UTILIZATION OF LEGACY SOIL DATA FOR DIGITAL SOIL MAPPING AND DATA DELIVERY FOR THE BUSIA AREA, KENYA

by

Joshua O. Minai

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Agronomy

West Lafayette, Indiana

December 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Darrell G. Schulze, Chair

Department of Agronomy

Dr. James J. Camberato

Department of Agronomy

Dr. Jie J. Shan

Department of Civil Engineering

Dr. Margaret W. Gitau

Department of Agricultural and Biological Engineering

Dr. Zamir Libohova

USDA-NRCS-National Soil Survey Center

Approved by:

Dr. Linda S. Lee

Program Head for the Ecological Sciences & Engineering Interdisciplinary Graduate Program

&

Dr. Ronald F. Turco

Department Head of Agronomy

ACKNOWLEDGMENTS

To my mentor, Dr. Joan Cuka Kagwanja, I owe it all to you. Many Thanks! Dr. Joan, I will never forget your efforts in 2009 when you challenged me to go to graduate school. It is through your efforts that this dream has come to pass. I vividly recall you allowing me to attend all the conferences, forums, and meetings that the Alliance for a Green Revolution in Africa (AGRA) was hosting so that I can realize that agriculture is not just taking a hoe and plowing the land. It is, however, a complex system where one has to understand the complexities of the value chains within the sector.

I am forever grateful to Prof. Gebisa Ejeta for making me consider Purdue University as an ideal university to undertake my graduate studies. From my first encounter with you during the second African Green Revolution Forum in Accra, Ghana, I was convinced that Purdue would be ideal for me even though I had never heard of it. I will be forever indebted to you because the knowledge, opportunities, and experiences that I have gained and shared with amazing people is something I will forever cherish.

I would like to express my sincere gratitude to my advisor Prof. Darrell G. Schulze for the continuous support of my Ph.D. study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this dissertation. I could not have imagined having a better advisor and mentor for my Ph.D. study. Words cannot express how lucky I am to have you as an advisor. The one thing I have learnt from you is having the patience and passion for research to be able to understand some of the most pressing agronomic challenges.

Besides my advisor, I would like to thank the rest of my research committee: Dr. Zamir Libohova and Professors Jie Shan, James Camberato, and Margaret Gitau for their insightful comments and encouragement, but also for the hard questions, which incented me to widen my research from various perspectives.

My sincere thanks also go to Dan and Darlene Childs who have always been there for me whenever I needed them to. I will never forget the pieces of advice both of you gave me, and still do, about life in general. Your support during my time at Purdue University will forever be imprinted in my happy memories. I can never ask for a better support family.

To Kossuth Street Baptist Church, May the Lord Almighty be there for each and every member of that church. The spiritual support has been a blessing to my life and I will forever be grateful.

I also wish to dedicate this work to my dad and mum, James Minai and Margaret Minai, who went through a lot to see me get the best education that I could ever ask for. Special thanks go to Evelyne Pambah and family. I will forever be grateful for the support that you gave me while undertaking my PhD research. To my sisters, aunts, uncles, and cousins, thank you all for being my rock and support throughout my education life. To aunties Rose, Jane, and Seline, and uncle Maurice, I will forever be grateful for your support during my high school education.

This work would not have been possible without the funding from the Purdue Center for Global Food Security. Sincere thanks to the Kenya Soil Survey staff including Mr. Lucas Tanui and Mr. Geoffrey Macharia, department head, for providing existing legacy soil data. Lastly, I thank Mr. Naman Nyabinda whose assistance was crucial for the reconnaissance visits to the study area.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AfSIS	Africa Soil Information Service.
AfSP	Africa Soil Profile.
AIC	Akaike Information Criteria.
AMPATH	Academic Model Providing Access to Healthcare.
AoF	Availability of Foothold for Roots
AoM	Availability of Moisture
AoN	Availability of Nutrients
ASD	Average Size Delineation.
C.I.	Confidence Interval.
Ca	Calcium.
CEC	Cation Exchange Capacity.
CIAT	International Center for Tropical Agriculture.
cLHS	Conditioned Latin Hypercube Sampling.
DEM	Digital Elevation Model.
DSM	Digital Soil Mapping.
EAMD	East Africa Metrological Department.
ESDAC	European Soil Data Center.
ESP	Exchangeable Sodium Percent.
ESRI	Environmental Systems Research Institute.
EuDASM	European Digital Archive of Soil Maps.
FAO	Food and Agriculture Organization.
FH	Flooding Hazard
FURP	Fertilizer Use Recommendation Project.
GCPs	Geodetic Control Points.
GDAL	Geospatial Data and Abstraction Library.
GIS	Geographical Information Systems.
GLASOD	Global Assessment of Human-induced Soil Degradation.

GPS	Global Positioning system.
IIASA	International Institute for Applied Systems Analysis.
IMR	Index of Maximum Reduction.
ISRIC	International Soil Reference and Information Centre.
ISRIC-WISE	ISRIC- World Inventory of Soil Emission Potentials.
ISSS	International Society of Soil Science.
JRC	Joint Research Centre.
K	Potassium.
KALRO	Kenya Agricultural and Livestock Research Organization.
KSS	Kenya Soil Survey.
m	Meters.
MASS	Modern Applied Statistics with S package in R Studio software
Mg	Magnesium.
MLA	Minimum Legible Area.
MLD	Minimum Legible Delineation.
mm	Millimeters.
Mn	Manganese.
MRRTF	Multiresolution Ridgetop Flatness.
MRVBF	Multiresolution Valley Bottom Flatness.
N	Nitrogen.
Na	Sodium.
NARL	National Agricultural Research Laboratories.
NASA	National Aeronautics and Space Administration.
NDVI	Normalized Difference Vegetation Index.
NRCS	National Resource Conservation Service
NRM	Natural Resources Management.
OK	Ordinary Kriging.
OLI	Operational Land Imager.
OM	Organic Matter.

Ox	Availability of oxygen in the root zone
P	Phosphorus.
PC	Principal Component.
PCA	Principal Component Analysis.
PDF	Portable Document Format.
pH	Potential Hydrogen.
RGB	Red Green Blue.
RMSE	Root Mean Square Error.
SAGA	System Automated Geoscientific Analyses.
SE	Susceptibility to Erosion
SH	Salinity hazard
SMLR	Stepwise Multiple Linear Regression.
SMU	Soil Map Unit.
SMUs	Soil Map Units.
SMW	Soil Map of the World.
SOC	Soil Organic Carbon.
Sod	Hazard of Sodicity
SoLIM	Soil Land Inference Model.
SOTER	Soil and Terrain.
SPC	Possibilities for Seedbed Preparation and Cultivation
SRTM	Shuttle Radar Topography Mission.
SSURGO	Soil Survey Geographic Database.
TAs	Terrain Attributes.
Temp	Temperature
TIRS	Thermal Infrared Sensor.
TPI	Topographic Position Index.
TWI	Topographic Wetness Index.
UNEP	United Nations Environmental Program.
UNESCO	United Nations Educational, Scientific and Cultural Organization.

UNFPA	United Nations Population Fund.
USDA	United States Department of Agriculture.
USGS	United States Geological Survey.
VESPER	Variogram Estimation and Spatial Prediction Plus Error.
WGS	World Geographic System.
WoSIS	World Soil Information Service.
WOSSAC	World Soil Survey Archive and Catalogue.
WRB	World Reference Base for soil resources.
WSS	Web Soil Survey

ABSTRACT

Much legacy soils data and soils information lies idle in libraries and archives and is largely unused, especially in developing countries like Kenya. We demonstrated the usefulness of a stepwise approach to bring legacy soils data ‘back to life’ using the 1980 *Reconnaissance Soil Map of the Busia Area (quarter degree sheet No. 101)* in western Kenya as an example. Three studies were conducted by using agronomic information, field observations, and laboratory data available in the published soil survey report as inputs to several digital soil mapping techniques.

In the first study, the agronomic information in the survey report was interpreted to generate 10 land quality maps. The maps represented the ability of the land to perform specific agronomic functions. Nineteen crop suitability maps that were not previously available were also generated.

In the second study, a dataset of 76 profile points mined from the survey report was used as input to three spatial prediction models for soil organic carbon (SOC) and texture. The three predictions models were (i) ordinary kriging, (ii) stepwise multiple linear regression, and (iii) the Soil Land Inference Model (SoLIM). Statistically, ordinary kriging performed better than SoLIM and stepwise multiple linear regression in predicting SOC (RMSE = 0.02), clay (RMSE = 0.32), and silt (RMSE = 0.10), whereas stepwise multiple linear regression performed better than SoLIM and ordinary kriging for predicting sand content (RSME = 0.11). Ordinary kriging had the narrowest 95% confidence interval while stepwise multiple linear regression had, the widest. From a pedological standpoint, SoLIM conformed better to the soil forming factors model than ordinary kriging and had a narrower confidence interval compared to stepwise multiple linear regression.

In the third study, rules generated from the map legend and map unit descriptions were used to generate a soil class map. Information about soil distribution and parent material from the

map unit polygon descriptions were combined with six terrain attributes, to generate a disaggregated fuzzy soil class map. The terrain attributes were multiresolution ridgetop flatness (MRRTF), multiresolution valley bottom flatness (MRVBF), topographic wetness index (TWI), topographic position index (TPI), planform curvature, and profile curvature. The final result was a soil class map with a spatial resolution of 30 m, an overall accuracy of 58% and a Kappa coefficient of 0.54.

Motivated by the wealth of soil agronomic information generated by this study, we successfully tested the feasibility of delivering this information in rural western Kenya using the cell phone-based Soil Explorer app (https://soilexplorer.net/). This study demonstrates that legacy soil data can play a critical role in providing sustainable solutions to some of the most pressing agronomic challenges currently facing Kenya and most African countries.

PREAMBLE

“Would it not be a great satisfaction to the king to know at a designated moment every year the number of his subjects, in total and by region, with all the resources, wealth & poverty of each place; [the number] of his nobility and ecclesiastics of all kinds, of men of the robe, of Catholics and of those of the other religion, all separated according to the place of their residence? ......[Would it not be] a useful and necessary pleasure for him to be able, in his own office, to review in an hour's time the present and past condition of a great realm of which he is the head, and be able himself to know with certitude in what consists his grandeur, his wealth, and his strengths?”

--*Marquis de Vauban, proposing an annual census to Louis XIV in 1686*--

CHAPTER 1. INTRODUCTION

Soils are essential for achieving food security and have the potential to help mitigate the negative impacts of climate change. Sustainable soil management, in turn, can contribute to the production of more and healthier food. The care, restoration, enhancement, and conservation of soil should therefore become a major global priority (Africa Progress Panel, 2015).

Africa's current population of ~1 billion is projected to increase to ~6 billion by 2100 (Gerland et al., 2004). As much as 70% of the current population depends directly on agriculture. There will be a tremendous need to produce more food on an essentially finite soil resource base. This can only happen when there is an in-depth understanding of existing soil resources.

Although information on soils for much of Africa is sparse, Kenya, fortunately, has considerable soils information in the form of traditional soil maps (Kenya Soil Survey, 1980), soil survey reports (i.e. Rachilo and Michieka, 1991), soil survey manuals (Kenya Soil Survey Staff, 1987), land evaluation frameworks, soil profile descriptions, and farm management handbooks (Jaetzold and Schmidt, 1982). These types of soil information are collectively known as *legacy soil data* (Zinck, 1995). Legacy soil data have been widely used as meaningful sources of soil information to support soil conservation or as major components of national environmental monitoring programs (McBratney et al., 2003; Odeh et al., 2012).

In developing countries, legacy soil data often remains idle in libraries as artifacts accumulating dust. The demand for soil data, however, is soaring (Cook et al., 2008). The probability of such data being lost through disasters, be it natural, manmade, political, or simply inattention is very high (Rossiter, 2008). Our visits to the Kenya Soil Survey (KSS) in the spring of 2016 confirmed this. (i) Legacy soil data is left seemingly unused and stored on library shelves, (ii) some are in private collections of retired soil scientists, (iii) while some data are in digital formats and, used only internally. One potential solution for the dissemination of such geographical data is to take advantage of advances in geo-information technology to take this data out of libraries, and make it available via portable digital devices. In addition to making the information more accessible, this may also reduce the possibility of loss of the paper data and the information it contains.

Legacy soil data contains considerable agronomic information that can help revitalize agriculture in countries with poor soils spatial data infrastructures (Zinck, 1995). This information often includes the spatial distribution of soils, land quality, crop suitability, geo-located soil profile information, geology, and land use. This sort of data can be used as baselines for long-term studies that assess changes of soil functional properties (Bellamy et al., 2005) and to model temporal trends of soil quality and soil processes (Baxter et al., 2006). Moreover, it can be used as a primary input for digital soil mapping (DSM), especially for countries with sparse soil data infrastructures (McBratney et al., 2003). Methodologies on how to effectively transform and utilize renewed legacy soil data have become a new area of research in soil science (Hengl et al., 2007). Rossiter (2008) proposed a stepwise process for collating legacy soil data where digital soil mapping concepts can play an important role in legacy soil data rescue and renewal through the use of medium to high resolution digital elevation models and derived terrain attributes.

The most common form of legacy soil data is the traditional soil map. Traditional soil mapping involves field observations followed by manual drawing of soil delineations onto aerial photographs or topographic maps. In the case of Kenya, this process was used to produce the Exploratory Soil Map and Agro-climatic Zone Map of Kenya, 1980 at a scale of 1:1,000,000 (Sombroek et al., 1982). This scale is equivalent to a resolution of ~300 meters. Resolution here refers to the smallest cell size on the ground. In this case, 300 m resolution refers to a 300 by 300 m area (an area of 9 hectares). The African Soil Information Service (AfSIS) is working to improve the spatial resolution of African soil maps using geostatistical approaches and their work so far has yielded a resolution of 250 meters (an area of 6.25 hectares).

Maps at these resolutions do not provide sufficient detail to inform management and land use decisions needed by rural smallholder farmers whose average landholding is less than a hectare (FAO, 2017). In addition, the uncertainty of such predictive maps is high and is not adequate to support cost-benefit based decisions at local levels. A more nuanced approach is needed to improve on the resolution of traditional maps to provide soils information to support agronomic decision-making. A digital soil mapping approach as described by Ashteker, (2014) resulted in resolutions as fine as 5 by 5 meters, which has sufficient resolution for farmers to use for management decisions.

All this will be in vain if such information does not reach the individuals making agronomic decisions. In addition to adapting the DSM outputs to specific user needs, this study aims to demonstrate the delivery of the information to potential end users via an easy to use cell phone app. We have already tested a possible approach using the cell phone network in rural western Kenya (Minai et al., 2016). This is an important accomplishment as it opens up the possibility of delivering timely and useful agronomic information to end users at low cost compared to traditional agricultural extension services.

1.1 Hypotheses

The need for spatial soil information is growing, particularly in digital formats that can readily be incorporated into geographical information systems (GIS) and analyzed with other spatial data (Lagacherie and McBratney, 2006). DSM offers not only the opportunity to map soil properties, but it also provides an opportunity to upgrade traditional soil maps (McBratney et al., 2003). DSM is defined as "the creation and population of spatial soil information systems by numerical models inferring the spatial and temporal variations of soil types and soil properties from observations and knowledge and from related environmental variables" (Lagacherie, 2008).

In this dissertation, I attempt to test two hypotheses. (1) Soil properties can be predicted without any additional field work through the utilization of existing legacy soil data and (2) a traditional soil map can be downscaled by utilizing soil forming factors together with terrain attributes derived from a digital elevation model (DEM).

1.1.1 Hypothesis 1

The prediction of soil properties can be achieved without any additional field work through the utilization of existing legacy soil data.

This hypothesis will be tested by using only soil profile data mined from an existing soil survey of a selected portion of western Kenya. The rapid advances made in georeferencing of soil samples and the growing availability of earth remote sensing data provided new opportunities for predicting soil functional properties using legacy soil data. Soil functional properties are those properties related to a soil's capacity to support essential ecosystem services such as primary productivity, nutrient and water retention, and resistance to soil erosion. DSM of soil functional properties, especially in regions with poor spatial data infrastructures such as Africa, is important for planning sustainable agricultural intensification and natural resources management (NRM) for rural smallholder farmers.

For instance, over the past decade, the AfSIS project has compiled two georeferenced soil profiles/samples datasets: (1) the Africa Soil Profiles database (Leenaars et al., 2014a) which contains legacy soil profile data and, (2) the Sentinel Sites database (Vågen et al, 2010) which contains newly collected topsoil data. Together, these two databases consist of over 85,000 samples, representing one soil sample for every 274 km^2. Using these soil point observations and measurements, and an extensive collection of global (SoilGrids1km) and continental (Africa) environmental covariates, ISRIC - World Soil Information Service, in collaboration with The Earth Institute, Columbia University, the World Agroforestry Centre, Nairobi, and the International Center for Tropical Agriculture (CIAT), have produced predictions of an array of soil functional properties for the whole African continent at an initial spatial resolution of 1km (Hengl et al., 2014). These predictions have since been downscaled to 250 m raster grids for either two or six standard soil depths (Hengl et al., 2017).

These initiatives show that state of the art baseline soil property map products can be developed cost-efficiently using DSM based on legacy soil data and can be updated when additional, data become available (Leenaars et al., 2017). It is unclear, however, what prediction model(s) are best suited and perform best with very limited soil data, which is the case for soil profile data mined from legacy soil data. Moreover, as already mentioned, the current spatial resolutions of 250 m resulting from prediction models is too low to help with practical land management for rural smallholder farmers in Africa whose landholding are typically less than a hectare (Sanchez et al., 2009).

1.1.2 Hypothesis 2

A traditional soil map can be downscaled by utilizing soil forming factors together with terrain attributes derived from a DEM.

Legacy soil data in the form of soil maps from traditional soil surveys describe soils as they appear in the field. They arise from traditional soil mapping approaches that are rooted in the catena concept to infer where specific soil types occur within landscapes. The catena concept refers to the repeating pattern of lateral soil variation over a hillslope (Milne, 1935). On steeper slopes, erosion and runoff is faster, so the soils there are typically thinner and drier. Conversely, on shallower slopes at the top or bottom of a catena, soils are usually deeper, resulting in soil types and soil characteristics that are different from those on the adjacent steeper slopes.

Using this concept, soil mappers have traditionally studied soil patterns in small-localized areas. Once they understood these patterns, they depicted the patterns by drawing polygons around areas of similar soil properties. These patterns were then extrapolated to larger spatial areas to map soils with similar soil properties. The problem with the traditional approach of depicting individual soil map units as discrete polygons with definite boundaries is that it simplifies the complex continuous distribution of soil types across a landscape (Zhu et al., 2004; Odgers et al., 2014). Specifically, soil map units of legacy soil maps were generalized to fit the amount of information the soil mapper could interpret from the available base maps and field observations. To correct for this, an appropriate method of better integrating relief attributes into the soil mapping process is needed (Penizek and Boruvka, 2008).

One way of doing this is by the use of terrain attributes (McBratney et al., 2003) to tie soil types described in soil map units with their slope position. For example, alluvial plains are relief units characterized by relatively flat areas around watercourses that have a concave transition to the surrounding upland landscape. Typical relief attributes of alluvial plains include: (1) high values of contributing area, (2) low slopes, (3) no or concave curvature, and (4) high values of the compound topographic index (Park et al., 2001). Delineation of alluvial plains using these terrain attributes can therefore give a more precise location of where alluvial soils occur within the landscape. Odgers et al. (2014) used such an approach to disaggregate a 1:250,000-scale legacy soil polygon map from Central Queensland, Australia to a 30 m raster grid resolution.

Disaggregation of soil map units involves “downscaling of information to produce new information at a finer scale than the original source” (McBratney, 1998), with the aim of mapping constituent soil classes of soil map units individually (Thompson et al., 2010). The result is a rasterized prediction of the spatial distribution of soil classes after disaggregating the soil polygon map units. Bui and Moran (2001) disaggregated soil associations mapped at a reconnaissance scale (1:500,000 – 1:250,000) using the legend description and terrain attributes to allocate the soils described for a particular soil map unit onto their respective landscape positions. This added further detail to the existing soil map and was used to improve (increase) their scale. Soil map unit disaggregation is important because the original soil map units, even though useful, do not contain enough detail to assist farmers in making sound agronomic decisions. Moreover, mapping soil classes as opposed to soil map units is of importance because specific land uses can be associated with specific soil classes (Brungard et al., 2015) but not necessarily with broad soil map units.

1.2 Objectives

The overall objective of this study was to bring legacy soil data for a selected portion of Kenya ‘back to life’ using DSM techniques. The specific objectives include:

a) Transform the best available legacy soil survey of a selected portion of Kenya into a digital format.
b) Make spatial predictions of selected soil functional properties by using data mined from the legacy soil survey using DSM techniques.
c) Improve the spatial resolution of the legacy soil map of the study area-using DSM techniques.
d) Develop a prototype platform that could deliver spatially explicit soil and agricultural information for the area of the legacy soil survey on a smart phone or tablet.

1.3 Dissertation Outline

This dissertation is organized into six chapters. This chapter, Chapter 1, describes the lack of use of legacy soil data and justifies the need for the ‘rescue’ of this legacy soils data for its use for digital soil mapping. Chapter 2 provides a literature review of the use of legacy soil data, highlights the attempts by various institutions and initiatives to rescue both legacy soil profile data and legacy soil survey reports, and identifies existing gaps in the literature. Chapter 3 demonstrates the stepwise process of bringing legacy soil data ‘back to life’. Chapter 4 illustrates how to make spatial predictions of selected soil functional properties mined from legacy soil data using specific soil interpolation methods. Chapter 5 discusses a proposed methodology for disaggregating a traditional soil polygon map to produce new information at a finer scale than the original source with an aim of mapping constituent soil classes within soil map units. Chapter 6 illustrates how to deliver spatially explicit soils and agricultural information on mobile devices to the end user. Chapter 7 summarizes the conclusions from the different chapters and provides recommendations for future study.

CHAPTER 2. LITERATURE REVIEW

This chapter reviews literature on legacy soil data highlighting some of the efforts used to rescue different types of legacy soil data. It also gives a history of soil mapping in Kenya and identifies some of the current efforts being made to map soil properties.

2.1 Legacy Soil Data

The world is full of soil resources produced over the past several decades (Arrouays et al., 2017). These exist as traditional soil maps, soil survey reports, soil survey manuals, land evaluation frameworks, soil profile descriptions, and farm management handbooks, collectively known as *legacy soil data* (Zinck, 1995).

Legacy soil data remains the backbone for present and future studies (Panagos et al., 2011). For instance, soil maps are resources used in a myriad of fields of science; they are among primary resources of soil resources used to: (i) monitor land degradation, improvement, and management, (ii) identify changes in land use and water resources, and (ii) predict climatic and environmental changes. The collection of new soil information, however, is costly. Fewer and fewer new field-based soil data are being collected and analyzed, and as a result, older data and information is increasingly sought after. Therefore, the preservation of legacy soil data is vital as they are the building blocks of most current studies requiring soil information. Ideally, users should have easy public access to the source material and available derived information.

Unfortunately, data, information and knowledge of the world soil resources are currently fragmented and even at risk of being lost or forgotten due to the cost of maintaining paper-based soil data holdings and archives, and the physical deterioration of these paper-based sources, especially in tropical climates. In addition, there are risks to storage buildings such as fire, storms etc. The loss of existing soils data would be a disaster not only because soil data are central to many of the major global issues the world is facing (McBratney et al., 2014; Amundson et al., 2015; Montanarella et al., 2016), but also because the tremendous resources that went into the collection and analysis of these data (Arrouays et al., 2017). Collection of comparable future soil data would certainly be cost prohibitive in many countries and not justifiable without first having made optimal use of data already available.

A visit to the Kenya Soil Survey department in the spring of 2016 confirmed this narrative. Legacy soil data remained idle in the Kenya soil survey library. Some data were in the collection of retired soil scientists. The probability of such data being lost was therefore very high and a lot of effort needs to be put to ensure they are not lost due to neglect or inattention.

Legacy soil data is a huge reservoir of soil information that can serve as inputs into DSM procedures or as evaluation data sets (Lagacherie, 2008). A quick look at the information contained within some of the existing legacy soil data from the Kenya Soil Survey showed that the legacy soil data often held information on the spatial distribution of soils, vegetation type, land evaluation keys, land quality assessments, geo-located soil profile data, geology, land use, crop suitability, soil fertility aspects, land management aspects, soil engineering properties, and soil erosion hazard. This information in itself is very useful as it can be used to support soil conservation or as major components of national environmental monitoring and hence revitalize agriculture (McBratney et al., 2003; Odeh et al., 2012). From soil maps, we can learn about the pedological context as well as some insight in the soil spatial variability, and since legacy soil data is also the history of soil mapping, several conclusions can be drawn on their relevance and their usage (Carré and Boettinger, 2008).

Legacy soil data arise from traditional soil survey (Bui and Moran, 2001). The methods of soil survey are generally empirical, based on conceptual models developed by the surveyors that correlate soil with underlying geology, landforms, vegetation, and air photo interpretations. In traditional soil survey, survey samples are located to confirm the surveyor's interpretation of the landscape and not in accordance with a statistical design. This automatically leads to bias at sampled locations. Carré et al. (2007) examined this problem in more detail. While soil observations collected in soil surveys pose a problem to the statistician, soil scientists recognize that the conceptual, mental models developed by soil surveyors in the past, represented in map legends and map boundaries, can be highly informative. The main challenge that legacy soil data pose is how to ensure that this information is effectively transferred into the DSM framework (Minasny et al., 2008).

There are two types of legacy soil data: (i) those existing as soil maps, and (ii) and those existing as data from individual soil profiles. Both occur in well-detailed soil survey reports. Soil maps provide a continuous representation of the soil pattern and can be used as soil covariates for DSM (Mayr and Palmer, 2006). In addition, soil maps can be used as a source for calibrating DSM procedures that consider the soil surveyor's knowledge if they are sampled to be representative of a larger region (Lagacherie et al., 1995). Existing soil profile data provide detailed information on many soil properties at different soil depths. Such data is often used as inputs into many statistical and geostatistical procedures to predict soil properties at unsampled locations (Carre and Girard, 2002; Hengl et al., 2004). A number of problems, however, usually hamper the use of legacy soil data. These include the unavailability of enough numeric data, lack of harmonization and imprecision of soil descriptions, imprecise georeferencing of soil profiles, and non-optimal location of soil data (Lagacherie, 2008). This has however not hindered the use of legacy soil data for DSM (Arrouays et al., 2017). A lot of efforts have been undertaken by various soil agencies to integrate soil data into coherent spatial data infrastructures (Dusart, 2005; Feuerherdt, 2006; Dobos, 2006; Daroussin et al., 2007).

2.2 Legacy Soil Data Rescue Efforts

2.2.1 Legacy Soil Report Rescue Efforts

Globally, there have been tremendous efforts by various institutions and initiatives to rescue legacy soil reports. Since 1966, the International Soil Reference and Information Center (ISRIC) has been compiling a large collection of articles, country reports, books, and maps with emphasis on developing countries (https://www.isric.org/explore/library, accessed 1/11/2019). The ISRIC library currently contains a collection of about 10,000 (digitized) maps and 17,000 reports and books, many of which can be accessed online. The subject emphasis is on soils, but related geographic information on climate, geology, geomorphology, land degradation, land use, and land suitability are also collected. The map collection contains mainly small-scale (1:250,000 or smaller) maps, the majority of which are accompanied by reports and related thematic and derived materials. A significant part of the ISRIC map collection was scanned at the European Commission's Joint Research Centre (JRC) as a foundation for the European Digital Archive of Soil Maps (EuDASM) whose main objective is to transfer soil information into digital formats, with the maximum resolution possible, to preserve the information of paper maps (https://esdac.jrc.ec.europa.eu/, accessed 9/16/2019). EuDASM provides access to an on-line collection of soil and related maps for Africa (Selvaradjou et al., 2005a), Asia (Selvaradjou et al., 2005b), Canada, Europe, Latin America and the Caribbean (Selvaradjou et al., 2005c), and the United States of America. Available maps can be downloaded and viewed on screen. More than 6,000 maps from 135 countries have been captured and are freely available through a user-friendly web-based interface, while over 30% of the available country reports are available as full text (PDF) (Panagos et al., 2011).

The Food and Agriculture Organization (FAO) land and water division has also made tremendous efforts to make legacy soil data available through the FAO soils portal (http://www.fao.org/soils-portal/en/ accessed 1/11/2019). Soil survey maps can be accessed through the FAO's *soil and legacy maps* web portal (http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/fao-soil-legacy-maps/en/ accessed 1/11/2019). It has a collection of up to 1,228 land legacy maps, mainly soil maps, but also land use, geology, and land cover maps. These maps were initially scanned as jpeg and then uploaded with standard metadata. Additionally, soil survey reports can be accessed through the FAO's *soil legacy report* portal (http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/soil-legacy-reports/en/ accessed 1/11/2019), which contains links to other institutes and organizations that have similar collections. These legacy reports resulted from the hundreds of soil survey field projects that were carried out since the 1950s and published as gray literature.

The World Soil Survey Archive and Catalogue (WOSSAC) also consists of an extensive collection of soil, land resource and land suitability surveys that were undertaken worldwide over the last 80 years by soil surveyors and scientists at the behest of the British government and others (https://www.wossac.com/ accessed 1/11/2019). This collection has been archived and catalogued using internationally recognized bibliographic standards (Hallet et al., 2006) and consists of 13,000 reports and maps available for consultation and use by the international soil science community. WOSSAC's mission is to provide a secure home for soil survey reports, maps, imagery, and photographs produced by British companies and surveyors overseas from 336 territories worldwide, with a view to ensuring their enduring availability and protection. WOSSAC concentrates solely on British-affiliated surveys in developing countries, but accommodates all soil survey materials that arrive irrespective of geographical coverage or source. The archive holds a wide variety of media, including map sheets; map albums; reports, books and monographs; satellite imagery on paper, film and in digital forms; GIS digital datasets; aerial photography; site photographs; micro-fiche; and survey electronic information (Hallet et al., 2011).

Within the United States, the United States Department of Agriculture (USDA) has published soil surveys reports since 1899 and archived many of these publications as PDF files. Most of the archived soil surveys include detailed soil maps. Printed copies of soil surveys reports are available at federal depository libraries and in some cases at USDA offices. These reports can be accessed through the USDA's Natural Resource Conservation Service web portal (https://www.nrcs.usda.gov/wps/portal/nrcs/soilsurvey/soils/survey/state/ accessed 7/10/2019). Web Soil Survey (WSS) (https://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm, accessed 9/16/2019) provides the most up to date soil data and information produced by the National Cooperative Soil Survey and is operated by the USDA Natural Resources Conservation Service (NRCS), which provides access to the largest natural resource information system in the world. NRCS has soil maps and data available online for more than 95% of the US's counties and anticipates having 100% in the near future. The site is updated and maintained online as the single authoritative source of soil survey information for the US and its territories.

2.2.2 Legacy Soil Profile Data Rescue Efforts

Even though soil survey reports have been scanned and made available online, many of the rescued soil survey reports do not contain geo-referenced soil profile information. Currently, there is much effort to capture and compile soil profile data mined from rescued soil survey reports into various databases. In this section, we highlight the efforts of various institutions and initiatives to rescue soil profile data from existing soil survey reports.

Since the 1980s, ISRIC has developed and managed a number of stand-alone soil profile databases that are freely available to the public. In 1986, ISRIC developed the Soil and Terrain (SOTER) database, a program initiated by FAO, the United Nations Environmental Program (UNEP), and ISRIC, under the auspices of the International Soil Science Society (ISSS) (van Englen and Dijkshoorn, 2012). The SOTER database is composed of a map that delineates the SOTER map units and a table with terrain and soil data that can be linked to the units of the map and provides data on key soil and terrain properties that are relevant inputs to agro-environmental applications such as food projection studies, climate studies, land evaluation or hydrological catchment modelling. The aim of the program was to develop a global SOTER database at scale 1:1,000,000 that was supposed to be the successor of the FAO-UNESCO Soil Map of the World (SMW). The SOTER database with global coverage, however, was never achieved, but SOTER databases were developed for various regions, countries and continents. Soil profile data in the SOTER databases is often incomplete, which hampers their applicability for quantitative studies. To overcome this, consistent taxotransfer rules, methodology for filling gaps in primary soil analytical data, were used to fill gaps in the SOTER soil profiles (Batjes, 2003). From these soil profiles, a consistent set of 18 soil properties were derived for depth intervals of 20 cm up to 100 cm depth. The soil properties include: organic C, total N, pH, CEC, base saturation, Al saturation, $CaCO_3$ and gypsum content, exchangeable Na, electric conductivity, bulk density and the sand, silt and clay fractions. The database contains 6,388 soil profiles (Ribeiro et al., 2018).

ISRIC also maintains the World Inventory of Soil Emission Potentials (ISRIC-WISE) soil profile database that was developed between 1991 and 1995 (Batjes and Bridges, 1994). The current version of the ISRIC-WISE database (ISRIC-WISE3) was compiled from a wide range of soil profile data collected by many soil professionals worldwide (Batjes, 2009). All profiles have been harmonized with respect to the original FAO Legend and FAO Revised legend of the Soil Map of the World (FAO-UNESCO, 1974; FAO, 1988). WISE3 holds selected attribute data for 10,253 soil profiles, with some 47,800 horizons, from 149 countries. Individual profiles have been sampled, described, and analyzed according to methods and standards in use in the originating countries. The primary data contained within the ISRIC-WISE3 project has been used for a range of applications such as the development of harmonized sets of derived soil properties for the soil types of the world, gap-filling in primary SOTER databases, global modelling of environmental change, analyses of global ecosystems, up-scaling and downscaling of greenhouse gas emissions, and crop simulation and agro-ecological zoning (ISRIC-WISE, 2008).

Significant progress has also been made in collecting and assembling legacy soil profile data for sub-Saharan Africa. This effort draws from decades of soil survey campaigns in Africa, converting the legacy profile data into formats that link the digital platforms. ISRIC-World Soil Information Service (WoSIS) has been compiling legacy soil profile data for Sub-Saharan Africa, as a project activity of the globally integrated-Africa Soil Information Service (AfSIS) project (Shepherd and Vågen, 2010). The Africa Soil Profiles database (AfSP) currently contains over 18,500 georeferenced legacy soil profile records for 40 countries (Leenaars et al., 2014b). The soil profile records were identified and collected from over 500 data sources, both digital and analogue, and the data were compiled and converted to a common standard and passed through basic quality rules and cleaning. Previously, such data would only be accessible through a myriad of sources and would not be standardized, hindering efforts to make the data shareable and usable. All records include soil profile layer attribute data and over 80% of the records include soil analytical data including, but not limited to, those specified by GlobalSoilMap. Soil attribute values were standardized according to e-SOTER conventions and validated according to routine rules. e-SOTER was proposed to overcome the shortcomings of SOTER and provide a regional platform that can be extended worldwide. The degree of validation, and associated reliability of the data varies because reference soil profile data that were previously and thoroughly validated are compiled together with non-reference soil profile data of lesser inherent representativeness. The database is continuously updated and growing, and milestone versions have been posted online and made available to the project and the public. This data has since been used to produce an array of both physical and chemical soil property maps at an initial resolution of 1 km, for 8 different soil properties, and currently at 250 m resolution for 35 different soil properties.

Due to the need for central global soil profile database, the World Soil Information Service (WoSIS) was developed in 2015 (Ribiero, et al., 2015). WoSIS contains "a centralized and user-focused database containing only validated and authorized data with a known and registered accuracy and quality" (Tempel et al., 2013). WoSIS is a server database for handling and managing multiple soil profile datasets in an integrated manner, subsequent to proper data screening, standardization, and ultimately harmonization of all the existing stand-alone soil profile databases. All data submitted for consideration in WoSIS are first preserved 'as is' in the ISRIC's world data center for soils as certified soils, data repository by the CoreTrustSeal Board (Edmunds et al., 2016). Subsequently, they are assessed for quality, standardized and, where possible, harmonized using consistent procedures. Special attention is paid to the selection of soil properties considered in the GlobalSoilMap specifications such as pH, SOC content, soil texture, bulk density, CEC, and soil water retention (Arrouays et al., 2014). WoSIS receives its data from a multitude of sources including the AfSP, ISRIC-WISE, and SOTER databases and other national and regional soil profile databases from around the world.

2.3 History of Soil Mapping in Kenya

Attempts to map the soils of Kenya date back to 1908 when there was a need for soil survey for settled areas (Aore, 1995). However, it was not until 1936 that the first soil map of Kenya was included in Milne's 'A Provisional Soil Map of East Africa' at a scale of 1:2,000,000 (Milne, 1936). Refinement of this map led to the publication of the 1:3,000,000 soil map of Kenya by Gethin-Jones and Scott in the first edition of the National Atlas of Kenya (Gethin-Jones and Scott, 1959; Survey of Kenya, 1959). The same map was included in the second and third editions of the National Atlas (Soil Survey of Kenya, 1962 and 1971). Basically, the same information as contained in the Atlas was used by Scott for the 1:4,000,000 soil map of East Africa in Morgan's book on people and natural resources of the region (Scott, 1969).

In all these attempts, soils were mapped following the catena sequence as developed by Milne (1936). This catena concept was taken a step further into a 'Land System' approach, which resulted in the preparation of the 'Land System Atlas for Western Kenya' at a scale of 1:500,000 (Scott et al., 1971). Making use of available information, FAO-UNESCO (1974) published the 'Soil Map of Africa' at a scale of 1:5,000,000 as part of 'Soil Map of the World' program.

The exploratory 'Soil Map of Kenya' at a scale of 1:1,000,000 (Sombroek et al., 1982) was the fourth effort to comprehensively map the soils of Kenya. Through a methodology developed by the Kenya Soil Survey from 1972 onwards termed '*physiographic soil survey'*, this exploratory soil map visualized the complex relation between landforms, geology, and soils. This map was used as a basis for compilation of the generalized soil map to be represented in the fourth edition of the National Atlas. In the process of generalization, the number of mapping units was reduced, the legends simplified, and the soil classification simplified to the highest level.

The legend of the generalized soil map represents a subdivision of the country into major landforms and soil mapping units (SMUs). Each SMU was described in terms of drainage condition, depth, color, texture, etc. using a descriptive terminology that was based on the FAO guidelines (FAO, 1977). Each SMU description was then followed by the soil classification according to FAO-UNESCO's legend for their Soil Map of the World (FAO-UNESCO, 1974).

More detailed soil surveys in Kenya started on an *ad hoc* basis in early 1950s mainly at a detailed or semi-detailed scale in areas earmarked for development. Soil surveys were mainly carried out by the chemistry department of the National Agricultural Research Laboratories (NARL) under the umbrella of the senior soil chemist. In the early 1960s, a soil survey unit was set up as part of the chemistry section. From 1972 onwards, soil survey in Kenya was considerably strengthened under a bilateral aid agreement between The Netherlands and the Kenya Soil Survey (KSS). KSS was mandated to conduct soil and other land resource surveys throughout Kenya to provide information about soil and land resources required for accelerated agricultural development. All these efforts led to the provision of traditional soil maps for Kenya at different scales (Kenya Soil Survey, 1984). Experienced soil surveyors who knew the area well spent much time in the field making auger observations at regular intervals, and in this way drew a field soil map that was later digitized and printed.

2.4 Current Soil Mapping Efforts

On a global scale, the only complete systematic survey of the world's soil resource remains the FAO-UNESCO Soil Map of the World completed in the 1970s at a scale of 1:5,000,000 (FAO-UNESCO, 1974). Modern technologies applied to it have permitted much more useful information to be derived from it than was originally envisaged. Publications such as World Soil Resources (FAO, 1991) and Potential Population Supporting Capacities of Lands in the Developing World (FAO/UNFPA/IIASA, 1982) demonstrated what can be done by systematically re-analyzing existing soils data. The Soil Map of the World is now available in digital form making it more useful compared to the paper version that is very large and hard to use (http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/ accessed 8/20/2019).

Nevertheless, the original maps are showing their age and need updating considering the wealth of more recent soils information (Purnell, 1995). This initiated the updating of the legend of the Soil Map of the World (FAO, 1988). A more ambitious and potentially valuable effort was to make a worldwide soils and terrain map at a scale of 1:1,000,000: SOTER project (ISRIC, 1986). Despite the widespread support from UNEP, FAO, ISRIC, ISSS, and some national soil surveys, financing for the whole project was not achieved (Purnell, 1995). At present, less than half the world has a complete, systematic published 1:1,000,000 scale soil map.

Another small-scale map is the Global Assessment of Human-induced Soil Degradation (GLASOD) that originally begun in the 1970s and was published in 1990 at a scale of 1:10,000,000 (Oldeman et al., 1991). The status of soil degradation was mapped within loosely defined physiographic units (polygons), based on expert judgment. The type, extent, degree, rate, and main causes of degradation were printed on a global map and documented in a downloadable database (https://www.isric.org/projects/global-assessment-human-induced-soil-degradation-glasod accessed 1/11/2019).

Currently, efforts to map soils at finer resolutions for Africa, including Kenya, have soared (Vågen et al., 2013). The African Soil Information Service (AfSIS) initially mapped selected physical and chemical soil properties at a resolution of 1 km (Hengl et al., 2014) and currently mapping has been downscaled to 250 m resolution (Hengl et al., 2017). Additionally, Hengl et al (2017) predicted the distribution of soil classes based on the World Reference Base (WRB) and the USDA classification systems for the entire world, Africa, and subsequently for Kenya at 250 m resolution.

All these efforts affirm that existing legacy soil maps, though not perfect, contain a wealth of information that can potentially serve as a starting point for revising soil legacy soil maps using advances made in GIS and soil modelling.

2.5 Challenges Associated with Traditional Soil Survey Methods

Traditional soil survey is anchored in the classic catena concept. Soil scientists first built a conceptual model associating soils with specific landscape positions. These models were then used together with photo-interpretations to identify and delineate soil-landscape units that were delineated by soil polygons. Although this process was very successful in providing information on soil variability within a landscape, the three main drawbacks associated with this approach include the polygon-based model itself, the manual mapping process, and the lack of documentation on the soil-landscape model.

2.5.1 The Polygon Based Model

With the polygon-based model, only soil bodies large enough to be drawn on the map by the cartographer can be shown. Therefore, the level of detail is limited by the scale of the map, not by what the soil scientist knows. In addition, soils in a given polygon are treated as homogenous bodies; changes in the soil property values occur only at the spatial boundaries of the polygons, which is almost never true. This creates an unrealistic representation of spatial variability of soil properties within the landscape.

2.5.2 Manual Mapping Process

The traditional process of mapping soils was mainly manual, in which surveyors delineated the extent of soil bodies based on visual interpretation of environmental conditions. It is very hard for soil mappers to correctly identify and delineate soil-landscape units using a few environmental data layers due to the limits of the human capacity for simultaneous visual perception of multiple variables. As a result, the delineation of soil-landscape units may not reflect the totality of knowledge processed by the surveyor. Most soil mappers base their soil-unit delineation solely on the visual interpretation of stereo aerial photographs. Subtle and gradual changes in environmental conditions are often difficult to discern via stereoscope and it is not unusual to misplace boundaries of soil polygons in the manual process of delineation. Thus, the mapping process is tedious and time consuming, and can be error prone and inconsistent.

2.5.3 Lack of Documentation

The lack of explicit documentation on soil landscape models for maps is a major limitation. In most soil survey reports, the landscape model is deeply buried within the survey text and needs careful interpretation to discern what kind soil landscape model used.

2.6 The Case for Digital Soil Mapping (DSM)

While digital soil maps are available for most parts of the world (Grunwald et al., 2011), for many areas, they exist at a very small scale (1:1,000,000 or coarser), and do not adequately represent soil variability in a way that is useful to non-pedologists (Sanchez et al., 2009). The majority of currently available digital soil maps are actually compilations of multiple legacy soil maps that were initially produced as hard-copy and subsequently digitized (Grunwald et al., 2011). For example, although the Harmonized World Soil Database version 1.2 is a digital product, it is based on paper maps that were later converted to vector-based polygon maps and later on rasterized (ISRIC, 2012).

In the U.S. the Soil Survey Geographic database (SSURGO) soil maps were originally produced as part of approximately 3,000 independent soil surveys. These individual maps were made at different times and have different scales. They were created using different mapping concepts, and they often use different soil components and different estimated property data to represent the same soil landscape features. Consequently, there are frequently artificial boundaries in the data associated with geopolitical boundaries caused by discontinuities in mapped soil properties and soil used and management interpretations. All these emphasize the point that digitizing existing paper maps is not DSM.

Hartemink et al. (2010) listed the limitations of most existing digitized soil survey maps: (i) they are static, (ii) they aggregate soil information into soil classes that are not readily compatible with quantitative applications, (iii) the information content has been overly generalized relative to the information on the regional soil resources that was collected to create the soil survey, (iv) they are improperly scaled, and (v) they represent the information as polygons that are not as readily combined with most other natural resource data that are raster-based. Similarly, Zhu (2006) emphasized that the spatial and attribute generalization of soil into discrete classes makes soil survey information incompatible with other forms of continuous spatial data for environmental modelling. All things considered, there is a tremendous potential for the DSM community to capitalize on the demand for better soil information by improving the quality of existing polygon-based soil maps and directly creating raster-based soil data of functional soil properties by utilizing soil properties existing within legacy soil data.

2.6.1 Digital Soil Mapping

Digital soil mapping refers to "the creation and population of spatial soil information systems by numerical models inferring the spatial and temporal variations of soil types and soil properties from observations and knowledge and from related environmental variables" (Lagacherie, 2008).

2.6.1.1 The Clorpt Model – State Factor Model

The scientific rationale for soil mapping has been the state factor, or clorpt model (Jenny, 1941, 1980) that was originally proposed by Dokuchaev (1883) and Hilgard (1906), and later articulated by Hans Jenny (Hudson, 1992). The model is formalized by the following equation,

$$S = f(cl, o, r, p, t, \ldots) \qquad [2.1]$$

where soil (S) is considered to be a function of climate (cl), organisms (o), relief (r), and parent material (p) acting through time (t) (Jenny, 1941, 1980). The ellipsis (.) in the model indicated that for additional unique factors that may be locally significant, such as atmospheric deposition (Thompson et al., 2012).

The *clorpt* equation illustrates that by correlating soil attributes with observable differences in one or more of the state factors, a function *(f)* or model can be developed that explains the relationship between the two that can be used to predict soil functional properties at new locations. An important distinction of the *clorpt* model is that "the factors are not formers, or creators, or forces; they are variables (state factors) that define the state of a soil system" (Jenny, 1961) meaning that the factors do not constitute pedogenic processes, but are factors of the environmental system which condition processes. To bridge the gap between factors and processes, the *clorpt* model is supplemented by additional models that are useful at explaining various processes at different scales. A notable example is the catena concept (Milne, 1936) that attributes soil variation along a hillslope sequence to erosion and deposition, hydrology, and stratigraphy.

2.6.1.2 The Scorpan Model – The Digital Soil Mapping Formula

Recently, McBratney et al. (2003) offered a revised formalization of the state factor model.

$$S = f(s, c, o, r, p, a, n) \quad [2.2]$$

where *S*, a set of soil attributes (S_a) or classes (S_c) is considered as a function of other known soil attributes or classes (*s*), climate (*c*), organisms (*o*), relief (*r*), parent materials (*p*), age or time (*a*), and spatial location or position (*n*). The *scorpan* equation also explicitly incorporates space (x, y coordinates) and time (~t). Thus, the *scorpan* model can be expanded as follows:

$$S[\mathrm{x,y,\sim t}] = f(s[\mathrm{x,y,\sim t}], c[\mathrm{x,y,\sim t}], o\,[\mathrm{x,y,\sim t}], r\,[\mathrm{x,y,\sim t}], p[\mathrm{x,y,\sim t}], a[\mathrm{x,y,\sim t}], [\mathrm{x,y}]) \quad [2.3]$$

This expansion indicates that *scorpan* is a geographic model, where the soil and factors are spatial layers that can be represented in a geographic information system. The *scorpan* model deviates from the *clorpt* model in that it is intended for quantitative spatial prediction, rather than explanation (McBratney et al., 2003).

The *scorpan* model has been used to predict soil properties for Sub-Saharan Africa using the AfSP Database for thirty-five (35) different soil functional properties at 2-7 standard depths: 0, 5, 15, 30, 60, 100, and 200 cm, following the vertical discretization as specified in the GlobalSoilMap specifications (Arrouays et al., 2014) at spatial resolutions of 1 km and 250 m. These maps can be accessed through this link https://data.isric.org/geonetwork/srv/eng/catalog.search#/search?resultType=details&sortBy=relevance&any=Africa%20Soil%20Profiles%20Database&fast=index&_content_type=json&from=11&to=20 (accessed 9/13/2019).

CHAPTER 3. RENEWAL OF ARCHIVAL LEGACY SOIL DATA: A CASE STUDY OF BUSIA AREA, KENYA

Abstract

Much older soils information, collectively known as 'legacy soil data' lies idle in libraries or in the personal collections of retired soil scientists. The probability of this legacy data being lost or destroyed is very high. We demonstrate the stepwise process of bringing legacy soils data 'back to life' using the *Reconnaissance Soil Survey of the Busia Area (quarter degree sheet No. 101)* in western Kenya as an example. The first step, *data archeology*, involves locating and cataloging legacy soil data from key institutions, which often requires numerous site visits and the assistance of individuals familiar with the desired data. The second step, *data rescue*, involves converting paper copies of data into a digital format by scanning the maps, narrative descriptions, and tables, and storing the information in a database. The third step, *data renewal*, consists of bringing the data to modern standards by taking advantage of technological and conceptual advances in geo-information technology. In our example, the resulting digital (scanned) soil map of the Busia area is a significant upgrade from the fragile paper map. Careful interpretation of the agronomic information available within the legacy soil survey allowed us to produce ten land quality maps showing the ability of the land to perform specific agronomic functions, and nineteen different crop suitability maps that were not available originally. Some of these maps will be made available in the Soil Explorer app and SoilExplorer.net website. These rescued maps and their associated tabular and narrative data, while useful themselves, also provide crucial inputs for generating more detailed soil maps using digital soil mapping techniques that were unavailable when the original mapping was conducted.

3.1 Introduction

The majority of current soil resources exist as traditional soil maps, soil survey reports, soil survey manuals, land evaluation frameworks, soil profile descriptions, and farm management handbooks, collectively known as legacy soil data (Zinck, 1995). These soil resource inventories have been widely used as meaningful sources of soil information to support soil conservation or as major components of national environmental monitoring (McBratney et al., 2003; Odeh et al., 2012; Cambule et al., 2015). Information on soils for much of Africa and most developing countries is sparse. Kenya fortunately has considerable soils information (Dijkshoorn, 2007).

Unfortunately, the majority of available legacy soil data often remains idle in libraries. The demand for soil data, however, is soaring (Cook et al., 2008). The probability of such data being lost through disasters, be it natural, manmade, political or simply inattention is very high (Rossiter, 2008). Our visits to the Kenya Soil Survey (KSS) in the spring of 2016 confirmed this. The majority of legacy soil data was left unused and stored in library shelves, some were in private collections of retired soil scientists, and those existing in digital format are largely unused or used only internally. One potential solution for the dissemination of such geographical data was to take advantage of advances in geo-information technology to take this data out of libraries and reduce the possibility of loss of the paper data and the information it contains.

This is mainly driven by the fact that a lot of effort and resources financial and human, went towards compiling, analyzing and publishing the data contained in legacy soil data. The information within such data often consists of spatial distribution of soils, land quality, crop suitability, geolocated soil profile information with their respective laboratory data, geology, and land use type. This sort of data can be analyzed and used as a primary input for digital soil mapping (DSM) especially for countries with sparse soil data infrastructures (McBratney et al., 2003; Baxter and Crawford, 2008; Krol, 2008). Since resources needed to replicate these studies is dwindling, the only justifiable thing to do is to fully use available soil data and then seek additional resources to study aspects not provided by the legacy data.

Ways on how to effectively utilize renewed legacy soil data has become a new area of research in soil science (Hengl et al., 2007). Currently, many discussions center on how to collate legacy soil data to enhance the re-use of such data to meet current demands (Dent and Ahmed, 1995; Ahmed and Dent, 1997; Baxter and Crawford, 2008; Rossiter, 2008; Dobos et al., 2010; Cambule et al., 2015). We observe, however, a need in the literature on how to renew legacy soil data, and how to interpret information stored in legacy soil surveys for additional agronomic information (Rossiter, 2008; Odeh et al., 2012; Cambule et al., 2015; Arrouays et al., 2017).

This study aimed at providing a conceptual framework for using geographical information techniques to support the renewal of legacy soil data by transforming the best available soil survey of a selected portion of Kenya into a digital format. To meet this objective, we followed the quality criteria described by Forbes et al. (1987) and used by Cambule et al. (2015) to guide the legacy soil data renewal to meet current and future demands for soil information.

3.2 Materials and Methods

3.2.1 Study Area

Busia area situated in the western part of the Republic of Kenya. It is bound by the equator to the south, latitude 0° 30’ N in the north, longitude 34° 30’ E in the east, and the Kenya-Uganda border to the west (Fig. 3.1). It has an acreage of 279,800 hectares including Lake Namboyo, Lake Kanyaboli, and part of Lake Victoria, which occupy 12,800 hectares. The climate of the area is characterized by a mean annual rainfall of 925 mm to 1990 mm. In drier areas bordering Lake Victoria, only one rainy season (March-May) is noticeable. The wetter areas have two rainy seasons, March – May and August – December (East Africa Metrological Department (EAMD), 1972).

The mean annual temperature ranges from 20 to 23 °C, with the mean annual maximum temperature ranging from 26 to 29 °C, whereas the mean annual minimum temperature ranges from 14 to 17 °C (EAMD, 1970). Dominant geological features are (1) igneous rocks (Samia hill series and andesites), (2) sedimentary rocks (Kavirondo series), (3) intrusive rocks (granites, dolerites, and felsites), and (4) quaternary superficial deposits (Rachilo and Michieka, 1991). The soils of the survey areas are classified according to the legend of the Soil map of the World (FAO-UNESCO, 1974), which has now been incorporated into the World Resource Base (WRB) for Soil Resources (Spaargaren and Deckers, 1998). The major soil classifications are Arenosols, Ferralsols, Nitosols, Luvisols, Acrisols, Vertisols, Gleysols, Histosols, Solonchaks, Fluvisols, Cambisols, and Lithosols (Rachilo and Michieka, 1991).

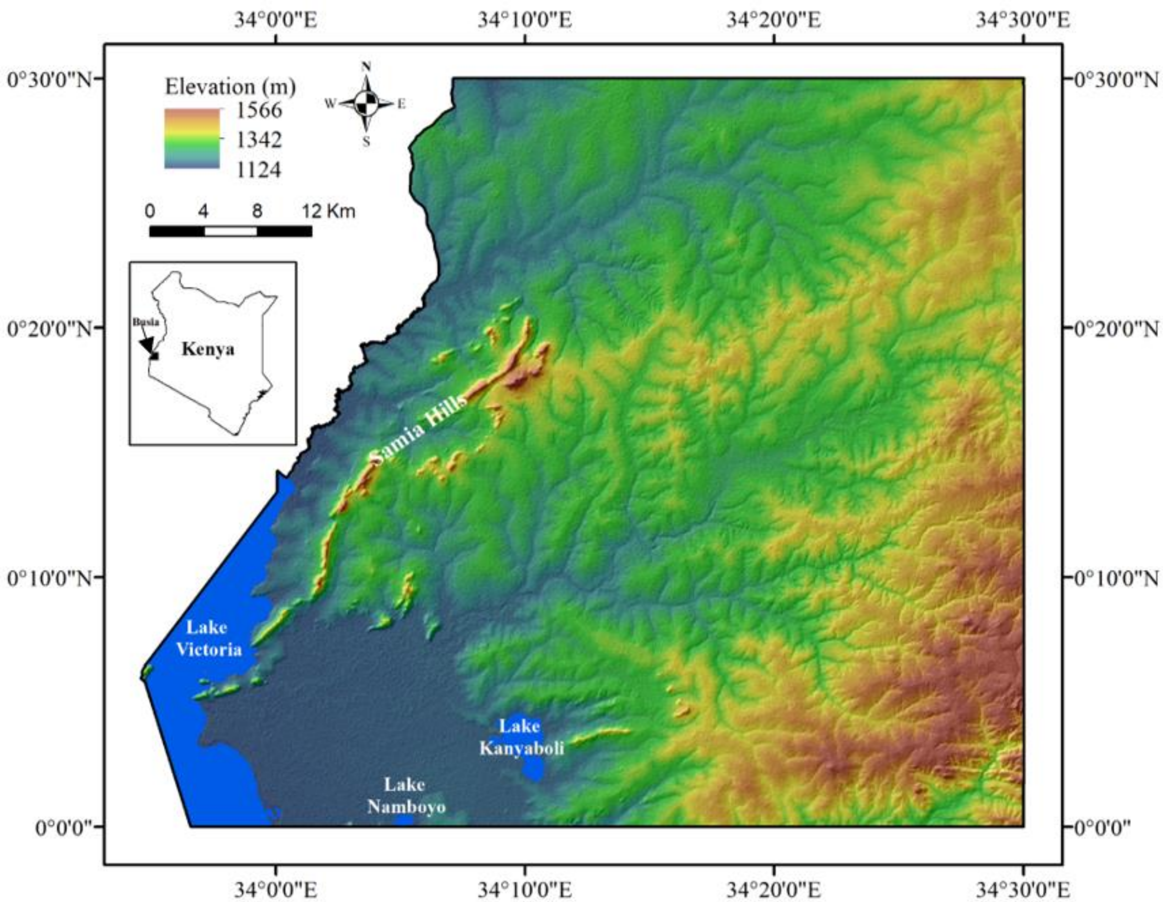


Figure 3.1: Geographical location of the Busia area.

3.3 Methodology

This study consisted of four phases. Phase 1 consisted of meetings with key partners and stakeholders in Kenya to identify a setting for this study. Phase 2 entailed locating and cataloging all historical legacy soil data for the study area by contacting various agricultural institutions through a process known as *data archeology*. All of the recovered legacy soil data for the study area were transformed to an up-to-date archival format by scanning or by direct entry into a database, a process known as *data rescue*. The quality of the legacy soil data was assessed using relevant adequacy criteria (Forbes et al., 1987; Goodchild and Hunter, 1997). Finally, the legacy soil data were brought to modern standards by taking advantage of technological and conceptual advances in geo-information technology through a process known as *data renewal*.

The renewal phase includes georeferencing soil polygons and point data and linking it with their associated attribute databases, use of auxiliary data such as multispectral images and derived terrain attributes, generation of metadata, which includes description of semantics used within the legacy soil data, and integration of the legacy soil data into an easily accessible geospatial data infrastructure. Data archeology, rescue, and renewal are terminology proposed by Rossiter et al. (2008) and are explored in detailed in this paper. Lastly, we use the renewed legacy soil data to generate additional agronomic information through interpretation of the soil information in the soil survey report.

3.3.1 Geodetic Control

One drawback of many legacy soil survey maps is lack of geodetic control points (Cambule et al., 2015). Four corners of the map that had clearly labeled latitude and longitude were used as control points. An additional forty control points were used to improve on georeferencing. These additional points were identified using clues such as road intersections and rivers that were clearly visible on the scanned map. The quality of this step was assessed by the absolute Root Mean Square Error (RMSE) of the control points.

3.3.2 Creation of GIS Layers

This step involved the conversion of the scanned soil map to a GIS layer. Fortunately, the KSS provided a digitized soil map of the area. However, a quick look at the GIS layer showed that the layer has was not correctly digitized. Not all soil map units were digitized, and most soil polygon boundaries were inaccurately drawn. The soil polygon map from the KSS was manually edited by going through the soil map unit boundaries and editing them to align with the digitized soil map in ArcMap. The attribute table was also edited and populated with attributes from the original soil map and survey report. The soil map and the survey report were also subjected to quality assessment following the adequacy criteria: scale and texture and legend, as described by Forbes et al. (1987) to determine the effectiveness of their use.

Map Scale and Texture

Forbes et al. (1987) defines map scale as the relation between distances on the map and corresponding distances on the ground whereas map texture refers to “the sizes and pattern of delineations on the map, and determines the map’s overall legibility”. The scale and texture of the Busia map were evaluated to assess the legibility and capability of the map to represent the smallest area of interest. These include the Minimum Legible Area (MLA) (Eqn. 3.1), which indicates the smallest land area that can be represented on the map at its published scale using the criterion of a Minimum Legible Delineation (MLD) of 0.4 cm^2 (Forbes et al., 1987). The Index of Maximum Reduction (IMR) was also used to assess the map (Eqn. 3.2), and it refers to the factor by which the scale of the map could be reduced before the average size delineation (ASD) would become equal to the MLD, i.e. before more than half of the map would become illegible. An IMR of 2.0 is considered optimal, that of 1.58 as minimally acceptable and an IMR of greater than 2 implies that the map is very legible. A large IMR implies that the survey area is represented on a map that is physically larger than necessary (Forbes et al., 1987). The ASD of a portion of a map is the arithmetic mean of the sizes of the delineations in that portion of the map (Eqn. 3.3). It is estimated for portions of a map with a given map texture by randomly sampling the map areas with circles or squares of known area, and converting the count of delineations in several of these areas (Laker, 1977). A transparent overlay with a 2.5 cm radius circle was used to count the number of delineations within the circle.

$$\mathrm{MLA} = \frac{(\frac{1}{\mathrm{RF}})^2}{2.5*10^8} \quad [3.1]$$

$$\mathrm{IMR} = \sqrt{2.5 \text{ x ASD}} \quad [3.2]$$

$$\mathrm{ASD}_{2.5 \text{ cm radius circle}} = \left\{\left(\frac{\text{sum of 5 counts}}{142}\right) - 0.1\right\} \quad [3.3]$$

Map Legend

The map legend identifies the map units, generally referring to a full description in the associated survey report and may also provide a brief description and various interpretations. The legend can be identified by the symbols printed inside the map unit polygons. The descriptive legend gives information about each map unit. The map unit names and definitions in descriptive and interpretative legends dictate the level of usefulness of the information. The map legend may be evaluated either in terms of specific use of the soil inventory or by a more general criterion, such as a soil classification system (Forbes et al., 1987). We evaluated the map unit information based on the soil classification. Information is considered adequate if the map unit description included the diagnostic information such as horizons, and properties, or the soil classification. The overall information quality of the Busia soil survey map was expressed by the proportion of the survey area evaluated as 'adequate' relative to the total number of units or area.

3.3.3 Integration of Remotely Sensed Data

The renewal of legacy soil data requires the integration of ancillary data such as remotely-sensed data (Rossiter, 2008). Such products include Landsat imagery, vegetation cover, and terrain attributes. Soil is in part related to topography and vegetation and therefore the boarders of some soil map units may be well depicted by these remotely-sensed data. This requires an overlay of the digitized map and the remotely sensed products. We used the 30m SRTM 1 Arc-Second

Global elevation data projected to the WGS84 Web Mercator (auxiliary sphere) coordinate system (USGS, 2017a) and compute the hillshade using the System Automated Geoscientific Analyses (SAGA) Geographical Information System (GIS) software (Conrad et al., 2015) to provide a grayscale 3-dimensional representation of the surface. The satellite imagery available in ArcMap was used a base map because (1) it is geometrically correct (2) it provides the best currently available up to date imagery of the study area from space, and (3) it is freely available. These two products were useful in georectification of the soil polygon map because map units in the study area represented physiographic units such as hills, valleys, swamps etc.

3.3.4 Metadata

This step involved the development of appropriate metadata to include key identification information such as spatial data source, spatial reference, attributes, information on data quality, and description(s) of methods used to renew the data. The metadata should also include the explanation of key semantics.

3.3.5 Interpretation of Soil Survey Data

Soil surveys can provide basic information on soil and land characteristics that can be useful for various purposes, for example, determining the suitability for various types of agriculture, range and forestry. This exercise involved interpretation of the land survey data from which land can be classified to show its use for a particular kind of land use in a process called land evaluation (Rachilo and Michieka, 1991).

For the Busia area, the ‘Framework for Land Evaluation’ prepared by FAO (FAO, 1976), was followed in the land evaluation process. Fundamental in this approach is that land can only be classified meaningfully for clearly defined land uses termed as land use alternatives types or land utilization types considered relevant for the survey area. Each land use type was defined by some specific, quantifiable factors that have a marked influence on the performance of the land that is integral for defining land utilization types devised in terms of crop suitability maps.

In many existing land evaluation systems, single land characteristics such as drainage conditions, or texture are used as a basis for diagnosis and for establishing suitability class to determining land use specifications, grouped together into land qualities. A land quality can be defined as the fitness of a given type of land for a specific type of use (FAO, 1976) without permanent damage (Soil Conservation Society of America, 1976).

The KSS prepared proposals for rating land qualities for the Busia area (internal communication Nos. 7 and 29 (Braun and van de Weg, (1977) and FURP (1988) respectively). In this rating system, land qualities were classified into three to five grades ranging from very low to very high based on the most limiting factor for the land qualities. The next step involved the establishment of specifications for the land qualities that will define the suitability class levels for each land use alternatives. We followed these steps to generate land quality and crop suitability maps of the Busia area. The current suitability evaluation of each soil map unit for each land use alternative was carried out by comparison of the rating of the land qualities of that particular soil map unit with specifications for each land use alternative.

3.4 Results and Discussions

3.4.1 Site Identification

We met with the Kenya Soil Survey (KSS), the Kenya Agricultural and Livestock Research Organization (KALRO), and the Academic Model Providing access to Healthcare (AMPATH) to identify a site for this study. Busia area was identified because it met the following criteria: (1) it had accessible detailed legacy soil data at a scale of 1:100,000 (Rachilo and Michieka, 1991), (2) agriculture is the main economic activity in the area (Onywere et al., 2011), (3) the area has high population and poverty densities (Kenya National Bureau of Statistics, 2010), and (5) we are familiar with the area.

3.4.2 Legacy Soil Data Archeology

Legacy soil data was obtained from the Kenya Soil Survey (KSS) in Nairobi, Kenya (Fig. 3.2). We obtained several soil survey reports in western Kenya and settled for the Busia area for the reasons stated above. Additionally, the main author is from the Busia area and familiar with the study area. The main soil survey report for the Busia area was published as the *Reconnaissance Soil Survey of the Busia Area (quarter degree sheet No. 101)* (Rachilo and Michieka, 1991). After studying this soil survey report, maps, and tables obtained during our first visit, we identified missing materials, which resulted in a second visit to the KSS to obtain missing information. Table 3.1 summarizes the list of soils information contained within the survey report. A reconnaissance site visit to the area was also conducted to familiarize ourselves with the area.

Figure 3.2: The Kenya Soil Survey library, June 6 2017. Photo by Joshua Minai.

3.4.3 Legacy Soil Data Rescue

The majority of the data in the soil survey report existed in a paper format. The survey report consists of the report itself, which was a published report of the soil resources of the Busia area. Additionally, there were three map sheets which include two soil maps of the Busia area, one in color (Fig. 3.3) and another in black and white (Appendix 1 and 2 in Rachilo and Michieka, 1991 respectively) and a black and white soil engineering map (Appendix 7 in Rachilo and Michieka, 1991). It also has two large folio sheets that include a land evaluation key (Appendix 3 in Rachilo and Michieka, 1991) and soil profile characteristics significant for the soil classification (Appendix 4 in Rachilo and Michieka, 1991) (Table 3.1).

Table 3.1: Soil information contained within the *Reconnaissance Soil Survey of the Busia Area*.

Type of Soil Information	Scale
a) Reconnaissance soil map of the Busia Area (colored)	1:100,000
b) Reconnaissance soil map of the Busia Area (black and white)	1:100,000
c) Soil engineering map of the Busia Area (in black and white)	1:100,000
d) Land evaluation key	
e) Soil profile characteristics significant for soil classification	
f) Soil profiles and analytical data descriptions	
g) Land quality ratings for soil mapping units	

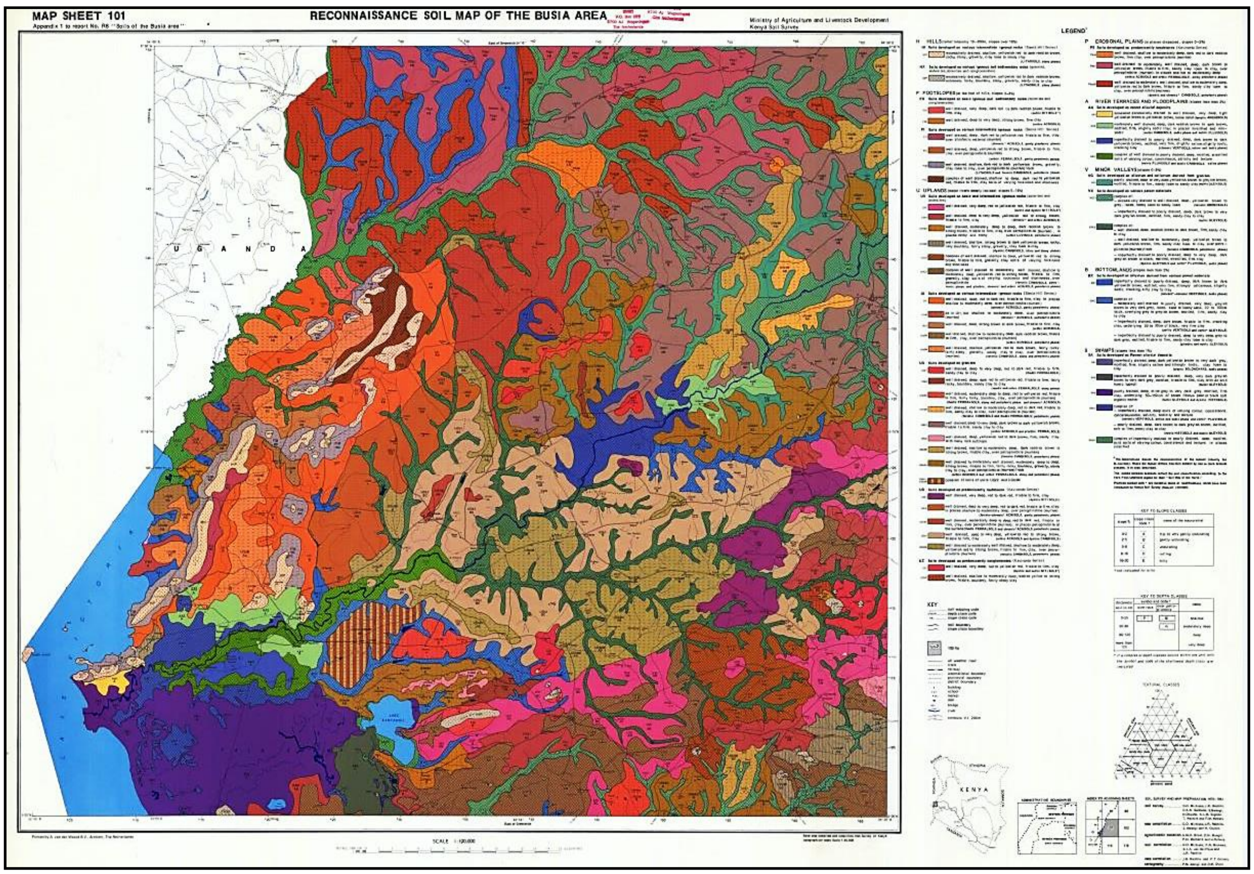


Figure 3.3: Rescued soil map of the Busia area at a scale of 1:100,000 from Panagos et al. (2011) used with permission.

3.4.4 Legacy Soil Data Renewal

Geodetic Control

The geodetic control was good: both geographic and grid coordinates were printed on the margin of the maps (Fig. 3.4). The map projection was not given explicitly, but conversations with the KSS GIS expert confirmed that Busia soil map was developed using the East Africa War System of Coordinates Traverse Mercator projection, belt I on the Arc 1960 datum. The reference ellipsoid was Clarke 1880. The reconnaissance soil map was first projected to Arc 1960 and then georeferenced using the four geodetic control points (GCPs) printed at the four corners of the map. This was then projected to the WGS84 Web Mercator (Auxiliary Sphere).

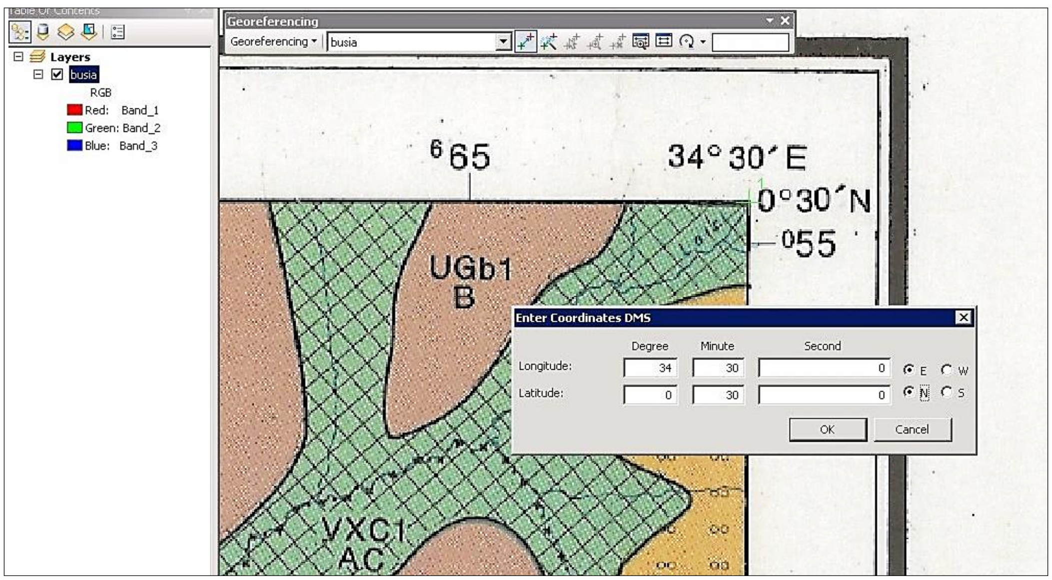


Figure 3.4: Geo-referencing of the reconnaissance soil map of the Busia area map using the top right corner geodetic control point. This geodetic control point is designated 34^0 30’E, 0^0 30’N.

For georectification purposes, satellite imagery was used as a base map because it is freely available and provides the best currently available, up to date, georeferenced imagery of the study area. Key features on the soil map such as road intersections and natural features such as rivers were clearly visible on the scanned map and were used as additional control points for georectification. Forty additional control points, evenly spread out within the study area were used for georectification. Using these GCPs, the map was transformed using the third order polynomial transformation. Forty-four GCPs were used and resulted to an RMSE of 6.63×10^{-4} m. A large number of GCPs is necessary to transform the map and achieve the best possible RMSE.

GIS Coverages

The digitized soil map showed that the soil map units were often inaccurately drawn and did not accurately delineate and capture key features such as islands and hills (Fig. 3.5). This is a common challenge with old paper maps as it was a common practice for drafters, not surveyors, to transfer soil boundaries by eye from field sheets to base maps. Without the soil surveyor's expert eye, knowledge, and experience, soil boundaries that followed obvious landscape features may not be produced correctly (Rossiter 2008). This is expected because base maps used by soil mappers available in the early 1980s were not as accurate compared to what is available today.

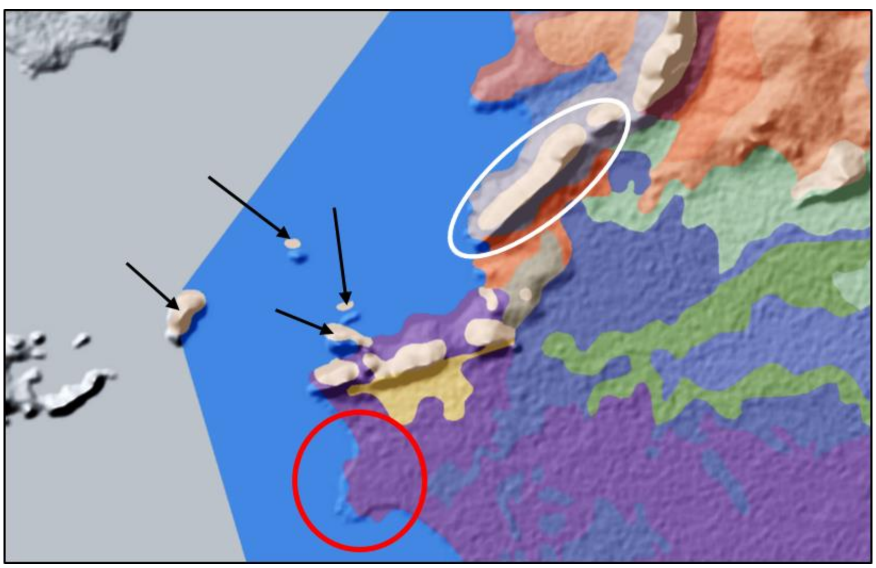


Figure 3.5: Using a hillshade to identify and correct inaccurately drawn soil map units. Black arrows show inaccurately drawn soil map units occurring in islands. Soil map unit within the red circle shows an incorrect boundary between that soil map unit and the water body. The soil map unit within the white boundary shows a map unit that is meant to represent soils of the hills. These discrepancies were manually adjusted by editing soil map unit boundaries in ArcMap version 10.6 (ESRI, 2018).

Quality Assessment

The guidelines described by (Forbes et al., 1987) were used to assess the quality of the legacy map data. These guidelines evaluate the scale, texture, and legend of the map.

Map scale and Texture

The minimum legible area for the Busia map was 40 ha and represents the smallest land area that can be represented on the map. The maximum location accuracy was 25 m meaning that the inherent uncertainty in the ground position of well-defined map points is 25 m. This directly affects the accuracy with which points on the ground may be represented. In order for a map scale to be adequate, the maximum location accuracy must be numerically smaller than the accuracy to which the user wishes to locate points on the ground and therefore depends on the intended uses of the survey. A well-defined ground point can be plotted to an accuracy of at best 0.25mm on the map sheet (Davis et al., 1981). The index of maximum reduction was 3.2. indicating that the map is very legible and that the scale of the map could be substantially reduced without impairing legibility.

Map Legend

Map units were explicitly labeled and categorized in the map legend (Table 3.2). The construction of the map unit legend indicated physiographic land types (such as hills, footslopes, uplands, etc.) based on physiographic photo-interpretation. These land types were further subdivided according to the underlying parent material on which the soils were developed, described as either the stratigraphy or just rocks such as dolerites, granites, etc. At the third level the map units were broken down and described based on important soil profile characteristics including drainage conditions, depth, color, consistency, texture, etc. (Rachilo and Michieka, 1991). This was then followed between brackets (dominant soils column in Table 1.2), by the classification of the main soils described according to the FAO/UNESCO nomenclature in the legend of the Soil Map of the World (FAO/UNESCO, 1974).

All the map units were explicitly described and therefore map units are considered to be 'adequately defined' because information within map units 'provides sufficient specific information relative to the land use so that the map unit's suitability may be determined' and 'are uniform in their suitability for the land use i.e. 85% of their total area will perform similarly for the use'. Map unit descriptions contained descriptions of acreage, agro-climatic zone, parent material, macro- and meso-relief, erosion, vegetation, land use, general soil description, color, texture, structure, consistence, chemical properties, clay mineralogy, diagnostic properties, and soil classification (Rachilo and Michieka, 1991).

Table 3.2: Samples of soil map legend tables of the Busia Area, at 1:100,000 scale.

Cartographic unit	Physiography	Geology	Soil depth (cm)	Soil characteristics	Dominant soils
HIP*	Hills	Igneous	0 – 50	Overlying hard rock	Lithosols (I)***
AA1**	River terraces and floodplains	Alluvium	-	-	Ferralic Arenosols (Qf) Chromic Vertisols (Vc)
PSb1	Plains	Sandstone	-	Brown	Orthic Acrisols (Ao) Orthic Ferralsols (Fo)
VXC2	Minor valleys	Various parent material	-	Complex	Ferralic Cambisols (Bf) Dystric Gleysols (Bd) Vertic Fluvisols (Jv)
UGr4M****	Uplands	Granite	0 – 50	Shallow and red, over petroplinthite	Rhodic Ferralsols (Fr)

* P – soils over hard rock.
**AA1 – integer number 1, 2, 4 indicate sequence of mapping units with almost identical features.
*** Major soil grouping (FAO, 1974).
**** ‘r’ – red soils at depth specified by the letter ‘M’ (M = shallow).

Integration of Remotely Sensed Data

Rossiter (2008) proposed the use of both satellite imagery and terrain attributes to adjust soil map unit boundaries. To ensure that each soil map unit captured their respective landscape features, satellite imagery and the hillshade were used to manually adjusting the polygon boundaries for each soil map unit. Satellite imagery proved useful in correcting soil map units that occurred on river terraces and swamps and for correcting boundaries between soil and water bodies, whereas the hillshade was used to adjust soil map unit boundaries occurring on islands and hills (Fig. 3.5). The result was an accurately georeferenced, digitized soil map of the Busia area with 348 polygons that belong to 52 different soil map units broadly grouped into eleven soil orders (Fig. 3.6). The majority of the soils are moderately deep-to-deep, yellowish red to reddish brown, non-calcareous and predominantly kaolinitic in clay composition with few weatherable minerals remaining and evidence of weak argillic horizon (Rachilo and Michieka, 1991).

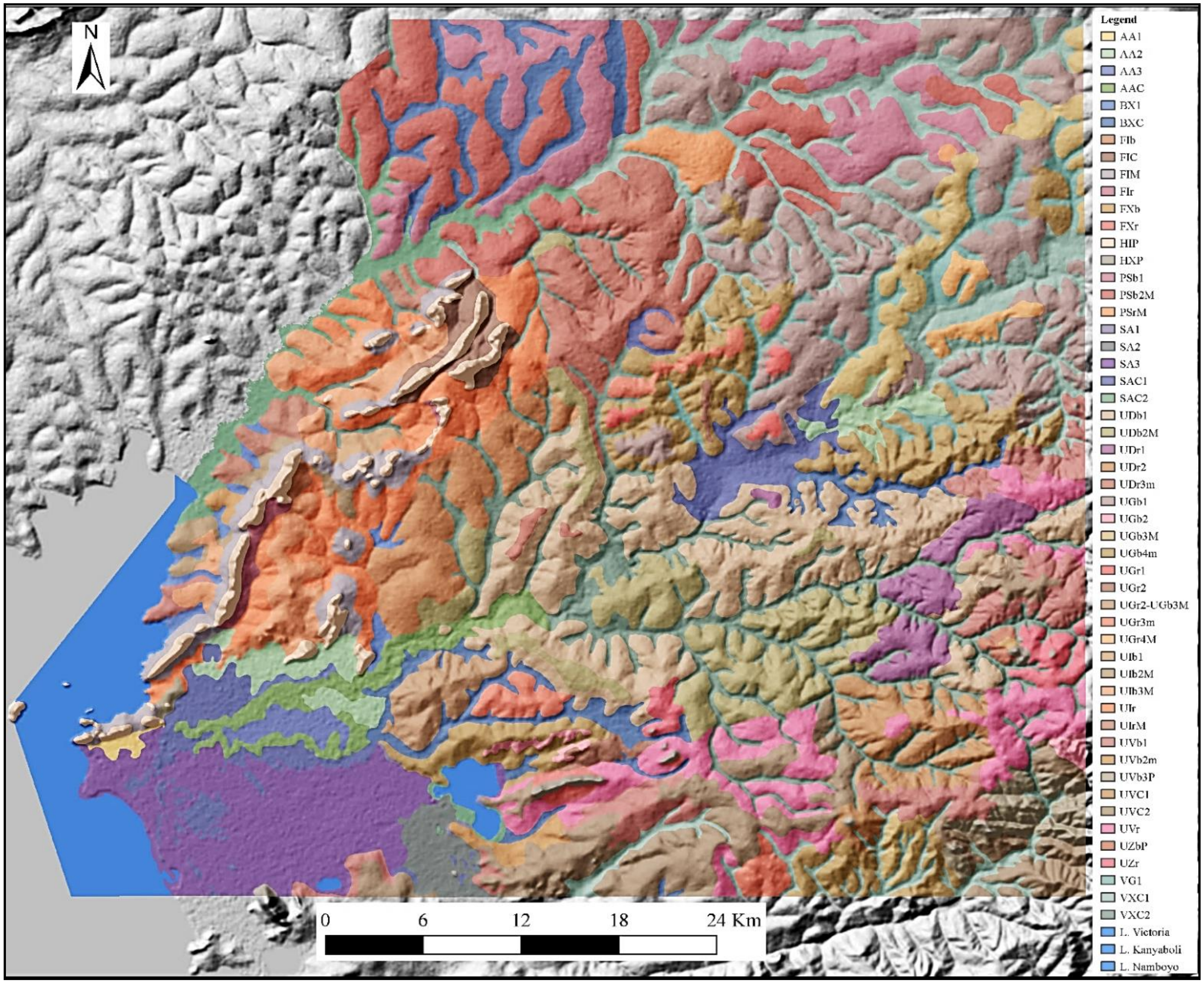


Figure 3.6: Digitized soil map of the Busia area. RGBs of soil map units similar to those used by Rachilo and Michieka (1991) were determined and used to reproduce the map and overlaid on the hillshade with transparency set to 45% for a 3-dimensional effect.

Metadata

Metadata included spatial data source, spatial reference, and processing steps. Table 3.3 shows the metadata stored internally as a text file for the renewed legacy soil map. This data is useful because it allows other users to access the data, evaluate its usefulness for their intended purposes, and assist other users to replicate renewal efforts for other legacy soil data.

Table 3.3: Metadata information in the GIS layer of the *Reconnaissance Soil Map of the Busia Area*.

Item	Detail	Description
Data source	Description	Data was obtained from the Kenya Soil Survey library.
ID information	Description	The soil map was created to enhance a systematic inventory of soil and land resources for multipurpose land use planning for the Busia area.
Spatial reference	Description	East Africa War System of the Coordinates Traverse Mercator projection belt I on the Arc 1960 datum. Reference ellipsoid was Clarke 1880.
Data quality	Processing steps	This GIS layer was created by (1) downloading the scanned map from the European Digital Archive of Soil Maps (EUDASM) (Panagos et al., 2011), then (2) georeferencing using satellite imagery and the hillshade as a base map. (3) The digital soil map was edited to ensure that polylines right on top of the soil line units' borders at very high magnification. (4) The attribute table was edited and repopulated with information from the legend and soil survey report.

3.4.5 Land Quality Maps

Careful interpretation of the Busia soil survey report showed that land quality maps could be generated depending on the agro-climatic zone within which a specific soil type occurs. We took, as an example of one land quality, the availability of moisture for plant growth, and demonstrate how to generate this land quality by interpreting legacy soil information in the Reconnaissance Soil Survey of the Busia area.

Adequate moisture is necessary for plant growth. For the Busia area, this land quality expresses the period of time that a plant has adequate available soil water to maintain its normal productive growth. It is measured in terms of the presence of humid months without limitations for the plant. The length of the accumulated growing months determined the suitability for a specific plant or crop. Suitability for specific plants was determined by the length of accumulated growing months (Rachilo and Michieka, 1991).

Three different rooting depths, 0-50, 0-80, and 0-120 cm, were used to determine the total easily available soil moisture storage capacity. This approach was explicitly described in the legacy soil survey report (Rachilo and Michieka, 1991). Soil moisture is climate dependent. The agro-climatic map of Kenya was used to map out the different agro-climatic zones in Busia area. Four different agro-climatic zones, zones I, II, III, IV, were mapped (Sombroek et al., 1982). Table 3.4 shows the relationship between soil depths and the length of the growing season for soils within agro-climatic zone one (I) only.

Table 3.4: Relationship of soil depth and length of growing season in agro-climatic zone I.

Soil depth (cm)	Growing months
0 - 50	11
0 - 80	11
0 - 120	$11\frac{1}{2}$

Source: Rachilo and Michieka, 1991.

The availability of moisture for plant growth was determined by two independent variables: soil type, herein soil map unit (Rachilo and Michieka, 1991), and agro-climatic zone (Sombroek et al., 1982). Spatial datasets of these two variables were joined using the Spatial Join function in the Overlay Tools within the Analysis Tools in ArcMap (ESRI, 2018). Since land quality ratings were determined by soil map units occurring within specific agro-climatic zones, Field Calculator was used to join the soil map units together with their respective slope classes, ‘SoilSlope’ and the agro-climatic zones within which they occur, ‘SoilCode’. For example, SoilCode Psb1/ABII (Fig. 3.7) represents soil map unit Psb1 of slope AB (0-5 %) and found in agro-climatic zone II. Rachilo and Michieka, (1991) worked out all the land quality ratings for each soil map unit occurring within the four agro-climatic zones for the Busia area (Fig. 3.8). A similar database was recreated using Microsoft Excel software version 2010 (Microsoft, 2010).

KEN_Busia_Join

Shape *	AgroZone	SOILMAPCOD	Slope	Shape_Length	Shape_Area	SoilSlope	SoilCode
Polygon	II	PSb1	AB	10823.137678	5115345.44545	PSb1/AB	PSb1/ABII
Polygon	II	UGb1	B	20626.940253	10413316.46892	UGb1/B	UGb1/BII
Polygon	II	VXC1	AC	18553.47732	2454692.115465	VXC1/AC	VXC1/ACII
Polygon	I	PSb1	AB	6721.366492	2070648.981915	PSb1/AB	PSb1/ABI

Figure 3.7: An excerpt from the attribute table resulted from the spatial join between the agro-climatic map of Kenya, ‘Agrozone’, and the Busia soil map ‘SOILMAPCOD’.

The Join and Relate function in ArcMap was used to join the recreated database with the digitized soil map of the Busia area. The SoilCode in the Busia soil map attribute table was used as a unique identifier to join with the recreated database Soil_Aczon unique identifier from Fig. 3.8. This resulted to a total of ten (10) different land qualities for the Busia area namely (i) temperature, (ii) availability of moisture for plant growth, (iii) availability of nutrients for plant growth, (iv) hazard of sodicity, (v) hazard of salinity, (vi) susceptibility to erosion, (vii) availability of oxygen in the root zone, (viii) flooding hazards, (ix) possibilities of seedbed preparation and cultivation (possibilities for the use of agricultural implements), and (x) availability of foothold for roots (Appendix A).

	A	B	C	D	E	F
1	MAPCODE	AgroZone	Soil_Aczon	AoM	Rating Criteria	Month(s) per gowing season:
2	PSb1/AB	I	PSb1/ABI	2	Moderate	9.5 - 11 months with plant available moisture
3	VXC1/A	II	VXC1/AII	2	Moderate	9.5 - 11 months with plant available moisture
4	VXC1/A	I	VXC1/AI	2	Moderate	9.5 - 11 months with plant available moisture
5	VXC2/AB	I	VXC2/ABI	1	High	>11 months with plant available moisture

Figure 3.8: An excerpt from the database of the availability of moisture. MAPCODE = individual soil-mapping units within Busia area. AgroZone = agro-climatic zone within which a soil map unit occurs. Soil_Aczon = similar to SoilCode in figure 7 above. AoM = land quality ratings. Rating Criteria = interpretations of land quality ratings. Month(s) per growing season = number of humid months without limitations for plant growth (Table 3.5).

Table 3.5: Ratings for the availability of moisture for plant growth.

Month(s) per growing season	Rating*
>11	1
9 - 5	2
6 - 9	3
4 - 5.5	4

* 1 = High, 2 = Moderate, 3 = Low; 4 = Very low.

Figure 3.9 is consistent with what we would expect on the landscape. Hills have low available moisture since soils are very shallow consisting of Lithosols with stony phases (Rachilo and Michieka, 1991). Conversely, river terraces and swamps have high available moisture since these are depositional areas. Having such a map is useful as it gives an idea of what crop types are best suited for such conditions. For instance, since hilly areas have very low available moisture for plant growth, forestry and/or grazing would be a suitable land use practice. On the other hand, plants require adequate available soil moisture for plant growth, but an excess may result cause root rot. Since soils in river terraces and swamps have a lot of available moisture, these areas can be used for ‘water loving’ plants such as rice.

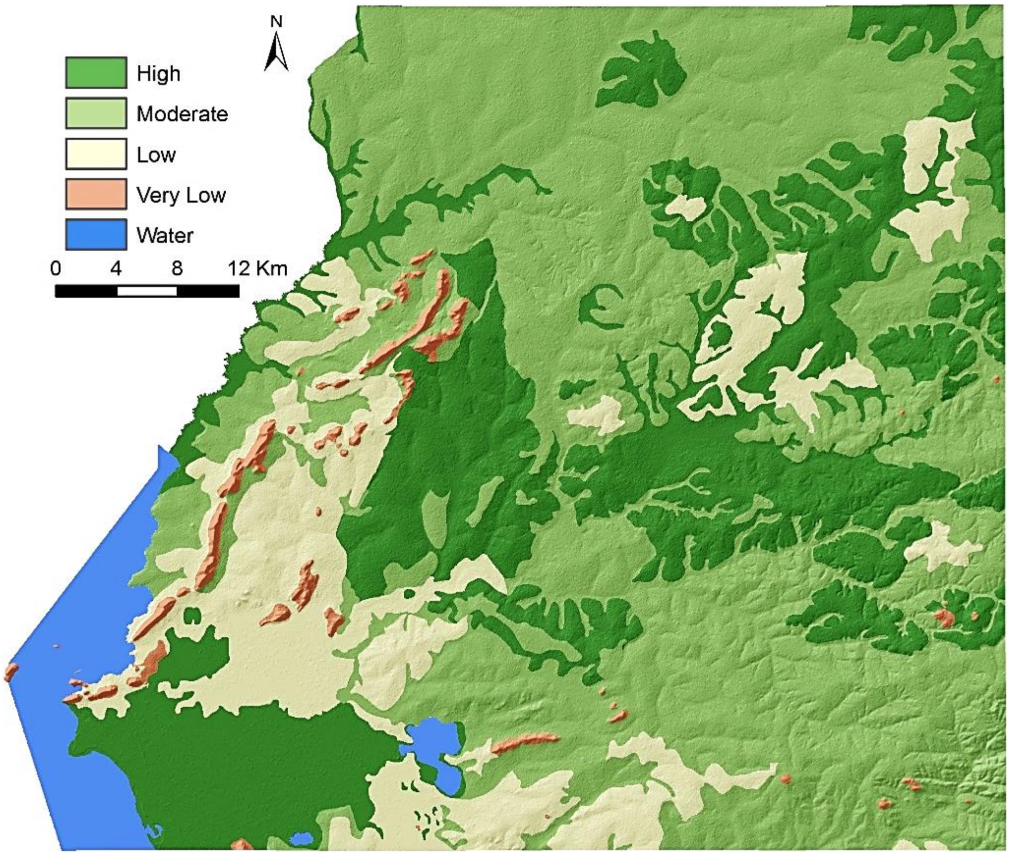


Figure 3.9: Availability of moisture for plant growth for the Busia Area.

Crop Suitability Maps

The ratings of all land qualities developed by Rachilo and Michieka (1991) for the Busia area were further used to determine suitability classes for specific crops. In this section, we use suitability for maize (*Zea mays* L.) as an example and demonstrate how land qualities are used to generate the maize crop suitability map. Maize was chosen because it is a staple crop within the study area and in Kenya (De Groote et al., 2010). Specific decision matrices developed for each soil map unit were used to rate Busia soils for the suitability for maize cultivation. A table of all the soil map units for the study area with their respective agro-climatic zones was generated in Microsoft Excel software version 2010 (Fig. 3.10).

1	SoilAEZ	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC	AoF
2	HIPI	3	4		1	1	1	1	1	4	3
3	HIPII	3	4		1	1	1	1	1	4	3
4	HIPIII	3	4		1	1	1	1	1	4	3

Figure 3.10: An excerpt from the recreated Busia land quality database. Temp = temperature. AoM = availability of moisture. AoN = Availability of nutrients. SH = Hazard of salinity. Sod = Hazard of sodicity. SE = Erosion hazard. Ox = Availability of oxygen. FH = Flooding hazard. SPC = Possibilities of seedbed preparation and cultivation. AoF = availability of foothold for roots. Numbers 1, 2, 3 and 4 represent land quality ratings while blanks indicate no data.

To determine specific suitability ratings for maize, decision matrices developed by Rachilo and Michieka (1991) were used to generate crop suitability classes for each soil map unit by utilizing *if then* statements created in RStudio software version 3.5.1 (RStudio Team, 2016). The Join and Relate function in ArcMap was used to combine the newly created suitability class database with the Busia soil map attribute table.

Table 3.6: Decision matrix for the suitability classification of soils for rainfed maize growing under the intermediate technology option.

Suitability Class	Land Qualities								
	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	1	1-2	2-3	1	1	1-2	1-2	1-2	1-2
Moderately Suitable (S2)	2, 3, 4	3	4	2-3	2-3	3-4	3-4	3-4	3-4
Marginally Suitable (S3)	5, 6	4	4	4	4	5	5	5	5
Unsuitable (NS)	7	4	4	4	5	5	5	5	5

Only two suitability classes for rainfed maize under intermediate technology, moderately suitable and unsuitable, were identified for the Busia area (Fig. 3.11) using the decision matrix in Table 3.6. Intermediate level of technology was defined as “that level of technology where certain inputs such as fertilizers, insecticides, and mechanized land preparation are used on a modest scale” (Rachilo and Michieka, 1991). Even though most of the arable land is still farmed at a traditional level of technology, the majority of farmers apply, to a certain extent, intermediate technology.

Over 85% of the Busia area is suitable for maize cultivation using the decision matrices set by Rachilo and Michieka (1991). Eighteen different crop suitability maps were generated including rainfed (1) sugarcane (*Saccharum officinarum*), (2) cabbages (*Brassica oleracea*), (3) kales (*Brassica oleracea*), (4) onions (*Allium cepa*), (5) tomatoes (*Lycopersicon esculentum* L.), (6) wetland and upland rice(*Oryza sativa*), (7) citrus guava (*Psidium guajava),* (8) cotton (*Gossypium hirsutum*), (9) groundnuts (*Arachis hypogaea*), (10) maize (*Zea mays* L.), (11) finger millet (*Eleusine coracana* L.), (12) cassava (*Manihot esculenta*), (13) common beans (*Phaseolus vulgaris*), (14) sunflower (*Helianthus annuus*), (15) Robusta coffee (*Coffea canephora* var. robusta), (16) forestry, (17) fodder crops, and (18) areas suitable for grazing (Appendix B).

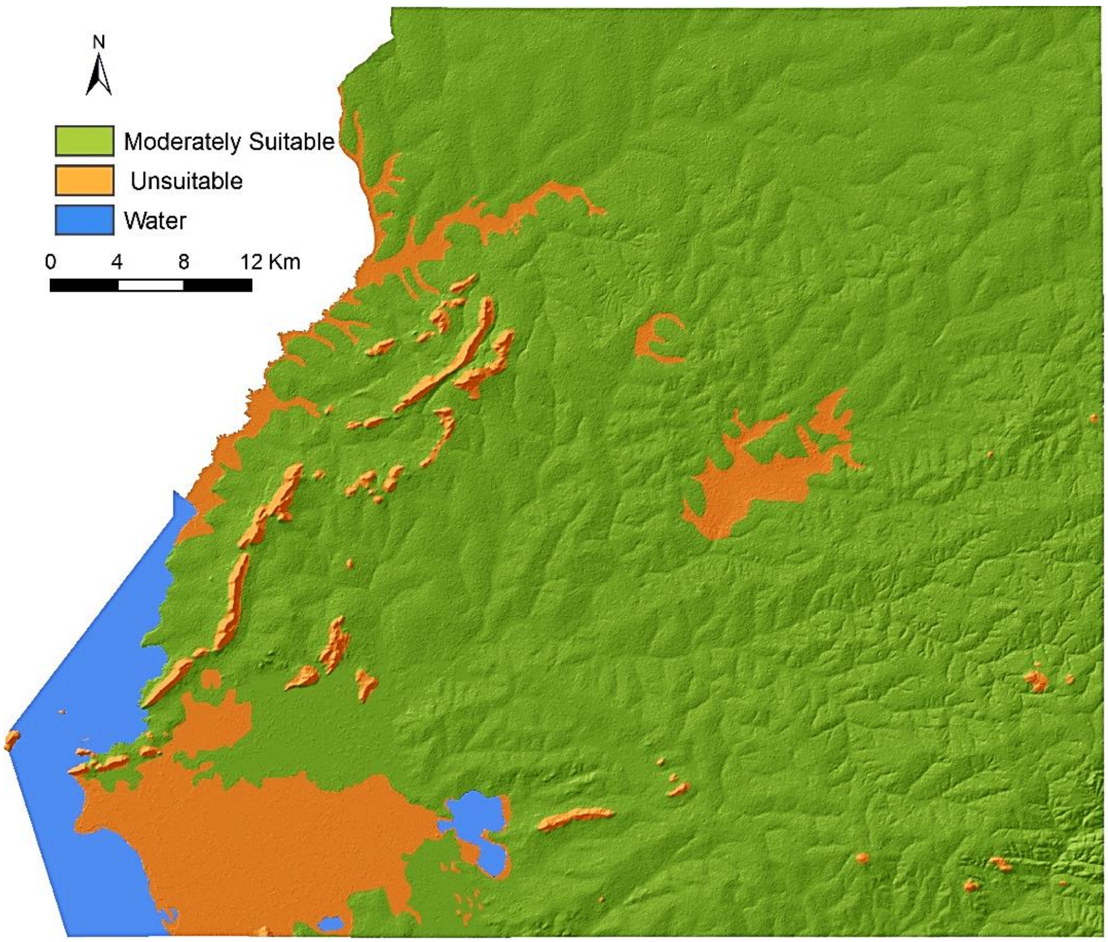


Figure 3.11: Maize (*Zea mays* L.) crop suitability for the Busia area.

3.5 Conclusion

The Reconnaissance Soil Survey of the Busia area was rescued and converted into digital format by taking advantage of key concepts of soil science and technological advancements made in GIS. Freely accessible data such as satellite imagery and terrain attributes generated from the NASA's 30m Shuttle Radar Topographic Mission (SRTM) were used to correct soil map unit polygon boundaries when editing the digitized soil polygon map. This legacy soil data was too information rich to be left moldering on the shelf. Specific decision matrices were used to generate a total of 10 land quality maps and 19 different crop suitability maps.

Rural smallholder farmers are keen to know what crops are best suited for their farm. We hope that by demonstrating how to mine useful agronomic data a from legacy soil survey, this effort will catalyze their use to provide sustainable solutions that can assist in addressing some of the most pressing agronomic challenges. Additionally, DSM can benefit from renewed legacy data as it provides the data such as soil observations and reality checks for validation in the form of interpreted polygons as well as the surveyor's concept of soil geography as revealed in the soil survey report.

CHAPTER 4. SPATIAL PREDICTION OF SOIL PROPERTIES OF THE BUSIA AREA, KENYA USING LEGACY SOIL DATA

Abstract

Legacy soil data (traditional soil maps, soil survey reports, soil survey manuals, land evaluation frameworks, soil profile descriptions, and farm management handbooks) often contains considerable agronomic information that can help revitalize agriculture in countries with poor soil spatial infrastructures. The objective of this study was to determine whether existing legacy soil data could be used to quantitively predict soil properties at a higher spatial resolution using digital soil mapping techniques without conducting additional field work. A dataset of 76 profile points mined from the Reconnaissance Soil Survey of the Busia area in western Kenya was used to compare three prediction models: (i) ordinary kriging, (ii) stepwise multiple linear regression (SMLR), and (iii) the Soil Land Inference Model (SoLIM) to predict soil organic carbon, clay, silt, and clay. Six principal components from twenty-three original covariates were used as predictors for SMLR. For SoLIM, six terrain attributes: (i) multiresolution ridgetop flatness (MRRTF), (ii) multiresolution valley bottom flatness (MRVBF), (iii) topographic wetness index (TWI), (iv) topographic position index (TPI), (v) planform curvature, and (vi) profile curvature were used to generate ideal soil types for soil property prediction. Predicted soil maps were at a resolution of 30 m, more suitable for smallholder farmers. From a statistical standpoint, ordinary kriging performed better than SoLIM and SMLR in predicting SOC (RMSE = 0.02), clay (RMSE = 0.32), and silt (RMSE = 0.10) whereas SMLR performed better than SoLIM and ordinary kriging for predicting sand content (RSME = 0.11). 95% C.I. ranges were narrowest from ordinary kriging, and widest from SMLR. However, from a pedological standpoint, SoLIM conformed better to soil

forming factors model compared to ordinary kriging, but had lower confidence interval ranges compared to SMLR.

4.1 Introduction

Legacy soil data is available for many parts of the world in the form of traditional soil maps, soil survey reports, soil survey manuals, land evaluation frameworks, soil profile descriptions, and farm management handbooks (Zinck, 1995; Arrouays et al., 2017). Legacy soil data has been, and is, widely used as the source of soil information to support soil conservation and environmental monitoring (McBratney et al., 2003; Odeh et al., 2012).

Legacy soil data, however, often remains idle in libraries, and the probability of this data being lost through natural, manmade, or political disasters, or simply neglect, is very high (Rossiter, 2008). The demand for soil data, however, is soaring (Cook et al., 2008), and this demand can often be met by legacy soil data. Legacy soil data may contain considerable agronomic information that, if accessible, can help revitalize agriculture in countries with poor soils spatial infrastructures (Zinck, 1995). The spatial distribution of soils, land quality, crop suitability, geo-located soil profile information, geology, and land use information that is often available in legacy soil data can be used as baselines for long-term studies to assess changes in soil properties (Bellamy et al., 2005), or to model temporal trends of soil quality and soil processes (Baxter et al, 2006). Legacy soil data can also be used as a primary input for digital soil mapping (DSM), especially for countries with sparse soil data infrastructures (McBratney et al., 2003). When financial resources are limited, legacy soil data is a reliable source of data for modelling soil variability without the need for major additional funding (Arrouays et al., 2017).

Although legacy soil data for much of Africa is sparse, Kenya, fortunately, has considerable soils information. Our visits to the Kenya Soil Survey in the spring of 2016 suggested that existing Kenyan legacy soil data was not used to its full potential. The majority of the data exist as paper maps and reports stored on library shelves, some data are in private collections of retired soil scientists, while digital formats seem to be used only internally.

The use of legacy data for DSM is determined by the type of data available. There are two broad groups of legacy soil data: (i) data existing as maps of soil types or soil properties, and (2) data existing as soil profile descriptions and tables of physical and chemical characterization data. When soil maps are the only source of data, they can be disaggregated and downscaled to produce new soil maps at a finer scale (McBratney, 1998) with the aim of mapping constituent classes of the soil map units individually (Thompson et al., 2010). On the other hand, when soil profile data is available, it can be used as inputs to many statistical, geostatistical, and machine learning methods to predict soil properties at unsampled locations (McBratney et al., 2003).

The work of the Africa Soil Information Service (AfSIS) is one example of the use of legacy soil data to map soil properties (Hengl et al., 2015). AfSIS developed the Africa Soil Profile Database (AfSP), a collection of rescued legacy soil profile data, together with additional soil data from 60 sentinel sites, for use in continent-wide digital soil mapping projects for Sub-Saharan Africa (Leenaars, 2012, 2013, 2014). The AfSP data has been used for the spatial prediction of an array of soil properties at a resolution of 250 m, or ~6.25 ha (Hengl et al., 2017). These maps are the first attempt to provide some level of detail for regional and/or local spatial distribution of soil properties to guide sustainable soil use and management decisions for rural smallholder farmers.

Two concerns still persist in the quest to predict soil properties using legacy soil data. First, it is not clear which of the various interpolation methods performs best with the very limited data that might be mined from legacy soil data. Second, the current spatial resolution of predicted maps is too coarse for practical land management for rural smallholder farmers in Africa whose landholdings are often less than a hectare (Sanchez et al., 2009).

Providing scientific answers to precision agriculture in quantitative terms is often very difficult because of lack of data (Council for Agricultural Science and Technology, 2019). This is a common problem in Africa where soils data is sparse, or in many cases nonexistent, making it difficult or impossible to make site specific recommendations on sustainable land use and management (Tully et al., 2015).

The objective of this study was to determine whether existing legacy soil data for a selected portion of Kenya could be used to quantitatively predict soil properties at a higher spatial resolution than current maps by using DSM techniques without conducting additional field work. We chose to evaluate three DSM techniques, ordinary kriging, stepwise multiple linear regression (SMLR), and the Soil Land Inference Model (SoLIM), to determine which would produce the “best” soil property maps. These techniques require two types of inputs: (1) point data from quantitatively analyzed soil profiles, and (2) rasterized spatial data derived from elevation models, remotely sensed imagery, and climate models.

4.2 Materials and Methods

4.2.1 Geographical Setting

This study focuses on the Busia area of western Kenya. It encompasses an ~30' x 30' quadrangle bound by the equator to the south, latitude 0° 30' N to the north, longitude 34° 30' E to the east, and the Kenya-Uganda border to the west (Fig. 4.1). The study area is comprised of about 279,800 hectares including Lake Namboyo, Lake Kanyaboli and part of Lake Victoria. The elevation ranges from 1127 on the shores of Lake Victoria to 1564 m in the Samia Hills. Mean annual rainfall increases from 925 mm in the southwest around Lake Victoria, to 1990 mm in the northeast. Average annual temperature ranges from 20 to 23° C. The Busia area was selected for this study because: (1) it has accessible legacy soil data at a scale of 1:100,000 (Rachilo and Michieka, 1991), (2) agriculture is the main economic activity (Onywere et al., 2011), (3) high population and poverty densities have strained existing natural resources (Kenya National Bureau of Statistics, 2010), and (4) we are personally familiar with the area and can draw on our own field observations for additional context.

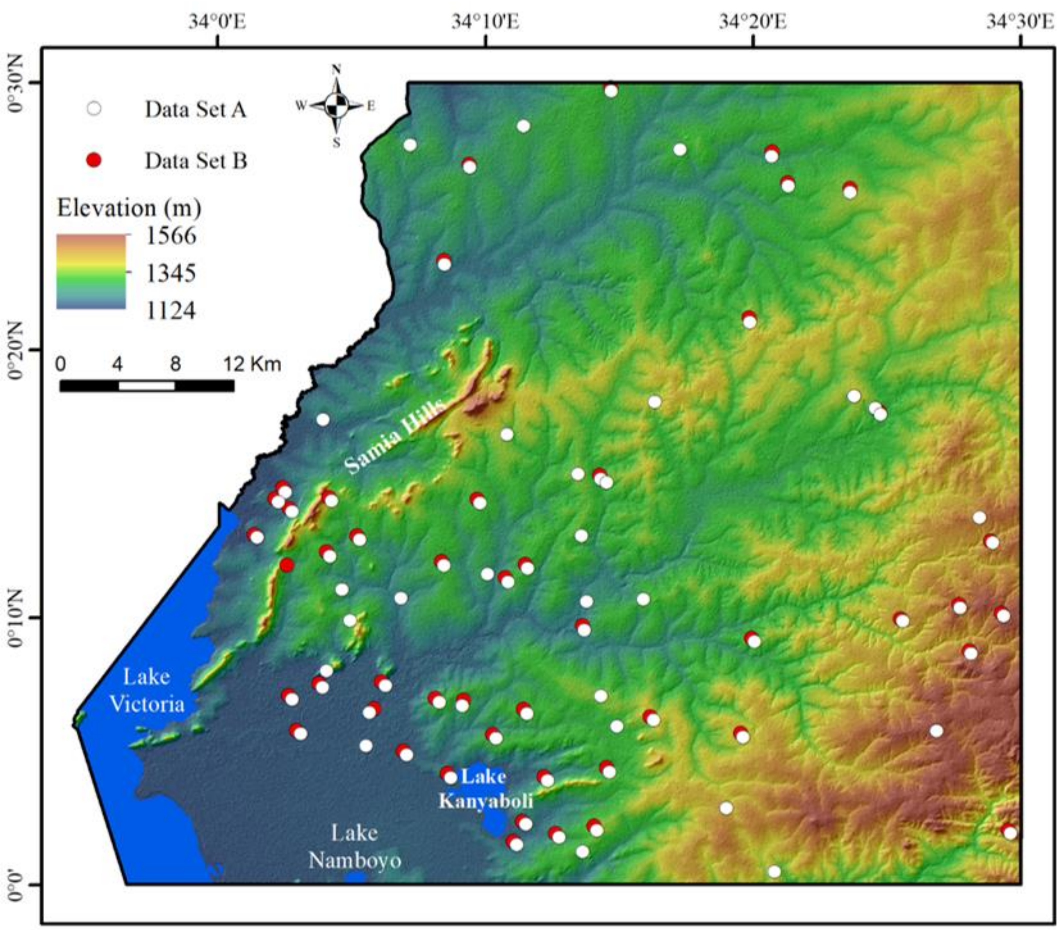


Figure 4.1: Geographical location of the Busia area.

4.2.2 Spatial Prediction Framework

The spatial prediction framework is summarized in Fig. 4.2 and consisted of data assembly, spatial prediction by ordinary kriging, spatial prediction by stepwise multiple linear regression analysis, spatial prediction using the Soil Land Inference Model (SoLIM), and a common evaluation framework. Although evaluation of the spatial predictions occurs at the end, we will describe the common evaluation framework first, followed by detailed descriptions of each of the prediction models.

4.2.3 Data Sources

Soil Profile Data Sources

Soil data were mined from the *Reconnaissance Soil Survey Report of the Busia Area (quarter degree sheet No. 101)* (Rachilo and Michieka, 1991). The available data consist of: (1) a soil map of the area that shows the spatial distribution of the soils at a scale of 1:100,000, (2) detailed soil chemical and physical laboratory data, (3) a land evaluation key used for determining the land use, and (4) a soil engineering map for determining development activities.

The data in this report are the result of fieldwork carried out by the Kenya Soil Survey mostly during October 1980 by two soil survey teams. Auger borings and profile pits were dug to depths ranging between 120 and 200 cm or to refusal, depending on the depth to the parent material. The sites were selected based on photo-interpretation and possible changes in land and/or soil characteristics as observed in the field. Soil profile descriptions were prepared according to the *Guidelines for Soil Profile Description* (FAO, 1977) and horizons were sampled for later chemical and physical analyses in the laboratory. Land and soil properties were recorded on standard soil profile description forms following recommendations of the Kenya Soil Survey. Two sets of soil chemical and physical laboratory data were available.

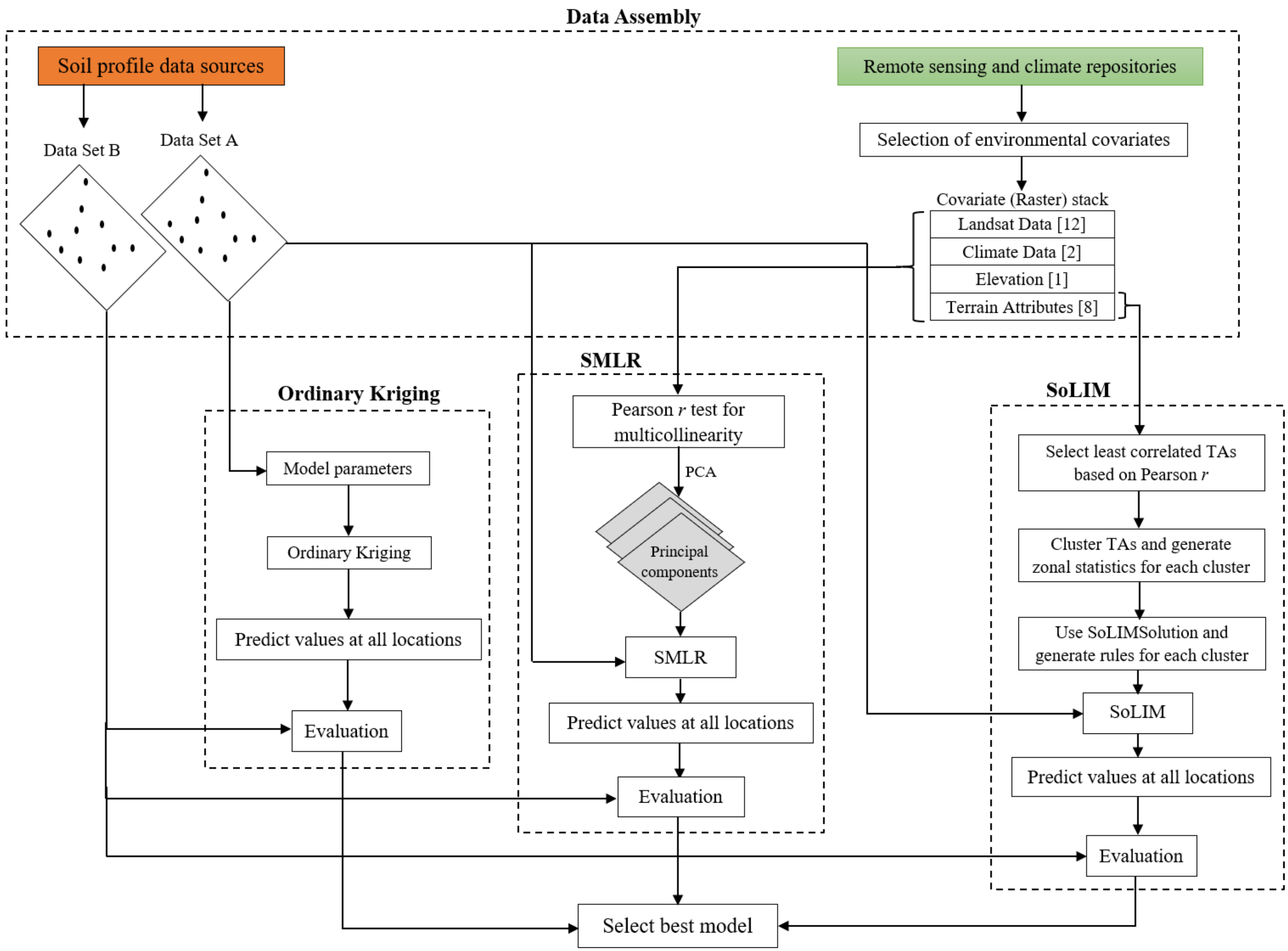


Figure 4.2: Schematic representation of legacy soil data-driven spatial prediction framework used for predicting soil properties.

Data Set A

(1) Data Set A consists of data for 76 georeferenced locations for which the A and B horizons were sampled, presumably by auger borings (Appendix 4 in Rachilo and Michieka, 1991). Composite samples of the A horizon were collected between depths of 0 - 30 cm for soil fertility analysis, while the B horizon was sampled to an unspecified depth. Soil properties recorded include: texture, Munsell color, structure, consistence, the presence or absence of clay cutans, clay type, bulk density, porosity, soil classification according to both the FAO-UNESCO Soil Map of the World and Soil Taxonomy (FAO/UNESCO, 1974), soil organic carbon (SOC), base saturation (at pH 7 and 8.2), exchangeable sodium percentage at pH 8.2 (ESP), and electrical conductivity (Rachilo and Michieka, 1991). The laboratory procedures for Both Data Set A and B are summarized in Table 4.1. For more details see Hinga et al. (1980). Of these properties, data for SOC for the A-Horizon, texture of the A and B horizons, B/A clay ratio, and soil classification were available for all of the 76 locations. Latitude and longitude for the sampling points were extracted from the scanned and georeferenced map, titled 'location of soil profile pits and augerings' in the survey report (Fig. 11 in Rachilo and Michieka, 1991).

Data Set A Descriptive Statistics

On average, SOC for the study was low and ranged between 0.18% and 2.86%. Ferralsols were the most common soils in the study area. The soils with relatively high SOC included (i) Nitosols, (ii) Cambisols, (iii) some soil mapping units found in Acrisols and Luvisols, (iv) some soil mapping units found in terraces, minor valleys, bottomlands, and swamps, which include Fluvisols, Gleysols, Histosols, Vertisols, and Solonchaks. Those with relatively low SOC include Acrisols, Fluvisols, Solonchaks, and Lithosols. On average, clay and sand contents were higher than silt content and clay content increased with depth as shown by the B/A clay ratio (Table 4.1).

Table 4.1: Descriptive statistics of selected soil properties for the A and B horizons of Data Set A.

Horizon	Soil Parameter	Mean	Minimum	Maximum	Std[1]	CV (%)[2]	Skew
A-Horizon	SOC (%)	1.16	0.18	2.86	0.54	46.6	0.68
	Clay (%)	43.0	8.0	78.0	16.1	37.4	0.14
	Silt (%)	20.1	4.0	39.0	9.0	44.8	-0.01
	Sand (%)	36.9	6.4	80.8	17.7	48.0	0.36
B-Horizon	Clay (%)	51.0	8.0	84.0	18.2	35.7	-0.23
	Silt (%)	16.1	3.6	39.6	7.8	48.4	0.64
	Sand (%)	32.9	6.4	76.6	18.5	56.3	0.45
	B/A Clay Ratio	1.21	0.4	1.8	0.2	17.8	-0.24

SOC = Soil organic carbon; [1] = Standard deviation; [2] = Coefficient of variation.

Data Set B

(2) Data Set B consists of detailed soil profile descriptions and analytical data for 48 georeferenced profiles presumably sampled from pits (pages 158 to 256 in Rachilo and Michieka, 1991). Fifteen soil properties were provided including pH (water and KCl), electrical conductivity, SOC, total N, CEC at pH 8.2, Ca, Mg, K, Na, exchangeable acidity, sum of cations, base saturation, texture, P, and Mn at different soil horizon depths. Latitude and longitude (in West Africa War System of Coordinates Traverse Mercator projection, belt I on the Arc 1960 datum) are recorded with the soil profile descriptions (see page159 in Rachilo and Michieka, 1991 as an example).

Fitting Mass Preserving Splines

The two datasets cannot be directly compared because Data Set A was sampled as a composite from 0 – 30 cm, while Data Set B was sampled by horizons with variable thickness. An equal area quadratic smoothing spline function (Ponce-Hernandez et al., 1986; Bishop et al., 1999; Malone et al., 2009) in the *ithir* package in R statistical software (Malone et al., 2017) was used to predict soil attributes for Data Set B over the interval 0 – 30 cm, allowing Data Set B to be compared with Data Set A.

Table 4.2: Soil laboratory analyses methods for the Busia soils (source: Rachilo and Michieka, 1991).

Soil Property	Method
Texture	Hydrometer method (Bouyoucos, 1962).
pH - H_2O or a 1*M* KCl	1:2.5 soil-water/salt suspension measured with a pH-meter.
Electrical conductivity ($EC_{2.5}$)	1:2.5 soil-water suspension. Soils with an $EC_{2.5}$ of over 0.85 mmhos/cm at 25°C, a saturation extract was prepared and the pH and EC were measured in the saturation extract.
Soil organic carbon	Walkley and Black method (Walkley and Black, 1934).
% N (on A-horizon only)	Semi-micro Kjeldahl method (Fawcett, 1954).
Cation exchange capacity (CEC)	Successive leaching of the sample with 1N ammonium acetate at pH 7.0 solution, 95% ethyl alcohol, 1N sodium acetate at pH 8.2 solution, 95% ethyl alcohol and lastly with ammonium acetate at pH 7.0 solution. CEC was determined in the last leachate by measuring the Na concentration with a flame-photometer (Tucker, 1974).
Exchangeable cations (Ca, Mg, K and Na)	Determined in the first ammonium leachates by flame photometer/atomic absorption spectrophotometer, respectively in the presence of a lanthanum chloride solution for Ca and Mg.
Exchange acidity (Hp)	Determined titrimetrically in an unbuffered 0.6 N barium chloride solution.
Mass analysis for available nutrients (on A-horizon only)	Soils were extracted by shaking for 1 hour at a 1:5 ratio with 0.1 N HCl/0.025 N H_2SO_4. Then Ca, Mg, K and Na determined with a flame-photometer, after an anion resin treatment for Ca was done. Mg, P and Mn were determined calorimetrically.
Mineralogical analysis	0.2μm fraction separated after pretreatments with H_2O_2 and dispersion with $(NaPO_3)_6$. X-ray diffraction analysis was carried out on samples saturated with Mg and K, using standard clay minerals for semi quantitative estimation. Peak area ratios, rather than peak height ratios were considered. Techniques involving solvation with ethylene glycol and heating to 500°C were employed as well. Apparatus - Phillips direct recording X-ray diffractometer, using copper, Kα radiation (Theisen et al., 1962, 1964).

Remote Sensing and Climate Repositories

Selection of Environmental Covariates

A raster stack of environmental covariates primarily based on remote sensing was used for predicting soil property maps (Fig. 4.3). These covariates were selected to represent the factors of soil formation and included the following. (1) The 1 Arc-Second (30 m) NASA shuttle Radar Topographic Mission (SRTM) global elevation dataset downloaded from the USGS Earth Explorer (USGS, 2017a). (2) Eight terrain attributes (TAs), namely, Multiresolution Index of Valley Bottom Flatness (MRVBF), Multiresolution Ridge Top Flatness (MRRTF), plan curvature, profile curvature, relief intensity, slope, topographic position index (TPI), and SAGA wetness index (TWI), all based on the 30 m SRTM elevation dataset. (3) Annual average precipitation and temperature derived from monthly averages over the years 1970 – 2000 obtained from WorldClim ver. 1 at https://www.worldclim.org, all at a spatial resolution of ~1 km (Hijmans et al., 2005). (4) All eleven bands from Landsat 8 OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor) level 2 images of identification number LC08_L1TP_170060_20171226_20180103_01_T1 of path 170 and row 60 with an acquisition date of 26 December 2017 downloaded from the USGS Earth Explorer (USGS, 2017b). These two sensors provide seasonal coverage of the global landmass at a spatial resolution of 30 m for the visible, near infrared, and short wavelength infrared; 100 m for the thermal infrared sensor; and 15 m for the panchromatic band. (5) The normalized difference vegetation index (NDVI) calculated from the Landsat 8 level two data as a proxy for the vegetation (organisms) soil forming factor. All soils data, Data Set A and Data Set B, environmental covariates, and intermediate rasters used in this study are available for download from Minai and Schulze (2019).

These twenty-three environmental covariates were projected to the WGS84 Web Mercator (Auxiliary Sphere) coordinate system. This coordinate system was deemed fit for this study due to its ability to preserve the shape of the area without significant distortion especially around the equator.

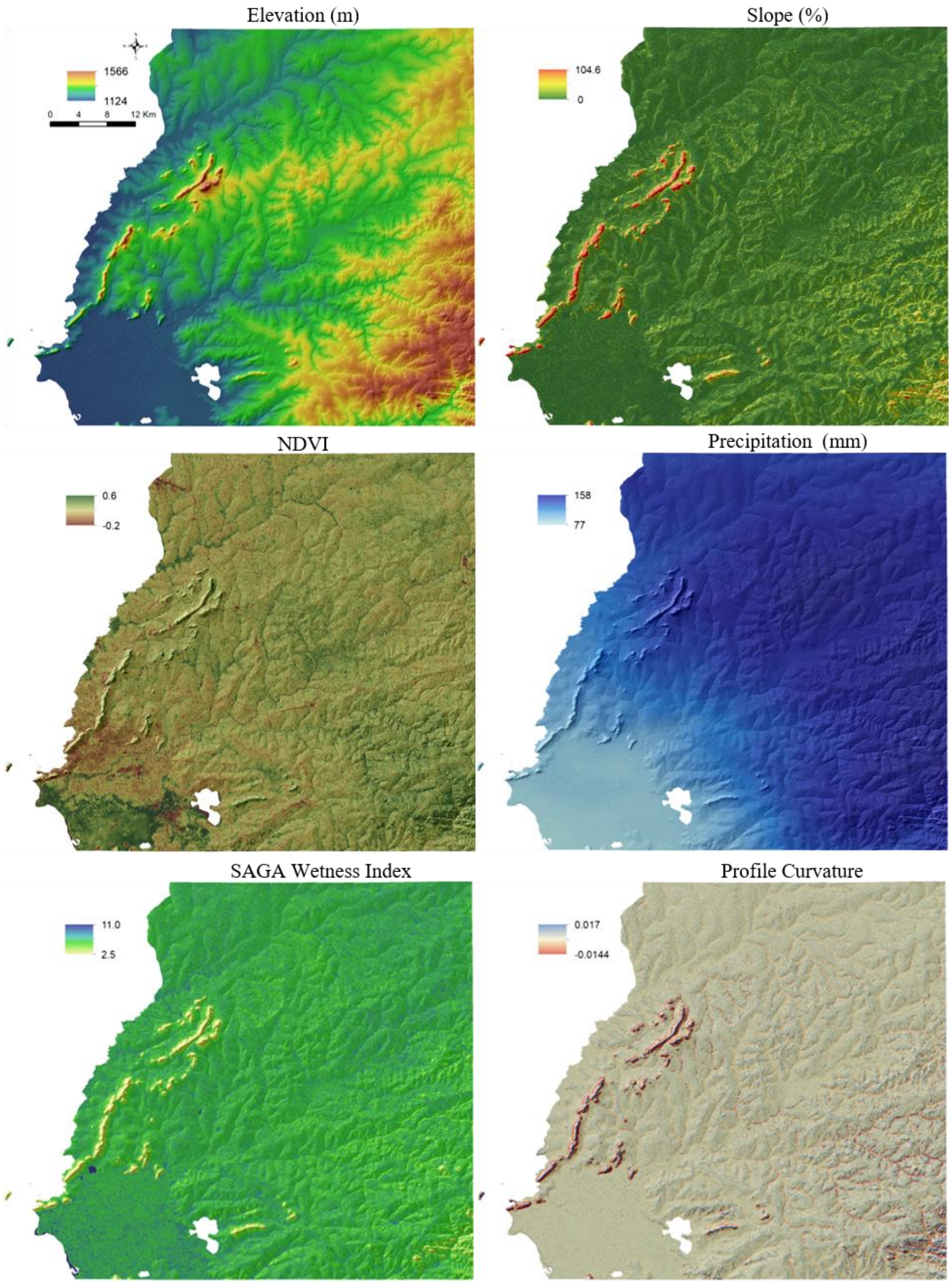


Figure 4.3: Examples covariates used to generate the soil property maps.

Data Processing

Study Area Buffer

The *Reconnaissance Soil Map of the Busia Area* was used to create a buffer of the study area.

Elevation

The DEM was preprocessed in ArcGIS 10.6. The Focal Statistics tool using a 5 x 5-pixel rectangular neighborhood around the central pixel was used to recalculate the mean of each pixel to reduce any noise and produce a smoother elevation surface. This study area buffer was used as a mask to extract the elevation raster. TAs indicated in section 4.2.3.2.1 were calculated using the System for Automated Geoscientific Analyses (SAGA-GIS) (Conrad, 2015).

Climate Data

The extract by mask tool in ArcGIS 10.6 was used to extract World Climate data using the buffer for the study area. These raster data were resampled using the bilinear interpolation resampling technique in ArcMap version 10.6 to a resolution of 30 m.

Landsat Data

The study area buffer was used to extract Landsat data using the extract by mask tool in ArcGIS 10.6. The resulting rasters were resampled using the bilinear interpolation resampling technique in ArcMap version 10.6 to a resolution of 30 m. The normalized difference vegetation index (NDVI) data was calculated using the raster calculator tool in ArcGIS 10.6 using the equation below:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}} \qquad [4.1]$$

where NIR is the near infrared band (Landsat band 5) and Red is the red band (Landsat Band 4).

4.2.4 Evaluation of Prediction Accuracy

The prediction accuracy for each numerical model was evaluated by cross-validation by looking at the differences between the observed and predicted values. The root mean square error (RMSE) (Eqn. 4.2), bias (Eqn. 4.3), correlation coefficient (r) (Eqn. 4.4), and concordance (ρ_c) (Eq. 4.5) performance statistics were computed:

$$\mathrm{RMSE} = \frac{\sqrt{\sum_{i=1}^{n}(\mathrm{pred}_i - \mathrm{obs}_i)^2}}{n} \qquad [4.2]$$

$$\mathrm{Bias} = \frac{\sum_{i=1}^{n}(\mathrm{pred}_i - obs_i)}{n} \qquad [4.3]$$

$$r = \frac{\sum_{i=1}^{n}(\mathrm{obs}_i - \overline{obs})(\mathrm{pred}_i - \overline{\mathrm{pred}})}{\sqrt{\sum_{i=1}^{n}(obs_i - \overline{\mathrm{obs}})^2}\sqrt{(\mathrm{pred}_i - \overline{pred})^2}} \qquad [4.4]$$

$$\rho_C = \frac{2\rho\sigma_{pred}\sigma_{obs}}{\sigma^2{}_{pred} + \sigma^2{}_{obs} + (\mu_{pred} - \mu_{obs})^2} \qquad [4.5]$$

where μ_{pred} and μ_{obs} are the means of the predicted and observed values respectively and σ^2_{pred} and σ^2_{obs} are the corresponding variances, ρ_c is the correlation coefficient (r) between the predictions and observations, and n equals the number of validation points, 48.

The uncertainty of the map prediction was estimated using the 95% confidence interval.

$$\mathrm{C.I.} = \bar{x} \pm z * \frac{\sigma}{\sqrt{n}} \qquad [4.6]$$

where $\bar{x}$ is the sample mean, *z* is the *z* score, σ is the standard deviation, and *n* is the sample size. The *z* score is dependent on the confidence interval range and sample size and is 1.96 for a C.I. of 95%.

4.3 Soil Prediction Models

The three prediction models were all derived from the basic s*corpan* model (Eqn. 2.2) used to predict selected soil properties. Data Set A was used for model fitting because of its larger sample size, whereas Data Set B was used for model evaluation. Spatial prediction was made for SOC and soil texture (percent sand, silt, and clay) for the A-horizon (0 - 30 cm) because only these two soil properties had complete data for all 76 locations in Data Set A.

The sections that follow will be ordered by the prediction model and discuss the preprocessing of the necessary data, the prediction model, the prediction results, and prediction evaluation. Finally, the 'best' prediction model will be selected.

4.3.1 Ordinary Kriging

Data Set A was the sole input for ordinary kriging. The data for SOC and sand, silt, and clay for the A horizon were subjected to the square root transformation to ensure they followed a normal distribution.

The Ordinary Kriging Model

Ordinary kriging is a prediction model whereby soil properties are predicted using the spatial arrangement of measured soil properties (Tobler, 1970). Ordinary kriging uses the equation

$$Z(x_0) = \sum_{i=1}^{n} \lambda_i . \mathrm{Z}(x_i) \qquad [4.7]$$

where $Z(x_0)$ is the soil property to be predicted at an *i*th location, λ_i is the unknown weight for the measured soil property at the *i*th location, x_0 is the prediction location, and n is the sample size. λ_i, depends on the: (1) fitted model to the measured points, (2) distance to the prediction location, and (3) spatial relationships among the measured values around the prediction location. The spatial dependency of soil properties was first modeled by describing the spatial structure of sampled points using the geostatistical technique of semi-variogram analysis. Calculation and modelling/fitting of the spatial structure was followed by kriging estimates for unsampled points and was performed using the variogram estimation and spatial prediction plus error (VESPER) software version 1.62 (Minasny et al., 2005).

Ordinary Kriging Results

Three spatial prediction structures were deemed fit for ordinary kriging soil property prediction: Gaussian, spherical, and exponential (Table 4.3).

Table 4. 3: Spatial prediction structures for ordinary kriging of selected soil properties for the A-horizon

Soil Property	C_0	C_1	$C_0/(C_0+C_1)$	Range (m)	RMSE	AIC	Model
SOC	0.3	103	0.0025	2,566,598	0.05	-67.8	Gaussian
Clay	100.2	160	0.3843	12,561	34.59	332.2	Spherical
Silt	68.5	10,000	0.0068	22,099,722	10.75	259.7	Exponential
Sand	242.5	10,000	0.0237	4,343,753	56.85	363.0	Exponential

SOC = Soil organic carbon; C_0 = Nugget; C_1 = Partial sill; RMSE = Root mean square error; AIC = Akaike information.

The predicted maps from ordinary kriging showed very little detail in terms of soil property variability (Fig. 4.4a and Fig. 4.5a, d, and g) and this is reflected by the fact that the overall ranges of the spatial dependence were very wide (Table 4.3). Ordinary kriging results underestimated the prediction of soil properties. The prediction ranges were 0.26% for SOC, 46.6% for clay, 3.4% for silt, and 18.6% for sand compared to Data Set A ranges which were 2.68% for SOC, 70.0% for clay, 74.4% for silt, and 35.0% for sand.

Ordinary Kriging Evaluation

Ordinary kriging is a special case of the *scorpan* model, where only the *n* (location) factor is considered. Predictions are made by modelling the spatial dependence between neighboring observations as a function of their distance only. The underestimation observed in kriging is due to the poor estimation of the spatial prediction models (Table 4.3). The ranges for SOC, silt, and sand were very large, indicating that the spatial structures do not effectively capture the variation within the study area. Ordinary kriging relies heavily on the data quality especially density and distribution. If the points used for modelling are biased, for instance if they do not capture the soil and landscape variability, ordinary kriging prediction might perform poorly (Hengl, 2009). For effective ordinary kriging, the study area must have an adequate and even distribution of data points for variogram modelling and point pairs must be available at various spacings. If the points represent only a portion of the study area, such as the case of the Busia area where most of the points occurred in the south western part of the study area (Fig.4.1), poor estimation of the model and spatial prediction will be expected. This explains the underestimation observed in kriging (Table 4.4).

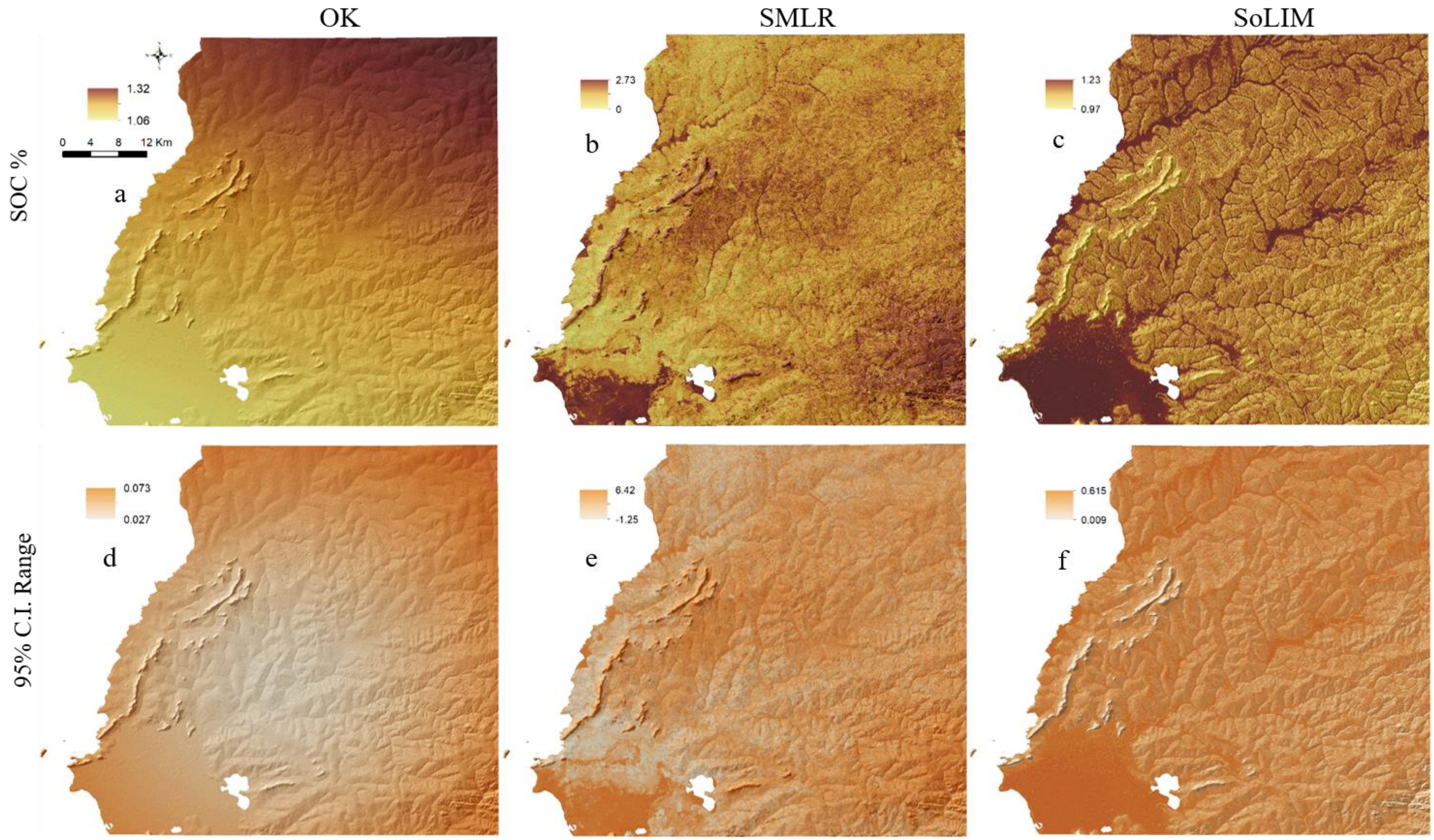


Figure 4.4: Predicted SOC and 95% C.I. ranges from three prediction models: OK = ordinary kriging, SMLR = stepwise multiple linear regression, and SoLIM = soil and landscape inference model.

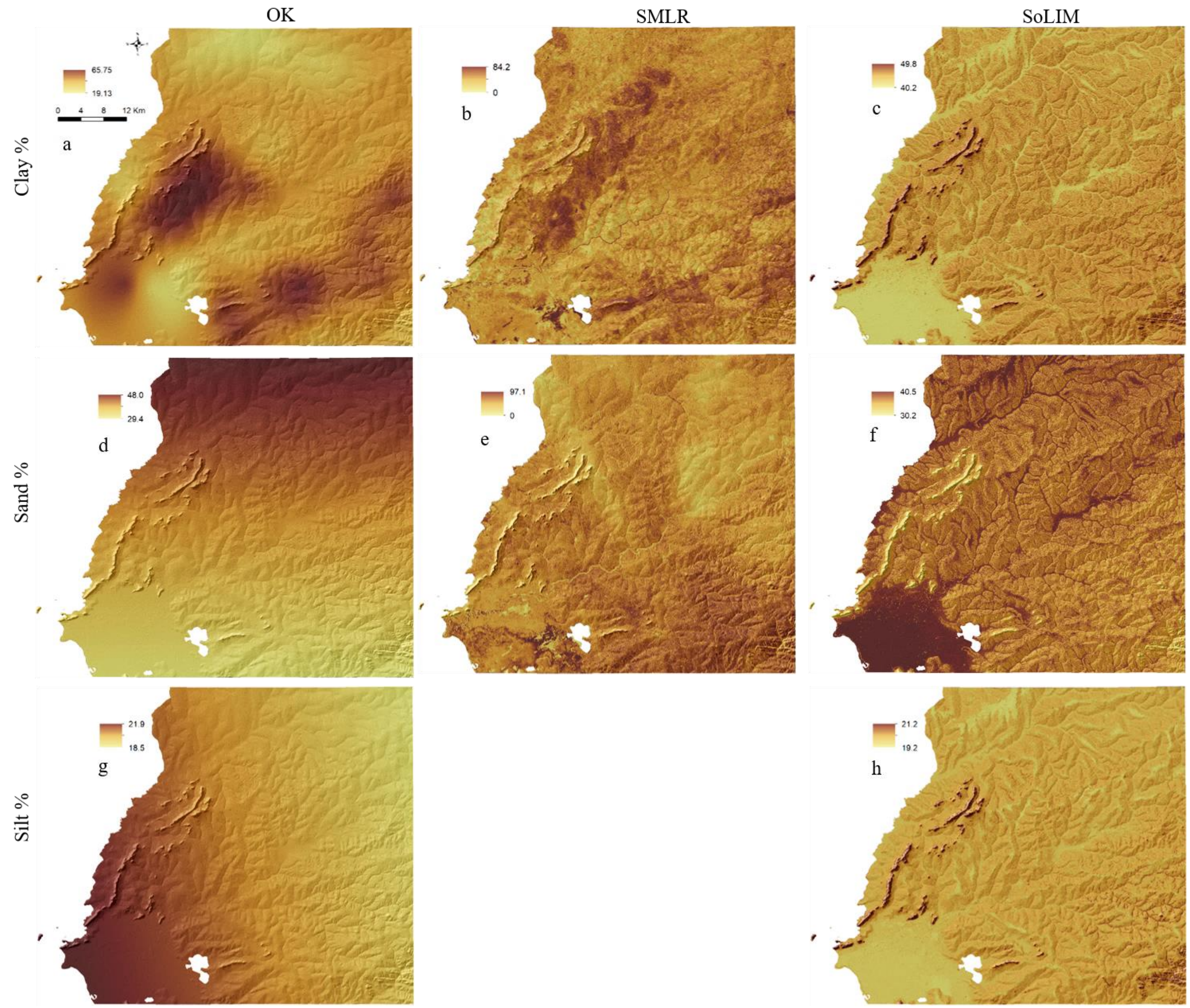


Figure 4.5: Predicted soil texture for the A horizon from three prediction models: OK = ordinary kriging, SMLR = stepwise multiple linear regression, and SoLIM = soil and landscape inference model.

Table 4.4: Prediction model evaluation.

Soil Property	Prediction Model	R^2	Concordance	RMSE	Bias
SOC	OK	0.09	0.07	0.02	0.002
	SoLIM	0.03	0.011	0.52	0.07
	SMLR	0.11	0.255	1.05	0.15
Clay	OK	0.80	0.797	0.32	-0.05
	SoLIM	0.09	0.010	17.1	-2.5
	SMLR	0.17	1.948	23.6	-4.1
Silt	OK	0.02	0.023	0.10	-0.01
	SoLIM	0.04	0.001	10.48	-1.51
Sand	OK	0.12	0.16	0.32	0.05
	SoLIM	0.0012	0.0004	26.98	3.89
	SMLR	0.0013	0.019	0.113	0.016

SOC = soil organic carbon; OK = ordinary kriging; SoLIM = soil land inference model; SMLR = stepwise multiple linear regression; RMSE = rot mean square error.

The variants of ordinary kriging that incorporate both the deterministic and stochastic components include cokriging and regression kriging. In comparison to other statistical and geostatistical models, Bishop and McBratney (2001) demonstrated regression kriging to be superior. Cokriging and regression kriging were not used in this study because of the low sampling density of the Busia calibration dataset (Fig. 4.1).

4.3.2 Stepwise Multiple Linear Regression

Pre-processing of Environmental Covariates for SMLR

Pearson Test for Multicollinearity

Most prediction models assume that predictor variables are independent of each other (Neter et al., 1996). Therefore, before environmental covariates could be used as predictors for stepwise multiple linear regression, there was need to account for multicollinearity effects between covariates (Hengl, 2009). Pearson correlation between 12 randomly selected covariates from the 23 covariates in the raster stack showed that the covariates were not entirely independent (Table 4.5). Many approaches have been proposed and used for multivariate soil prediction mapping to optimize sample locations and to ensure that landscape variability is captured in the sampling scheme (Hengl et al., 2004, 2007; Hengl, 2009; Vašát et al., 2010). However, to ensure that multivariate covariates are independent of each other, Gobin (2000) used principal components instead of the original environmental covariates as predictors to improve on the prediction for soil-landscape modelling.

Principal Component Analysis

Accordingly, all the original 23 environmental covariates within the raster stack were subjected to a standardized principal component analysis (PCA) to generate a small number of standardized linear combinations that capture most of the variation within the raster stack as a whole (Crawley, 2012). RStudio version 3.5.1 was used to conduct a standardized PCA using the *RStoolbox* package (Leutner and Horning, 2017). Standardization used a Z score expressed as:

$$Z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad [4.8]$$

where Z_{ij} is the Z score of pixel i in the covariate layer j, x_{ij} is the untransformed value of pixel i of covariate layer j, μ_j is the mean of the covariate layer j, and σ_j is the standard deviation of the covariate layer j.

The Kaiser rule, which recommends that PCs with eigenvalues greater than one should be retained (Kaiser, 1958), was used to determine the appropriate number of PCs. Six of the twenty-three PCs accounted for 80.9% of the variation within the raster stack (Table 4.6) and were retained (Fig. 4.6).

Table 4.5: Pearson correlation coefficient matrix between some selected soil covariates at 76-soil profile pit locations (Data Set A).

Covariate	DEM	MRRTF	MRVBF	NDVI	PCurv	PPTN	PrCurv	RI	Slope	TEMP	TPI	TWI
DEM	1.00											
MRRTF	0.19	1.00										
MRVBF	-0.40**	0.19	1.00									
NDVI	0.47**	0.02	-0.18	1.00								
PCurv	0.26	0.08	-0.19	0.10	1.00							
PPTN	0.76**	0.12	-0.32	0.50**	0.13	1.00						
PrCurv	0.34	0.12	-0.12	0.06	0.56**	0.28	1.00					
RI	0.16	-0.39**	-0.50**	0.13	0.19	-0.03	-0.18	1.00				
Slope	0.02	-0.27	-0.43**	0.03	0.19	-0.09	-0.23	0.67**	1.00			
TEMP	-0.94**	-0.11	0.36	-0.50**	-0.19	-0.79**	-0.30	-0.12	0.04	1.00		
TPI	0.39**	0.15	-0.34	0.21	0.65**	0.29	0.51**	0.28	0.13	-0.35	1.00	
TWI	-0.06	0.24	0.50**	-0.12	-0.32	0.01	0.05	-0.56**	-0.85**	0.00	-0.31	1.00

Significance level: ** = $p \leq 0.001$, $\alpha = 0.05$. DEM = digital elevation model, MRRTF = Multiresolution Ridge Top Flatness, MRVBF = Multiresolution resolution valley bottom flatness, NDVI = Normalized Difference Vegetation Index, PCurv = Planform Curvature, PrCurv = Profile Curvature, PPTN = Precipitation, RI = Relief Intensity, TEMP = Temperature, TPI = Topographic Position Index, TWI = SAGA Wetness Index.

Interpretation of principal components was based on finding which covariate layers are most strongly correlated with each PC, i.e., which of the loadings are large in magnitude and the farthest from zero in the either the positive or negative direction. The highest loadings were based on those that were highest for each PC, with loadings greater than 0.340 (Table 4.6) indicating the most important variable(s) in each PC.

PC1 accounted for 32.7% of the total variance within the raster stack. Landsat Band 4 was the most important metric and had the highest loading of -0.345. PC2 accounted for 16.1% of the total variance. Highest loadings were from MRVBF (-0.426), temperature (-0.348), and relief intensity (0.341). High PC2 values were observed on erosional areas that had high relief intensities, whereas low PC2 values were observed in depositional areas such as alluvial plains, bottomlands, and swamps where the values of both MRVBF and temperature were low. PC3 accounted for 10.4% of the total variance with the highest negative loadings from planform curvature (-0.505), profile curvature (-0.509), and TPI (-0.561). High PC3 values were observed in areas where the values of these TAs decreased. PC4 described 8.4% of the total variance with the highest negative loading from slope (-0.421). High PC4 values were observed in areas with steep slopes such as valleys and hillsides whereas low PC4 values were observed in areas with very gentle slopes such as uplands. PC5 accounted for 7.5% of the total variance with the strongest negative loadings from Landsat Band 5 (-0.651), whereas PC6 accounted 5.2% of the total variance with the strongest negative loading from Landsat band 9 (-0.588).

In summary, the 23 original covariates were transformed into 6 independent soil predictive components whose attributes make them ideal to improve on soil spatial prediction (Fig. 4.6). These attributes include (1) the ability to explicitly capture soil forming factors, (2) the ability to account for the majority of the variation within the raster stack, (3) independence and orthogonality to each other, and (4) reduction to a smaller dataset (six) compared to the twenty-three covariates initially selected for soil and landscape modelling.

Table 4.6: Proportion of variance contributed by the first six principal components (PCs) and Eigenvalues (loading) within each PC.

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Standard Deviation	2.7	1.9	1.5	1.4	1.3	1.1
Proportion of Variance (%)	32.7	16.1	10.4	8.9	7.5	5.2
Cumulative Proportion (%)	32.7	48.8	59.3	68.2	75.7	80.9
Environmental Covariate	**Eigenvector Values**					
Landsat Band 1	-0.321	-0.074	-0.005	-0.008	-0.203	-0.213
Landsat Band 2	-0.333	-0.037	0.012	0.018	-0.194	-0.178
Landsat Band 3	-0.328	0.029	0.069	0.055	-0.258	-0.067
Landsat Band 4	**-0.345**	0.091	0.027	0.041	-0.036	-0.023
Landsat Band 5	0.046	-0.01	0.164	0.094	**-0.615**	**0.381**
Landsat Band 6	-0.319	0.111	0.088	0.082	-0.131	0.144
Landsat Band 7	-0.340	0.109	0.046	0.058	-0.018	0.078
Landsat Band 8	-0.288	0.052	0.040	0.034	-0.117	-0.005
Landsat Band 9	-0.004	0.058	0.065	0.321	-0.008	**-0.588**
Landsat Band 10	-0.267	0.092	-0.041	-0.044	0.332	0.311
Landsat Band 11	-0.260	0.117	-0.030	-0.031	0.336	0.336
Elevation	0.122	**0.387**	0.078	0.276	0.024	0.090
MRRTF	-0.022	-0.165	-0.224	0.259	0.046	0.004
MRVBF	0.012	**-0.426**	0.002	-0.013	-0.066	-0.109
NDVI	0.263	-0.059	0.084	0.034	**-0.369**	0.285
Planform Curvature	0.010	0.123	**-0.505**	0.011	-0.137	-0.036
Precipitation	0.086	0.311	0.162	**0.417**	0.049	-0.123
Profile Curvature	0.001	0.078	**-0.509**	0.170	-0.099	0.083
RI	0.033	0.341	0.045	-0.335	-0.060	-0.036
Slope	0.057	0.311	-0.022	**-0.421**	-0.126	-0.193
Temp	-0.141	**-0.348**	-0.159	-0.298	-0.017	-0.051
TPI	0.005	0.128	**-0.561**	0.105	-0.115	0.030
TWI	-0.033	-0.319	0.055	**0.363**	0.119	0.155

MRRTF = Multiresolution Ridge Top Flatness, MRVBF = Multiresolution Valley Bottom Flatness, NDVI = Normalized Difference Vegetation Index, RI = Relief Intensity, TPI = Topographic Position Index, TWI = SAGA Wetness Index. Highlighted values represent loadings greater than 0.340 indicating the most important variable(s) in each PC.

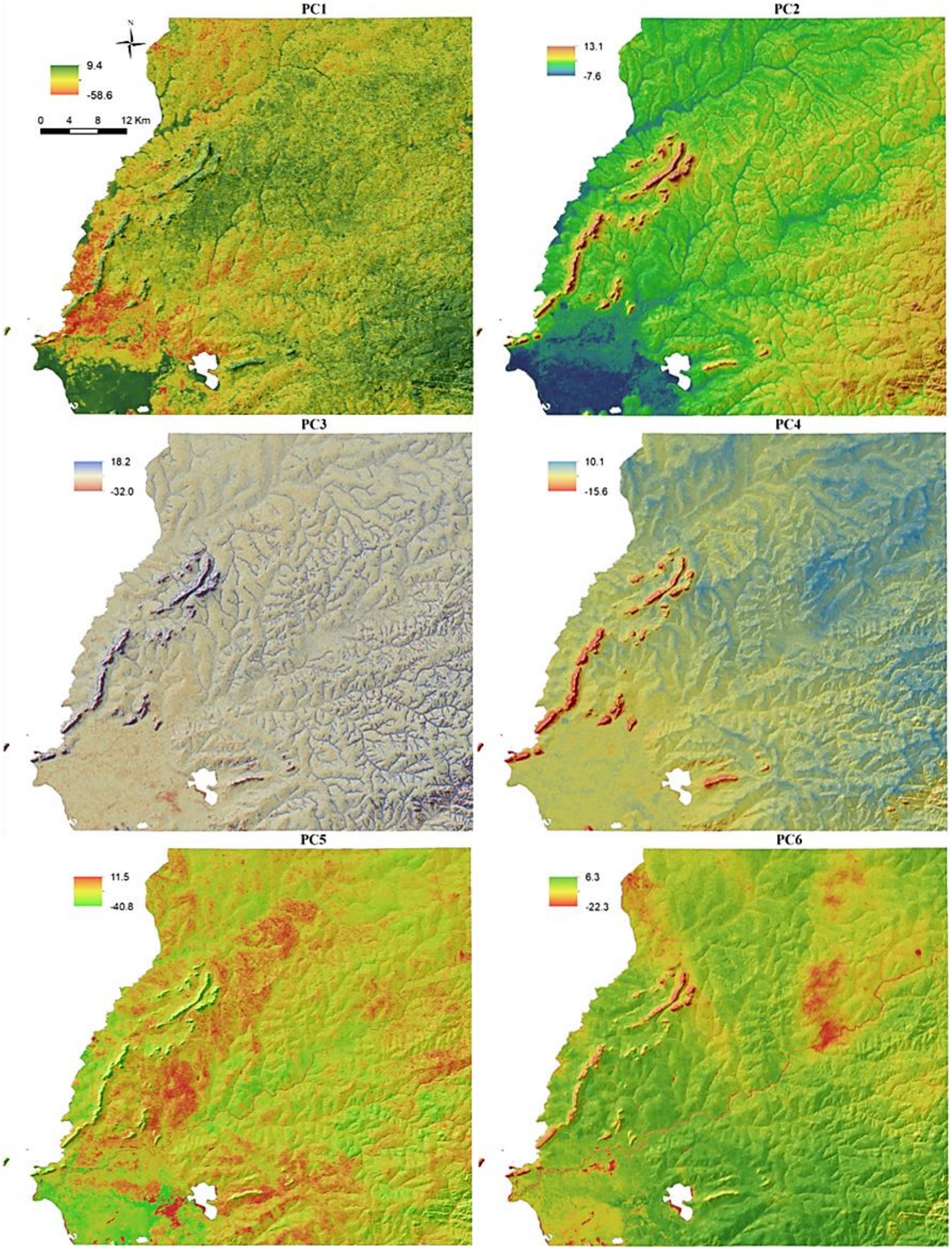


Figure 4.6: Principal component rasters used to predict soil properties for the Busia area.

Stepwise Multiple Linear Regression Model

Both forward and backward selection SMLR were used to predict soil properties using the *MASS* package in R (Ripley et al., 2013). The six principal components rasters described in section 2.3.1 above and Data Set A were the inputs into the SMLR. The significance of the principle components for predicting soil properties was assessed using the Akaike Information Criteria (AIC).

$$\mathrm{AIC} = 2\mathrm{K} + Nln(L) \quad [4.9]$$

where K is the number of the estimated parameters included in the model (in this case the 6 principal components), and L is the maximized likelihood function for the estimated model (Akaike, 1974). In cases, however, where a number of models have similarly low AICs, the one with the fewest predictor variables was chosen.

Stepwise Multiple Linear Regression Results

Among the six PCs, PC4 was the only significant predictor for SOC based on the AIC criterion, during SMLR modelling (Eqn. 4.10). PCs 1, 3, and 5 were significant in predicting clay (Eqn. 4.11) whereas PCs 1, 3, 5, and 6 were significant in predicting sand (Eqn. 4.12). No PC predicted silt.

$$\mathrm{lm}\ (\mathrm{formula} = \sqrt{\mathrm{SOC}}) = 1.07 + 0.02 * \mathrm{PC4} \quad [4.10]$$

$$\mathrm{lm}\left(\mathrm{formula} = \sqrt{\mathrm{Clay}}\right) = 6.56 + 0.11 * \mathrm{PC1} + 0.26 * \mathrm{PC3} + 0.26 * \mathrm{PC5} \quad [4.11]$$

$$\mathrm{lm}\left(\mathrm{formula} = \sqrt{\mathrm{Sand}}\right) = 5.69 - 0.16 * \mathrm{PC1} - 0.30 * \mathrm{PC3} - 0.29 * \mathrm{PC5} - 0.34 * \mathrm{PC6} \quad [4.12]$$

SMLR results overestimated the predicted soil properties (Fig. 4.4b and Fig. 4.5b and e). The prediction ranges were 2.73% for SOC, 84.2% for clay, and 97.1% for sand compared to Data Set A ranges which were 2.68% for SOC, 70.0% for clay, and 35.0% for sand.

Stepwise Multiple Linear Regression Evaluation

SMLR assumes that there is one global linear model that can explain the variation of the selected soil property within the study area. The drawback with this assumption is that with low sampling density, an appropriate relationship between a selected soil property and soil forming factors cannot be achieved. Predictors must have a consistent physical relationship with the target variable in all parts of the study area, otherwise predictions can be biased. The overestimation observed in SMLR may be due to the large number of predictors (6 principal components) and a small sample size (76). Pearson correlation between Data Set A and selected covariates showed very weak correlations of less than 0.5 (Table 4.5). Therefore, the established SMLR models between the point observations and selected environmental covariates used to predict soil properties may not explain all the variability within the study area.

This also explains the wide C.I. ranges of the soil prediction maps (Figs. 4.4e, 4.7b and e). Strong relationships can only be achieved when there is an adequate number of sample points evenly distributed throughout the study area capable of capturing the variability. This is a common challenge when using soil property data from legacy soil data: the lack of an appropriate statistical design when sampling. Results from SMLR should therefore be taken with caution. Nevertheless, SMLR was used to predict soil properties because: (1) it has the ability to manage large numbers of potential predictor variables; (2) the model can be fine-tuned to choose the best predictor variables from the available options; and (3) it is faster than other automatic model selection methods.

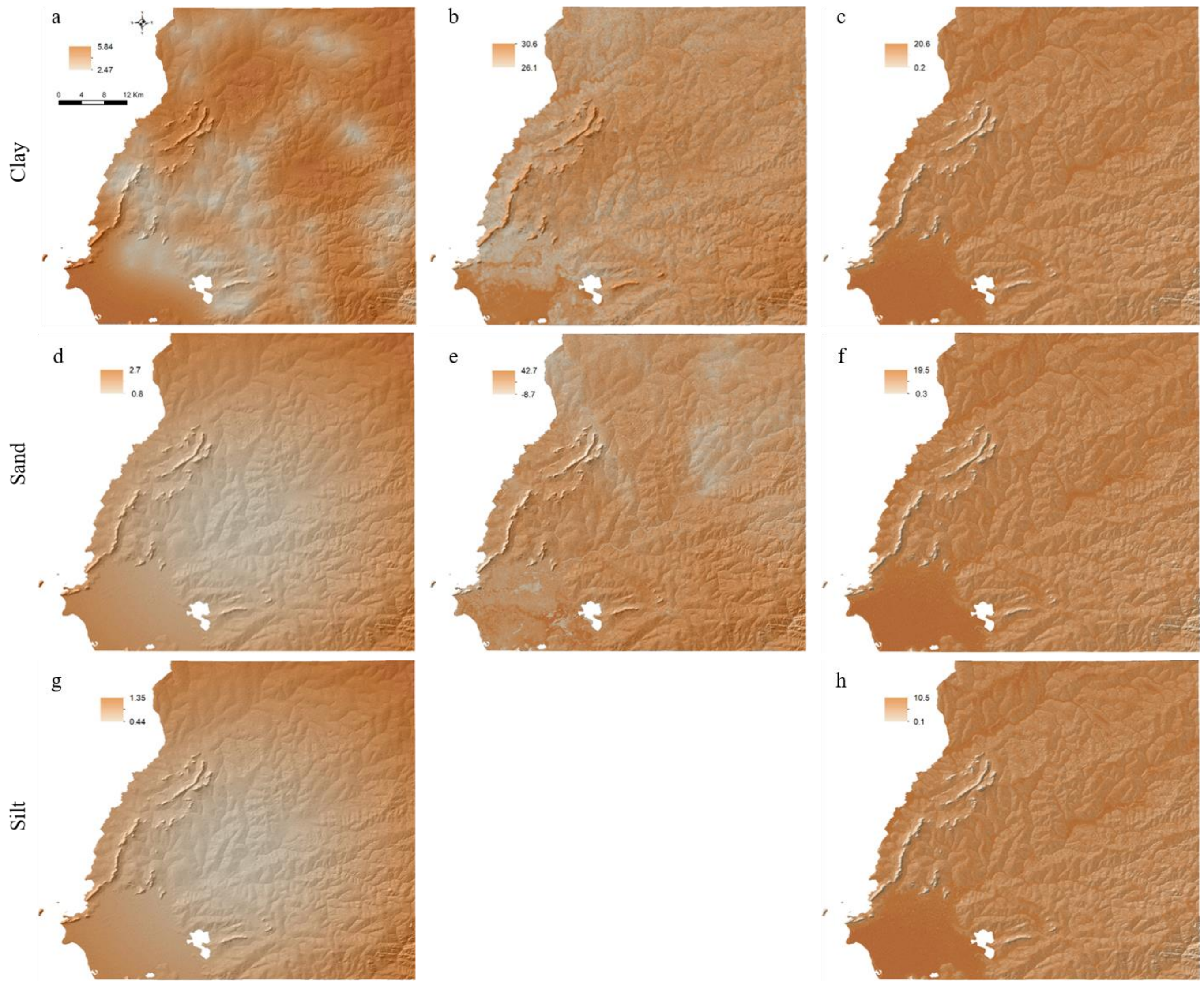


Figure 4.7: 95% C.I. range maps for the predicted soil texture maps from Figure 4.5.

4.3.3 The Soil Land Inference Model (SoLIM)

When mapping soil properties using SoLIM, only TAs are used because SoLIM assumes that soil property variability is solely explained by topography.

Preprocessing of Environmental Covariates for SoLIM

Selection of Least Correlated Terrain Attributes

The most relevant TAs were selected by omitting highly correlated TAs (Table 4.5). This resulted in six TAs for input into the SoLIM model, MRVBF, MRRTF, planform curvature, profile curvature, TPI, and TWI.

Terrain Attributes Clustering and Generation of Zonal Statistics

SAGA-GIS (Conrad et al., 2015) was used to perform the K-means clustering on the six TAs to partition them into k ($\leq n$) sets using the hill climbing method, where n is the number of TAs used for clustering (Rubin, 1967). TA clustering was intended to mimic the geometry of 'fully developed slopes' observed within the landscape (Wood, 1942; King, 1957). These clusters were considered to be 'ideal soil types' as discussed in section 4.3.3.2 below. To determine the sets of rules to use for each 'ideal soil type', zonal statistics for each cluster were calculated in SAGA-GIS to obtain the range of values for each TA to generate rule values for soil type landscape relationships in the study area (Table 4.7).

SoLIM Prediction Model

SoLIM is based on the premise that soil properties can be inferred from soil-related environmental conditions. It requires TAs that depicts the environmental conditions indicative of soil conditions (Zhu et al., 2001). SoLIM's inference engine links the TAs with a knowledge base to calculate similarity values. An example of soil-environment relationship knowledge could be expressed as rule sets like, "If the elevation is 1000 m and slope is 12%, then the soil is Soil Type 1". The inference engine uses the TAs to identify all the locations where elevation values are 1000 m and slope values are 12%, and then assigns full membership (similarity) to those locations because the soils at these locations are typical cases of Soil Type 1.

Not all locations, however, will satisfy the conditions set for Soil Type 1. For instance, Soil Type 1 occurs in areas with elevation from 500 m to 1500 m and slope from 6% to 18%. This does not mean that all places within this range of values will have soils similar to Soil Type 1. SoLIM acknowledges that places within this range will be more or less similar to another soil type, Soil Type 2, depending on the environmental variables. Also, soils in areas just a bit outside this range may still bear some similarity to Soil Type 1. For these locations, SoLIM will assign partial membership values based on how similar the environmental conditions at other locations are to the conditions stated above. This is accomplished by adopting a rule expressed as a function that defines how changes in an environmental variable affect the optimality of that environmental variable for a specific soil type. The optimality function describes how the similarity of a typical soil type changes as the environmental conditions deviate from the ideal conditions. This procedure is then repeated for all defined soil types, yielding a vector of similarity values (membership values) for each pixel.

SoLIM predicts soil properties using a linear and additive weighting function (Zhu et al., 2001) to predict soil properties for each pixel in the database.

$$D_{ij} = \frac{\sum_{k=1}^{n} S_{ij}^{k} \cdot D^{k}}{\sum_{k=1}^{n} S_{ij}^{k}} \tag{4.13}$$

where D_{ij} is the soil property at site *(i,j)*, S_{ij}^k is the similarity measure between the soil at site *(i,j)* and soil type k, D^k is the prescribed soil property of soil type k, and n is the total number of prescribed soil types in the study area.

Table 4.7: Rules used for terrain-based clustering (soil-landscape relationship).

Cluster	**MRRTF**			**MRVBF**			**Plan Curvature**			**Profile Curvature**			**TPI**			**TWI**		
	Mean	**Min**	**Max**	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**
1	0.45	-0.94	6.85	1.44	-0.89	7.72	-0.000063	-0.00626	0.0047	-0.000072	-0.0059	0.0045	-0.44	-11.91	9.95	6.03	2.50	11.18
2	0.08	-0.28	3.23	0.01	-0.05	1.05	0.003246	-0.00805	0.0193	0.003353	-0.0090	0.0191	27.31	-14.83	88.19	3.50	2.01	8.84
3	1.42	-1.52	8.06	0.78	-0.64	7.38	0.000375	-0.00676	0.0118	0.00036	-0.0066	0.0069	2.83	-12.03	33.59	5.82	2.02	10.91
4	0.06	-0.28	6.20	1.68	-1.33	8.31	-0.000521	-0.00851	0.0119	-0.000834	-0.0152	0.0064	-5.63	-70.34	18.16	5.80	1.81	11.83
5	2.00	-1.82	8.53	5.50	-1.39	8.63	0.000012	-0.00187	0.0022	-0.000026	-0.0025	0.0026	-0.13	-9.78	9.87	6.44	3.08	11.04

The six TAs selected for soil and landscape analysis (section 4.3.3.1.1) and the mean soil property values of Data Set A occurring in each soil type were inputs into the SoLIM model. The similarity vectors were developed from a knowledge-based strategy in SoLIMSolutions 2015 software using the concept of Jenny (1941, 1980), which postulates that relationships exist between soils and their formative environment, which in turn can be used to predict soil properties.

SoLIM Results

The five clusters (Fig. 4.8) resulted from k-means clustering were used as 'ideal soil types' for computing similarity vector values for the study area. These clusters mimicked different slope positions viz. summit (cluster 2), shoulder (cluster 3), backslope (cluster 1), toeslope (cluster 4), and footslope (cluster5) (Figs.4.8 and 4.9). These slope positions directly and/or indirectly influence geomorphic processes such as weathering, pedogenesis, and water runoff which in turn influence erosion-sedimentation determining the type of soils and landforms that evolve from these processes (Ruhe, 1975).

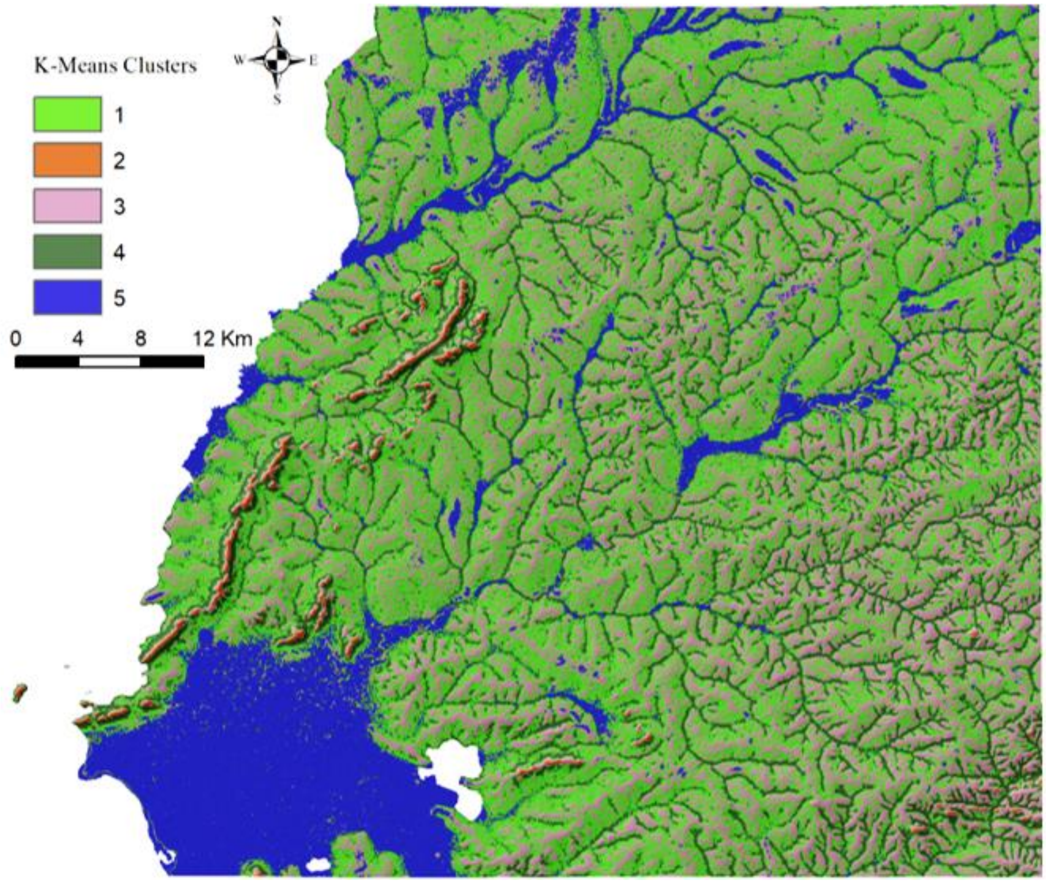


Figure 4. 8: K-Means clusters representing ‘ideal soil types’ within the landscape.

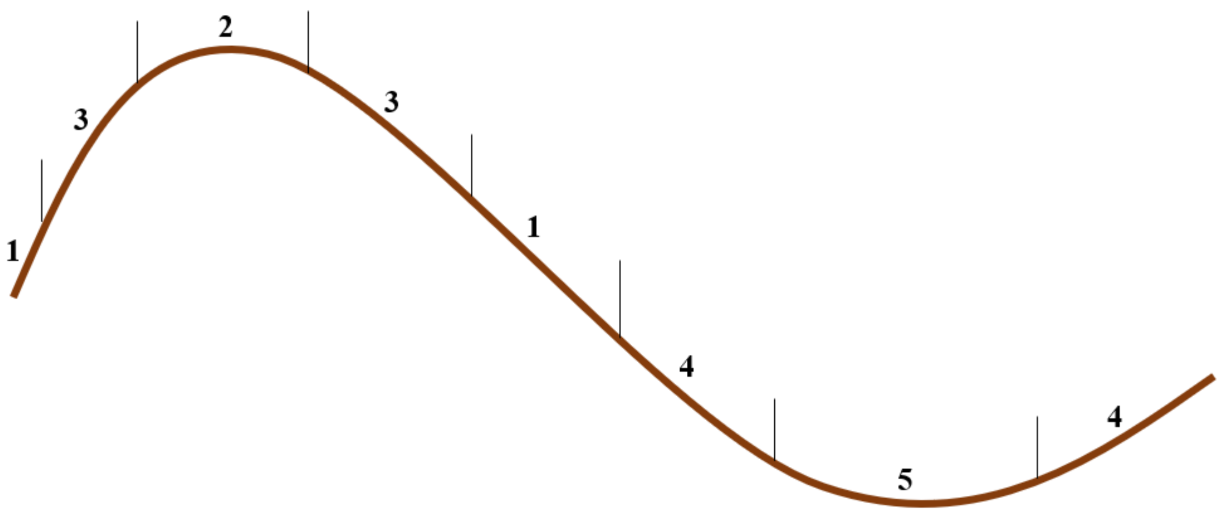


Figure 4. 9: Schematic representation of where each cluster occurred on ‘fully developed slopes’ within Busia landscape: 1 = Backslope (BS), 2 = Summit (SU), 3 = Shoulder (SH), 4 = Footslope (FS), and 5 = Toeslope (TS) (after Wysocki et al., 2000 and Schoeneberger et al., 2012).

SoLIM greatly underestimated the prediction of soil properties (Fig. 4.4c, Fig. 4.5c, f, and h). The prediction ranges were 0.26% for SOC, 9.6% for clay, 1.9% for silt, and 10.3% for sand compared to Data Set A ranges which were 2.68% for SOC, 70.0% for clay, 74.4% for silt, and 35.0% for sand.

SoLIM Evaluation

SoLIM heavily underestimated the prediction of properties because the model used the mean of all the soil property values found within each 'ideal soil type' for prediction instead of the individual soil property values (Table 4.8). This collapsed all the variation within the soil property intended for prediction into one value for each 'ideal soil type' and explained the very low prediction ranges.

Table 4.8: Soil properties within each 'ideal soil type' used for SoLIM prediction.

Ideal Soil Type	% SOC	% Clay	% Silt	% Sand
1	0.995	46.824	18.753	34.424
2	0.970	50.000	20.000	30.000
3	1.118	42.000	19.003	38.997
4	1.245	44.083	21.250	34.667
5	1.218	31.667	18.533	49.800

SoLIM uses the expert knowledge of an experienced soil scientist to formalize the relationships between soil characteristics and environmental covariates. This explains the level of detail from SoLIM predictions relative to ordinary kriging (Fig. 4.4c, 4.5 c, f, and h). The drawback is that a soil scientist's expert knowledge is subjective and lacks statistical grounds of inference. In addition, the accuracy of the clusters generated when identifying different slope positions (Fig. 4.9) that occur within the landscape depend on the resolution of the DEM. A 30 m DEM might not be suitable to generate clusters in areas with very gentle relief such as swamps. A higher resolution

DEM would presumably lead to better results. Moreover, adequate distribution of the sampled data points within slope positions is necessary for effective soil property prediction. Some clusters were over represented with sampled data points, whereas other clusters were under represented, thus affecting the prediction results. This explains the low correlation between the prediction and observed values (Fig. 4.10). Also, a greater number of soil types that capture the range of soil property variability would potentially increase the range of predictions values. However, the ability to assign property values for each soil type could be diminished with sparse point data.

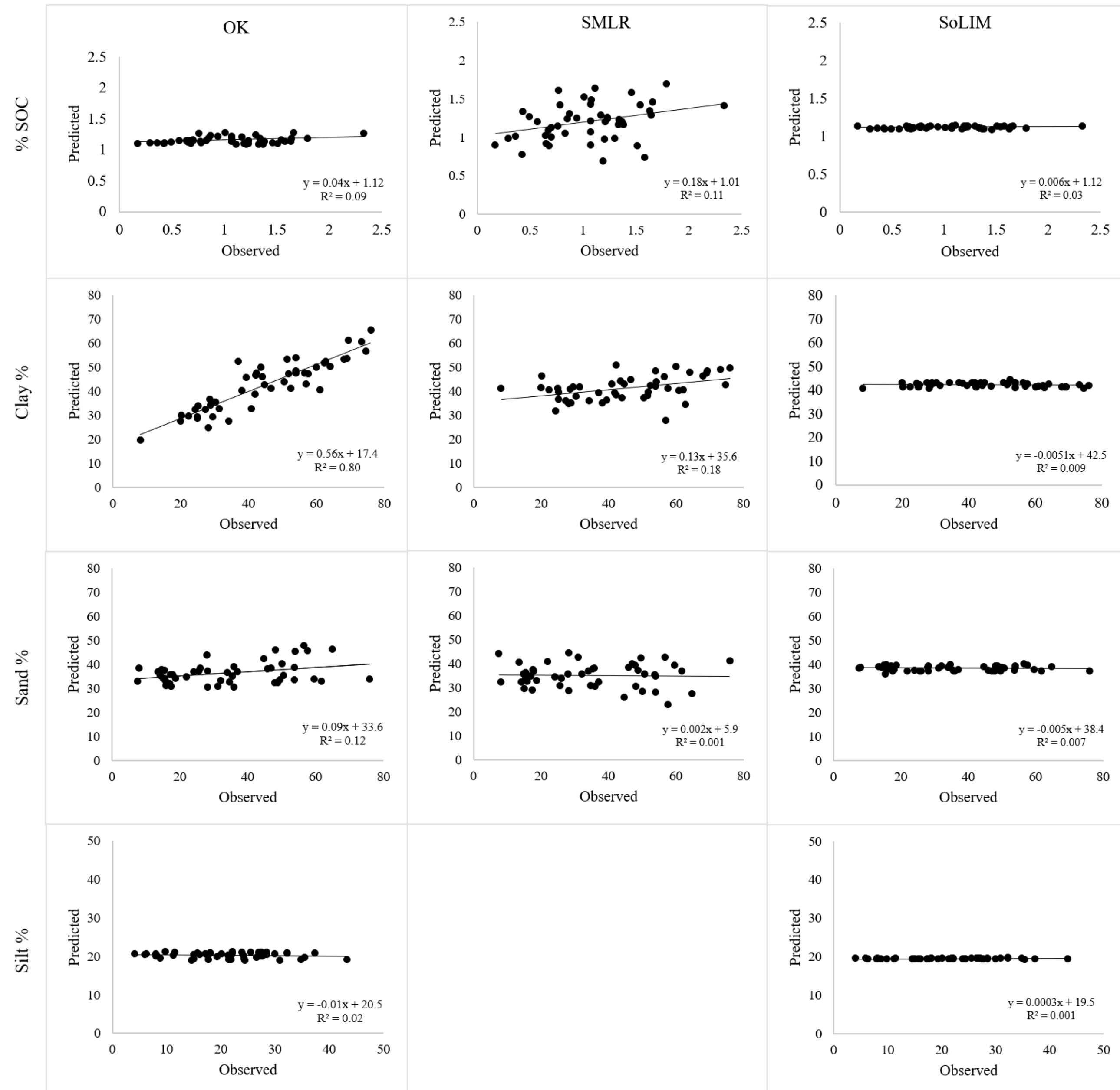


Figure 4.10: Correlation (density) plots produced as a result of cross-validation. See also Table 4.4 for more results.

4.3.4 Model Tradeoffs

From a statistical standpoint, ordinary kriging performed better than SoLIM and SMLR in predicting SOC (RMSE = 0.02), clay (RMSE = 0.32), and silt (RMSE = 0.10) whereas SMLR performed better than SoLIM and ordinary kriging for predicting sand content (RSME = 0.11) (Table 4.4).

The 95% C.I. ranges were narrowest for ordinary kriging (Fig. 4.4d and Fig. 4.7a, d, and g), followed by SoLIM (Figs. 4.4f, 4.7e, f, and h), and widest for SMLR (Figs. 4.4e, 4.7b and e) indicating that we are more confident of ordinary kriging predictions and least confident about the predictions from SMLR. From a pedological standpoint, SoLIM provided a better prediction model because it not only incorporates more soil forming factors compared to ordinary kriging but also has low C.I. ranges compared to SMLR. For example, SOC prediction from SMLR (Fig. 4.4b) predicted high OC on the sharp ridges in the southwestern part of the study area, which does not make pedological sense. This is supported by the large C.I. range on these areas (Fig. 4.4e). SoLIM on the other hand predicts low SOC on these ridges (Fig. 4.4c) with low C.I. ranges (Fig. 4.4f).

However, regardless of the variety of models used to predict soil properties, Austin et al. (2003) emphasized that in ecological modelling, the most important consideration is not the statistical model employed, but the ecological knowledge and statistical skill of the analyst. Minasny and McBratney (2007) likewise concluded that improved spatial prediction of soil characteristics will result from accumulating better soil data, rather than more sophisticated statistical models.

4.4 Discussion

The objective of this study was to determine whether existing legacy soil data for a selected portion of Kenya can be used to predict soil properties at a higher spatial resolution than current maps by using DSM techniques without conducting additional field work. Currently, the best soil property maps for the Busia area based on DSM techniques are at 250 m resolution (Hengl et al., 2017). Results from our study predicted soil properties at 30 m resolution, which should be more appropriate for smallholder farmers whose average landholding is less than a hectare (Sanchez et al., 2009). For research studies such as of nutrient cycling, finer resolution might be needed also.

DSM, however, requires a different sampling strategy which would be driven more by sampling the feature space rather than finding the extent or characterizing individual soil polygons.

A number of legitimate concerns, however, exist when using legacy soil data. The most immediate and significant concern is the data quality that affects the accuracy and usability of the initial soil property predictions. We used a calibration dataset, Data Set A, that was sampled to support soil classification (Appendix 4 in Rachilo and Michieka, 1991). The criteria used to determine the selected location of profile pits were not described explicitly in the soil survey report, although we can assume that the goal was to sample the most extensively cultivated soil types. Rachilo and Michieka (1991) stratified the Busia area into eight geomorphic units *viz.* hills, footslopes, uplands, plains, river terraces and floodplains, minor valleys, bottomlands, and swamps. Data Set A, however, is biased towards the uplands where 67% of the sampled points occur. Hills and swamps had no representative samples. Similarly, using the slope positions, the point density by slope position showed that 96% of the dataset occurred on shoulders, backslopes, and footslopes. Only one sample occurred on a summit within the entire study area which was not adequate to make useful predictions. This challenge has been highlighted by our predictions which exhibited lower accuracy on summits, as expected, due to the low sampling density.

On the other hand, for Data Set B used for evaluation, we applied an equal area quadratic smoothing spline function (section 2.2.1.2.1) to generate continuous depth function of soil properties. The approach not only reduced the variability (Odgers et al., 2012), but introduced additional errors during the evaluation of the prediction models. Compounded by the fact that this dataset was also not collected using any statistical design and/or probability and sampling criteria, the results from model evaluation possibly carry the same sampling bias (Savtchenko, 2004).

The initial bias of sampling designs coupled with the scarcity of the points highlight one of the major limitations of the methods used in our study. Sparse coverage of quantitative observations could introduce considerable spatial uncertainty (Mayr et al., 2010). Most of the methods used in this study are data demanding. Accurate prediction of soil properties, especially for spatial interpolation methods, depends not only on a higher sampling density but also on the spatial distribution. Our study fell short on meeting both criteria, and thus may not have captured the variability. A limitation of any model development is that the training dataset must contain and represent the full range of variation in the landscape of the study area for which the predictions are made (McBratney et al., 2003). This limitation was obvious for all the prediction models used in our study.

With regards to the input data, the prediction models used in this study can be improved or replaced once better sampling schemes are identified, and adequate sampling density is achieved. The fact that Data Set A did not capture the spatial variation within the study area, constitutes a future challenge to select a similar number of sample sites that can better capture the spatial variation. To enable this, we propose using the Conditioned Latin Hypercube Sampling technique (cLHS) (Minansy and McBratney, 2006). Given a set of environmental covariates, the cLHS provides a full coverage of the range of each environmental covariate by maximally stratifying the marginal distribution. The cLHS not only selects site locations that can be used to optimize the fitting of a spatial regression model but also are representative of the total spatial variation for the targeted soil property or properties (Minasny and McBratney, 2010a). The cLHS was run using the *clhs* package in R Studio (Roudier, 2017). In this example, the six PCs were used as the environmental covariates with 10,000 iterations to achieve an optimal solution. Sampling sites chosen by cLHS are given in Fig. 4.11.

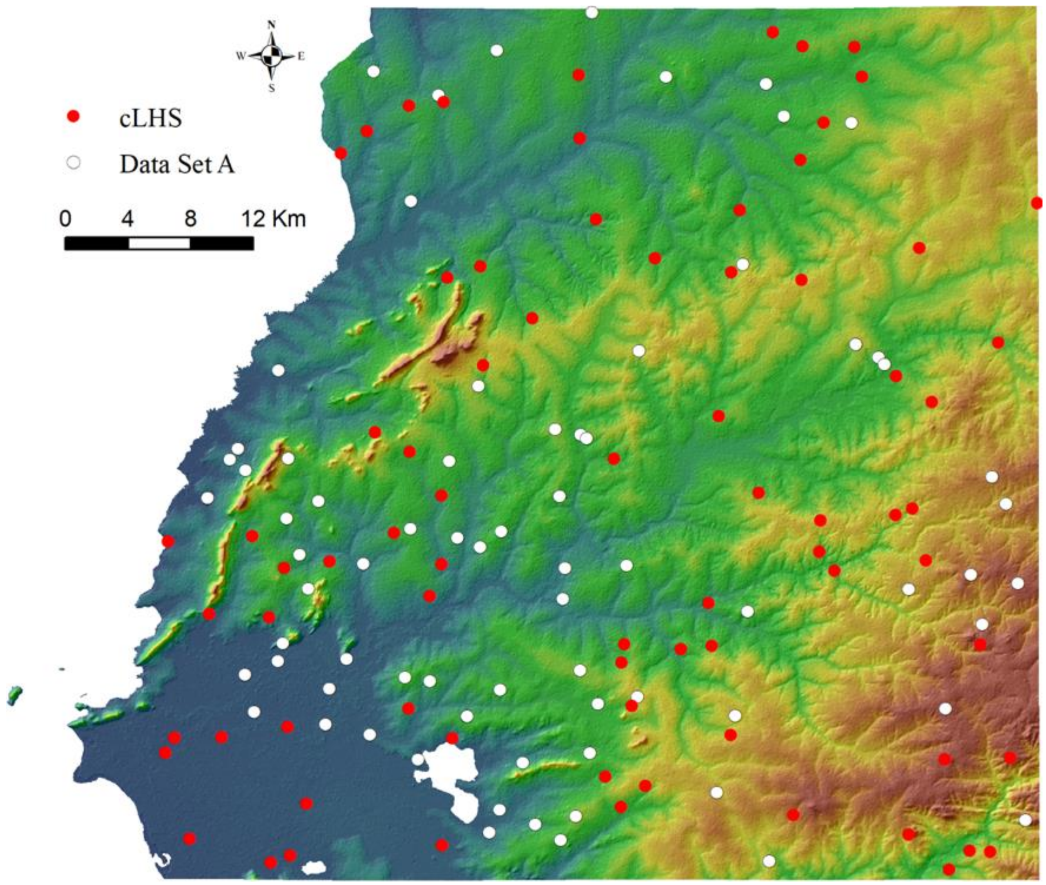


Figure 4. 11: Location of sample points determined by cLHS (red) relative to Data Set A (white).

The assumption when predicting soil properties for the Busia area is that they are strongly dependent on topography. A profile graph of a section of soil map unit UGb3M found in the uplands, where 67% of the calibration data set occurred, shows high topographic variability with no trend over very short distances (Fig. 4.12). Although the topography varied substantially over short distances, parent material differences, which were not identified in this study, might have also introduced another source of variability that contributed to poor predictions.

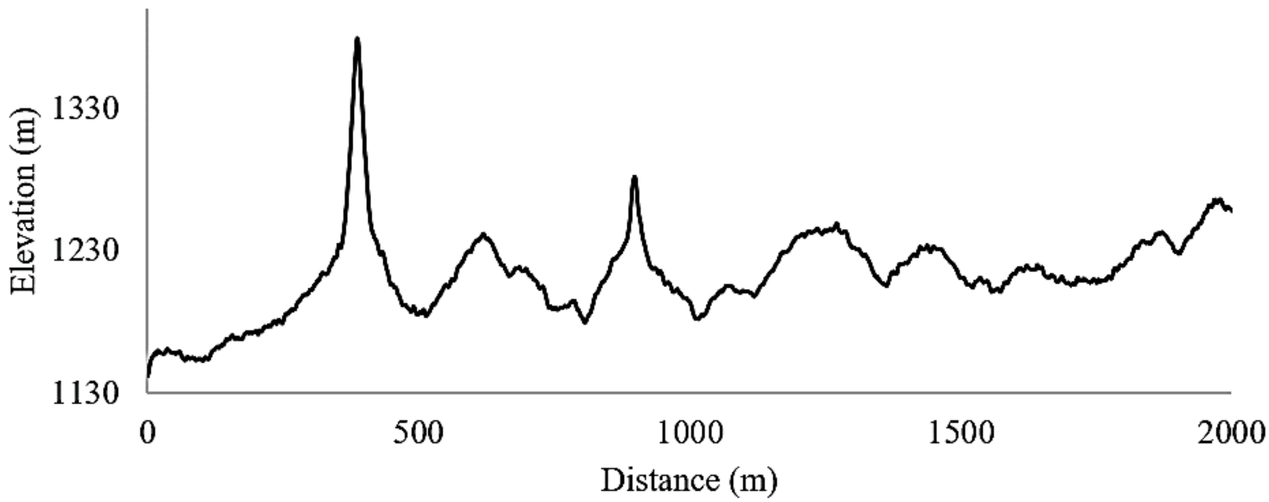


Figure 4. 12. UGb3M profile graph.

Even though ordinary kriging performed better when evaluating the prediction models, this result should be taken with a lot of caution. The large ranges resulted from fitting the spatial prediction models showed that there was weak spatial dependence (Table 4.3). Significant proportion of spatial variation for most soil properties occurs within relatively short distances (Heuvelink et al., 2001; Antonić et al., 2003; Scharlemann et al., 2014). It is therefore unreasonable to expect for any model to explain more than 50% of the total observed variation (Hengl et al., 2014).

The DSM techniques used in this study provide a blue-print for future soil property mapping using data from legacy soil data. These results offer the best possible soil property predictions at finer resolutions (30 m raster grids) than the ones currently available (250 m raster grids). Similar prediction models and soil and landscape relationships can be refined and used in future studies. The expectation is that these issues and concerns can gradually be addressed in a systematic way as more data become available.

4.4.1 Conclusion

This study demonstrates use of legacy soil data for predicting SOC, clay, silt, and sand and emphasizes the importance of such finite resources. One challenge, however, still remains when utilizing such data: low sampling density. We demonstrated that DSM can be used to predict soil properties mined from legacy soil data at finer spatial resolutions, 30 m for this study. We hope that by demonstrating how to effectively utilize legacy soil data for soil property prediction, we will encourage efforts to map specific soil properties in areas with existing legacy soil data.

CHAPTER 5. DISAGGREGATION OF THE 1:100,000 RECONNAISSANCE SOIL MAP OF THE BUSIA AREA USING A SOIL LANDSCAPE RULE-BASED APPROACH

Abstract

A 1:100,000 soil map of the Busia area in western Kenya published in 1991 is the main source of soil information for the area, but the map is outdated and unsuitable for new emerging demands. We disaggregated the *Reconnaissance Soil Map of the Busia Area (quarter sheet No. 101)* into individual soil classes to produce a soil class map that may better meet current needs. The soil landscape rule-based model was used to disaggregate the soil map units by exploiting information in the map legend and the map unit descriptions. These descriptions were used to generate rules that were applied to a fuzzy soil class map generated from a parent material map and a K-means cluster map generated from six terrain attributes, namely, multiresolution ridgetop flatness (MRRTF), multiresolution valley bottom flatness (MRVBF), topographic wetness index (TWI), topographic position index (TPI), planform curvature, and profile curvature. The result was a soil class map with a spatial resolution of 30 m with an overall accuracy of 58% and a Kappa coefficient of 0.54. The soil landscape rule-based approach provides an opportunity to disaggregate traditional soil maps in cases where there is no adequate dataset. The drawback with this approach, however, is that it relies heavily on the resolution of the DEM and the type of information contained within the soil survey report describing the soil classes within the soil map units.

5.1 Introduction

A 1:100,000 soil map of the Busia area in western Kenya published in 1991 (Rachilo and Michieka, 1991) is the main source of soil information for the area. It is used for a variety of purposes, such as agriculture and environmental policy making and soil and water conservation planning. Almost three decades later, however, the map has become outdated and unsuitable for emerging demands. Large-scale changes in land use and water resource management have occurred in the area since the soil survey was conducted, and the existing soil map does not provide the level of detailed needed today.

Although the need for updating this map is recognized, soil surveys are very expensive, labor intensive, and time consuming (Arrouays et al., 2017). One way of updating existing soil maps is by using digital soil mapping (DSM) techniques (McBratney et al., 2003) with data extracted from existing legacy soil survey reports (Bui et al., 1999; Bui and Moran, 2001; Odgers et al., 2014).

Traditional, polygon-based soil maps often describe soil classes within the map units but lack explicit indications of where particular soil classes occur. Mapping soil classes as opposed to map units is important because specific soil classes can effectively inform and improve sustainable soil use and management at a finer scale (Brungard et al., 2015). Disaggregating broad soil map units into individual soil classes has the potential to make existing soil maps more useful for meeting current needs.

5.1.1 Spatial Disaggregation

Spatial disaggregation of soil classes within existing map units refers to downscaling of information to produce new information at a finer scale than the original source (McBratney, 1998; Thompson et al., 2010). The result is a rasterized prediction of the spatial distribution of soil classes occurring within the original soil map units.

Different methods have been tested in an attempt to disaggregate traditional soil maps. These include: (1) the soil landscape rule-based approach (Bui et al., 1999; Bui and Moran, 2001;

Zhu et al., 2007, 2008; Thompson et al., 2010; Nauman et al., 2012), (2) decision trees / rule induction (Bui et al. 1999; Bui and Moran, 2001; Thompson et al., 2010; Wei et al., 2010; Häring et al., 2012; Subburayalu et al., 2014; Odgers et al., 2014), (3) area to point kriging (Kerry et al., 2012), (4) multinomial logistic regression (Hengl et al., 2007; Kempen et al., 2009; Hengl et al., 2017), (5) regression kriging (Hengl et al., 2007), and probabilistic approach (Cheney et al., 2016). The majority of these methods, however, require an adequate calibration dataset of point data for effective disaggregation.

In cases where there is a lack of adequate point data, legacy soil survey reports remain the only source of information for disaggregating soil map units (Minasny and McBratney, 2010b). Legacy soil data in the form of soil maps and associated reports from traditional soil mapping approaches are rooted in the catena concept formalized by Milne (1936). These maps divide the landscape into map units consisting of one or more soil classes that occur in predictable and repeating patterns (Soil Science Division Staff, 2017).

The soil landscape model offers an opportunity to disaggregate soil map units by exploiting information in the map legend and the soil map unit descriptions (Bui and Moran, 1991; Lagacherie et al., 1995; Mayr et al., 2001). These descriptions reflect the mental models used by soil surveyors when the map was made, and they describe the association between soil map units and other environmental and spatial data.

The objective of this study was to spatially disaggregate the best available soil map for the Busia area of western Kenya into individual soil classes to produce a soil map that may better meet current needs. The procedures developed for this study may also be useful for disaggregating similar soil maps available for other parts of Kenya.

From our initial evaluation of the soil map and associated report for the Busia area, we concluded that there were insufficient point data to apply any of the statistically based procedures mentioned above and that the soil landscape rule-based approach was the most likely to result in a useful soil map.

The soil landscape rule-based approach for disaggregating soil map units has four major steps (Fig. 5.1). First, the mental soil landscape models used by the original soil surveyor(s) must be extracted from the available descriptive information and quantified by a set of rules. Second, a digital model must be developed that represents the landscape and environmental conditions of the landscape. Third, the rules that describe the locations of different soil classes in the landscape must be applied to the digital model to place the soil classes in their most likely position(s) in the landscape. Fourth, the resulting map should be evaluated in some way.

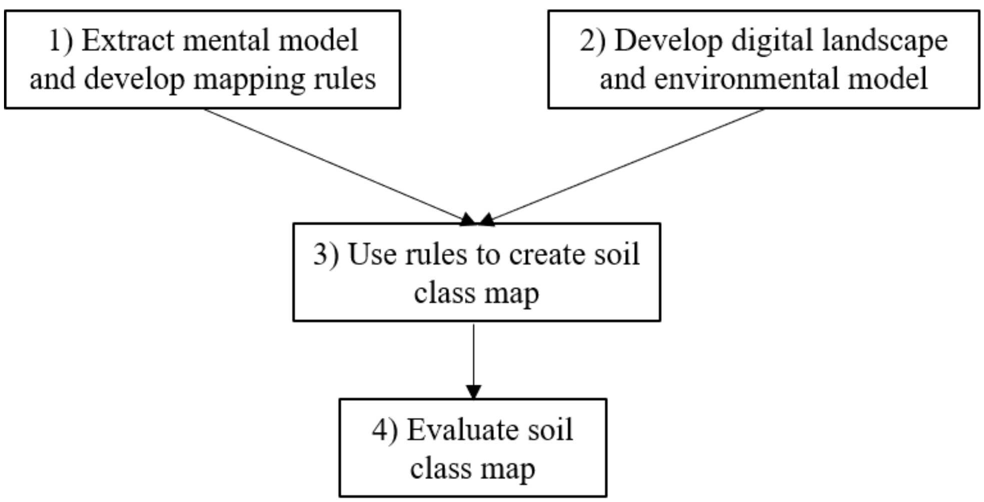


Figure 5.1: Soil landscape rule-based model for disaggregating soil map units.

The distribution of soil classes within soil map units like hydrological processes, is often assumed to be either stochastic or deterministic (Seyfried and Wicox, 1995), i.e., the distribution of soil classes is either random or it follows a certain pattern. In soil surveys, these situations are described as associations and complexes, respectively (Soil Science Division Staff, 2017).

Deterministic patterns can be associated with specific landscape positions that are captured in the map unit descriptions but are not spatially represented on the map (Bui and Moran, 2001; Nauman et al., 2012). Stochastic patterns cannot be associated with specific landscape positions. Examination of the legend and map unit descriptions for the Busia area soil map showed that the map units were described deterministically, providing some assurance that the soil landscape rule-based approach would likely be successful.

Digital elevation model data was available for the area and sufficient information was available in the soil survey report to extract a soil parent material map. Thus, it would be possible to develop a digital model to represent the landscape and environmental conditions. Sufficient data, therefore, were available for the objective of this paper to apply the soil landscape rule-based approach to disaggregating the Busia area soil map.

5.1 Materials and Methods

5.1.1 Study Area

The study area is located in the western part of Kenya and bound by the equator to the south, latitude 0° 30’ N to the north, longitude 34° 30’ E to the east and the Kenya-Ugandan border in the west. It has an area of 2,798 km^2 and an elevation ranging from 1,127 to 1,564 m. Mean annual rainfall increases to the northeast from 925 to 1,990 mm. The average annual temperature ranges from 20 to 23° C.

5.1.2 Data Sources

The *Reconnaissance Soil Map of the Busia Area (quarter degree sheet No. 101)* at a scale of 1:100,000 (Rachilo and Michieka, 1991), the “Busia soil survey report,” was used as the primary source of data. The digital elevation model (DEM) data was from the 1 Arc-Second (30 m) Shuttle

Radar Topography Mission (SRTM) global elevation dataset (USGS, 2017). It was projected to the WGS84 Web Mercator (auxiliary sphere) coordinate system before further processing. This coordinate system was chosen for this study because of its ability to preserve the shape of the area without distortion, especially around the equator (ESRI, 2010).

Soil Tabular Data

The Busia soil survey report contains tabular descriptions and corresponding soil classifications according to the FAO-UNESCO nomenclature for the Soil Map of the World (FAO/UNESCO, 1974) for 76 geo-located soil profile pits occurring within specific soil map units (Appendix 4 in Rachilo and Michieka, 1991).

Soil Spatial Data

The soil polygon map that is part of the Busia soil survey report comprises 348 polygons that belong to 52 soil map units (Fig. 5.3). The map units are soil associations consisting of one or more soil classes explicitly described within the Busia soil survey report.

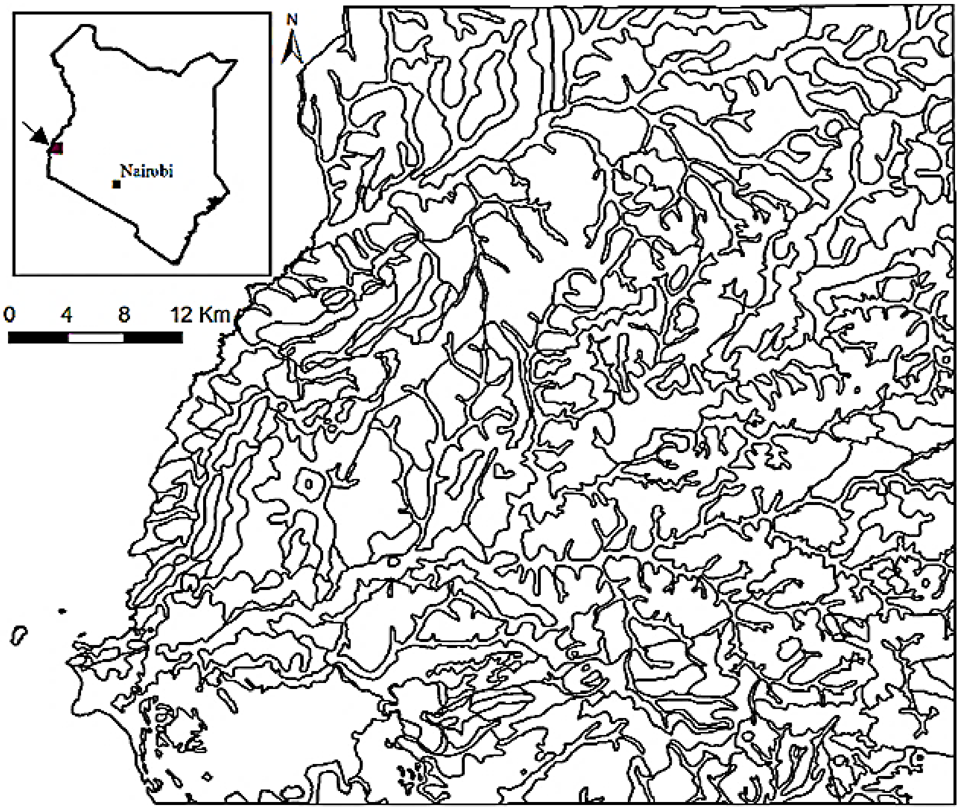


Figure 5.2: Busia soil polygon map and location within Kenya (arrow).

Systematic Nomenclature of the Soil Map Units

The legend for the soil map units consisted of three categories. The highest category was the physiographic unit on which the soil map unit occurs, such as hills, foothills, uplands, etc. At the second level, the physiographic units were subdivided according to the underlying parent material such as igneous rocks, granite, alluvium, etc. At the third level, the soil map units were subdivided based on important characteristics of the soil profile, such as drainage conditions, depth, color, consistency, texture, etc. This was followed by the soil classification using the Legend for the Soil Map of the World. For example, a soil map unit designated as "HIP (Lithosols, stony phase)" is described as soils found on hills (H), of igneous parent material (I), with very shallow soils overlying bedrock with soil depth ranging between 0-50 cm (P). Such descriptions make it easier to disaggregate soil map units.

Soil Classes

Twenty-six different soil classes occur within Busia area and they are grouped hierarchically into soil orders in Table 5.1. The most common soil orders in the Busia area are Cambisols (21.4%), Acrisols (16.1%), Gleysols (14.3%), and Ferralsols (14.3%).

Table 5.1: Soil classes of the Busia area.

Soil order	**Soil classes**	**Frequency (%) of occurrence within soil map units**
Acrisols	Chromic Acrisols Orthic Acrisols	16.1
Cambisols	Chromic Cambisols Dystric Cambisols Eutric Cambisols Ferralic Cambisols Vertic Cambisols	21.4
Vertisols	Chromic Vertisols Pellic Vertisols	7.1
Gleysols	Dystric Gleysols Eutric Gleysols Humic Gleysols Plinthic Gleysols Vertic Gleysols	14.3
Lithosols	Dystric Lithosols	7.1
Nitosols	Dystric Nitosols Eutric Nitosols	7.1
Fluvisols	Eutric Fluvisols Vertic Fluvisols	3.6
Histosols	Eutric Histosols	1.8
Arenosols	Ferralic Arenosols	3.6
Solonchaks	Gleyic Solonchaks	1.8
Ferralsols	Orthic Ferralsols Plinthic Ferralsols Rhodic Ferralsols	14.3
Luvisols	Orthic Luvisols	1.8

5.1.3 Soil Land Inference Model (SoLIM)

The soil land inference model (SoLIM) was used to map fuzzy soil classes occurring within the Busia landscape. SoLIM is based on the concept that soil classes can be spatially inferred from soil-related environmental conditions. This approach employs a set of rules depict environmental conditions indicative of soil types. SoLIM uses an inference engine to link a GIS database, which stores the environmental covariates, with a knowledge base to calculate similarity values (Zhu and Burt, 2011).

For example, soil-environment relationship knowledge could be the statement, "If the elevation is 1000 m, slope is 12%, and parent is type X, then Soil Class A is most likely to occur." In this case, the inference engine will use the GIS database to identify all the locations where these conditions are met, and then assign full membership to Soil Class A for those locations where soils are typical for Soil Class A. Not all locations in the area, however, will meet the conditions perfectly for Soil Class A. For example, "Soil Class A occurs in areas with elevation from 500 m to 1500 m and slope from 6% to 18% and underlying granite parent material." This does not mean that all places with this range of values will have the same soil. Instead, SoLIM acknowledges that places within that range will be more or less similar to another soil class, Soil Class B, depending on the values of the environmental variables for Soil Class B. In addition, soils in areas just a bit outside of the range may still bear some similarity to Soil Class A; these soils will not be perfect examples of typical Soil Class A, but they will not be totally dissimilar to Soil Class A either.

For these locations, which constitute the majority of the landscape, SoLIM will assign partial membership values based on how similar the environmental conditions at other locations are to the conditions stated above. This is accomplished by adopting a rule which is expressed as a function that defines how changes in an environmental variable affect the optimality of that specific soil class occurring at that particular location. This procedure is repeated for all defined soil classes, yielding a vector of similarity values for each pixel.

5.1.4 Disaggregation Approach

The disaggregation approach was achieved using 12 steps (Fig. 5.3). These steps expand on the soil landscape rule-based approach for disaggregating soil map units introduced in section 1.1 and illustrated in Fig. 5.1.

Extract mental model and develop mapping rules

Create List of Soil Classes from Soil Survey Report

The soil map units in the Busia soil survey report have information on the total area, agro-climatic zone, parent material, meso- and macro-relief, erosion, land use, surface rockiness, general soil description, color, texture, structure, consistence, chemical properties, clay minerology, diagnostic properties, and soil classification. A list of all soil classes in the Busia soil survey report was first created (step 1 in Fig. 5.3, Table 5.1).

Extract Soil Class Definitions from The Legend of The Soil Map of The World

The legend of the Soil Map of the World (FAO/UNESCO, 1974) was used to extract definitions of the soil classes (step 2 in Fig. 5.3). The Busia soil survey report was also used (i) as an additional source of information in cases where soil classes were explicitly defined, (ii) in cases where soil class definitions did not exist within the legend of the Soil Map of the World, and (iii) to generate a separate database showing the soil classes within each soil map unit.

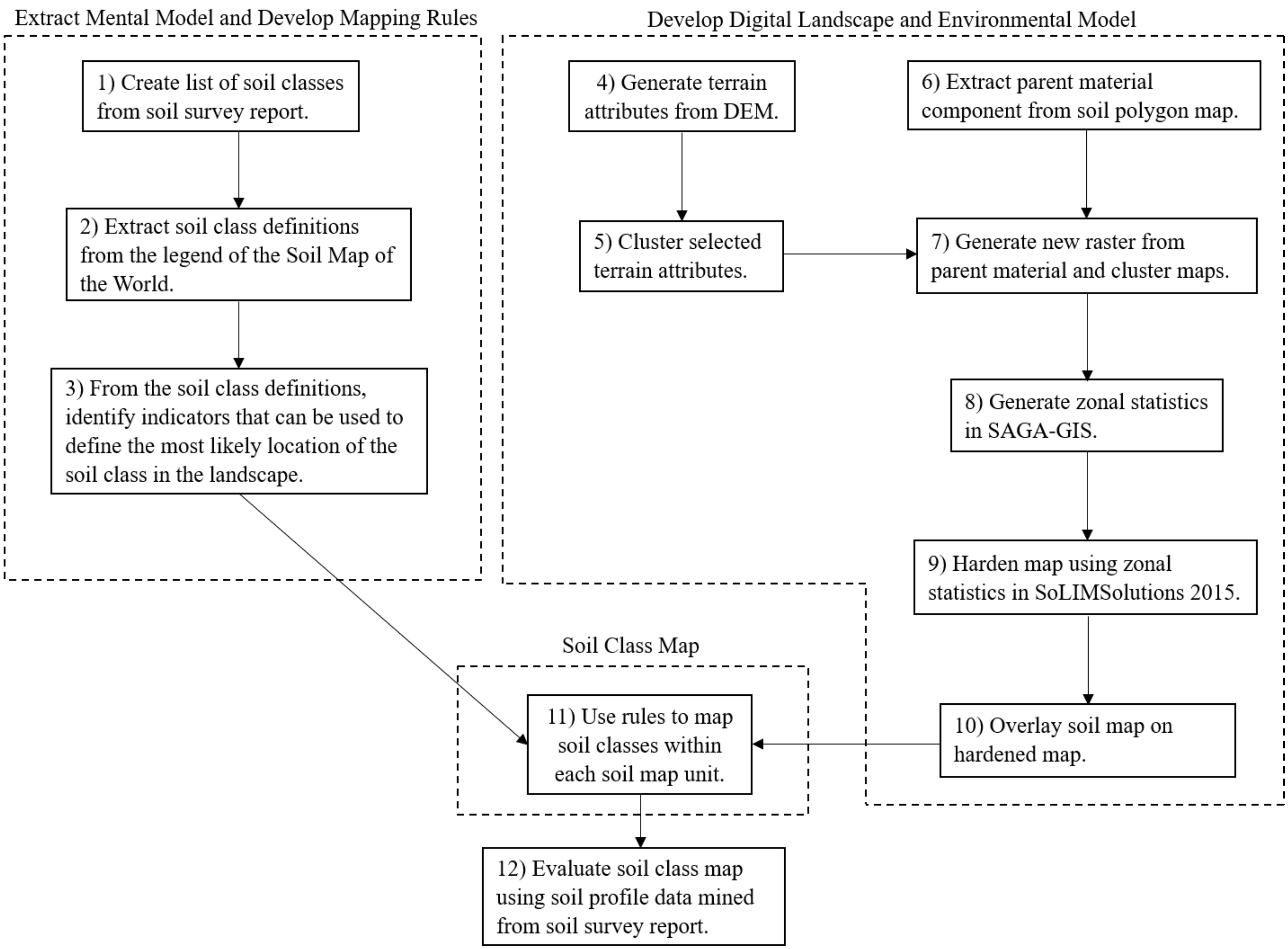


Figure 5.3: Schematic flow chart for dissaggregating of soil map units using legend and map unit desciptions.

Develop Soil Class Mapping Rules

Soil class definitions contained indicators that could be used to associate these soil classes to specific landscape positions (step 3 in Fig. 5.3). For example, according to the soil map legend, Eutric Nitosols are described as having "a base saturation of 50% or more throughout the argillic B horizon within 125 cm of the surface." Within the landscape, such soil classes will most likely be found on depositional areas where soluble materials washed from upslope areas will likely accumulate. Similarly, Dystric Nitosols are described to have "a base saturation of less than 50% in at least a part of the argillic B horizon within 125 cm of the surface and lack high OM content." Within the landscape, such soil classes will most likely be associated with upslope areas where soluble nutrients are washed away.

In very few cases, soil classes within the Busia soil survey report were explicitly described. For example, within the valleys, the report describes soil classes occuring in specific soil map units as "soils of the valley sides" and "soils of the valley bottom." In these cases, the descriptions were used to associate these soil classes with their corresponding slope positions. Valley sides are areas that occur on shoulders and backslopes whereas the bottomlands are areas that occur on footslopes and toeslopes. The process of sifting through the soil class definitions in the legend and soil map unit descriptions and associating soil classes with specific landscape positions led to the generation of rules for mapping each soil class.

Examples of Rules to Associate Soil Classes Within Specific Map Units to Landscape Positions

Table 5.2 shows an example of part of the database created to generate rules to map soil classes to their most likely landscape positions within specific soil map units. We discuss the generation of rules for three map units only.

Table 5.2: Examples of rules for mapping soil classes within soil map units. Italicized text are key indicators of where soil classes would most likely occur within the landscape.

Map unit	Soil classes	Busia soil survey description	Soil class definition according to the soil map of the world	Slope position 'soil class mapping rules'
VXC1	Ferralic Arenosols	'soils of the *valley sides*'	*Excessively drained to well drained, deep yellowish brown* to grey, loose, loamy sand to sandy loam.	Shoulder and backslope.
	Eutric Gleysols	'soils of the valley *bottom*'	Have *a base saturation of ≥ 50% at least between 20 and 50 cm from the surface*; *imperfectly drained to poorly drained, deep dark brown, mottled*, firm, sandy clay to clay.	Footslope and toeslope.
SAC1	Chromic Vertisol		Vertisols having *moist chromas of 1.5 or more* dominant in the soil matrix.	Summit, shoulder, and backslope.
	Eutric Histosols		Histosols having a pH of 5.5 or more at least between 20 and 50 cm from the surface; lacking permafrost within 200 cm of the surface.	Footslope.
	Humic Gleysols		Gleysols having an *umbric A horizon* or a dystric histic H horizon; *lacking plinthite within 125 cm of the surface*; lacking permafrost within 200 cm of the surface.	Footslope.
	Vertic Fluvisols		Fluvisols with *vertic* properties.	Toeslope.
UVC2	Chromic Acrisols		Acrisols with *high chroma values.*	Summit, shoulder, and backslope.
	Ferralic Cambisols		*Ochric A horizon* and a cambic B horizon with ferralic properties; *lack vertic properties*; *lack hydromorphic properties* within 100 cm of the surface; lack permafrost within 200 cm of the surface.	Summit, shoulder, and backslope.
	Orthic Acrisols		*Ochric A horizon*; lacking ferric properties; lack high organic matter content in the B horizon; lack hydromorphic properties within 50 cm of the surface.	Summit, shoulder, and backslope.
	Plinthic Acrisols		Acrisols with *plinthite within 125 cm of the surface.*	Summit, shoulder, and backslope.

Map Unit VXC1

This map unit occurs on valleys with granites and grits as the underlying parent material and is an association of two classes, Ferralic Arenosols and Eutric Gleysols. The Busia soil survey describes Ferralic Arenosols as "soils of the valley sides" whereas Eutric Gleysols are described as "soils of the valley bottom" (Table 5.2). In this example, the explicit soil class descriptions were sufficient to associate these soil classes to specific landscape positions. "Valley sides" and 'valley bottoms' are key slope position indicators. "Valley sides" are associated with shoulders and backslopes characterized by steep slopes with soil drainage class ranging between excessively well drained to well drained. Erosion is quite intense and therefore soils are shallow (Figs. 5.4 and 5.5). Conversely, "valley bottoms" are associated with footslopes and toeslopes or bottomlands characterized with gentle topography. Soils within these slope positions are poorly drained because water accumulates (Figs. 5.4 and 5.5).

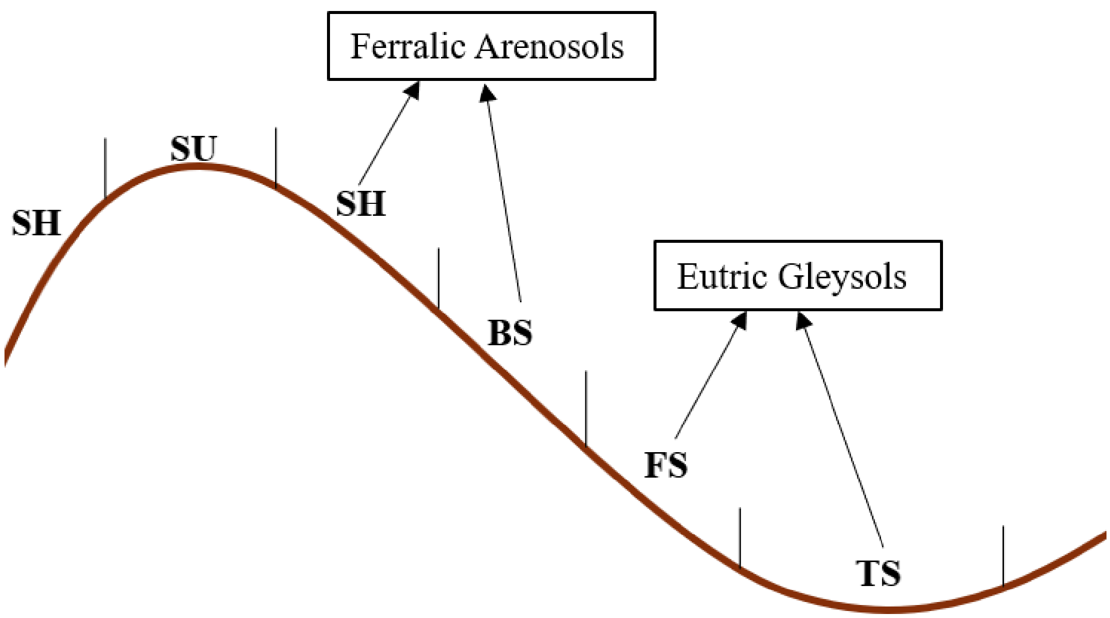


Figure 5. 4: Variation of soil classes within soil map unit VXC1.

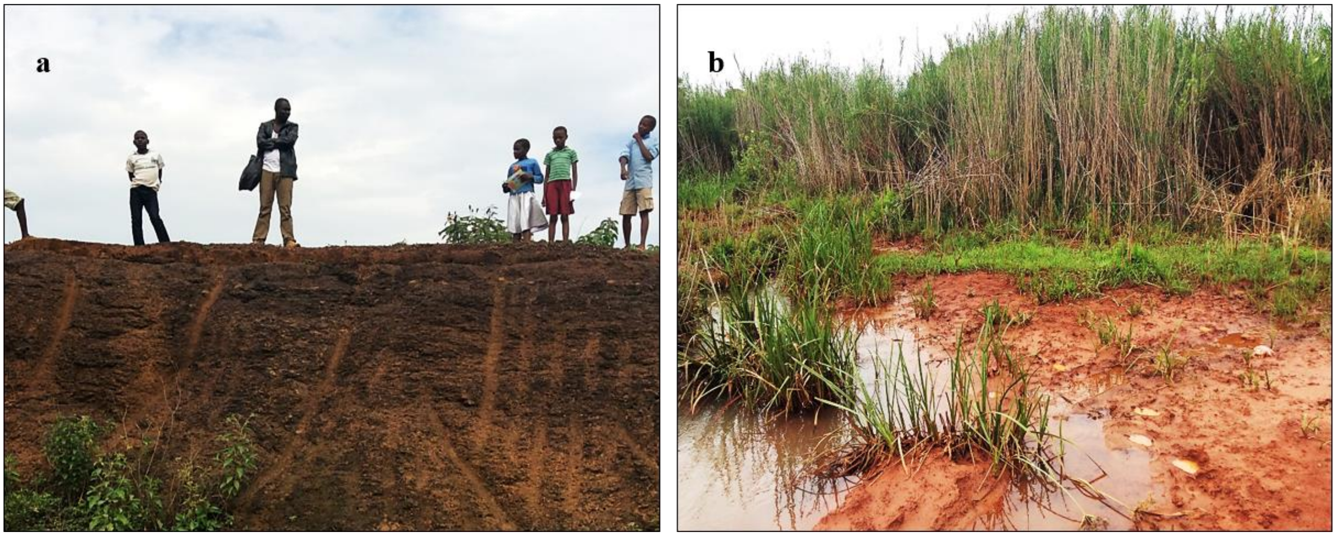


Figure 5. 5: (a) very shallow Ferralic Arenosols overlying petroplinthite on the valley sides and (b) poorly drained Eutric Gleysols found on the valley bottoms.

Map Unit SAC1

Map unit SAC1 occurs in swamps of alluvial parent material. Soils classes in this map unit include Chromic Vertisols, Vertic Fluvisols, Eutric Histosols, and Humic Gleysols (Table 5.2). There was no explicit description of this map unit in the Busia soil survey report, and therefore only the definitions of these soil classes from the legend of Soil Map of the World were used. Chromic Vertisols are described as Vertisols that have high chroma values. High chroma colors (yellow and red to brown) are a result of iron oxide minerals that are commonly found in well drained areas where soils are not seasonally saturated and reduced and are commonly associated with summits, shoulders, and backslopes. There was no unique indicator from the Eutric Histosols definition that could be used to associate it with a specific landscape position. In this case, the formative element “eutric,” was used. Eutric is derived from the Greek word *eu*, meaning good, eutrophic, and/or fertile with a high base saturation (Buol et al., 2011). Fertile soils with high base saturations are common in areas where there are enough soil nutrients for plants and commonly associated with depositional areas where soil nutrients have washed in from upslope areas. Within the landscape, this soil class would most likely occur on footslopes. Similarly, Humic Gleysols are described as having an umbric epipedon, characterized by a dark colored surface horizon with low natural base saturation. The dark surface soil color is the result of organic matter accumulation. Gleysols are saturated with water and reduced at some period during the year and there are indicators that reduction of Fe occurs. Typical slope position for this soil class would be footslopes where there is periodic flooding.

On the other hand, Vertic Fluvisols are soils developed from recent alluvial deposits with vertic properties. Vertic soil properties are commonly associated with smectitic clays, which are sticky and subject to shrinking and swelling on drying and wetting. Conditions conducive for the formation and stability of smectites in soil environments include high amounts of Si and Mg activity, basic pH, and poor drainage (Reid-Soukup and Ulery, 2002). Poor drainage is associated with toeslopes, which are almost always under aquic conditions (Fig. 5.6).

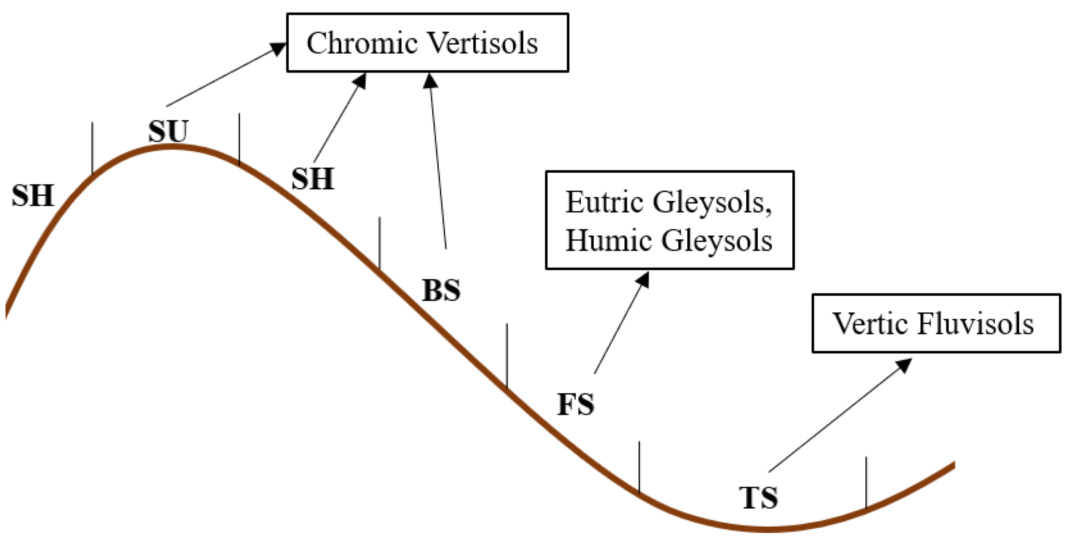


Figure 5.6: Variation of soil classes within soil map unit SAC1.

Map Unit UVC2

Map unit UVC2 occurs on uplands with an underlying dolerite parent material and includes four soil classes: Orthic Acrisols, Plinthic Acrisols, Chromic Acrisols, and Ferralic Cambisols (Table 5.2). Chromic Acrisols are described as Acrisols with chromas >1.5 and an ochric epipedon. High chroma (yellow and red to brown colors) result from iron oxide minerals found in well-drained areas. Typical slope position of such soil classes would be summits, shoulders and backslopes. Similarly, Ferralic Cambisols are described as having an ochric epipedon that lack both hydromorphic and vertic properties. Typical slope positions would be shoulders and backslopes. Orthic Acrisols also have an ochric epipedon and lack hydromorphic properties similar to the Ferralic Cambisols and would be found on similar landscape positions. Plinthic Acrisols are described as Acrisols characterized by the presence of plinthite within 125 cm below the soil surface. Conditions necessary for plinthite formation are common on summits, shoulders, and backslopes (van Wambeke, 1992; Eze et al., 2014). This map unit could not be disaggregated because soil classes occurred on similar landscape positions; i.e., this soil map unit is a complex rather than an association (Fig. 5.7). In cases where complexes occurred, soil classes were "lumped" together.

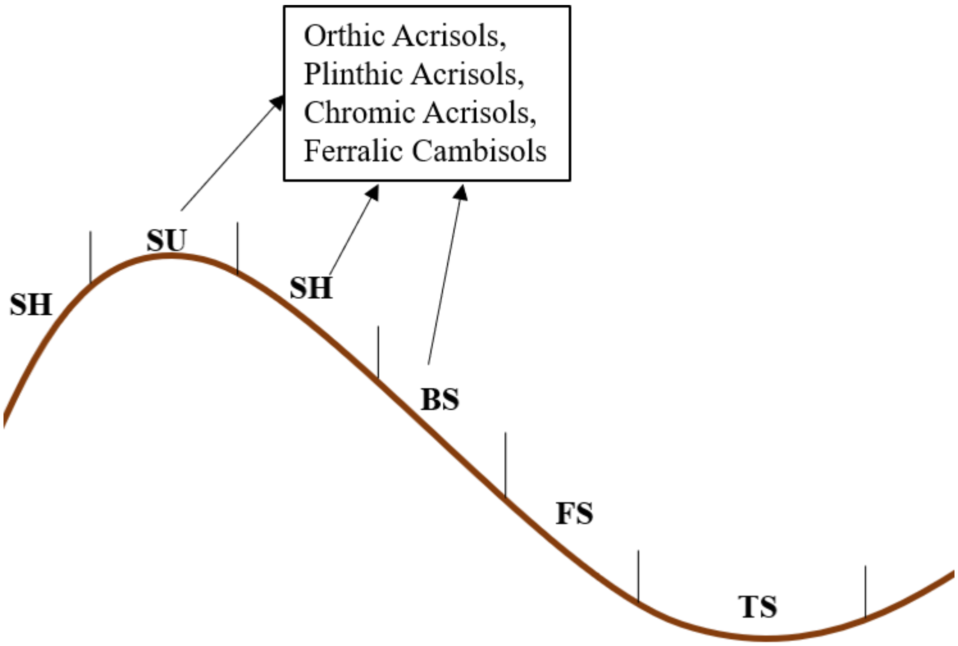


Figure 5.7: Variation of soil classes within soil map unit UVC2.

Develop Digital Landscape and Environmental Model

Generate Terrain Attributes from Digital Environmental Model

Terrain attributes were computed using the System for Automated Geoscientific Analyses (SAGA) Geographical Information System (GIS) software (Conrad et al., 2015) using the 30 m DEM data. Eight terrain attributes were computed: Multiresolution Index of Valley Bottom Flatness (MRVBF), Multiresolution Ridge Top Flatness (MRRTF), planform curvature, profile curvature, relief intensity, slope, topographic position index (TPI), and SAGA topographic wetness index (TWI) (step 4 in Fig. 5.3). Pearson correlation coefficients between the 8 selected terrain attributes at the 76 soil profile pit locations showed that they were not entirely independent. Only terrain attributes with correlation values of less than 0.70 were selected for use in the landscape model, resulting in six terrain attributes namely MRVBF, MRRTF, planform curvature, profile curvature, TPI, and TWI.

Terrain Attribute Clustering

The 6 selected terrain attributes were subjected to K-means clustering (step 5 in Fig. 5.3) performed in SAGA-GIS using the hill-climbing method (Rubin, 1967). This process was intended to mimic the geometry of "fully developed slopes" observed within the landscape (Fig 5.8) (Wood, 1942; King, 1957).

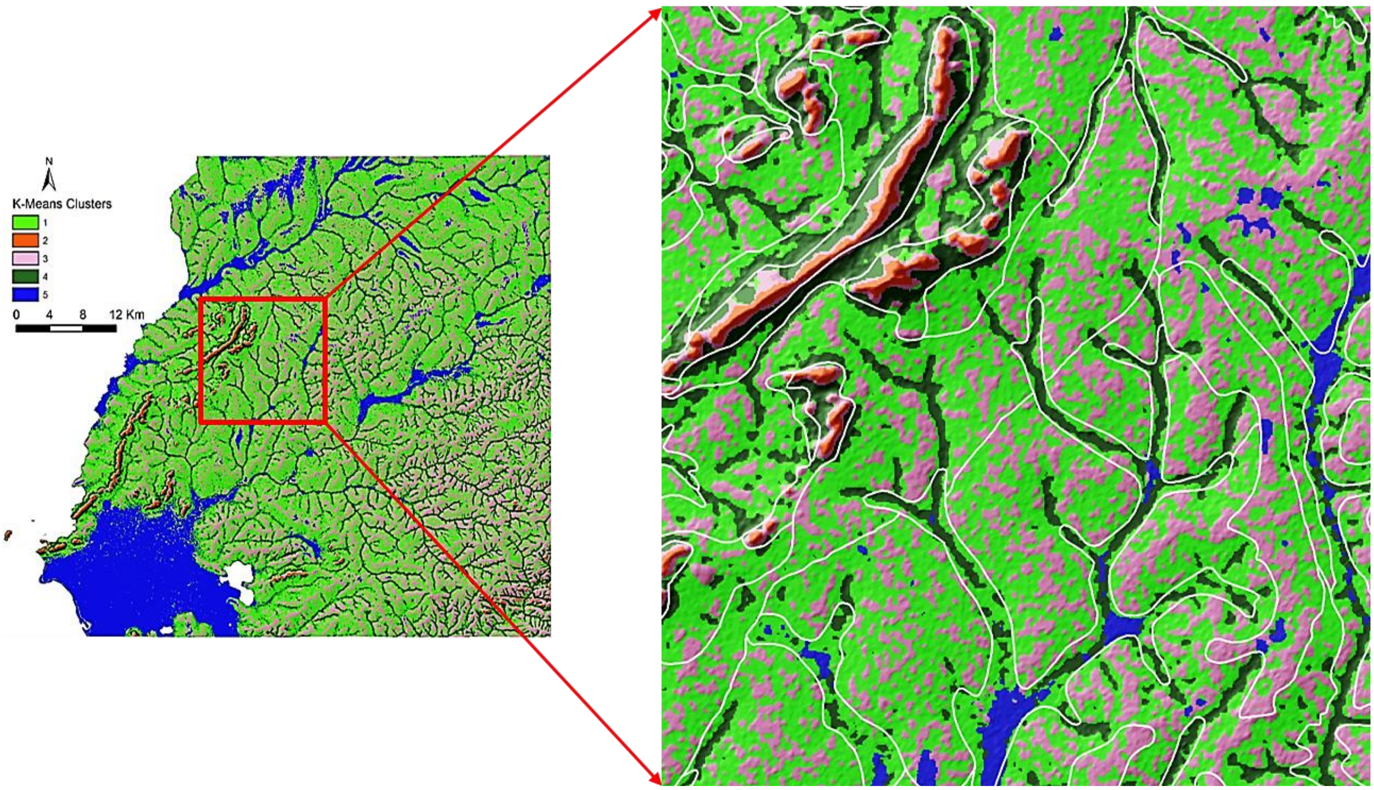


Figure 5.8: On the left, K-means clusters representing different slope positions within the landscape. On the right, a close-up of the cluster map. White lines are soil map unit boundaries from the Busia survey report.

Examination of the K-means cluster map by overlaying it on the satellite imagery showed that the clusters correspond to different slope positions viz. summit, shoulder, backslope, toeslope, and footslope (Fig. 5.9). These slope positions influence geomorphic processes, such as weathering, pedogenesis, and soil water movement, which relate to erosion-sedimentation processes; thus, they determine soil classes and landforms that evolve from these processes (Ruhe, 1975). This theory applies directly to soils classes occurring on landscapes and matches the catena concept (Milne, 1936).

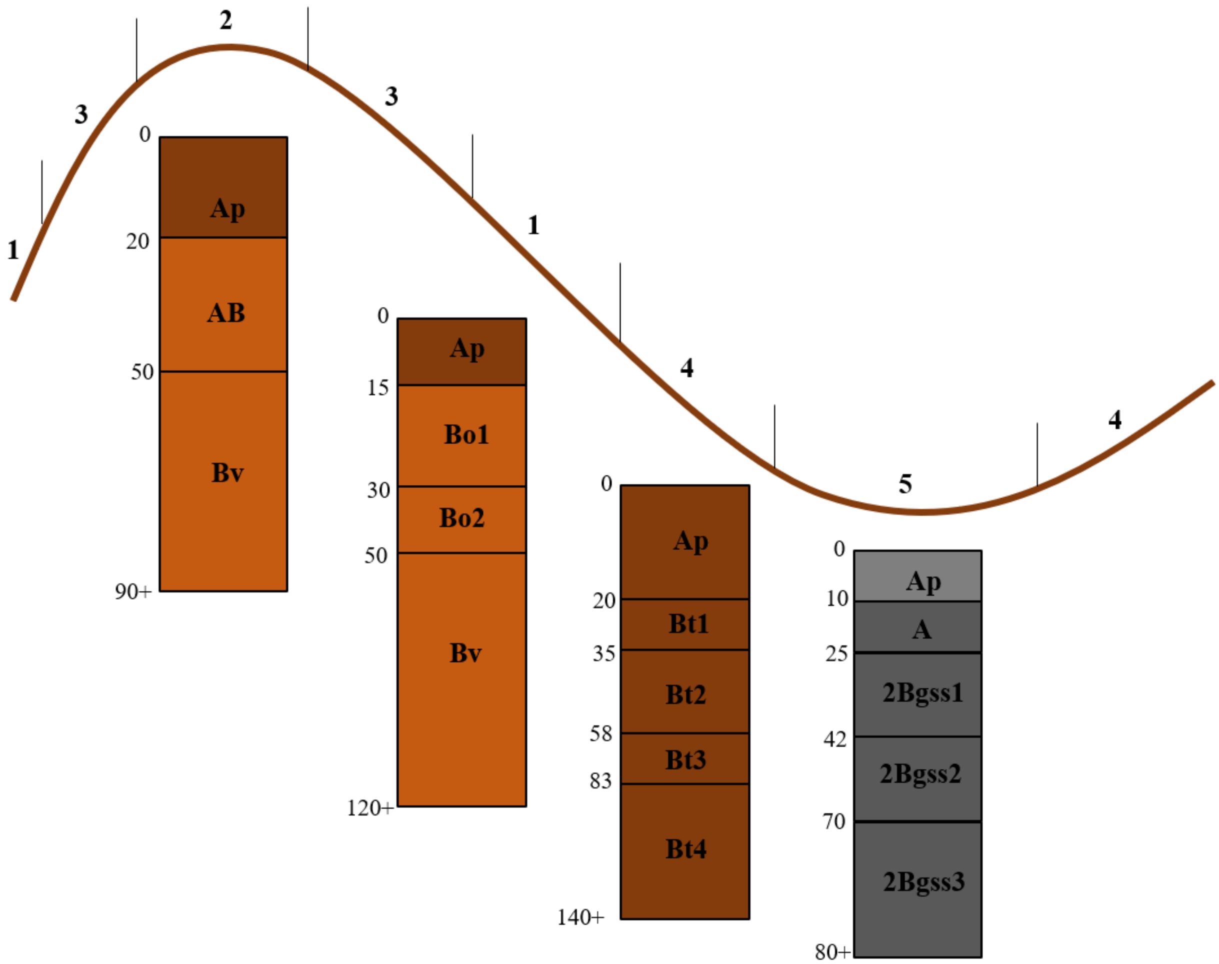


Figure 5.9: Schematic representation of where each cluster occurred on 'fully developed slopes' within the Busia landscape: 1 = Backslope, 2 = Summit, 3 = Shoulder, 4 = Footslope, and 5 = Toeslope (after Wysocki et al., 2000 and Schoeneberger et al., 2012). Summits, shoulders, and backslopes have shallow well drained soils overlying petroplinthite. Soils on footslopes range from deep to very deep with well-developed soils. Soils on the toeslopes are poorly drained and may have vertic properties.

Extract Parent Material Component from Soil Polygon Map

A soil parent material map was extracted from the Busia soil polygon map and rasterized to 30 m resolution (step 6 in Fig 4). This was possible because, as described in section 2.2.2.1, information on the soil parent material was imbedded in the design of the soil map units (Fig. 5.10).

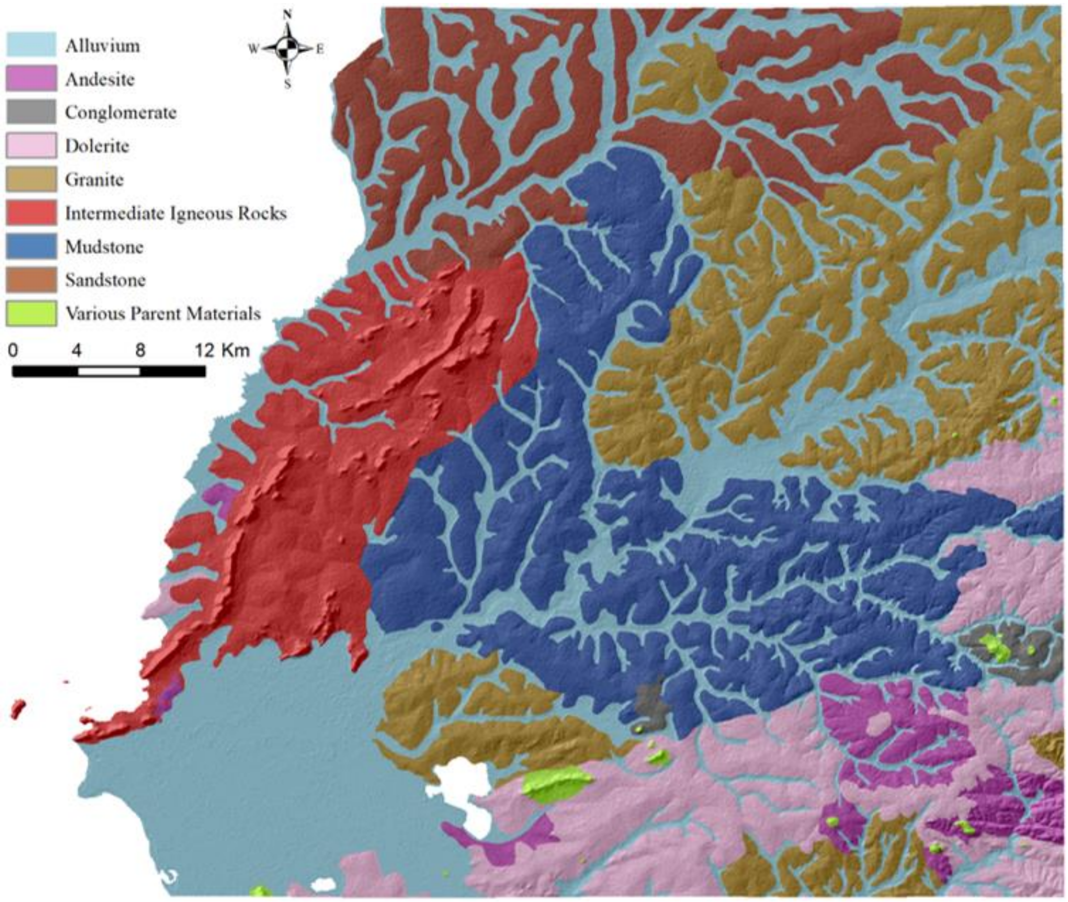


Figure 5.10: Busia soil parent material map.

Generate New Raster from Parent Material and Cluster Maps

The K-means cluster map (Fig. 5.8) was combined with the parent material map (Fig. 5.10) using the raster calculator in ArcMap version 10.6 to generate a new raster map showing all landscape positions occurring within each parent material (step 7 in Fig. 5.3 and Fig. 5.11). This new raster could have up to 45 classes (9 parent materials x 5 slope positions), but since not all combinations of parent material and slope are represented, there were only 44 classes. This process mimics the “physiographic soil survey” methodology developed by the Kenya Soil Survey, which helps to delineate and assist in visualizing the complex relation between landforms, parent material, and soils (Muchena et al., 1982). In the legend, the first digit indicates the parent material (Table 5.3) and the second digit indicates the slope position (Fig. 5.9). For example, code 11 represents soil classes that occur on alluvial parent material (code 10) and occur on backslopes (code 1). Similarly, code 65 represents soil classes occurring on igneous parent material (code 60) and toeslopes (5).

Table 5.3: Parent material codes

Geology	Code number
Alluvium	10
Andesite	20
Conglomerate	30
Dolerite	40
Granite	50
Intermediate igneous rocks	60
Mudstone	70
Sandstone	80
Various parent materials	90

Generate Zonal Statistics in SAGA-GIS

Zonal statistics were computed for each fuzzy soil class in SAGA-GIS (Conrad et al., 2015) (step 8 in Fig. 5.3) to determine the general distribution of each terrain attribute within each fuzzy soil class (Table 5.4). The statistical values were used to develop rules for associating, spatially and quantitively, soils classes within their respective slope positions (see section 5.2.4.2.2).

Harden Fuzzy Soil Class Map

This process used the zonal statistics computed for each fuzzy soil class (Table 5.4) and assigned each pixel an overall membership to the class that had the highest value in the similarity value (Zhu et al., 1996, 2010). This was computed using SoLIM Solutions 2015 software (step 9 in Fig. 5.3 and Fig. 5.12).

Overlay Soil Map on Hardened Map

For visualization, the original soil polygon map was overlaid on the hardened soil class map (step 10 in Fig. 5.3 and Fig. 5.12).

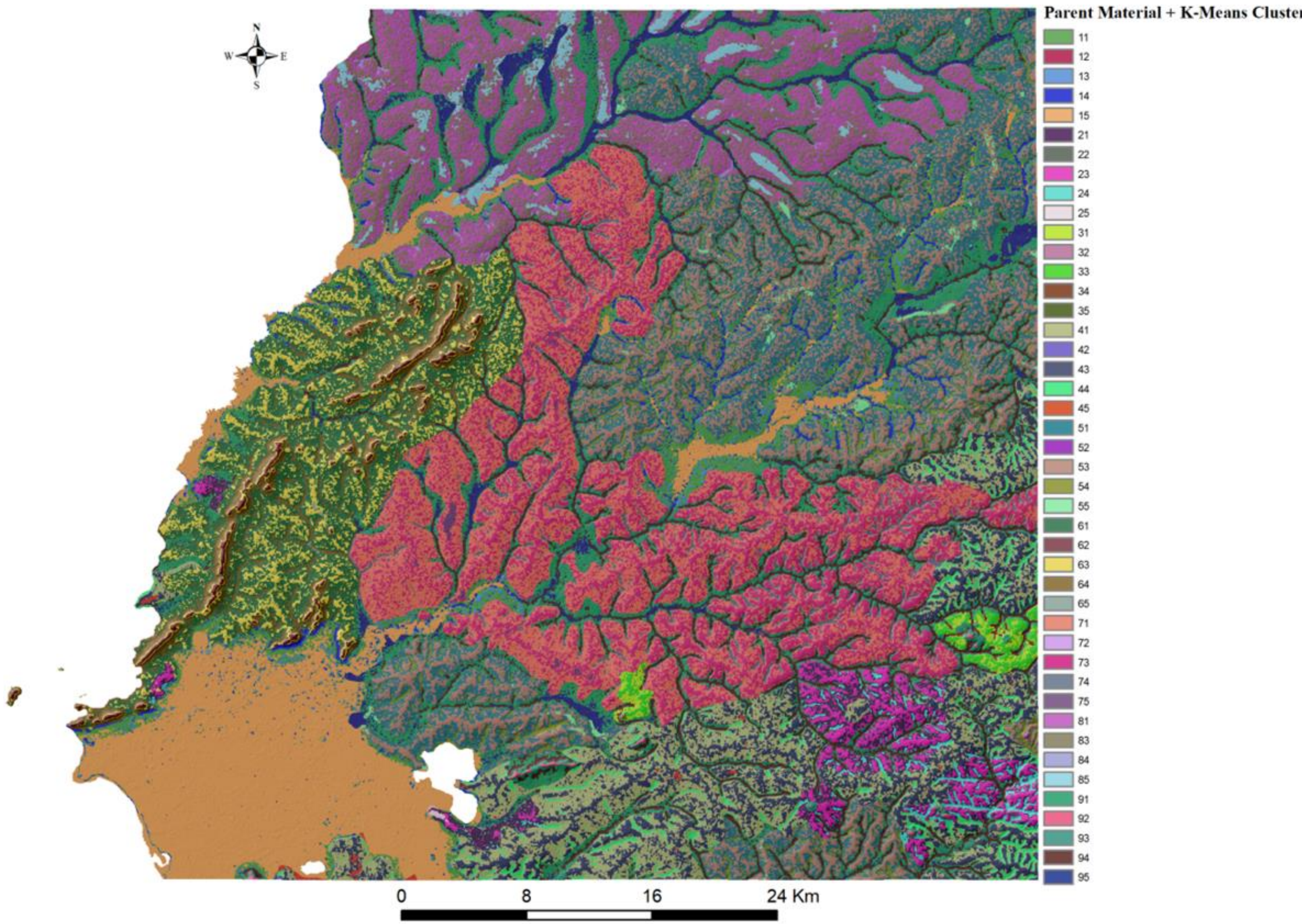


Figure 5.11: Fuzzy soil class map resulted from combining the parent material and the K-means cluster maps.

Table 5.4: Zonal statistics for hardening the fuzzy soil class map (soil landscape relationship).

Soil Class	MRRTF		MRVBF		PlanCurv		ProfCurv		TPI		TWI	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
11	0.251	0.551	2.124	1.213	-0.00003	0.00033	-0.00017	0.00036	-1.022	1.449	6.197	0.614
12	0.000	0.000	0.000	0.000	0.00382	0.00122	0.00267	0.00191	18.028	3.834	3.082	0.093
13	1.256	1.101	0.983	1.190	0.00040	0.00039	0.00031	0.00039	2.758	1.474	5.844	0.694
14	0.126	0.640	2.790	1.234	-0.00033	0.00042	-0.00063	0.00058	-4.833	2.106	6.188	0.684
15	1.857	1.664	5.921	1.123	0.00002	0.00024	-0.00002	0.00027	-0.042	1.562	6.451	0.759
21	0.198	0.558	1.074	0.729	-0.00010	0.00049	-0.00007	0.00044	-0.565	1.457	5.816	0.606
22	0.541	0.724	0.063	0.121	0.00231	0.00150	0.00303	0.00134	19.518	2.994	4.354	0.632
23	0.671	1.027	0.518	0.491	0.00056	0.00072	0.00054	0.00068	4.260	3.063	5.424	0.766
24	0.017	0.054	0.727	0.713	-0.00098	0.00106	-0.00103	0.00105	-7.383	4.264	5.492	0.714
25	3.227	1.395	5.192	1.093	0.00007	0.00032	0.00017	0.00042	1.170	1.706	6.559	0.734
31	0.266	0.766	1.081	0.605	-0.00011	0.00044	-0.00002	0.00045	-0.383	1.395	5.818	0.537
32	0.257	0.446	0.002	0.003	0.00297	0.00015	0.00094	0.00034	17.834	1.424	3.832	0.325
33	1.029	1.380	0.679	0.513	0.00039	0.00048	0.00042	0.00050	3.405	1.904	5.707	0.643
34	0.020	0.060	0.764	0.622	-0.00085	0.00080	-0.00082	0.00075	-5.980	2.742	5.585	0.629
35	4.759	0.097	2.529	0.282	-0.00035	0.00036	0.00004	0.00023	-0.848	0.906	6.126	0.310
41	0.337	0.863	1.238	0.665	-0.00009	0.00040	-0.00005	0.00040	-0.436	1.397	5.956	0.544
42	0.400	0.547	0.017	0.033	0.00229	0.00115	0.00229	0.00128	19.094	3.380	3.993	0.601
43	1.255	1.427	0.710	0.566	0.00039	0.00048	0.00039	0.00047	3.138	1.947	5.798	0.702
44	0.038	0.233	1.304	0.915	-0.00076	0.00071	-0.00071	0.00062	-5.763	2.533	5.878	0.598
45	3.866	1.067	3.989	1.657	-0.00002	0.00029	-0.00003	0.00034	0.009	1.714	6.507	0.677
51	0.490	1.030	1.337	0.661	-0.00009	0.00039	-0.00005	0.00039	-0.345	1.373	6.003	0.541
52	0.281	0.458	0.004	0.008	0.00305	0.00121	0.00268	0.00121	22.474	5.061	3.892	0.538
53	1.573	1.527	0.770	0.611	0.00035	0.00042	0.00035	0.00042	2.830	1.665	5.922	0.686
54	0.039	0.219	1.598	0.887	-0.00070	0.00059	-0.00064	0.00050	-5.236	1.967	6.055	0.506
55	3.646	1.521	3.600	1.162	0.00001	0.00033	0.00004	0.00036	0.425	1.627	6.481	0.673
61	0.303	0.732	1.245	0.725	-0.00004	0.00041	-0.00009	0.00044	-0.578	1.467	5.882	0.649
62	0.056	0.203	0.004	0.033	0.00369	0.00242	0.00372	0.00253	29.983	11.471	3.426	0.419
63	1.119	1.297	0.595	0.592	0.00048	0.00078	0.00028	0.00072	3.364	2.915	5.567	1.057
64	0.032	0.110	0.947	1.042	-0.00047	0.00088	-0.00120	0.00109	-7.179	4.726	4.974	1.028
65	2.162	1.556	4.250	0.880	0.00004	0.00032	0.00002	0.00033	0.153	1.713	6.428	0.644
71	0.474	1.039	1.332	0.654	-0.00009	0.00040	-0.00005	0.00040	-0.349	1.397	5.973	0.530
73	1.439	1.457	0.724	0.578	0.00035	0.00042	0.00036	0.00042	2.820	1.582	5.891	0.665
74	0.040	0.262	1.437	0.913	-0.00077	0.00063	-0.00071	0.00056	-5.511	2.188	5.921	0.471
75	3.831	1.409	3.530	1.093	0.00000	0.00032	0.00000	0.00035	0.004	1.597	6.426	0.652
81	0.893	1.413	1.624	0.654	-0.00006	0.00034	-0.00004	0.00036	-0.251	1.299	6.125	0.553
83	2.244	1.585	0.811	0.704	0.00027	0.00034	0.00029	0.00035	2.236	1.286	6.081	0.679
84	0.126	0.647	1.992	0.942	-0.00058	0.00047	-0.00055	0.00040	-4.637	1.338	6.077	0.468
85	3.811	1.519	4.456	1.462	0.00002	0.00031	0.00002	0.00033	0.289	1.514	6.365	0.636

Table 5.4 continued

Soil Class	MRRTF		MRVBF		PlanCurv		ProfCurv		TPI		TWI	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
91	0.187	0.401	1.737	1.205	-0.00002	0.00038	-0.00017	0.0004	-0.999	1.478	6.054	0.603
92	0.259	0.495	0.020	0.061	0.00272	0.00164	0.00285	0.00168	23.347	6.455	3.900	0.508
93	0.681	0.888	0.595	0.707	0.00052	0.00054	0.00041	0.00054	3.204	2.311	5.591	0.760
94	0.046	0.250	2.277	1.388	-0.00046	0.00057	-0.00080	0.00059	-5.529	2.250	6.081	0.604
95	1.555	2.007	4.665	0.914	-0.00001	0.00029	-0.00009	0.00032	-0.756	1.647	6.514	0.686

Soil class = specific soil type due to the weathering of a specific parent material occurring on a specific topographic position. The first digit represents geology/parent material. The second digit represent slope positions (1 – backslope; 2 – summit; 3 – shoulder; 4 – footslope; 5 – toeslope). For example, soil class 11 denotes a soil of alluvial parent material (code 10) occurring on backslope (code 1). Similarly, soil class 65 denotes a soil of igneous parent material (code 60) found on toeslopes/ bottomlands (code 5).

Soil Class Map

Use Rules to Map Soil Classes within each Soil Map Unit

To map out the soil classes within each soil map unit, each soil map unit was clipped from the hardened soil map. The rules generated in step 3 of Fig. 5.3 were then used to map soil classes to the hardened soil classes. For example, the HIP map unit in Fig. 5.12 has one soil class, namely Lithosols, stony phase. All hardened soil classes occurring on summits (62), shoulders (63), and backslopes (61) were remapped as Lithosols. This mapping process was repeated separately for each of the 52 soil map units because different soil classes occurred in different soil map units. The 52 disaggregated raster maps were then merged together again to generate the final soil class map for the study area (step 11 in Fig. 5.3; Fig. 5.12).

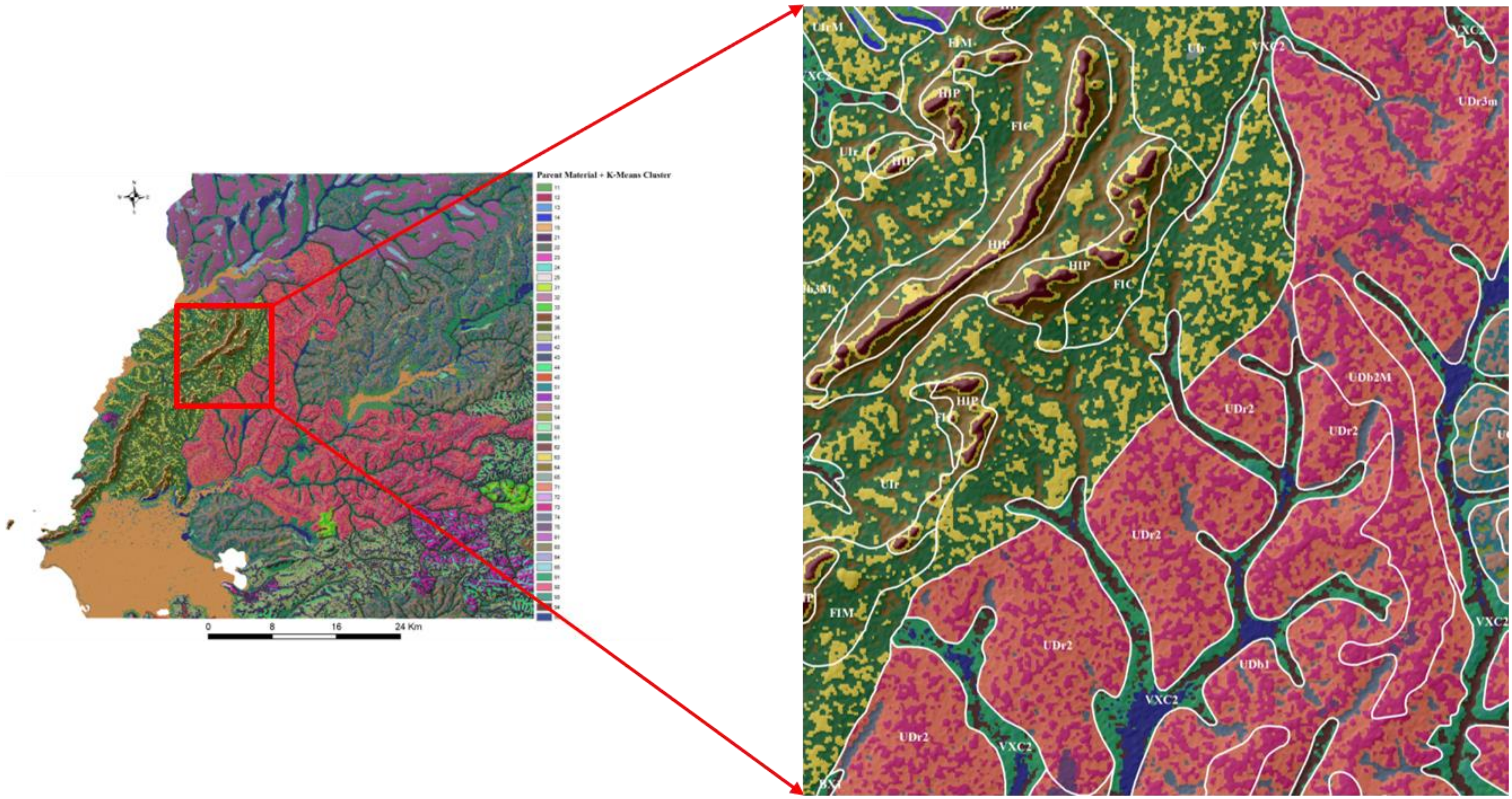


Figure 5.12: A close up of the Busia soil polygon map overlaid on the hardened soil class map.

Evaluate Soil Class Map using Soil Profile Data Mined from Soil Survey Report

The disaggregated soil class map (Fig. 5.13) was evaluated (step 12 in Fig. 5.3) by using the Kappa coefficient, K, to measure the fidelity between the observed soil classes for the 76 soil profiles (see section 5.2.2.1) and predicted soil classes from the disaggregated soil map (Fig. 5.13) (Landis and Koch, 1977). The Kappa coefficient is based on the difference between how much agreement there is between what is actually present, the "observed" agreement, and how much agreement would be expected to be present by chance alone. The "expected," agreement, K, is calculated from equation:

$$\mathrm{K} = \frac{\mathrm{P_o} - \mathrm{P_e}}{1 - \mathrm{P_e}} \qquad [5.1]$$

where P_o is the overall or observed accuracy, and P_e is the expected accuracy. Expected accuracy is calculated from the equation below:

$$\mathrm{P_e} = \sum_{i=1}^{n} \left(\frac{colsum_i}{N}\right) + \left(\frac{rowsum_i}{N}\right) \qquad [5.2]$$

where *N* is the total number of observations, *n* is the number of classes, $colsum_i$ refers to the total number of observations for each soil class, and $rowsum_i$ refers to the total number of predictions for each soil class.

5.2 Results

5.2.1 Disaggregated Soil Class Map

Not all soil map units could be disaggregated. Some soil map units had complex soil classes (see section 2.4.1.3.1.3). In such cases, soil classes had to be "lumped" together. Of the 76 soil profile data, only 48 could be used for evaluation because they occurred on completely disaggregated soil map units. The overall accuracy of the disaggregated soil class map was 58%

with a Kappa coefficient of 0.54, indicating moderate agreement between the actual soil classes and the predicted soil classes, i.e., how much agreement was actually present, compared to how much agreement would be expected by chance alone (Fig. 5.13).

These results, however, do not take into consideration similarity between soil classes. For example, if a soil class is predicted as a Pellic Vertisol while the observation made was a Chromic Vertisol, there is very little distinction between the two soil classes except for color. Pellic subgroups are darker that the Chromic subgroups. Statistically, the prediction is wrong, but the soil classes are both Vertisols and exhibit shrink swell properties, which in this case, are more important than slight differences in color. A similar case applies to Orthic and Chromic Acrisols.

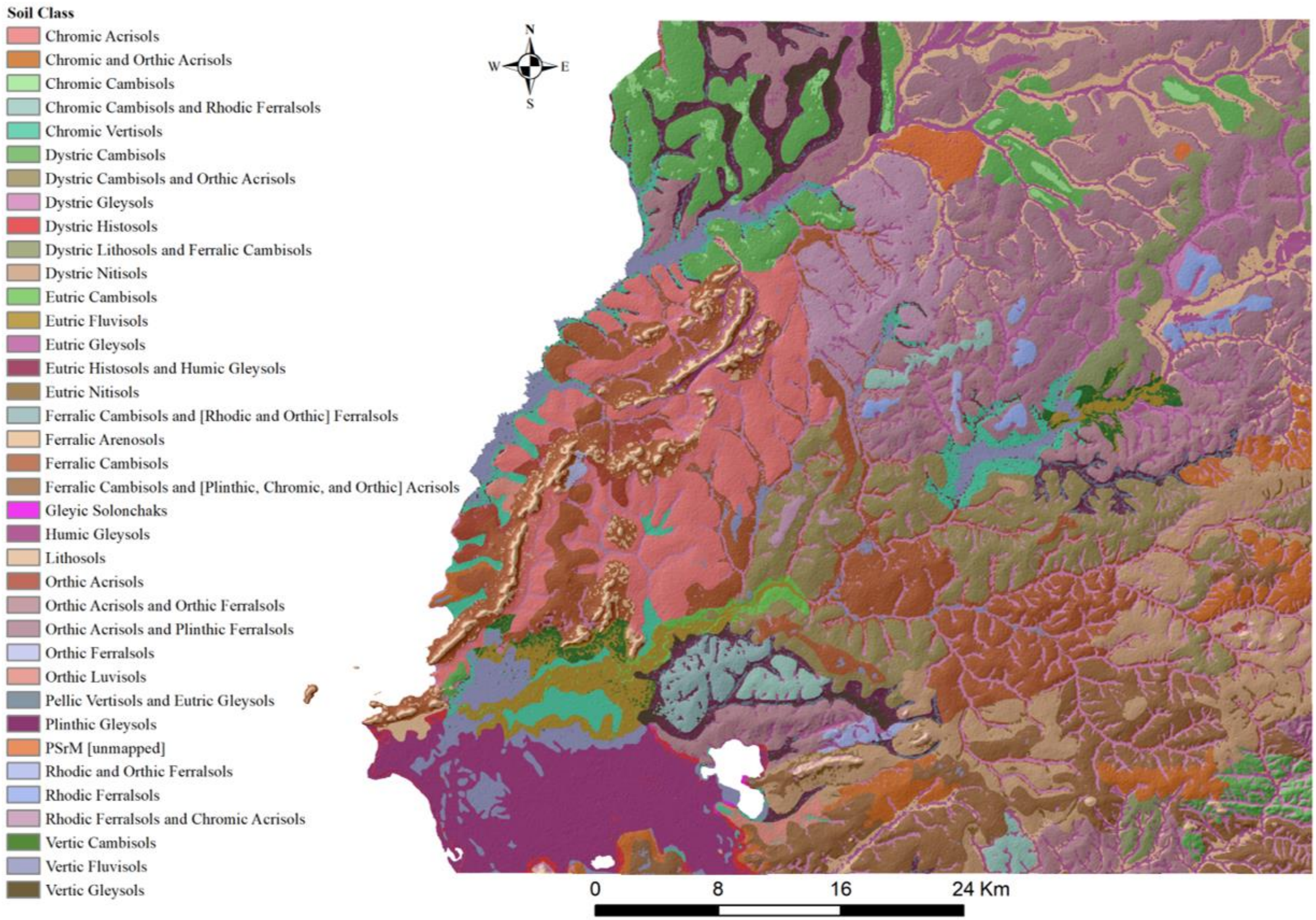


Figure 5.13: Disaggregated soil class map of the Busia area.

5.3 Discussion

In this study we demonstrate how information within a legacy soil survey report can be used to disaggregate a traditional soil polygon map into its component soil classes using a DSM approach. Our approach focused on using the information describing map unit polygons to create a raster-based map of soil classes across the landscape.

A number of challenges arose from this approach. The order in which we discuss each challenge does not reflect its importance. We start with the descriptions of the soil classes within the legacy soil data.

The first challenge was that disaggregation relied heavily on the descriptions within the legacy soil survey report because these descriptions were used to discern where a given soil class was most likely to occur on the landscape. Three types of information existed within the legacy survey report. First, in most cases, soil class descriptions were not explicit, and as a result the soil class' definitions based on the legend of the Soil Map of the World (FAO/UNESCO, 1974) had to be used to discern where soil classes would most likely occur on the landscape. This may have introduced errors since our interpretation of where the soil classes occurred with the map units may have differed from that of the soil scientists who made the original survey. Second, in two soil map units, soil classes were explicitly described in such a way that the information could be used to disaggregate soil map units using the map unit description alone (see section 5.2.4.1.3.1.1 for an example). This was an ideal situation with regards to disaggregation. Third, in five cases, soil classes occurring within soil map units were not mentioned within the soil survey report. In these cases, the homosoil methodology (Mallavan et al., 2010) was implemented. This involved extrapolating soil information from areas with similar soil forming factors to areas of interest where no prior soil information existed. For example, within the legacy soil survey report, map unit SAC1 is described as having four soil classes (section 5.2.4.1.3.1.2) whereas soil classes existing within map unit SAC2 were not mentioned. Since both soils have similar soil-forming factors, i.e., occur in swamps and are of alluvial parent material, the soil classes that occur within SAC1 were assumed to likely occur in SAC2 as well. One soil map unit (PSrM), however, could not be disaggregated because not only was the soil class(es) existing within this map unit not mentioned but also the homosoil approach could not be applied.

The second challenge was related to surveyor bias and the amount of detail that the surveyors described. Soil surveyors are commonly described as either "splitters" or "lumpers" (McKenzie et al., 2008). Splitters try to map all the variation they observe in the field, whereas lumpers believe it is important to place closely associated soil classes together as a single class because the soil classes may function similarly and there is less risk of being incorrect if they are mapped together. Both cases were observed (sections 5.2.4.1.3.1.1 and 5.2.4.1.3.1.3). The main challenge was to disaggregate more than three soil classes within a soil map unit, and this was related to the resolution of the 30 m DEM used in this study. Areas with high relief, such as hills and valleys, performed best because in these cases it was possible to associate soil classes with their respective slope positions. In areas where the relief was gentle, as in swamps and uplands, the DEM failed to provide enough topographic detail for disaggregation.

The disaggregation of the Busia soil map also highlighted two other challenges associated with disaggregating a traditional soil map. Firstly, with the polygon-based model, only soil bodies that were large enough to be delineated and labeled clearly were shown on the resulting map. Therefore, the level of detail was limited by the scale of the map, not by what the soil scientist observed. This was evident for soil classes within soil map units that extended to other adjacent map units. This is understandable because traditional soil mapping was mainly manual, in which soil surveyors had to delineate the extent of soil bodies based on visual interpretation of environmental conditions using aerial photographs or fairly low-resolution topographic maps. As a result, the delineation of soil landscape units may not reflect the complete knowledge of the surveyor. Secondly, soil polygons were depicted as discrete polygon units within definite boundaries (Zhu et al., 2004; Odgers et al., 2014). This affected the disaggregation process because soil classes occurring within a soil map unit were assumed to occur only within the geographic extent of that specific polygon.

5.2.2 Conclusion

Spatial disaggregation provides an opportunity for representing the variation of soil types within soil map units. This study presented an approach for disaggregating a traditional soil polygon map, the *Reconnaissance Soil Map of the Busia Area*, utilizing digital soil mapping techniques together with descriptions within a soil survey report. It was possible to disaggregate areas of interest by careful interpretations of the soil class descriptions within the soil survey report, but the approach relies heavily on the resolution of the DEM and the type of information available within the soil survey report.

CHAPTER 6. DELIVERY OF SPATIALLY EXPLICIT SOILS INFORMATION FOR THE BUSIA AREA, KENYA.

Abstract

As population continues to rise, smallholder farmers in Sub-Saharan Africa face increasing challenges for obtaining sufficient food, fiber, and fuel. Efficient and effective extension services are needed to improve Africa's agriculture. This study explored how to leverage the *Reconnaissance Soil Survey of the Busia Area (quarter degree sheet No. 101)* to deliver useful agronomic information via an easy to use cellphone app**.** We tested the feasibility of delivering one of these maps in the field in rural western Kenya using the Soil Explorer app installed on an iPad Mini that accessed a server via the cell phone network. The Soil Explorer platform provides the ability to deliver soils information on the go. Voluminous reports and unwieldy maps were reduced to portable maps at one's fingertips that can be zoomed, panned, queried to provide information to end users. Mobile information delivery platforms like Soil Explorer open up the possibility of delivering timely and useful agronomic information to farmers at low cost compared to traditional agricultural extension.

6.1 Introduction

Africa currently has the highest prevalence of the world's undernourished people and has among the lowest agricultural yields in the world (Africa Progress Panel, 2015). By 2050, the continent is expected double in population, when it will be home to about one quarter of the World's people (United Nations Department of Economic and Social Affairs Population Division, 2019). As population continues to rise, smallholder farmers in Sub-Saharan Africa will face increasing challenges to obtaining sufficient food, fiber, and fuel. More than half of Africa's population is directly or indirectly involved in agricultural production (Alliance for a Green Revolution in Africa, 2018). Many are smallholder farmers who own small plots of land and grow one or two cash crops that rely almost exclusively on family labor. These farmers often lack adequate access to the resources needed to increase yields on their farms.

Technology and innovation offer unique opportunities to ease the access of farmers to resources and information. In Kenya, for example, innovations such as M-Pesa (Mas and Radcliffe, 2010), Hello Tractor (Ströh de Martínez et al., 2016), iCow (http://www.icow.co.ke/ accessed on 2/15/2019), m-farm (West, 2012), and Kilimo Salama (Kilimo Salama, 2011), demonstrate the key role technology can play in increasing agricultural resources such as access of rural smallholder farmers to finance, market, labor, and seeds.

Even though these technological advances are addressing some of the agronomic challenges facing the continent, the majority of the technologies and innovations are tailored in such a way that they contain technical information that farmers cannot understand. Farmers just want to know what crop(s) is/are suitable for planting on their farms or what fertilizers to use to increase yields (personal conversations with Dr. Joseph J. Mamlin, director emeritus, AMPATH, Eldoret, Kenya, 4/20/2016).

Technologies, however, that deliver soil information still lag behind. Available soils information in Kenya hosted by government agencies and various research institutions is mostly free for public access, although some requires an access fee. This information, however, is often under-used or even neglected because soil maps, legends and reports are not presented in an accessible, purpose-oriented, user friendly format. Soil maps, for example, are frequently difficult to use because of excessive cartographic detail that obscures the more general soil distribution patterns and potentials. In some cases, the style of presentation and reproduction, as black and white copies, lacks appeal, leading the user to underestimate the quality of information available. Soil maps also get misused for solving problems and making decisions outside the range of the project objectives (Zinck, 1995). Map legends and soil survey reports are difficult to comprehend because of the complex language used to name soils. Non-soil specialist users should be given only interpretive maps with simple legends for specific purposes (Geitner et al., 2017).

If farmers have access to useful soil information such as how to farm a given soil type, crop productivity could increase significantly. There is therefore an urgent need to get the right information into the hands of those who need it, when they need it, and in a way that they can use it. To address this challenge, farmers need accurate and timely agronomic information on crop suitability, fertilizer application, and sustainable land use at low or no cost. Archived soils data contains considerable agronomic information that can be repackaged in a simple manner to inform farmers to make agronomically sound decisions on their farms that will increase yields and therefore household income (Minai et al., 2018).

6.2 The Soil Explorer App and Website

The Soil Explorer mobile app and website (https://soilexplorer.net/) was originally developed to aid in teaching in soil, crop, and environmental sciences courses to undergraduate and graduate students at Purdue University within the Department of Agronomy in a way that they can understand (Schulze et al., 2010). Soil Explorer allows anyone, anywhere in the world to access information about soils, landscapes, and natural and man-made features. Maps for ten US states, Kenya, and the Arequipa region of Peru are currently available. One of the focuses of Soil Explorer is on the spatial aspects of soil properties, i.e. how soil properties are distributed over large areas. The app allows one to see and understand spatial patterns within the landscape without spending years mapping soils in the field. It consists of maps from different sources, all of which are georeferenced so that each point on each map corresponds to its equivalent latitude and longitude on the Earth's surface. The app successfully delivers soils information in a way that students can understand the key concepts of soils and landscapes. This study exploits the Soil Explorer platform to deliver spatially explicit soils information on soils, land use, crop suitability, predicted soil property maps, and the disaggregated soil map for the Busia area in a manner that is easy to understand.

6.3 Methodology

6.3.1 Soil Data Source

Soil data was mined from the *Reconnaissance Soil Survey Report of the Busia Area (quarter degree sheet No. 101)* (Rachilo and Michieka, 1991). The Busia legacy soil data consisted of: (1) a soil map of the area that shows the spatial distribution of the soils, (2) detailed geo-located soil laboratory data for a number of profile pits at specific soil depths or for specific soil horizons, (3) a land evaluation key, and (4) crop suitability maps which can be used to determine the potential land use of an area.

6.3.2 Soil Data Capture and Display

Descriptive text within the Busia soil survey report was studied and a list of categories of information that could be mined from the report was made. These categories included physiographic position, total acreage, agro-climatic zone, parent material, meso and micro reliefs, erosion, land use, surface rockiness, general soil description, soil color, soil texture, consistence, chemical properties, clay minerology, diagnostic properties, and soil classification. An example of what is available is shown in Fig. 6.1.

Mapping unit HIP

Total area	:	4180 ha.
Agro-climatic zone	:	I, II, III, IV
Parent material	:	Various intermediate igneous rocks (Samia Hill series)
Relief, macro	:	Hilly, slopes 16 to over 30%
Relief, meso	:	Some inactive termite mounds, 50 m apart
Erosion	:	Slight
Vegetation	:	Bushland
Land use	:	In places cleared and cultivated (cassava and sorghum).
Surface stoniness	:	Stony
Surface rockiness	:	Rocky
Soils, general	:	The soils are relatively young in their stage of weathering and development. They consist of excessively drained, shallow, brown dark reddish to yellowish red, rocky and stony, gravelly clay loam to sandy clay. Profile development is often of an ACR sequence.
Colour	:	Yellowish red to dark reddish brown (5YR 4/6 to 5YR 3/4) throughout.
Texture	:	Clay loam to sandy clay. There is a considerable amount of rocks, stones and gravel in the profile.
Structure	:	Moderate, subangular blocky.
Soil classification	:	LITHOSOLS, stony phase.

Figure 6.1: Categories of soil information within map unit HIP (Rachilo and Michieka,1991).

All these categories were entered into a spreadsheet as column headers. Given that the structure of information available for each map unit was the same, each row in the spreadsheet was assigned to a soil map unit. These rows were then filled with their corresponding information mined from the map unit descriptions (Fig. 6.2).

Soil Unit:	Physiographic Position:	Total Area:	Agro-climatic Zone:	Parent Material:	Macro-Relief:
HIP	Hill	4180 ha.	I, II, III, IV	Various intermediate igneous rocks (Samia Hill Series).	Hilly, slopes 16 to over 30%.
HXP	Hill	1120 ha.	I, II, III, IV	Various igneous and sedimentary rocks, like granites, ar	Hilly, slopes over 16%.
FXr	Footslope	810 ha.	II, III	Dolerites and conglomerates.	Gently undulating, slopes 4-5%, convex, more than 500 m long and regular.
FXb	Footslope	100 ha.	I	Conglomerates.	Gently undulating to undulating, slopes 3-5%, more than 400 m long, linear and regular.
Flr	Footslope	1080 ha.	I, II	Intermediate igneous rocks (Samia Hill series).	Gently undulating to undulating, slopes 4-5%, 300-400 m long, concave and regular.
Flb	Footslope	250 ha.	II	Intermediate igneous rocks (Samia Hill series).	Undulating, slopes 6%, 400 m long, linear and regular.
FIM	Footslope	4790 ha.	I, II, III, IV	Andesites and tuffs.	Gently undulating, slopes 4-5%; 500-600 m long, concave and regular pattern.
FIC	Footslope	1640 ha.	I, II	Colluvium from various igneous rocks.	Gently undulating to undulating, slopes 3-8%, 300-600 m long, convex and regular.
UVr	Upland	9800 ha.	I, II, III	Dolerites.	Very gently undulating to undulating/slopes 2-8%; 300 to more than 500 m long; linear and regular.
UVb1	Upland	7080 ha.	I, II, III	Dolerites.	Gently undulating to undulating, slopes 2-8%, linear to convex, regular pattern.
UVb2m	Upland	1590 ha.	III	Intermediate volcanic rocks: Andesites.	Gently undulating, slopes 2-5%, 100 m long, linear and of regular form.
UVb3P	Upland	2410 ha.	I, II, IV	Andesites.	Rolling to hilly, slopes 9 to more than 16%, 250-500 m long and irregular.
UVC1	Upland	4290 ha.	I, II	Andesites.	Very gently undulating to undulating, slopes 2-8%, more than 200 m long, linear to convex regular pattern.
UVC2	Upland	8670 ha	I, II, III, IV	Dolerites.	Very gently undulating to undulating slopes 2-8%, more than 500 m long, linear to convex, irregular pattern.
Ulr	Upland	5450 ha.	I, II, III, IV	Intermediate igneous rocks (Samia Hill series).	Very gently undulating to undulating, slopes 2-8%, more than 200 m long, convex and of regular pattern.
UlrM	Upland	2255 ha.	I, II, III	Intermediate igneous rocks (Samia Hill series).	Flat to gently undulating, slopes 0-5%.
Ulb1	Upland	340 ha.	II	Intermediate igneous rocks (Samia Hill series).	Flat to gently undulating, linear and regular pattern.
Ulb2M	Upland	1780 ha.	II, III	Intermediate rocks (Samia Hill Series).	Very gently undulating to gently undulating, slopes, 1-3%, 300-400 m long, linear and irregular.
Ulb3M	Upland	7310 ha.	I, II, III	Intermediate igneous rocks (Samia Hill Series).	Flat to rolling, slopes 1-10%, 300-500 m long, linear to convex and regular.
UGrl	Upland	640 ha.	I	Granites.	Flat to gently undulating, slopes 0-4%, 300-400m long, linear and regular.
UGr2	Upland	550 ha.	II	Granites.	Gently undulating, slopes 2-5%, 250-400 m long, linear to convex and regular.
UGr3m	Upland	16130 ha.	I, II	Granites.	Flat to gently undulating, slopes 0-5%, 500 m long, linear and regular.
UGr4M	Upland	1600 ha.	I	Granites.	Flat to gently undulating slope 1-4%, linear to convex and irregular, 200-500 m long.
UGb1	Upland	16041 ha.	I, II	Granites.	Gently undulating to undulating, slope 2-7%, linear to concave and regular.
UGb2	Upland	570 ha.	I, II	Granites.	Gently undulating to rolling, slopes 3-10%, 60-150 m long, linear to convex and irregular pattern.
UGb3M	Upland	6820 ha.	I, II	Granites.	Very gently undulating to gently undulating, slope 1-5%, linear to convex, regular, 100-300-m long.

Figure 6.2: Soil information mined and organized in a spreadsheet. The first column represents the soil map unit. Each row corresponds to a soil map unit with corresponding information mined from the survey report.

Data captured in each row was reformatted into a Hypertext Markup Language (HTML) text string that would display as a popup in Soil Explorer app (Fig. 6.3).

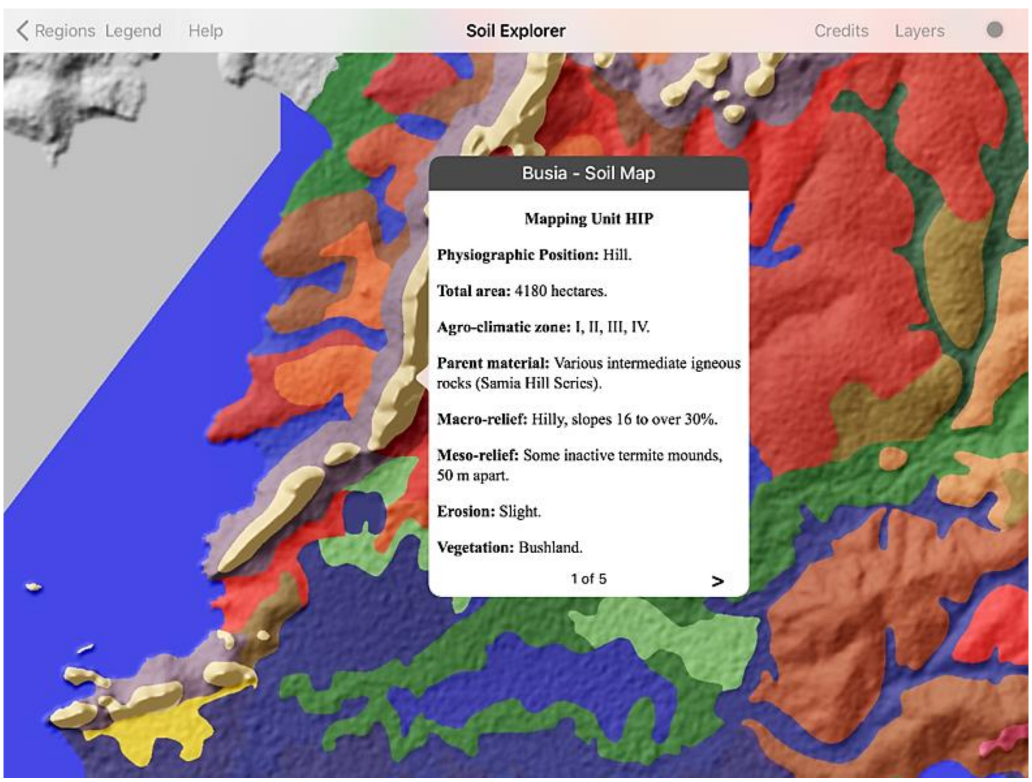


Figure 6.3: Captured soil information displayed on a portable device showing specific soil information of a soil map unit.

This approach was used to display available information for the Busia area including land quality maps, crop suitability maps, predicted soil property maps, and disaggregated soils map of the area using the Soil Explorer app (Fig. 6.4).

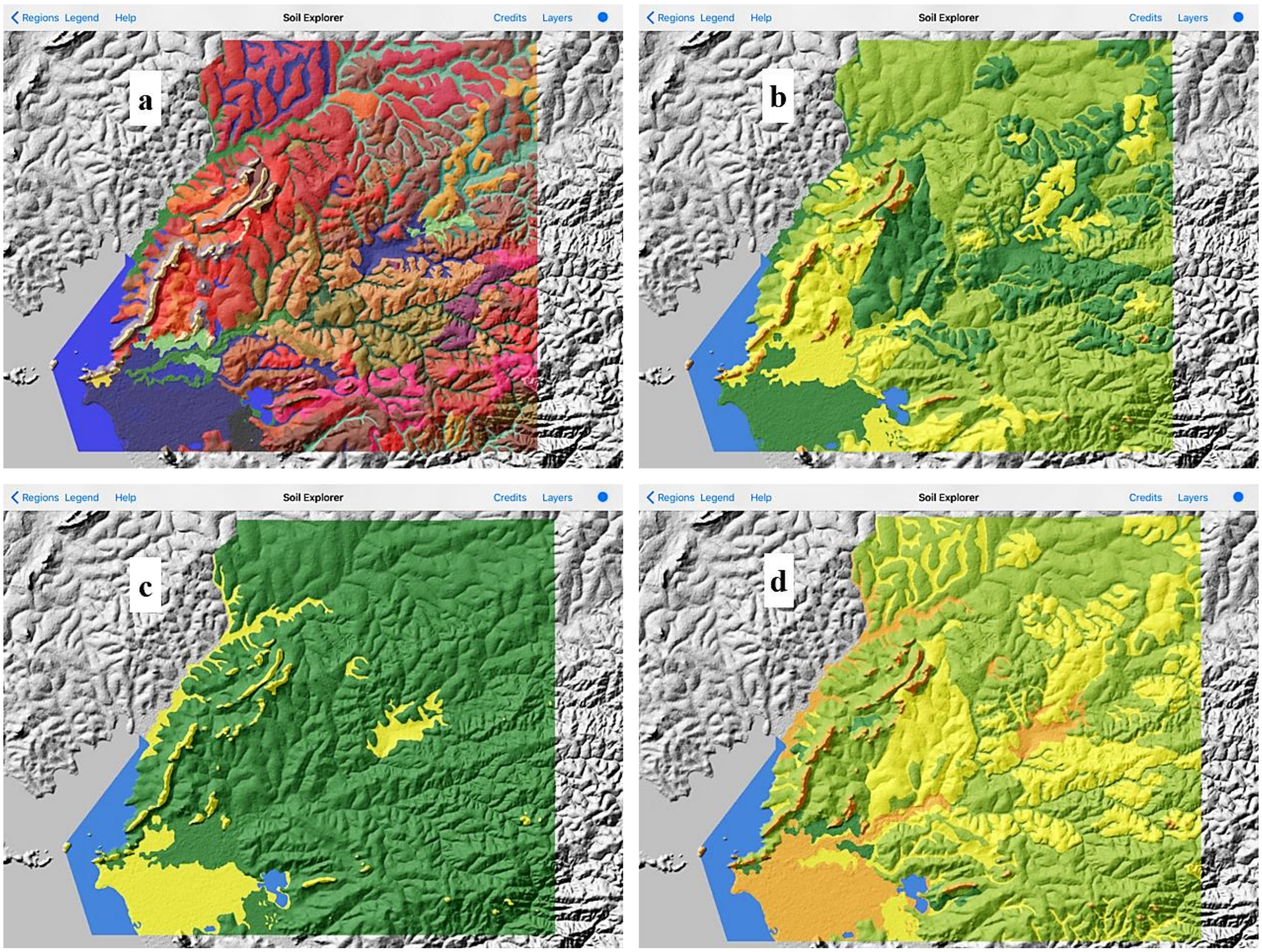


Figure 6.4: Examples of soils information displayed using the Soil Explorer app. (a) Reconnaissance soil map of the Busia Area, (b) availability of moisture for plant growth (c), suitability classification of soils for rainfed maize (*Zea mays* L.), and (d) suitability classification of soils for rainfed beans (*Phaseolus spp.).*

6.1 Results

The feasibility of this approach was successfully tested in rural western Kenya (Minai et al., 2016) using the Soil Explorer app on an iPad Mini equipped with a cellular modem (Fig. 6.5). Broadband is available over as much as 30% of Kenya, with up to 88% of Kenya's population having access to the Internet through their cellphones, thanks to cheaper data plans and the ubiquitous use of mobile money platforms like Mpesa (Communications Authority of Kenya, 2017). This makes the applicability of this approach feasible because the infrastructure is available.

Figure 6.5: Testing the feasibility of the Soil Explorer app in rural western Kenya. Photo by D.G. Schulze, CC-BY 4.0.

6.3.3 Implications of the Soil Explorer Platform

The Soil Explorer platform provides the ability to deliver soils information on the go. Traditional methods of storing soils information in books stored in libraries hinders ease of access of information to the rural smallholder farmers who need it. Soils information commonly exists in big voluminous soil survey reports (i.e., Rachilo and Micheieka, 1991) that are hard to use for non-soil scientists. Mobile apps like Soil Explorer make it possible to take information currently locked up in paper soil survey reports and maps and reformat it and display it in a manner that is easier to comprehend. Voluminous reports and unwieldy maps are reduced to portable maps at one's fingertips that can be zoomed, panned, and queried to provide information to end users.

Agricultural outreach encompasses a wide range of supportive programs that exist around farmers to help them to utilize research findings and newer agricultural innovations and technologies. It includes training, advisory services and technology transfer schemes. The challenge is that current extension services in Kenya are demand driven. Even if the information is freely available, farmers have to request the information from an extension officer. In some cases, farmers have to pay for the information from an extension officer. In both of these scenarios, the access of information to the farmer is reduced. Mobile information delivery platforms like Soil Explorer open up the possibility of delivering timely and useful agronomic information to farmers at low cost compared to traditional agricultural extension. The ability of a farmer to access agronomic information at the touch of a button could be a game changer with respect to current agricultural extension services.

6.4 Conclusion

We leveraged existing legacy soils data to impart useful agronomic data via an easy to use mobile app currently available for both Android and Apple products. By doing so, we proved that spatially explicit, legacy soils data could be delivered via a portable electronic device.

CHAPTER 7. CONCLUSION AND RECOMMENDATIONS

7.1 Introduction

“Data and information are essential building blocks of science. Many types of data, including extant historical data which have newly appreciated scientific importance for the analysis of changes over time, are not being used for research because they are not available in digital formats” (International Council for Science, 2004).

In Africa, the primary concern is long-term food security and food production through smallholder farmers, while conserving the environment and sustaining the capacity of the land resource base: soil (Gobin, 2000). Information on existing soil resources, especially for Africa, is sparse. Kenya fortunately has considerable soils information in the form of traditional soil maps, soil survey reports, soil survey manuals, land evaluation frameworks, soil profile descriptions, and farm management handbooks, collectively known as legacy soil data. These types of data, however, remain in libraries in analogue formats and the probability of such data being lost is very high (Arrouays et al., 2017). Legacy data can be used as meaningful sources of soil information to support digital soil mapping and/or as major components of national environmental programs.

The overall objective of this research was to bring legacy soil data for a selected portion of Kenya ‘back to life’ using digital soil mapping techniques. Methodologies were developed and evaluated for the Busia area in western Kenya to achieve this objective. The Busia area was selected as a suitable setting for this study because: (1) it has accessible legacy soil data at a scale of 1:100,000, (2) agriculture is the main economic activity, (3) high population and poverty densities have strained existing natural resources, and (4) we are personally familiar with the area and can draw on our own field observations for additional context. The specific conclusions of this thesis are grouped into four major parts distinguished on the basis of the methodologies developed in the respective chapters.

7.1.1 Renewal of Archival Legacy Soil Data

The first specific objective was to transform the best available legacy soil survey of a selected portion of Kenya into a digital format (Chapter 3). The *Reconnaissance Soil Survey of the Busia Area (quarter degree sheet No. 101)* was used as the legacy soil data for this study. This legacy soil data was brought back to life in three steps. The first step, *data archeology*, involved locating and cataloging all historical legacy soil data for the Busia area by contacting various agricultural institutions. This process required numerous site visits and assistance of individuals familiar with the desired legacy soil data. The second step, *data rescue*, entailed converting paper copies of data into a digital format by scanning the maps, narrative descriptions, and tables, and storing the information in a database. In the third step, *data renewal*, all the renewed data was brought to modern standards by taking advantage of technological and conceptual advances in geo-information technology.

Careful interpretation of the agronomic information contained within the Busia soil survey report identified the decision matrices used by the original soil surveyors to generate land quality maps and associated crop suitability maps based on the soil map (Rachilo and Michieka, 1991) and the agro-climatic zone map of Kenya (Sombroek et al., 1982). These decision matrixes were carefully reinterpreted and used to generate 10 land quality maps (Appendix A) that showed the ability of the land to perform specific functions without being degraded. Additionally, based on the land quality maps, 19 crop suitability maps (Appendix B) were also generated for the Busia area.

7.1.2 Spatial Prediction of Soil Properties from Archival Legacy Soil Data

The second specific objective was to make spatial predictions of selected soil functional properties by using DSM techniques by using soil profile data mined from the legacy soil survey report (Chapter 4). Soil properties mined from the Busia soil survey report, together with a set of environmental covariates, were used to test three interpolation models: ordinary kriging, stepwise multiple linear regression, and the Soil Land Inference Model. Two types of point soil data were used in this study (Section 4.2.3). Data Set A, which was a collection of data for the A horizons of 76 sampling points, was used as the calibration dataset, and Data Set B, a collection of 48 detailed soil profile descriptions and analytical data, was used as the evaluation dataset.

All three prediction models were based on the *scorpan* model (Section 2.6.1.2) that postulates that one can predict soil properties based on a set of soil forming factors and a numerical model (McBratney et al., 2003). The three prediction models differed in that for ordinary kriging only the soil property data was used to predict soil properties, for the stepwise multiple linear regression, both soil property data and a collection of 23 environmental covariates were used for prediction, and for the Soil Land Inference Model, both the soil property data and a selection of terrain attributes were used for prediction. Environmental covariates were carefully selected based on a review of the literature and reflected the soil forming factors (Section 4.2.3).

Soil organic carbon and soil texture were predicted at a resolution of 30 m. This resolution was better than the currently available 250 m resolution from the AfSIS project (Hengl et al., 2017). Statistically, ordinary kriging performed best in predicting soil organic carbon, clay, and silt, but there was almost no spatial detail. On the other hand, stepwise multiple linear regression performed best in predicting sand. The 95% confidence interval maps showed that the predictions made from ordinary kriging had the highest confidence, while the predictions made from stepwise multiple linear regression had the lowest confidence. Pedologically, the Soil Land Inference Model (SoLIM) was best in capturing soil variability that followed landscape positions.

The limitation, however, of using legacy soil data for digital soil mapping in this study was that the number of data points available limits the use of the numerical models. The three prediction models require more points than were available in order to create a strong relationship between the calibration dataset and the environmental/ terrain attributes used for model calibration. This is a major limitation when using soil property data mined from legacy soil data.

7.1.3 Disaggregation of a Traditional Soil Polygon Map Using a Soil Landscape Rule-Based Approach

The third specific objective was to improve the spatial resolution of the legacy soil map of the study area-using DSM techniques (Chapter 5). The lack of an adequate calibration dataset limited the use of statistical models to downscale the soil polygon map. Therefore, the legend and soil map unit descriptions were used to generate rules that were used to associate soil types with the most likely landscape positions. The soil landscape rule-based model was used to map soil types within soil map units in a process known as disaggregation. This was achieved by exploiting the information in the map legend and the soil map unit descriptions. These rules were applied to a fuzzy soil class map generated from a parent material map and a K-means cluster map and six terrain attributes, namely, multiresolution valley bottom flatness, multi resolution ridgetop flatness, topographic wetness index, topographic positin index, planform curvature, and profile curvature. Terrain attributes were generated from the 30 m SRTM digital elevation model using SAGA-GIS. The result was a soil class map at a resolution of 30 m with an overall accuracy of 58% and a Kappa coefficient of 0.54. The challenge, however, with the soil landscape rule-based method was that it relies heavily on the soil class descriptions contained within the soil survey report and the resolution of the digital elevation model.

7.1.4 Delivery of Spatially Explicit Soils Information

The fourth specific objective was to develop a prototype platform that could deliver spatially explicit soil and agricultural information for the area of the legacy soil survey on a smart phone or tablet (Chapter 6). Digital agronomic soil maps, including the renewed Busia soil map, generated land quality maps, crop suitability maps, and predicted soil property maps, were all put on a portable electronic device using the Soil Explorer app and tested in rural western Kenya. This was a proof of concept that agronomic information can be relayed to the end-user when they need it and in a form that they can understand. The mobile device used in this study was an Apple iPad mini that used the cellphone network to access the data on a server at Purdue University.

7.2 General Recommendations

Legacy soil data can provide basic information on soil and land characteristics useful for various purposes such as determining the suitability for various types of agriculture. The ability to generate maps that demarcate which soils are suitable for specific crops, which soils are susceptible to specific limiting factors such as flooding, salinization, soil erosion etc. (Appendix A), is key to not only improving agricultural production within a region, but also to bolstering efforts to bringing legacy soil survey reports to modern standards using the latest GIS technologies (Chapter 3).

Soil data and reliable soil maps are imperative for environmental management, conservation, and policy. Soil property data existing within legacy soil data may not have necessarily been collected for used in digital soil mapping. In cases, however, where the legacy soil data is geo-located, methods are required to utilize these historical soil data to produce quantitative soil property maps to assess spatial and temporal trends and also to determine where future sampling is required (see Chapter 4). The predictions serve as a basis for judgment about land use and management for areas ranging from small scale (30 m pixel) to large scale (Busia area). These predictions, however, must be evaluated along with economic, social, and environmental considerations before they can be used to make valid recommendations for land use and management.

Legacy soil survey reports are an important part of information used to make workable plans for land management. Such information must, however, be interpreted and presented in a format that is useable to end users. This is because current soil survey reports, even though useful, exist in formats that are not user friendly to non-soil scientists. Mobile platforms such as the Soil Explorer app make it possible to not only relay interpretive maps to end users, but also open the possibility of delivering timely and useful agronomic information to rural smallholder farmers at low cost compared to traditional extension approaches (Chapter 5).

Efforts should, therefore, be made to develop a central open-access portal for Kenya where all existing legacy soil data can be accessed. The literature review revealed that there is no central point where such information can currently be accessed (Chapter 2). There is a lot of duplicity in where to access such data. This makes it hard to find the authoritative data source one is looking for. In certain cases, legacy soil data can exist in more than three different databases. This shouldn't be the case. An authoritative central database should be created by the relevant ministry. A very good example is that of the United States' Department of Agriculture - Natural Resource Conservation Service archived soil survey database (https://www.nrcs.usda.gov/wps/portal/nrcs/soilsurvey/soils/survey/state/ accessed 9/8/2019). In this database, legacy soil survey reports are digitally archived according to the state, and within each state, legacy soil survey reports are archived according to county. Such organization eases the access of such data. The Kenya soil survey should strive to have all existing legacy soil survey data in a central database.

The limitation of legacy soil survey reports is that information contained within such legacy data cannot replace site-specific details that require onsite investigation. Nevertheless, legacy soil data is a valuable tool when onsite data collection is not feasible or cost prohibitive, or as a tool for planning future onsite investigations. Understanding the capability and limitations of the different types of legacy soil data is essential for making the best conservation-planning decisions.

7.2.1 Specific Recommendations

Legacy soil data are at risk of not being used for scientific research because the majority are not available in digital formats or are in danger of being lost because the media on which they are recorded may decay, become corrupted, or be superseded by new software. Therefore, soil scientists need to inventory major collections of extant legacy soil data and set priorities for their rescue and permanent preservation.

For digital soil mapping, more soil property data needs to be mined from existing legacy soil data for other parts of Kenya and added to the Africa Soil Profile (AfSP) database to promote the use of existing soil property data.

More research is needed on how to best deliver spatially explicit soils information using mobile platforms. Success in the use of the Soil Explorer app to deliver agronomic information in rural western Kenya demonstrated that there is a huge potential in delivering soils information using mobile platforms.

Legacy soil data security and integrity must be addressed in the context of procedures for data management. Professional information technology staff is required for both archives and data centers if the maintenance of adequate system and database security and integrity is to be achieved.

REFERENCES

Africa Progress Panel. (2015). Power people planet: seizing Africa's energy and climate opportunities: Africa progress report 2015. Geneva, Switzerland, p.182. Retrieved from https://reliefweb.int/sites/reliefweb.int/files/resources/APP_REPORT_2015_FINAL_low1.pdf [accessed on 8/26/2019].

Ahmed, F. B., & Dent, D. L. (1997). Resurrection of soil surveys: a case study of the acid sulphate soils of The Gambia. II. Added value from spatial statistics. Soil Use and Management, 11:69-76.

Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19:716-723.

Alliance for a Green Revolution in Africa. (AGRA). (2018). Impact, progress, partnerships: 2018 annual report. Nairobi, Kenya, p. 84. Retrieved from https://agra.org/ar-2018/wp-content/uploads/2019/07/AGRA-Annual-Report_v18_FINAL_Print-Ready_LR.pdf [accessed on 8/26/2019].

Amundson, R., Berhe, A. A., Hopmans, J. W., Olson, C., Sztein, A. E., & Sparks, D. L. (2015). Soil and human security in the 21st century. Science, 348:261071-1 – 1261071-6.

Antonić, O., Pernar, N., & Jelaska, S. D. (2003). Spatial distribution of main forest soil groups in Croatia as a function of basic pedogenetic factors. Ecological modelling, 170:363-371.

Aore, W. W. (1995). Kenya. *In* Soil survey: perspectives and strategies for the 21st century (No. 21). Food & Agriculture Org. Rome.

Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B., Hong, S. Y., Lagacherie, P., Lelyk, G., McBratney, A. B., McKenzie, N. J., dL Mendonca-Santos, M., Minasny, B., Montanarella, L. Odeh, I. O. A., Sanchez, P. A., Thompson J. A., & Zhang, G-L. (2014). GlobalSoilMap: Toward a fine-resolution global grid of soil properties. Academic Press, 125:93-134.

Arrouays, D., Leenaars, J. G., Richer-de-Forges, A. C., Adhikari, K., Ballabio, C., Greve, M., Grundy, M., Guerrero, E., Hempel, J., Hengl, T., Heuvelink, G., Batjes, N., Carvalho, E., Hartemink, A., Hewitt, A., Hong, S-Y., Krasilnikov, P., Lagacherie, P., Lelyk, Libohova, Z., Lilly, A., McBratney, A., McKenzie, N., Vasquez, G. M., Mulder, V. L., Minasny, B., Montanarella, L., Odeh, I., Padarian, J., Poggio, L., Roudier, P., Saby, N., Savin, I., Searle, R., Solbovy, V., Thompson, J., Smith, S., Suleman, Y., Vintila, R., Rossel, R. V., Wilson, P., Zhang, G-L., Swerts, M., Oorts, K., Karklins, A., Feng, L., Navarro, A. R. I., Levin, A., Laktioniva, T., Dell'Acqua, M., Suvannang, N., Ruam, W., Prasad, J., Patil, N., Husnjak, S., Pasztor, L., Okx, J., Hallett, S., Keay, C., Farewell, T., Lilja, H., Juilleret, J., Marx, S., Takata, Y., Kazuyuki, Y., Mansuy, N., Panagos, P., Liedekerke, M. V., Skalsky, R., Sobocka, J., Kobza, J., Eftekhari, K., Alavipanah, S. K., Moussadek, R., Badraoui, M., Da Silva, M., Paterson, G., Goncalves, M. da C., Theocharopoulos, S., Yemefack., M., Tedou, S., Vrscaj, B., Grob, U., Kozak, J., Boruvka, L., Dobos, E., Taboada, M., Moretti, L., & Rodriguez, D. (2017). Soil legacy data rescue via GlobalSoilMap and other international and national initiatives. GeoResJ, 14:1-19.

Ashtekar, J. M. (2014). Digital methods for field scale soil mapping (Doctoral dissertation, Purdue University), p. 148.

Austin, M. P., Belbin, L., Meyers, J. A., Doherty, M. D., & Luoto, M. (2006). Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory. Ecological Modelling, 199:197-216.

Batjes, N. H. (2003). A taxotransfer rule based approach for filling gaps in measured soil data in primary SOTER databases (version 1.0). Technical report 2003/03 ISRIC World Soil Information, Wageningen.

Batjes, N. H. (2009). Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. Soil Use and Management, 25:124-127.

Batjes, N. H., & Bridges, E. M. (1994). Potential emissions of radioactively active gases from soil to atmosphere with special reference to methane: development of a global database (WISE). Journal of Geophysical Research: Atmospheres, 99:16479-16489.

Baxter, S. J., & Crawford, D. M. (2008). Incorporating legacy soil pH databases into digital soil maps, pp.311-318. In: A. E. Hartemink, A. B. McBratney, M. L. Mendonça-Santos (Eds.), Digital Soil Mapping with Limited Data. Springer, The Nethernlands.

Baxter, S. J., Oliver, M. A., & Archer, J. R. (2006). The spatial distribution and variation of available Phosphorus in agricultural topsoil in England and Wales in 1971, 1981, 1991 and 2001, 31:477-486. In: P Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in Soil Science

Bellamy, P. H., Loveland, P. J., Bradley, R. I., Lark, R. M., & Kirk, G. J. (2005). Carbon losses from all soils across England and Wales 1978–2003. Nature, 437: 245-248.

Bishop, T. F. A., & McBratney, A. B. (2001). A comparison of prediction methods for the creation of field-extent soil property maps. Geoderma, 103:149-160.

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

Bouyoucos, G. J. (1962). Hydrometer Method Improved for Making Particle Size Analyses of Soils 1. Agronomy journal, 54(5), 464-465.

Braun, H. M. H., & Van de Weg, R. F. (1977). Proposals for rating of land qualities. Internal Communications, (7). Kenya Soil Survey, Nairobi.

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

Bui, E. N., & Moran, C. J. (2001). Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. Geoderma, 103:79-94.

Bui, E. N., Loughhead, A., & Corner, R. (1999). Extracting soil-landscape rules from previous soil surveys. Soil Research, 37:495-508.

Buol, S. W., Southard, R. J., Graham, R. C., & McDaniel, P. A. (2011). Soil genesis and classification. John Wiley & Sons., pp.531.

Cambule, A. H., Rossiter, D. G., Stoorvogel, J. J., & Smaling, E. M. A. (2015). Rescue and renewal of legacy soil resource inventories: A case study of the Limpopo National Park, Mozambique. Catena, 125:169-182.

Carré, F., & Boettinger, J. L. (2008). Synthesis and priorities for future work in digital soil mapping. Pages 399-403 *in* A. E. Hartemink, A. B. McBratney, and M. L. Mendonc¸a Santos, eds. Digital soil mapping with limited data. Springer, New York, NY.

Carré, 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.

Carré, F., McBratney, A. B., & Minasny, B. (2007). Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. Geoderma, 141:1-14.

Chaney, N., E. F. Wood, J. Hempel, A. B. McBratney, T. W. Nauman, C. Brungard, &

Clarke, K. C. (1880). Geodesy. John Wiley & Sons, Ltd.

Communications Authority of Kenya. (2017). Annual Report for the Financial Year 2016-2017, pp.91. Retrieved from https://ca.go.ke/wp-content/uploads/2018/04/Annual-Report-for-the-Financial-Year-2016-2017.pdf [accessed on 6/5/2019].

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Boehner, J. (2015): System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev., 8, 1991-2007, doi:10.5194/gmd-8-1991-2015. URL http://www.geosci-model-dev.net/8/1991/2015/gmd-8-1991-2015.html [accessed on 2/11/2019].

Cook, S. E., Jarvis, A., & Gonzalez, J. P. (2008). A new global demand for digital soil information, pp.31-41. In: Hartemink, A. E., McBratney, A., Mendonça-Santos, M. L. (Eds.) Digital Soil Mapping with Limited Data. Springer, Verlag, the Netherlands.

Council for Agricultural Science and Technology (CAST). (2019). Enabling Open-source Data Networks in Public Agricultural Research. CAST Commentary QTA2019-1. CAST, Ames, Iowa.

Crawley, M. J. (2012). The R book. John Wiley & Sons, pp.1051.

Daroussin, J., King, D., Le Bas, C., Vrščaj, B., Dobos, E., & Montanarella, L. (2006). The Soil Geographical Database of Eurasia at Scale 1: 1,000,000: history and perspective in digital soil mapping. Developments in Soil Science, 31:55-602.

Davis, R.E., Foote, F. S., Anderson, J.M., & Mikhail, E.M. (1981). Surveying, Theory and Practice, 6th Edition. McGraw-Hill, New York, pp. 992.

De Groote, H., Vanlauwe, B., Rutto, E., Odhiambo, G. D., Kanampiu, F., & Khan, Z. R. (2010). Economic analysis of different options in integrated pest and soil fertility management in maize systems of Western Kenya. Agricultural Economics, 41:471-482.

Dent, D. L., & Ahmed, F. B. (1995). Resurrection of soil surveys: a case study of the acid sulphate soils of The Gambia. I. Data validation, taxonomic and mapping units. Soil Use and Management, 11:69-76.

Dijkshoorn, J. A. (2007). Soil and terrain database for Kenya (ver. 2.0) (KENSOTER). ISRIC – World Soil Information, Wageningen.

Dobos, E. (2006). Digital Soil Mapping as a support to production of functional maps — EUR 22123 EN, Office for Official Publication of the European Communities, Luxemburg, pp. 68.

Dobos, E., Bialkó, T., Micheli, E., & Kobza, J. (2010). Legacy soil data harmonization and database development. Progress in Soil Science, 2:309-323.

Dokuchaev, V. V. (1883). The Russian chernozem report to the free economic society. Imperial University of St. Petersburg: St. Petersburg) [in Russian].

Dusart, J. (2005). Adapting soil mapping practices to the proposed INSPIRE guidelines. DSM 2004 Montpellier 13-17 September 2004. Elsevier.

EAMD. (1970). Temperature for stations in East Africa; Part 2, Kenya. EAAFRA, Nairobi.

EAMD. (1972). Summary of rainfall in Kenya. EAMD, Nairobi.

Edmunds, R., Hervé, L., Rickards, L., Trilsbeek, P., & Vardigan, M. (2016). Core trustworthy data repositories requirements. Zenodo. Retrieved from https://zenodo.org/record/168411 [accessed on 9/5/2019].

Environmental Systems Research Institute (ESRI). (2018). ArcGIS. Version 10.6.

ESRI, (2010). ESRI Web Mercator: The Standard for Sharing Data on the Web Redlands, CA. Retrieved from https://www.esri.com/events/seminars/bettermaps/~/media/files/pdfs/events/seminars/bettermaps/materials/pdfs/webmercatorsmnrbrochure.pdf [accessed on 9/3/2019].

Eze, P. N., Udeigwe, T. K., & Meadows, M. E. (2014). Plinthite and its associated evolutionary forms in soils and landscapes: a review. Pedosphere, 24:153-166.

FAO. (1976). A Framework for Land Evaluation. FAO, Rome: Soils Bulletin 32. FAO, Rome, pp. 72.

FAO. (1977). Guidelines for Soil Profile Description. 2nd edition. FAO, Rome.

FAO. (1988). FAO/UNESCO Soil Map of the World, Revised Legend. World Soil Resources Report, 60, pp.110. Retrieved from https://www.isric.org/sites/default/files/isric_report_1988_01.pdf [accessed on 9/11/2019].

FAO. (2017). FAOSTAT database collections. Food and Agriculture Organization of the United Nations. FAO, Rome, Italy. Retrieved from http://www.fao.org/faostat/en/#home [accessed on 4/11/2017].

FAO/ UNESCO. (1974). Soil Map of the world 1:5M., vol. 1, The Legend. Food and Agriculture Organization (FAO), United Nations Educational Scientific and Cultural Organization (UNESCO), Paris, pp. 65.

FAO/UNFPA/IIASA. (1982). Potential population supporting capacities of lands in the developing world. Project FPA/INT/513, FAO, Rome.

FAO-UNESCO. (1974). Soil Map of the World, 1:5,000,000. Vol. 1: Legend. UNESCO, Paris, France.

Fawcett, J. K. (1954). The semi-micro kjeldahl method for the determination of nitrogen. The Journal of medical laboratory technology, 12(1), 1-22.

Fertilizer Use Recommendation Project (FURP). (1988). Description of first priority trial sites in the various districts, Ministry of Agriculture/NAL/FURP, Final Report-Annex III: Vol. 4, Siaya district to Vol. 31, Kwale district. National Agricultural Laboratories, Nairobi, Kenya, pp. 140.

Feuerherdt, C., & Robinson, N. (2006). Storage, maintenance and extraction of digital soil data. In: P. Lagacherie, A.B. McBratney and M. Voltz (Eds.), Digital Soil Mapping, an introductory perspective. Developments in Soil Science, 31:87-96.

Forbes, T., Rossiter, D., & van Wambeke, A. (1987). Guidelines for evaluating the adequacy of soil resource inventories. (SMSS: Technical Monograph: Soil management support services; Vol. 4). Ithaca; Washington, D.C.: Cornell University, Department of Agronomy, New York State College of Agriculture and Life Sciences; U.S. Department of Agriculture, Soil Conservation Service, Soil Management Support Service (SMSS), pp.51.

Geitner, C., Baruck, J., Freppaz, M., Godone, D., Grashey-Jansen, S., Gruber, F. E., Heinrich, K., Papritz, A., Simon, A., Stanchi, S., Traidl, R., von Albertini, N. & Vrščaj, B. (2017). Soil and Land Use in the Alps—Challenges and Examples of Soil-Survey and Soil-Data Use to Support Sustainable Development, pp.221-292. In: Pereira P, Brevik E, Muñoz -Rojas M and Miller B (eds), Soil Mapping and Process Modeling for Sustainable Land Use Management. Elsevier, Amsterdam.

Gerland, P., Raftery, A. E., Ševčíková, H., Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B.K., Chunn, J., Lalic, N. & Bay, G. (2014). World population stabilization unlikely this century. Science 346:234-237.

Gethin-Jones, G. H., & Scott, R. M. (1959). Soil map of Kenya. Atlas of Kenya, 1st ed. Nairobi.

Gobin, A. (2000). Participatory and spatial-modelling methods for land resources analysis. Doctoral dissertation, Katholieke Universiteit Leuven, pp. 344.

Goodchild, M. F., & Hunter, G. J. (1997). A simple positional accuracy measure for linear features. International journal of geographical information science, 11:299-306.

Grunwald, S., Thompson, J. A., & Boettinger, J. L. (2011). Digital soil mapping and modeling at continental scales: Finding solutions for global issues. Soil Science Society of America Journal, 75:1201-1213.

Hallett, S. H., Baillie, I. C., Kerr, B., & Truckell, I. G. (2011). Development of the World Soil Survey Archive and Catalogue (WOSSAC). Commission on the History, Philosophy and Sociology of Soil Science, 18:14-17.

Hallett, S. H., Bullock, P., & Baillie, I. (2006). Towards a world soil survey archive and catalogue. Soil use and management, 22:227-228.

Häring, T., Dietz, E., Osenstetter, S., Koschitzki, T., & Schröder, B. (2012). Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. Geoderma, 185:37-47.

Harrell Jr, F. E., & Harrell Jr, M. F. E. (2018). Package 'Hmisc'. R Foundation for Statistical Computing.

Hartemink, A. E., Hempel, J., Lagacherie, P., McBratney, A., McKenzie, N., MacMillan, R. A., Minasny, B., Montanarella, L., de Mendonça Santos, M. L., Sanchez, P., & Walsh, M. (2010). GlobalSoilMap.net: A new digital soil map of the world, pp.423-428. In: J.L. Boettinger et al., editors, Digital soil mapping: Bridging research, environmental application, and operation. Springer, Dordrecht, the Netherlands.

Hengl, T. (2009). A practical guide to geostatistical mapping of environmental variables, Luxembourg: Office for Official Publication of the European Communities., pp. 165.

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Guevara, M. A. (2017). SoilGrids250m: Global gridded soil information based on machine learning. PLOS ONE, 12: e0169748.

Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J. G. B., Walsh, M. G., & Gonzalez, M. R. (2014). SoilGrids1km—global soil information based on automated mapping. PLOS ONE, 9: e105992.

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

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

Hengl, T., Heuvelink, G. B., Kempen, B., Leenaars, J. G., Walsh, M. G., Shepherd, K. D., Sila, A., MacMillan, R. A., de Jesus, J. M., Tamene, L., & Tondoh, J. E. (2015). Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PloS one, 10: e0125814.

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., & Webster, R. (2001). Modelling soil variation: past, present, and future. Geoderma, 100:269-301.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology: A Journal of the Royal Meteorological Society, 25:1965-1978.

Hilgard, E. W. (1906). Soils; their formation, properties, composition, and relations to climate and plant growth in the humid and arid regions. 593 pp., illus. New York and London.

Hinga, G., Muchena, F. N., & Njihia, C. M. (1980). Physical and chemical methods of soil analysis. National Agricultural Research Laboratories, Nairobi, Kenya.

Hudson, B. D. (1992). The soil survey as paradigm-based science. Soil Science Society of America Journal, 56:836-841.

ISRIC, I. (2012). JRC: Harmonized World Soil Database (version 1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria. Retrieved from http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/ [accessed on 9/5/2019].

ISRIC. (1986). World soil and terrain digital database at a scale of 1:1,000,000 (SOTER). Project Proposal. Wageningen.

ISRIC-WISE. (2008). References and examples of WISE database applications. PDF ISRIC 2007. ISRIC-world soil database.

International Council for Science. (2004). ICSU Report of the CSPR Assessment Panel on Scientific Data and Information. Retrieved from http://lasp.colorado.edu/media/projects/egy/files/PAA_Data_and_Information.pdf [accessed on 11/19/2019].

Jaetzold, R., & Schmidt, H. (1982). Farm Management Handbook of Kenya, vol. II. Natural Conditions and Farm Management Information. Part A. West Kenya. Kenya Ministry of Agriculture, Nairobi, Kenya, pp. 728.

Jenny, H. (1941). Factors of Soil Formation: A System of Quantitative Pedology. McGraw-Hill, New York, pp.199.

Jenny, H. (1961). Derivation of state factor equations of soils and ecosystems. Soil Sci. Soc. Am. Proc. 25:385–388.

Jenny, H. (1980). Soil genesis with ecological perspectives. Springer-Verlag, New York, New York, USA.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. Psychometrika, 23:187-200.

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

Kenya National Bureau of Statistics. (2010). Kenya National Bureau of Statistics (KNBS) Kenya 2009 Population and Housing Census Highlights. KNBS Government of Kenya, Nairobi, pp. 93. Retrieved from https://international.ipums.org/international/resources/enum_materials_pdf/enum_instruct_ke2009a.pdf [accessed on 9/5/2019].

Kenya National Bureau of Statistics. (2010). Kenya National Bureau of Statistics (KNBS) Kenya 2009 Population and Housing Census Highlights. KNBS Government of Kenya, Nairobi.

Kenya Soil Survey Staff. (1987). Manual for Soil Survey and Land Evaluation. Volume 1. Miscellaneous paper No. M24, Vol. 1. Ministry of Agriculture, Nairobi, Kenya.

Kenya Soil Survey. (1980). List of publications of the Kenya Soil Survey from 1972 Onwards. Ministry of Agriculture (Nairobi). Kenya Soil Survey. Ministry of Agriculture, pp. 27.

Kenya Soil Survey. (1984). List of Publications of the Kenya Soil Survey from 1972 Onwards. Ministry of Agriculture and Livestock Development, National Agricultural Laboratories, Kenya Soil Survey, Nairobi, pp. 37.

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

Kilimo Salama. (2011). Kilimo Salama: since we cannot control the weather. Online Press Room. Retrieved from https://kilimosalama.wordpress.com [accessed on 7/11/2019].

King, C. (1957). The uniformitarian nature of hillslopes. Transactions of the Edinburgh Geological Society, 17:81-102.

Krol, B. G. C. M. (2008). Towards a data quality management framework for digital soil mapping with limited data, pp.137-149. In: A.E. Hartemink, A.B. McBratney, M.L. Mendonca-Santos (Eds.), Digital soil mapping with limited data. Springer, Dordrecht.

Lagacherie, P. (2008). Digital soil mapping: a state of the art, pp. 3-14. In: Hartemink, A. E., McBratney, A. B., Mendonca-Santos, M. L. (Eds.) Digital soil mapping with limited data. Springer, Dordrecht.

Lagacherie, P., & McBratney, A. B. (2006). Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. Developments in Soil Science, 31:3-22.

Lagacherie, P., Legros, J. P., & Burfough, P. A. (1995). A soil survey procedure using the knowledge of soil pattern established on a previously mapped reference area. Geoderma, 65:283-301.

Laker, M. C. (1977). Selection of a method for determination of map intensities of soil maps. Proceedings of a workshop in Cornell University. 1977-1978, SMSS Tech. Mon. No. 1:339-349.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33:159-174.

Leenaars, J. G. B. (2012). Africa Soil Profiles Database, version 1.0: a compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa (with dataset). Wageningen, the Netherlands: Africa Soil Information Service (AfSIS) project and ISRIC — World Soil Information, 45 pp. Available: http://www.isric.org. ISRIC report 2012/03.

Leenaars, J. G. B. (2013). Africa Soil Profiles Database, Version 1.1. A compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa (with dataset). Africa Soil Information Service (AfSIS) project (No. 2013/03). ISRIC-World Soil Information.

Leenaars, J. G. B., Kempen, B., van Oostrum, A. J. M., & Batjes, N. H. (2014a). Africa Soil Profiles Database: A compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa, pp.51-57. In: Arrouays, D., McKenzie, N., Hempel, J., de Forge, A. R., McBratney, A. (Eds.). GlobalSoilMap — Basis of the Global Spatial Information System, Taylor & Francis Group, Orléans.

Leenaars, J. G. B., Kwabena, N. A., Silatsa, F. B. T., Fening, J. O., Yemefack, M., Hengl, T., & Heuvelink, G. B. M. (2017). World soil information developing from global, continental and national initiatives. Paper presented at GlobalSoilMap 2017 conference, Moscow, Russian Federation.

Leenaars, J. G. B., van Oostrum, A. J. M., & Ruiperez Gonzalez, M. (2014). Africa soil profiles database version 1.2. A compilation of georeferenced and standardized legacy soil profile data for Sub-Saharan Africa (with dataset). Wageningen: ISRIC Report 2014/01; 2014. ISRIC-World Soil Information, Wageningen.

Leenaars, J. G., Van Oostrum, A. J. M., & Ruiperez Gonzalez, M. (2014b). Africa soil profiles database version 1.2. A compilation of georeferenced and standardized legacy soil profile data for Sub-Saharan Africa (with dataset). Wageningen: ISRIC Report 2014/01; 2014. ISRIC-World Soil Information, Wageningen.

Leutner, B., & Horning, N. (2017). RStoolbox: tools for remote sensing data analysis. R package version 0.1,7.

Mallavan, B. P., Minasny, B., & McBratney, A. B. (2010). Homosoil, a methodology for quantitative extrapolation of soil information across the globe. pp. 137-149. In: J. L. Boettinger, D. Howell, A. C. Moore, A. E. Hartemink, S. Kienast-Brown (Eds.), Digital Soil Mapping. Springer, Dordrecht.

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

Malone, B. P., Minasny, B., & McBratney, A. B. (2017). Using R for digital soil mapping. Basel, Switzerland: Springer International Publishing, p.271.

Mas, I., & Radcliffe, D. (2010). Mobile payments go viral: M-PESA in Kenya. pp.353-369. In: Punam Chuhan-Pole, Manka Angwafo (Eds.), Yes Africa Can: Success Stories from a Dynamic Continent, World Bank, Washington, D.C.

Mayr, T. R., Palmer, R., Lawley, R., & Fletcher, P. (2001). New methods of soil mapping. Final Report: SR0120. National Soil Resources Institute, Silsoe, UK.

Mayr, T., & Palmer, B. (2006). Digital soil mapping: an England and Wales perspectives. Developments in soil science, 31:365-618.

Mayr, T., Rivas-Casado, M., Bellamy, P., Palmer, R., Zawadzka, J., & Corstanje, R. (2010). Two methods for using legacy data in digital soil mapping, pp.191-202. In Digital Soil Mapping. Springer, Dordrecht.

McBratney, A. B. (1998). Some considerations on methods for spatially aggregating and disaggregating soil information. Nutrient Cycling in Agroecosystems, 50:51-62.

McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. Geoderma, 117:3-52.

McBratney, A., Field, D. J., & Koch, A. (2014). The dimensions of soil security. Geoderma, 213:203-213.

McKenzie, N. J., Grundy, M. J., Webster, R., & Ringrose-Voase, A. J. (Eds.) (2008). Guidelines for surveying soil and land resources (2nd ed.). Collingwood, Australia: CSIRO Publishing.

Microsoft. (2010). Microsoft Excel [computer software].

Milne, G. (1936). A provisional soil map of East Africa (Kenya, Uganda, Tanganyika, and Zanzibar) with explanatory memoir. East Africa Agricultural Research Station, Amani, Tanganyika Territory, pp. 317.

Milne, G. G. (1935). Some suggested units of classification and mapping particularly for East African soils. Soil Research, 4:183-198.

Minai, J. O., & Schulze, D. G. (2019). Busia Digital Soil Mapping. Purdue University Research Repository. doi:10.4231/E81F-NX21.

Minai, J. O., Ngunjiri, M. N., Schulze, D. G., Owens, P. R., Serrem., C., & Nyabinda, N. (2016). Delivery of Spatial Explicit Information in Western Kenya. Abstracts, Soil Science Society of America International Annual Meeting, November 6-9, 2016, Phoenix, AZ. Retrieved from https://scisoc.confex.com/crops/2016am/webprogram/Paper100855.html [Accessed on 8/23/2019].

Minai, J. O., Schulze, D. G., & Ngunjiri, M. W. (2018). Renewal and Delivery of Spatially Explicit Soils Information in Western Kenya. Poster presentation. Scale-Up Conference. Purdue University, West Lafayette, IN. Sept. 25-27, 2018. Retrieved from https://ag.purdue.edu/scaleup/Documents/Poster%20Abstracts.pdf [accessed on 8/23/2019].

Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. Computers & geosciences, 32:1378-1388.

Minasny, B., & McBratney, A. B. (2007). Spatial prediction of soil properties using EBLUP with the Matérn covariance function. Geoderma, 140:324-336.

Minasny, B., & McBratney, A. B. (2010a). Conditioned Latin hypercube sampling for calibrating soil sensor data to soil properties, pp. 111-119. In: Proximal soil sensing. Springer, Dordrecht.

Minasny, B., & McBratney, A. B. (2010b). Methodologies for global soil mapping, pp.429–436. In: Boettinger JL, Howell DW, Moore AC, Hartemink AE, Kienast-Brown S. (Eds.), Digital soil mapping: Bridging Research, Environmental Application, and Operation, Springer, volume 2 of Progress in Soil Science.

Minasny, B., McBratney, A. B., & Lark, R. M. (2008). Digital soil mapping technologies for countries with sparse data infrastructures, pp.15-30. In: Hartemink, A. E., McBratney, A. B., Mendonca-Santos M. L. (Eds.), Digital soil mapping with limited data. Springer, Dordrecht, the Netherlands.

Minasny, B., McBratney, A. B., & Whelan, B. M. (2005). VESPER version 1.62. Australian Centre for Precision Agriculture, McMillan Building A05, the University of Sydney, NSW 2006. URL [https://sydney.edu.au/agriculture/pal/software/vesper.shtml accessed on 2/13/2019].

Montanarella, L., Pennock, D. J., McKenzie, N., Badraoui, M., Chude, V., Baptista, I., Mamo, T., Yamefack, M., Aulakh, M. S., Yagi, K., Hong, S. Y., Vijarnsorn, P., Zhang, G., Arrouays, D., Black, H., Krasilnikov, P., Sobocka, J., Alegre, J., Herinque, C. R., Mendoca-Santos, M. L., Taboada, M., Espinosa-Victoria, D., Alshankiti, A., AlaviPanah,

S. K., Elsheikh, E. A. E. M., Hempel, J., Arbestin, M. C., Nachtergaele, F., & Vargas, R. (2016). World's soils are under threat. Soil, 2:79-82.

Moore, T. R. (1979). Rainfall erosivity in east Africa. Geografiska Annaler: Series A, Physical Geography, 61(3-4), 147-156.

Muchena, F. N., van der Pouw, B. J. A., & Siderius, W. (1982). An outline of the major soils in Kenya (with a generalized soil map at scale 1:3,000,000). Republic of Kenya, Ministry of Agriculture, National Agricultural Laboratories, Kenya Soil Survey.

N. P. Odgers (2016), POLARIS: A 30-meter Probabilistic Soil Series Map of the Contiguous United States, Geoderma, 274:54–67.

Nauman, T. W., Thompson, J. A., Odgers, N. P., & Libohova, Z. (2012). Fuzzy disaggregation of conventional soil maps using database knowledge extraction to produce soil property maps, pp.203-207. In: B. Minasny, B. Malone, A.B. McBratney (Eds.), Digital Soil Assessments and Beyond: Proceedings of the Fifth Global Workshop on Digital Soil Mapping, Sydney, Australia.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). Applied linear statistical models. Richard D. Irwin, Inc., Homewood, IL, p.318

Odeh, I. O., Leenaars, J., Hartemink, A., & Amapu, I. (2012). The challenges of collating legacy data for digital mapping of Nigerian soils, pp.453-458. In: B. Minansy, B.P. Malone, A. B. McBratney (Eds.), Proceedings of the Fifth Global Workshop on Digital Soil Mapping, Sydney, Australia, 10–13 April 2012, CRC Press/Balkema, The Netherlands.

Odgers, N. P., Libohova, Z., & Thompson, J. A. (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, 189:153-163.

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

Oldeman, L. R., Hakkeling, R. U., & Sombroek, W. G. (1991). World Map on the Status of Human-Induced Soil Degradation. An Explanatory Note. Global Assessment of Soil Degradation. GLASOD. ISRIC - Winand Centre - ISSS - FAO - ITC, Wageningen. pp. 41.

Onywere, S. M., Getenga, Z. M., Mwakalila, S. S., Twesigye, C. K., & Nakiranda, J. K. (2011). Assessing the challenge of settlement in Budalangi and Yala swamp area in Western Kenya using landsat satellite imagery. Open Environmental Engineering, 4:97–104.

Panagos, P., Jones, A., Bosco, C., & Kumar, P. S. (2011). European digital archive on soil maps (EuDASM): preserving important soil data for public free access. International Journal of Digital Earth, 4:434-443.

Park, S. J., McSweeney, K., & Lowery, B. (2001). Identification of the spatial distribution of soils using a process-based terrain characterization. Geoderma, 103:249-272.

Penizek, V., & Boruvka, L. (2008). The Digital Terrain Model as a tool for improved delineation of alluvial soils, pp.319-326. In: A. Hartemink, A. McBratney and M. Mendoça-Santos (Eds.), Digital Soil Mapping with Limited Data. Wageningen, Sydney, Río de Janeiro: Springer.

Ponce-Hernandez, R., Marriott, F. H. C., & Beckett, P. H. T. (1986). An improved method for reconstructing a soil profile from analyses of a small number of samples. European Journal of Soil Science, 37:455-467.

Purnell, M. F. (1995). Soil survey information supply and demand: international policies and stimulation programs. World Soil Resources Report 80, FAO, Rome, pp. 3--35.

Rachilo, J.R., & Michieka, D.O. (1991). Reconnaissance Soil Survey of the Busia Area (quarter degree sheet No. 101). Kenya Soil Survey, pp. 269.

Reid-Soukup, D. A., & Ulery, A. L. (2002). Smectites, pp.467-499. In: J.B Dixon, D.G. Schulze (Eds.) Soil mineralogy with environmental applications, SSSA, Madison, WI.

Ribeiro, E., Batjes, N. H., & van Oostrum, A. J. M. (2018). World Soil Information Service (WoSIS)-Towards the standardization and harmonization of world soil data. Procedures manual 2018, Report 2018/01, ISRIC – World Soil Information, Wageningen, pp.166.

Ribeiro, E., Batjes, N. H., Leenaars, J. G., van Oostrum, A., & de Jesus, J. M. (2015). Towards the standardization and harmonization of world soil data: Procedures Manual ISRIC World Soil Information Service (WoSIS version 2.0). ISRIC World Soil Information, Wageningen, pp. 110.

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package ‘MASS’. CRAN Repository.

Rossiter, D. G. (2008). Digital soil mapping as a component of data renewal for areas with sparse soil data infrastructures, pp.69-80. In: A. Hartemink, A. McBratney and M. Mendoça-Santos (Eds.), Digital Soil Mapping with Limited Data. Springer, the Netherlands.

Roudier, P. (2017). Package “clhs.” R Package, pp. 1–7. doi:10.1201/b12728-46

RStudio Team. (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA [http://www.rstudio.com/].

Rubin, J. (1967). Optimal classification into groups: an approach for solving the taxonomy problem. Journal of Theoretical Biology, 15:103-144.

Ruhe, R. V. (1975). Climatic geomorphology and fully developed slopes. Catena, 2:309-320.

Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., de Lourdes Mendonça-Santos, M. and Minasny, B. (2009). Digital soil map of the world. Science, 325:680-681.

Savtchenko, A., Ouzounov, D., Ahmad, S., Acker, J., Leptoukh, G., Koziana, J., & Nickless, D. (2004). Terra and Aqua MODIS products available from NASA GES DAAC. Advances in Space Research, 34:710-714.

Scharlemann, J. P., Tanner, E. V., Hiederer, R., & Kapos, V. (2014). Global soil carbon: understanding and managing the largest terrestrial carbon pool. Carbon Management, 5:81-91.

Schoeneberger, P.J., D.A. Wysocki, and E.C. Benham, and Soil Survey Staff. (2012). Field book for describing and sampling soils. Version 3.0. U.S. Department of Agriculture, Natural Resource Conservation Service, Lincoln, NE, pp. 300.

Schulze, D. G., Owens, P. R., Mitzman, S., Snyder, L. U., Van Scoyoc, G., Steinhardt, G. C., Miller, C. C., Bracke. M. S. Glotzbach, R. N. J., Kocur, L. A., Benes, B. Neely, T., Wiggington, M., and Nielson, R. (2010). Isee - Integrating spatial educational experiences into soil, crop, and environmental sciences. In 19th World Congress of Soil Science, Soils for a Changing World. Brisbane, Australia.

Scott, R. M. (1969). East African Soils. In: Morgan, M. T. W. (Ed.), 1969, East Africa, its peoples and resources. Oxford University Press, London-Nairobi (revised edition of: Russel, E. W., 1962, Natural Resources of East Africa. D. A. Hawkins, London.

Scott, R. M., Webster, R., & Lawrance, C. J. (1971). A land system atlas of Western Kenya. A land system atlas of Western Kenya. Christchurch, Hampshire, UK, pp. 382.

Selvaradjou, S. K., Montanarella, L., Spaargaren, O., Dent, D., Filippi, N., & Dominik, S. (2005b). European digital archive of soil maps (EuDASM)–Metadata of the Soil Maps of Asia. Office of the Official Publications of the European Communities, Luxembourg.

Selvaradjou, S. K., Montanarella, L., Spaargaren, O., Dent, D., Filippi, N. and Reuter, H.I. (2005c). European Digital Archive of Soil Maps (EuDASM) - Metadata of the Soil Maps of Latin America and Caribbean islands. EUR 21821 EN, Office for Official Publications of the European Communities, Luxembourg, pp. 219.

Selvaradjou, S. K., Montanarella, L., Spaargaren. O. and Dent. D. (2005a). European Digital Archive of Soil Maps (EuDASM) - Soil Maps of Africa. EUR 21657 EN, Office for Official Publications of the European Communities, Luxembourg, pp. 386.

Seyfried, M. S., and B. P. Wilcox. (1995). Scale and the nature of spatial variability: field examples having implications for hydrologic modeling. Water Resources Research, 31:173–184.

Shepherd, K., & Vågen, T. (2010). A Globally Integrated African Soil Information Service (AFSIS), http://africasoils.net/ [accessed 1/15/2019].

Soil Conservation Society of America. (1976). Resource conservation glossary, 2nd edition. Soil Conservation Society of America, Ankeny, Iowa, pp.63.

Soil Science Division Staff. (2017). Soil survey manual. C. Ditzler, K. Scheffe, and H.C. Monger (Eds.). USDA Handbook 18. Government Printing Office, Washington, D.C, pp. 315.

Sombroek, W. G., Braun, H. M. H., & Van der Pouw, B. J. A. (1982). Exploratory soil map and agro-climatic zone map of Kenya, 1980. Scale 1: 1,000,000. Exploratory Soil Survey Report No. E1. Kenya Soil Survey, Nairobi, Kenya.

Spaargaren, O. C., & Deckers, J. (1998). The world reference base for soil resources, pp.21-28. In Soils of Tropical Forest Ecosystems. Springer, Berlin, Heidelberg.

Ströh de Martínez, C., Feddersen, M., & Speicher, A. (2016). Food security in sub-Saharan Africa: A fresh look on agricultural mechanization; how adapted financial solutions can make a difference. Bonn: Deutsches Institut für Entwicklungspolitik (DIE), p.171. Retrieved from https://www.die-gdi.de/uploads/media/Study_91.pdf [accessed on 8/26/2019].

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

Survey of Kenya (1959). Atlas of Kenya: A Comprehensive Series of New and Authentic Maps Prepared from the National Survey and Other Governmental Sources with Gazetteer and Notes on Pronunciation & Spelling. 1st Ed. Government of Kenya.

Survey of Kenya (1962). Atlas of Kenya: A Comprehensive Series of New and Authentic Maps Prepared from the National Survey and Other Governmental Sources with Gazetteer and Notes on Pronunciation & Spelling. 2nd Ed. Government of Kenya.

Survey of Kenya (1970). Atlas of Kenya: A Comprehensive Series of New and Authentic Maps Prepared from the National Survey and Other Governmental Sources with Gazetteer and Notes on Pronunciation & Spelling. 3rd Ed. Government of Kenya.

Tempel, P., van Kraalingen, D. W. G., de Jesus, J. M., & Reuter, H. I. (2013). Towards an ISRIC World Soil Information Service (WoSIS ver. 1.0) ISRIC Report 2013/02, ISRIC-World Soil Information, Wageningen, pp. 188.

Theisen, A. A., & Bellis, E. (1964). Quantitative analysis of clay mineral mixtures by X-ray diffraction. Nature, 204:1228-1230.

Theisen, A. A., & Harward, M. E. (1962). A Paste Method for Preparation of Slides for Clay Mineral Identification by X-ray Diffraction 1. Soil Science Society of America Journal, 26:90-91.

Thompson, J. A., Prescott, T., Moore, C. A., Bell, J., Kautz, D., Hempel, J., Waltman, W. S., & Perry, H., C. (2010). Regional approach to soil property mapping using legacy data and spatial disaggregation techniques, pp.1-4. In 19th World Congress of Soil Science. U.S. Department of Agriculture.

Thompson, J. A., Roecker, S., Grunwald, S., & Owens, P. R. (2012). Digital soil mapping: Interactions with and applications for hydropedology. Lin, H. S. (Ed.), Hydropedology: Synergistic Integration of Pedology and Hydrology, Academic Press, Waltham, USA, pp. 665-709.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic geography, 46:234-240.

Tucker, B. M. (1974). Laboratory procedures for cation exchange measurements on soils (No. 23 Tech. Pape).

Tully, K., Sullivan, C., Weil, R., & Sanchez, P. (2015). The state of soil degradation in Sub-Saharan Africa: Baselines, trajectories, and solutions. U.S. Department of Agriculture. Sustainability,7:6523-6552.

United Nations, Department of Economic and Social Affairs, Population Division. (2019). World Population Prospects 2019: Highlights (ST/ESA/SER.A/423), p.46. Retrieved from https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf [Accessed on 6/3/2019].

USGS. (2017a). NASA Shuttle Radar Topography Mission (SRTM) Global 1 arc second dataset (SRTMGL1), Digital Elevation Version 3. Retrieved from https://earthexplorer.usgs.gov/. [accessed on 4/11/2017].

USGS. (2017b). Landsat 8 OLI aerial imagery Level 1. URL: https://earthexplorer.usgs.gov/.

Vågen, T. G., Shepherd, K. D., Walsh, M. G., Winowiecki, L., Desta, L. T., & Tondoh, J. E. (2010). AfSIS Technical Specifications. Soil Health Surveillance. Version, 1.

Vågen, T. G., Winowiecki, L. A., Tondoh, J. E., & Desta, L. T. (2013). Africa Soil Information Service (AfSIS)-Soil Health Mapping, https://doi.org/10.7910/DVN/2JUBRA, Harvard Dataverse, V2.

Van Engelen, V. W. P., & Dijkshoorn, J. A. (2012). Global and National Soils and Terrain Databases (SOTER). Procedures Manual, version 2.0, draft for comments. ISRIC Report 2012/04, ISRIC–World Soil Information, Wageningen. ISRIC Report, 4, pp. 196. Retrieved from https://www.isric.org/sites/default/files/isric_report_2012_04.pdf [accessed on 9/5/2019].

van Wambeke, A. (1992). Soils of the tropics: properties and appraisal. McGraw Hill, New York, NY.

Vašát, R., Heuvelink, G. B. M., & Borůvka, L. (2010). Sampling design optimization for multivariate soil mapping. Geoderma, 155:147-153.

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

Wei, S., McBratney, A., Hempel, J., Minasny, B., Malone, B., D'Avello, T., Burras, L., & Thompson, J., (2010). Digital harmonization of adjacent analogue soil survey areas–4 Iowa counties, pp.1-6. In: Proceedings World Congress of Soil Science: Soil solutions for a changing world, 19th, Brisbane, Australia.

West, D. M. (2012). How mobile technology is driving global entrepreneurship. Brookings Policy Report.

Wood, A. (1942): The development of hillslopes. Geologists Association Proceedings 53:128-138.

Wysocki, D.A., P.J. Schoeneberger, and H.E. LaGarry. 2000. Geomorphology of soil landscapes, pp. E5–E40. In: M.E. Sumner, editor, Handbook of soil science. CRC Press, Boca Raton, FL.

Zhu, A. X., & Burt, J. E. (2011). Digital Soil Mapping Software: SoLIM Solutions 2013. Reference Manual. Institute of Geographic Sciences. Natural Resources Research-Chinese Academy of Sciences. The University of Wisconsin-Madison. Retreived from https://solim.geography.wisc.edu/software/index.htm [accessed on 2/11/2019].

Zhu, A. X., Band, L. E., Dutton, B., & Nimlos, T. J. (1996). Automated soil inference under fuzzy logic. Ecological Modelling, 90:123-145.

Zhu, A. X., Burt, J. E., Moore, A. C., Smith, M. P., Liu, J., & Qi, F. (2007). SoLIM: A New Technology for Soil Mapping Using GIS, Expert Knowledge & Fuzzy Logic. Overview of SoLIM Programme. University of Wisconsin.

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

Zhu, A. X., Moore, A. C., Smith, M. P., Liu, J., Burt, J. E., Qi, F., Simonson, D., Hempel, J. and Lubich, K. (2004). Advance in information technology for soil surveys: the SoLIM effort, pp.25-42. In: H. Eswaran, P. Vijarnsorn, T. Vearasilp, E. Padmanabhan (Eds.), Innovative Techniques in Soil Survey: Developing the Foundation for a New Generation of Soil Resource Inventories and Their Utilization. Land Development Department, Chattuchak, Bangkok, Thailand.

Zhu, A. X., Qi, F., Moore, A., & Burt, J. E. (2010). Prediction of soil properties using fuzzy membership values. Geoderma, 158:199-206.

Zhu, A. X., Yang, L., Li, B., Qin, C., English, E., Burt, J. E., & Zhou, C. (2008). Purposive sampling for digital soil mapping for areas with limited data, pp. 233-245. In: A. E.

Hartemink, A. B. McBratney, M.L. Mendonca Santos (Eds.), Digital soil mapping with limited data. Springer-Verlag, New York.

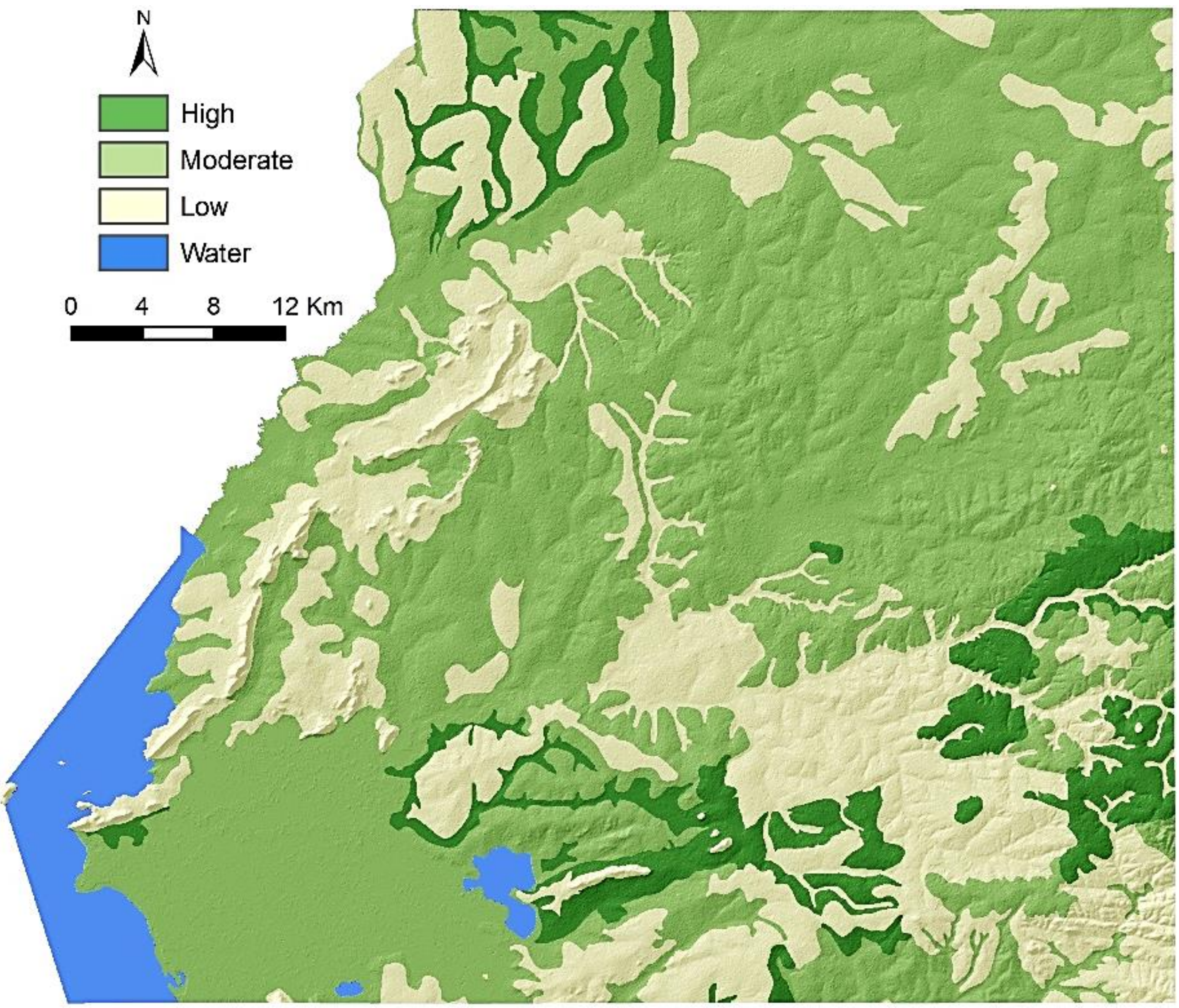
Zhu, A.X. (2006). Fuzzy logic models in soil science, pp. 215-239. In: Sabine Grunwald and Mary E. Collins (Eds.) Environmental soil-landscape modeling: geographic information technologies and pedometrics, CRC/Taylor & Francis, New York.

Zinck, J. A. (Ed.). (1995). Soil survey: perspectives and strategies for the 21st century: an international workshop for heads of national soil survey organizations, November 1992, Enschede: (No. 21). FAO/ITC, Rome, pp. 132.

APPENDIX A: BUSIA LAND QUALITY MAPS

A.1: Availability of Foothold for Roots (AoF).

This land quality is related to the depth of the soil and can be evaluated according to soil depth being overserved. The rating criteria used for this study was not included in the survey report. This land quality, however, reflects the ability of the soil to have deep well drained soils to hold up plants for growth. Areas with high ratings for this land quality indicating the areas with deep soils whereas those with shallow soils have a low rating.

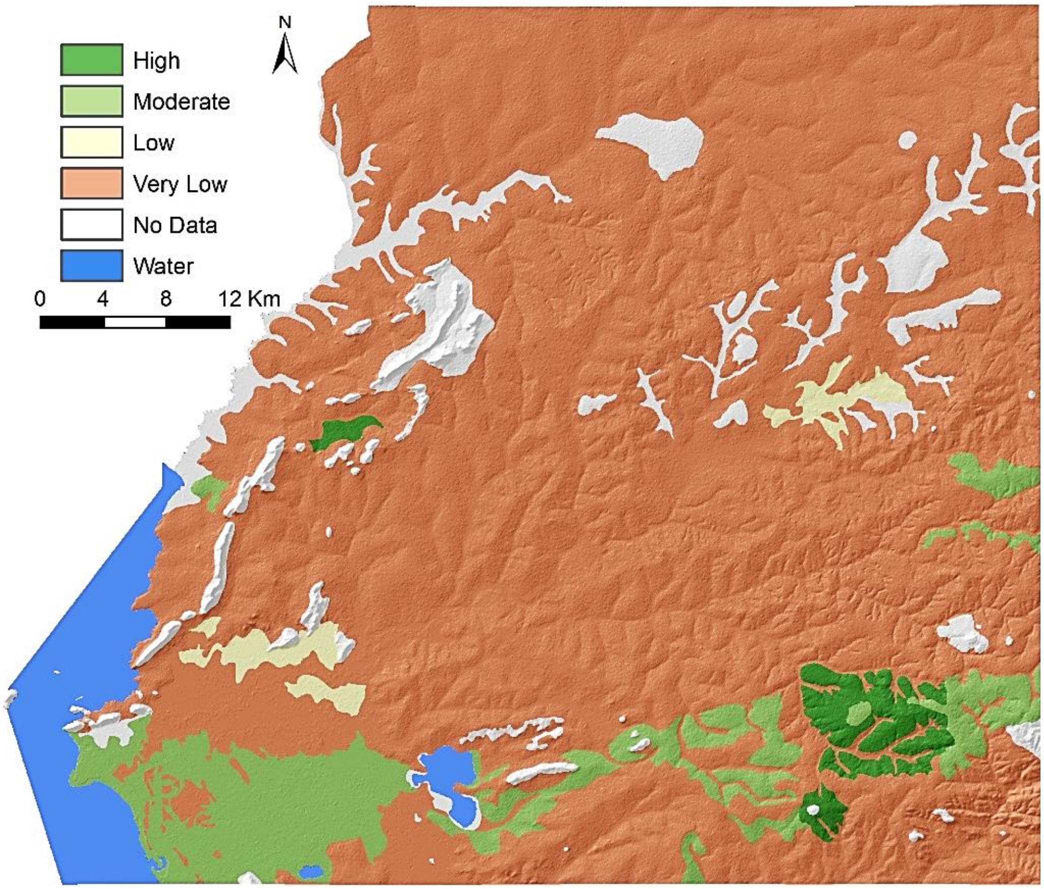


A.2: Availability of Nutrients for Plant Growth (AoN).

This land quality represents the ability of soils to supply the required plant nutrients for optimum growth. For the availability of nutrients, the topsoil was taken from 0 to 30 cm (Table A.2). Salinity and sodicity were not considered within the fertility rating system.

Table A.1: Rating land quality availability of nutrients for plant growth

Rating	**CEC me/100g**	**%SOC in temperature zone 1,2,3**	**Available P (ppm)**		**Exch. K**	**Exch. Ca**	**Exch. Mg**	**pH-H_2O (1:2.5)**
			Mehlich	**Olsen 1,2,3**				
1. High	>16	>2.0	>60	>20	>0.5	>6.0	>3.0	5.6-6.8
2. Moderate	6-16	1.2-2.0	21-60	11-20	0.21-0.5	3.0-6.0	1.1-3.0	4.8-5.5 or 6.9-7.5
3. Low	3-5.9	0.5-1.0	10-20	5-10	0.10-0.20	1.0-2.9	0.5-1.0	4.0- 4.7
4. Very low	<3	<0.5	<10	<5	<0.10	<1.0	<0.5	<4.0 or >8.7

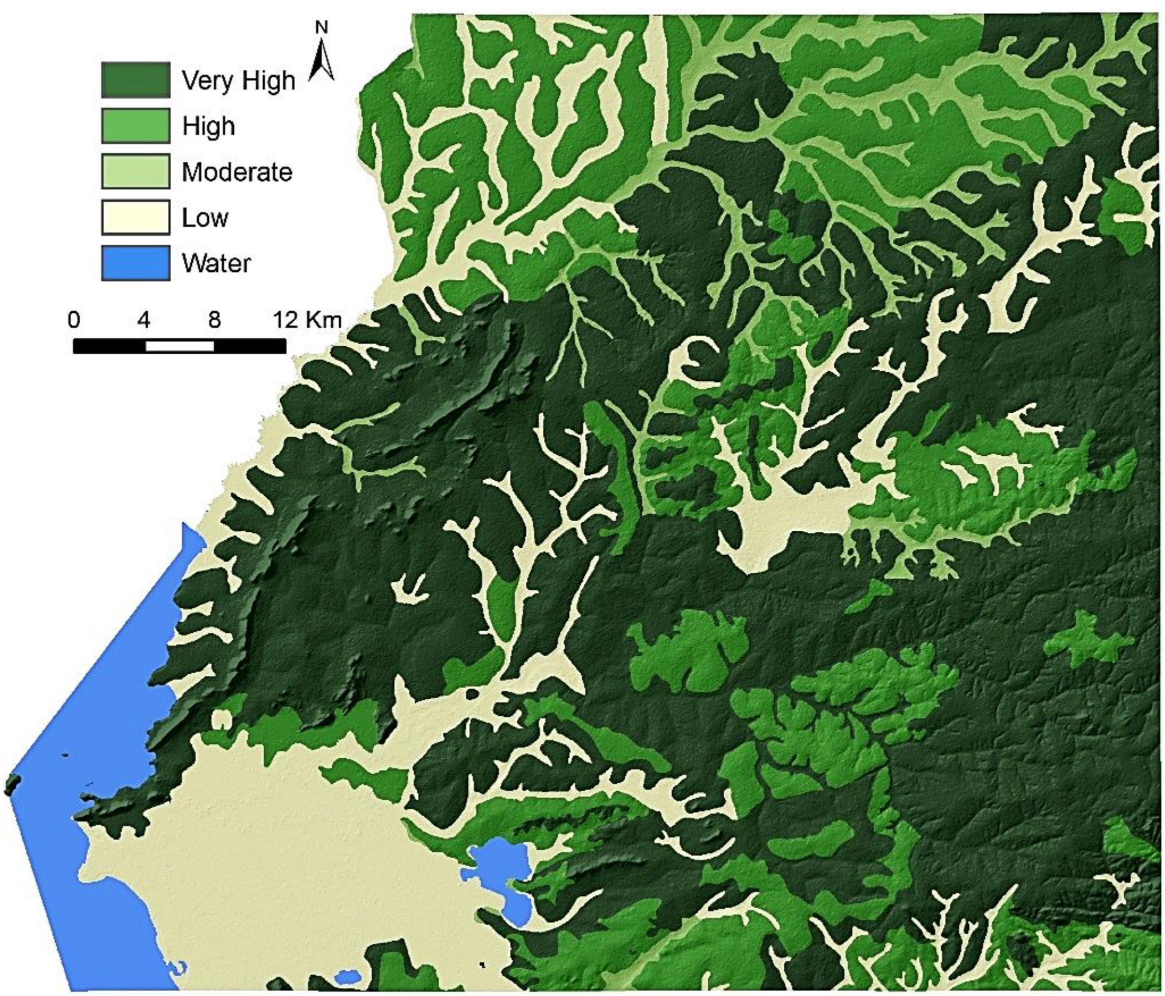


A.3: Availability of Oxygen in the Root Zone (Ox).

Availability of oxygen to the plant roots is described in terms of drainage conditions of the soil (Table A.3). Mottles and dark grey colors in the subsoil are used as indicators of poor drainage, and are usually qualitatively regarded as the basis for rating drainage conditions.

Table A.2: Rating land quality availability of oxygen

Rating	Soil draining class
1. Very High	Well to excessively drained
2. High	Moderately well drained
3. Moderate	Imperfectly drained
4. Low	Poorly drained
5. Very Low	Very poorly drained

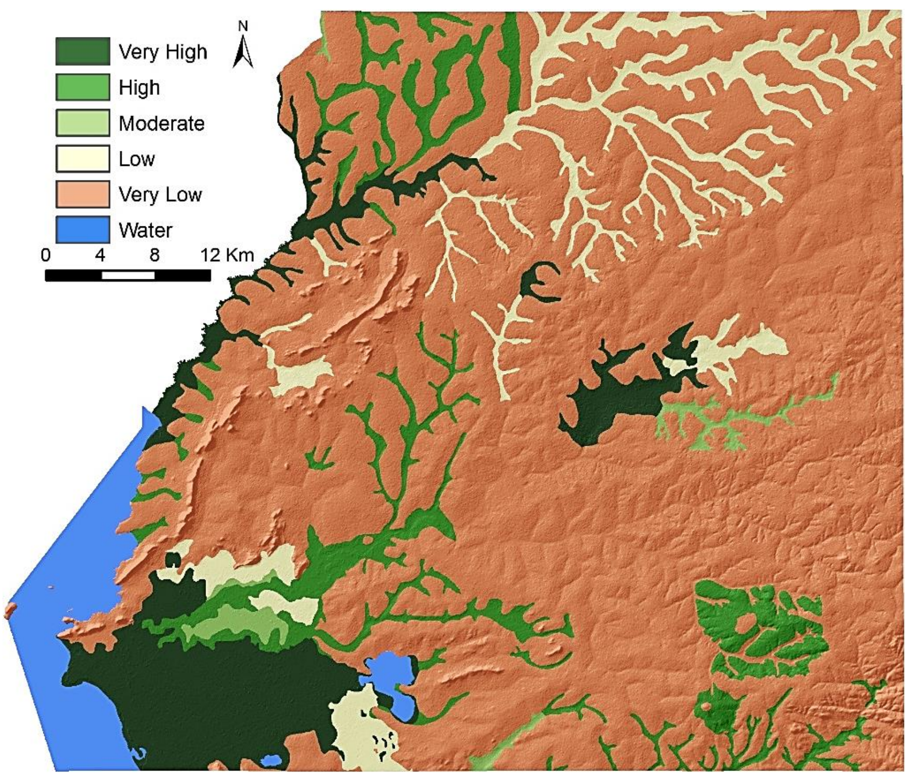


A.4: Flooding Hazard (FH).

Flooding hazard shows the damage resulted by floods through the mechanical forces of moving water. The frequency of the occurrence of the floods during the period in which the plants are on the field was an important criterion that was used. Depth of the flooding as well might have to be considered for the assessment of the risk. The effects of pounding resulting in phenomena similar to drainage problems are described and evaluated under the land quality availability of Oxygen.

Table A.3: Rating land quality, Flooding Hazard during the growing season.

Ratting	Flooding frequency	Inundation frequency/duration
1. Very low	every 10 years or more	None
2. Low	every 5 to 10 years	1-2 months, every 3-5 years
3. Moderate	every 3 to 5 years	2-3 months in 5 out of 10 years
4. High	every 1 to 3 years	2-4 months, almost every year
5. Very high	every year	more than 4 months every year

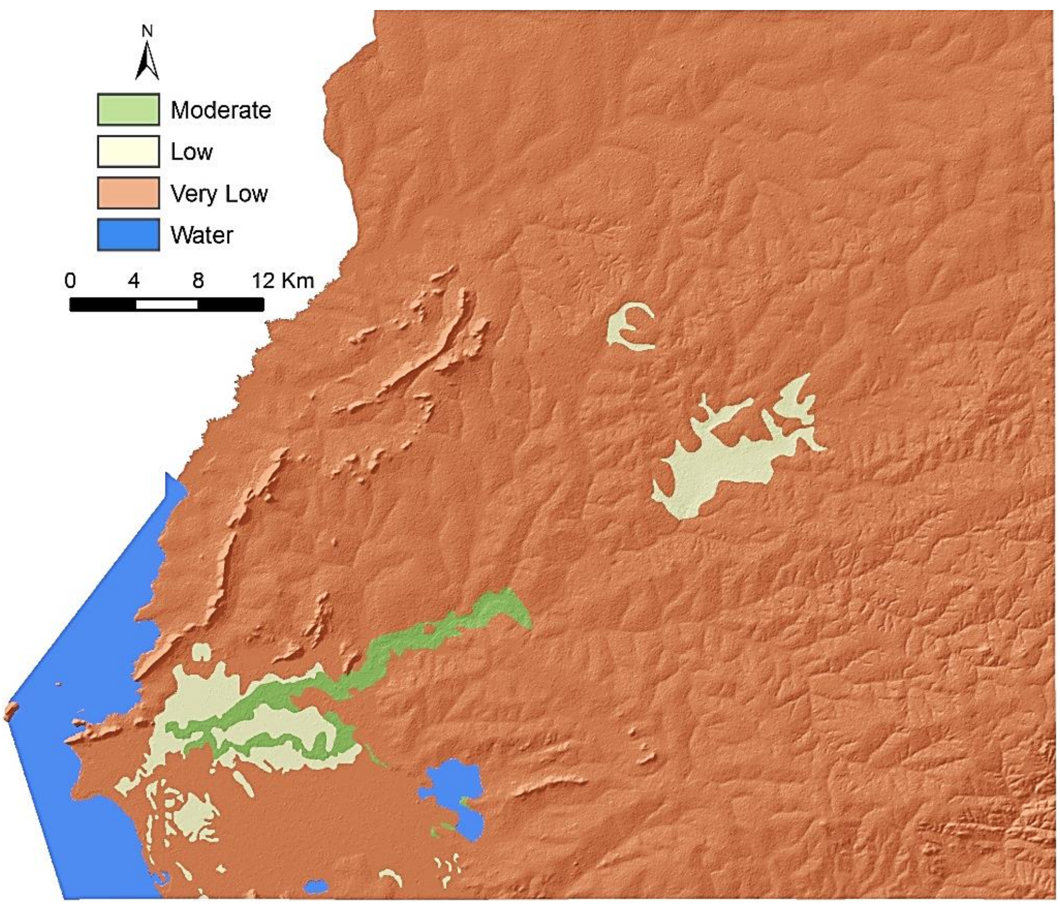


A.5: Hazard of Salinity (SH).

High levels of salt in the soil may lead to the dying of plant roots owing to the difficulty with which plant roots absorb water from the soil. Two different depth zones were used to for the interpretation. Within each depth zone, the highest value of saturation extract was classified as being the determinant critical characteristic for this land quality (Table A.5).

Table A.4: Rating land quality hazard of salinity

Ratting	0 – 30 cm	30 – 100 cm
1. Very low	< 2.0	< 4.0
2. Low	2.0 – 4.0	4.0 – 8.0
3. Moderate	4.1 - 8.0	8.1 – 15.0
4. High	8.1 – 15.0	15.1 – 30.0
5. Very high	>15.0	> 30.0

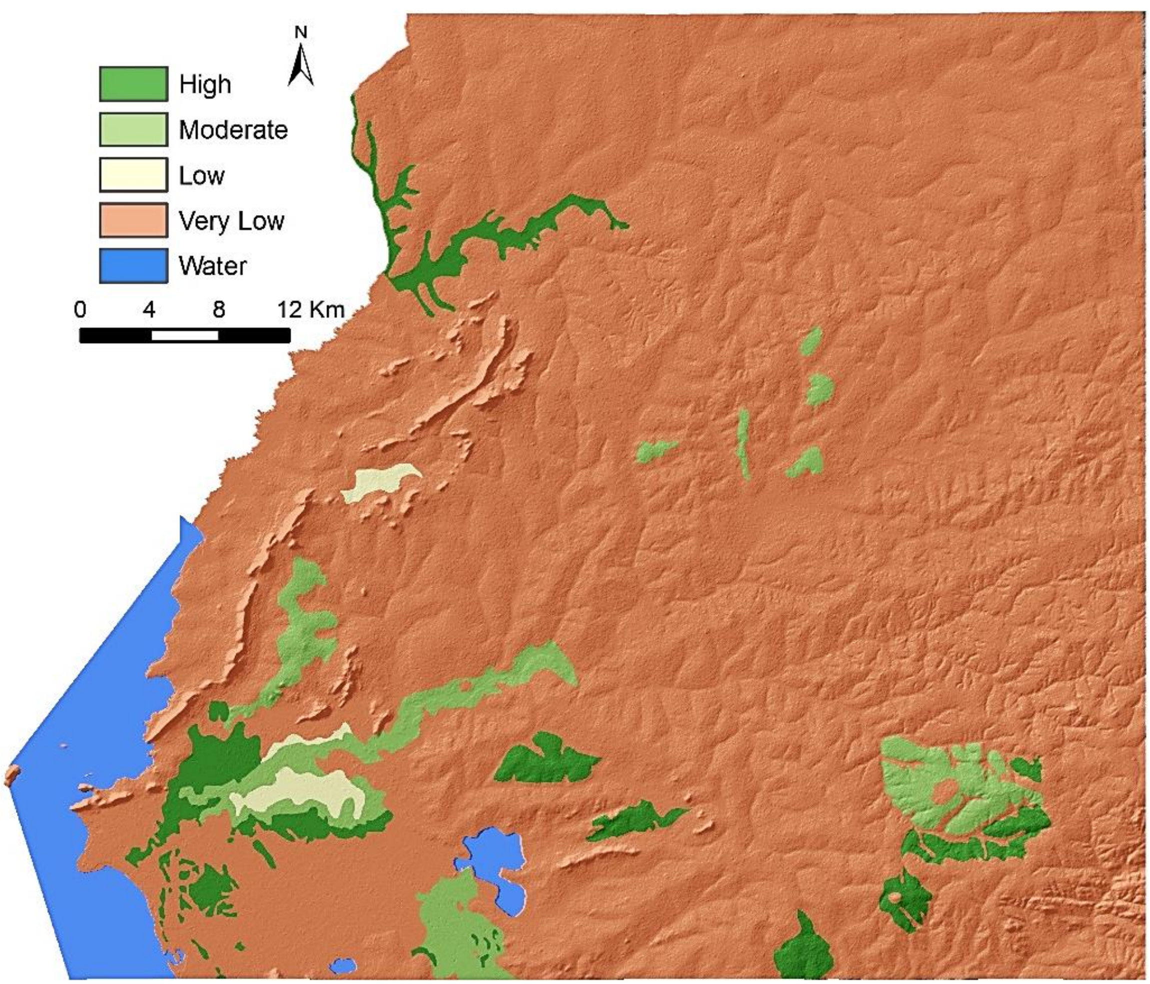


A.6: Hazard of Sodicity (Sod).

High levels of sodium in the soil will cause dispersion of structural aggregates which may result in poor aeration in the soil. This land quality was determined by the most limiting factor method, applied to the two different soil depth zones. Within each depth, the highest value of the ESP was classified, and the most limiting rating used as the final classification (Table A.6).

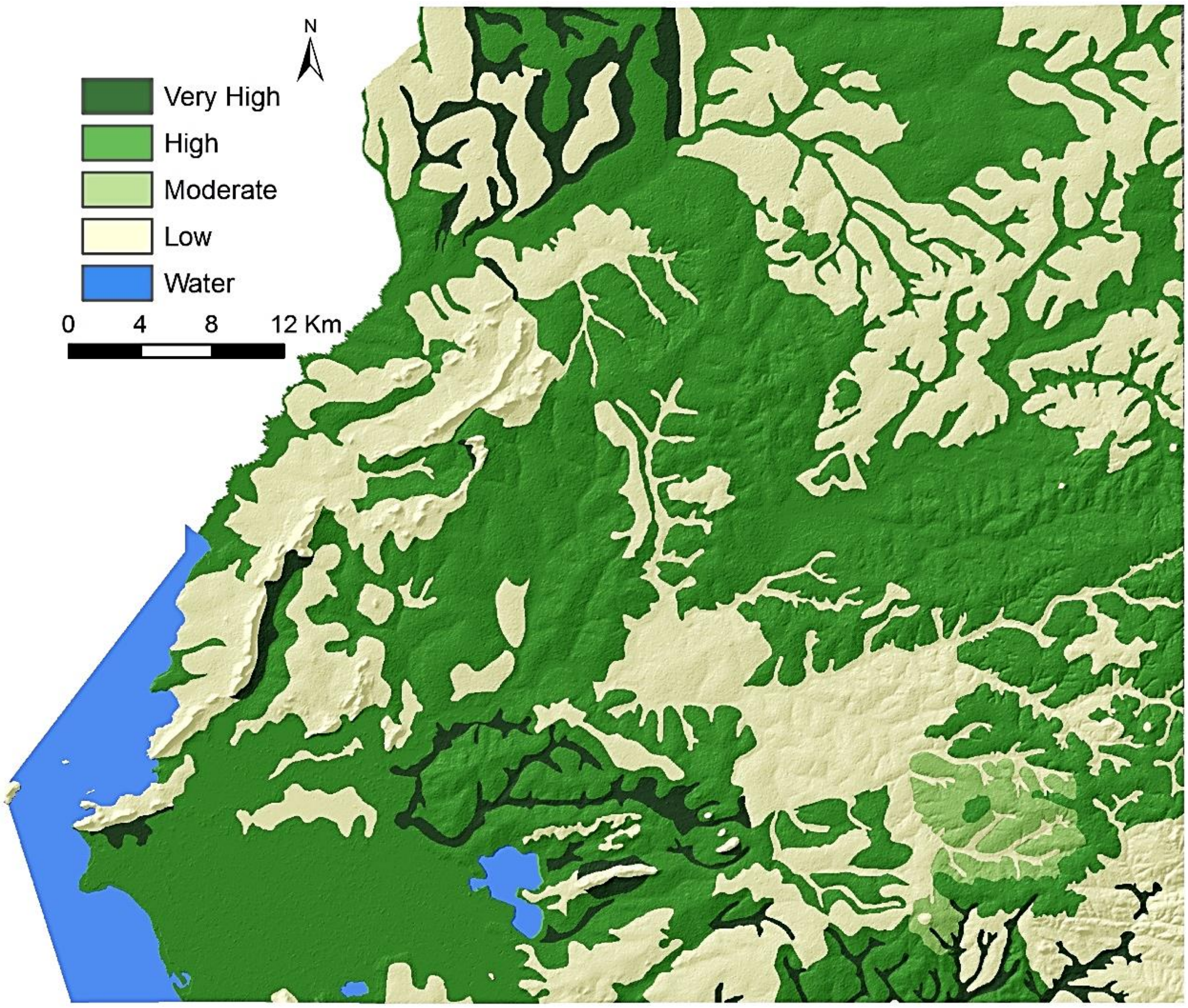
Table A.5: Rating land quality hazard of sodicity

Ratting	0 – 30 cm	30 – 100 cm
1. Very low	< 6.0	< 6.0
2. Low	6.0 – 10.0	6.0 – 15.0
3. Moderate	10.1 -15.0	15.1 – 40.0
4. High	15.1 – 40.0	>40.0
5. Very high	>40.0	>40.0



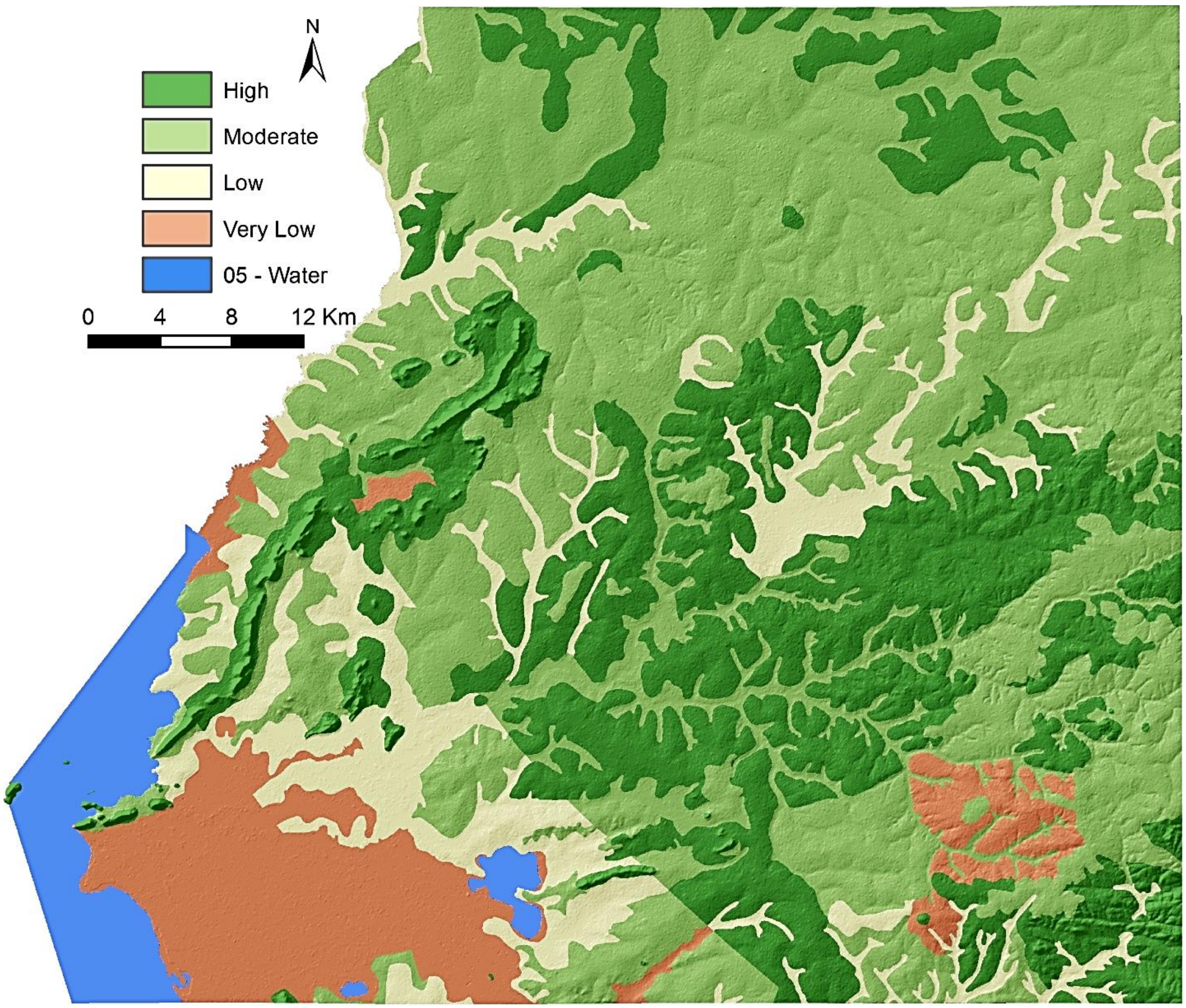
A.7: Possibilities for Seedbed Preparation and Cultivation (SPC).

This land quality reflects the possibilities for seedbed preparation and cultivation as far as the use of different kinds of implements is concerned. Low and high level of technology implements from hoe to tractor for ploughing were considered for assessment. The factors considered to be of major importance were: steepness of slope, stoniness or rockiness of the topsoil, depth of soil, consistence, presence of termite mounds, and size and form of the fields.



A.7: Susceptibility to Erosion (SE).

The following factors were considered to generate this land quality: the Moore's equation (Moore, 1979) relating to climate factor (rainfall erosivity), slope, and the soil. This land quality was obtained through the simulation of the individual factor ratings and the final result was expressed in terms of very high, high, moderate, low, and very low resistance to erosion for bare surfaces. The conditions of the plant cover will be assessed within the context of the land utilization type requirements.



APPENDIX B: BUSIA CROP SUITABILITY MAPS

Most of the arable land in the survey area are farmed in a traditional way. Some farmers, however, already apply intermediate technology. It is hoped that agriculture in the survey area will have to be intensified to cope with the rapid population growth and resulting land pressure. Therefore, smallholder rainfed 'intermediate technology' was used to develop crop suitability maps for the Busia area. Intermediate level of technology is defined as '*that level of technology where certain inputs such as fertilizers, insecticides, mechanized land preparation are used on a modest scale'* (Rachilo and Michieka, 1991). Crops growing under intermediate technology implies:

- Timely and proper seedbed preparation.
- Use of selected (improved) planting material.
- Better crop spacing.
- Weed control measures.
- A certain measure of disease and pest control.
- Use of ox-plough and sometime hired tractor service.
- Timely harvesting.

B.1: Suitability Classification of Soils for Rainfed Beans.

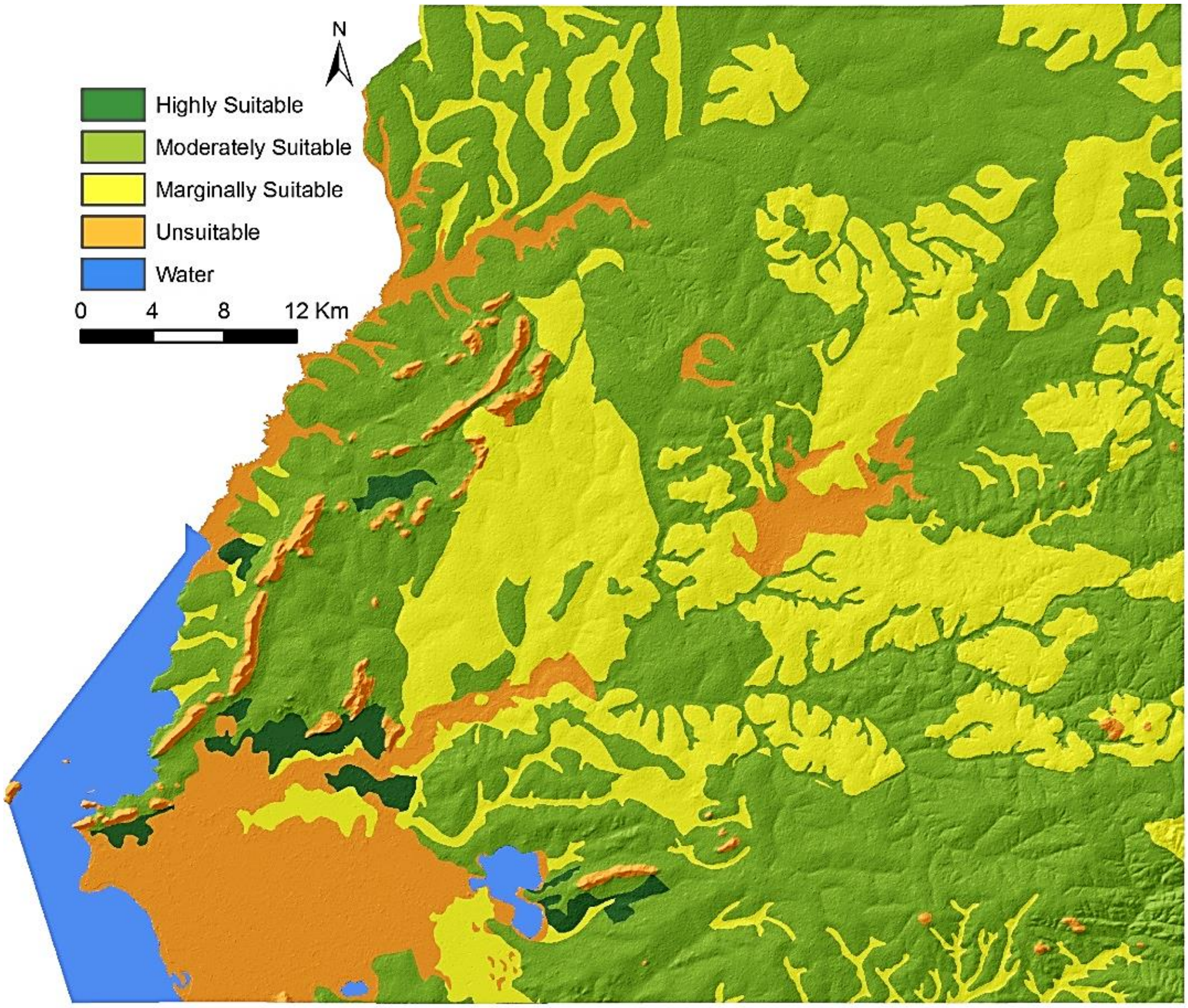


Table B.1: Decision matrix for the suitability classification of soils for rainfed beans growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	3,4	2-3	2-3	1	1	1-2	1-2	1	1-2
Moderately Suitable (S2)	2,5,6	4	4	1	1	3-4	3	2-3	3-4
Marginally Suitable (S3)	1,7,8	1	4	2	2-3	5	4	4	5
Unsuitable (NS)	9	1	4	3-4	4-5	5	5	5	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation.

B.2: Suitability Classification of Soils for Rainfed Cabbages and Kales.

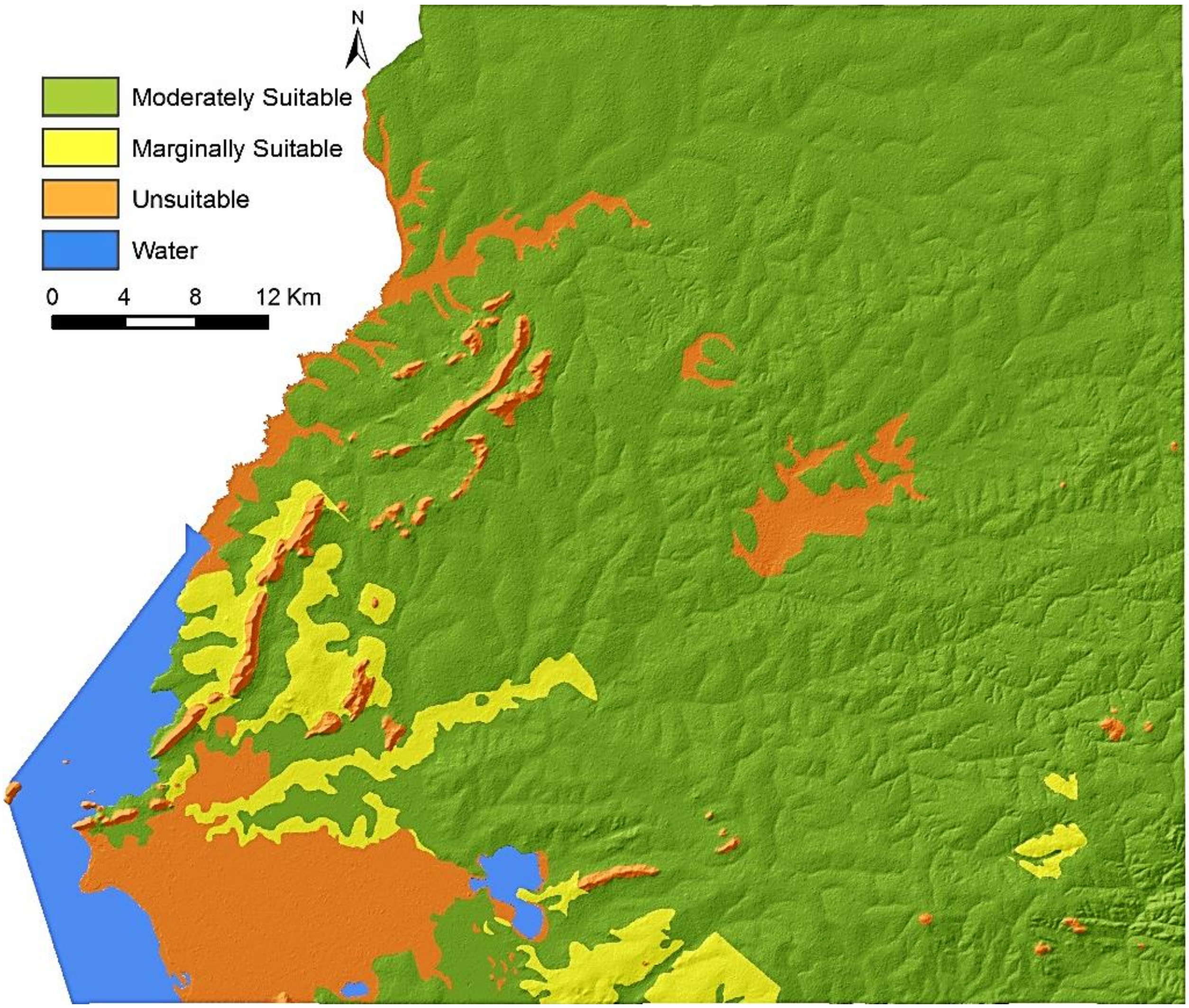


Table B.2: Decision matrix for the suitability classification of soils for rainfed cabbages and kales growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	4, 5	2	2-3	1	-	1-2	1-2	1-2	1-2
Moderately Suitable (S2)	3, 6	3	4	2	-	3-4	3-4	3-4	3-4
Marginally Suitable (S3)	2, 7	4, 1	4	3	-	5	5	5	5
Unsuitable (NS)	1, 8, 9	4, 5	4	4	-	5	5	5	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation.

B.3: Suitability Classification of Soils for Rainfed Cassava.

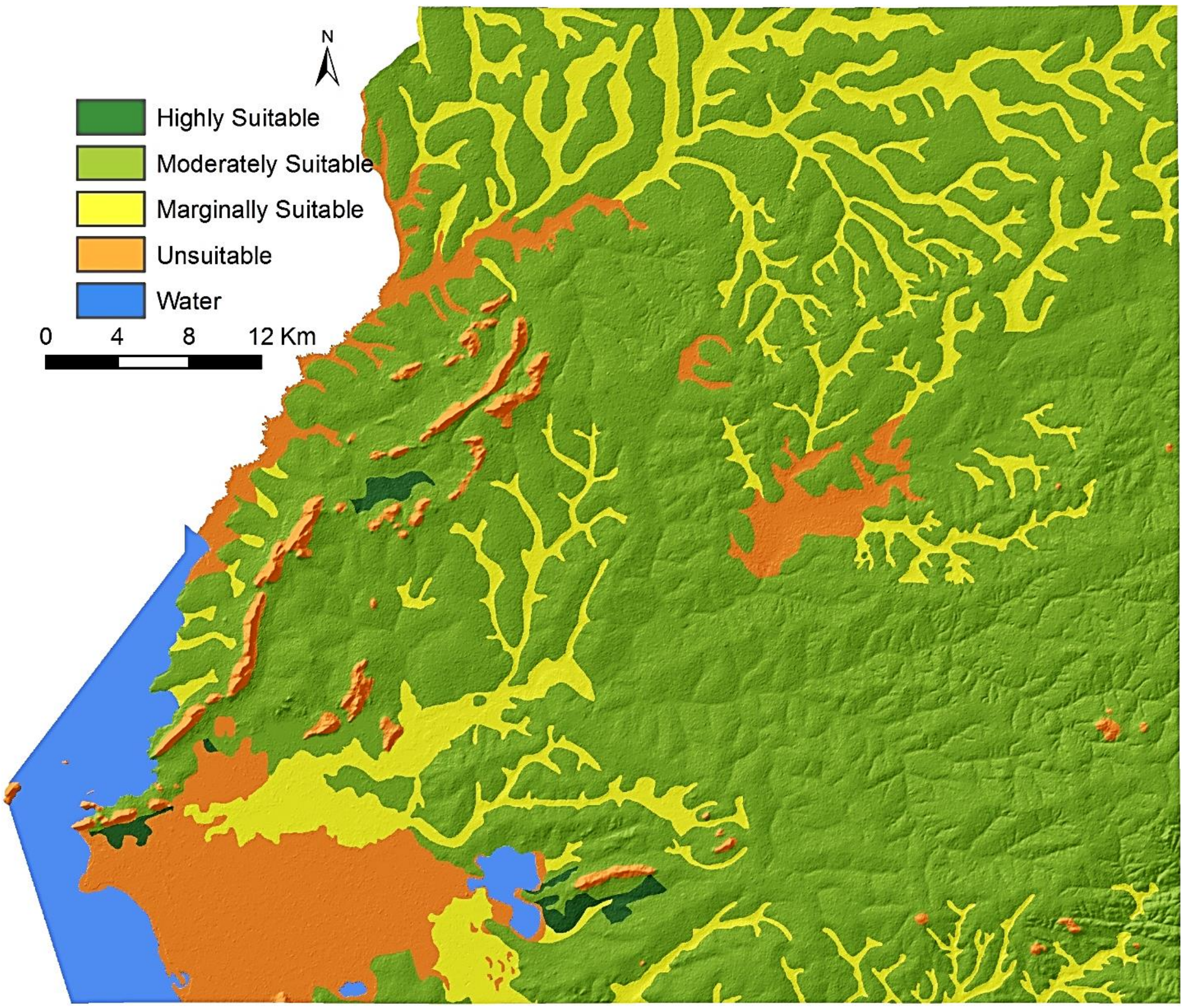


Table B.3: Decision matrix for the suitability classification of soils for rainfed cassava growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	1,2,3	1-2	2-3	1	1	1-2	1-2	1	1-2
Moderately Suitable (S2)	4	3	4	2-3	2-3	3-4	3	2	3-4
Marginally Suitable (S3)	5	4	4	4	4	5	4	3-4	5
Unsuitable (NS)	6	4	4	4	5	5	5	5	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation.

B.4: Suitability Classification of Soils for Rainfed Citrus Guava.

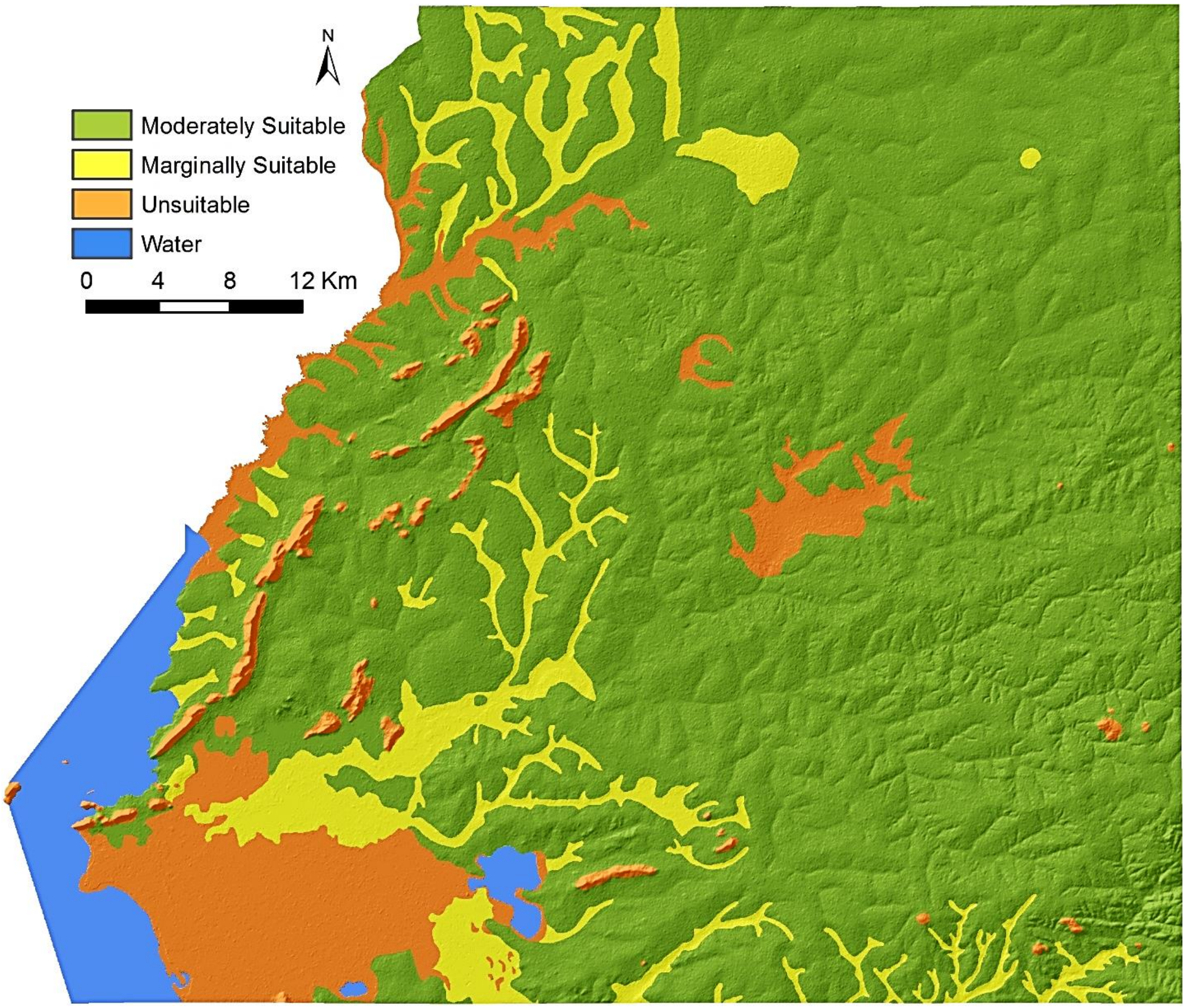


Table B.4: Decision matrix for the suitability classification of soils for rainfed citrus guava growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	SoD	SE	Ox	FH	SPC
Highly Suitable(S1)	2	1-2	2-3	1	1	1-2	1	1	1
Moderately Suitable (S2)	1, 3, 4	3	4	2 - 3	2	3-4	2-3	2-3	3-4
Marginally Suitable (S3)	5, 6, 7	4	4	4	3 - 4	5	4	4	5
Unsuitable (NS)	8	4	4	4	5	5	5	5	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation.

B.5: Suitability Classification of Soils for Rainfed Cotton.

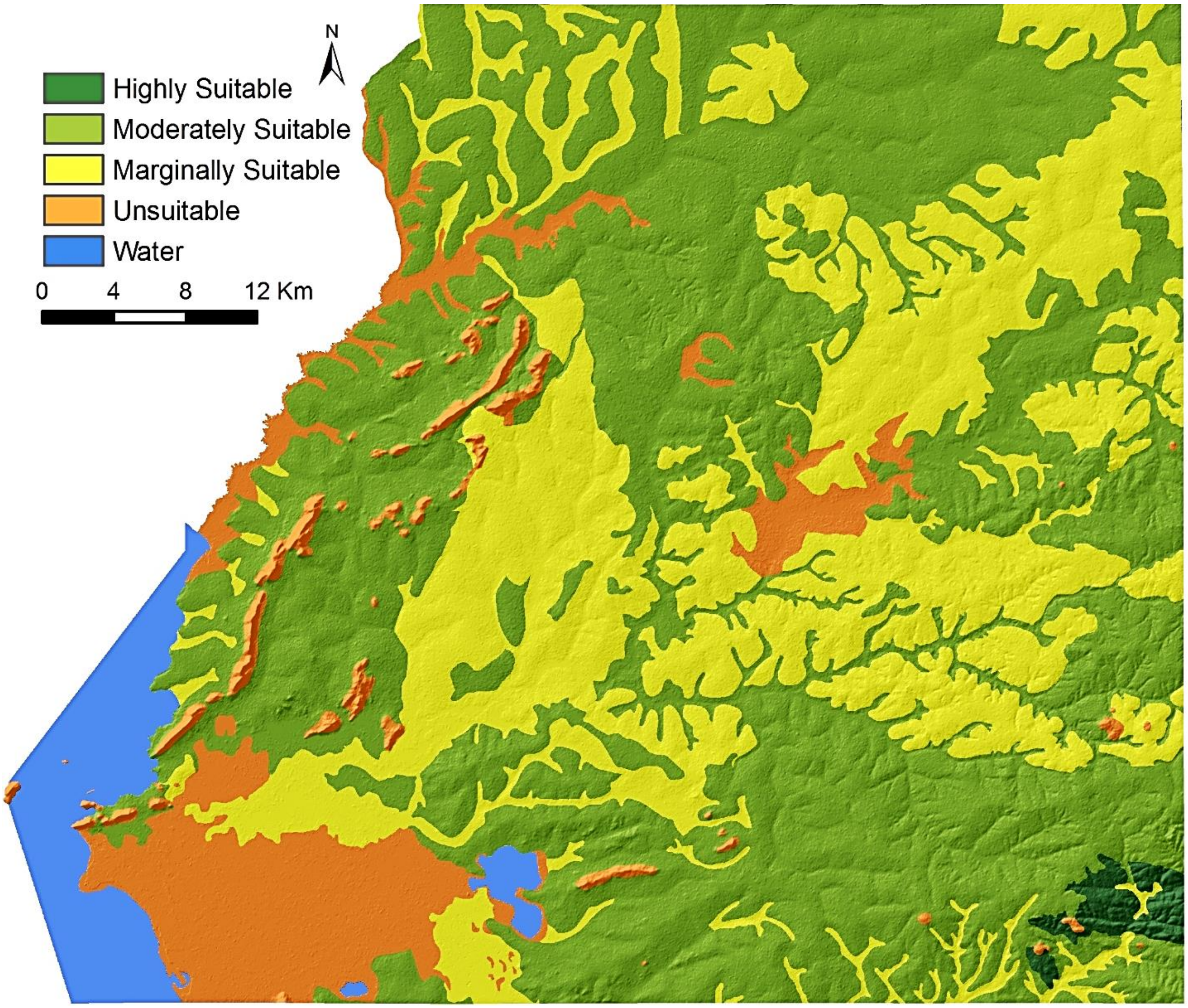


Table B.5: Decision matrix for the suitability classification of soils for rainfed cotton growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	1,2,3	2	2-3	1-2	1	1-2	1-2	1	1-2
Moderately Suitable (S2)	4	3	4	3	2-3	3-4	3	2-3	3-4
Marginally Suitable (S3)	5	4,1*	4	4	4	5	4	4	5
Unsuitable (NS)	6	1*	4	4	5	5	5	5	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation. *Too much moisture is not good for cotton growing may encourage much vegetative growth, at the expense of boll formation, especially in the mono-modal varieties.

B.6: Suitability Classification of Soils for Rainfed Fingermillet.

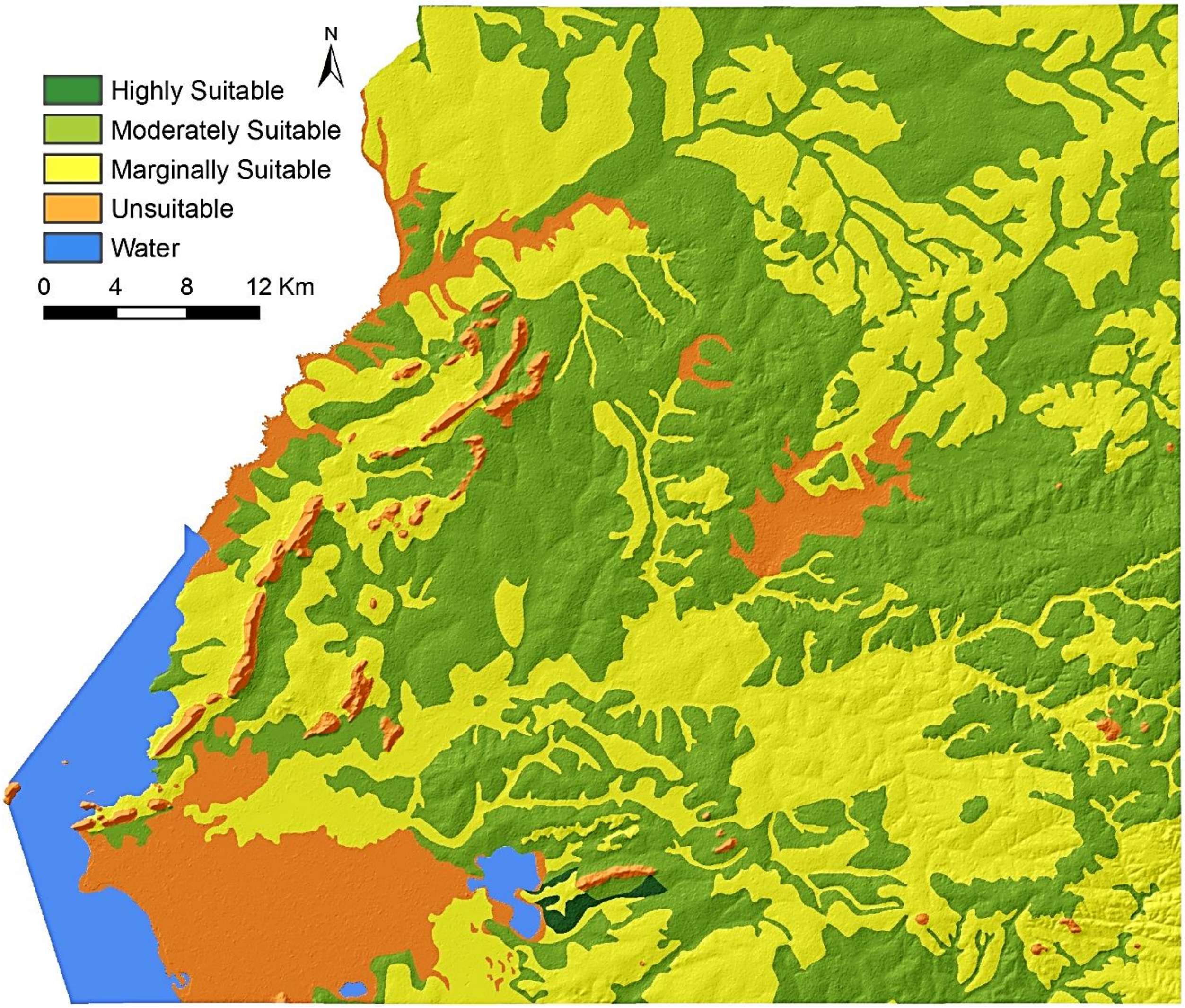


Table B.6: Decision matrix for the suitability classification of soils for rainfed fingermillet growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	1, 2, 3	1-2	2-3	1-2	1-2	1-2	1-2	1-2	1
Moderately Suitable (S2)	4	3	4	3	3	3-4	3	3	2
Marginally Suitable (S3)	5	4	4	4	4	5	4	4	3-4
Unsuitable (NS)	6	4	4	4	5	5	5	5	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation.

B.7: Suitability Classification of Soils for Fodder Crops.

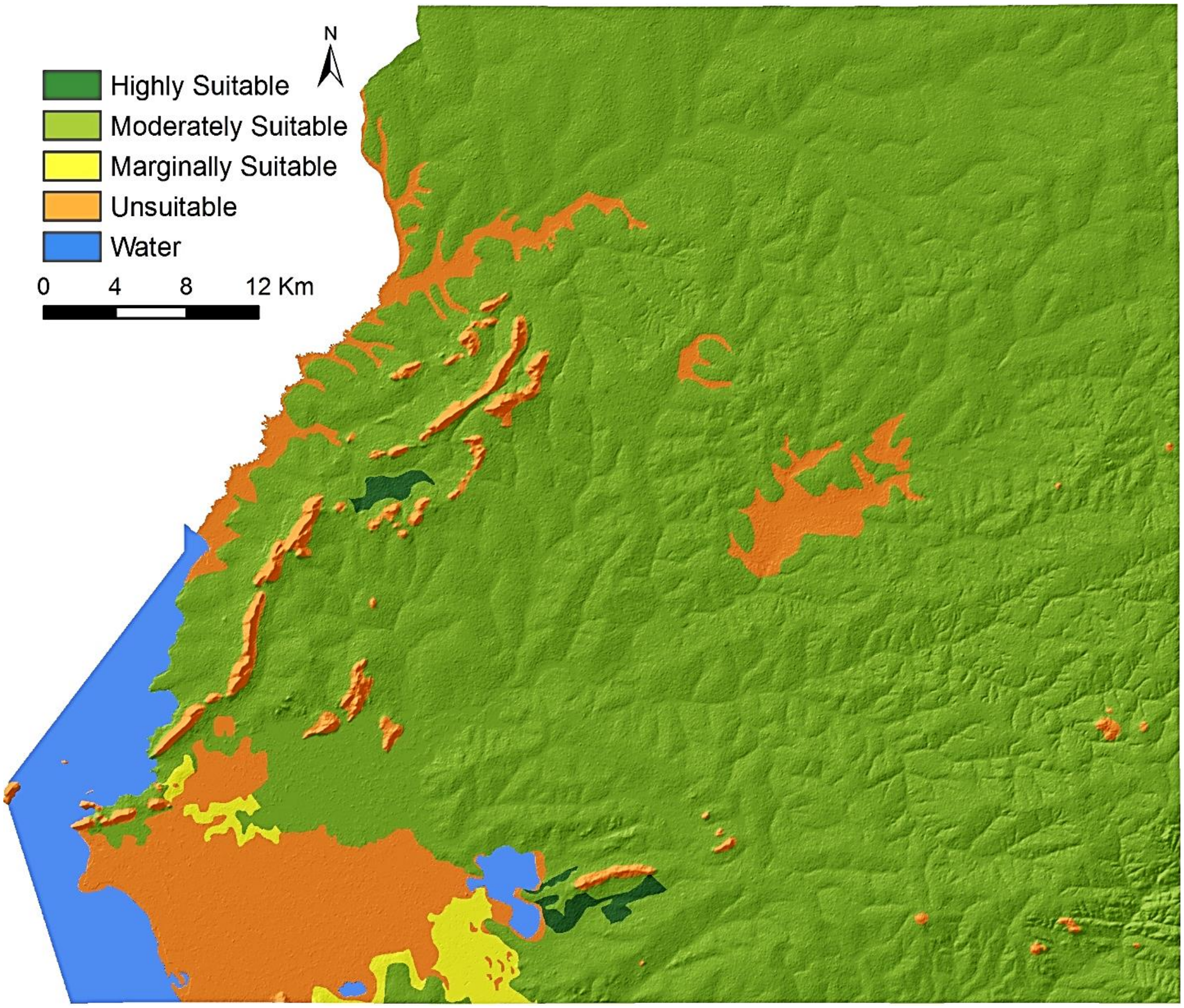


Table B.7: Decision matrix for the suitability classification of soils for rainfed fodder crops (nappier grass and maize) growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	1,2,3	1-2	2-3	1	1	1-2	1-2	1-2	1-2
Moderately Suitable (S2)	4	3	4	2-3	2-3	3-4	3-4	3-4	3-4
Marginally Suitable (S3)	4,5,6	4	4	4	4	5	5	5	5
Unsuitable (NS)	7	4	4	4	5	5	5	5	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation.

B.8: Suitability Classification of Soils for Forestry.

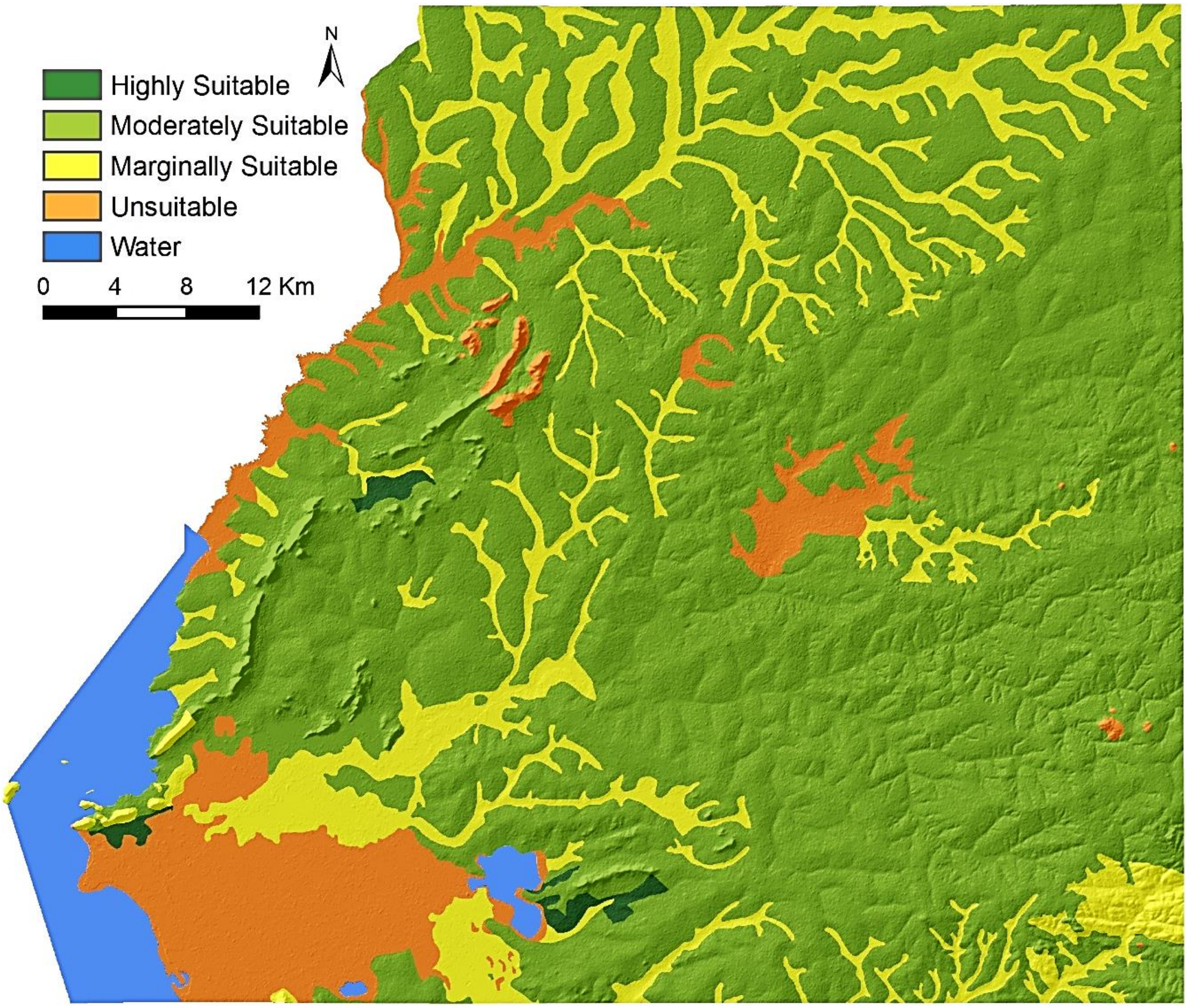


Table B.8: Decision matrix for the suitability classification of soils for rainfed forest* growing, under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC	AoF
Highly Suitable(S1)	2,3,6	2-3	2-3	1	1	1-2	1	1	1-2	1-2
Moderately Suitable (S2)	4,5	4	4	2-3	2-3	3-4	2	2	3-4	3
Marginally Suitable (S3)	1,7,8	4	4	4	4	5	3-4	3-4	5	3
Unsuitable (NS)	9	4	4	4	5	5	5	5	5	3

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation. *Forest are mainly restricted to hills and trees are of indigenous species. Some exotic tree species are planted around homesteads for timber and act as windbreaks.

B.9: Suitability Classification of Soils for Grazing.

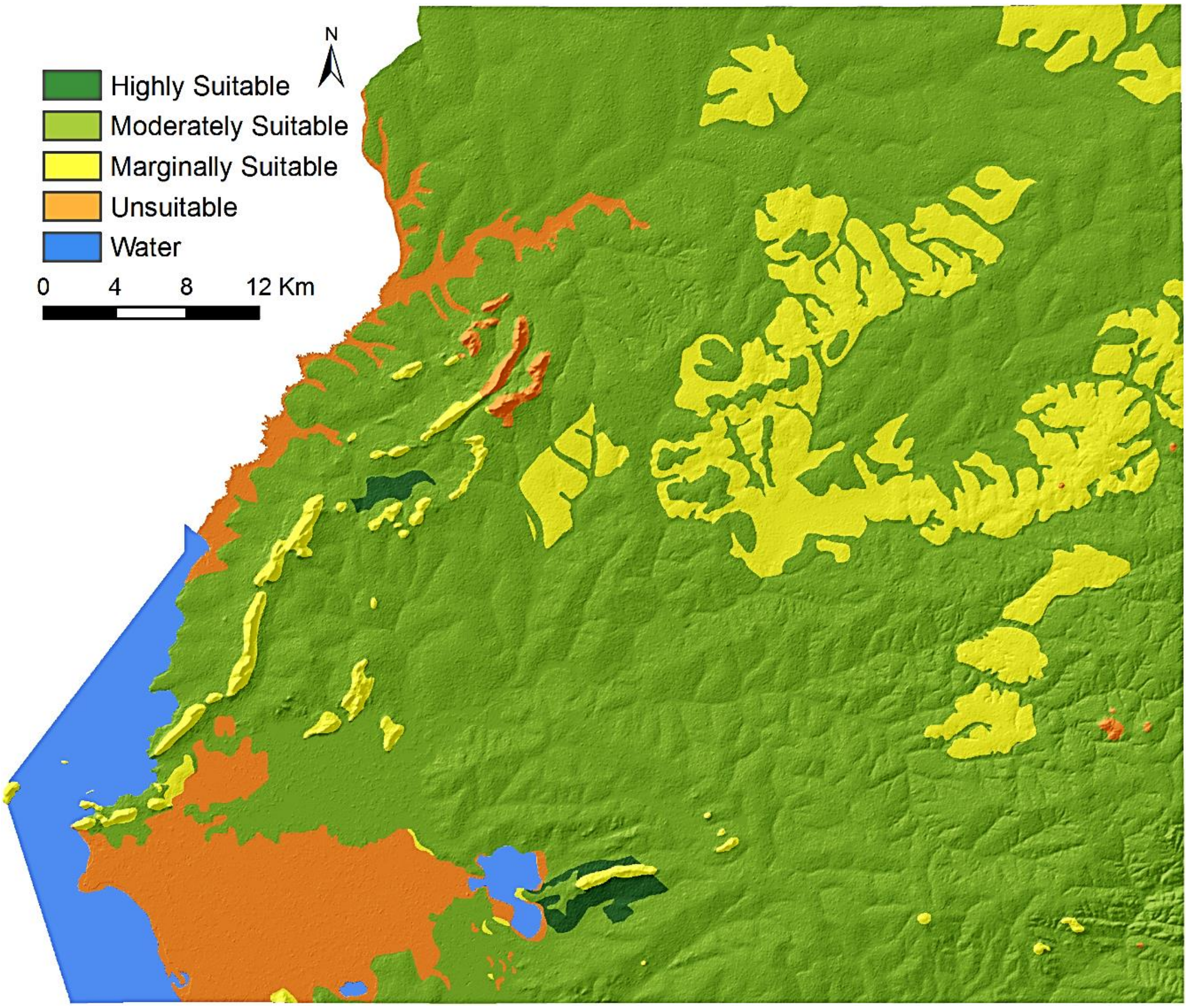


Table B.9: Decision matrix for the suitability classification of soils for grazing.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	1,2,3	1-2	2-3	1-2	1-2	1-2	1-2	1-2	n.a[1)]
Moderately Suitable (S2)	4	3	4	3	3-4	3-4	3-4	3-4	n.a
Marginally Suitable (S3)	5,6	4	4	4	5	5	5	5	n.a
Unsuitable (NS)	7	4	4	4	5	5	5	5	n.a

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation; [1)]n.a = not applicable, since the grazing is mainly free grazing in fallow and wet or bottom land areas, SPC is not considered for it.

B.10: Suitability Classification of Soils for Rainfed Groundnuts.

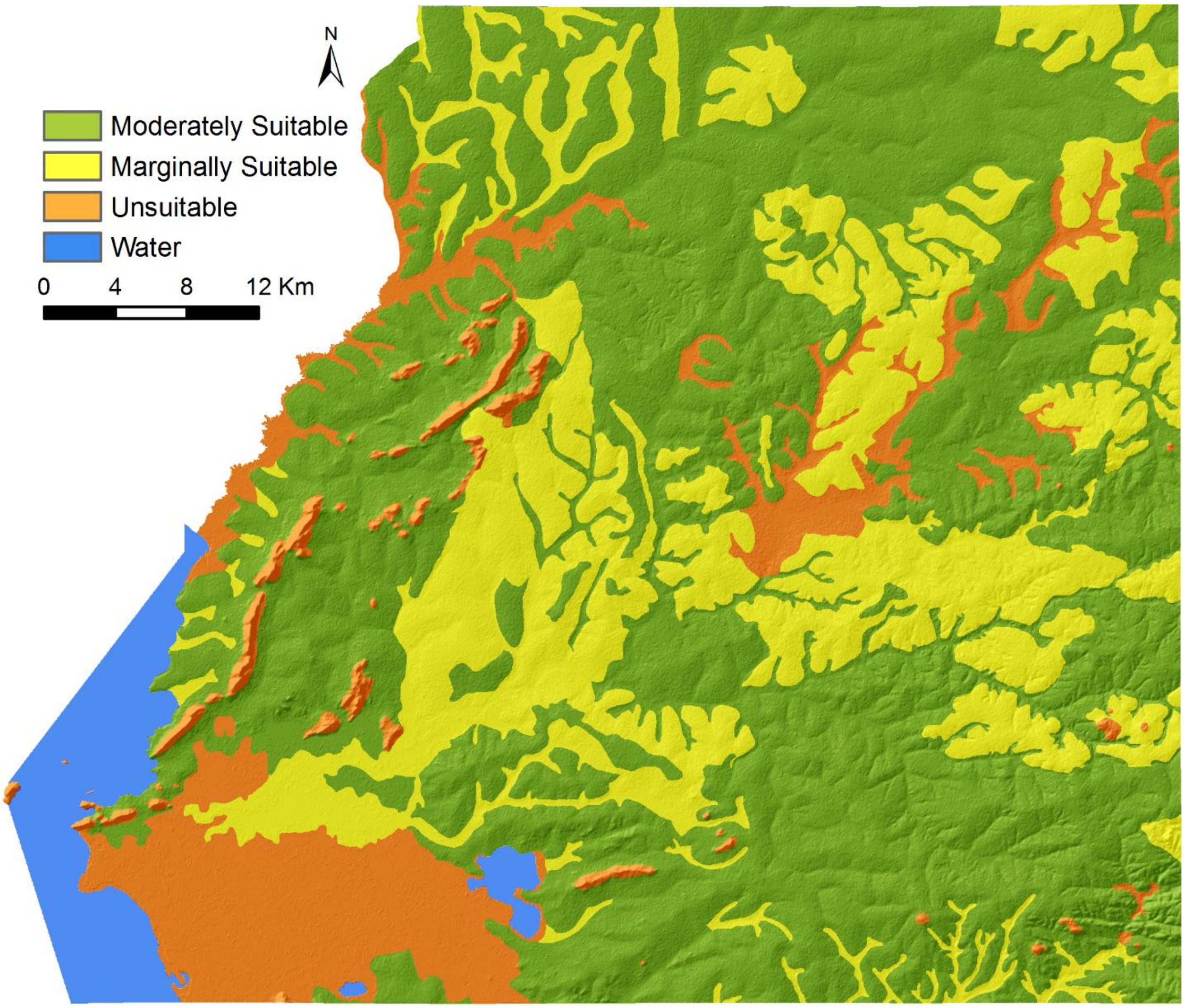


Table B.10: Decision matrix for the suitability classification of soils for rainfed groundnuts growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	1, 2	4	2-3	1-2	1	1-2	1	1	1-2
Moderately Suitable (S2)	3	3 -2	4	3	2-3	3-4	1	2	3-4
Marginally Suitable (S3)	4	1	4	4	4	5	2-3	3-4	5
Unsuitable (NS)	5	1	4	4	5	5	4-5	5	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation. In this case, the highest number is the best rating and the lowest is the worst.

B.11: Suitability Classification of Soils for Rainfed Onion.

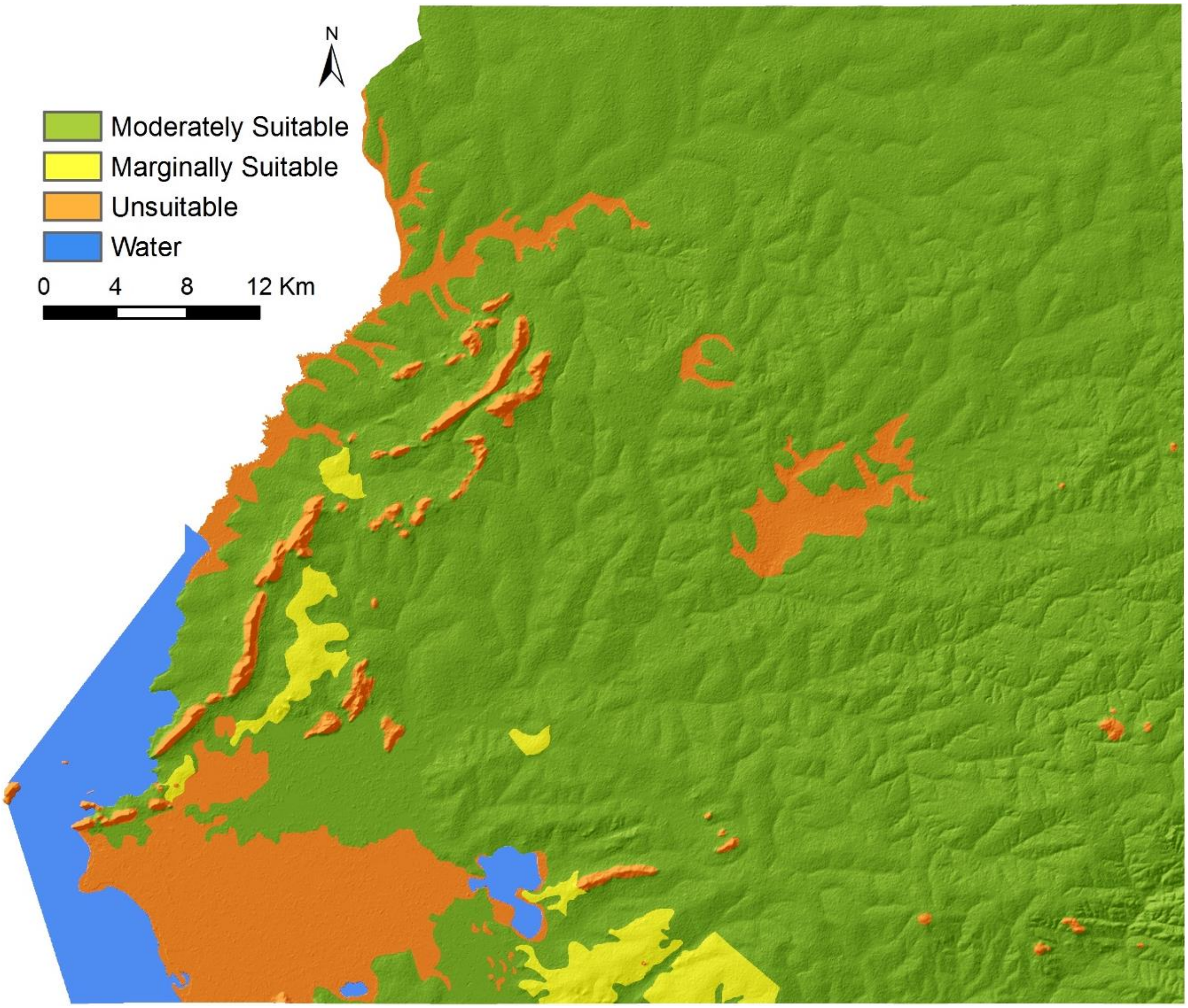


Table B.11: Decision matrix for the suitability classification of soils for rainfed onions growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	4, 5	1-2	2-3	1	1	1-2	1-2	1-2	1-2
Moderately Suitable (S2)	3, 6	3	4	2-3	2-3	3-4	3-4	3-4	3-4
Marginally Suitable (S3)	2, 7	4	4	4	4	5	5	5	5
Unsuitable (NS)	1, 8, 9	5	4	4	5	5	5	5	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation.

B.12: Suitability Classification of Soils for Robusta Coffee.

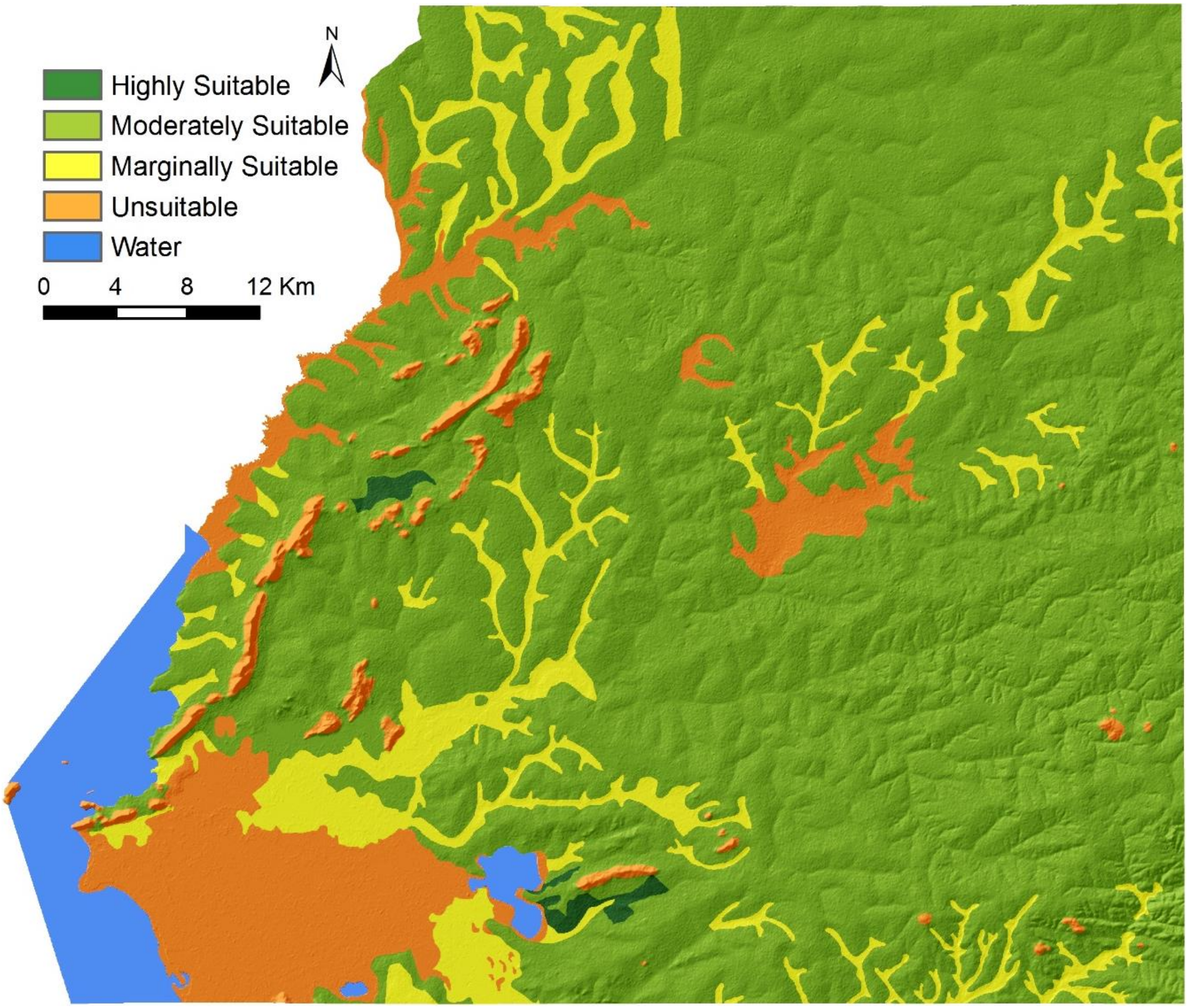


Table B.12: Decision matrix for the suitability classification of soils for rainfed Robusta Coffee growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	3,4	1-2	2-3	1	1	1-2	1-2	1	1-2
Moderately Suitable (S2)	2,5	3	4	1	1	3-4	3	2-3	3-4
Marginally Suitable (S3)	1,6,7	3	4	2	2-3	5	4	4	5
Unsuitable (NS)	8	4	4	3-4	4-5	5	5	5	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation.

B.13: Suitability Classification of Soils for Sorghum.

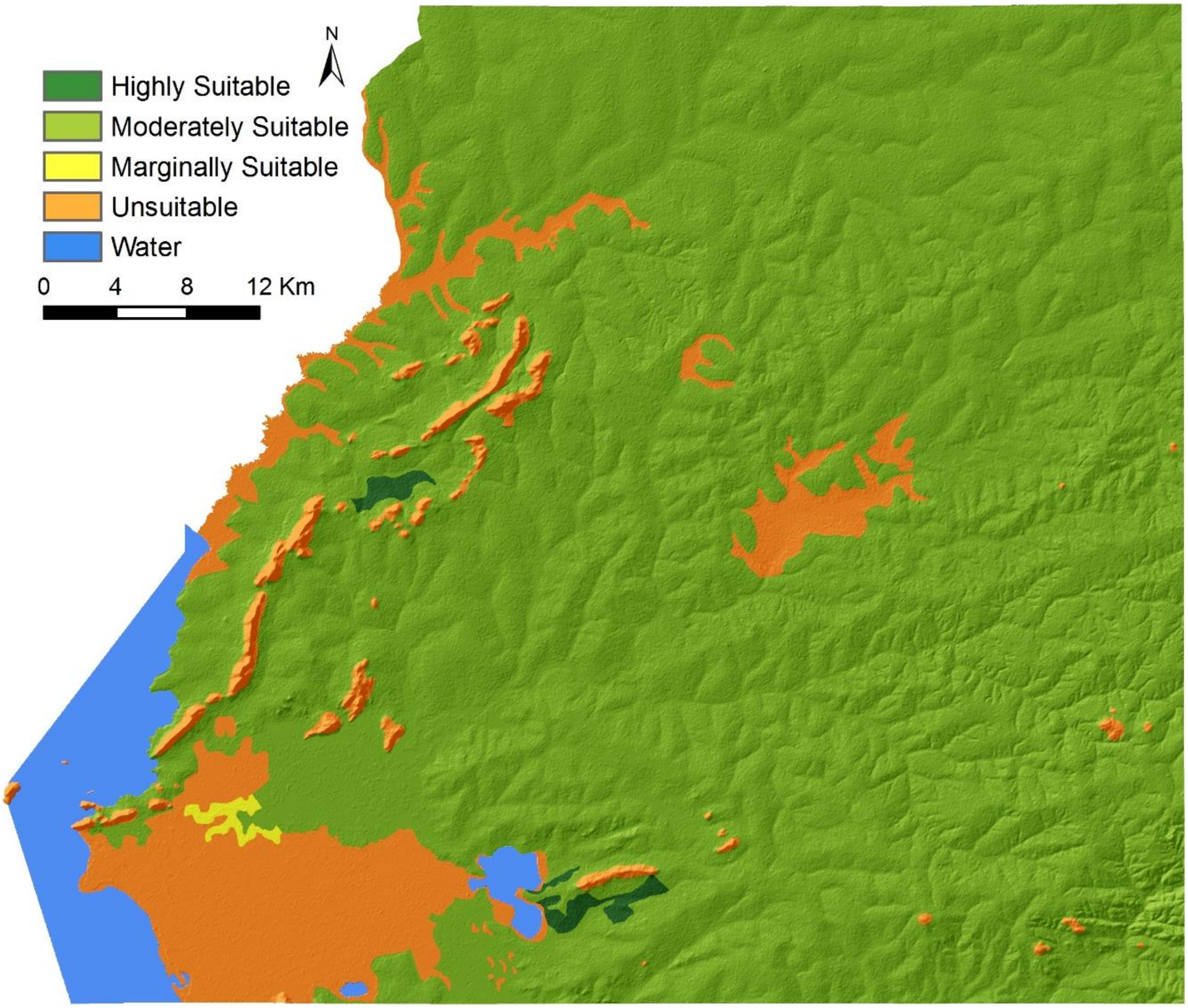


Table B.13: Decision matrix for the suitability classification of soils for rainfed sorghum growing intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	1, 2, 3	1-2	2-3	1-2	1-2	1-2	1-2	1-2	1-2
Moderately Suitable (S2)	4	3	4	3	3-4	3-4	3-4	3-4	3-4
Marginally Suitable (S3)	5	4	4	3	5	5	5	5	5
Unsuitable (NS)	6	4	4	4	5	5	5	5	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation.

B.14: Suitability Classification of Soils for Sugarcane.

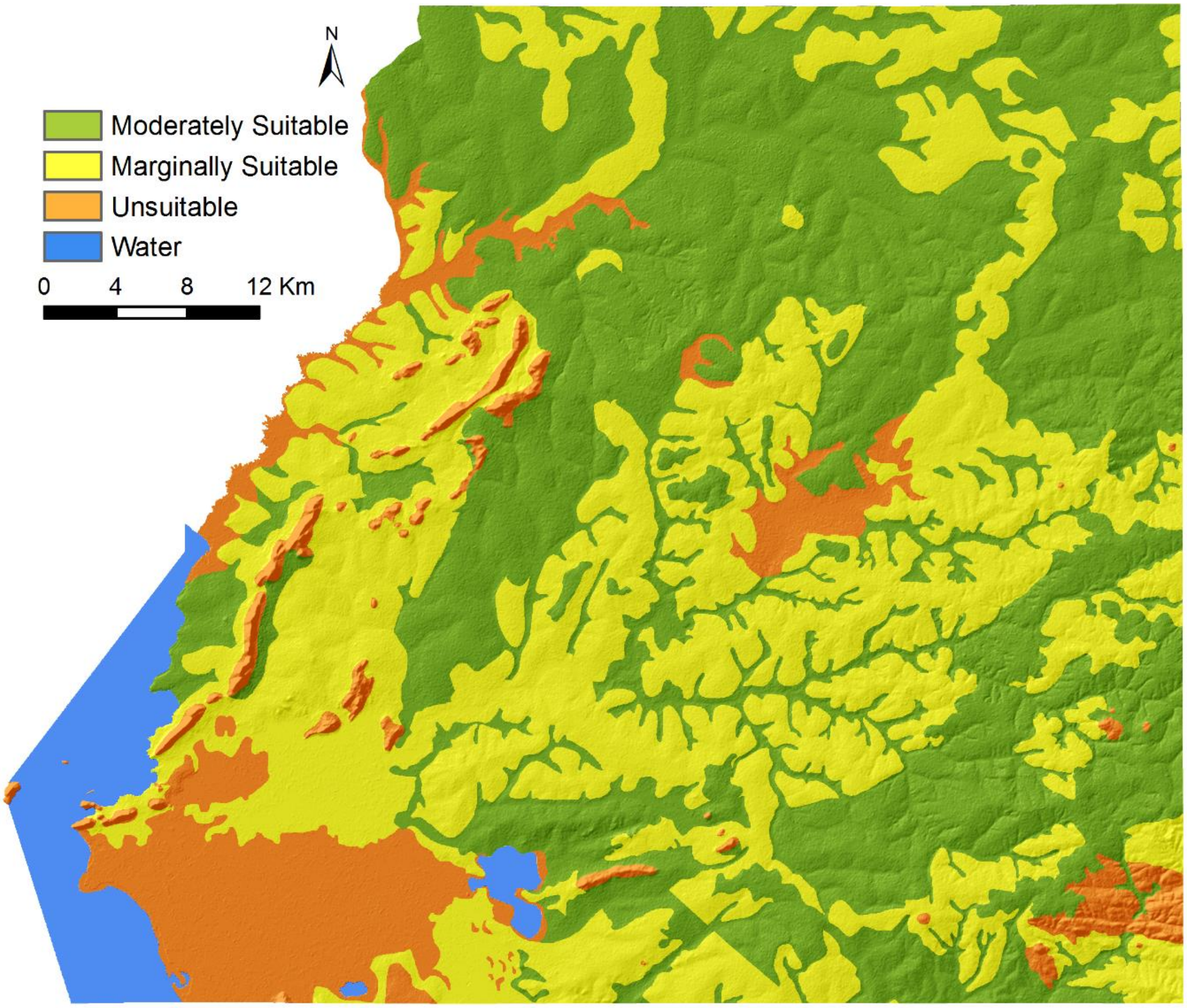


Table B.14: Decision matrix for the suitability classification of soils for rainfed Sugarcane growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	1, 2	1	2-3	1	1	1	1-2	1-2	1-2
Moderately Suitable (S2)	3, 4	2	4	2	2	2-3	3-4	3-4	3-4
Marginally Suitable (S3)	5, 6	3	4	3	3	4	5	5	5
Unsuitable (NS)	7	4	4	4	4-5	4	5	5	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation.

B.15: Suitability Classification of Soils for Sunflower.

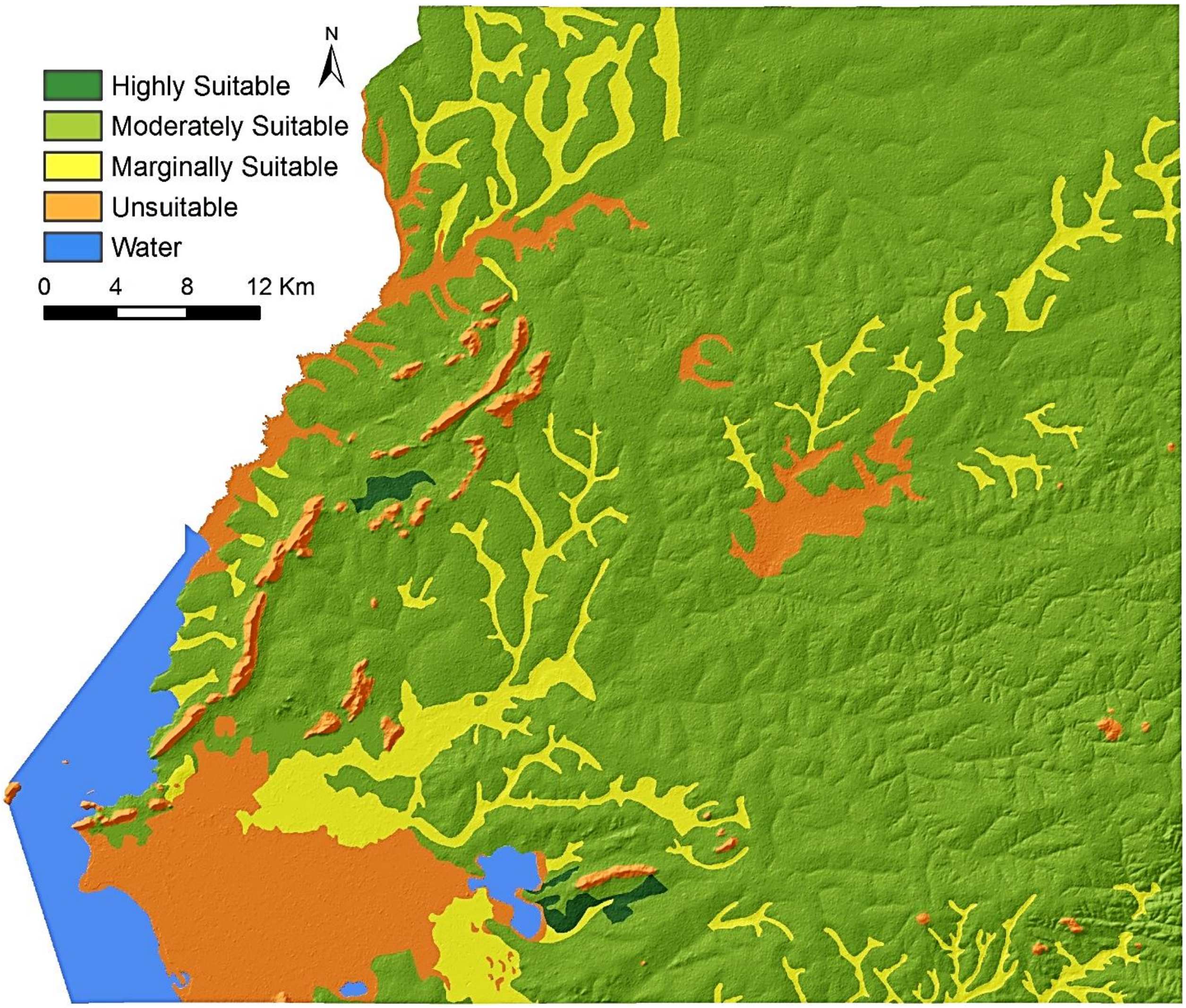


Table B.15: Decision matrix for the suitability classification of soils for rainfed sunflower growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	2,3,4	1-2	2-3	1	1	1-2	1-2	1-2	1-2
Moderately Suitable (S2)	1,5,6	3	4	2-3	2-3	3-4	3	3	3-4
Marginally Suitable (S3)	7	4	4	4	4	5	4	4	5
Unsuitable (NS)	8,9	4	4	4	5	5	5	5	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation.

B.16: Suitability Classification of Soils for Tomato.

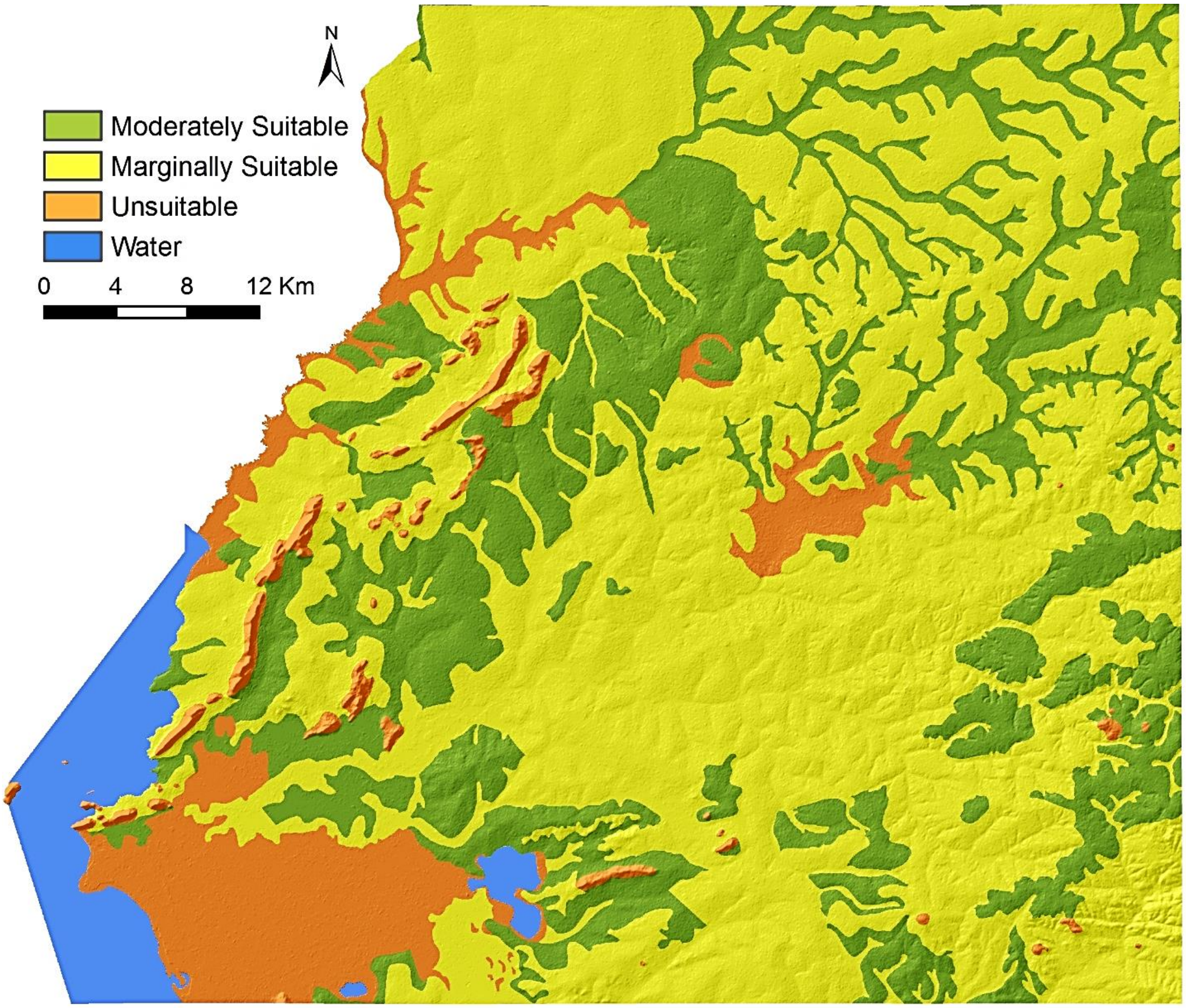


Table B.16: Decision matrix for the suitability classification of soils for rainfed tomato growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	2, 3, 4	1-2	2-3	1	1	1	1	1-2	1
Moderately Suitable (S2)	1, 5	3	4	2	3	2-3	2-3	2-3	2-3
Marginally Suitable (S3)	6	4	4	3	4	4	4	4	4
Unsuitable (NS)	7, 8, 9	4	4	4	5	5	5	5	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation.

B.17: Suitability Classification of Soils for Upland Rice.

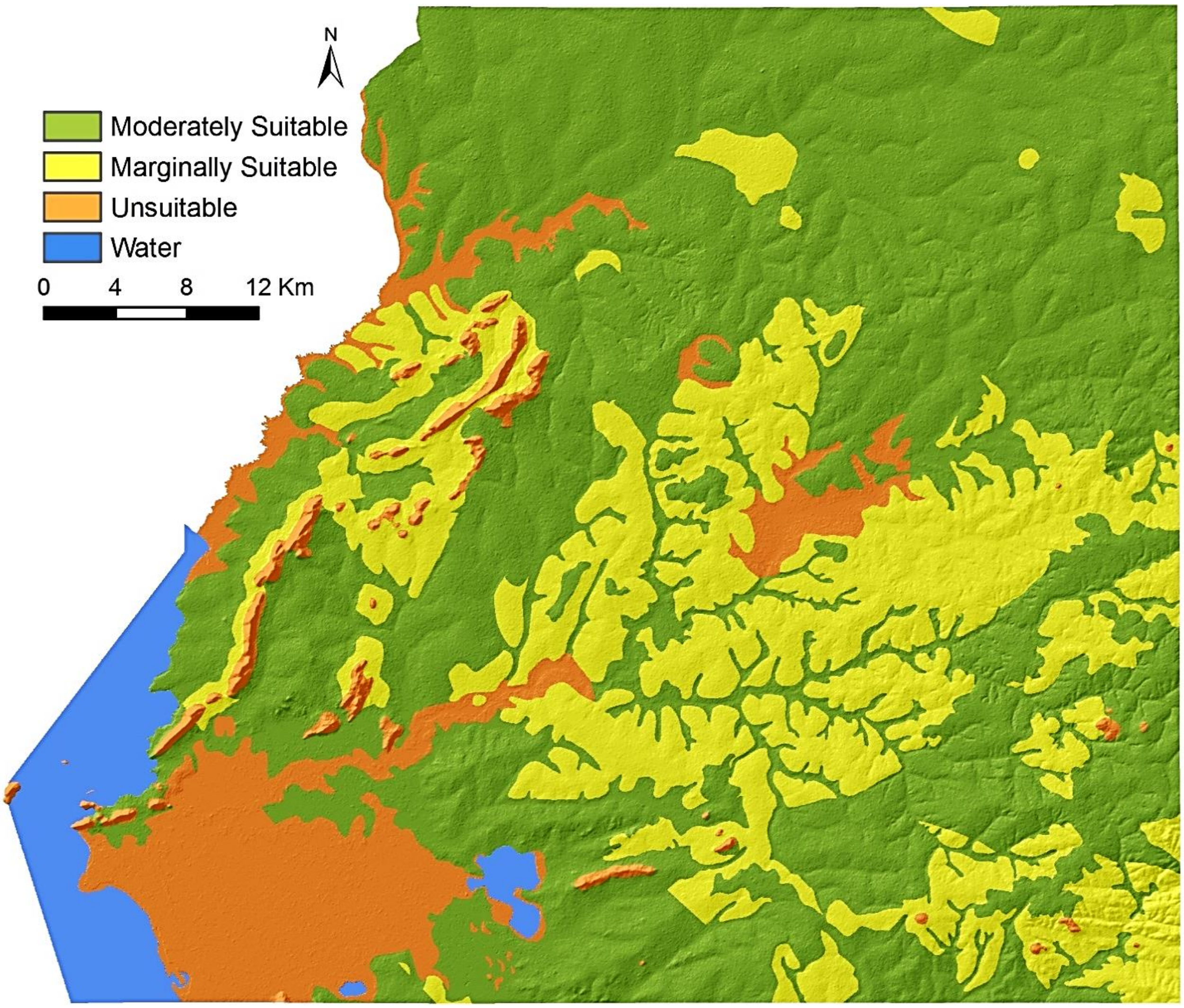


Table B.17: Decision matrix for the suitability classification of soils for upland rice growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	2	3	3	1	1-2	1	1-2	1-2	1-2
Moderately Suitable (S2)	3	2	4	2	3	2-3	3	3-4	3-4
Marginally Suitable (S3)	4	1	4	2	4	4	5	3	5
Unsuitable (NS)	5	1,4	3-4	5	5	5	5	5	5

The rating for upland rice is reversed with rating 3 being the best and 1 being the worst for land quality AoM. For land quality Ox, for wetland rice, the best rating is 5 and the worst is 1. In the case of land quality FH for wetland rice, the best rating is 5 and the worst is 2-1.

B.18: Suitability Classification of Soils for Wetland Rice.

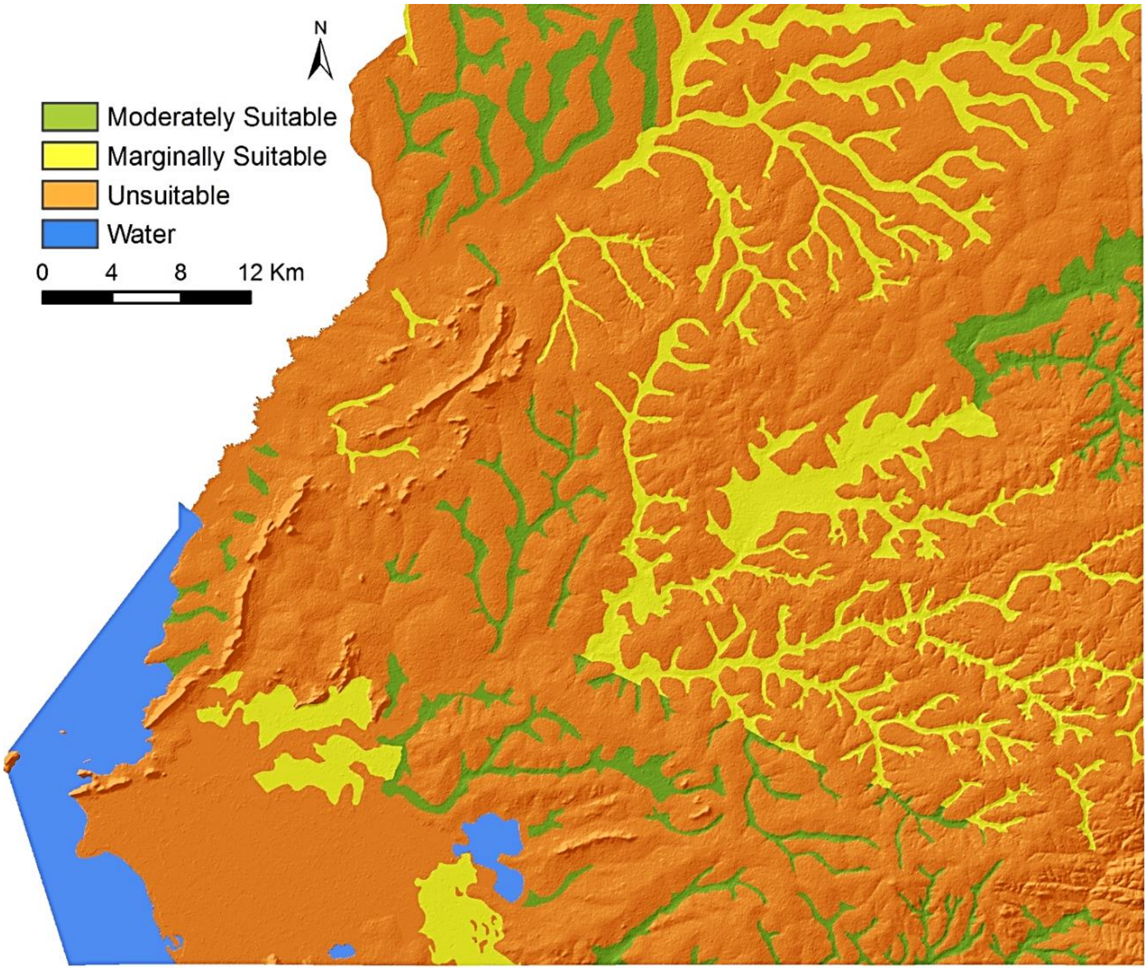


Table B.18: Decision matrix for the suitability classification of soils for wetland rice growing under intermediate technology.

Suitability Class	Temp	AoM	AoN	SH	Sod	SE	Ox	FH	SPC
Highly Suitable(S1)	1,2	1	2-3	1	1	1	5	5	1-
Moderately Suitable (S2)	3	2	4	2	3	2-3	4	4	3-4
Marginally Suitable (S3)	4	3	4	2	4	4	2-3	2-3	5
Unsuitable (NS)	5	4	4	5	5	1	1	1	5

Temp = Temperature; AoM = Availability of moisture; AoN = Availability of nutrients; SH = salinity hazard; Sod = Sodicity; SE = susceptibility of erosion; Ox = oxygen; FH = Flooding hazard; SPC = suitability for seedbed preparation. The rating for upland rice is reversed with rating 3 being the best and 1 being the worst for land quality AoM. For land quality Ox, for wetland rice, the best rating is 5 and the worst is 1. In the case of land quality FH for wetland rice, the best rating is 5 and the worst is 2-1.