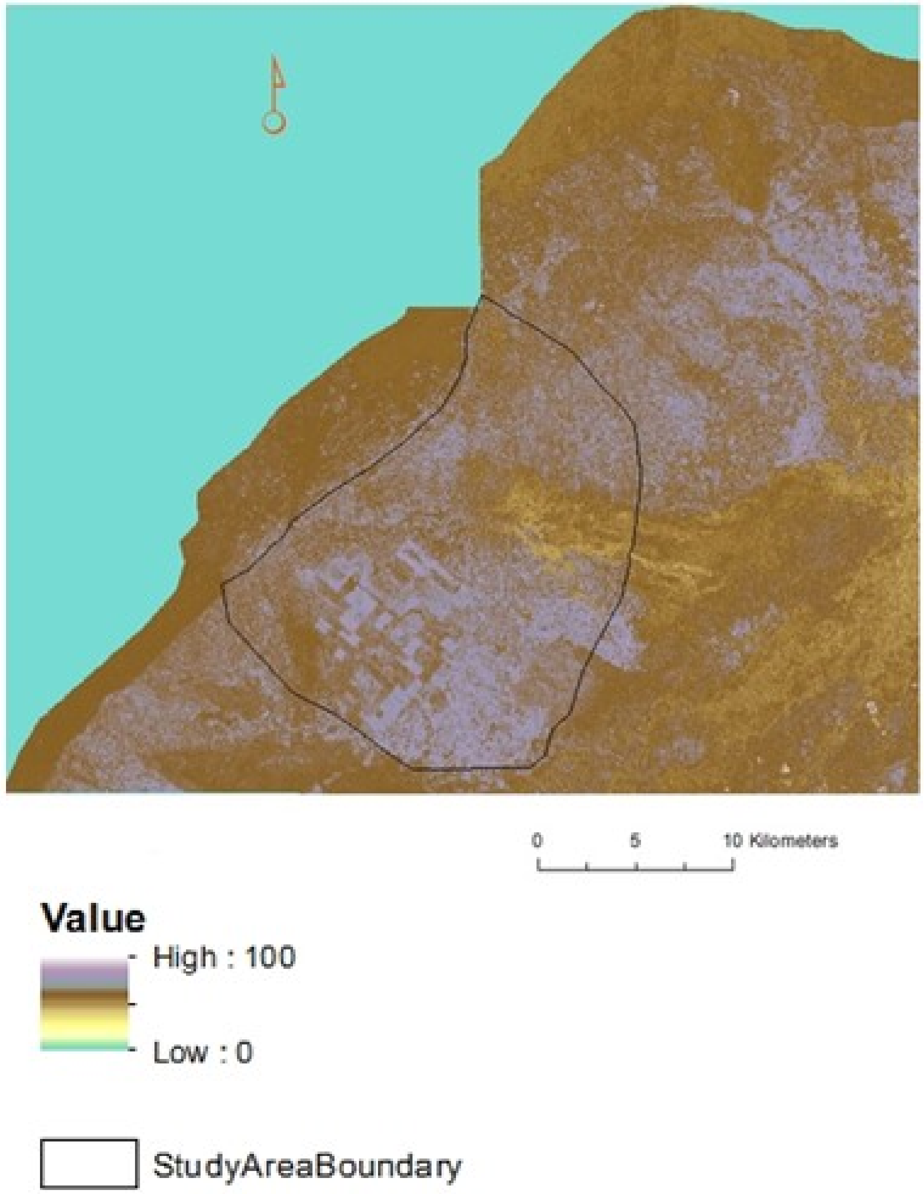


Figure 4.3. Soil enhancement ratio determined from band3/band4 ratio of RapidEye satellite image

r

The digital elevation models (DEMs) and its derivatives have been used to provide ancillary data in remote sensing image analysis and data for landscape modeling (Giles and Franklin, 1996). In this study, two DEMs of almost equal resolution were used with intent to compare the quality of their outputs for digital soil mapping of a relatively flat area in Kilombero Valley. Derivatives were calculated from ASTER DEM with a 30 m resolution and 1-Arc resolution SRTM DEM after filling depression using Planchon, and Darboux (2001) algorithm in Whitebox GAT 3.2. Some of DEM derivatives such as aspect were not used in this study because they were deemed of little contribution to soil formation due to the flatness of the study area. All the derivatives were calculated in Whitebox Geospatial Analysis Tool 3.2 software. The derivatives are listed below:

Slope Gradient

Slope gradient is defined by a plane tangent to a topographic surface at a point (Burrough, 1986). It measures the rate of change of elevation in the direction of the steepest descent. In this study, slope was estimated using Horn's (1981) 3rd-order finite difference method.

Plan Curvature

This is defined as the rate of change in aspect along a contour line, from a digital elevation model (DEM). Plan curvature characterizes convergence or divergence of water as it flows within the landscape (Gallant and Wilson, 2000). Plan curvature is negative for diverging flow along ridges and positive for convergent areas, e.g. along valley bottoms. It was calculated according to the method by Gallant and Wilson (2000).

Profile Curvature

This is the rate of change of slope along a downslope flow line. It is used to characterize changes in flow velocity and sediment transport processes (Gallant and Wilson, 2000). Convex flow profiles have negative values while concave slopes have positive values. The tool uses method by Gallant and Wilson, (2000) to calculate profile curvature.

Tangential Curvature

Tangential curvature is the curvature of an inclined plan perpendicular to both the direction of flow and the surface (Gallant and Wilson, 2000). It is computed as plan curvature multiplied by sine of the slope angle. The Whitebox tool reports curvature in degrees multiplied by 100 for easier interpretation.

Total Curvature

This is the curvature of the surface itself and not the curvature of intersecting lines such as profile and plan curvatures. The values of total curvature range from positive to negative, with zero values indicating that the surface is either flat, or that the convexity in one direction is perfectly balanced by the concavity in the opposite direction. The tool uses algorithm which is the same formula for the calculation of total curvature as presented by Gallant and Wilson (2000).

Relative Stream Power Index

The relative stream power index is directly related to the stream power if the discharge is directly proportional to upslope contributing area. The index is calculated as

$$RSP = A_s{}^p \; x \tan(B)$$

Where, A_s is the specific catchment area (the upslope contributing area per unit contour length), B is the local slope gradient in degrees; and p is a user-defined exponent term that controls the location-specific relation between contributing area and discharge.

Sediment Transport Index

In Whitebox GAT software, sediment transport index is calculated following description by Moore and Burch (1986). The index combines upslope contributing area (As), under the assumption that contributing area is directly related to discharge, and slope (B). The index is calculated as:

$$STI = (m + 1) \; x \; (A_s \,/22.13)^m \; x \sin(B/0.0896)^n$$

Where As is the specific catchment area (i.e. the upslope contributing area per unit contour length) B is the local slope gradient in degrees; the contributing area exponent, *m*, is usually set to 0.4 and the slope exponent, *n*, is usually set to 1.4.

Wetness Index

The wetness index describes the propensity for a site to be saturated to the surface given its contributing area and local slope characteristics. It is calculated as:

$$WI = ln(A_s/tan(S))$$

Where A_s is the specific catchment area (i.e. the upslope contributing area per unit contour length), and S is slope measured in degrees. Grid cells with a slope of zero are assigned an arbitrary high value (32767) in the output image to compensate for the fact that division by zero is infinity. These very flat sites coincide with the wettest parts of the landscape.

Deviation from Mean Elevation

This is the difference between the elevation of each grid cell and the mean elevation of the centering local neighbourhood normalized by standard deviation. This attribute measures the relative topographic position as a fraction of local relief, and so is normalized to the local surface roughness.

Difference from Mean Elevation

This is the difference between the elevation of each grid cell in an input DEM and the average elevation in the local neighbourhood. The local neighbourhood is defined as a circular area of a user-specified dimension surrounding each grid cell.

Topographic Ruggedness Index

This is a measure of local topographic relief. It calculates the root-mean-square-deviation (RMSD) for each grid cell in a digital elevation model (DEM), calculating the residuals (i.e. elevation differences) between a grid cell and its eight neighbours.

Flow Accumulation

The flow accumulation grid (i.e. contributing area) was generated using the D-infinity algorithm (Tarboton, 1997). This algorithm is an example of a multiple-flow-direction (MFD) method because the flow entering each grid cell is routed to one or two downslope neighbours, i.e. flow divergence is permitted.

4.2.4 Training Sets Sampling

Training sets were extracted from the x,y locations of the 33 soil observations which were used to predict the 13 soil clusters (refer chapter 3 for clustering process). The 'extract values at XY coordinate' tool in Whitebox GAT was used to extract the attribute values for the training dataset. Two sets of training dataset were prepared: one in which the topographic data were extracted from ASTER DEM and the other from SRTM DEM. Both sets included the same satellite image derivatives and soil datasets.

4.2.5 Running of the machine learning algorithm

Two algorithms were used in this study with an objective of picking up the results from the algorithm which give a better predicted soil cluster map. The algorithms J48 and Random Forest (RF) were ran in Machine Learning Framework known as WEKA (Witten and Frank, 2005).

J48 is an Open Source version of Quinlan's C 4.5 algorithm (Quinlan, 1993), that uses divide-and-conquer approach to learn decision trees. The algorithm recursively splits a training set into nodes based on a measure of impurity at the nodes. The entropy measure called gain ratio is used to measure impurity, where nodes with less impurity are desired.

The RF technique uses bootstrap samples to build multiple decision trees (Breiman, 1996). Breiman (2001) defined RF as a classifier consisting of a collection of tree structured classifiers (‘forests’) {h(x, Θk), k =1, ...} where {Θk} are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x. The number of predictors used to find the best split at each node is a randomly chosen subset of the total number of predictors (Prasad et al., 2006). The advantage of the RF algorithm is that it does not over-fit the data.

The training data set were input into WEKA as csv (comma delimited) file, where the two algorithms were used to perform decision tree analysis (DTA) to get prediction models for both SRTM and ASTER DEM datasets.

4.2.5 Mapping of the Predicted Soil Clusters

To be able to map the predicted soil clusters, a 30 m SRTM DEM raster was converted to points shape file using ‘Raster to Vector Points’ tool in Whitebox GAT software. The points generated correspond with grid cells center points. A total of 324,082 points were generated. The ‘Export Geography Markup Language (GML) tool was used to export the shape file to an GML file from which XY coordinates for each point were extracted and stored in an excel file. The XY coordinates for each point were then employed to extract attributes values for the SRTM and ASTER DEM derivatives using 'Extract Values at XY Coordinate' tool in Whitebox GAT, and 'Extract Multi Values to Point' spatial analyst tool in ArcMap 10.1 (ESRI, 2010) software. The Andersons soil units from a polygon map classes for the points were extracted by overlaying the point layer to the soil map layer and performing spatial join tool to join the attribute values of the two layers.

The J48 and Random Forest models were ran using each training set and the predictions of each model on the test set were obtained in a text file. The prediction output file for each model was then imported back into excel and was queried against the original test area file containing the X and Y coordinates to get the corresponding locational information for each instance. The new file with the predictions and locations for each model was used to create point files in ArcGIS which were later converted to a prediction map for each model, in the form of a raster grid.

4.2.6 Validation of Predicted Soil Cluster Maps

Two methods were used to validate the predicted soil cluster maps. The first one was by using a statistical paired sample t test and the second one was by visual analysis.

By Paired Sample t Test

A total of 24 validation points were randomly sampled and georeferenced. The soil samples from the validation points were analysed for topsoil pH, organic carbon, available phosphorus, and particle size distribution in the soil laboratory.

A point shape file was created from these verification points and then was overlaid on predicted soil cluster maps in ArcMap 10.1 (ESRI, 2010). Using identification tool in the software, the corresponding clusters for each verification point were recorded. Predicted topsoil values for each soil cluster were then extracted from the soil clustering output for each validation parameter. By so doing, actual value and predicted value for each point were assembled for the predicted maps. The paired sample t test was used to asses if there was significant difference in the properties values.

By Visualization

In this method, the soil cluster maps produced were compared with the legacy map to see how well the different soil units and soil clusters were predicted. Special attention was given to the map unit boundaries of the predicted and legacy maps. Pixels contiguity was also assessed.

4.3 Results and Discussion

4.3.1 Ranking of Training Data Attributes

Before training the data for decision tree analysis, information gain attribute evaluation was used to select and rank the input training data for both ASTER and SRTM DEMs sets. The results showed that, of the 18 training attributes, the legacy map soil units (AU) attribute was highly ranked in ASTER DEM training data set. The information gain for this attribute was 1.764, while the other 17 attributes, each had a value of zero. In SRTM training set, AU was also highly ranked, again with information gain value 1.764. Difference from mean elevation (DFME) attribute was ranked second in SRTM training set with information gain value of 0.785. The rest of the attributes (16) in the dataset had information gain values of zero.

The high ranking of the AU from both datasets suggests that they highly represent features which are result of the factors that influenced soil genesis in the study area. These units were delimited using aerial photograph interpretation. Low ranking of the other attributes suggests that these variables were poorly related to the factors that influenced soil genesis in these areas. Unsurprisingly, satellite image derivatives such as land cover classes; soil enhancement ratios and vegetation indices were not ranked high despite their fine resolution of 5 m. This could be because the major land use type in the study site is agriculture and the reflectance obtained could have just reflected differences in growth stages of vegetation or land clearances at the time of exposure, not soil variability. For DEM related derivatives, the explanation could be the scale from which the attributes were derived. For our relatively flat alluvial area, a 30 m resolution may not be a correct scale to capture the soil-landscape relationships. Soils in an area represent multiple soil forming factors operating over long periods of time. We therefore face a challenge of choosing optimal predictive environmental variables that are truly linked to the soil forming processes.

Both training sets were then run using Random Forest (RF) and J48 decision tree classifier algorithms to construct their respective models. The observations were not initially split to training and testing datasets for model accuracy testing because of fewer observations. Splitting observations into training/testing sets is a common practice (Tesfa et al., 2009; Subburayalu and Slater, 2013), but in case of few observations splitting may not be done (Brungard et al., 2015).Testing of the model accuracy was instead done using the training sets because splitting would result in too few observations for accurate model training. Evaluation of model built using ASTER training data set on RF showed that the algorithm was able to classify all 33 instances correctly. In comparison, J48 algorithm was able to classify 20 instances (about 61%) correctly. Using SRTM training data, the model developed by RF algorithm predicted correctly 32 instances (about 97%), while the model by J48 predicted correctly 19 instances (about 58%).

The comparison of the two learners suggests that RF performed better that J48. In a study by Subburayalu and Slater (2013), RF outperformed J48 in prediction of soil series, and the authors suggested that the learner has a prospect of becoming a useful learning algorithm for digital soil mapping.

4.3.2 Predicted Soil Clusters

Results showed that J48 learner predicted 8 clusters out of possible 13 clusters. This was for both training data sets; ASTER and SRTM. Soil clusters predicted by J48 learner in both training data sets are S56, S58, S59, S60, S61, S62 and S65. In addition S68 was predicted only in ASTER training data results, while S64 was predicted only in SRTM data. The RF learner predicted all 13 possible clusters. These are S56, S57, S58, S59, S60, S61, S62, S63, S64, S65, S66, S67 and S68. Predicted clusters are shown in Table 4.2.

The higher prediction rate of RF compared to J48 could be attributed to how differently the two learners operate. RF is an ensemble (forest) of bagged classification trees. The classification trees in RF are independent and the classification of samples does not depend upon previous trees in the ensemble (Kuhn and Johnson, 2013). Because of this, the learner performs better in terms of prediction of majority of soils and showing resistance to the extreme changes in the data such as noise (Subburayalu and Slater, 2013) compared to J48 which is a single tree classifier.

Table 4.2. Soil clusters predicted for each training data set and learner. S56 to S68 are predicted soil clusters. (More description is given after the Table)

J48_ASTER	J48_SRTM	RF_ASTER	RF_ASTER
S56	S56	S56	S56
S58	S58	S57	S57
S59	S59	S58	S58
S60	S60	S59	S59
S61	S61	S60	S60
S62	S62	S61	S61
S65	S64	S62	S62
S68	S65	S63	S63
		S64	S64
		S65	S65
		S66	S66
		S67	S67
		S68	S68

Note:
J48_ASTER = Soil clusters predicted by using J48 learner on ASTER DEM based dataset
J48_SRTM = Soil clusters predicted by using J48 learner on SRTM DEM based dataset
RF_ASTER = Soil clusters predicted by using Random Forest learner on ASTER DEM based dataset,
RF_SRTM = Soil clusters predicted by using Random Forest learner on SRTM DEM based dataset

Another interesting observation is that J48 learner tended to predict soil clusters which have higher constituent membership and leave those with few constituent members. The soil clusters which were not predicted by the learner are those which have either 1 or 2 members (S57, S63, S66, and S67 for both data sets, S68 for SRTM and S64 for ASTER). The clusters with 3 or above constituent members were predicted by the J48 algorithm in both SRTM and ASTER data sets. Since the number of members in a cluster also corresponded to the number of training sets in that soil cluster, it can therefore be concluded that J48 requires more training data than Random Forest to be able to make a prediction of a soil cluster.

4.3.3 Spatial Prediction of Soil Cluster Maps

The predicted soil cluster maps are displayed on Figures 4.4, 4.5, 4.6 and 4.7.

The dominance of Anderson's (Legacy map) soil units have been demonstrated in all four predictions: ASTER (RF and J48) and SRTM (RF and J48). In all predictions, the predicted soil clusters generally followed boundaries prescribed by the Anderson soil units. The Andersons units are shown in Figure 4.8. The dominance of the Anderson's units may suggest the thoroughness of the work done to demarcate the soil units.

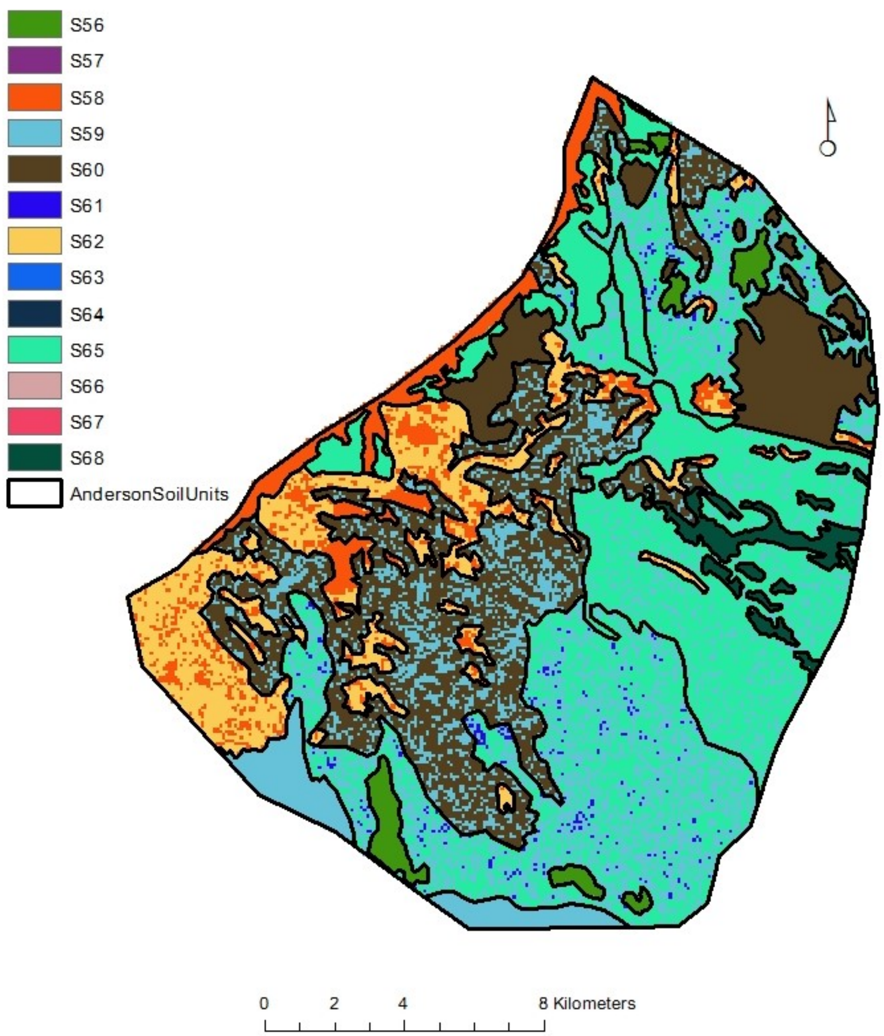


Figure 4.4. Map of soil clusters predicted using J48 learner on ASTER DEM based data set. S56 to S68 represent predicted soil clusters

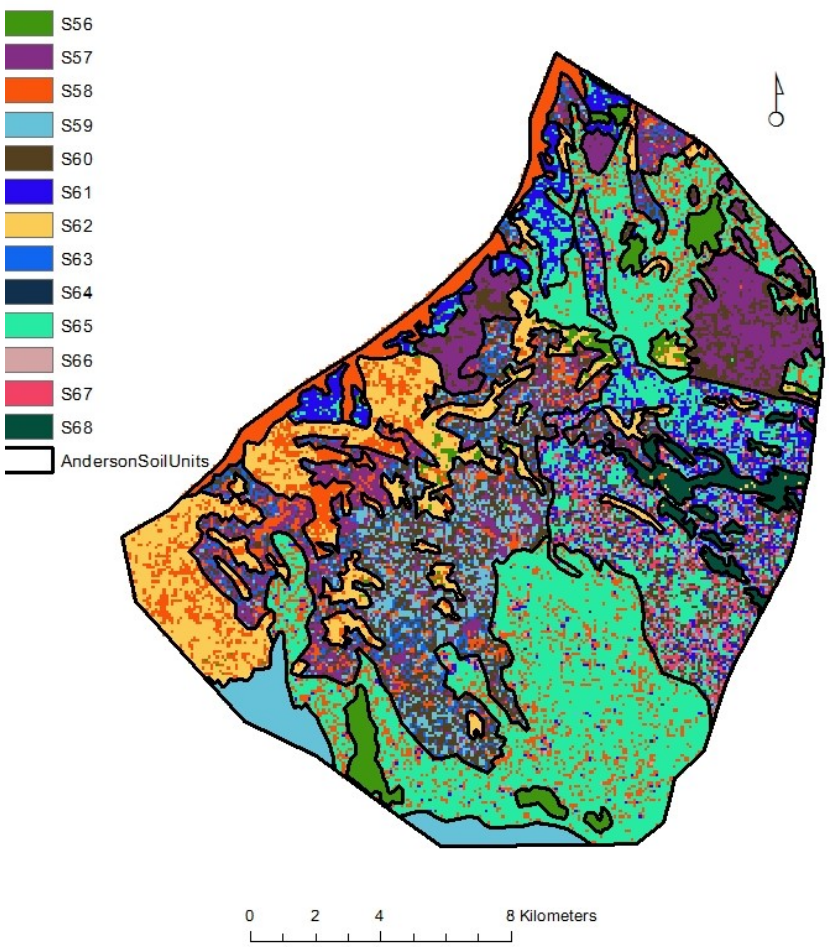


Figure 4.5. Map of soil clusters predicted using Random Forest learner on ASTER DEM based data set. S56 to S68 represent predicted soil clusters

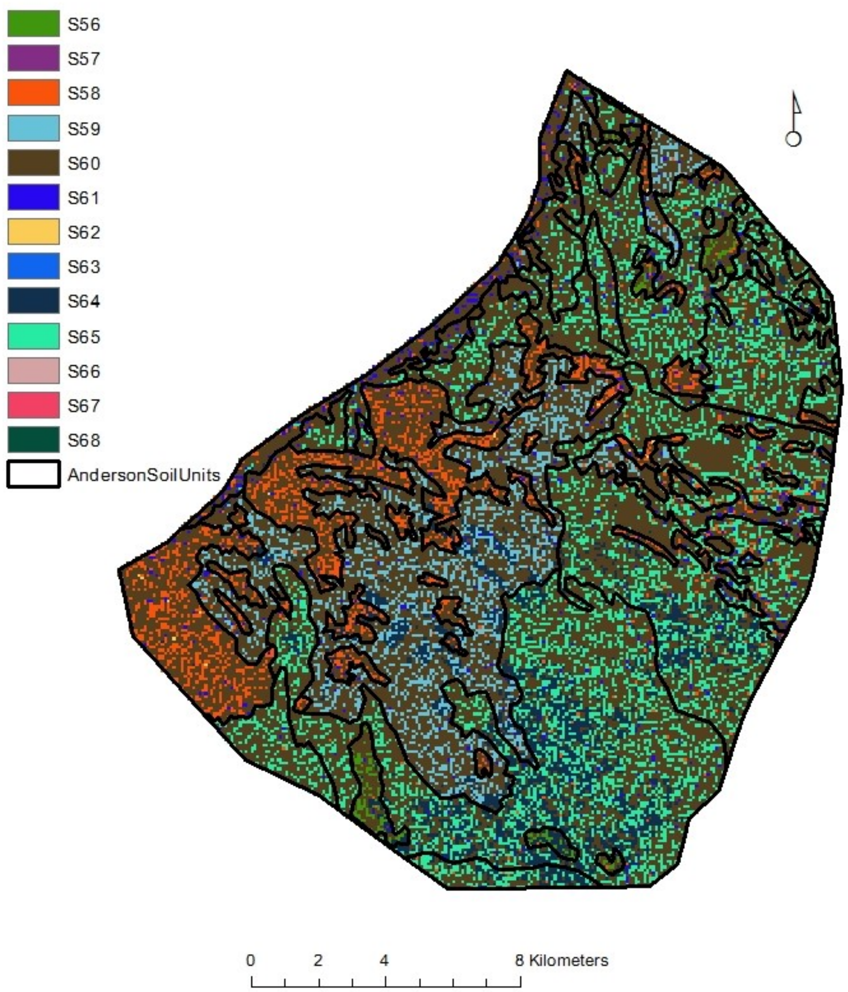


Figure 4.6. Map of soil clusters predicted using J48 learner on SRTM DEM based data set. S56 to S68 represent predicted soil clusters

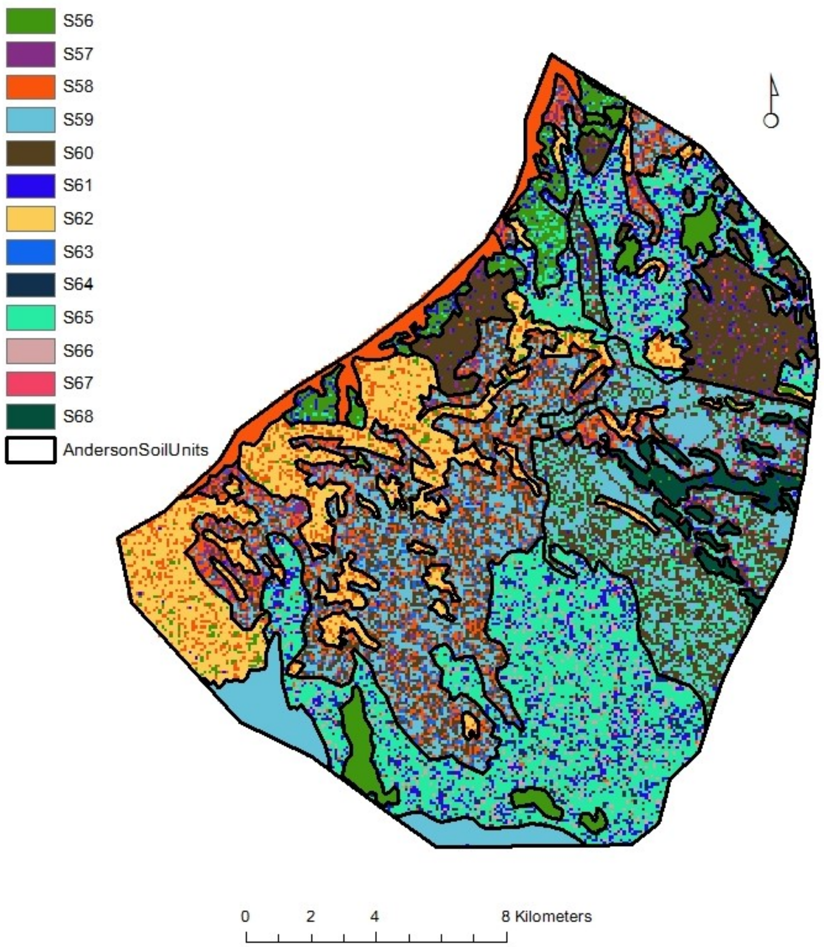


Figure 4.7. Map of soil clusters predicted using Random Forest learner on SRTM DEM based data set. S56 to S68 represent predicted soil clusters

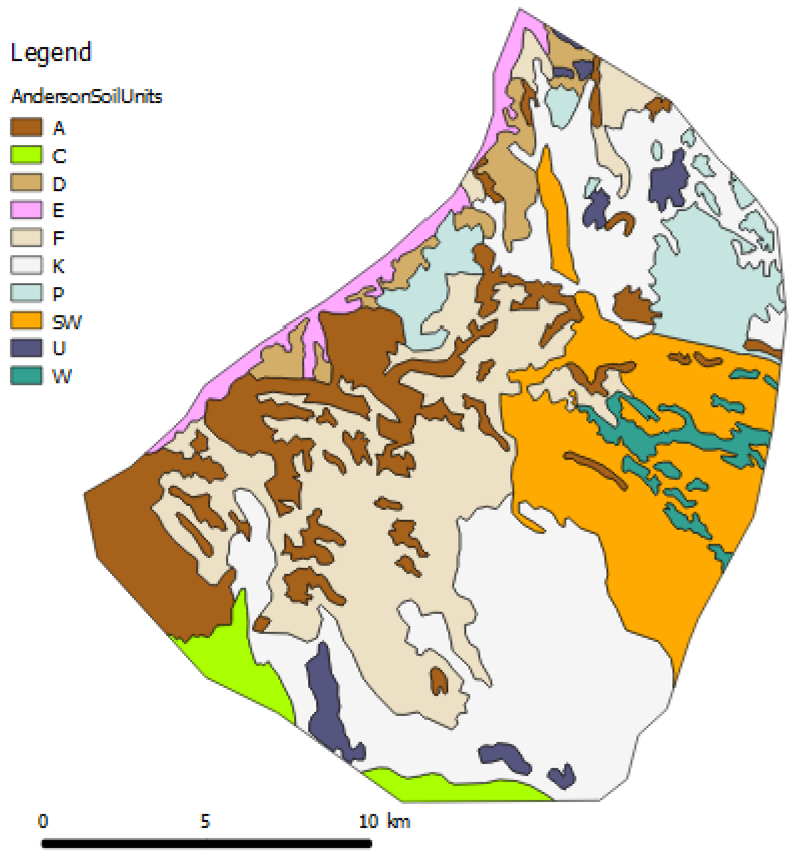


Figure 4.8. Legacy map soil units' boundaries. Letter symbols on the map key represent soil units

Dominance of well demarcated soil units in decision tree based predictions has been observed in other studies. For example, in a machine learning based prediction, detailed soil maps were the strongest predictor of actual biological soil crust in the Canyonland National Park (Brungard and Boettinger, 2012).

The dominance can also suggest the inability of the derived attributes to explain the distribution of the soil clusters. This is partly due to the resolution of the DEMs. While 30 m resolution DEM can be useful to assess influence of terrain on soil forming processes in sloping areas, its performance in relatively flat lands such as this study site is somehow limited. Digital elevation model with higher resolution might have given different results.

The great improvement on Anderson’s soil map units by this work is the mapping of different soil clusters within the units which were described as pure units. This is described in detail in the forthcoming section. The corresponding components predicted by each learner for each of the legacy map’s soil unit are shown in Table 4.3.

Table 4.3. Comparisons of soil clusters predictions on legacy map soil units for the studied learners and DEM data sets. Capital letters on first column represent letter symbols for legacy map soil units. S56 to S68 represent symbols for predicted soil clusters

Legacy map soil unit	Predicted soil clusters			
	SRTM training data set		ASTER training data set	
	J48 learner	Random Forest learner	J48 learner	Random Forest learner
A	Complex of S60, S58, few pixels of S61	S62 dominates; inclusions: S58, S56	Complex of S62, S58. More of S62	Complex of S62, S58. More of S62
C	Complex of S60, S59	S59	S59	S59
D	S60 dominant; S61, S65 minor components	S56 dominant, S61 minor inclusion	S65	S61 dominant, S65 inclusion
E	S60 dominant; S58, S61 minor component	S58	S58	S58
F	Complex of S59, S60, S63. More S60	Complex of S58, S59, S60, S63	Complex of S59, S60. More of S60	Complex of S57, S59, S63
K	Complex of S60, S63, S65	S65 dominant; S61, S63, S66 minor inclusions	S65 dominant; S59, S63 minor inclusions	S65 dominant; S58, S60, S61 minor inclusions
P	Dominantly complex of S60, S65. Minor inclusions: S58, S61	Dominantly S60. Few pixels: S58, S61, S66, S67	S60	S57 dominant; S 60 minor inclusions
SW	Complex of S60, S61, S63, S65. Comparatively more of S60	Complex of S59, S60. In lower proportions: S61, S63, S67	S65 dominant; few pixels of S61	Complex of S57, S59, S61, S63, S66. Much pixel coherence: S59
U	Complex of S56, S60, S61. Comparatively more of S60	S56	S56	S56
W	Dominantly S60; inclusions: S58, S65	S68 dominant; few pixels of S56	S68	Dominantly S68; few pixels of S58, S62

Comparison shows that when SRTM training data was used, J48 tended to predict complexes or higher number of soil clusters while Random Forest tended to predict dominant or fewer clusters for the same unit. For example, for soil unit C, J48 predicted it is occupied by complex of S60 and S59, almost each of them having the same share of the unit, while Random forest predicted that the unit was pure S59. Another example is unit E, where J48 predicted dominance of S60 with some minor inclusions of S58 and S61 while Random Forest predicts dominance of soil cluster S58. Predictions in soil units P, U and W show the same trend.

In contrast, the two learners generally predicted the same when ASTER training dataset was used. For example, both J48 and Random Forest predicted that soil unit A is occupied by S62 and S58 with S62 showing some dominance. Soil unit C was predicted to be covered by S59, E by S58, U by S56 and W dominantly by S68 by both learners. A few exceptions such as unit P in which J48 predicted dominance of S60 while Random Forest predicted S60 as a minor inclusion in an S57 dominated unit were also noted.

Another observation was that RF and J48 learners both generally predicted some common clusters presence in soil units for the SRTM data, with a few missing or added. For example, S58 is common for both J48 and RF for unit A, S59 for unit C, S61 for unit D, S58 for unit E, S59, S60 and S63 for unit F, and clusters S63 and S65 for unit K. The same observation was displayed by both learners in ASTER data, but in a higher consistence compared to SRTM data. Good examples for this are units A, C, E, K and U.

It was also observed that both J48 and RF generally predicted same clusters for both SRTM and ASTER data. However, RF learner was more consistent than J48 for predicting similar clusters for both data sets. Good examples are soil cluster predictions for soil units E, U and W where almost exactly the same predictions for SRTM and ASTER data were done using RF learner.

Differences in prediction of clusters could be attributed not only to the superiority of one learner to another, but also the number of training instances falling in the soil unit. As shown earlier, J48 requires more training instances to be able to predict a soil cluster. Generally, it has been observed that enough sampling is needed to separate clusters in feature space. Studies conducted suggest that soil classes with lower sampling frequencies are predicted less accurately. Such studies include Hengl et al. (2007), Stum et al. (2010), Barthold et al. (2013) and Brungard et al. (2015).

Predictions could also potentially be improved through the use of additional environmental covariates such as climatic parameters or spatial predictions of specific soil properties (e.g surface texture) from digital soil mapping (McBratney et al., 2003; Sanchez et al., 2009).

4.3.4 Validation of the Soil Clusters Predictions

To assess the correctness of cluster prediction, a paired sample t test was done. A total of 23 georeferenced verification points were randomly sample in the study area (Figure 4.9). The soil samples from the verification points were analysed for topsoil pH, organic carbon, available phosphorus, and particle size distribution in the soil laboratory.

A point shape file was created from these verification points and then was overlaid on predicted soil cluster maps in ArcMap 10.1 (ESRI, 2010). Using identification tool in the software, the corresponding cluster for each verification point was recorded. Predicted topsoil values for each soil cluster were then extracted from the soil clustering output for each validation parameter. By so doing, actual value and predicted value for each point were assembled for all predicted maps.

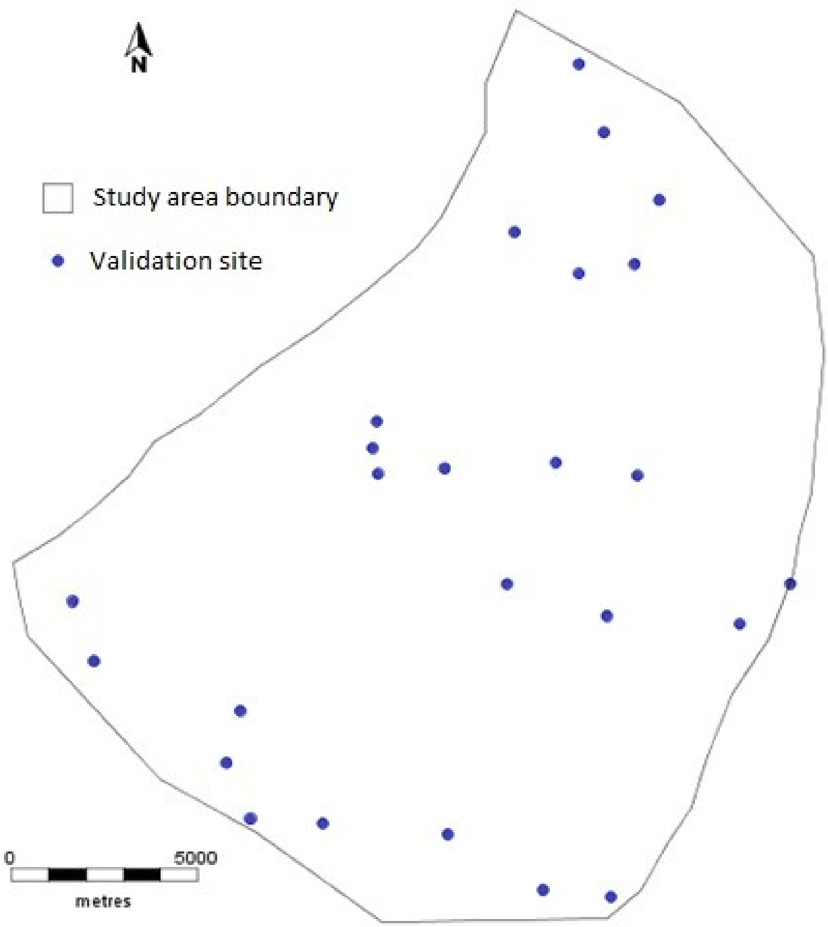


Figure 4.9. Locations of validation sampling points

The paired sample t test gives inferences about the difference between two population means for paired data. Two assumptions are necessary for carrying out this comparison:

1. The samples must be paired, that is the sample selected from the first population is related to the corresponding sample from the second population.
2. The differences of the pairs follow a normal distribution if there are less than 30 pairs.

The tested hypothesis was;

H_o: $\mu_{actual} - \mu_{predicted} = 0$

H_a: $\mu_{actual} - \mu_{predicted} \neq 0$

at the significance level $\alpha = 0.05$.

Since the sample size is not large enough (less than 30), it was necessary to check whether the differences between the paired samples follow a normal distribution. Table 4.4 summarises normality test results. Only parameters which passed the test were used for the paired t test analysis.

Table 4.4. Normality test results for validation parameters

Verification parameter	Predicting data source	Learner used	Normality test status
pH	SRTM	RF	Fail
		J48	Pass
	ASTER	RF	Pass
		J48	Pass
Organic carbon	SRTM	RF	Fail
		J48	Fail
	ASTER	RF	Fail
		J48	Fail
Available phosphorus	SRTM	RF	Fail
		J48	Fail
	ASTER	RF	Pass
		J48	Fail
Sand content	SRTM	RF	Pass
		J48	Pass
	ASTER	RF	Pass
		J48	Pass
Clay content	SRTM	RF	Pass
		J48	Pass
	ASTER	RF	Pass
		J48	Pass

Results show that there was no significant difference between selected predicted and validation points for both RF and J48 learners when using SRTM DEM data set (Table 4.5). The same results are shown when ASTER DEM data set was used, except when available phosphorus was used for validation on RF learner.

However, these validation results should be interpreted with caution. This is because visually, the four predicted soil cluster maps (Figures 4.4 through 4.7) show some obvious differences. For example, while RF predicted 13 soil clusters, J48 learner predicted only 8 soil clusters for both the test data sets. Also, while J48 learner resulted in more map contiguity for ASTER data set – likely due to its inability to predict clusters with fewer members – it resulted in less contiguous map when SRTM data set was used for RF prediction.

Likely, sampling more validation points would have been more informative. It should also be noted that while spatial soil cluster prediction used only the topsoil attributes, the numerical clustering process did not only take into consideration the topsoil properties. The metrics involved all profile horizons to assign a soil profile to a given soil cluster. Hence, validation using information from the entire soil profile instead of the topsoil would probably give more informative results.

The selection of the validation parameters was based on their stability in the landscape. Soil pH, soil organic carbon, available phosphorus, and soil texture are not expected to change over short period between the sampling periods. However, these are not the only parameters which were used to cluster the soils. Other parameters such as exchangeable bases, CEC, EC, zinc, iron, copper, manganese and total nitrogen were used. More informative validation results could be obtained when some of these were used.

Table 4.5. T-test of mean difference = 0 (vs not = 0) for actual vs predicted cluster topsoil values for selected validation soil attributes

Predicting data source	Learner	Verification attribute	P - value
SRTM	Random Forest (RF)	Sand	0.303
		Clay	0.592
	J48	Sand	0.278
		Clay	0.147
		pH	0.834
ASTER	Random Forest (RF)	Sand	0.113
		Clay	0.342
		pH	0.958
		Available phosphorus	0.000
	J48	Sand	0.307
		Clay	0.398
		pH	0.121

4.4 Conclusion

This chapter assessed performance of J48 and Random Forest (RF) learners in predicting soil clusters using two data sets, one based on derivatives from ASTER DEM and the other from the same derivatives but from SRTM DEM. Soil units from a legacy soil map and derivatives from RapidEye satellite image were included in each data set.

Results showed that J48 learner predicted 8 clusters out of possible 13 clusters for both data sets while RF was able to predict all 13 possible soil clusters. It was noted that J48 learner tended to predict soil clusters which have higher training data representation and unable to predict those with few training data. RF was able to predict all clusters, regardless of the number of the training instances in the soil cluster.

All predicted soil cluster maps demonstrated dominance of the legacy map soil units as a predictor. The dominance of the units may suggest the thoroughness of the work done to demarcate the legacy soil units, but can also suggest the limitations of the derived terrain and satellite image attributes to explain the distribution of the soil clusters. This could be due to the resolution of the DEMs and the somewhat uniform land use.

Comparison showed that J48 tended to predict complexes or higher number of soil clusters for the same legacy map soil unit while RF tended to predict dominant or fewer clusters for the same unit. It was also observed that both J48 and RF generally predicted same clusters for both SRTM and ASTER data. However, RF learner was more consistent than J48 for predicting similar clusters for each data set.

Differences in prediction of clusters could be attributed to the superiority of one learner to another in dealing with quantity of training data. RF learner showed higher prediction rates and resulted to better soil cluster prediction when it was used with SRTM training data set.

The study demonstrated that the predictions were able to add more details on the legacy soil map units; hence these methods are suggested in updating legacy soil map.

4.5 References

Acres, B.D. 1983. The soils of Tabora Region. Tabora Rural Integrated Development Project, Tanzania. Project Report No.67, Land Resources Development Centre, UK. ODA.

Baker, R.M. 1970. The Soils of Tanzania. FAO. Dar es Salaam, Tanzania.

Barthold, F.K., M. Wiesmeier, L. Breuer, H.G. Frede, J. Wu, and F.B. Blank. 2013. Land use and climate control the spatial distribution of soil types in the grasslands of Inner Mongolia. Jourrnal of Arid Environments 88:194–205.

Baxter, S.J., and D.M. Crawford. 2008. Incorporating legacy soil pH databases into digital soil maps. In: A.A. Hartemink, A. McBratney, and M.L. Mendonca-Santos, editors, Digital soil mapping with limited data. Springer. p. 311-318.

Behrens, T., H. Förster, T. Scholten, U. Steinrücken, E. Spies and M. Goldschmitt. 2005. Digital soil mapping using artificial neural networks. Journal of Plant Nutrition and Soil Science 168:21-33.

Bell, J.C., D.F. Grigal, and P.C. Bates. 2000. A soil–terrain model for estimating spatial patterns of soil organic carbon. In: J.P. Wilson, and J.C Gallant, editors, Terrain analysis: principles and applications. John Wiley & Sons, Inc. New York, p. 295– 310.

Boettinger, J.L., D.W. Howell, A.C. Moore, A.E. Hartemink, and S. Kienast-Brown. 2010. Digital soil mapping: Bridging research, environmental application, and operation. Springer-Verlag, Dordrecht, the Netherlands.

Boettinger, J.L., R.D Ramsey, J.M. Bodily, N.J. Cole, S. Kienast-Brown, S.J.Nield, A.M. Saunders, and A.K. Stum. 2008. Landsat spectra data for digital soil mapping. In: A.A. Hartemink, A. McBratney, and M.L. Mendonca-Santos, editors, Digital soil mapping with limited data. Springer. p. 193-202.

Breiman, L. 1996. Bagging predictors. Machine Learning 24:123-140.

Breiman, L. 2001. Random Forests. Machine Learning 45:5-32.

Brungard, C.B., and J.L. Boettinger. 2012. Spatial prediction of biological soil crust classes; value added DSM from soil survey. In: B. Minasny, B.P. Malone, and A. McBratney, editors, Digital soil assessments and beyond: Proceedings of the 5th global workshop on digital soil mapping. CRC Press, Sydney. p. 57–60.

Brungard, C.W., J.L. Boettinger, M.C. Duniway, S.A. Wills, and T.C. Edwards Jr. 2015 Machine learning for predicting soil classes in three semi-arid landscapes Geoderma 239–240:68–83.

Burrough, P.A. 1986. Principles of geographical information systems for land resources assesment. Oxford University Press.

Calton, W.E. 1954. An experimental pedological map of Tanganyika. 2nd Inter-African Soils Conference 237-240.

Carré, F. and M. Jacobson. 2009. Numerical classification of soil profile data using distance metrics. Geoderma. 148:336–345

Carre´, F., Girard, M.C. 2002. Quantitative mapping of soil types based on regression kriging of taxonomic distances with landform and land cover attributes. Geoderma 110:241–263.

Cook, S.E., A. Jarvis and J.P. Gonzalez. 2008. A new global demand for digital soil information. In: A.A. Hartemink, A. McBratney, and M.L. Mendonca-Santos, editors, Digital soil mapping with limited data. Springer. p. 31-41.

D'Hoore, J.L. 1964. Soil Map of Africa. Commission for Technical Cooperation in Africa, Lagos, Publication no. 23.

De Pauw, E. 1984. Soil Physiography and Agro-ecological Zones of Tanzania. Consultants Final Report for Crop Monitoring and Early Warning Systems Project. KILIMO/FAO.

Deckers, J., O. Spaargaren, and S. Dondeyne. 2002. Soil survey as a basis for land evaluation. Encyclopedia of Life Support Systems.

Eschweiler, J.A. 1998. SOTER database – Tanzania. FAO, Rome.

ESRI. 2010. ArcGIS - A Complete Integrated System. Environmental Systems Research Institute, Inc., Redlands, California. http://esri.com/arcgis.

FAO. 1961. The Rufiji Basin Tanganyika. FAO Exp. Techn. Ass. Progr. No. 1269 Vol. 7. Rome.

Gallant, J.C., and J.P. Wilson. 2000. Primary topographic attributes. In: J. P. Wilson and J. C. Gallant, editors, Terrain analysis: Principles and applications. John Wiley & Sons, New York. p. 51-86

Gholap, J., A. Ingole, J. Gohil, S. Gargade and V. Attar. 2012. Soil data analysis using classification techniques and soil attribute prediction. International Journal of Computer Science Issues 9(3):415-418.

Giles, P.T., and S.E. Franklin. 1996. Comparison of derivative topographic surfaces of a DEM generated from stereoscopic SPOT images with field measurements. Photogrammetric Engineering & Remote Sensing 62(10):1165-1171.

Gitelson, A.A., Y.J. Kaufman, and M.N. Merzlyak. 1999. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sensing of Environment 58:289–298.

Hansen, M.K., D.J. Brown, P.E. Dennison, S.A Graves, and R.S. Bricklemyer. 2009. Inductively mapping expert-derived soil-landscape units within dampo wetland catena using multispectral and topographic data. Geoderma 150 (1-2):72-84.

Haring, T., E. Dietz, S. Osenstetter, T. Koschitzki, and B Schroder. 2012. Spatial disaggregation of complex soil map units: a decision tree based approach in Bavarian forest soils. Geoderma 185-186:37-47.

Hathout, S.A. 1983. Soil Atlas of Tanzania. Tanzania Publishing House, Dar es Salaam.

Hengl, T., Toomanian, N., Reuter, H.I., Malakouti, M.J., 2007. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. Geoderma 140:417–427.

Heuvelink, G.B.M., and R. Webster. 2001. Modelling soil variation: past, present, and future. Geoderma 100, 269–301.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965–1978.

Horn, B.K.P. 1981. Hill shading and the reflectance map. Proceedings of the IEEE 69(1):14-47.

Jacobson, M., and F. Carré. 2006. OSACA version 1.0. Land Management and Natural Hazards Unit, Institute for Environment and Sustainability. European Commission, Italy.

Jenny, H., 1941. Factors of soil formation: a system of quantitative pedology. McGraw-Hill, New York.

Kauzeni, A.S., I.S. Kikula, S.A. Mohamed, J.G. Lyimo, and D.B. Dalal-Clayton. 1993. Land use planning and resource assessment in Tanzania: a case study. IIED Environmental Planning Issues No. 3 IRA Research Paper No. 35.

Kempen, B., D.J. Brus, G.B.M. Heuvelink, and J.J. Stoorvogel. 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. Geoderma 151:311–326.

Kerry, R., P. Goovaerts, B.B. Rawlings, and B.P. Marchant. 2012. Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. Geoderma 170:347-358.

Kiss, J.J., de Jong, E., Martz, L.W., 1988. The distribution of fallout Cesium-137 in southern Saskatchewan, Canada. Journal of Environmental Quality 17:445–452.

Lagacherie, P., A.B. McBratney, and M. Voltz. 2007. Digital soil mapping: An introductory perspective. Elsevier, Amsterdam.

Lilburne, L., A. Hewitt, P. McIntosh, and I. Lynn. 1998. GIS-driven models of soil properties in the high country of the south island. 10th Colloquium of the Spatial Information Research Centre, University of Otago, New Zealand, p. 173–180.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Byers. 2011. Big data: The next frontier for innovation, competition, and productivity. Technical report, McKinsey Global Institute. San Francisco, CA.

McBratney, A.B., M.L. Mendonca Santos, and B. Minasny 2003. On digital soil mapping. Geoderma 117:3–52.

McBratney, A.B., I.O.A. Odeh, T.F.A. Bishop, M.S. Dunbar, and T.M. Shatar. 2000. An overview of pedometric techniques for use in soil survey. Geoderma 97:293–327.

Milne, G. 1936. A Reconnaissance journey through parts of Tanganyika Territory. Journal of Ecology 35:192-265.

Mitchell, T.D. and P.D. Jones. 2005. An improved method of construction a database of monthly climate observation and associated high-resolution grid. International Journal of Climatology 25:693-712.

Moore, I. and G. Burch. 1986. Physical basis of the length-slope factor in the Universal Soil Loss Equation. Soil Science Society of America Journal 50:1294–1298.

Odeh, I.O.A., A.B. McBratney, and D.J. Chittleborough. 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67:15–225.

Odgers, N.P., W. Sun, A.B. McBratney, B. Minasny, and D. Clifford. 2014. Disaggregating soil map units through resampled classification trees. Geoderma 214:91-100.

Pachepsky, Y.A., D.J. Timlin, and W.J. Rawls. 2001. Soil water retention as related to topographic variables. Soil Science Society of America Journal 65:1787–1795.

Park, S.J., and L.G. Vlek. 2002. Prediction of three-dimensional soil spatial variability: a comparison of three environmental correlation techniques. Geoderma 109:117–140.

Planchon, O. and F. Darboux. 2001. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. Catena 46:159–176.

Prasad, A.M., L.R. Iverson and A. Liaw. 2006. Newer classification and regression tree techniques: Bagging and Random Forest for ecological prediction. Ecosystems 9:181-199.

QGIS Development Team. 2014. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Quinlan, J.R. 1993. C4.5: Programs for machine learning. Morgan Kaufmann, San Mateo.

Ramesh, V., and K. Ramar. 2011. Classification of agricultural land soils: a data mining approach. Agricultural Journal 6(3):82-86.

Rondeaux, G., M. Steven and F. Baret. 1996. Optimization of Soil Adjusted Vegetation Indices. Remote Sensing of Environment 55:95-107.

Sahebjalal, E. and K. Dashtekian. 2013. Analysis of land use-land covers changes using normalized difference vegetation index (NDVI) differencing and classification methods. African Journal of Research Agricultural 8(37):4614-4622.

Samki, J.K. 1977. Provisional Soils Map of Tanzania. Survey and Mapping Division, Dar es Salaam.

Samki, J.K. 1982. Soils Map of Tanzania. National Soil Service. Tanga, Tanzania.

Samki, J.K., and J.F. Harrop. 1984. Fertilizer recommendations in Tanzania on a district by district basis. NSS ARI Mlingano, Tanzania.

Sanchez, P.A., S. Ahamed, F. Carré, A.E. Hartemink, J. Hempel, J. Huising, P. Lagacherie, A.B. McBratney, N.J. McKenzie, M.L. Mendonca-Santos, B. Minasny, L. Montanarella, P. Okoth, C.A. Palm, J.D. Sachs, K.D. Shepherd, T.G. Vagen, B. Vanlauwe, M.G. Walsh, L.A. Winowiecki, and G.L. Zhang. 2009. Digital soil map of the world. Science 325:680–681.

Schaetzl R.J., and S. Anderson. 2005. Soils: genesis and geomorphology. Cambridge University Press, Cambridge, UK.

Scull, P., J. Franklin, O.A. Chadwick and D. McArthur. 2003. Predictive soil mapping: A review. Progress in Physical Geography 27:171-197.

Stum, A.K., Boettinger, J.L., White, M.A., Ramsey, R.D., 2010. Random forests applied as a soil spatial predictive model in arid utah, in: Boettinger, J.L., D.W. Howell, A.C. Moore, A.E. Hartemink, and S. Kienast-Brown (Eds.), Digital Soil Mapping: Bridging Research, Environmental Application, and Operation, Springer, Dordrecht, 179–190.

Subburayalu, S.K. and B.K. Slater. 2013. Soil series mapping by knowledge discovery from an Ohio county soil map. Soil Science Society of America Journal 77:1254–1268.

Subburayalu, S.K., I. Jenhani, and B.K. Slater. 2014. Disaggregation of component soil series on an Ohio County soil survey map using possibilistic decision trees. Geoderma 213:334–345.

Tarboton, D.G. 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resources Research 33(2):309–319.

Tesfa, T.K., D.G. Tarboton, D.G. Chandler, and J.P. McNamara. 2009. Modeling soil depth from topographic and land cover attributes. Water Resource Research Volume 45, W10438. doi:10.1029/2008WR007474.

Turetta, A.P.D. M.L.Mendonca-Santos, L.H.C. Anjos, and R.L.L. Berbara. 2008. Spatial-temporal changes in land cover, soil properties and carbon stocks in Rio de Janeiro. In: A.A. Hartemink, A. McBratney, and M.L. Mendonca-Santos, editors, Digital soil mapping with limited data. Springer. p. 261-271.

Viscarra-Rossel, R.A., McBratney, A.B., 2008. Diffuse reflectance spectroscopy as a tool for digital soil mapping. In: A.A. Hartemink, A. McBratney, and M.L. Mendonca-Santos, editors, Digital soil mapping with limited data. Springer. p. 165-172.

Wickama, J.M.W. 1997. Pedological investigation and characterization in Kitanda Village, Mbinga District, Tanzania. MSc Thesis, Sokoine University of Agriculture, Morogoro, Tanzania.

Wilson, J.P., and J.C. Gallant 2000. Terrain analysis: Principles and applications. John Wiley & Sons Inc., New York.

Witten, I.H., and E. Frank. 2005. Data mining: Practical machine learning tools and techniques. Second edition. Elsevier Inc., San Francisco, CA.

Witten, I.H., E. Frank, and M.A. Hall. 2011. Data mining: Practical machine learning tools and echniques. Morgan Kaufmann, Burlington.

Yang, L., Y. Jiao, S. Fahmy, A.X. Zhu, S. Hann, J.E. Burt, and F. Qi. 2011. Updating conventional soil maps through digital soil mapping. Soil Science Society of America Journal 75:1044–1053.

Zheng, D., Hunt, E.R., Running, S.W., 1996. Comparison of available soil water capacity estimated from topography and soil series information. Landscape Ecology 11:3–14.

Zhu, A. and Band, L. 1994: A knowledge-based approach to data integration for soil mapping. Canadian Journal of Remote Sensing 20:408–18.

Zhu, A.X., B. Hudson, J. Burt, K. Lubich, and D. Simonson. 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Science Society of America Journal 65:1463–1472.

Chapter 5: Scoring Land Evaluation Criteria using Analytical Hierarchy Process (AHP)

5.1 Introduction

5.1.1 Land Evaluation and Global Food Security

The process of evaluation and grouping of specific areas of land in terms of their suitability for a defined use is called land use suitability analysis (Trung et al., 2006). The analysis aims at identifying the most appropriate spatial pattern for future land uses according to specific requirements, preferences, or predictors of some activity (Malczewski, 2004).

Land resources are increasingly becoming scarce due to increased population, land degradation and climate change (Mueller et al, 2010; Elaalem et al., 2011). The scarcity compounds the challenge humankind faces of ensuring global and regional food security and to increase food production to support the growing population. Sustainable use of soils is a critical factor in improving food production, especially in Sub Saharan Africa because it is where poverty and food shortages are currently more experienced and where the population growth is very fast. A recent report by UNICEF estimates that by the end of this century 40% of the world's people will be living in Africa, compared to the current’s 15% (You et al., 2014).

Sustainable increase in food production will help to tackle poverty and land degradation problems in this part of the world. Following this observation, food security is one of the top priorities in developing countries (Elaalem et al., 2011). To make these priorities realistic, the evaluation of arable land and agricultural potential in these countries is needed in order to support current and future agricultural uses. The evaluation of the land will help in identifying suitable areas for agricultural land use and other land uses, which are preludes to identification of appropriate managements required to optimize and sustain the land use types. Involving current and potential land users, especially in agricultural land use planning process have been suggested as a methodology toward sustainable land uses.

5.1.2 Land Evaluation: a Multi-Criteria Decision Making (MCDM) Process

Several factors are used in the analysis to decide if a piece of land is suitable for a particular land use, or comparably better than another type of land use. It includes consideration of not only inherent capacity of a land unit to support a specific land use sustainably, but also the socio-economic and environmental costs. For example, deciding on a suitability of a land for a certain crop growth the planners need to consider, among other criteria; soil properties (fertility status, depth), accessibility of the area, labour availability, diseases prevalence (crops and human diseases), climatic conditions (precipitation, temperature, evapotranspiration), and topographic factors (slope gradient, elevation, aspect) (Messing et al, 2003; Perveen et al., 2008; Kuria, et al., 2011; Samanta et al., 2011; Elsheikh et al., 2013). For example, studies on deciding about location of a banana distribution warehouse in a farm will require analysis of accessibility to the area, distance, cost, security of the region, and local acceptance of the company (Garcia et al., 2014); while analysis of a place to locate gravel and sand pit will require criteria like ground water protection, distance to the metropolitan, thickness of the exploitable material, and soil agricultural productivity to be considered (Marinoni and Hoppe, 2006). Therefore, in land use analysis a decision is a result of a comparison of one or more alternatives with respect to one or more criteria that we consider relevant for the decision. Dealing with many criteria in making decision requires multicriteria decision making (MCDM) approaches (Xu and Yang, 2001).

The MCDM approaches can be used to identify a single most preferred option, to rank options, to short-list a limited number of options for subsequent detailed appraisal, and to distinguish acceptable from unacceptable possibilities (Department for Communities and Local Government, 2009). Two types of criteria are used along with MCDM: those which need to be minimized and those which require to be maximized. An area with the minimum agricultural productivity is the best choice when locating area suitable for gravel and sand pit location, while the area with maximum possible thickness of the sand (or gravel) layer are the best choice for this land use (Marinoni and Hoppe, 2006). The MCDM methods can be classified as compensatory or non-compensatory (Hwang and Yoon, 1981). In non-compensatory methods unfavourable value in one attribute cannot be offset by a favourable value in other attributes. The non-compensatory techniques are not very useful for general decision making (Xu and Yang, 2001). Compensatory methods permit tradeoffs between attributes. Scoring methods falls into compensatory methods. In scoring methods, a score is used to express the decision maker's preference. It transforms attribute values into a common preference scale so that comparisons between different attributes become possible. The Analytical Hierarchy Process (AHP) method, which calculates the scores for each alternative based on pairwise comparisons (Saaty, 1988) is among the most popular methods in this category (Xu and Yang, 2001; Marinoni and Hoppe, 2006; Saaty, 2008; Elaalem et al., 2011).

5.1.3 Analytic Hierarchy Process (AHP) for MCDM

Among the criteria employed in decision making through MCDM methods, some are considered as more and some as less important. Weights are assigned to the criteria reflecting their relative importance. Higher weights are assigned to criteria which have been determined by the decision makers as more important than the others in the list of identified criteria needed for a specific decision making. Estimation of weights (scores)needs developed methods which will effectively guide the process to avoid biasness and give quality and desired results to aid in decision making.

AHP is one of the preferred scoring methods. This method is superior to many other weighting methods because it can deal with inconsistent judgments by providing a measure of inconsistency. AHP can also be integrated into other software to provide greater flexibility and accuracy. For example, several multi-criterial land use suitability analyses studies have been successfully accomplished by combining AHP and GIS (Marinoni and Hoppe, 2006; Cherif Ahmed et al., 2007; Perveen et al., 2008; Kihoro et al., 2013).

The AHP method tackles the weighting based on the hierarchical relationship among the objectives and their attributes (Malczewski, 1999). During the process, the objectives are defined at the higher level followed by identification of the attributes contributing to the objective at the lower level. Decision-makers (normally domain experts) then make pairwise comparisons between the attributes and develop relative weights using Saaty scale which range from 1 to 9 (Table 5.1). Consistency ratio (CR) is calculated to assess consistency of the experts' preferences. The pairwise comparison is repeated if CR is above 10% (Saaty and Vargas, 1991).

Table 5.1. Fundamental Saaty's scale for comparative judgments (Saaty and Vargas, 1991)

Numerical value	Preference level
1	Equally preferred
2	Equally to moderately preferred
3	Moderately preferred
4	Moderately to strongly preferred
5	Strongly preferred
6	Strongly to very strongly preferred
7	Very strong preferred
8	Very strong to extremely preferred
9	Extremely preferred
Reciprocals	Values for inverse comparison

5.1.4 Objective of the Study

In this study we use AHP methods to determined weights for criteria identified for rice suitability analysis in Kilombero Valley, Tanzania. The identified criteria are soil physical properties (soil texture), soil chemical fertility (soil pH, soil macro-, and soil micro-nutrients), accessibility of the land, distance to rice market, terrain of the land (slope gradient), and surface water resources.

The domain experts used in this study are local farmers and extension staff. Each group performed its own AHP and later they worked together as a joint group to produce a final preference matrix. Results from the three groups are compared to get insights on how the farmers and extension staffs differ in perception about important factors for rice production in their area. The output of the joint group will be used in GIS to develop a suitability model for rice production in Kilombero Valley in the next Chapter of this document.

5.2 Methods

5.2.1 Study Area

The study was conducted in Mngeta, Mchombe, Njage, and Mbingu area of Kilombero Valley. The study site is located in UTM zone 37 south, occupying the area lying between 9064697 and 9089031 m northings and 175422 to 197033 m eastings. More details about the study area are given in Chapter 1.

5.2.2 Identification of the Land Evaluation Criteria

Five extension officers and four lead farmers from the villages falling in the study area participated in this exercise which was organized at Mchombe village. Six criteria were identified as important for rice production in the study area in a focused group discussion session comprising both farmers and extension staffs. These were:

i. Soil physical properties. For this study; soil physical properties were represented by soil texture. Soil texture was selected because of its influence on other properties such as water infiltration rates, bulk density and water holding capacity. Soil depth was also determined, but it was found not to be a limiting factor throughout the study area because the soils were deep.

ii. Soil chemical fertility. In this work, the attributes which were included in soil chemical fertility criteria included soil pH, soil organic carbon, soil micronutrients and macronutrients.

iii. Accessibility. For this criterion, reference was made on how easily people can reach their fields any time during farm management activities such as land preparation, sowing, weeding, and harvesting.

iv. Distance to market. This criterion referred to distance from the farms to village centres or subtowns. The shorter the distance to the market the higher the suitability of the area for rice production.

v. Surface water resources. This criterion covered distribution of rivers and channels. Distances to rivers and streams were related to amount and duration of floods which are requied for lowland rice production. Amount and distribution of rainfall was not considered because there was no marked variability within the study area to influence suitability differences for rice growing. Rivers and streams were important, especially because they receive and transport water to the study area from rains falling in their sources and along the way.

vi. Terrain. The idea behind this criterion was on how fast the surficial water can move or be retained in the field. Slope gradient was used to guide the ranking based on terrain (topography).

5.2.3 Criteria Scoring

To familiarize with the exercise, after the general description of pairwise preference matrix, we started by every participant doing the pairwise preference matrix individually. Questions were asked and explanations were given to make sure everybody knew what to do. The second step was to do the pairwise preference matrix in two groups. Farmers formed their own group and extension staffs formed theirs. Lastly, a joint group was formed and was also asked to prepare a preferential matrix of the identified criteria. The verbal terms of the fundamental Saaty's scale (Table 5.1) were used to assess the preference between two compared criteria at each instance in the matrix. The scale was then used to translate the verbal judgment by means of the fundamental scale 1 to 9.

The preference matrices were input in BPMSG AHP priority calculator (Goepel, 2014). The CR and Principal Eigen values were calculated and the weights were assigned using this calculator. Percentages were calculated and were used to assess the differences in perceptions on importance of one criterion over another between farmers and extension staffs based on the calculated weights.

It was necessary to check if the pairwise comparison has been consistent in order to accept the results of the weighting. Consistency Ratio (CR) is a measure of how much variation is allowed and must be less than 10% (Saaty and Vargas, 1991). Revisions of the preference matrices were done for all three groups when CR was above 10%.
CR is defined as the ratio of the consistency index CI to an average consistency index RI, thus

$$CR = \frac{CI}{RI} \qquad \text{Eq. 5.1}$$

Values for RI with *n* order of matrix are developed by Saaty and Vargas (1991). CI can be directly calculated from the preference matrix using Eq. 2 below

$$CI = \frac{\lambda_{max} - n}{n-1} \qquad \text{Eq. 5.2}$$

Where λ_{max} is greatest Eigen value of preference matrix, and *n* is order of matrix.

5.3 Results and Discussion

5.3.1 AHP Criteria Scores by Extension Staff

The decision matrix suggested by the extension staffs group on their first sitting is shown on Table 5.2. The CR for this matrix was 10.1%, slightly larger than suggested consistence value. The decision matrix was revised despite the CR value being marginal. The revised matrix gave CR value of 7.1% (Table 5.3), thus the weights were taken since it was lea 10%.

The criteria weights calculated from the revised matrix and their respective rankings are shown on Table 5.4. Results show surface water resources criterion was ranked the highest by the extension staffs compared to other identified criteria by scoring 41.4%. It was followed by soil chemical fertility (31.6%) and soil physical properties (13.9%). Distance to market and accessibility of the farms were given the lowest two priorities by scoring 2.5% and 3.9% respectively.

Table 5.2. Rice suitability analysis criteria preference matrix of extension staff before revision

Criteria	Soil physical properties	Soil chemical fertility	Accessibility	Distance to market	Surface water	Terrain
Soil physical properties	1	0.33	5	7	0.14	5
Soil chemical fertility	3	1	7	9	0.5	9
Accessibility	0.2	0.14	1	2	0.17	0.33
Distance to market	0.14	0.11	0.5	1	0.11	0.25
Surface water resources	7	2	6	9	1	7
Terrain	0.2	0.11	3	4	0.14	1

Table 5.3. Rice suitability analysis criteria preference matrix of extension staffs after revision

Criteria	Soil physical properties	Soil chemical fertility	Accessibility	Distance to market	Surface water	Terrain
Soil physical properties	1	0.33	5	7	0.2	3
Soil chemical fertility	3	1	7	9	0.5	9
Accessibility	0.2	0.14	1	2	0.17	0.33
Distance to market	0.14	0.11	0.5	1	0.11	0.25
Surface water resources	5	2	6	9	1	7
Terrain	0.33	0.11	3	4	0.14	1

Table 5.4. Criteria weights and ranks for rice suitability analysis derived from revised extension officer's preference matrix

Criteria	Weight	Rank
Soil physical properties	0.139	3
Soil chemical fertility	0.316	2
Accessibility	0.039	5
Distance to market	0.025	6
Surface water resources	0.414	1
Terrain	0.066	4

5.3.2 AHP Criteria Scores by Lead Farmers

The farmers' group first preference matrix (Table 5.5) had a CR of 11.5%. The revision of the matrix was done for some of the criteria in order to improve the consistence. The revised preference matrix had a CR of 7.3 and is shown on Table 5.6. The criteria weights derived from the revised matrix are shown on Table 5.7. The order of ranking of the criteria from the highest priority to the lowest was surface water resources, soil chemical fertility, soil physical properties, terrain, accessibility, and distance to market, in that order. This ranking of criteria by farmers group was similar to that of extension staffs group (Table 5.4). However, there are differences in weights given to some of the criteria by each group, indicating differences in perceptions about the importance of each criterion on rice productivity in Kilombero Valley.

The weights of the highest and the lowest two ranked criteria were not very different between the two groups. Extension staff gave surface water resources a priority of 41.4%, while farmers gave it 41.9%. Accessibility was given priority of 3.9% by extension staffs while farmers gave it 3.3%, meanwhile distance to market was given priorities of 2.5% and 2.6% by extension staffs and farmers respectively. While to extension staffs soil chemical fertility deserved 31.6% priority, to farmers the criteria was not as important and they gave it a 24.5% priority. Farmers thought soil physical properties (soil depth and soil texture) deserved priority closer to what they gave to soil chemical fertility. They therefore gave soil physical properties a priority of 20% while extension staffs gave it a priority of 13.9%, much lower than the soil chemical fertility. Farmers also gave more weight to terrain (7.6% priority) than extension staffs who gave it a priority of 6.6%.

Table 5.5. Farmers' group rice suitability analysis criteria preference matrix before revision

Criteria	Soil physical properties	Soil chemical fertility	Accessibility	Distance to market	Surface water	Terrain
Soil physical properties	1	1	9	9	0.14	3
Soil chemical fertility	1	1	7	7	1	4
Accessibility	0.11	0.14	1	3	0.11	0.33
Distance to market	0.11	0.14	0.33	1	0.14	0.2
Surface water resources	7	1	9	7	1	7
Terrain	0.33	0.25	3	5	0.14	1

Table 5.6. Farmers' group revised rice suitability analysis criteria preference matrix

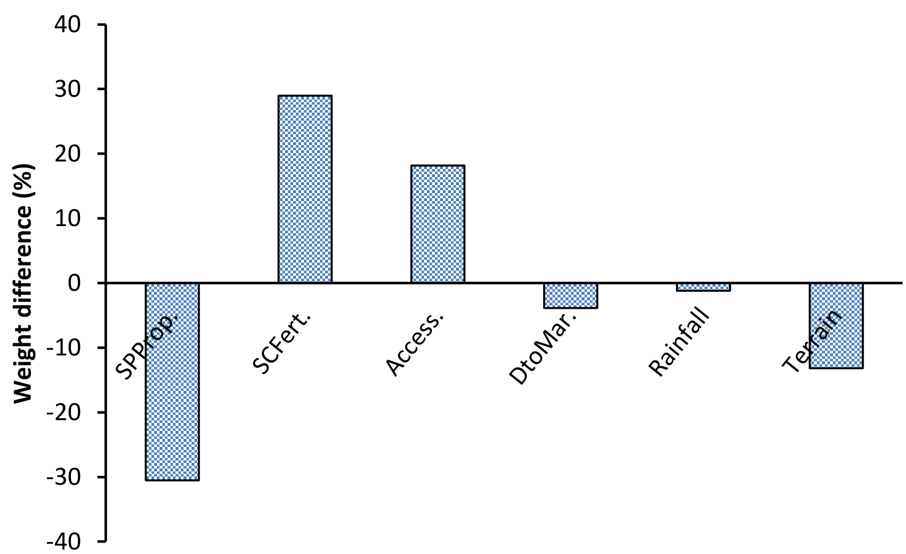
Criteria	Soil physical properties	Soil chemical fertility	Accessibility	Distance to market	Surface water	Terrain
Soil physical properties	1	1	9	9	0.2	3
Soil chemical fertility	1	1	7	7	1	4
Accessibility	0.11	0.14	1	2	0.11	0.33
Distance to market	0.11	0.14	0.5	1	0.12	0.2
Surface water resources	5	1	9	8	1	7
Terrain	0.33	0.25	3	5	0.14	1

Table 5.7. Criteria weights and ranks for rice suitability analysis derived from farmers' revised decision matrix

Criteria	Weight	Rank
Soil physical properties	0.2	3
Soil chemical fertility	0.245	2
Accessibility	0.033	5
Distance to market	0.026	6
Surface water resources	0.419	1
Terrain	0.076	4

The magnitudes and directions of differences in perceptions per each criterion for the two groups as indicated by differences in the criteria weights are shown in Figure 5.1. The extension staffs group perception of the importance of soil physical properties on rice production in Kilombero Valley was lower by 30.5% compared to that of the farmers group. Their opinion about the importance of soil chemical fertility was higher by 29% compare to the farmers' while that for accessibility was also higher by 18.2%. The extensions staffs perception of importance was lower than the farmers' perception on terrain by 13.2% and distance from farm to market place by 3.8%. Farmers and extension staffs appear to agree on the importance of surface water resources as extension staffs scoring of the criteria was only 1.3% below that of the farmers.

It was difficult to attribute the differences in perceptions and consequently prioritization between farmers and extension staffs. Despite comparatively low level of formal education of the farmers to that of extension staffs, they appeared to quickly and comfortably grasp the whole AHP exercise and relate it to their farming activities. This was confirmed by the low inconsistency results from their first preference matrix (CR = 11.5%). Both extension staffs and farmers participate in rice production by owning farms. However, the farmers participate directly in the farming practices by providing labour, leading the family labour force, and occasionally working with hired casual labour. The extension staffs have lesser time to do the day to day management of their farms compared to farmers because of the employment commitments. They, in most times, hire casual labour and are therefore likely to have less 'on-hand' experience on rice farming compared to the farmers. However, the extension staffs have attained higher formal education and have specialized courses in agriculture. This can explain why they ranked soil fertility higher by 29% compared to farmers. It can therefore be suggested that, the differences in perceptions and prioritization could be stemming from the differences on their levels of formal education and the actual time every group is spending on real rice farming.



SPProp. = Soil physical properties
SCFert. = Soil chemical fertility
Access. = Accessibility
DtoMar. = Distance to market
Rainfall = Surface water resources
Terrain = Terrain

Figure 5.1. Differences in rice suitability analysis criteria scoring between extension staff and lead farmers

5.3.3 AHP Criteria Scores by Joint Lead Farmers and Extension Staff Group

To get overall weightings of identified criteria for land use suitability analysis in Kilombero Valley, it was then necessary to get them from a combined group consisting of both farmers and extension staffs. Evaluation of the first decision matrix from this group (Table 5.8) showed high inconsistency by resulting to a CR of 23.5%. This new and high inconsistency suggested the difficulty the group faced coming with new compromises after they had done that in their own individual groups. Due to this high inconsistence, it was necessary to revise the decision matrix. The final decision matrix for this joint group is shown on Table 5.9. The CR of the final decision matrix is 7.9%, which allowed us to use the resultant criteria weights. The calculated criteria weights and ranks based on the revised decision matrix are shown on Table 5.10.

The results of criteria ranking by this joint group were similar to those ranked by extension staffs and farmers groups separately (Table 5.11). This shows general agreement on importance of one criterion over the other. However, there were differences on weights given for each criterion, differing from both the farmers and the extension staffs groups.

Table 5.8. Farmers and extension staff joint group's rice suitability analysis criteria preference matrix before revision

Criteria	Soil physical properties	Soil chemical fertility	Accessibility	Distance to market	Surface water	Terrain
Soil physical properties	1	0.2	9	7	0.17	6
Soil chemical fertility	5	1	7	9	0.17	6
Accessibility	0.11	0.14	1	5	0.14	0.2
Distance to market	0.14	0.11	0.2	1	0.11	0.33
Surface water resources	6	6	7	9	1	7
Terrain	0.17	0.17	5	3	0.14	1

Table 5.9. Farmers and extension staff joint group's revised rice suitability analysis criteria preference matrix

Criteria	Soil physical properties	Soil chemical fertility	Accessibility	Distance to market	Surface water	Terrain
Soil physical properties	1	0.5	9	7	0.25	6
Soil chemical fertility	2	1	7	9	0.25	6
Accessibility	0.11	0.14	1	2	0.14	0.5
Distance to market	0.14	0.11	0.5	1	0.11	0.33
Surface water resources	4	4	7	9	1	7
Terrain	0.17	0.17	2	3	0.14	1

Table 5.10. Farmers and extension staff joint group rice suitability analysis criteria weights and ranks

Criteria	Weight	Rank
Soil physical properties	0.19	3
Soil chemical fertility	0.234	2
Accessibility	0.036	5
Distance to market	0.025	6
Surface water resources	0.462	1
Terrain	0.052	4

Table 5.11. Summary of weights and ranks of criteria from extension staff group, farmers group, and joint group for rice suitability analysis

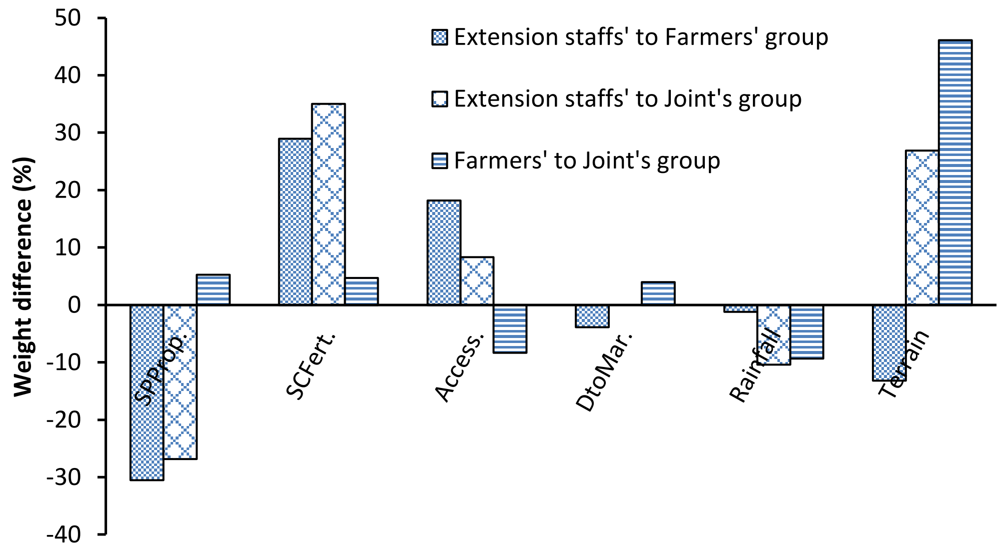
Criteria	Extension staff group		Farmers group		Group of both	
	Weight	Rank	Weight	Rank	Weight	Rank
Soil physical properties	0.139	3	0.2	3	0.19	3
Soil chemical fertility	0.316	2	0.245	2	0.234	2
Accessibility	0.039	5	0.033	5	0.036	5
Distance to market	0.025	6	0.026	6	0.025	6
Surafce water resources	0.414	1	0.419	1	0.462	1
Terrain	0.066	4	0.076	4	0.052	4

Percentage differences in criteria weights between the joint group and the other two groups are shown on Table 5.12 and depicted on Figure 5.2. The farmers' group prioritization of soil physical properties criteria was higher by 5% while that of extension staff was lower by 26.8% compared to the joint group prioritization of the same criterion. On the soil chemical fertility criterion, farmers' scoring was higher by 4.7% while that of the extension staffs was higher by 35% over the joint group's scoring. It can be observed that the farmers' group scores for both physical and chemical soil properties were very close to the joint group's scores while those of the extension staffs group showed bigger deviations. On the importance of accessibility, farmers' criteria were lower by 8.3% while those of extension staffs were higher by 8.3%. The extension staffs' perception of the importance of distance to market criteria was the same as that for the joint group, while that of farmers group was up by 4%. There was no much difference between the farmers and extension staffs differences against the joint group on the groups' priorities given to the surface water resources criteria. The farmers' weight was lower by 9.3% while that of extension staff was lower by 10.4%. The joint groups' results suggest that the farmers group overemphasized the importance of terrain by 46.2% while the extension staffs did the same by 26.9%.

It can be concluded from the results that famers' weights were generally very closer to the joint group's weights except for the terrain criterion. The differences between farmers' weights and joint group's weights are less than 10% for five out of six criteria, while only two criteria have their differences below 10% for the extension staffs weightings. This suggests that the farmers' decision matrix was superior to the extension officers' decision matrix. It also suggests that farmers are able to argue their case and their decisions are not influenced by superiority of formal education and status of the extension officers. The results also suggest that extension officers of Kilombero Valley are open to learning and accept challenges from their farmers.

Table 5.12. Percentage differences in weights of criteria between extension staffs, farmers and joint groups for rice suitability analysis

	Soil physical properties	Soil chemical fertility	Accessibility	Distance to market	Surface water	Terrain
Extension staffs' to Farmers' group	-30.5	29.0	18.2	-3.8	-1.2	-13.2
Extension staffs' to Joint group	-26.8	35.0	8.3	0.0	-10.4	26.9
Farmers' to Joint group	5.3	4.7	-8.3	4.0	-9.3	46.2



SPProp. = Soil physical properties
SCFert. = Soil chemical fertility
Access. = Accessibility
DtoMar. = Distance to market
Rainfall = Surface water resources
Terrain = Terrain

Figure 5.2. Differences in criteria weights for rice suitability analysis between famers group, extension staff group, and joint group

5.4 Conclusion

Analytic Hierarchy Process (AHP) was used to assign weights of the identified criteria to be later applied in multi-criteria land evaluation for rice crop suitability analysis. Extension staff and lead farmers from the study area were used to score the criteria. The two groups scored the criteria as separate groups and later they joined to provide the joint group scores.

Results show surface water resources was ranked the highest by the extension staffs compared to other identified criteria by scoring 41.4%. It was followed by soil chemical fertility (31.6%) and soil physical properties (13.9%). Distance to market and accessibility of the farms were given the lowest two priorities by scoring 2.5% and 3.9% respectively.

The ranking of the criteria by farmers group was similar to that of extension staffs group. However, the actual weights given to the criteria were different. The extension staffs group perception of the importance of soil depth and soil physical properties on rice production was lower by 30.5% compared to that of the farmers group. Their opinion about the importance of soil chemical fertility was higher by 29% compare to the farmers' while that for accessibility was also higher by 18.2%. The extensions staffs perception of importance was lower than the farmers' perception on terrain by 13.2% and distance from farm to market place by 3.8%. Farmers and extension staffs appear to agree on the importance of surface water resources as extension staffs scoring of the criteria was only 1.3% below that of the farmers.

The results of criteria ranking by the joint group were similar to those ranked by extension staffs and farmers groups separately. This shows general agreement on importance of one criterion over the other. However, again, there were differences on weights given for each criterion, differing from both the farmers and the extension staffs groups. The famers' weights were generally very closer to the joint group's weights except for the terrain criterion. The differences between farmers' weights and joint group's weights are less than 10% for five out of six criteria, while only two criteria have their differences below 10% for the extension staffs weightings. This suggests that the farmers' decision matrix was superior to the extension officers' decision matrix.

5.5 References

Cherif Ahmed, A. O., R. Nagasawa, K. Hattori, D. Chongo, and M.F. Perveen. 2007. Analytical Hierarchic Process in conjunction with GIS for identification of suitable sites for water harvesting in the oasis areas: Case study of the Oasis zone of Adrar, Northern Mauritania. Journal of Applied Sciences 7(19):2911-2917.

You, D., L. Hug, and D. Anthony. 2014. Generation 2030 – Africa: Child demographics in Africa. UNICEF, New York.

Department for Communities and Local Government. 2009. Multi-criteria analysis: a manual. Communities and Local Government Publications, West Yorkshire, UK.

Elaalem, M., A. Comber, and P. Fisher. 2011. A Comparison of fuzzy AHP and Ideal Point Methods for evaluating land suitability. Transactions in GIS 15(3):329–346.

Elsheikh, R., A. Rashid, B.M. Shariff, F. Amiri, N.B. Ahmad, S.K. Balasundram, and M.A.M. Soom. 2013. Agriculture Land Suitability Evaluator (ALSE): A decision and planning support tool for tropical and subtropical crops. Computers and Electronics in Agriculture 93:98–110.

García, J.L., A. Alvarado, J. Blanco, E. Jiménez, A.A. Maldonado, and G. Cortés. 2014. Multi-attribute evaluation and selection of sites for agricultural product warehouses based on an Analytic Hierarchy Process. Computers and Electronics in Agriculture 100:60–69.

Goepel, K.D. 2014. BPMSG AHP priority calculator [Version 2014-02-09 (c)]. Rational decision making made easy. http://bpmsg.com/academic/ahp.php. Visited: December, 2014.

Hwang C. L., and K. Yoon. 1981. Multiple attribute decision making methods and applications, Springer-Verlag, Berlin.

Jones A., V. Stolbovoy, E. Rusco, A.R. Gentile, C. Gardi, B. Marechal, and L. Montanarella. 2009. Climate change in Europe. 2. Impact on soil. a review. Agronomy for Sustainable Development 29:423–432.

Kihoro, J., N.J. Bosco, and H. Murage. 2013. Suitability analysis for rice growing sites using a multicriteria evaluation and GIS approach in Great Mwea region, Kenya. SpringerPlus 2013, 2:265. http://www.springerplus.com/content/2/1/265

Kuria, D., D. Ngari, and E. Waithaka. 2011. Using geographic information systems (GIS) to determine land suitability for rice crop growing in the Tana delta. Journal of Geography and Regional Planning 4(9):525-532.

Malczewski, J. 1999. GIS and Multicriteria Decision Analysis. John Wiley and Sons, New York.

Malczewski, J. 2004. GIS-based land-use suitability analysis: a critical Overview. Progress in Planning 62:3–65.

Marinoni, O., and A. Hoppe. 2006. Using the analytical hierarchy process to support sustainable use of geo-resources in metropolitan areas. Journal of Systems Science and Systems Engineering 15(2):154-164.

Messing, I., M.H. Fagerstrom, L. Chen, and B. Fu. 2003. Criteria for land suitability evaluation in a small catchment on the Loess Plateau in China. Catena 54:215–234.

Mueller, L., U. Schindler, W. Mirschel, T.G. Shepherd, B.C. Ball, K. Helming, J. Rogasik, F. Eulenstein, and H. Wiggering. 2010. Assessing the productivity function of soils. A review. Agronomy for Sustainable Development 30:601–614.

Perveen, F., R. Nagasawa, A.O.C. Ahmed, I. Uddin, and R. Kimura. 2008. Integrating biophysical and socio-economic data using GIS for land evaluation of wheat cultivation: A case study in north-west Bangladesh. Journal of Food, Agriculture and Environment 6(2):432-437.

Saaty T. L. 1988. The Analytic Hierarchy Process. University of Pittsburgh, Pennsylvania.

Saaty, T.L. 2008. Decision making with the analytic hierarchy process. International Journal of Services Sciences 1: 83–97.

Saaty, T.L., and L.G. Vargas. 1991. Prediction, projection and forecasting. Kluwer Academic Publishers, Dordrecht.

Samanta, S., B. Pal, and D.K. Pal. 2011. Land suitability analysis for rice cultivation based on multi-criteria decision approach through GIS. International Journal of Science and Emerging Technologies 2(1):12-20.

Trung, N.H., L.Q. Tri, M.E.F. van Mensvoort, and A.K. Bregt. 2006. Application of GIS in land-use planning: a case study in the Coastal Mekong Delta of Vietnam.In: Proceedings of the International Symposium on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences, Ho Chi Minh, Vietnam, 1-11 November. p. 6.

Vogel, R. 2008. A software framework for GIS-based multiple criteria evaluation of land suitability. In Proceedings of the Eleventh AGILE International Conference on Geographic Information Science, Girona, Spain: p. 1–32.

Xu, L., and J. Yang. 2001. Introduction to multi-criteria decision making and the evidential reasoning approach. Working paper No. 0106. Manchester School of Management, Manchester, UK.

Chapter 6: Multi-criteria Land Evaluation for Rice Suitability Analysis

6.1 Introduction

6.1.1 Land and its Importance

In its most common use, the term land refers to the part of the earth's surface that is not covered by water. But FAO (1995) gives land a broader definition which takes into account not only the physical entity, but also the associated physio-biotic and socio-economic as well. It defines land as "*a delineable area of the earth's terrestrial surface, encompassing all attributes of the biosphere immediately above or below this surface including those of the near-surface climate the soil and terrain forms, the surface hydrology (including shallow lakes, rivers, marshes, and swamps), the near-surface sedimentary layers and associated groundwater reserve, the plant and animal populations, the human settlement pattern and physical results of past and present human activity (terracing, water storage or drainage structures, roads, buildings, etc.).*" This definition is applicable to many land use planning activities.

The importance of land cannot be overemphasized:

- It supports production of biomass that provides food, fodder, fibre, fuel, timber and other biotic materials for human and other organisms use
- It provides biological habitats and gene reserves for plants, animals and micro-organisms.
- It acts as source and sink of greenhouse gases, thus regulating climate
- It influences quality and quantity of surface and ground water resources by regulating storage and flow
- Raw materials and minerals are naturally stored in land
- Natural and enhanced decontamination of hazardous compounds are supported by land
- Human settlements, and other structures are built on land
- Land acts as an archive where evidence of the cultural history of mankind and past climatic conditions are stored and protected.
- Land provides space for the transport of people and materials

The need, availability, and suitability of the land for these functions vary over the landscape. This necessitates inventory and evaluation of land resource for its suitability and management options for optimization and sustainability of the land functions. This result to allocation of land to land use types in a process known as land use planning.

The function of land use planning is to guide decisions on land use in such a way that the resources of the environment are put to the most beneficial use for man, whiles at the same time conserving those resources for the future (Beek et al., 1997). The land use planning is informed by land evaluation process.

6.1.2 Land Evaluation

Land evaluation is concerned with the assessment of land performance for specific land utilization purposes and provides a rational basis for taking land use decisions based on analysis of relations between the land use and land (FAO, 1985). In agriculture, land evaluation is a prerequisite to achieving optimum utilization of the available land resources for sustainable agricultural production (Perveen et al., 2008). The principle purpose of agriculture land suitability evaluation is to predict the potential and limitation of the land for crop production (Pan and Pan, 2012).

6.1.3. Land Evaluation Frameworks

Realizing the importance of land evaluation and land suitability in agriculture, the United Nations Food and Agricultural Organization developed a framework for land suitability (FAO, 1976). The framework involves construction of matching tables or transfer functions and subsequent calculations of suitability. Several land evaluation systems have been developed worldwide. Some of them are highlighted below.

In Sumatra, the land evaluation computer system (LECS) which based on the FAO framework for predicting local crop yields was developed (Wood and Dent, 1983). LECS has been used to assess the land suitability for a variety of crops but area specificity and simplicity have been mentioned as the constraints for its adoption (Elsheikh et al., 2013).

Micro-LEIS is another system created for land evaluation (Hoobler et al., 2003). It is an integrated system for land data transfer and agro-ecological land evaluation which is integrated with GIS (Rosa et al., 2009). The system provides a computer-based set of tools for an orderly arrangement and practical interpretation of land resources and agricultural management data. The limitation with this system is that the user cannot build a personal expert system (Nwer, 2006).

The automated land evaluation system (ALES) allows land evaluators to build expert systems (Rossiter and Wambeke, 1997). This system was developed for land evaluation according to the method presented in the Food and Agriculture Organization

Framework for land evaluation. It can be linked to socioeconomic evaluation, and has no fixed list for land characteristics or land use requirements (Elsheikh et al., 2013). However, the system cannot display maps and has no GIS functions (Rossiter and Wambeke, 1997). In a comparison study in Kilosa, Tanzania by Kimaro et al (2001), ALES was found to perform better than LECS, despite both of them having limitations mentioned above.

Kalogirou (2002) developed land evaluation using an Intelligent Geographical Information System (LEIGIS). The implementation of LEIGIS includes models for general cultivation and is limited for specific five crops: wheat, barley, maize, seed cotton, and sugar beet. The system also does not include climate, an important parameter for crop production.

6.1.4 Multi-criteria Land Evaluation

Among frequently used frameworks for land evaluation is the Multi-Criteria Evaluation (MCE) framework (Chen et al., 2008; Maddahi et al., 2014). The MCE is used because land evaluation process involves analysis of multiple land resources criteria such as climate, topography and soil properties, and socio-economic criteria such as accessibility, disease prevalence, and manpower availability.

The primary issue in MCE is concerned with how to combine the information from several criteria to form a single index of evaluation (Rahman and Saha, 2008). Several procedures can be used to achieve MCE:

Boolean overlay: In this procedure criteria are reduced to logical statements of suitability and then combined by means of logical operators such as intersection (AND) and union (OR).

weighted linear combination: in this procedure continuous criteria are standardized to a common numeric range and then combined by means of a weighted average. Criteria are combined by applying a weight to each followed by a summation of the results to yield a suitability map that may then be masked by one or more Boolean *constraints* to accommodate qualitative criteria.

ordered weighted average: This method offers a complete spectrum of decision strategies along the primary dimensions of degree of trade off involved and degree of risk in the solution. However, the simple Boolean operations sometimes are not suitable because they do not provide sufficient flexibility required for the analysis.

fuzzy technique : This approach combines decision or classification scores from multiple information sources into a single composite score by applying a fuzzy integral with respect to a designated fuzzy measure, representing differential weighting of scores derived from a variety of information sources. This technique results to a better real simulation than the Boolean method.

However, for continuous factors, a weighted linear combination is the most commonly used method (Zopounidis and Doumpos, 2002).

6.1.5 Analytical Hierarchy Process (AHP) as a Weighted Linear Combination Method for MCE

The Analytical hierarchy process (AHP) is a widely used method in decision-making (Ananda and Herath, 2003; Marinoni and Hoppe, 2006; Cherif Ahmed et al., 2007; Rahman and Saha, 2007; Perveen et al., 2008; Kihoro et al., 2013). The method tackles the weighting based on the hierarchical relationship among the objectives and their attributes (Malczewski, 1999; 2004). During the process, the objectives are defined at the higher level followed by identification of the attributes contributing to the objective at the lower level. Decision-makers (normally domain experts) then make pairwise comparisons between the attributes and develop relative weights using Saaty scale which range from 1 to 9. Consistency ratio (CR) is calculated to assess consistency of the experts' preferences. The pairwise comparison is repeated if CR is above 10% (Saaty and Vargas, 1991).

AHP method is superior to many other weighting methods because it can deal with inconsistent judgments by providing a measure of inconsistency. AHP can also be integrated into other software to provide greater flexibility and accuracy. For example, several multicriterial land use suitability analyses studies have been successfully accomplished by combining AHP and Geographical Information System - GIS (Marinoni and Hoppe, 2006; Cherif Ahmed et al., 2007; Kihoro et al., 2013). The integration of MCE in a GIS provides a powerful spatial decision support system which offers the opportunity to efficiently produce these land evaluation and crop suitability maps to aid in land use planning decisions (Mendas and Delali, 2012).

6.1.6 Objective of the Study

The objective of this study was to apply a weighted multi-criteria land evaluation procedure to undertake suitability analysis of a part of Kilombero Valley, Tanzania for rice production.

6.2 Methods

6.2.1 The Study Area

The study was conducted in Kilombero Valley, Tanzania. The valley is about 300 km east of Indian Ocean covering about 11000 km^2. The study site covered about 300 km^2 within the valley. The study site is located in UTM zone 37 south. It occupies northings from 9064697 to 9089031 m and eastings from 175422 to 197033 m. More description of the study area can be found in the introductory chapter of this document.

6.2.2 Identifying Land Evaluation Parameters

The first thing in land evaluation for crop suitability analysis is to identify parameter which will be used in the analysis. These are normally the attributes which its quality and/or quantity affects production of the crop. Literature review and opinions from crop experts, local agronomists and farmers were used in this study to prepare a list of parameters which were analysed for rice growth suitability. Land attributes identified were topography, surface water resource, accessibility, distance to markets, and soil properties. These land attributes were derived from different sources, mapped, classified, weighed, and finally combined in GIS environment to produce suitability map. Figure 6.1. Summarizes the procedures used for this study.

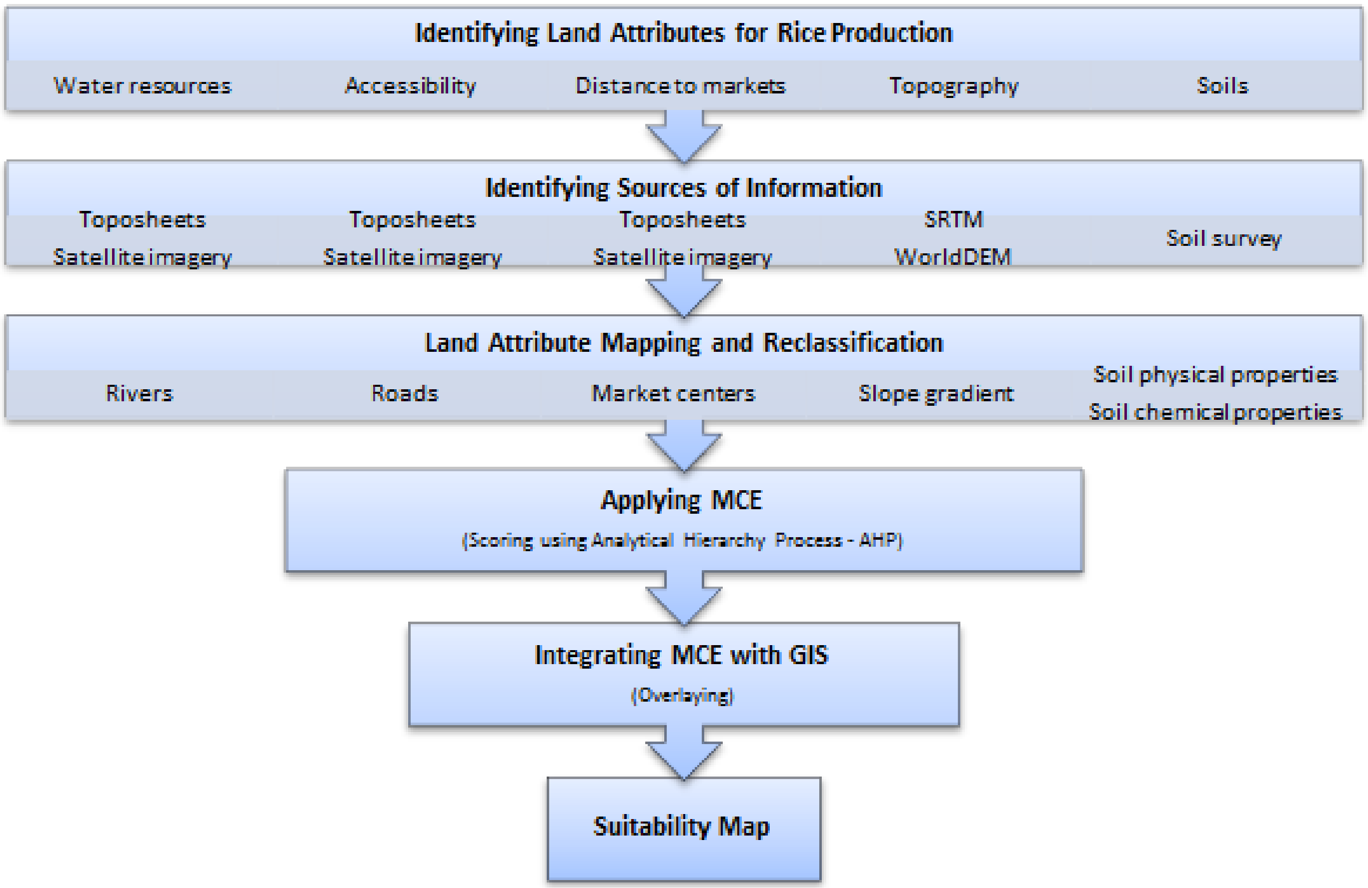


Figure 6.1. Workflow of methodology followed in this study

6.2.3 Land Attributes Spatial Information Generation and Classification

Procedures for generating spatial information for each criterion are described under their subheadings below. Since the criteria are measured in different dimensions or ranges of values, it was then necessary to convert them into one common scale by doing reclassification. The values of classes were derived based on the meaning of assigned class values of soil pH (Table 6.1). Values of a parameter indicating suitable conditions for rice production were assigned higher class values, consistent with the pH scale. For example, since low pH values are unfavourable for rice growth, they have been assigned low class values compared to the neutral pH values.

Table 6.1. Assigned values for soil pH classes (modified from Soil Survey Staff (1993)).

Denomination	pH range	Meaning	Assigned class value
Ultra acid	< 3.5	extremely low	0
Extreme acid	3.5–4.4	very low	1
Very strong acid	4.5–5.0	low	2
Strong acid	5.1–5.5	low - medium	3
Moderate acid	5.6–6.0	medium	4
Slight acid	6.1–6.5	high	5
Neutral	6.6–7.3	very high	6

Surface Water Resources

In the Kilombero Valley, rice is grown on flooded fields. Flooding in this valley is a function of rainfall and rivers which receive waters from the Udzungwa catchment. Given the small size of the study area, the influence of rainfall was assumed to be uniform throughout the area, and thus would not have varied influence for suitability analysis. Instead, rivers were used to map the surface water resource.

Major rivers were digitized as polylines from 1:50,000 scale Topographic maps sheets 234/1, 234/2, and 234/3 (Survey and Mapping Division, 1983) and 5 m resolution RapidEye satellite imagery. Onscreen digitization was done in ArcMap 10.1 (ESRI, 2010) software to produce surface water resource map.

Buffer tool in ArcMap was used to divide distances from the rivers. Buffers were created at 500 m, 1000 m, 1500 m, 2000 m, 2500 m, 3000 m, and lastly 4000 m. the shortest distance was assigned the highest class value (6) while the furthest distance was assigned the lowest class value (0). Classes near the rivers were given higher values, consistent with the idea that areas near the rivers and channels have more chances of being flooded and experiencing longer flooding durations than the distant ones. Since paddy rice prefer flooded condition, it is sensible to assign higher class values to areas with higher flooded chances and durations.

Accessibility

Accessibility to the rice fields in Kilombero valley is via footpaths and seasonal untarmacked roads. The roads were digitized as polylines from 1:50,000 scale Topographic maps sheets 234/1, 234/2, and 234/3 (Survey and Mapping Division, 1983) and 5 m resolution RapidEye satellite imagery. Onscreen digitization was done in ArcMap 10.1 (ESRI, 2010) software. Buffers were then created on the roads to produce 6 linear distance intervals map. The buffers were at distances 1000 m, 2000 m, 3000 m, 4000 m, 5000 m, and 6000 m. The shortest distance to the road was given a value of 6 while the furthest distance was given class value of 1. The greater the distance from the road, the lower the class value because of the cost and time used to reach the field and transporting the crops.

Distance to Markets

During harvesting period, buyers set buying posts and send agents to the village centers. These village centers therefore acts as market centers. Village centers coordinates were recorded during the field work and were mapped as points in ArcMap 10.1 (ESRI, 2010) to produce market centers' map. Buffers were created around the market centers to create distance spatial data. The distances were divided into 7 classes. The class distances were 1 km, 2 km, 3 km, 4 km, 5 km, 6 km, and 7 km. The 1 km distance was given the highest class value (6), while the last distance class was given the lowest class value (0). The classes near the market centers were thus given higher values.

Topography

1 arc (approximately 30 m resolution) Shuttle Radar Topography Mission (SRTM) terrain model (USGS, 2000) was used to derive the slope of the study area. The terrain model was preprocessed by filling sinks using algorithm developed by Planchon and Darboux (2001). The slope of each cell was determined by the rate of elevation change in the direction of steepest descent using algorithm in Whitebox Geospatial Analysis Tool following the formula by Horn (1981). Slope gradient raster map in degrees was reclassified to 6 classes: 0 -1, 1 – 2, 2 – 5, 5 – 8, 8 – 15, and >15. The lowest slope gradient class was given class value of 6, while the steepest class was given class value of 1. Lower slope gradients are given higher class values because surface runoff decreases with slope gradient.

Soil Properties

Soil properties were retrieved from a digitally predicted soil cluster map (Refer Chapter 4). Topsoil attribute values of the soil clusters generated during numerical clustering process (Refer Chapter 3) were used. Two reclass maps were developed from soil data: one for soil physical properties and the other for soil chemical properties.

Soil Physical Properties

To get soil physical properties reclass map, topsoil sand content of the soil clusters was reclassified. Initially, soil depth was also included in soil physical properties analysis, but later dropped because all soil observations showed there was no soil depth limitation. Higher sand contents in soil leads to increased infiltration rates. The sand classes were 48 – 55%, 55 – 65%, 65 – 75%, 75 – 85%, and > 85%. The class with lowest sand content was given a high class value of 5, while the highest sand content was given class value of 1. Higher infiltration rates are not good for lowland rice because it needs water ponding to thrive better. Classes with low sand contents were therefore assigned higher class values.

Soil Chemical Properties

The soil chemical properties reclass map was developed after overlay of reclassified pH, CEC, OC, TN, available P, exchangeable K, and Zn maps. Soil pH, CEC and organic carbon were selected because they are among stable indicators of soil quality. N, P and K were selected because these have been identified as the major limiting plant nutrients in Tanzania. Zn was selected from the list of studied micronutrients because Cu, Fe, and Mn levels were generally above the limiting values and below the toxic levels. Guide to ratings of CEC, OC, TN, P, and K were from Msanya et al. (2001), Zn from De Data (1989), soil pH from Soil Survey Staff (1993), and CEC from Hazelton and Murphy (2007).

Soil pH classes were 5.1 – 5.5, 5.5 – 6.0, 6.0 – 6.5, and 6.5 – 7.3. These classes were assigned class values from 3 to 6 (Table 6.1).

The CEC classes were 6 – 12 cmol(c)/kg, 12 – 25 cmol(c)/kg, 25 – 40 cmol(c)/kg, and > 40 cmol(c)/kg. The class with lowest CEC was given class value of 2, while the highest was assigned class value of 5.

The OC classes were 0 – 0.6%, 0.6 – 1.25%, 1.25 – 2.5%, 2.5 – 3.5%, and > 3.5%. The lowest class range was assigned class value of 1; the highest class range was assigned 6.

Class ranges for Total nitrogen (TN) were 0.01 – 0.1%, 0.1 – 0.2%, 0.2 – 0.4%, and 0.4 – 0.6%. The assigned class values were 1, 2, 3 and 5 respectively.

Available phosphorus classes were 1 – 7 mg/kg, 7 – 20 mg/kg, and > 20 mg/k. the assigned class values were respectively 2, 4, and 5.

The exchangeable potassium (K) class ranges were 0.03 – 0.13 cmol(c)/kg, 0.13 – 0.25 cmol(c)/kg, 0.25 – 0.5 cmol(c)/kg. These classes were assigned class values of 1, 2, and 3 respectively. The K levels were generally low for high class value assignment.

The classes for Zn were 0.1 – 0.5 mg/kg, 0.5 – 0.8 mg/kg, and 0.8 – 2 mg/kg. The assigned class values were respectively 2, 4, and 5.

6.2.4 Criteria Weighing

The criteria weights were assigned using a method called Analytic Hierarchy Process (AHP) developed by Saaty (1980). This method has been suggested for multi –criteria evaluation because of its flexibility and ability to provide a measure for consistence. Local extension staff and lead farmers were used to assign weights to the identified criteria in a procedure detailed in Chapter 5. The weights of the criteria used in this study as derived by local extension staff and lead farmers are summarized in Table 6.2.

Table 6.2. Weights and ranks assigned by a group of lead farmers and extension staff to the criteria identified for rice suitability analysis

Criteria	Assigned AHP derived weight	Rank
Soil physical properties	0.19	3
Soil chemical fertility	0.234	2
Accessibility	0.036	5
Distance to market	0.025	6
Surface water resources	0.462	1
Topography	0.052	4

6.2.5 Predicting Rice Suitability Map

Weighted overlay (MCE) tool in Whitebox GAT was used to produce the rice suitability map. Weights assigned to each criterin through AHP (Table 6.2) were employed to dictate contribution of each criteria layer in the overlay process. The final map class values ranged from 1.2703 to 5.3163. The map was then reclassified into classes values and meaning falling into that range based on the soil pH common scale (Table 6.1); 1 – 2 = 1 (very low suitability), 2 – 3 = 2 (low suitability), 3 – 4 = 3 (low to medium suitability), 4 – 5 = 4 (medium suitability), and 5 – 6 = 5 (high suitability). Area for each suitability class was calculated using Area tool in the Whitebox GAT.

6.3 Results and Discussion

6.3.1 Accessibility Suitability

The rice suitability of the study area based on accessibility is shown on Figure 6.2. Most of the area (59 %) was classified as very highly suitable (Table 6.3), meaning most fields are connected by roads and paths and are accessible when the area is not flooded. However, it should be noted that during flooding most of the areas are mainly accessed only on foot, power tillers and by bicycles. Heavy vehicles such as trucks and tractors cannot easily pass especially in the spots with clayey soils. Apart from flooding, the other accessibility limitation is lack of bridges to facilitate river crossing. The accessibility map was developed based on the distance to the roads, but did not consider the roads conditions and if they are passable all year round.

The accessibility was not perceived as a very important criterion in the ranking process where it was given a weight of 0.036 in a 0 – 1 scale (Table 6.2). This could be because the farmers and local extension staff see that most activities heavily requiring mechanization are done before and after flooding. Currently mostly mechanized rice production activities in the study area are land preparation where tractors are increasingly used and transportation of harvested rice where trucks and other vehicles are used. However, this perception might change when more mechanization in other agronomic practices such as weed control will be employed.

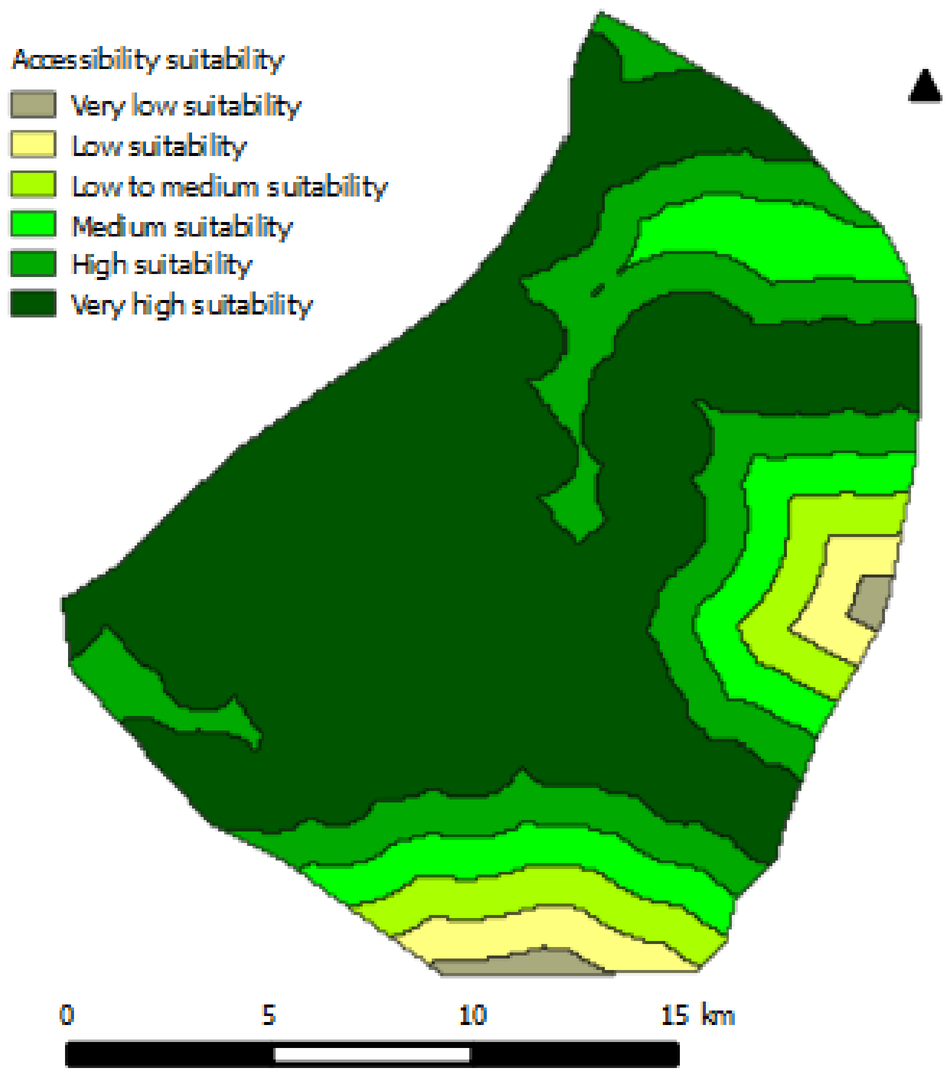


Figure 6.2. Rice production suitability based on accessibility

Table 6.3. Percentage of areas suitable for rice production based on accessibility

Class Value	Meaning	Area (m^2)	% of total area
1	very low suitability	2,891,700	1.0
2	low suitability	11,072,700	3.7
3	low - medium suitability	17,196,300	5.7
4	medium suitability	32,205,600	10.8
5	high suitability	58,077,000	19.4
6	very high suitability	178,062,300	59.5

6.3.2 Market Accessibility Suitability

The market access suitability was mapped based on the distance to the market centers as described in the methodology section. Areas closer to the market centers have higher suitability. About 40% of the study area is located where it can be categorized as having high market suitability. About 25% are medium suitable, while about 35% are generally marginally suitable by being far from market centers (Table 6.4). The distribution of the market access suitability area is shown on Figure 6.3.

Despite the importance of distance to market centers, the dynamics of the markets in the study area are greatly influenced by the yield and amount of rice in the market. When yields are generally good, the competition among the buyers is low and farm gate prices are low. Some farmers would opt to store and wait, anticipating for a better price in a near future. Some farmers are compelled to sell, nevertheless, in order to meet financial needs such as paying harvesting and transportation costs. In some instances, like in 2013/14 season, the country was flooded with low-taxed cheap rice imported from Asian countries where production cost is low due to mechanization and exploitation of economies of scale. The farm gate prices fall drastically in such instances. However, the importance of distance to market holds as those closer to the market will get relatively better prices than those who will need to incur additional transport costs.

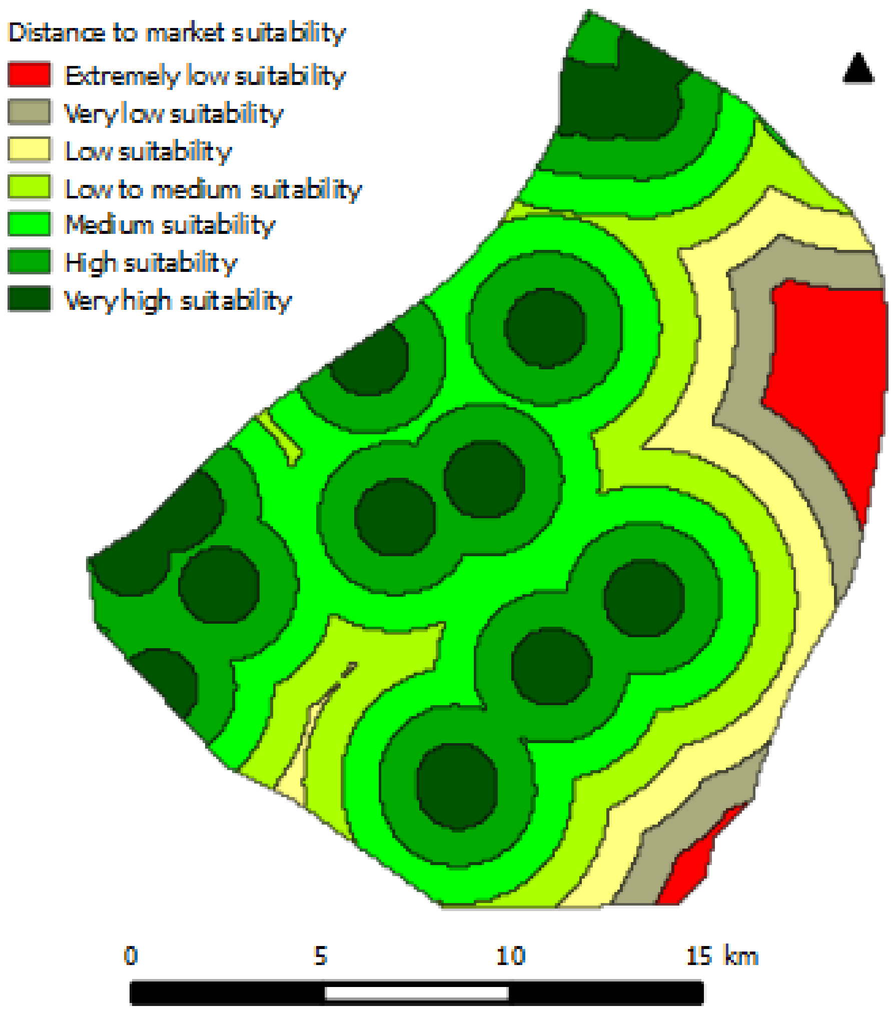


Figure 6.3. Rice production suitability based on distance to market

Table 6.4. Percentage of area suitable for rice production based on distance to market

Class Value	Meaning	Area (m^2)	% of total area
0	extremely low suitability	15,568,200	5.2
1	very low suitability	17,795,700	5.9
2	low suitability	26,470,800	8.8
3	low - medium suitability	44,396,100	14.8
4	medium suitability	74,876,400	25.0
5	high suitability	79,979,400	26.7
6	very high suitability	40,419,000	13.5

6.3.3 Topographic Suitability

Topographic suitability was derived from slope gradient. Other topographic derivatives such as slope aspect and elevation were not used because of the relative flatness of the study area. The topographic suitability is displayed on Figure 6.4. The dominant slope gradient range is 2 – 5 degree. This occupies 59% of the study area and is classified as medium suitable for rice production. About 35% of the area has slope gradient below 2 degrees and is classified as highly suitable for rice production (Table 6.5).

The topographic suitability is classified based on influence of slope gradient to surface runoff and water loss. Steeper slopes are known to favour faster surface runoff restricting water ponding as opposed to gentle slopes. However, this should be interpreted in conjunction with the ability of the soil to retain surface water even if runoff is restricted by gentle slopes. Soil texture is one of the properties which need to be studied because it gives a picture on water infiltration potential. Due to this, soil texture has been used to develop soil physical properties suitability map for this study.

In the study area, the relatively high slope areas are used for settlement and growth of upland crops such as maize, beans, cocoa, banana, and a variety of trees for fruits and timber. These areas are also used for grazing, especially when rice fields are flooded and the crop is in the field.

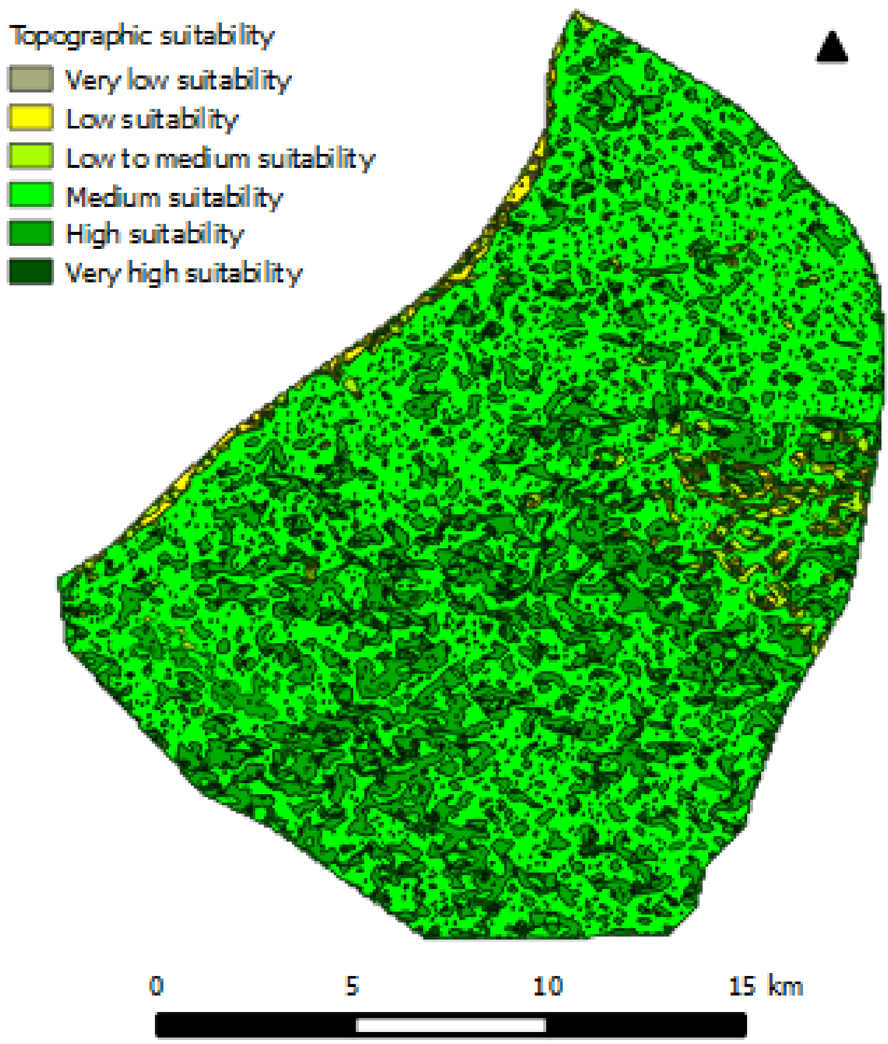


Figure 6.4. Rice production suitability based on topography

Table 6.5. Percent of areas suitable for rice production based on topography

Class Value	Meaning	Area (m^2)	% of total area
1	very low suitability	226,800	0.1
2	low suitability	4,341,600	1.4
3	low - medium suitability	11,939,400	4.0
4	medium suitability	177,916,500	59.4
5	high suitability	96,705,900	32.3
6	very high suitability	8,367,300	2.8

6.3.4 Surface Water Resources Suitability

The surface water resource suitability is displayed on Figure 6.5. About 15% of the study area is classified as being of low suitability to rice production based on their distance from the surface water resources (Table 6.6). These areas are likely to have less chances of being flooded, also are likely to have short durations of flooding because of the small amount of flood water they are likely to get due to being far from the water resources.

Other factors such as infiltration which is dictated by soil physical properties and slope gradients are also important in deciding the duration and amount of flooding. So, interpretation of flood related suitability of rice production need to consider in unison, surface water resources, topography, and soil physical properties. All these criteria were considered at the end for the analysis of the suitability of rice production in this study.

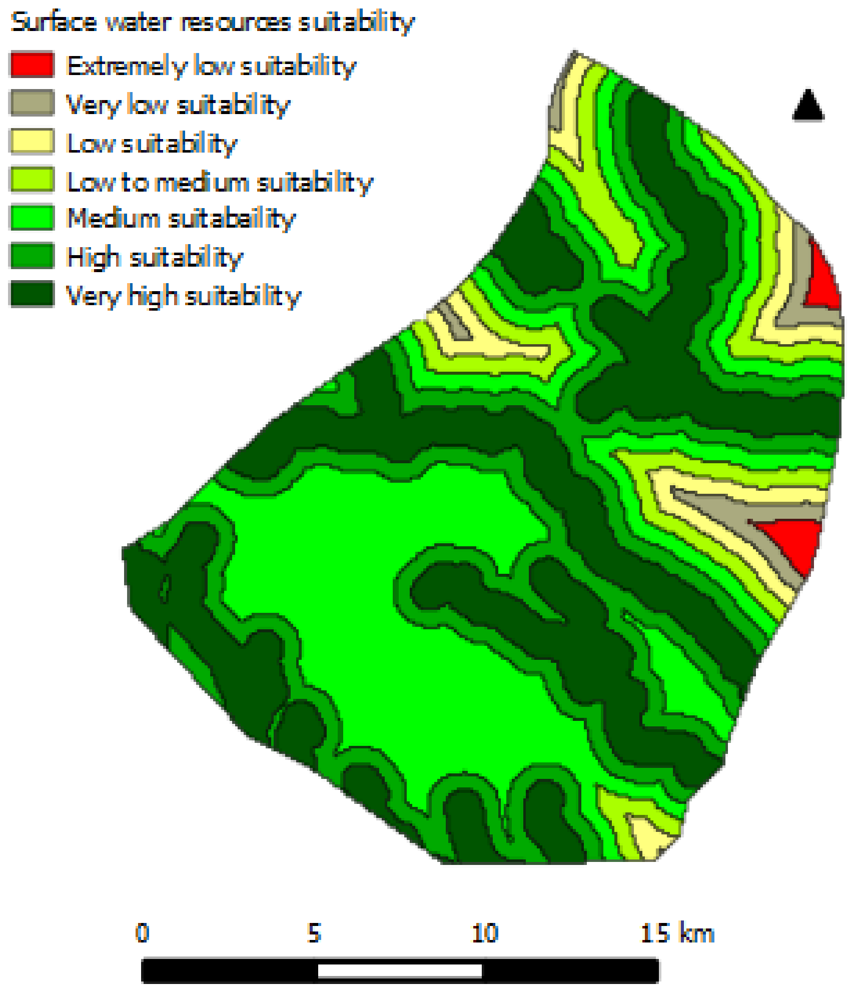


Figure 6.5. Rice production suitability based on distance to surface water resources

Table 6.6. Percentage of areas suitable for rice production based on their distance to surface water resources

Class Value	Meaning	Area (m^2)	% of total area
0	extremely low suitability	2,932,200	1.0
1	very low suitability	6,836,400	2.3
2	low suitability	13,721,400	4.6
3	low - medium suitability	22,923,000	7.7
4	medium suitability	92,680,200	30.9
5	high suitability	70,923,600	23.7
6	very high suitability	89,488,800	29.9

6.3.5 Soil Physical Properties Suitability

Figure 6.6 displays distribution of soil physical properties suitability classes in the study area. As described in the methodology section, this layer was developed using sand content because other soil physical properties such as soil effective depth were not variable to influence suitability rankings. Other soil physical properties such as bulk density and infiltration rates were not determined and are generally related to soil texture.

Most of the study area (56%) was estimated to have sand contents which are low to medium suitable for low land rice production (Table 6.7). Only about 10% of the area is having finer textures which would hold water relatively well for paddy rice which prefers ponding conditions. Obviously, this criterion looked soil texture on the side of holding ponding of water, and neglected the easiness of workability for high sand content soils. However, the experience show farmers would prefer flooded fields because the condition also helps to manage weeds.

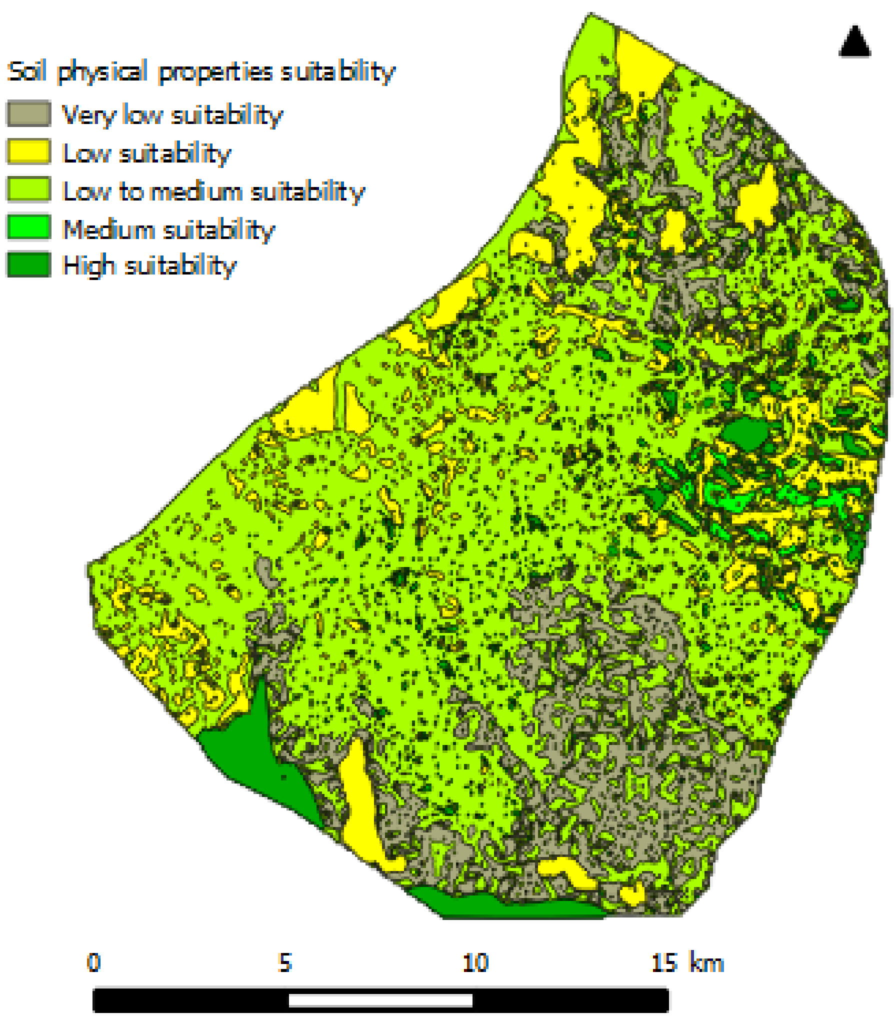


Figure 6.6. Rice production suitability based on soil physical properties

Table 6.7. Percentage of areas suitable for rice production based on soil physical properties

Class Value	Meaning	Area (m^2)	% of total area
1	very low suitability	46,364,400	15.5
2	low suitability	53,549,100	17.9
3	low - medium suitability	168,949,800	56.4
4	medium suitability	8,375,400	2.8
5	high suitability	22,266,900	7.4

6.3.6 Soil Chemical Fertility Suitability

The spatial distribution of the determine soil chemical fertility suitability based on soil pH, organic carbon, CEC, total nitrogen, available phosphorus, exchangeable potassium and zinc is shown on Figure 6.7. The soil chemical suitability ranges from low to medium suitability. About 63% of the area is classified as of medium soil fertility suitability, while the rest have lower suitability (Table 6.8).

The areas which showed low suitability are generally those closer to the escarpment. Historically, the areas closer to the escarpment were put into agriculture earlier than those close to the main Kilombero River at the center of the valley. Kato (2007) narrates that early famers started rice fields close to their settlement which were ideally located on higher altitudes closer to the Escarpment on the western side and closer to the Mahenge Mountains on the eastern side of the valley. As productivity decreased due to fertility degradation and weed infestation, also due to increased population, new fields were opened gradually towards the center of the valley. Relatively higher fertility could be explained by the time the fields have been put into agriculture.

Fertility management in the study area by application of fertilizers and conservation agriculture is very low (Massawe and Amury, 2012). Use of fertlizers is low because of lack of awareness and affordability by small scale farmers. Another factor is the fluctuation of the rice prices, such that farmers are not sure of getting returns if the farm gate prices are low. There is a general lack of knowledge about use of conservation agriculture. Disc plowing is very popular currently, residues are burned or grazed on, and a few farmers apply blanket recommended N, P, K, and S containing fertilizers.

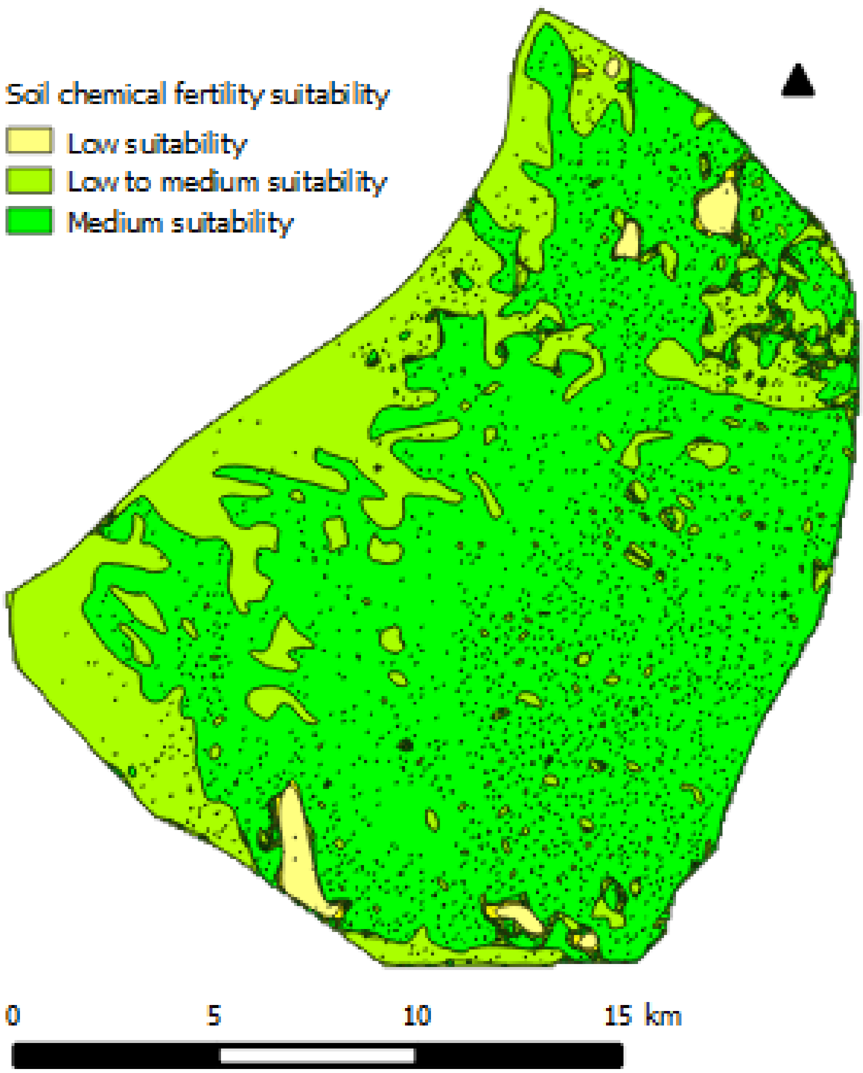


Figure 6.7. Rice production suitability based on soil chemical fertility

Table 6.8. Percentage of areas suitable for rice production based on soil chemical fertility

Class Value	Meaning	Area (m^2)	% of total area
2	low suitability	5,443,200	1.8
3	low - medium suitability	103,939,200	34.7
4	medium suitability	190,123,200	63.5

6.3.7 Overall Rice Suitability

The overall suitability of the study area for rice production based on the 6 identified criteria (accessibility, markets, topography, water resources, soil physical properties and soil chemical properties) is displayed on Figure 6.8. About 8% of the area is classified as having low suitability for rice production while only 2% is highly suitable. The majority of the area (about 89%) is classified as having medium suitability for rice production (Table 6.9).

Being a lowland rice production area, the dominance of surface water resources for the suitability of an area is clearly seen. Areas closer to the water sources have generally higher suitability values, a result of higher weights given to the criterion by the farmers and extension staff (Table 6.2).

The dominance of the water resource implies that the suitability can be greatly improved by managing the water resources. This may include constructions of soil bunds to restrict surface lateral movement of water once they have entered the fields. Construction of irrigation schemes will also be very useful. However, the major challenge here is the soil textures which are generally coarse resulting to higher infiltration rates (Bonarius, 1975).

Improved fertility management can also have big influence in rice production suitability of the study area. Soil testing and site specific application of fertilizers will be very beneficial. As such, except for soil pH, CEC, and OC, the rest of the soil chemical fertility limitations can be addressed and improved within a short period of time by application of correct type of fertilizer in required amount and placement technique. Efforts are being done in educating small scale holders about the importance of fertilizer applications and fertility management through Farmers Field Schools (FFS). The result of this will be improved fertility management and use of improved seeds. The training can also include other beneficial agronomic practices such as minimum tillage to improve soil organic matter and CEC build ups

Access to rice fields and markets can be greatly improved by improving roads and bridges infrastructures. This can be local or central government efforts which would greatly improve food production and hence reduce food insecurity.

This multi-criteria evaluation has given us some light about the suitability of the rice in the study area. A number of criteria or details were left out because of logistics or unavailability of the data. Factors like type and extent of weed infestation, access to microfinances, relative importance of one market center to another, or access to extension services would have added value to this suitability analysis study.

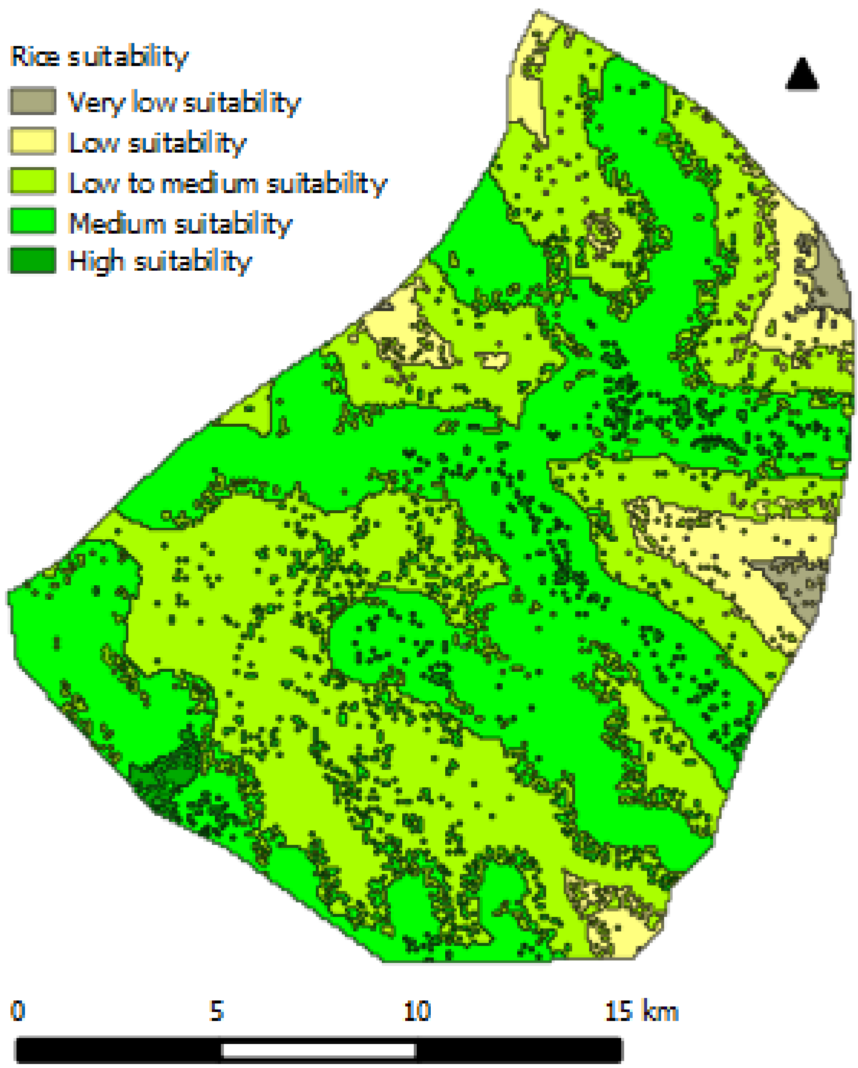


Figure 6.8. Rice production suitability based on the six identified suitability analysis criteria

Table 6.9. Percent of areas suitable for rice production based on the six identified rice suitability analysis criteria

Class Value	Meaning	Area (m^2)	% of total area
1	very low suitability	3,102,300	1.0
2	low suitability	22,396,500	7.5
3	low - medium suitability	127,113,300	42.4
4	medium suitability	141,669,000	47.3
5	high suitability	5,216,400	1.7

6.4 Conclusion

The suitability of land in the study area for rice production was analysed using six criteria, which were identified by local extension staff and farmers and through literature search. The criteria were accessibility, distance to markets, topography, surface water resources, soil physical properties, and soil chemical properties.

About 8% of the area is classified as having low suitability for rice production while only 2% is highly suitable. The majority of the area (about 89%) is classified as having medium suitability for rice production.

The suitability decision was dominated by the surface water resource criterion. The criterion was assigned relatively higher weight compared to other identified criteria, understandably due to reliance of the low land rice production to flooding conditions. The areas close to the rivers are thus relatively more suitable for rice production. The dominance of the water resource implies that the suitability can be greatly improved by managing the water resources. This may include constructions of soil bunds to restrict surface lateral movement of water once they have entered the fields. Construction of irrigation schemes will also be very useful. However, the major challenge here is the soil textures which are generally coarse resulting to higher infiltration rates.

Improved fertility management can also have big influence in rice production suitability of the study area. Soil testing and site specific application of fertilizers will be very beneficial. As such, except for soil pH, CEC, and OC, the rest of the soil chemical fertility limitations can be addressed and improved within a short period of time by application of appropriate type of fertilizer in appropriate amount, time and placement technique.

Access to rice fields and markets can be greatly improved by improving roads and bridges infrastructures. This can be local or central government efforts which would greatly improve food production by allowing all year round accessibility of the production by farm machinery and vehicles, hence reduce food insecurity.

A number of criteria or details were left out because of logistics or unavailability of the data. Factors like type and extent of weed infestation, access to microfinances, relative importance of one market center to another, or access to extension services would have added value to this suitability analysis study. However, the spatial information provided gives a guidance as to where and what interventions are needed in order to improve rice productivity in Mngeta, Mchombe, and Njage areas in Kilombero Valley, Tanzania.

6.5 References

Ananda, J., and G. Herath. 2003. The use of analytic hierarchy process to incorporate stakeholder preferences into regional forest planning. Forest Policy and Economics 5:13–26.

Beek, K.J., K. de Bie, and P. Driessen. 1997. Land information and land evaluation for land use planning and sustainable land management. International Institute for Aerospace Survey and Earth Sciences (ITC). The Netherlands.

Bonarius, H. 1975. Physical properties of soils in the Kiolmbero Valley, Tanzania. German Agency for Technical Cooperation (GTZ), Germany.

Chen, Y., K. Khan, and Z. Padar. 2008. Irrigation intensification or extensification assessment: a GIS-based spatial fuzzy multi-criteria evaluation. In: Proceedings of the 8th international symposium on spatial accuracy assessment in natural resources and environmental sciences, 25–27 June, Shanghai, P.R. China. p. 309–318.

Cherif Ahmed, A. O., R. Nagasawa, K. Hattori, D. Chongo, and M.F. Perveen. 2007. Analytical Hierarchic Process in conjunction with GIS for identification of suitable sites for water harvesting in the oasis areas: Case study of the Oasis zone of Adrar, Northern Mauritania. Journal of Applied Sciences 7(19):2911-2917.

De Data, S.K. 1989. Rice. In: D.L. Plucknett, and H.B. Sprague, editors, Detecting mineral nutrient deficiencies in tropical and temperate crops. Westview Press Inc. Boulder, CO. p. 41-51.

Elsheikh, R., A. Rashid, B.M. Shariff, F. Amiri, N.B. Ahmad, S.K. Balasundram, M.A.M. Soom. 2013. Agriculture Land Suitability Evaluator (ALSE): A decision and planning support tool for tropical and subtropical crops. Computers and Electronics in Agriculture 93:98–110.

ESRI. 2010. ArcGIS - A Complete Integrated System. Environmental Systems Research Institute, Inc., Redlands, California.

FAO 1976. A framework for land evaluation. Soils Bulletin 3, Rome.

FAO 1985. Guidelines: Land Evaluation for Irrigated Agriculture. FAO Soils Bulletin 55. Food and Agriculture Organization of the United Nations, Rome.

FAO 1995. Planning for sustainable use of land resources; Towards a new approach. Land and Water Bulletin 2, Rome.

FAO. 1976. A framework for land evaluation. Food and Agriculture Organization of the United Nations, Soils Bulletin 32. FAO, Rome.

Hazelton, P., and B. Murphy. 2007. Interpreting soil test results. CSIRO Publishing, Collingwood Victoria, Australia.

Hoobler, B.M., G.F. Vance, J.D. Hamerlinck, L.C. Munn, and J.A. Hayward. 2003. Applications of land evaluation and site assessment (LESA) and a geographical information system in East Part Couny, Wyoming. Journal of Soil and Water Conservation 58:105–112.

Horn, B.K.P. 1981. Hill shading and the reflectance map. Proceedings of the IEEE 69(1):14-47.

Kalogirou, S. 2002. Expert systems and GIS: an application of land suitability evaluation. Computers, Environment and Urban Systems 26:89–120.

Kato, F. 2007. Development of a major rice cultivation area in the Kilombero Valley, Tanzania. African Study Monographs, Supplement 36: 3-18.

Kihoro, J., N.J. Bosco, and H. Murage. 2013. Suitability analysis for rice growing sites using a multicriteria evaluation and GIS approach in Great Mwea region, Kenya. SpringerPlus 2013, 2:265. http://www.springerplus.com/content/2/1/265.

Kimaro, D.N., B.M. Msanya, S. ArakI, and H. Otsuka. 2001. Application of computerized land evaluation systems in Tanzania: a case study in Kilosa District, Morogoro Region. UNISWA Journal of Agriculture 10:40-50.

Maddahi, Z., A. Jalalian, M.M.K. Zarkesh, and N. Honarjo. 2014. Land suitability analysis for rice cultivation using multi criteria evaluation approach and GIS. European Journal of Experimental Biology 4(3):639-648.

Malczewski, J. 1999. GIS and Multicriteria Decision Analysis. John Wiley and Sons, New York.

Malczewski, J. 2004. GIS-based land-use suitability analysis: a critical overview. Progress in Planning 62:3–65.

Marinoni, O., and A. Hoppe. 2006. Using the analytical hierarchy process to support sustainable use of geo-resources in metropolitan areas. Journal of Systems Science and Systems Engineering 15(2):154-164.

Massawe, B.H.J. and N.A. Amury. 2012. Report on agronomic practices and soil fertility analysis for improved rice production in Kilombero and Wami Valleys in Tanzania. USAID Feed the Future NAFAKA staples value chain activity project.

Mendas, A. and A. Delali. 2012. Integration of multi-criteria decision analysis in GIS to develop land suitability for agriculture: application to durum wheat cultivation in the region of Mleta in Algeria. Computers and Electronics in Agriculture 83:117–126.

Msanya, B.M., D.N. Kimaro, E.P. Kileo, G.G. Kimbi, and S.B. Mwango. 2001. Land suitability evaluation for the production of food crops and extensive grazing: a case study of Wami Plains in Morogoro Rural District, Tanzania. Soils and Land Resources of Morogoro Rural and Urban Districts, Volume 1. Department of Soil Science, Faculty of Agriculture, Sokoine University of Agriculture, Morogoro, Tanzania.

Nwer, B.A.B. 2006. The Application of land evaluation technique in the north-east of Libya. Faculty of Environment. Cranfield University, Silsoe.

Pan, G., and J. Pan. 2012. Research in crop land suitability analysis based on GIS. Computer and Computing Technologies in Agriculture 365:314–325.

Perveen, F., R. Nagasawa, A.O.C. Ahmed, I. Uddin, and R. Kimura. 2008. Integrating biophysical and socio-economic data using GIS for land evaluation of wheat cultivation: A case study in north-west Bangladesh. Journal of Food, Agriculture and Environment 6(2):432-437.

Planchon, O., and F. Darboux. 2001. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. Catena 46, 159–176.

Rahman, R., and S. K. Saha. 2008. Remote sensing, spatial multi criteria evaluation (SMCE) and analytical hierarchy process (AHP) in optimal cropping pattern planning for a flood prone area. Journal of Spatial Science 53(2):161–177.

Rosa, D.D.L., M. Anaya-Romero, E. Diaz-Pereira, N. Heredia, and F. Shahbazi. 2009. Soil specific agro-ecological strategies for sustainable land use. A case study by using MicroLEIS DSS in Sevilla province. Land Use Policy 6:1055–1065.

Rossiter, D.G. and A.R.V. Wambeke. 1997. Automated Land Evaluation System ALES Version 4.65 User's Manual. Cornell University.

Saaty, T.L. 1980. The analytic hierarchy process, McGraw-Hill, New York.

Saaty, T.L., and L.G. Vargas. 1991. Prediction, projection and forecasting. Kluwer Academic Publishers, Dordrecht.

Soil Survey Division Staff. 1993. Soil survey. Soil Conservation Service, U.S. Department of Agriculture Handbook 18.

Survey and Mapping Division. 1983. Mbingo 1:50000 Toposheet No. 234/2. Ministry of Lands, Housing and Urban Development, Dar es Salaam, Tanzania.

Survey and Mapping Division. 1983. Mngeta 1:50000 Toposheet No. 234/3. Ministry of Lands, Housing and Urban Development, Dar es Salaam, Tanzania.

Survey and Mapping Division. 1983. Njagi 1:50000 Toposheet No. 234/1. Ministry of Lands, Housing and Urban Development, Dar es Salaam, Tanzania.

USGS. 2000. Shuttle Radar Topography Mission, 1 Arc Second scene SRTM. http://earthexplorer.usgs.gov/ page visited 24 October, 2014.

Wood, S.R., and F.J. Dent. 1983. LECS. A land evaluation computer system methodology. Bogor: Ministry of Agriculture/PNUD/FAO, Centre for Soil Research, Indonesia.

Zopounidis, C., and M. Doumpos. 2002. Multicriteria classification and sorting methods: a literature review. European Journal of Operational Research 138:229-246.

Chapter 7: Conclusion

This dissertation has performed a GIS-based multi-criteria land evaluation for rice production in Kilombero Valley, Tanzania. The dissertation is organized into chapters in which an earlier chapter provides information, which is needed and used in the following chapter.

In chapter 2, characterization and classification of soils of the study area was done using field and laboratory methods.

The dominant soil textural classes were found to be sandy clay loams and sandy loams accounting for about 66% of the studied soil horizons. Clays accounted for 13% and sands accounted for only 3%. Due to dominance of medium to coarse soil fractions, higher infiltration rates and hydraulic conductivities, as well as leaching of base cations are expected.

Soil pH values were found to vary widely spatially and vertically, ranging from 4.7 to 7.2, with topsoil values predominantly being around 5.8. The soil organic carbon values were generally low where levels are predominantly below 4%. Generally, the organic carbon levels decreased with soil depth indicating accumulation in the surface horizons. Some pedons showed irregular decrease/distributions of organic carbon with depth, a common indication of alluvial soils.

The soil CEC values were generally low with many pedons having values below 20 cmol(+)/kg in the surface horizons, and below 15 cmol(+)/kg in their subsurface horizons. The vertical variability showed that the CEC generally decreases with soil depth, following the trends displayed by organic carbon.

Gleying properties were observed in most part of the study area supporting records of repeated flooding and drying of the landscape. Mostly, the upper most gleyed horizons were observed to be Bg horizons followed by Btg, Cg, ABg, Bssg, and BCg in that order. Gleying was not observed in the topsoils because of disturbances caused by tillage.

The soils of the study area classified into Inceptisols, Entisols, Alfisols, Mollisols and Vertisols soil orders. Inceptisols and Entisols were the dominant soil orders, represented by 31 out of 42 classified pedons. Generally, the soils of Kilombero valley represent properties common for alluvial soils worldwide.

In chapter 3, distance metrics were used to numerically cluster the horizons and pedons described and sampled in chapter 2. A total of 11 horizon clusters and their respective centroids to represent modal horizons were generated and tested for pedological consistency. The predicted horizon clusters demonstrated attributes relationships expected in pedology.

Further, OSACA was used to cluster the soil profiles. A total of 13 solum clusters with their modal pedons attributes were generated. Variability in trends and values with soil depth were observed and confirmed that the clustering process was able to separate different soil types present in the study area. Different soil forming processes were also picked. For example cyclic deposition of sediments was shown by irregular decrease in soil organic carbon with depth. Likewise, clay illuviation to form Bt horizons was shown by increased clay content in the subsoil on stable landscapes.

In chapter 4, the soil clusters numerically classified in chapter 3 were predicted and mapped in areas where observation was not done using machine learning based digital soil mapping techniques. Two decision tree algorithms: J48 and Random Forest (RF) were used to train the data derived from two digital elevation models (DEMs), satellite image and a legacy map.

It was noted that J48 learner tended to predict soil clusters represented by higher training instances (samples), leaving out those with lower representation while RF was able to predict all clusters, regardless of the number of training instances falling in the soil cluster. Results suggested that using RF learner on SRTM data set produced better predictive soil cluster maps. Therefore, the soil cluster map predicted from RF trained SRTM data set was used in chapter 6 as soil information input in multi-criteria land evaluation for rice suitability analysis.

Chapter 5 was dedicated to identification of other inputs (criteria) required for land evaluation for rice suitability apart from soil information. Local experts and lead farmers from the study area were used. The chapter also reports on the Analytical Hierarchy Process (AHP) which was used to assign weights and rank the criteria, ready for the multicriteria analysis. The identified criteria were accessibility, distance to rice markets, topography, surface water resources, and soil properties. Surface water resources and soil properties were given higher weights in the AHP exercise compared to other criteria.

The GIS based multi-criteria decision analysis was done in chapter 6. Results show that about 8% of the area was classified as having low suitability for rice production while only 2% is highly suitable. The majority of the area (about 89%) was classified as having medium suitability for rice production.

Since the suitability decision was dominated by the surface water resource criterion, the suitability can be greatly improved by managing the water resources. However, the major challenge here is the soil textures which are generally coarse resulting to higher infiltration rates. Improved fertility management can also have big influence in rice production suitability of the study area. Soil testing and site specific application of fertilizers will be very beneficial. Access to rice fields and markets can be greatly improved by improving roads and bridges infrastructures.

The results of this works are expected to aid rice farmers, extension staff, land use planner and other stakeholders to make informed decisions which will help to improve rice productivity in Mngeta, Mchombe, and Njage areas in Kilombero Valley, Tanzania.

References

Acres, B.D. 1983. The soils of Tabora Region. Tabora Rural Integrated Development Project, Tanzania. Project Report No.67, Land Resources Development Centre, UK. ODA.

Alves, C., and D. Ross. 2010. Examining phosphorus contributions from alluvial soils – a comparison of three Vermont, U.S.A. River corridors. In Proceedings of 19th World Congress of Soil Science, Soil Solutions for a Changing World 1 – 6 August, Brisbane, Australia. 228 – 231.

Ananda, J., and G. Herath. 2003. The use of analytic hierarchy process to incorporate stakeholder preferences into regional forest planning. Forest Policy and Economics 5:13–26.

Baker, R.M. 1970. The Soils of Tanzania. FAO. Dar es Salaam, Tanzania.
Balasubramanian, V., M. Sie, and R.J. Hijmans, K. Otsuka. 2007. Increasing rice production in Sub - Saharan Africa: challenges and opportunities. Advances in Agronomy 94:55-133.

Barthold, F.K., M. Wiesmeier, L. Breuer, H.G. Frede, J. Wu, and F.B. Blank. 2013. Land use and climate control the spatial distribution of soil types in the grasslands of Inner Mongolia. Journal of Arid Environments 88:194–205.

Baxter, S.J., and D.M. Crawford. 2008. Incorporating legacy soil pH databases into digital soil maps. In: A.A. Hartemink, A. McBratney, and M.L. Mendonca-Santos, editors, Digital soil mapping with limited data. Springer. p. 311-318.

Beck, A.D. 1964. The Kilombero Valley of south-central Tanganyika. East Africa Geography Review 2:37-43.

Beek, K.J., K. de Bie, and P. Driessen. 1997. Land information and land evaluation for land use planning and sustainable land management. International Institute for Aerospace Survey and Earth Sciences (ITC). The Netherlands.

Behrens, T., H. Förster, T. Scholten, U. Steinrücken, E. Spies and M. Goldschmitt. 2005. Digital soil mapping using artificial neural networks. Journal of Plant Nutrition and Soil Science 168:21-33.

Bell, J.C., D.F. Grigal, and P.C. Bates. 2000. A soil–terrain model for estimating spatial patterns of soil organic carbon. In: J.P. Wilson, and J.C Gallant, editors, Terrain analysis: principles and applications. John Wiley & Sons, Inc. New York, p. 295– 310.

Bidwell, O.W., and F.D. Hole. 1964a. An experiment in the numerical classification of some Kansas soils. Soil Science Society of America Proceedings 28:263-268.

Bidwell, O.W., and F.D. Hole. 1964b. Numerical taxonomy and soil classification. Soil Science 97: 58-62.

Bidwell, O.W., L.F. Marcus, and P.K. Sarkar. 1964. Numerical classification of soils by electronic computer. Transactions of International Congress of Soil Science 5:933-941.

Bishop, T.F.A., McBratney, A.B., and G.M. Laslett. 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91:27–45.

Boettinger, J.L., D.W. Howell, A.C. Moore, A.E. Hartemink, and S. Kienast-Brown. 2010. Digital soil mapping: Bridging research, environmental application, and operation. Springer-Verlag, Dordrecht, the Netherlands.

Boettinger, J.L., R.D Ramsey, J.M. Bodily, N.J. Cole, S. Kienast-Brown, S.J.Nield, A.M. Saunders, and A.K. Stum. 2008. Landsat spectra data for digital soil mapping. In: A.A. Hartemink, A. McBratney, and M.L. Mendonca-Santos, editors, Digital soil mapping with limited data. Springer. p. 193-202.

Bonarius, H. 1975. Physical properties of soils in the Kiolmbero Valley, Tanzania. German Agency for Technical Cooperation (GTZ), Germany.

Brady, N.C. and R. Weil. 2002. The Nature and Properties of Soils. Thirteenth Edition, Prentice Hall, Upper Saddle River, NJ.

Brady, N.C., and R.R. Weil. 2010. Elements of the Nature and Properties of Soils. Pearson Hall, NY.

Bray, R.H., and L.T. Kurtz. 1945. Determination of total, organic and available forms of phosphorus in soils. Soil Science 59:39-45.

Breiman, L. 1996. Bagging predictors. Machine Learning 24:123-140.

Breiman, L. 2001. Random Forests. Machine Learning 45:5-32.

Bremner, J.M., and C.S. Mulvaney. 1982. Total nitrogen. In: A.L. Page, R.H.Miller, and D.R. Keeney, editors, Methods of soil analysis, Part 2; Chemical and mineralogical properties, 2nd edition. American Society of Agronomy, Madison, Wisconsin. p. 595-624.

Bruland, G.L., and C.J. Richardson. 2004. Hydrologic gradient and topsoil additions affect soil properties of Virginia Created wetland. Soil Science Society of America Journal 68:2069 – 2077.

Brungard, C.B., and J.L. Boettinger. 2012. Spatial prediction of biological soil crust classes; value added DSM from soil survey. In: B. Minasny, B.P. Malone, and A. McBratney, editors, Digital soil assessments and beyond: Proceedings of the 5th global workshop on digital soil mapping. CRC Press, Sydney. p. 57–60.

Brungard, C.W., J.L. Boettinger, M.C. Duniway, S.A. Wills, and T.C. Edwards Jr. 2015 Machine learning for predicting soil classes in three semi-arid landscapes Geoderma 239–240:68–83.

Bucheyeki, T.L., E. Shennkalwa, D. Kadadi, and J. Lobulu. 2011. Assessment of rice production constraints and farmers preferences in Nzega and Igunga Districts. Journal of Advances in Developmental Research 2(1): 30-37.

Buol, S.W., R.J. Southard, R.C. Graham, and P.A. McDaniel. 2002. Soil Genesis and classification. 5th Edition. Iowa State University Press, Ames, IA.

Burrough, P.A. 1986. Principles of geographical information systems for land resources assesment. Oxford University Press.

Calton, W.E. 1954. An experimental pedological map of Tanganyika. 2nd Inter-African Soils Conference 237-240.

Carpenter, A.J. 1978. The history of rice in Africa. In: I.W. Buddenhagen, and G.J. Persley, editors, Rice in Africa. Academic Press, New York. p. 3–10.

Carré, F., and A.B. McBratney. 2005. Digital terron mapping. Geoderma 128:340-353.

Carré, F., and M. Jacobson. 2009. Numerical classification of soil profile data using distance metrics. Geoderma 148:336–345.

Carre´, F., Girard, M.C. 2002. Quantitative mapping of soil types based on regression kriging of taxonomic distances with landform and land cover attributes. Geoderma 110:241–263.

Chen, Y., K. Khan, and Z. Padar. 2008. Irrigation intensification or extensification assessment: a GIS-based spatial fuzzy multi-criteria evaluation. In: Proceedings of the 8th international symposium on spatial accuracy assessment in natural resources and environmental sciences, 25–27 June, Shanghai, P.R. China. p. 309–318.

Cherif Ahmed, A. O., R. Nagasawa, K. Hattori, D. Chongo, and M.F. Perveen. 2007. Analytical Hierarchic Process in conjunction with GIS for identification of suitable sites for water harvesting in the oasis areas: Case study of the Oasis zone of Adrar, Northern Mauritania. Journal of Applied Sciences 7(19):2911-2917.

Cimmery, V. 2007. User guide for SAGA (version 2.0). Vern Cimmery, Hamburg.

Cook, S.E., A. Jarvis and J.P. Gonzalez. 2008. A new global demand for digital soil information. In: A.A. Hartemink, A. McBratney, and M.L. Mendonca-Santos, editors, Digital soil mapping with limited data. Springer. p. 31-41.

D'Hoore, J.L. 1964. Soil Map of Africa. Commission for Technical Cooperation in Africa, Lagos, Publication no. 23.

De Data, S.K. 1989. Rice. In: D.L. Plucknett, and H.B. Sprague, editors, Detecting mineral nutrient deficiencies in tropical and temperate crops. Westview Press Inc. Boulder, CO. p. 41-51.

De Pauw, E. 1984. Soil Physiography and Agro-ecological Zones of Tanzania. Consultants Final Report for Crop Monitoring and Early Warning Systems Project. KILIMO/FAO.

Deckers, J., O. Spaargaren, and S. Dondeyne. 2002. Soil survey as a basis for land evaluation. Encyclopedia of Life Support Systems.

Department for Communities and Local Government. 2009. Multi-criteria analysis: a manual. Communities and Local Government Publications, West Yorkshire, UK.

Diday, E. 1971. Une nouvelle méthode en classification automatique et reconnaissance des formes: la méthode des nuées dynamiques. Revue de Statistique Applique. 19:283-300. In French.

Dobermann, A., and T. Fairhurst. 2000. Rice: nutrient disorders & nutrient management. Oxford Graphics Printers Pte Ltd. Singapore.

Egbuchua, C.N. 2011. A pedological characterization of soils of the River Niger alluvial complex of Nigeria. Nigerian Journal of Soil and Environmental Research 9:65-70.

Elaalem, M., A. Comber, and P. Fisher. 2011. A Comparison of fuzzy AHP and Ideal Point Methods for evaluating land suitability. Transactions in GIS 15(3):329–346.

Elsheikh, R., A. Rashid, B.M. Shariff, F. Amiri, N.B. Ahmad, S.K. Balasundram, and M.A.M. Soom. 2013. Agriculture Land Suitability Evaluator (ALSE): A decision and planning support tool for tropical and subtropical crops. Computers and Electronics in Agriculture 93:98–110.

Eschweiler, J.A. 1998. SOTER database – Tanzania. FAO, Rome.

ESRI. 2002. What's New in ArcView 3.1, 3.2, and 3.3. Environmental Systems Research Institute, Inc., Redlands, California.

ESRI. 2010. ArcGIS - A Complete Integrated System. Environmental Systems Research Institute, Inc., Redlands, California.

FAO 1976. A framework for land evaluation. Soils Bulletin 3, Rome.

FAO 1985. Guidelines: Land Evaluation for Irrigated Agriculture. FAO Soils Bulletin 55. Food and Agriculture Organization of the United Nations, Rome.

FAO 1995. Planning for sustainable use of land resources; Towards a new approach. Land and Water Bulletin 2, Rome.

FAO. 1961. The Rufiji Basin Tanganyika. FAO Exp. Techn. Ass. Progr. No. 1269 Vol. 7. Rome.

FAO. 1976. A framework for land evaluation. Food and Agriculture Organization of the United Nations, Soils Bulletin 32. FAO, Rome.

FAO. 2004. International year of rice 2004: rice is life. FAO, Rome

FAO. 2006. Guidelines for Soil Description. FAO, Rome.

Fridland, V.M. 1974. Structure of the soil mantle. Geoderma 12:35-41.

Gallant, J.C., and J.P. Wilson. 2000. Primary topographic attributes. In: J. P. Wilson and J. C. Gallant, editors, Terrain analysis: Principles and applications. John Wiley & Sons, New York. p. 51-86

García, J.L., A. Alvarado, J. Blanco, E. Jiménez, A.A. Maldonado, and G. Cortés. 2014. Multi-attribute evaluation and selection of sites for agricultural product warehouses based on an Analytic Hierarchy Process. Computers and Electronics in Agriculture 100:60–69.

Gee, G. W., and J.W. Bauder. 1986. Particle-size analysis. In: A. Klute, editor, Methods of soil analysis: Part 1—physical and mineralogical methods. SSSA Book Series 5.1 Soil Science Society of America, American Society of Agronomy. p. 383–411.

Geological Survey of Tanganyika. 1963. Chombe: Quarter degree sheet No. 234. Geological Survey Division, Dodoma.

Gervais-Beaulac, V., D. Saint-Laurent, J.S. Berthelot, and M. Mesfioui. 2013. Organic carbon distribution in alluvial soils according to different flood risk zones. Journal of Soil Science and Environmental Management 4(8):169-177.

Gholap, J., A. Ingole, J. Gohil, S. Gargade and V. Attar. 2012. Soil data analysis using classification techniques and soil attribute prediction. International Journal of Computer Science Issues 9(3):415-418.

Giles, P.T., and S.E. Franklin. 1996. Comparison of derivative topographic surfaces of a DEM generated from stereoscopic SPOT images with field measurements. Photogrammetric Engineering & Remote Sensing 62(10):1165-1171.

Gill, A. 2003. Ancient Egyptians: The Kingdom of the Pharaohs brought to Life. Harper Collins Entertainment, London, United Kingdom.

Gitelson, A.A., Y.J. Kaufman, and M.N. Merzlyak. 1999. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sensing of Environment 58:289–298.

Goepel, K.D. 2014. BPMSG AHP priority calculator [Version 2014-02-09 (c)]. Rational decision making made easy. http://bpmsg.com/academic/ahp.php. Visited: December, 2014.

Haifa. 2008. Nutritional recommendations for rice. Haifa, Israel.

Hansen, M.K., D.J. Brown, P.E. Dennison, S.A Graves, and R.S. Bricklemyer. 2009. Inductively mapping expert-derived soil-landscape units within dampo wetland catena using multispectral and topographic data. Geoderma 150 (1-2):72-84.

Haring, T., E. Dietz, S. Osenstetter, T. Koschitzki, and B Schroder. 2012. Spatial disaggregation of complex soil map units: a decision tree based approach in Bavarian forest soils. Geoderma 185-186:37-47.

Hathout, S.A. 1983. Soil Atlas of Tanzania. Tanzania Publishing House, Dar es Salaam.

Hazelton, P., and B. Murphy. 2007. Interpreting soil test results. CSIRO Publishing, Collingwood Victoria, Australia.

He, Y. 2013. mColorConverter. Apple Inc. CA, USA.

Hengl, T., Toomanian, N., Reuter, H.I., Malakouti, M.J., 2007. Methods to interpolate soil categorical variables from profile observations: lessons from Iran. Geoderma 140:417–427.

Heuvelink, G.B.M., and R. Webster. 2001. Modelling soil variation: past, present, and future. Geoderma 100, 269–301.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965–1978.

Hole, F.D., and M. Hironaka. 1960. An experiment in ordination of some soil profiles. Soil Science Society of America Proceedings 24: 309-312.

Holmes, R.S. and W.E. Hearn. 1942. Chemical and physical properties of some of the important alluvial soils of the Mississippi drainage basin. Technical bulletin No, 833. United States Department of Agriculture, Washington, D. C.

Hoobler, B.M., G.F. Vance, J.D. Hamerlinck, L.C. Munn, and J.A. Hayward. 2003. Applications of land evaluation and site assessment (LESA) and a geographical information system in East Part Couny, Wyoming. Journal of Soil and Water Conservation 58:105–112.

Horn, B.K.P. 1981. Hill shading and the reflectance map. Proceedings of the IEEE 69(1):14-47.

Hoskins, B.R. 1997. Soil testing handbook for professionals in agriculture, horticulture, nutrient and residuals management. Third edition, University of Maine, Maine, USA.

Hwang C. L., and K. Yoon. 1981. Multiple attribute decision making methods and applications, Springer-Verlag, Berlin.

Indexmundi. 2014. United Republic of Tanzania, milled rice production by year. http://www.indexmundi.com/agriculture/?country=tz&commodity=milled-rice&graph=production. Last visited May 2014.

Iqbal, J., J.A. Thomasson, J.N. Jenkins, P.R. Owens, and F.D. Whisler. 2005. Spatial variability analysis of soil physical properties of alluvial soils. Soil Science Society of America Journal 69:1-14

IUSS Working Group WRB. 2014. World reference base for soil resources 2006. 2nd edition. World Soil Resources Reports No. 103. FAO, Rome.

Jacobson, M., and F. Carré. 2006. OSACA version 1.0. Land Management and Natural Hazards Unit, Institute for Environment and Sustainability. European Commission, Italy.

Jamer, T., J. Hill, H. Lavée, and P. Sarah. 2010. Mapping topsoil organic carbon in non agricultural semi-arid and arid ecosystems of Israel. Photogrammetric Engineering and Remote Sensing 76(1): 85-94.

Jenny, H., 1941. Factors of soil formation: a system of quantitative pedology. McGraw-Hill, New York.

Jones A., V. Stolbovoy, E. Rusco, A.R. Gentile, C. Gardi, B. Marechal, and L. Montanarella. 2009. Climate change in Europe. 2. Impact on soil. a review. Agronomy for Sustainable Development 29:423–432.

Kalogirou, S. 2002. Expert systems and GIS: an application of land suitability evaluation. Computers, Environment and Urban Systems 26:89–120.

Kato, F. 2007. Development of a major rice cultivation area in the Kilombero Valley, Tanzania. African Study Monographs, Supplement 36: 3-18.

Kauzeni, A.S., I.S. Kikula, S.A. Mohamed, J.G. Lyimo, and D.B. Dalal-Clayton. 1993. Land use planning and resource assessment in Tanzania: a case study. IIED Environmental Planning Issues No. 3 IRA Research Paper No. 35.

Kemker, C. 2014. Sediment transport and deposition: Fundamentals of environmental measurements. Fondriest Environmental, Inc. http://www.fondriest.com/environmental-measurements/parameters/hydrology/sediment-transport-deposition/. Last visited November, 2014.

Kempen, B., D.J. Brus, G.B.M. Heuvelink, and J.J. Stoorvogel. 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. Geoderma 151:311–326.

Kerry, R., P. Goovaerts, B.B. Rawlings, and B.P. Marchant. 2012. Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. Geoderma 170:347-358.

Kihoro, J., N.J. Bosco, and H. Murage. 2013. Suitability analysis for rice growing sites using a multicriteria evaluation and GIS approach in Great Mwea region, Kenya. SpringerPlus 2013, 2:265. http://www.springerplus.com/content/2/1/265.

Kimaro, D.N., B.M. Msanya, S. ArakI, and H. Otsuka. 2001. Application of computerized land evaluation systems in Tanzania: a case study in Kilosa District, Morogoro Region. UNISWA Journal of Agriculture 10:40-50.

Kiss, J.J., de Jong, E., Martz, L.W., 1988. The distribution of fallout Cesium-137 in southern Saskatchewan, Canada. Journal of Environmental Quality 17:445–452.

Kohnke, H. 1968. Soil Physiscs. McGraw Hill Book Company, New York.

Kuria, D., D. Ngari, and E. Waithaka. 2011. Using geographic information systems (GIS) to determine land suitability for rice crop growing in the Tana delta. Journal of Geography and Regional Planning 4(9):525-532.

Kuria, D., D. Ngari, and E. Waithaka. 2011. Using geographic information systems (GIS) to determine land suitability for rice crop growing in the Tana delta. Journal of Geography and Regional Planning 4(9):525-532.

Lagacherie, P., A.B. McBratney, and M. Voltz. 2007. Digital soil mapping: An introductory perspective. Elsevier, Amsterdam.

Lal, R. and M.K. Shukla 2004, Principles of Soil Physics. Marcel Dekker, New York.

Landon, J.R. (1991). Booker tropical soil manual. A handbook for soil survey and agricultural land evaluation in the tropics and subtropics. Longman Scientific & Technical Publishers. Essex.

Lilburne, L., A. Hewitt, P. McIntosh, and I. Lynn. 1998. GIS-driven models of soil properties in the high country of the south island. 10th Colloquium of the Spatial Information Research Centre, University of Otago, New Zealand, p. 173–180.

Lin, G.C.S. 1997. Red capitalism in South China: Growth and development of the Pearl River Delta. UBC Press, Vancouver, Canada.

Lindsay, W.L,. and W.A. Norvell. 1978. Development of a DTPA soil test for zinc, iron, manganese, and copper. Soil Science Society of America Journal 42:421-428.

Loxton, R.P. 1953. Kilombero land use survey - Block A. Unpublished report to the Tanganyika Government.

Loxton, R.P. 1954. Kilombero land use survey- Blocks B and C. Unpublished report to the Tanganyika Government.

Ly, P., L.S. Jensen, T.B. Bruun, D. Rutz, and A. de Neergaard. 2012. The system of rice intensification: Adapted practices, reported outcomes and their relevance in Cambodia. Agricultural Systems 113:16–27.

Maddahi, Z., A. Jalalian, M.M.K. Zarkesh, and N. Honarjo. 2014. Land suitability analysis for rice cultivation using multi criteria evaluation approach and GIS. European Journal of Experimental Biology 4(3):639-648.

Malczewski, J. 1999. GIS and Multicriteria Decision Analysis. John Wiley and Sons, New York.

Malczewski, J. 2004. GIS-based land-use suitability analysis: a critical Overview. Progress in Planning 62:3–65.

Malone, B. 2013. Use of R for digital soil mapping. Soil Security Laboratory, University of Sydney, Australia.

Malone, B.P., McBratney, A.B., Minasny, B., and G.M. Laslett. 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154:138–152.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Byers. 2011. Big data: The next frontier for innovation, competition, and productivity. Technical report, McKinsey Global Institute. San Francisco, CA.

Marinoni, O., and A. Hoppe. 2006. Using the analytical hierarchy process to support sustainable use of geo-resources in metropolitan areas. Journal of Systems Science and Systems Engineering 15(2):154-164.

Massawe, B.H.J., and N.A. Amury. 2012. Report on agronomic practices and soil fertility analysis for improved rice production in Kilombero and Wami Valleys in Tanzania. USAID Feed the Future NAFAKA staples value chain activity project. Morogoro, Tanzania.

McBratney, A., B. Minasny, and R.V. Rossel. 2010. Numerical soil classification: A missed, but not a lost, opportunity. 19th World Congress of Soil Science, Soil Solutions for a Changing World, 1 – 6 August, Brisbane, Australia.

McBratney, A.B., and I.O.A. Odeh. 1997. Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma 77:85-113.

McBratney, A.B., I.O.A. Odeh, T.F.A. Bishop, M.S. Dunbar, and T.M. Shatar. 2000. An overview of pedometric techniques for use in soil survey. Geoderma 97:293–327.

McBratney, A.B., M.L. Mendonca Santos, and B. Minasny 2003. On digital soil mapping. Geoderma 117:3–52.

McLean, E.O. 1986. Soil pH and lime requirement. In: A.L. Page, R.H.Miller, and D.R. Keeney, editors, Methods of soil analysis, part 2; chemical and mineralogical properties, 2nd edition. American Society of Agronomy, Madison. p. 199-223.

Meertens, H.C.C., L.J. Ndege, and P.M. Lupeja. 1999. The cultivation of rainfed, lowland rice in Sukumaland, Tanzania. Agriculture, Ecosystems and Environment 76:31–45.

Mendas, A., and A. Delali. 2012. Integration of MultiCriteria Decision Analysis in GIS to develop land suitability for agriculture: application to durum wheat cultivation in the region of Mleta in Algeria. Computers and Electronics in Agriculture 83:117–126.

Messing, I., M.H. Fagerstrom, L. Chen, and B. Fu. 2003. Criteria for land suitability evaluation in a small catchment on the Loess Plateau in China. Catena 54:215–234.

Milne, G. 1936. A Reconnaissance journey through parts of Tanganyika Territory. Journal of Ecology 35:192-265.

Minasny, B., and A.B. McBratney. 2007. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. Geoderma 142:285–293.

Minitab. 2004. Meet Minitab: Minitab user guide. Minitab Inc., USA.

Mitchell, T.D. and P.D. Jones. 2005. An improved method of construction a database of monthly climate observation and associated high-resolution grid. International Journal of Climatology 25:693-712.

Moore, I. and G. Burch. 1986. Physical basis of the length-slope factor in the Universal Soil Loss Equation. Soil Science Society of America Journal 50:1294–1298.

Moore, I.D., P.E. Gessler, G.A. Nielsen, and G.A. Peterson. 1993. Soil attributes prediction using terrain analysis. Soil Science Society of America Journal 57: 443–452.

Msanya, B.M., D.N. Kimaro, E.P. Kileo, G.G. Kimbi, and S.B. Mwango. 2001. Land suitability evaluation for the production of food crops and extensive grazing: a case study of Wami Plains in Morogoro Rural District, Tanzania. Soils and Land Resources of Morogoro Rural and Urban Districts, Volume 1. Department of Soil Science, Faculty of Agriculture, Sokoine University of Agriculture, Morogoro, Tanzania.

Msanya, B.M., J.P. Magoggo, and H. Otsuka. 2002. Development of soil surveys in Tanzania. Pedologist 46(1):79-88.

Mueller, L., U. Schindler, W. Mirschel, T.G. Shepherd, B.C. Ball, K. Helming, J. Rogasik, F. Eulenstein, and H. Wiggering. 2010. Assessing the productivity function of soils. A review. Agronomy for Sustainable Development 30:601–614.

Muir, J.W., H.G.M. Hardie,R.H.E. Inkson, and A.J.B. Anderson. 1970. The classification of soil profiles by traditional and numerical methods. Geordema 4: 81 – 90.

Munsell Color Company. 1992. Munsell soil color charts. Munsell Color Co. Inc. Baltimore.

Nelson, D.W., and L.E. Sommers. 1982. Total carbon, organic carbon and organic matter. In: A.L. Page, R.H.Miller, and D.R. Keeney, editors, Methods of soil analysis, part 2; chemical and mineralogical properties, 2nd edition. American Society of Agronomy, Wisconsin. p. 539-579.

Nwer, B.A.B. 2006. The Application of land evaluation technique in the north-east of Libya. Faculty of Environment. Cranfield University, Silsoe.

Odeh, I.O.A., A.B. McBratney, and D.J. Chittleborough. 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67:15–225.

Odgers, N.P., A.B. McBratney, and B. Minasny. 2011a. Bottom-up digital soil mapping I. Soil layer classes. Geoderma 163:38–44.

Odgers, N.P., A.B. McBratney, and B. Minasny. 2011b. Bottom-up digital soil mapping II. Soil series classes. Geoderma 163:30–37.

Odgers, N.P., W. Sun, A.B. McBratney, B. Minasny, and D. Clifford. 2014. Disaggregating soil map units through resampled classification trees. Geoderma 214:91-100.

Odgers, N.P., Z. Libohova, and J.A. Thompson. 2012. Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. Geoderma 189–190:153–163.

Pachepsky, Y.A., D.J. Timlin, and W.J. Rawls. 2001. Soil water retention as related to topographic variables. Soil Science Society of America Journal 65:1787–1795.

Pan, G. and J. Pan. 2012. Research in crop land suitability analysis based on GIS. Computer and Computing Technologies in Agriculture 365:314–325.

Park, S.J., and L.G. Vlek. 2002. Prediction of three-dimensional soil spatial variability: a comparison of three environmental correlation techniques. Geoderma 109:117–140.

Perveen, F., R. Nagasawa, A.O.C. Ahmed, I. Uddin, and R. Kimura. 2008. Integrating biophysical and socio-economic data using GIS for land evaluation of wheat cultivation: A case study in north-west Bangladesh. Journal of Food, Agriculture and Environment 6(2):432-437.

Planchon, O. and F. Darboux. 2001. A fast, simple and versatile algorithm to fill the depressions of digital elevation models. Catena 46:159–176.

Prasad, A.M., L.R. Iverson and A. Liaw. 2006. Newer classification and regression tree techniques: Bagging and Random Forest for ecological prediction. Ecosystems 9:181-199.

QGIS Development Team. 2014. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Quinlan, J.R. 1993. C4.5: Programs for machine learning. Morgan Kaufmann, San Mateo.

Rahman, R., and S. K. Saha. 2008. Remote sensing, spatial multi criteria evaluation (SMCE) and analytical hierarchy process (AHP) in optimal cropping pattern planning for a flood prone area. Journal of Spatial Science 53(2):161–177.

Ramesh, V., and K. Ramar. 2011. Classification of agricultural land soils: a data mining approach. Agricultural Journal 6(3):82-86.

Rayner, J.H. 1966. Classification of soils by numerical methods. European Journal of Soil Science 17:79-92.

Rhoades, J.D. 1982. Soluble salts. In: A.L. Page, R.H.Miller, and D.R. Keeney, editors, Methods of soil analysis. Part 2, chemical and mineralogical properties, 2nd edition. American Society of Agronomy. Wisconsin. p. 149-158; 167-180.

Rizzo, R., J.A. M. Demattê, and F.S. Terra. 2014. Using numerical classification of profiles based on VIS-NIR spectra to distinguish soils from the Piracicaba region, Brazil. Revista Brasileira de Ciência do Solo 38:372-385.

Rondeaux, G., M. Steven and F. Baret. 1996. Optimization of Soil Adjusted Vegetation Indices. Remote Sensing of Environment 55:95-107.

Rosa, D.D.L., M. Anaya-Romero, E. Diaz-Pereira, N. Heredia, and F. Shahbazi. 2009. Soil specific agro-ecological strategies for sustainable land use. A case study by using MicroLEIS DSS in Sevilla province. Land Use Policy 6:1055–1065.

Rossiter, D.G. and A.R.V. Wambeke. 1997. Automated Land Evaluation System ALES Version 4.65 User’s Manual. Cornell University.

Rural Livelihood Development Company-RLDC. 2011. Rice sector strategy. Dodoma, Tanzania.

Saaty T. L. 1988. The Analytic Hierarchy Process. University of Pittsburgh, Pennsylvania.

Saaty, T.L. 1980. The analytic hierarchy process. McGraw-Hill, New York.

Saaty, T.L. 2008. Decision making with the analytic hierarchy process. International Journal of Services Sciences 1: 83–97.

Saaty, T.L., and L.G. Vargas. 1991. Prediction, projection and forecasting. Kluwer Academic Publishers, Dordrecht.

Sahebjalal, E. and K. Dashtekian. 2013. Analysis of land use-land covers changes using normalized difference vegetation index (NDVI) differencing and classification methods. African Journal of Research Agricultural 8(37):4614-4622.

Sahrawat, K.L. 2012. Soil fertility in flooded and non-flooded irrigated rice systems. Archives of Agronomy and Soil Science 58(4):423-436.

Samanta, S., B. Pal, and D.K. Pal. 2011. Land suitability analysis for rice cultivation based on multi-criteria decision approach through GIS. International Journal of Science and Emerging Technologies 2(1):12-20.

Samki, J.K. 1977. Provisional Soils Map of Tanzania. Survey and Mapping Division, Dar es Salaam.

Samki, J.K. 1982. Soils Map of Tanzania. National Soil Service. Tanga, Tanzania.

Samki, J.K., and J.F. Harrop. 1984. Fertilizer recommendations in Tanzania on a district by district basis. NSS ARI Mlingano, Tanzania.

Sanchez, P.A., S. Ahamed, F. Carré, A.E. Hartemink, J. Hempel, J. Huising, P. Lagacherie, A.B. McBratney, N.J. McKenzie, M.L. Mendonca-Santos, B. Minasny, L. Montanarella, P. Okoth, C.A. Palm, J.D. Sachs, K.D. Shepherd, T.G. Vagen, B. Vanlauwe, M.G. Walsh, L.A. Winowiecki, and G.L. Zhang. 2009. Digital soil map of the world. Science 325:680–681.

Schaetzl R.J., and S. Anderson. 2005. Soils: genesis and geomorphology. Cambridge University Press, Cambridge, UK.

Schaetzl, R., and S. Anderson. 2010. Soil genesis and geomorphology. Cambridge University Press, Cambridge, UK.

Schoeneberger, P.J., D.A. Wysocki, E.C. Benham, and Soil Survey Staff. 2012. Field book for describing and sampling soils, Version 3.0. Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE.

Schollenberger, C.J., and R.H. Simon. 1945. Determination of exchange capacity and exchangeable bases in soils-ammonium acetate method. Soil Science 59:13-24.

Scull, P., J. Franklin, O.A. Chadwick and D. McArthur. 2003. Predictive soil mapping: A review. Progress in Physical Geography 27:171-197.

Shaw, R.J., K.J. Coughlan, and L.C. Bell. 1998. Root zone sodicity. In; Eds M.E. Sumner and R. Naidu. Sodic soils: Distribution, properties, management, and environmental consequences. Oxford University Press, New York. p. 95–106.

Soil Survey Division Staff. 1993. Soil survey. Soil Conservation Service, U.S. Department of Agriculture Handbook 18.

Soil Survey Staff (2014), Keys to soil taxonomy. 12th Edition. USDA, Natural Resource Conservation Service, USA.

Southard, J. 2006. Introduction to fluid motions, sediment transport, and current-generated sedimentary structures. MIT Open Courseware: Massachusetts Institute of Technology. USA.

Stum, A.K., Boettinger, J.L., White, M.A., Ramsey, R.D., 2010. Random forests applied as a soil spatial predictive model in arid utah, in: Boettinger, J.L., D.W. Howell, A.C. Moore, A.E. Hartemink, and S. Kienast-Brown (Eds.), Digital Soil Mapping: Bridging Research, Environmental Application, and Operation, Springer, Dordrecht, 179–190.

Subburayalu, S.K. and B.K. Slater. 2013. Soil series mapping by knowledge discovery from an Ohio county soil map. Soil Science Society of America Journal 77:1254–1268.

Subburayalu, S.K., I. Jenhani, and B.K. Slater. 2014. Disaggregation of component soil series on an Ohio County soil survey map using possibilistic decision trees. Geoderma 213:334–345.

Survey and Mapping Division. 1983. Mbingo 1:50000 Toposheet No. 234/2. Ministry of Lands, Housing and Urban Development, Dar es Salaam, Tanzania.

Survey and Mapping Division. 1983. Mngeta 1:50000 Toposheet No. 234/3. Ministry of Lands, Housing and Urban Development, Dar es Salaam, Tanzania.

Survey and Mapping Division. 1983. Njagi 1:50000 Toposheet No. 234/1. Ministry of Lands, Housing and Urban Development, Dar es Salaam, Tanzania.

Sweeney, M.R., E.V. McDonald, and C.E. Markley. 2013. Alluvial sediment or playas: what is the dominant source of sand and silt in desert soil vesicular A horizons, southwest USA. Journal of Geophysical Research: Earth Surface 118:1–19.

Tarboton, D.G. 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resources Research 33(2):309–319.

Tesfa, T.K., D.G. Tarboton, D.G. Chandler, and J.P. McNamara. 2009. Modeling soil depth from topographic and land cover attributes. Water Resource Research Volume 45, W10438. doi:10.1029/2008WR007474.

Thien, S.J. 1979. A flow diagram for teaching texture by feel analysis. Journal of Agronomic Education 8:54-55.

Thomas, G.W. 1982. Exchangeable cations. In: A.L. Page, R.H.Miller, and D.R. Keeney, editors, Methods of soil analysis. Part 2, chemical and mineralogical properties, 2nd edition. American Society of Agronomy. Wisconsin. p. 595-624.

TIC. 2013. Tanzania investment guide 2013-14. Tanzania Investment Center, Dar es Salaam.

Trung, N.H., L.Q. Tri, M.E.F. van Mensvoort, and A.K. Bregt. 2006. Application of GIS in land-use planning: a case study in the Coastal Mekong Delta of Vietnam.In: Proceedings of the International Symposium on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences, Ho Chi Minh, Vietnam, 1-11 November. p. 6.

Turetta, A.P.D. M.L.Mendonca-Santos, L.H.C. Anjos, and R.L.L. Berbara. 2008. Spatial-temporal changes in land cover, soil properties and carbon stocks in Rio de Janeiro. In: A.A. Hartemink, A. McBratney, and M.L. Mendonca-Santos, editors, Digital soil mapping with limited data. Springer. p. 261-271.

USGS. 2000. Shuttle Radar Topography Mission, 1 Arc Second scene SRTM. http://earthexplorer.usgs.gov/ page visited 24 October, 2014.

Viscarra-Rossel, R.A., McBratney, A.B., 2008. Diffuse reflectance spectroscopy as a tool for digital soil mapping. In: A.A. Hartemink, A. McBratney, and M.L. Mendonca-Santos, editors, Digital soil mapping with limited data. Springer. p. 165-172.

Vogel, R. 2008. A software framework for GIS-based multiple criteria evaluation of land suitability. In Proceedings of the Eleventh AGILE International Conference on Geographic Information Science, Girona, Spain: p. 1–32.

Watanabe, F.S., and S.R. Olsen. 1965. Test of ascorbic acid method for determining phosphorus in water and NaHCO3 extracts from soil. Soil Science Society of America 29:677-678.

Wickama, J.M.W. 1997. Pedological investigation and characterization in Kitanda Village, Mbinga District, Tanzania. MSc Thesis, Sokoine University of Agriculture, Morogoro, Tanzania.

Wilson, J.P., and J.C. Gallant 2000. Terrain analysis: Principles and applications. John Wiley & Sons Inc., New York.

Witten, I.H., and E. Frank. 2005. Data mining: Practical machine learning tools and techniques. Second edition. Elsevier Inc., San Francisco, CA.

Witten, I.H., E. Frank, and M.A. Hall. 2011. Data mining: Practical machine learning tools and echniques. Morgan Kaufmann, Burlington.

Wood, S.R., and F.J. Dent. 1983. LECS. A land evaluation computer system methodology. Bogor: Ministry of Agriculture/PNUD/FAO, Centre for Soil Research, Indonesia.

Xu, L., and J. Yang. 2001. Introduction to multi-criteria decision making and the evidential reasoning approach. Working paper No. 0106. Manchester School of Management, Manchester, UK.

Yang, L., Y. Jiao, S. Fahmy, A.X. Zhu, S. Hann, J.E. Burt, and F. Qi. 2011. Updating conventional soil maps through digital soil mapping. Soil Science Society of America Journal 75:1044–1053.

You, D., L. Hug, and D. Anthony. 2014. Generation 2030 – Africa: Child demographics in Africa. UNICEF, New York.

Zheng, D., Hunt, E.R., Running, S.W., 1996. Comparison of available soil water capacity estimated from topography and soil series information. Landscape Ecology 11:3–14.

Zhu, A. and Band, L. 1994: A knowledge-based approach to data integration for soil mapping. Canadian Journal of Remote Sensing 20:408–18.

Zhu, A.X., B. Hudson, J. Burt, K. Lubich, and D. Simonson. 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Science Society of America Journal 65:1463–1472.

Zopounidis, C., and M. Doumpos. 2002. Multicriteria classification and sorting methods: a literature review. European Journal of Operational Research 138:229-246.