APPLICATION OF HYPER-TEMPORAL NDVI DATA IN GRASSLAND MAPPING AND BIOMASS ESTIMATION IN THE MASAI MARA ECOSYSTEM, KENYA

OJWANG' DENNIS ONYANGO
February, 2015

SUPERVISORS:

Dr. Ir. C.A.J.M. de Bie
Dr. A. G. Toxopeus

APPLICATION OF HYPER-TEMPORAL NDVI DATA IN GRASSLAND MAPPING AND BIOMASS ESTIMATION IN THE MASAI MARA ECOSYSTEM, KENYA

OJWANG' DENNIS ONYANGO
Enschede, The Netherlands, February, 2015

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Natural Resources Management

SUPERVISORS:
Dr. Ir. C.A.J.M. de Bie
Dr. A. G. Toxopeus

THESIS ASSESSMENT BOARD:
Prof. Dr. A.K. Skidmore (Chair)
Prof. Dr. V.G. Jetten (External Examiner, University of Twente - ITC – AES)
Dr. Ir. C.A.J.M. de Bie (1st Supervisor)
Dr. A. G. Toxopeus (2nd Supervisor)

ABSTRACT

Rangeland vegetation mapping and assessment of its productivity is an integral aspect of ecosystem management. This study aims to map grasslands of Masai Mara ecosystem and estimate above ground grass biomass for rangeland monitoring and management. A review of previous vegetation maps show that there is need for a new mapping approach that solves the problem of misinterpretation of remote sensing data. Misinterpretation results from local distribution of rainfall which is highly variable in space and time in this area. Highly variable rainfall also affect rangeland seasonal productivity of forage in the rangeland. Therefore, a reliable model for estimating rangeland biomass that is not rainfall dependent is required. The methods used in mapping vegetation cover in this study involved; i) unsupervised image classification through ISODATA clustering and, ii) calculations of NDVI image stack statistics. Analysis of hyper-temporal Modis terra NDVI data produced classified NDVI and NDVI image statistics SD, Median and Trend. The image analysis outputs were used to design a sample scheme for fieldwork. Random stratified sampling was then followed to gather vegetation and biomass samples during fieldwork. Field samples were therefore analysed and used to characterize NDVI Classes into meaningful vegetation cover types. Biomass samples collected using quadrat and clipping technique were used to train biomass prediction model. Linear regression modelling technique was used to determine a statistically significant ($p<0.05$) model for predicting grass biomass. The statistical analysis also involved correlation coefficient calculation between measured grass biomass and explanatory variables SD, Median, Trend, distance to Bomas, animal density and NDVI as at Oct 2014. Root Mean Square Error (RMSE) was calculated for the model and used to assess its accuracy in prediction. The prediction model was validated using secondary data that was collected in Sept/Oct 2006 to check if the predicted values differ statistically to the 2006 measured data. The results of this study showed that it is possible to map vegetation cover through NDVI-derived data such as SD, Median and Trend. The mapping procedure distinguished the area into six cover units (A, B, C, D, E and F). However, some of the differences that are easily detected through remote sensing are not clearly distinguishable through field percent cover estimates because of overlaps in cover estimations. The differences in mapped cover types were investigated through a statistical test of difference using field measured grass biomass. A Kruskal-Wallis test reveal that mean of biomass measurements are significantly different between vegetation cover units; C – E, D – E, and E – F. Statistical results from Spearman's Rank correlation tests revealed that grass biomass is significantly correlated to variables SD, distance to Bomas and to animal density. Linear relationship also exist between grass biomass and NDVI though not significant. Significant model coefficients explaining biomass ($R^2 = 0.653$, $N=42$) was developed and used in predicting biomass. A Wilcoxon signed-rank test was done to compare between the model estimates and historical biomass measurements of 2006 and the results show that the two biomass datasets are not identical. This study concluded that the most reliable mapping approach to the effect of highly variable rainfall is through NDVI-Derived image products which measure the behaviour of vegetation over a longer period of time and not weather but climate dependent. However, it does not perform well in overlapping percentage cover estimates. This study have also demonstrated that SD, Bomas and NDVI measurements are key factors associated to measurable grass biomass and the approach used is not comparable to the one provided by IRLI, 2006 for this area since the results of the two studies are statistically significantly different. This study therefore recommends that, future studies should consider SD of NDVI more in vegetation cover mapping, assess biomass in different seasons with successive data and also include soil and herbivore grazing intensity in order to get an improved biomass prediction model.

Keywords: Vegetation cover units, mapping, grass, NDVI, SD, Median, prediction, biomass, Masai Mara, hyper-temporal

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my two supervisors Dr. Kees de Bie and Dr. Bert Toxopeus for your patience and dedication in supervising me in every step of this study. I enjoyed your company, suggestions and support during my fieldwork in Kenya. Your critical comments, insights and support during the entire study is immensely appreciated. Thanks to Dr. Bert for your effort in coordinating and ensuring that fieldwork was a success and all required documents were processed in time. Thanks to Dr. Kees for ensuring and checking that remote sensing and other data were properly processed for the success of this study. Thanks to both of you for inspiring me throughout this process of a challenging scientific study.

I am very grateful to Kenya Wildlife Service (KWS) for granting me the permission to conduct this study in the Masai Mara National Reserve. Thanks to head of ecosystems and landscapes conservation, Dr. Kanga, I really appreciate your assistance. Thanks to ILRI, specifically Dr. Said Mohammed for your effort in ensuring that this study got the necessary data from ILRI.

I would like to thank the entire ITC-University of Twente for the opportunity to follow a worthwhile MSc programme. It is a dream come true to study in the world renowned Geoinformation and Earth Observation institution. I am also thankful to the Dutch government for funding my studies through the Netherlands Fellowship Programme (NFP). It is highly appreciated and shall forever remain part of me, The Dutch.

Special thanks to Centre for Training and Integrated Research in Arid and Semi-Arid Lands Developments (CETRAD) team in Nanyuki, Kenya. I specifically would like to pass my gratitude to the director, Dr. B.P. Kiteme, deputy director, Mr. Samoka Ongwae, head of GIS and Remote Sensing department, Mr. E. Njuguna and also special thanks to colleague Mrs Caroline Ouko for your encouragements and prayers. Thanks to all of you for your enthusiasm and effort to ensure that I succeeded in my studies.

I would also like to thank all my ITC classmates especially those who we worked with under the framework of MaMaSe project; Jared Buoga, Vella Kwamboka and Maina Ben. Thanks to you and thanks to the MaMaSe project. Thanks to those who worked with us all through during fieldwork; Ms Alya Debie, Mr. John Kimeu and Loureen Akinyi. Thanks to Kenneth Rutto of Masai Mara University for your dedication in ensuring my samples were well processed.

To the Kenyan community in the Netherlands, thank you so much. You made me feel at home through the light moments we had. Thanks to David O, Josyline M, Beryl N, Irene M, Felix O and Enock O. You always encouraged me especially when everything seemed tough academically. Thanks to Festus Ihwangi for always being there to share ideas and give insights on various issues.

I am also indebted to my parents for bringing me up and ensuring that I went to school. Thanks to all my family members for your continuous support, encouragement and prayers during the entire period of my study. I love all of you. Special thanks to Christina Barasa for it was a challenge as our son Jayden was born just weeks after I left for studies. You stood strong, encouraged and gave me moral support to do my best in my studies. Lots of love to you.

Thanks to God Almighty for enabling me through this study. All the Glory back you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

ASAVGOL: Adaptive Savitzky-Golay

DRSRS: Department of Remote Sensing and Resource Survey

EVI: Enhanced Vegetation Index

GIS: GeoInformation System

GPS: Global Positioning Satellite

ILRI: International Livestock Research Institute

ISODATA: Iterative Self-Organizing Data Analysis Technique

KWS: Kenya Wildlife Service

LOOCV: Leave One Out Cross Validation

MaMaSe: Mau Mara Serengeti Sustainable Water Initiative

MODIS: Moderate Resolution Imaging Spectroradiometer

MSAVI: Modified Soil-Adjusted Vegetation Index

NASA: National Aeronautics and Space Administration

NDVI: Normalized Difference Vegetation Index

RMSE: Root Mean Square Error

RVI: Ratio Vegetation Index

SAVI: Soil-Adjusted Vegetation Index

SD: Standard Deviation of time series NDVI

SOP: Standard Operations Procedure

SRTM: Shuttle Radar Topography Mission

VIF: Variance Inflation Factor

1. INTRODUCTION

1.1. Background

Rangelands are in remote areas of the world with low human population densities and cover about 50% of the earth's total surface area (Laliberte, Winters, & Rango, 2011). They comprise of grasslands, woodlands, wetlands and shrub lands. Generally, rangelands consist of shrub, grass, savannah and sparse woody vegetation. Throughout these areas, there are variations in vegetation types, climate, animal species and management systems that make rangelands vary in terms of their biological and economic productivity (Menke & Eric Bradford, 1992). Grasslands are essential parts of rangeland ecosystems in that, grassland biomass are key to maintaining the services of a rangeland ecosystem. In many countries, rangelands form part of protected areas and are a home to many wildlife species.

The main use of rangelands is to provide forage for both grazing wildlife and livestock. Grassland production is determined by the amount and timing of rainfall, soil type, temperature and fire (Yeganeh, Khajedein, Amiri, & Shariff, 2012). Present trends of climate change cause variations in vegetation compositional state (Boorman, 1997) and this have implications on ecological systems and wildlife species distribution (Mundia & Murayama, 2009). As human population increases, the biodiversity in turn face stiff competition for the shrinking resources (Mundia & Murayama, 2009) causing many rangelands to be degraded.

In East Africa, rangeland ecosystems have been undergoing unprecedented period of change, which have implications on their sustainability to wildlife and human beings. Climate variability and land-use changes have devastating effects on rangeland ecosystems and wildlife. Environmental changes and economic developments in and around rangelands contribute to wildlife disturbance, loss of biodiversity, pollution and reduction in food and water supply to wildlife.

Masai Mara ecosystem, located South-west of Kenya, is a reserve surrounded by group ranches and conservancies and embodies many of the current issues in biodiversity conservation (Mundia & Murayama, 2009). This ecosystem is increasingly getting transformed by the agro-pastoral communities adjacent to it (Homewood et al., 2001). In addition, commercial farming and tourism activities transform and threaten the sustainable living of pastoral people and wildlife. These threats can be seen as effects of grazing observable in vegetation properties such as cover, cover fractions, plant species diversity and herbage production (Yang & Guo, 2011) in the grasslands. Development of lodges and urban centres are among some of the landuse changes that have an impact to this reserve. Due to these developments, wildlife are disturbed and restricted to this ecosystem leading to competition for water and food within and around the reserve.

Conservation and protection of rangelands require a clear understanding of local and temporal distribution of grasses for wildlife grazing. Masai Mara ecosystem is an immense rangeland and may prove difficult to assess its resources by just ground observation techniques. Remote sensing offers reliable techniques for monitoring, assessing and estimating rangeland productivity over time. Through satellite imagery data, evaluation and comparison of vegetation cover changes have been possible and it has proved to be a very useful tool for estimating grass production (Biro, Pradhan, Buchroithner, & Makeschin, 2013).

Previous studies have showed that vegetation indices derived from remotely sensed data are correlated to grass or forage production in the rangelands (Yeganeh et al., 2012). Indices such as NDVI, SAVI and RVI have been used in such studies to show the relationship between grazers' distribution and forage production. This relationship between NDVI and vegetation phenology have been used in mapping and discriminating rangeland changes as could be observed on satellite image data (de Bie, Mobushir, Toxopeus, Venus, & Skidmore, 2008). However, the maps in most cases relied on a one time image data to map vegetation cover. Correlation of grass production and vegetation indices have shown that above ground grass biomass production vary from time to time (Yeganeh et al., 2012) depending on many affecting factors. The variations in biomass can be mapped and explained through hyper-temporal remote sensing image data collected over an area for different growing seasons.

This study explored the potentials of hyper-temporal 16-day Modis Terra NDVI data in vegetation cover mapping and above ground grass biomass estimation. Through field assessment of vegetation composition, structure, density and percentage cover, classification of NDVI data was possible for vegetation mapping. Field samples of above ground grass biomass measurements were used to train a model for biomass estimates for the entire study area. The aim of this study was to map grassland vegetation cover and estimate rangeland grass biomass through a reliable statistical model for this ecosystem.

This study was carried out under the framework of MaMaSe Sustainable Water Initiative project. The initiative aims at improving water safety and security in the Mara River Basin to support structural poverty reduction, sustainable economic growth and conservation of the basin's ecosystems. The initiative is through the financial support of the Netherlands Embassy in Nairobi, Kenya. It consists of broad-based public-private partnership including international and Kenyan government agencies, civil society, and private sector, NGO, and knowledge institutions. This study was in line with one of the of MaMaSe objectives which was to ensure that key forest and savannah ecosystems are protected or restored and wildlife get access to habitats and water resources needed at different times of the year, especially during drought years.

1.2. Problem statement

Masai Mara ecosystem was originally used for pastoral livestock and wildlife grazing. In recent years, the ecosystem has experienced major changes in landuse and tenure as a result of increased human settlement (Ottichilo, 2000). Originally, the land in this ecosystem was owned by the indigenous Masai people as communal land held in trust for them as trust land by Narok county council. In mid-1960's, changes in land tenure system begun which have led to fragmentation and conversion of the rangelands in the northern part of the ecosystem to arable agricultural land. The areas that used to be for wildlife grazing during dry season are now reduced to farming lands because the soils there are fairly fertile and moisture is favourable for crop growing. Conversion of parts of Mara ecosystem to farmlands, fencing and different landuse practices together with other environmental factors such as drought have caused disturbance to animals and their movements leading to a reduction in their populations in the recent decades (Ottichilo, 2000). The reduction in wildlife grazing areas as a result of land fragmentation and fencing of what used to be pasture land have restricted animals to limited areas for grazing and water points causing competition for the life-supporting habitat services found in the ecosystem.

Sustainable utilization of grass resources within this ecosystem by pastoralists, ranchers and wildlife calls for an understanding about the changes going on in the grazing areas at any particular time. In Masai Mara ecosystem, the number of negative changes in dominant grazing areas are attributed to overgrazing and ever rising human activities (Ottichilo, 2000) and this require new scientific knowledge to manage from time to time. There has been previous studies that attempted to quantify biomass accurately using traditional methods, (Boutton & Tieszen, 1983 and Boutton, Tieszen, & Imbamba, 1988). These attempts were based on statistical correlations of biomass to rainfall in most cases so as to make biomass estimations. Since rainfall in this area is highly variable in patterns over space and time (Ogutu, Piepho, Dublin, Bhola, & Reid, 2008), it is still a problem to come up with a good model for estimating biomass.

Rainfall has been assumed to follow a bimodal distribution across Masai Mara Ecosystem, however, local distribution is highly variable in space and time leading to a very high level of misinterpretation of remote sensing images. Some land cover maps exist for this area, for instance, the one done by Mundia and Murayama (2009) was made through classification of images that are climate representative and not weather representative. This poses a challenge in mapping vegetation due to rainfall variability in this area. Classification of hyper-temporal NDVI image data offers an opportunity to overcome this challenge in mapping vegetation. Little has been done using NDVI-derived parameters to map vegetation in this ecosystem. The irregularity in the peaks of NDVI profile over the past 14 years illustrate growth in vegetation due to variability in precipitation in every growing season, see Figure 1.

Application of remote sensing has been immense in recent times in monitoring rangeland pasture. In order to assess the capabilities of remote sensing technologies, satellite data have been used to assess changes in range vegetation phenology by considering vegetation indices such as NDVI, SAVI, RVI, EVI and MSAVI. Spatial variability of grazing animals over pasture land in Masai Mara have been assessed and mapped using satellite data (Oindo, 2008) so as to influence decisions on rangeland exploitation and management as well. Understanding of local and temporal distribution of grasses through properties measurable by remote sensing is more reliable than traditional methods which are limited to local areas (Yeganeh et al., 2012). Modis multi-temporal image data have been applied in forage production analysis and according to Yeganeh et al. (2012) NDVI, SAVI and RVI indices have moderate correlations with forage over time. Remote sensing techniques have not been fully exploited to make quantitative measurements of biomass production and assessments of the changes in space over time for this area. However, there are a few studies, mentioned earlier, that have reported biomass production estimates for this ecosystem mainly through traditional estimation methods.

Lack of a proper map for grassland vegetation that matter in this ecosystem, lack of proper model for biomass estimation and monitoring are the key problems that this study seeks to address by analysing hyper-temporal NDVI images to map and provide an understanding to the spatio-temporal dynamics of rangeland vegetation and production in relation to utilization practices by the people and wildlife.

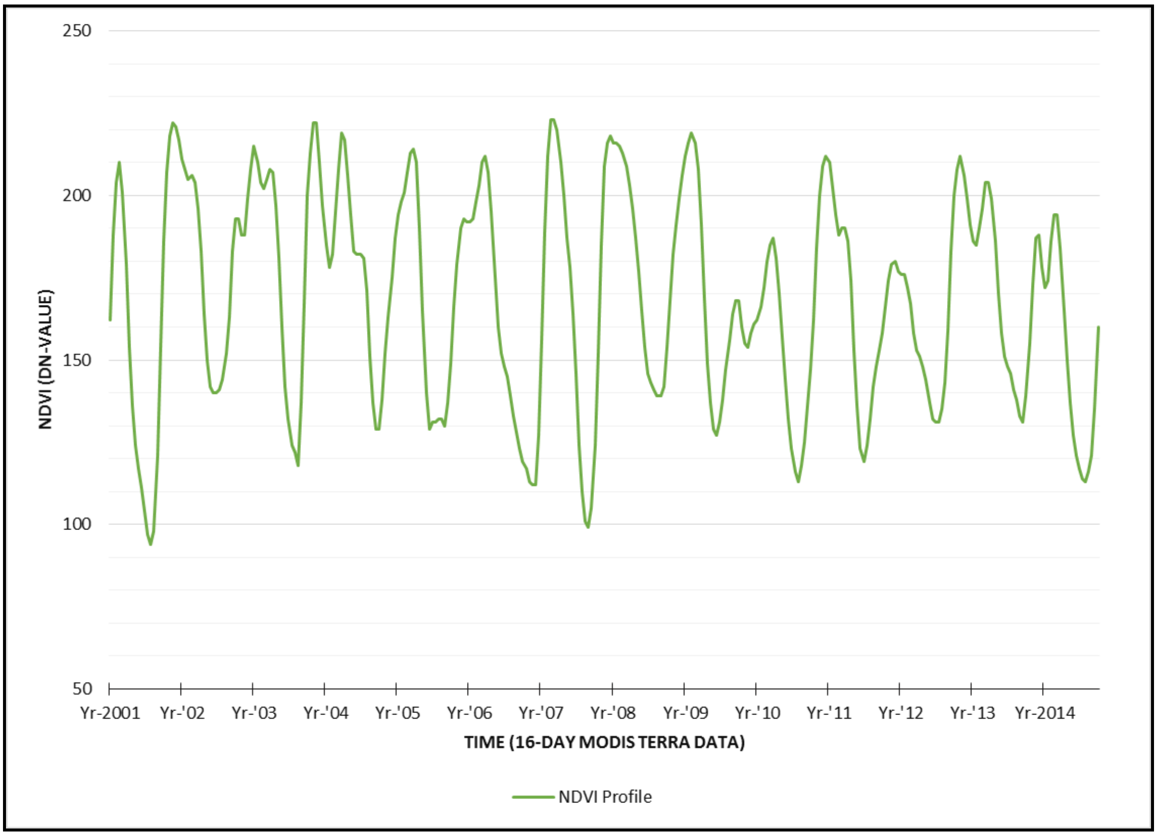


Figure 1: A profile of NDVI over the past 14 years showing irregular peaks in every growing season

1.3. Study Area Description

Masai Mara Ecosystem is located to the south-west of Kenya, between 34^{0} 45' E to 36^{0} 00'E and 0^{0} 45'S to 2^{0} 00'S and it covers an area of about $6500km^{2}$. The study area is roughly triangular in shape, see Figure 2 and can be divided into three range units based on their biogeography and climate (Stelfox et al., 1986). These range units are the Mara and the National Reserve (which are composed of mainly Themeda grasslands), Loita plains (composed of dwarf shrub and *Acacia drepanolobium* grassland) and Siana (mainly hills and plains supporting Croton bush and other woody species interspersed with grasslands). In wet season, Loita plains formed the best part of the range for most grazers in the Masai Mara Ecosystem while during the dry season, the Mara unit formed the most part of the range for most grazing animals (Said, Skidmore, Leeuw, & Prins, 2003).

Annual rainfall distribution in this area is bimodal, characterized by two rainy seasons and two dry seasons (Ottichilo, 2000). The wet season occurs in the months of March to May with its peak in April and the main dry season is from mid-June to mid-October (Stelfox et al., 1986). The area receives rainfall ranging between 600mm to 1000mm (Lamprey & Reid, 2004) with the lowest rain received in the eastern side and highest in the western side where climate is influenced by the Lake Victoria weather system. The soil moisture in the ecosystem is sufficient to sustain grass growth during the dry period. These characteristics make the area to be best described to be in eco-climatic zone IV (Lamprey & Reid, 2004), that is, the semi-arid to sub-humid zone.

The vegetation in this area is supported by the soils weathered from 'phonolitic tuffs' derived from volcanic ash with moderately high fertility. The grassland plains of the Mara range dominated by *Themeda triandra* and *Pennisetum* species support majority populations of the grazing wildlife and livestock.

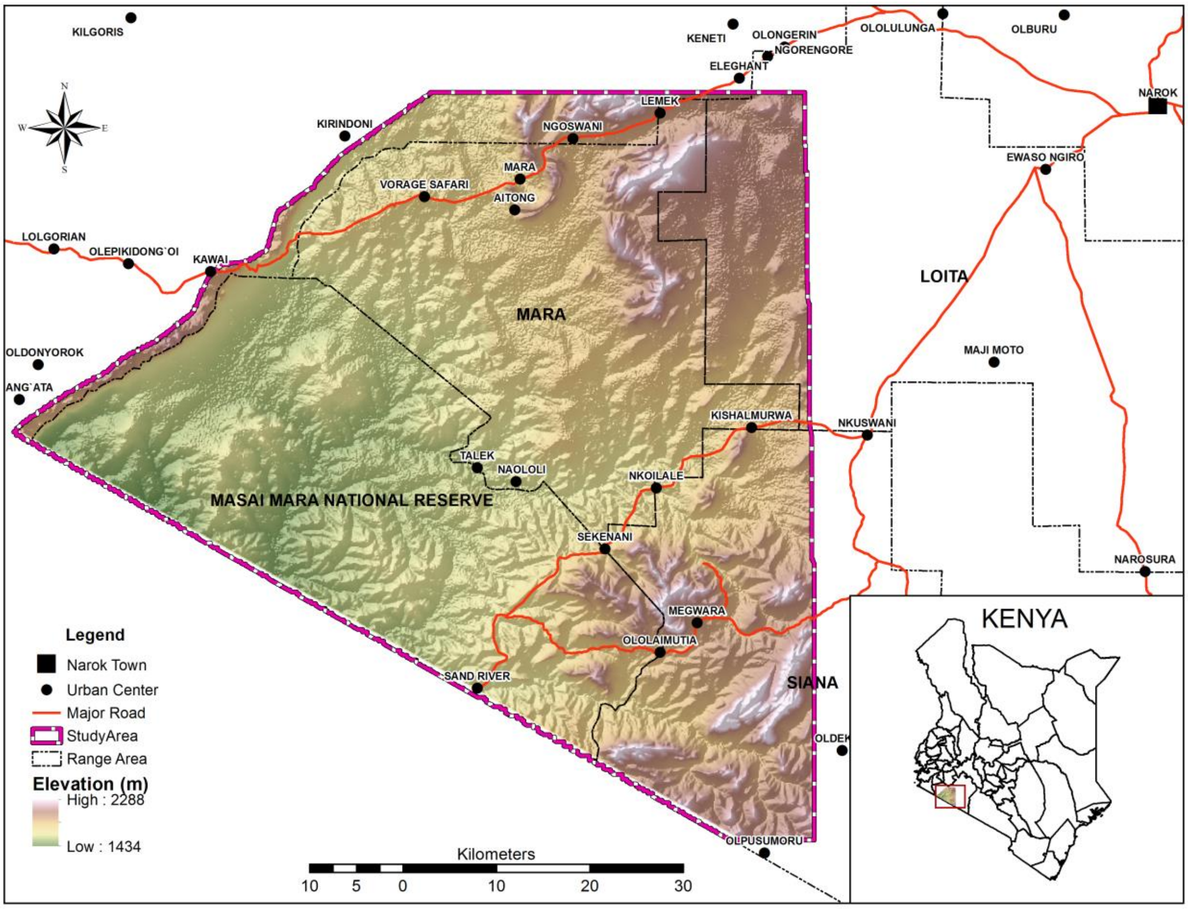


Figure 2: Study area map and its geographical location in Kenya

1.4. Research objectives

The main objective of this study is to use hyper-temporal NDVI data to map grasslands and estimate above ground grass biomass in the Rangelands of Masai Mara ecosystem. In order to realize this objective, the following specific objectives are used;

i. To map grassland cover types of the ecosystem
ii. To determine a model for estimating above ground grass biomass
iii. To evaluate the relationship between grass biomass and Modis NDVI data

1.5. Research questions

This research was designed to help find answers to the following questions

i. Can Modis NDVI data effectively distinguish complexes in Grassland vegetation cover?
ii. What are the differences in actual above ground grass biomass by grassland cover type?
iii. What are the main factors associated with actual measurable grass biomass in Masai Mara Ecosystem?
iv. How do model estimated grass biomass compare to actual historical biomass data?

1.6. Hypothesis

i. Standing biomass quantity by can significantly be explained by NDVI, NDVI derived data, animal density and distance to Bomas.

H_0: $\beta_1 = \beta_2 = \beta_2 \ldots = \beta_n = 0$
H_1: At least one of the β is not 0

ii. Measured grass biomass is not statistically the same in all the vegetation cover types.

H_0: There is no difference in measured grass biomass by vegetation cover type
H_1: There is a difference in measured grass biomass by vegetation cover type

iii. NDVI-derived biomass as predicted for Oct 2006 and biomass as measured by ILRI in Sep/Oct 2006 are identical.

H_0: Model estimated grass biomass and actual biomass measurements by ILRI are not identical
H_1: Model estimated grass biomass and actual biomass measurement by ILRI are identical

1.7. Assumptions

i. Grass cover types did not change over time
ii. Relative wildlife abundance and distribution did not change over the 15 year study period

1.8. Conceptual Diagram

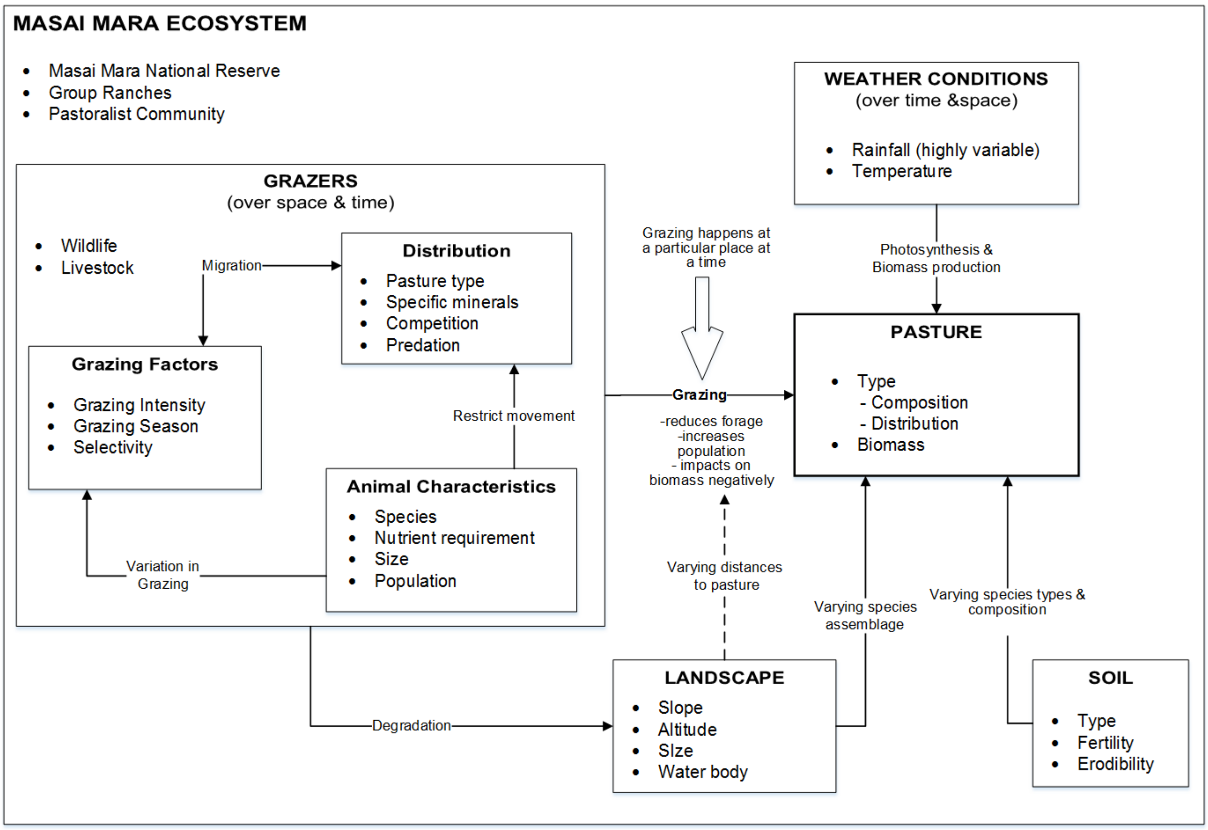


Figure 3: A conceptual diagram illustrating interactions between Grazers, Pasture and influencing factors

The schematic representation of the concept running through in this study is as shown in Figure 3 above. The interactions in the ecosystem is conceptualized as those around primary producers which in this case is pasture. Pasture is considered in terms of type and biomass. Type refers to grasses composition and their distribution across the ecosystem. Grasses types can be annual or perennial. In this concept, biomass refers to the total weight possessed by grasses in a given land cover area and is transferable from one trophic level to another through a natural process; grazing. Different landscape conditions, soil type, fertility and weather conditions affect or cause variability in the amount of pasture available for the grazing animals at any point in space and time. Grazers modify landscape to an extent thus affecting pasture conditions in space and time depending on their population, distribution and their grazing behaviour.

2. LITERATURE REVIEW

2.1. Introduction

This chapter reviews various scientific materials; published and non-published, gleaned and found to be relevant to this study. It gives a critical discussion of what have been done by other researchers showing different arguments, theories and approaches used by the researchers looking into issues regarding vegetation mapping, grassland biomass estimation and monitoring through remote sensing data.

2.2. Grassland Mapping

Grasslands are extensive parts of rangelands and in Eastern Africa, they are all found in the tropics. Most of the grasslands are in Arid and semi-arid areas and the vegetation are tolerant to the semi-desert environment (Grasslands of the world). The grasslands have been grazed over the past many years by wildlife and livestock and due to increase in human population and activities, the grasslands get encroached more so as to meet the ever increasing human demands. Rangelands are degraded by human activities and also selective removal and trampling by grazing animals thus changing grass species found in an ecosystem (Peterson, Price, & Martinko, 1998) and this happens when there is high grazing intensity.

Mapping and monitoring is an important part of any process required for maintaining rangelands. Remote sensing images have proved to be very important recently in vegetation mapping activities (Xie, Sha, & Yu, 2008). The process of getting information about vegetation types and species by interpreting satellite imagery is referred to as vegetation mapping from remote sensing according to Xie et al. (2008). Mapping of vegetation through remote sensing is argued by Sha et al. (2008) that it is not easy and is a big challenge to get satisfactory classification with fine biotic details of vegetation from low and medium resolution imagery like those from Landsat TM and Modis sensors. However, Xie et al. (2008) highlights that Modis as a low resolution data source is applicable for mapping vegetation at a large scale making mapping at global, continental or national scales possible. They further say that the repeat time of Modis Terra satellite and Aqua of one to two days is important for mapping vegetation over time, which is a monitoring aspect. Monitoring needs data measurements about vegetation taken in a sequence over a time interval thus referred to as time series data. Time series data from Modis is reliable and have similar spectral matching techniques to hyperspectral data and the techniques have similar potentials in applications for identifying land cover classes from historical image data (Gumma et al., 2014).

It is important to note that researchers have examined the use of maps in land cover change and monitoring. Sequential production of reliable vegetation maps depicting change or trend over time have been explored using various techniques. Mapping techniques have varied from one researcher to another. De Bie et al. (2008) in their analysis of hyper-temporal images for crop mapping argue that many researchers have done land cover mapping by interpreting single time frame multi-spectral images leaving out of the map the aspect of high temporal variation, a major characteristic of vegetation. To effectively map and monitor grasslands, it is important to define map units of interests depending on the behaviour of vegetation in time as can be measured by the satellite sensors de Bie et al. (2008). Different vegetation types and species show different spectral profiles that are distinct and useful for discrimination and mapping. In a study carried out by Jakubauskas et al. (2002) found out that grasslands show a unimodal phonological pattern same as corn. They further explain that similar species of vegetation (crops) tend to have similar temporal profiles in their phenology making them easy to stratify as a single unit.

Zhang et al. (2003) in monitoring vegetation phenology explains that MODIS-based vegetation growth, production and seasonal variation estimates show spatio-temporal patterns that are related to land cover types. They further argue that in order to maintain or increase the percentage cover of grass species, minimize the percentage cover of invasive species, maintain structural diversity of native ecosystems and to improve their composition, hyper-temporal measurements of vegetation phenology is very important.

Remote sensing has been used to map and discriminate these areas. However, there is no or limited, if any, studies that explains properly how to map vegetation in areas with highly variable rainfall patterns using remote sensing data. Hyper-temporal remote sensing add time stamp to spatial variability of vegetation measured but high variability in rainfall many at times make temporal profiles of similar vegetation differ from time to time causing a mix-up of vegetation types especially when mapping using multi-temporal images as de Bie et al. (2008) explains.

2.3. Hyper-temporal image analysis techniques

Hyper-temporal remote sensing data is one of the primary data sources used in many GIS related studies today. It consist of same image taken at regular time intervals to help study highly dynamic phenomenon (Boyd & Danson, 2006), that is, they use many different time periods of the same image. Therefore, various approaches of extracting information from the time series NDVI data have been applied by various researchers.

Jakubauskas et al. (2002) explains application of the harmonic (Fourier) analysis technique as a method that reduces a complex raw curve of a time-dependent periodic phenomenon into sinusoidal waves, each wave defined by unique amplitude and phase values referred to as harmonics. As an improvement to the technique proposed by Jakubauskas et al. (2001), they illustrated harmonics as a technique that represent a periodic, repeating pattern of a phenomenon with a unique set of height, wavelength, and phase angle. Since the amplitude of a harmonic corresponds to the magnitude of surface greenness (NDVI), the phenological pattern of vegetation over multi-year period can be evaluated as illustrated by Jakubauskas et al. (2002). Each harmonic designates the number of cycles completed by a wave form over the defined period which vegetation is being assessed. Using variance of a harmonic as determined by the magnitude of its amplitude Jakubauskas & Legates, (2000) were able to characterize overall vegetation greenness by amplitude and phase values derived from NDVI biweekly composites. They found out using this method that grasslands and shrublands have lower additive term values than drylands and irrigated farmlands.

A stack of time series data have also been analysed through unsupervised classification method where features have been characterized as patterns or points in a *d*-dimension space. Using an unsupervised classification, that is, the Iterative Self-Organizing Data Analysis Technique (ISODATA) of Erdas-Imagine software, the time series image stack have been segmented to produce clusters (Beltran-abaunza, 2009). The clusters are then subjected to subsequent calculations and analysis for both the minimum and average divergence values. This helps in evaluating the separability so as to assist determine the choice of the optimal number of clusters that generalize the time series data (Skidmore et al., 2003; de Bie et al., 2008; Ali et al., 2012 and Jiang et al., 2013). The decision of the optimal number of clusters is reached through visual inspection and every class identified in this kind of analysis is supported by a temporal NDVI profile expressing trend in vegetation.

Since the coarser spatial resolution of Modis images result into mixed pixels that make it difficult to differentiate between vegetation classes, using one image to do cover classification result into loss of information (Lobell & Asner, 2004). Using hyper-temporal image analysis techniques help solve this problem of mixed pixels (Wang, Ge, & Li, 2013). However, the decisions about the trend within a pixel and classification still rely more on the expert knowledge even though separability statistics proves to be a more reliable decision support technique.

2.4. Grass biomass estimation

Grass and grass production is an integral part of rangeland ecosystem. This is because grass is essential in maintaining rangelands and its services (Jin et al., 2014). Vegetation biomass have been estimated using different methods including, visual (Redjadj et al., 2012; Waite, 1994), harvesting (O. Sala, Deregibus, Schlichter, & Alipe, 1981), capacitance meter (Terry, Hunter, & Swindel, 1981), spectral image data (Tucker, 1979) among others. It is important to know how much the productivity of a rangeland is and quantify the production so as to monitor trends over time. This is because herbage distribution determines the distribution of grazing animals across the ecosystem (Yu et al., 2010). Sala et al. (1988) argues that the spatial pattern of above ground biomass productivity follows a similar gradient to isohyets. This implies that the biomass production is influenced by the amount and distribution of rainfall across an ecosystem which has greater variations from year to year. However, this pattern may not be consistent in ecosystems where rainfall is highly variable. This productivity gradient have been demonstrated using coarse spatial resolution satellite data with high temporal resolution to show seasonal dynamics of vegetation (Justice, Townshend, Holben, & Tucker, 2007).

Biomass estimates is required as part of any grassland monitoring process. Remote sensing presents itself as a tool for grassland monitoring because it provides a timely and synoptic view of grassland conditions. Traditional (Sala et al., 1988) and modern techniques both use statistical approaches to estimate biomass. In recent times, researches have focussed on estimating above ground biomass through correlation and other statistical techniques relating biomass to vegetation indices measurable through remote sensing (Cho et al, 2007, Yu et al., 2010, Jin et al., 2014 and Zhao et al., 2014). Regression analysis is a technique that has been used extensively in these studies to model and predict biomass.

2.5. Hyper-temporal NDVI vs Rainfall measurements for biomass estimation

The relationships between peak biomass and rainfall have been analysed through various studies to prove that rainfall is linearly related to biomass production (Sala et al., 1988, Wylie et al, 1992 and Ran et al, 2006). Sala et al. (1988) argues that the pattern of biomass production in an area is largely accounted for by annual precipitation, accounting for 90% of variations in biomass estimates. In a study that used multi-temporal NDVI data to quantify vegetation change, (Elmore, Mustard, Manning, & Lobell, 2000), discussed NDVI as a simple and reliable measure of greenness that showed correlation with field measurements. However, they further say that, the relationship that existed was less robust especially when measurements were subsequently taken to estimate change. They then attributed this to documented soil brightness and precipitation. Hyper-temporal image data have been used to quantify and map vegetation change because of its strength in repeat cycle (time) of the same image scene (de Bie et al., 2008) and this takes care of rainfall variability as a factor causing non-robustness of NDVI in estimating productivity.

3. MATERIALS AND METHODS

This chapter describes all the materials and methods used to undertake this study. The datasets, tools and techniques used to collect and process the data are explained in various sections of the chapter.

3.1. Data and Materials

3.1.1. Primary Data

The main purpose of primary data in this study was for grassland cover mapping and above ground grass biomass estimation. The primary data were sourced through online download and field sampling. The dataset include, 16-day Modis Terra NDVI and field samples for biomass measurements and percentage cover for vegetation. Table 1 summarizes out all the datasets that were used in this study.

MODIS Terra NDVI Images (2000 – 2014)

Hyper-temporal 16-day Modis Terra NDVI was the primary satellite data used in this study. This data was downloaded from an online source, Reverb, operated by NASA and was used as the basis for grassland mapping and as an indicator of grass biomass. In every plant's growth, phenology changes over time depending on species and growth conditions thus chlorophyll variability in different tissues lead to fluctuations in biomass (Tucker, 1979). These changes are detectable from the NDVI measured by satellite sensors. Since grass biomass vary in space and time, it was important to study vegetation phenology using a reliable and repeatable technique that will provide accurate and timely information on spatio-temporal coverage. This therefore required that a hyper-temporal data, a 16-day Modis Terra NDVI data be used in this study for grassland mapping and biomass estimation. The 16-day Modis Terra NDVI is provided as images averaging to 23 images per year and translating to a large image stack of 15 year NDVI data used in this study.

Grass Biomass Measurements

Spatial distribution of grasslands and grass species relates to the amount of biomass data that can be measured at any point in time (Jin et al., 2014). In this study, primary biomass data was collected in a fieldwork exercise where for above ground grass biomass, grass was clipped through quadrat method and weighed. Fresh and dry weights measurements were recorded in grams for every quadrat clipped. All the grass biomass samples were referenced to Lat/Long GPS point for further spatial analysis.

Vegetation Percentage Cover Estimates

Vegetation cover data such as plant species, percentage cover and height were collected as important characteristics of every sample site. In this study, vegetation cover was considered as any green vegetated area which can be monitored through a sensor viewing from any direction. The vegetation species on the other hand were referred to as plants of a certain scientific assemblage observed at ever sample point.

3.1.2. Secondary Data

Historical Biomass Data

Secondary biomass data is a historical biomass measurement collected from the same study area in September to October 2006. The data was obtained from International Livestock Research Institute (ILRI) and their sampling method was based on a 0.25 by 0.5 m quadrats along a systematically placed line transect see distribution in Figure 4. Biomass measurements of eight quadrats for every sampling transect were summed up to get biomass in grams per square meter. The data was provided in excel spreadsheet containing variables; spatial XY, dry biomass weight in grams and quadrat information.

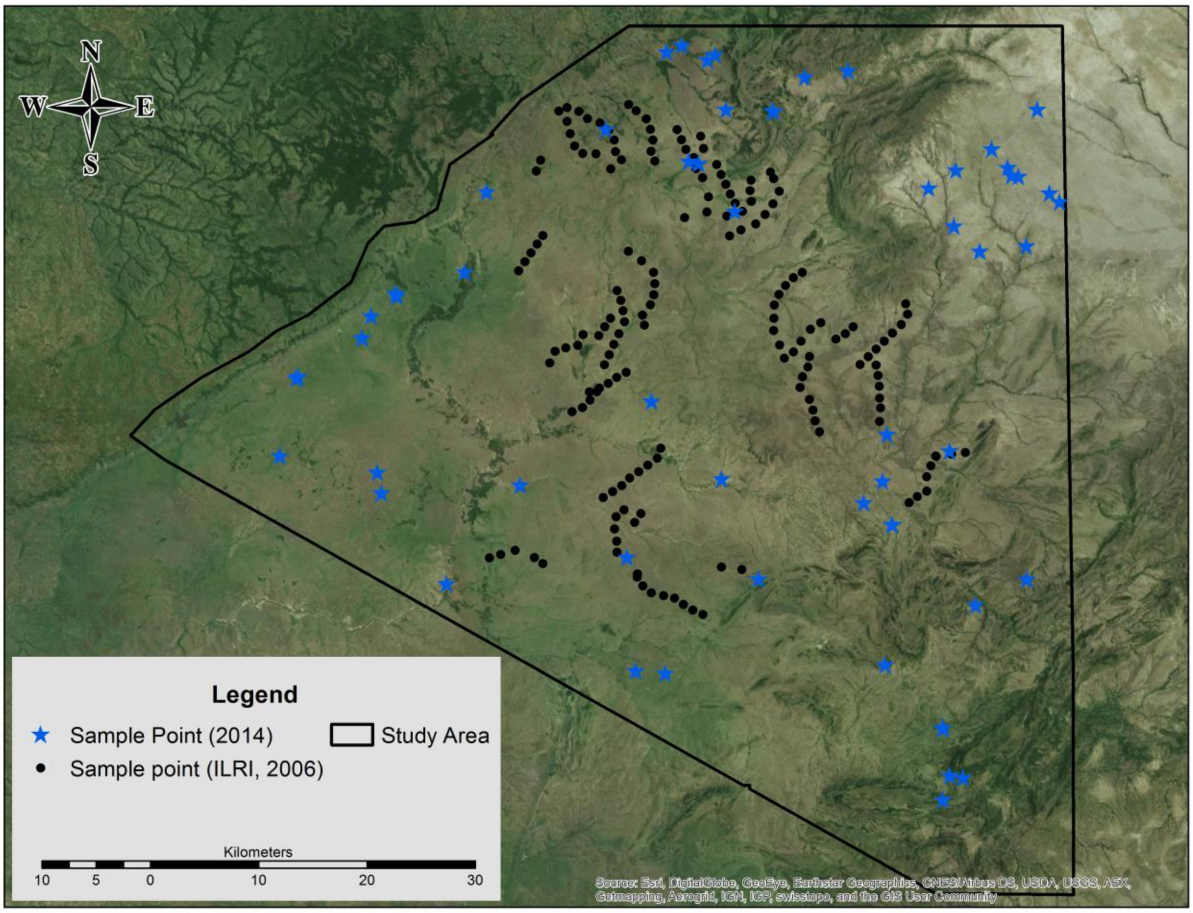


Figure 4: Historical samples distribution in the Ecosystem vs 2014 sample scheme

Animal Census Data

Animal census data for the entire ecosystem was collected as a secondary data from ILRI Kenya. The data is dated as at the year 2010. This animal count data was used to relate abundances and densities of wildlife and livestock to grass biomass. The section of interest in this dataset was on the grazing population of wildlife and livestock found in the ecosystem. The data was provided in a 5km by 5km grid of Esri shapefile with attributes required.

Bomas Data

Bomas are temporarily fenced areas put up by pastoralists and consist of around fence of thorns where the Masai corral their livestock at night. This data was obtained through google earth image and used in this study.

Other Datasets

Other datasets used in this study include Digital Elevation Model (DEM), slope, roads, farms and conservancy boundaries. Roads and farm boundaries were digitized from google earth images. DEM was downloaded from SRTM online source and used as altitude data and for deriving slope.

Table 1: A table listing and describing data used

	DATA	FORMAT	DESCRIPTION	YEAR COLLECTED	SOURCE
1	Modis Terra NDVI (2000 - 2014)	Raster (tiff or .img)	NDVI image data used as the primary data	2014	NASA through Reverb
2	Grass Biomass Measurements	point shapefile, table	biomass at sample locations measured and recorded in grams	2014	Fieldwork
3	Historical Biomass Measurements	Table	Contains biomass measured using quadrats method	Sept-Oct 2006	ILRI
4	Vegetation Samples	Table	Collected during fieldwork and used in mapping	2014	Fieldwork
5	Animal Census	Vector (.shp) or dbf	Animal counts useful in calculating density per vegetation class	2010	DRSRS – obtained through ILRI
6	Bomas	KML/Shapefile	Point data file collected from Google earth by digitizing	2014	Google Earth
7	DEM	Raster (tiff or .img)	Useful in explaining homogeneity in vegetation strata for sampling	NA	SRTM online source
8	Roads and Farms	Vector (.shp)	Used in designing sample scheme and field navigation plan	NA	Digitized google earth images

3.1.3. Materials

Materials used include field equipment, computer hardware and software. Field equipment were, maps, 1x1m metal quadrat, shears, clippers, sample bags, labels, digital photo camera, iPAQ, Garmin GPS and weighing scale. Some of the equipment were borrowed from ITC while others were acquired in Kenya.
Computer software used include ArcGIS version 10.2, ERDAS-Imagine, ENVI (Modified ASAVGOL), Ms Excel, R, SPSS, Ms Word, Ms Visio, Mendeley for citation and referencing and Ms PowerPoint.

3.2. Methods

This section of the report describes the rationale for the application of different techniques used to conduct field sampling, identify mapping units, select and analyse satellite images and secondary data so as to realize the objectives of this research. Key steps followed in this study are summarized as shown in Figure 5 below. Details of each step are explained further in various sub-sections in this chapter and sub-flowcharts provided where possible.

As MaMaSe Research team, all work under sub-sections 3.2.1 and 3.2.2 were carried out together with each researcher taking lead in areas most relevant and specific to their individual topic.

General Flowchart of Methods

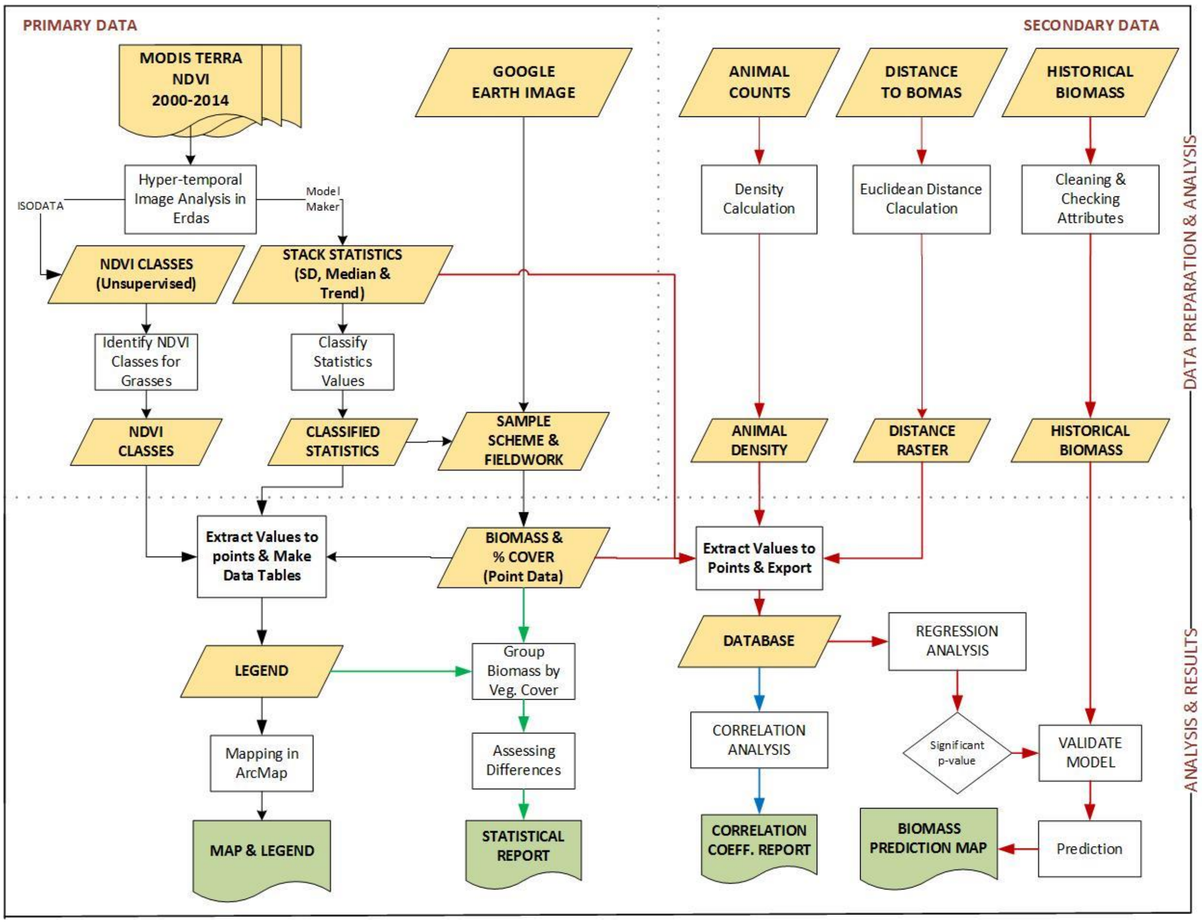


Figure 5: A Flow chart illustrating schematically the general methods followed in the study

3.2.1. Analysis of Modis Terra NDVI Data

3.2.1.1. Noise removal

Normalized Difference Vegetation Index (NDVI) collected for Moderate Resolution Imaging Spectroradiometer (MODIS) is based on the relative values in the Red (R) and near infrared (NIR) wavelengths and it is correlated to vegetation greenness and biomass production (Yeganeh et al., 2012). The formula for NDVI is (NIR - Red)/ (NIR + Red). Downloaded MODIS NDVI data come with cloud or noise that should be removed before any meaningful analysis can be done to the data. De-clouding and outlier removal was done using a filter.

Filtering is a process that was run on the hyper-temporal NDVI stack using modified Adaptive Savitzky-Golay filter (ASAVGOL). Modified ASAVGOL software was used to reduce noise in the data by forcing an upper envelope in the stacked hyper-temporal NDVI data. Figure 6 illustrates through two curves of NDVI information of a pixel of noisy data and a smooth curve after filtering process. This filtering technique uses a simplified least squares procedure to smoothen and differentiate the NDVI data and during the fitting process, it allows iterations to the data (Beltran-abaunza, 2009).

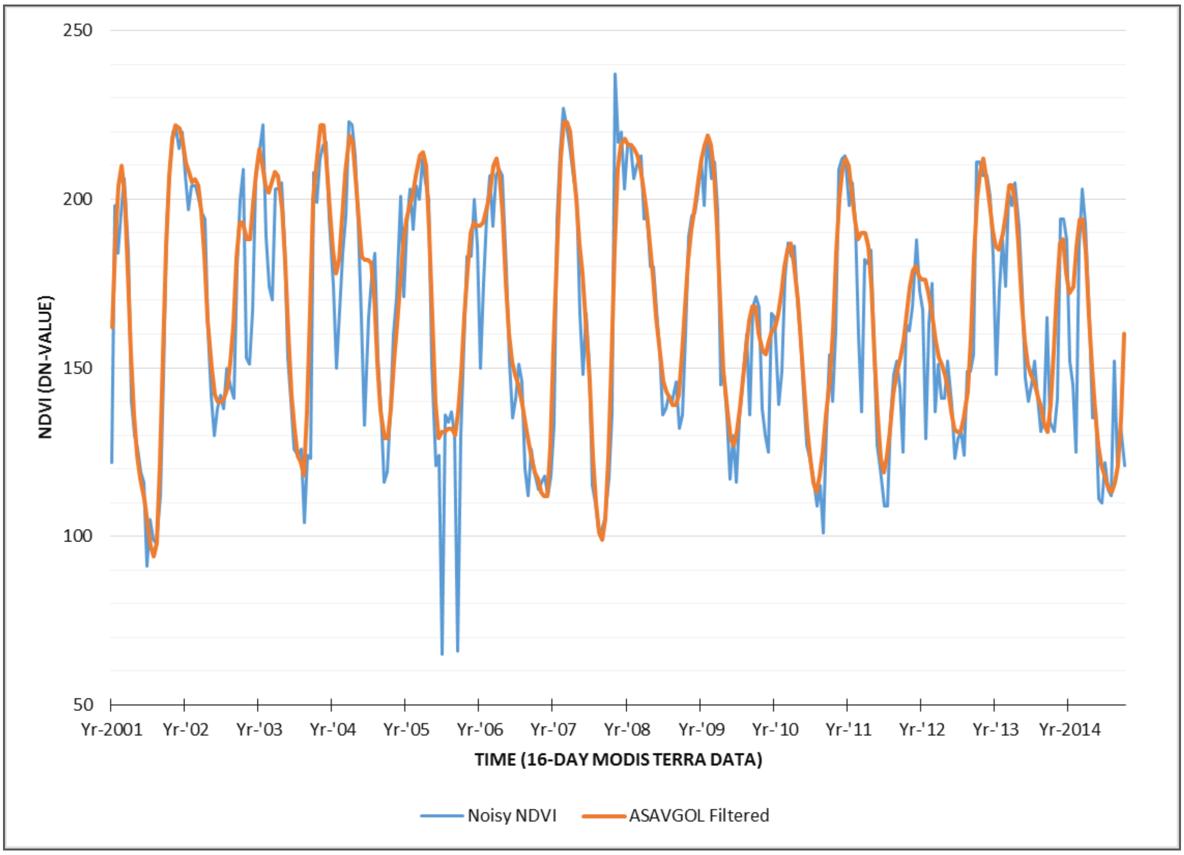


Figure 6: Curves of filtered NDVI time series together with Noisy NDVI; plots of a pixel location.

3.2.1.2. Unsupervised Classification

Unsupervised classification of all the stacked NDVI data layers was done using ISODATA clustering algorithm of Erdas-Imagine software to generate a map with pre-defined number of classes. ISODATA forms clusters by using minimum spectral distance formula to the data (Arai, 2007). The method was used in this study in a way similar to that by de Bie et al. (2008) where the maximum number of iterations used during clustering was 50, rule of thumb requires that this be half number of classes and convergence threshold was set to 1.0 so that the classification do not stop earlier than 50 iterations. The iterations were performed across the entire classification and by initializing means along diagonal axis, the algorithm was set to generate, for a start 10 classes creating a map and a signature file. This procedure was then repeated for up to 200 classes, though it took a long time doing in batch, results were achieved.

Once the data had been clustered, the next step was to identify the 'best' map with significant number of classes where the number of classes are kept low and same time avoid losing important information. By using divergence separability statistics, the 'best' map was selected from 191 maps that had classes ranging from 10 to 200 classes. In order to compare separability between classes, divergence statistical measure of distance was used. Through separability divergence (class separability) of all the generated cluster signatures, minimum and average separability data were produced. This process was done on cluster signatures for 10 class map and repeating it one-by-one to the last map and putting all the generated data-pairs in excel spreadsheet. From this data in excel, a graph of minimum and average separability (Figure 7) was made and the peak in average and minimum divergence indicated the number of classes for the specific place. Based on these peaks of minimum and average divergence, the most adequate number of classes for further analysis of NDVI data was picked as 71 classes.

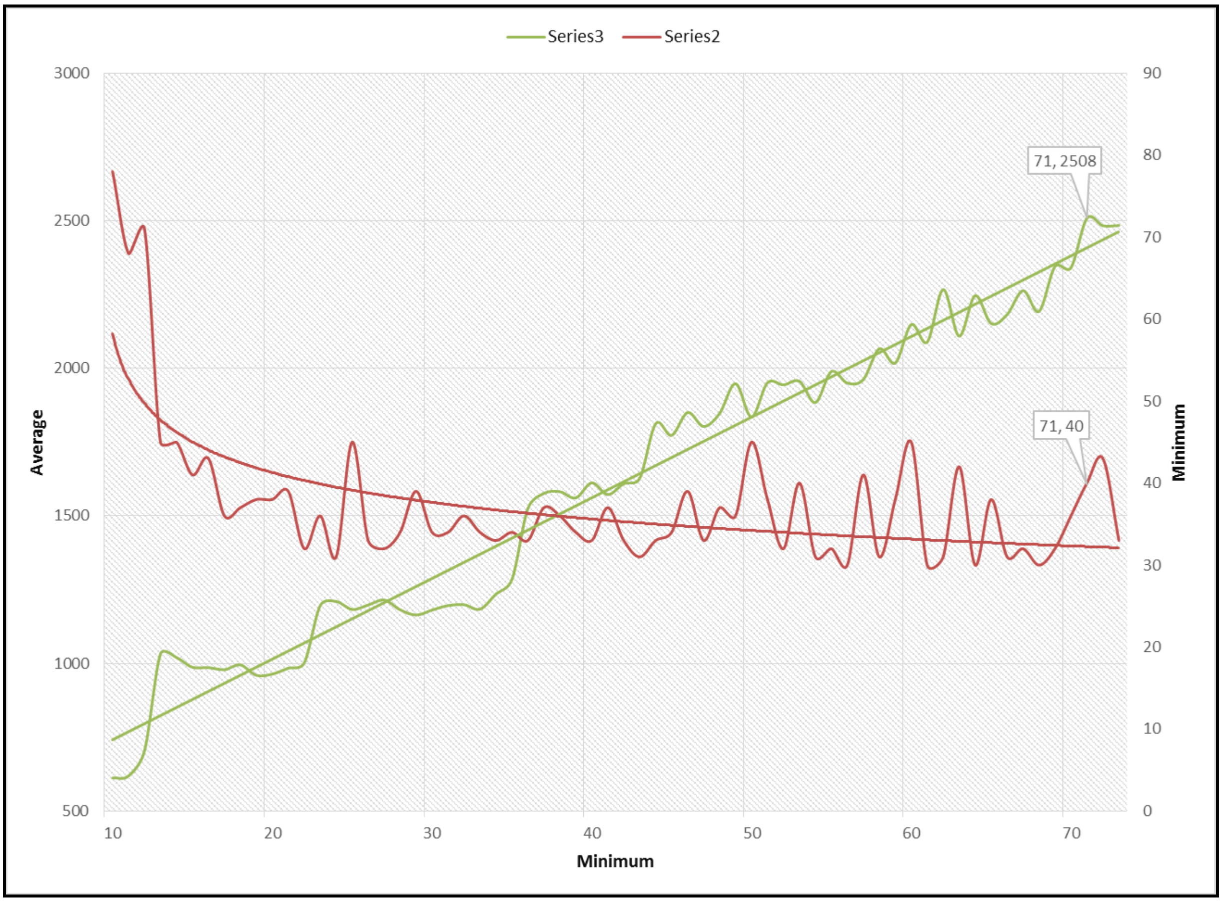


Figure 7: Curves of separability statistics; series 2 is the minimum and series 3 is the average divergence

3.2.1.3. Modis Data to NDVI Statistics Analysis

Stack NDVI data was analysed statistically through 'Stack Statistics tool' in Erdas Imagine software. The software embed stack statistics tools in Model Maker Functions which were used to make a model that calculated SD and Median from the time series stack NDVI data. Statistics maps calculated were classified using different classification methods in ArcGIS and by examining their histogram distributions and break values, meaningful number of classes was reached. Through verification and checking on google earth images, the maps were re-classified further to make classes of interest. The classification procedure for each of the statistics images is as follows;

i) Standard Deviation of NDVI (SD)

Natural breaks method was used to classify Standard deviation map into 8 classes (Figure 8b) in ArcGIS. This method is based on natural groupings of the data values. The method was used to identify class breaks that best group the standard deviation data with similar values and at the same time maximized the differences between classes. The classes were further related visually to google earth images whereby classes 1 and 2 of SD related to woody vegetation, 3 to degraded rangeland or grassland, 4 and 5 to a good condition rangeland or grassland and classes 6, 7 and 8 related to agricultural fields or bare land. Therefore, the data was reclassified into five distinguishable classes 2, 3, 4, 5 and 6, shown in Figure 8(b). Standard deviation map represents variation in vegetation phenology over a longer period of time and those areas with similar long term pattern are considered to belong to same vegetation cover type.

(a) (b)

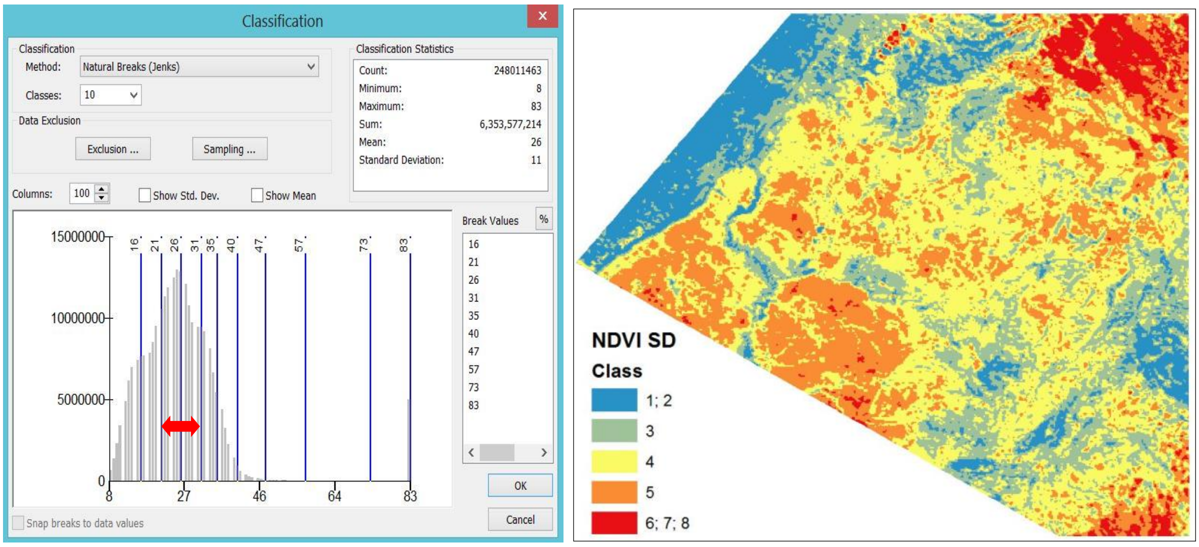


Figure 8: Classification of Standard deviation values; (a) grouping through histogram (b) classified SD

ii) Median of NDVI

A classification process known as Standard deviation in ArcGIS was applied on Median NDVI values as shown in Figure 9(a). ArcMap calculates class breaks with equal value ranges that are proportional to standard deviation which in this analysis, interval size of ¼ was specified as a fraction of standard deviations. In reference to "ArcGIS Help 10.1", smaller fractions of standard deviation generates more classes leading to classification of the data into 20 classes of which classes 12 to 17 were found to be the most relevant to this study. Classes 11 and lower relates to agriculture land and classes 18 and higher related to mixes of trees/woody vegetation/perennial crops. The data was therefore reclassified to nine classes from 10 to 18 as shown in Figure 9(b).

This classification process was useful to aid distinguish different vegetation classes depending on different combinations with classified NDVI standard deviation and Median.

(a) (b)

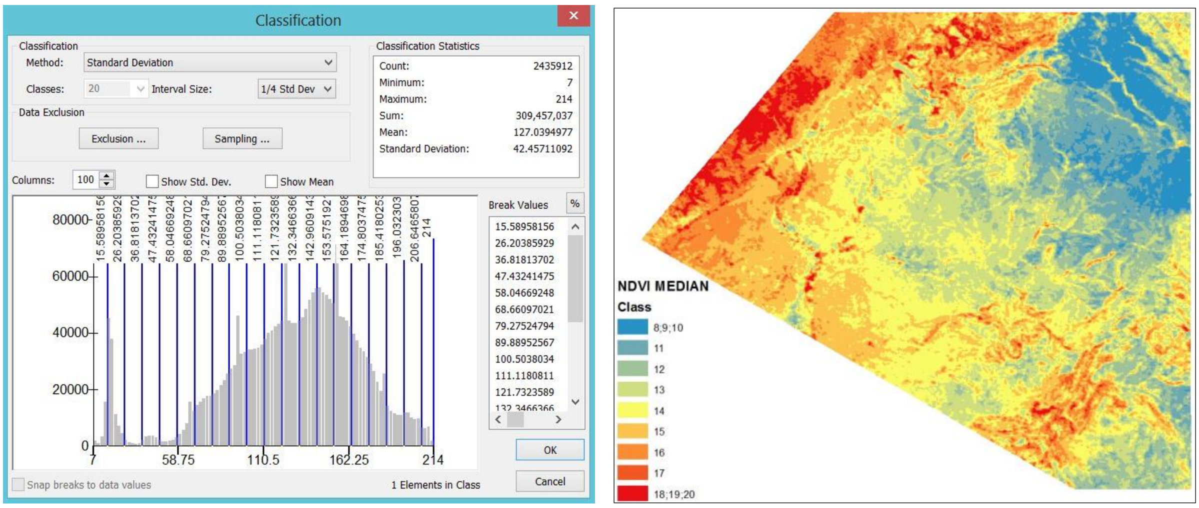


Figure 9: Classification of NDVI Median values; (a) grouping through histogram (b) classified Median

iii) Trend of NDVI

Trend analysis was done in ENVI IDL software where calculations of probability of change for the NDVI time series was possible. Trend break analysis indicate that a trend in time series change between positive to negative values (Schucknecht et al, 2013), where negative values represent pixels that over time are changing negatively; can be related to degradation. Pixels with zero values mean no change. Trend values that are positive values represent positive change and can be related to good condition in vegetation.

Classification of trend data in ArcGIS was similar to that applied to Median data. The classification method, Standard Deviation was used to classify Trend values to 7 classes as shown in Figure 10 where classes 1, 2 and 3 related to degrading rangeland, pasture or land that is being converted to agriculture. Classes 4 and higher related to relatively stable rangelands, pasture or agricultural land. Therefore, the data was further reclassified into 2 classes; 3 and 4, see Figure 10(b).

(a) (b)

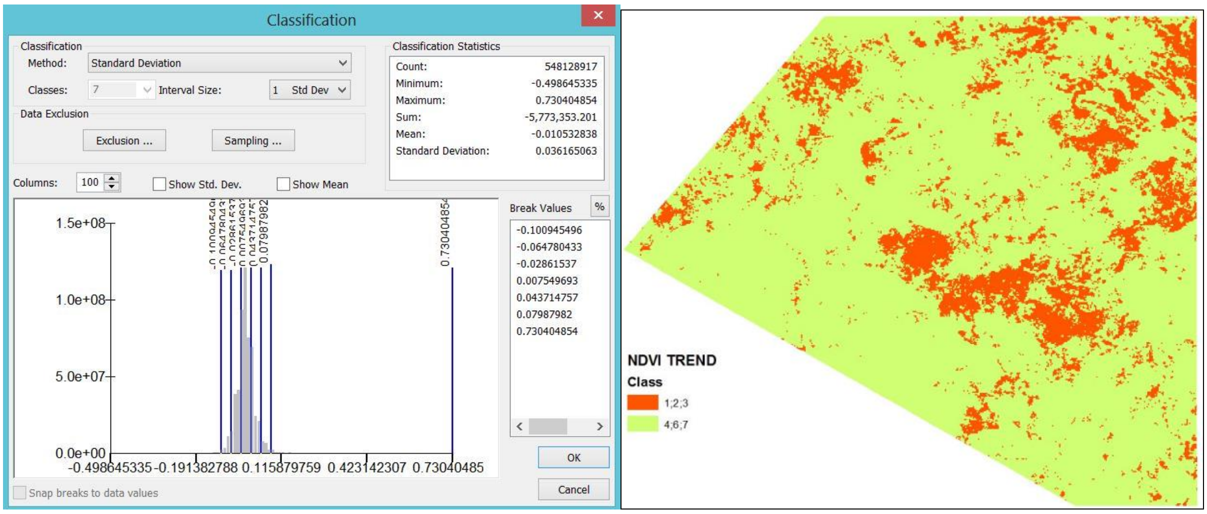


Figure 10: Classification of Trend values; (a) grouping through histogram (b) classified Trend Map

3.2.2. Sample Scheme and Fieldwork

Pre-Fieldwork

In fieldwork preparation, 16-day Modis-Terra NDVI data was downloaded and analysed. A stack of hyper-temporal time series images from Modis Terra was analysed to produce the three statistics; Median, SD and Trend. Classified NDVI statistics maps were used to define the sample units uniquely. In this study, sampling was avoided in settlement areas and through google earth images, farmlands were digitized and used for fieldwork preparation. A buffer of 500m from roads was created and then intersected with a union of the three classified statistical outputs of NDVI. A spatial query was then executed in ArcGIS such that areas that were non-agricultural, and were within 500m from roads were selected to form part of the sample areas. Field maps showing classes of NDVI trends, standard deviation, median and google earth images were prepared and printed for use in the field. Maps were also prepared in digital form and loaded in iPAQ. A shapefile of the study area, roads, towns, sample area, median, SD and Google earth images were transferred into iPAQ to help in navigation during fieldwork.

Stratified Random Sampling

A stratified random sampling was followed in this study based on variability as could be detected through classified long term NDVI data, that is, NDVI standard deviation statistics. The area was first divided into homogenous strata according to the identified NDVI variability after intersection between SD, median and trend that generated about 6000 polygons. Out of these polygons, those that had areas less than 20 hectares were excluded resulting to 1110 polygons remaining. Single Part (polygons that were adjacent to each other and shared same combinations of attributes) procedure was done to merge polygons that were close to each other resulting to 672 polygons. These polygons formed the final strata from which random sampling was done to obtain 50 sample units which were at least 20 hectares in size, were within 500 meters from the road and captured all the variability detectable through NDVI statistics. These formed the final sample units considered (Figure 11) within which sampling was done in the field.

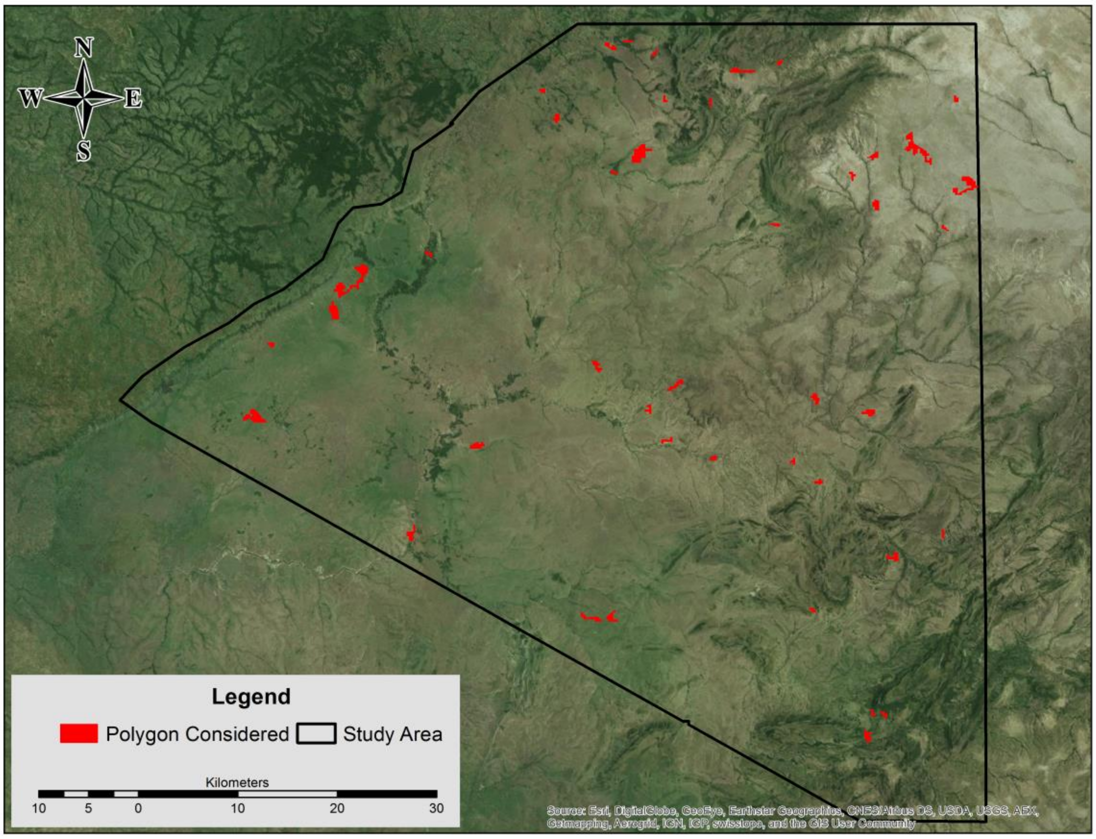


Figure 11: Considered sample scheme for 2014 survey

Vegetation Percent Cover Sampling

Vegetation sampling for composition, percentage cover, canopy height estimation and other vegetation characteristics was carried out during fieldwork exercise. Vegetation cover estimations involved observing and estimating percentage cover for different cover types present in a sample unit including trees, shrubs, herbs and grasses for every single sample unit that selected in the sample scheme. Percentage estimation of plant cover was guided by the scarcity or abundance of all the species and the degree of heterogeneity in their distribution in every sample unit.

In areas where vegetation was homogenous or near-homogenous in the sample units, all the vegetation species found there accounted for near 100% cover estimates. In complex units, each cover type present was estimated by percentage and all the observed cover types accounted for a total of 100% or more depending on the amount of overlaps observed.

In the field, for estimation purposes, a tree was defined as a wooded vegetation, single trunk and at least 1.5m tall. High shrub is a wooded vegetation characterized by several main stems arising from the base and height between 0.5 – 2m. Low shrubs are similar to high shrub in characteristics except that its height range from 0 – 0.5m. Herb are non-wooded, soft stem plants, seed-bearing with canopy rising few centimetres above ground. Grass was defined during percentage cover estimation as herbaceous plant with jointed stems and spiky, pointed, narrow leaves. Percentage cover for each vegetation type was based on team consensus after individual visual and expert judgement.

Grass Biomass Sampling and Drying

During fieldwork, standing biomass samples were collected through quadrat clipping and fresh weights taken at every sample point. At every point, first was to check for grass cover estimates depending on complexes observed (see Figure 12). A decision on whether the grass area was heterogeneous or homogenously covered by grass was made through observations and team consensus. If it was uniformly covered by grass then one quadrat clip was enough but if otherwise, that is a complex of many cover types, then two or more quadrat clips were made.

The clipping technique involved placing of a 1m by 1m quadrat at random in grass areas in every sample unit. All the grasses that were bound by the quadrat were clipped, sorted, that is, dead grass, soil particles and any other litter that might have been collected during clipping were removed before fresh weights of green grass were taken and recorded. After taking the fresh weights of green grass, samples were kept in sample bags and taken for oven drying.

(a) (b)

Figure 12: Complexes in sample units (a) a complex of grass and bare; (b) a complex of grass, bush and trees

Aboveground Grass biomass samples were clipped using shears and other small clippers, see Figure 13(a). Clipping was done to an extent that almost nothing remained within every quadrat. During clipping, it was not possible to achieve 100% clips of grass from every quadrat because of the conditions of grass and nature of complexes at every sample point. As a research team, we did put effort and clipped to at least 95% of all the grasses bound by the one by one meter quadrat.

Sample sorting involved separating brown/dead grass from green, see Figure 13(b). The weights of green grass samples were recorded in grams at a precision of two decimal places. Grass biomass samples were kept in a clearly labelled sample bags then taken to Masai Mara University Laboratory in Narok for oven drying. Drying was done at a temperature of 65^{0} C for five hours. These drying conditions were appropriate according to the university laboratory SOP (K. Rutto, in a discussion) and also these samples were collected at the end of dry season and had stayed for some days in open air before the actual oven drying took place. Dried samples were weighed and readings recorded as were for the case of fresh weights.

(a) (b)

Figure 13: Pictures showing biomass clipping activity (a); and sorted sample (b)

3.2.3. Field Data Pre-processing

Vegetation samples data sheets and biomass measurements records were entered in Ms excel spreadsheet with sample numbers verified and referenced to their corresponding GPS coordinates. A point shapefile was created with the attributes of the dataset appended. The points were overlaid with sample scheme file and google earth images to check that they are proper and corresponds to the expected sample area. Once the coding system had been checked, the table was ready further analysis. For the complex sample units, vegetation percentage cover estimates were processed through weighting averages using the complex proportion estimates for each unit. For instance, a sample unit of two dominant cover types, say 70% grass and 30% bush, had two samples taken which were then weighted or reduced to a proportion represented by the complex percentages which then adds to form a single value for the sample unit. This was done so as to ensure representativeness of the differences as could be seen in the field.

Attributes were added to field data points from the NDVI data. Attributes SD, Median, Trend, Animal density and distance to Bomas were extracted using a tool called 'Extract values to points' in spatial analyst tools of ArcGIS. This process is pixel based depending on the location of the points on a 2-D space. Biomass sample data were also joined to the points to form part of the major database.

3.2.4. Vegetation Cover Mapping

Grassland vegetation cover mapping process involved extraction and classification of vegetation types from a complex of multiple land cover types; grasses, trees, shrubs, herbs, bare soil and others. Modis NDVI images were used to perform vegetation cover mapping by describing NDVI classes from temporal changes of vegetation measurable through remote sensing data. NDVI spectral patterns were established from the stack NDVI to help characterize similar combinations of NDVI SD, median and trend that describe different NDVI classes that are unique for every vegetation cover type. This was possible through spectral profiles as shown in Figure 14. Classes of the unsupervised classification, as could be identified from spectral profiles, together with their corresponding SD, Median and Trend values, were used to characterize the differences in percentage cover as observed in the field into distinct NDVI classes.

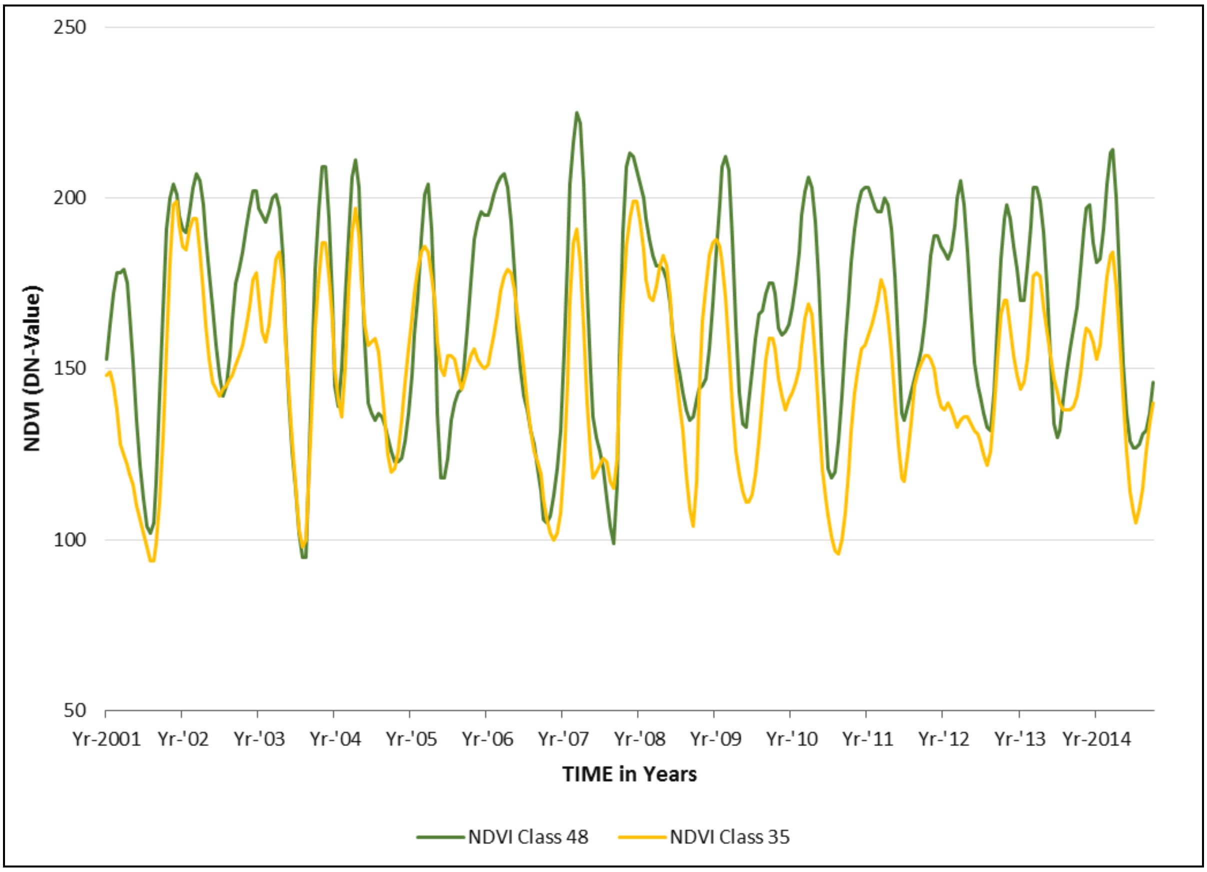


Figure 14: A legend figure of 2 NDVI profiles used to detect differences between NDVI classes

The mapping technique used here explored the relationship between percentage cover estimates and classified Modis NDVI data to distinguish different cover types. The relationship between SD and Median seem to have negative linear relationship whereby areas with high SD corresponded to those with low median values. This is to say that areas with SD class 2, which relates to woody cover, corresponded to high median values-class 17 and 18 compare Figure 8b and Figure 9b. Considering this relationship and the cover percent estimates acquired from the field, unsupervised 71 classes were identified and re-grouped to form distinct vegetation cover types.

The approach considered a database of field sample cover estimates, NDVI classes, SD and Median values which were useful in making 'between-class' differentiation. This vegetation cover mapping method divided grass cover percentage estimates into strata depending on different combinations of NDVI class, SD and Median in relation to other cover estimates that could be identified within NDVI classes. This mapping approach is summarized in a flow chart shown in Figure 15 below.

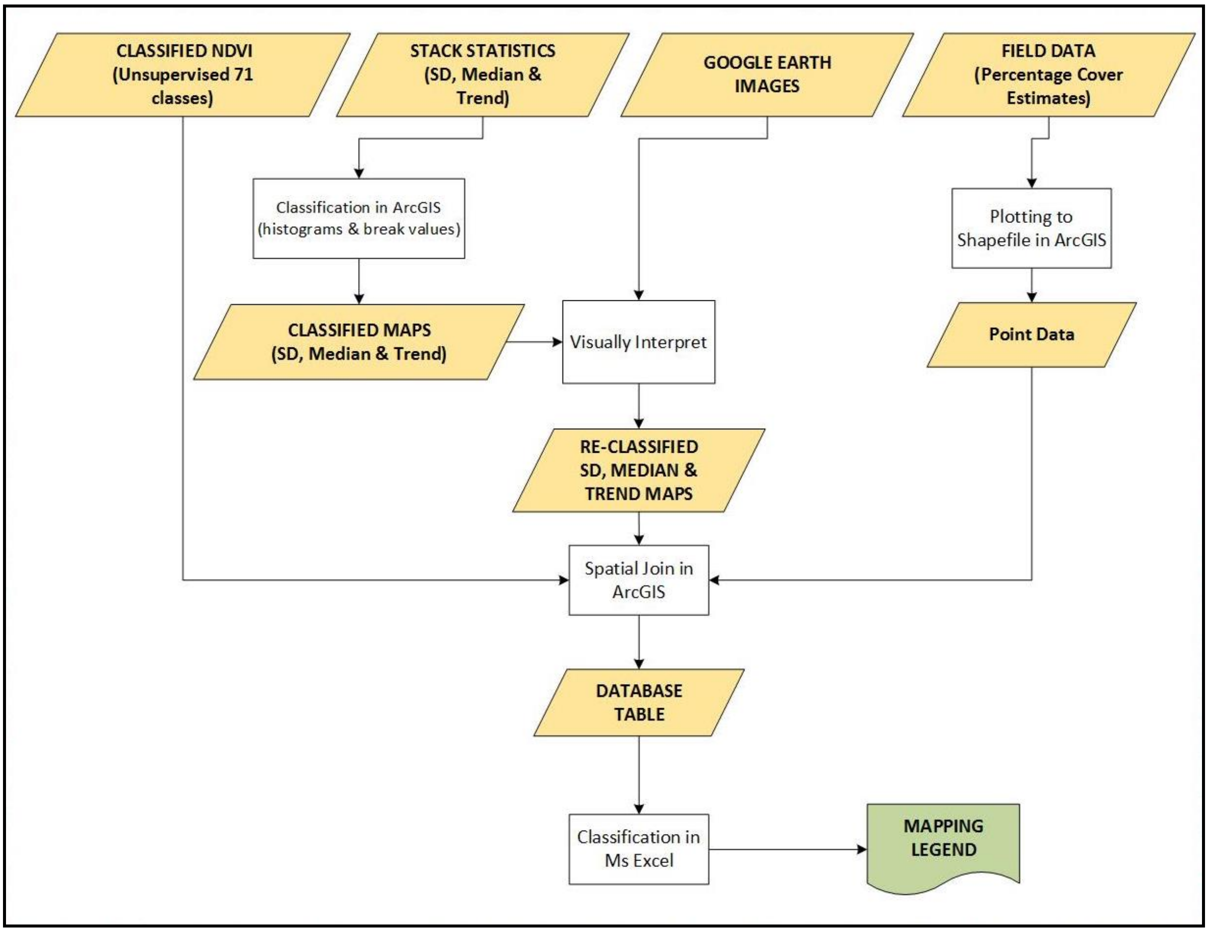


Figure 15: A flowchart of vegetation cover mapping process

3.2.5. Assessing the differences in measured grass biomass by grassland cover type

The measured grass biomass samples were stratified into groups according to the mapped grassland cover units. Since the distributions of biomass for each of the strata were not normal, a non-parametric test was used to compare multiple independent samples of biomass testing the null hypothesis that the average biomass weights in each vegetation cover units were different (Ruxton & Beauchamp, 2008). It was necessary to test if there were variations in the means of grass biomass within each of the mapped vegetation cover types so as to determine forage production dissimilarities in different parts of the ecosystem.

Tukey and Kramer (Nemenyi) test with Tukey-Distance approximation for independent samples post-hoc was done after the Kruskall-Wallis test to determine which group of means of grass biomass measurements had statistically significant differences from the others. Kruskal-Wallis is a non-parametric method preferred for comparisons of more than two independent samples (Delaney & Vargha, 1998).

3.2.6. Statistical Analysis and Model Development

Correlation and Regression analysis

Model development required preparation of dependent (response) and independent (predictor) variables as the first step. Field biomass measurement was used as response variable to a list of predictor variables including; results from the analysis of hyper-temporal images (SD, Median and Trend), NDVI measurements (as at October 2014), animal abundance and distance to Bomas.

A distribution model was developed through statistical approach once the data had been prepared. Regression technique was applied to simulate the relationship between biomass (response) and multiple independent predictor variables. The first step was to do correlation and collinearity analysis so as to detect the relationships between biomass and the multiple predictor variables and also to check if the predictor variables were related. A non-parametric technique called Spearman's Rank-Order correlation was used to test the null hypothesis that the correlation between biomass and the predictor variables was equal to zero against the alternative hypothesis that the correlation is greater than zero. Spearman's Rank-Order was used because the response variable (biomass) was not normally distributed (Puth, Neuhäuser, & Ruxton, 2015). The variables are said to be correlated if the association is linear, i.e., can be represented by a straight line on a scatterplot. The correlation can be positive, zero or negative with values ranging from -1 to +1. Spearman's coefficient of rank correlation, rho, is calculated by applying the following formula (Puth et al., 2015);

$$rho = 1 - \left(\frac{6\sum \quad di^2}{n(n^2-1)}\right)$$

Where d_i is difference in ranks given to the two variable values for each item of the rank; n is the sample size

In order to achieve a good regression model, a collinearity check was first done in the predictor variables using variance inflation factor (VIF) technique so as to determine if two or more variables tell the same information about the measured biomass. As a rule of thumb, VIF values greater or equal to 10 signifies a high collinearity in the predictor variables and such variables are always excluded from the model development.

The multiple independent predictor variables were then regressed on biomass measurements following a backward stepwise procedure where the most parsimonious set of predictors that were most effective in predicting biomass were considered in the final model. All the independent variables were first entered in a simultaneous way into the analysis to develop a model and then by removing independent variables one at a time, considering their significance (p-values) in explaining biomass variation, subsequent models were built until a significant model ($p<0.05$) and best R^2 was reached. The model was established in R environment (a package for statistical analysis).

The multiple linear regression equation fitted through the data was adapted from (Field, 2009). The basic equation used was in the form of model estimated statistical coefficients in a mathematical equation that modelled the response variable from a list of predictor variables. The basic equation;

$\text{Outcome}_i = (\text{Model}_i) + \text{error}_i$

Addition of all the predictor variables give the model: $Y_i = \beta_0 + \beta_1.X_1 + \beta_2.X_2 + \beta_3.X_3 \ldots + \beta_n.X_n + \dot{\varepsilon}_i$

Where Y is the response variable; β_0 is the Line Intercept and $\beta_{1, 2, 3 \ldots n}$ are coefficients of the predictor variables and $X_{1, 2, 3 \ldots n}$ are the predictors and $\dot{\varepsilon}_i$ is the difference between the observed Y value for the i^{th} participant.

Modelling assumptions and Testing

The four principle assumptions of linear regression modelling as explained by Marill, (2004) were considered and tested during the modelling process. The response and predictor variables were assumed to be linearly related with some elements of additivity, i.e., the relationship is a straight line function and the effect of different predictor variables on the expected predicted value of the response variable are additive. The other assumptions were that there is statistical independence of errors, homoscedasticity of errors (variations of data points around the fitted regression line assumed to be constant) and that the errors are normally distributed. Assumption of statistical independence means that as the research continued to remove each of the independent variables during the stepwise procedure, the deviation of each data point from the fitted line had no relationship on to the value of another data point in the dataset (Osborne & Waters, 2002 and Marill, 2004).

In this modelling process, it was inevitable for this study to check whether the parameters of the fitted regression model were distinguishable or undistinguishable from zero. A t test for parameters was done to test for null hypothesis that the parameters of the model are equal to zero or the alternative hypothesis is held. The assumptions were tested for a decision to accept or reject the model generated, i.e., the tests were used here to enable make a decision on the applicability of the model. The assumption of normality of the model residuals was checked using normal QQ plots (distribution is normal if the points on the QQ plot fall close to the diagonal reference line), histogram (should not be too skewed) and Shapiro-Wilks test. A Shapiro-Wilks test value greater that 0.05 was acceptable (Razali & Wah, 2011) for the errors to be assumed to be normally distributed. Linearity of predictor variables to response variable was diagnosed through a plot of residuals versus predicted values of the model. Durbin-Watson statistic and a plot of residual autocorrelations was used to test if there existed violation of independence, i.e., to test for a significant residual autocorrelation where Durbin-Watson statistics close to 2 was acceptable (Akter, 2014) for statistical independence of residuals. Homogeneity of variance of the residuals is one of the main assumptions in the regression analysis. However, heteroscedasticity tests were pointless in this study since it doesn't bias coefficient estimates but only make standard error incorrect. This study cared much about the fit of the model and heteroscedasticity did not matter much.

Model validation

A complete Leave One Out Cross Validation (LOOCV) method was used to calculate RMSE for the model since it was of no statistical significance to split the dataset for validation because the samples were few. A complete approach for LOOCV was done so as to give each data record an equal chance as a testing case. In this technique, regression process is done in iterations equal to number of samples. Cross validation does not use the entire data set when building a model. It leaves out some records before it models the data; removed data also refered to as testing set. The records left are refered to as training set which the technique uses to build the model. The testing set on the other hand are used to test the perfomance of the model built. This method provides unbiased estimation of prediction error for model selection. Accurate model have the lowest RMSE value for prediction. Using DAAG library in R, a function CVlm is called to perform the k-fold cross-validation which generates sum of squares as result. Square root of the overall result is what was then compared to the RSE of the original model developed in the stepwise procedure.

Once the assumptions were tested and decisions reached, the model was then evaluated so as to assess the best fit of the model. The significance of the model was checked before its application in biomass estimation. Significant p-value (<0.05) and good R-squared as well as low residual standard error enabled decision whether to use the model for prediction or not.

3.2.7. Spatial prediction of grass biomass

This analysis was done for the entire ecosystem to explore spatially the distribution of measurable grass biomass. The function generated through regression procedure was used to make spatial predictions for biomass in grams/m^2. The formula was applied on NDVI image, SD, and Distance to Bomas map (significant variables selected through the stepwise procedure) in ArcGIS using Raster Calculator tool to get a prediction map of biomass in grams per square meter.

Using raster calculator tool in ArcGIS, a mathematical calculation was applied to raster files of the predictor variables using model coefficients generated. The output image was then converted to vector for further analysis. The output biomass vector was stratified based on vegetation cover map developed through a process described in section 3.2.4 of this document. The attribute table of the intersection map between Biomass estimate and grassland cover map was exported to Ms Excel for statistical analysis and summaries.

3.2.8. Comparison of predicted biomass estimates to historical measured biomass data

This analysis was done on secondary data that was obtained from ILRI. The sample points for this data were used to extract values of predicted grass biomass (Oct 2006) for purposes of comparison. Wilcoxon signed rank test with continuity correction test is unpaired test that that is very useful to test for statistical mean differences in paired samples (Fay & Proschan, 2010) and it was used to compare mean of the estimated biomass to corresponding means of samples collected on different dates.

The data was paired since it was from the same geographic coordinate or the two samples were picked. The sample from ILRI were paired with extracted values from estimated biomass then the dataset compared through a nonparametric test. Without assuming normality of the data, it was tested at 0.05 significance level to check if estimated biomass for October 2006 and measured biomass for the year 2006 have identical data distributions.

4. RESULTS AND DISCUSSION

4.1. Results

4.1.1. Descriptive Statistics of Field Data

The study designed fieldwork to collect data about above ground grass biomass and vegetation cover. The descriptive statistics were used to summarize field data were minimum, lower and upper quartiles, median, mean and maximum values. Table 2 shows the statistics for biomass weights (g/m^2) as well as for percentage cover of different sampled cover types. There were 50 sample units out of which, biomass samples were possible to collect in 42 units. Data distribution boxplots for the field data are as shown in Figure 16 and their distribution histograms in Appendix 1. Outliers were easily detected using the summaries and graphic plots. By examining the summary table together with boxplots, biomass data appeared to be skewed to the right.

Table 2: A summary table of the field data

Summary	Fresh Biomass	Dry Biomass	Grass Cover	Tree Cover	High Shrub	Low Shrub	Herbs	Bare soil
Minimum	22.87	19.09	0.6	0	0	0	0	0
1st Quartile	68.35	44.58	31.52	0	0	0	0	3.6
Median	134.92	79.52	54.3	0	2.5	0	0	30.8
Mean	149.77	104.32	50.85	1.933	12.18	1.094	0.973	31.4
3rd Quartile	186.91	141.26	68.33	0.85	16.9	0.575	0	58.85
Maximum	386.32	288.24	97.8	48.3	95	19.5	15	73.1

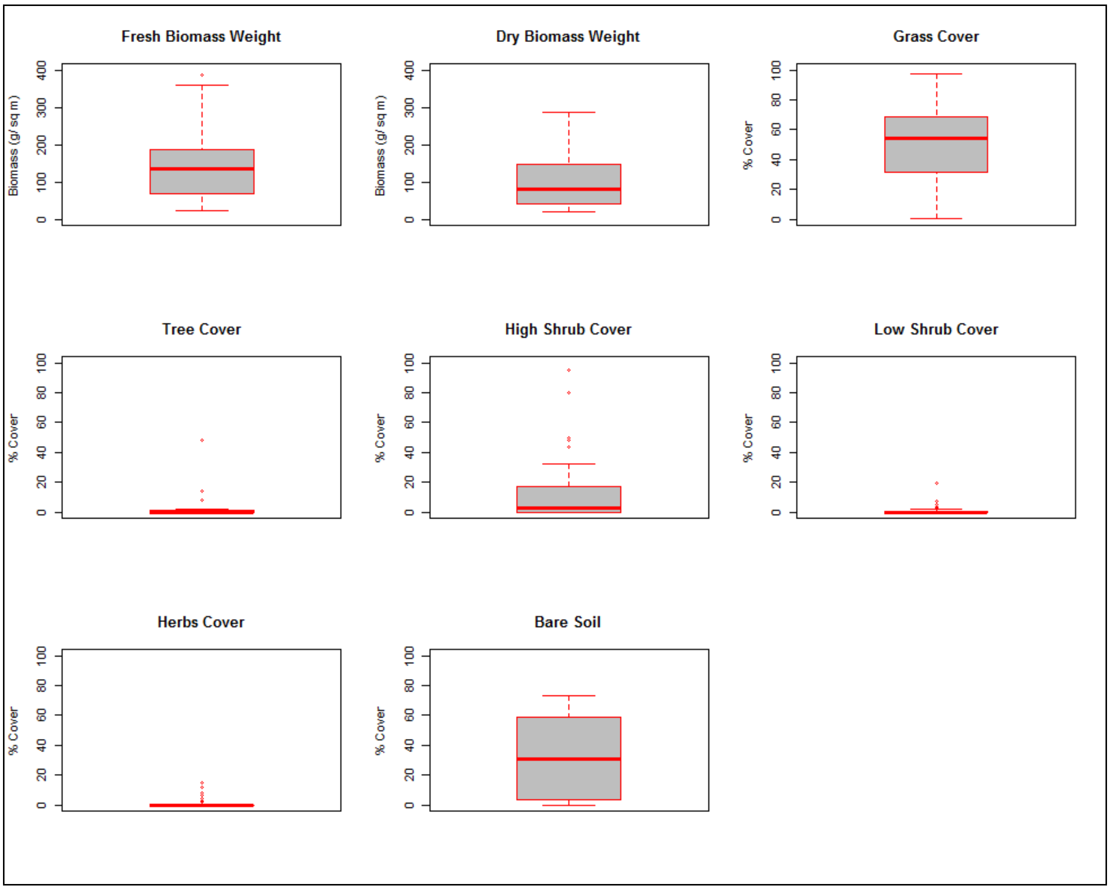


Figure 16: Distribution boxplots for the field data

4.1.2. Grassland classification map and legend

The study collected data and acquired information regarding cover types for the ecosystem (Appendix 1). The information was used to characterize the area into specific classification units of NDVI to make them discernible from remotely sensed Modis NDVI data. Each unit in the map reveal characteristics of a given vegetation type distinguishable through NDVI, Median, SD and Trend and the attributes collected in the field.

The NDVI data enabled the grassland cover types be divided into six distinguishable cover units, each unit uniquely identified by NDVI classes and NDVI derivatives. Percentage cover types as shown in the map legend and corresponding NDVI derived information are shown in Table 3. The vegetation cover classes were identified as units A, B, C, D, E and F as shown in Figure 17. The mapping result also show that cover unit C constitute the largest percentage cover of about 34% followed by unit E of about 26% of the total area. The category 'Others' represent NDVI classes that were not sampled and had unique combinations of Median and SD that were distinct from any of the sampled classes.

Table 3: Mapping units by NDVI-Derived information and percentage cover

					% Cover					
UNIT	NDVI CLASS	MEDIAN	SD	TREND	Tree & High Shrub	Low Shrubs	Grass	Litter	Bare	Stony
A	18	10	3,4,5	3,4	6	2	31	0	67	0
B	25	11,12	2,3,4,5	3,4	7	1	58	0	41	0
C	35	12,13,14	2,3,4,5	3,4	8	1	38	1	59	2
D	39	13,14	3,4,5	3,4	13	1	**51**	0	**47**	0
E	48	14,15,16	3,4,5,6	3,4	1	1	**70**	20	9	0
F	39,52,57	15,16, 17,18	2,3,4	4	**59**	2	**31**	13	49	3
Other	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

The differences in vegetation cover are distinguished through different combinations of NDVI class, Median, SD and Trend such that they show the differences in vegetation types as observed in the field. For instance, Unit B and F in the map indicate that they units are different vegetation types when we visually compare the observed cover percentage. Comparing Median values for different cover units as depicted in the map, it follows that unit B has low median while F has high median. There are overlaps in the combinations of NDVI derived data to define field cover estimates. Remote sensing successfully distinguished vegetation cover in NDVI classes depending on the spectral reflectance of the vegetation. However, the map result doesn't show proper categorization of the observed cover percentage since overlaps exist in reality. Each mapping unit is shown to have at least a dominant cover but through remote sensing data, the differences are not easy to see in Median, SD and Trend combinations. Median to an extent show the differences but still some overlaps could be seen.

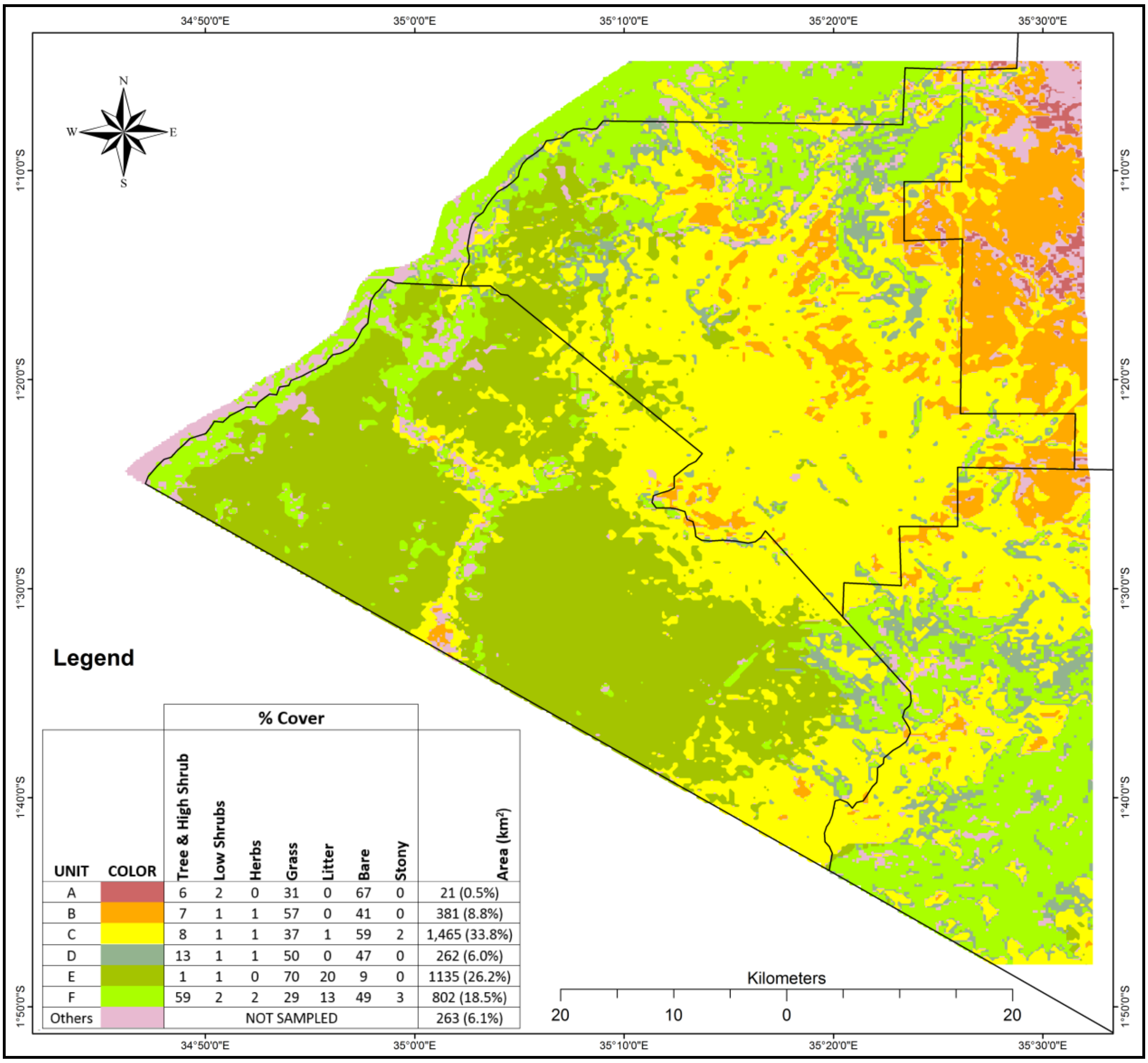


Figure 17: A Vegetation Cover Map of Masai Mara Ecosystem

The units A, B and C (Classes 18, 25 and 35 respectively) are overlapping in cover, making it difficult to distinguish percentage vegetation cover through remote sensing. The study decided to keep them as separate because the classes are distinctively different from remote sensing. The average annual vegetation behaviour in this area is shown as average long term spectral response are shown in Figure 18. The figure shows that the peak of growing season is around May-June and lowest NDVI measurements around Sept-Oct, with curves very distinct for every vegetation unit.

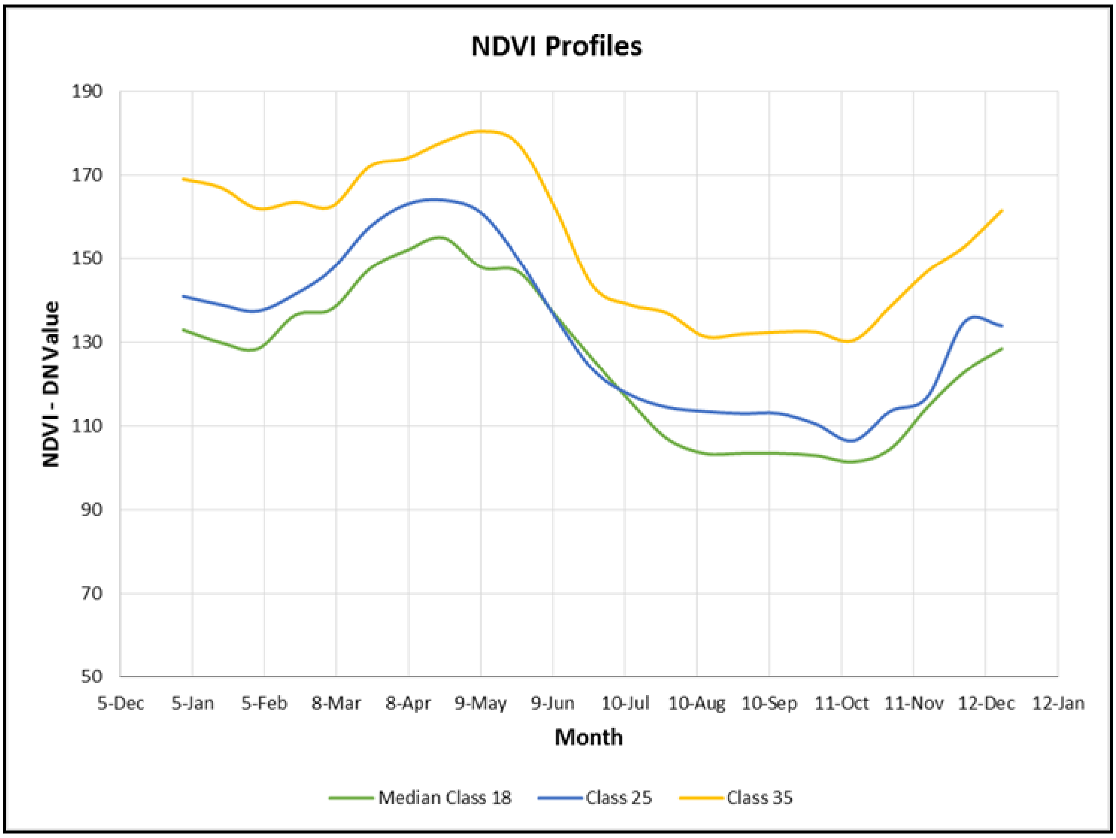


Figure 18: Spectral profiles of Classes 18, 25 and 35; annual monthly averages

4.1.3. Comparison of actual grass biomass by cover type

Measured grass biomass was grouped and compared by mapped vegetation cover units. A comparison of multiple independent samples was carried out using Kruskal-Wallis to test if there were statistically significant differences in biomass samples measured within the mapped cover units A, B, C, D and F.

H_0: $\mu_1=\mu_2=\mu_3=\mu_4=\mu_5=\mu_6$

H_1: at least one of the means is different

The result of the Kruskal-Wallis test statistics is highly significant (Kruskal-Wallis chi-squared = 23.1442, df = 5, p-value = 0.0003168) at a significance level $p \leq 0.05$, leading to a conclusion that the means of biomass measurements collected within each of the six identified cover classes are not equal. However, it is not possible to tell through Kruskal-Wallis test which of the group or groups are different from the others. A distribution boxplot is created as shown in Figure 19 below to create an overview of the data within each of the groups.

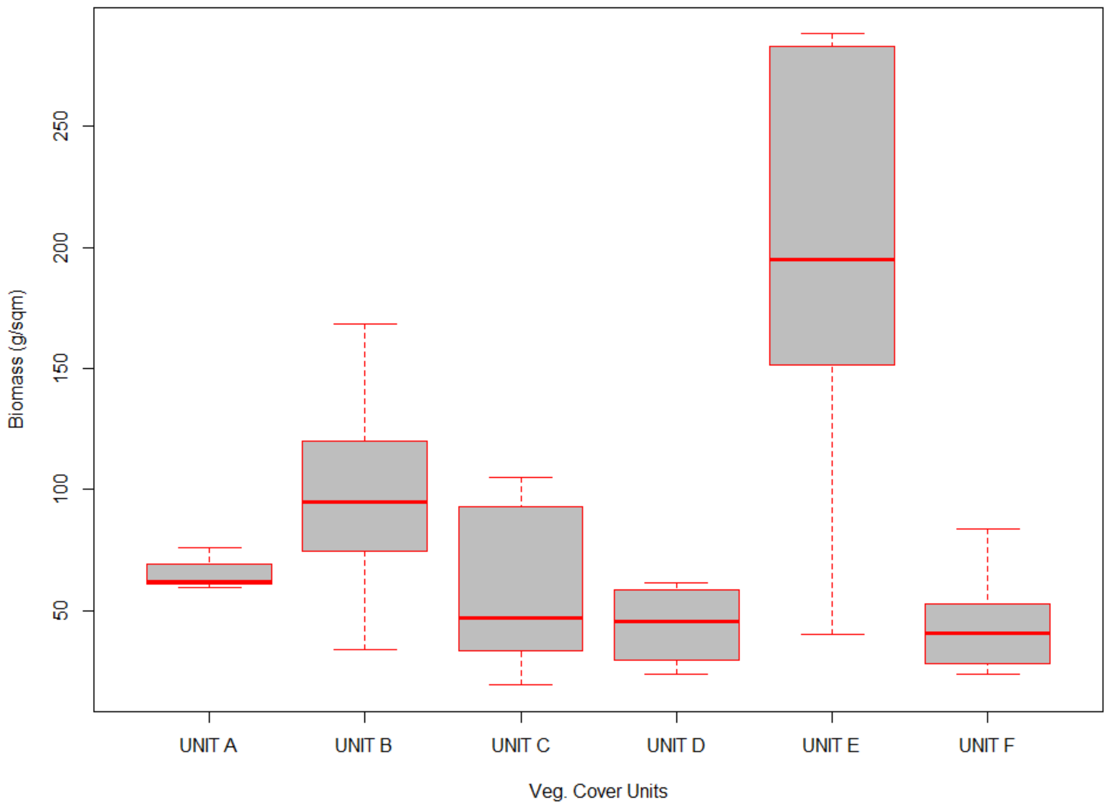


Figure 19: Boxplots of Biomass by grassland vegetation cover types

Since at least one of the group means is different from the others, a post hoc analysis was carried out for pairwise comparisons of the samples. The results reveal that mean of biomass measurements are significantly different between vegetation cover units; C – E, D – E, and E – F, as shown in a matrix containing a pairwise comparisons p-values (Table 4).

Table 4: Pairwise comparisons using Tukey and Kramer (Nemenyi) test with Tukey-Dist. approximation for independent samples

	Unit A	Unit B	Unit C	Unit D	Unit E
Unit B	0.9460	-	-	-	-
Unit C	0.9991	0.4617	-	-	-
Unit D	0.9659	0.3055	0.9911	-	-
Unit E	0.2886	0.4496	**0.0066**	**0.0089**	-
Unit F	0.9753	0.2832	0.9953	1.0000	**0.0052**

4.1.4. Correlation of measured grass biomass to explanatory variables

Analysis of associations between measured biomass was done using Spearman's Rank-Order (rho) test. It was important to assess linear relationships between biomass and the associated variables before so as to understand the strength of their relationships. Since biomass data (Shapiro: W = 0.8439, p-value = 4.426e-05) does not assume normality in distribution, a non-parametric test of association was done to establish the strength of associations between biomass measurements and variables SD, Median, Trend, NDVI Oct. 2014, , distance to Bomas and grazers density. Table 5 presents results from Spearman's Rank-Order (rho) correlation analysis with an indication of the significance of the association (p-value ≤ 0.05).

Table 5: Spearman's Rank Correlation Coefficients

	Spearman's Rank	
	Estimate (rho)	p-value
SD	0.76	0.000
MEDIAN	0.14	0.377
TREND	-0.02	0.914
NDVI Oct 2014	0.10	0.545
Distance to Bomas	0.59	0.000
Grazers Density	-0.38	0.021

The type of relationship is indicated by a positive or negative sign meaning that a linear regression line can be fit through the data as shown on the scatterplots in Figure 20. The results of Spearman's Rank show that SD has a strong positive correlation with measured grass biomass while Trend has almost no associations with biomass. Grazer's density have a moderate negative association to measured grass biomass.

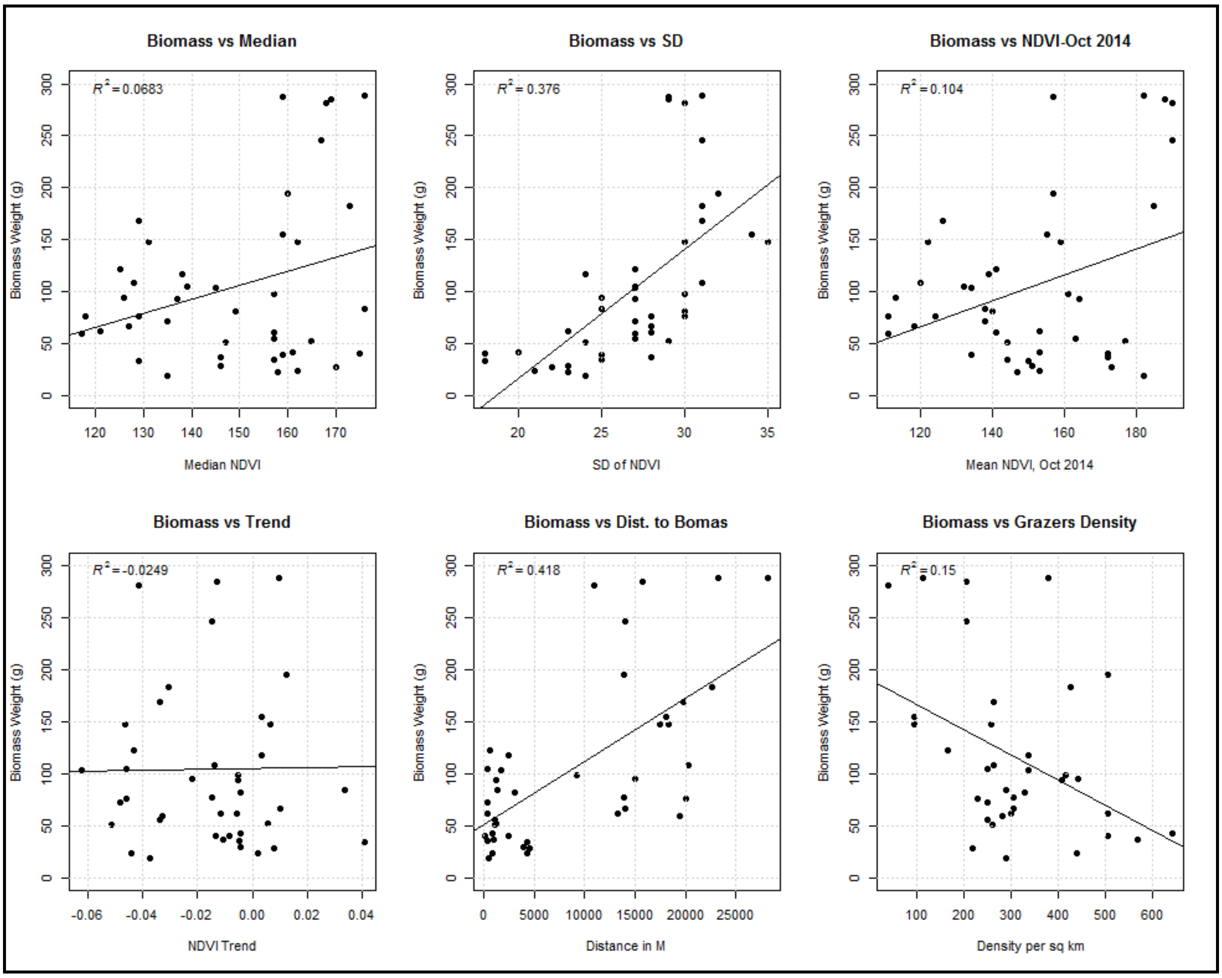


Figure 20: Scatter plots of biomass and correlation variables with lines of best fit

4.1.5. Multiple Linear Regression Analysis for Biomass prediction

In order to have a robust regression model, it was an important step to evaluate collinearity in the predictor variables. A check on correlation between the predictor variables was done using Spearman's rho correlation coefficients. In Table 6, significantly (α = 0.05) correlated variables are stared indicating that the data should be checked for possible multi-collinearity. The results of the multi-collinearity evaluation reveal that there is no collinearity between the predictor variables even though there exist some sort of relationships between them. The calculated Variance Inflated Factors (VIF) were all less than 10, indicating no collinearity between the independent variables. The relationships were also inspected visually using scatterplots (Appendix 7) and from the graphs and VIF calculation, it was concluded that there was no serious collinearity issue between the predictor variables.

Table 6: Correlation Matrix Table of Spearman's rho estimates

	SD	Median	Trend	Oct. NDVI	Grazers Density
Median	0.10				
Trend	-0.01	0.39*			
Oct. NDVI	0.09	0.70***	0.26		
Grazers Density	-0.41*	-0.08	0.20	-0.17	
Distance to Bomas (m)	0.60***	-0.08	0.19	-0.05	-0.35

The Regression Model

A stepwise linear regression analysis was conducted to evaluate the relationship of sample biomass measurements with the predictor variables so as to generate a linear regression function for biomass prediction. The results of this analysis indicate that the quality of the prediction (R) of biomass by the independent variables (SD, NDVI by Oct 2014 and Distance to Bomas) is 80.8% indicating a good level of prediction. The results also show the coefficient of determination (R^2) for this model as 0.653. This means that the predictors accounts for about 65.3% of variability of the biomass data with adjusted R^2 of 62.5%. The difference between R^2 and Adj. R^2 indicate that the regression equation generated maybe over-fitted to the data by 2.8% meaning that the upward bias in R^2 is not very big in terms of percentage.

Table 7: A table of Model coefficients explaining biomass (R^2 = 0.653, N=42)

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
(Constant)	-303.170	74.937		-4.046	.000
SD	6.484	2.371	.326	2.734	.009
NDVI Oct 2014	1.289	.335	.370	3.848	.000
Distance to Bomas (m)	.004	.001	.483	4.045	.000

The anova test was done and the results ($F_{(3, 38)}$ = 23.8; $p < 0.0005$) show that the model is a good fit for the data, an indication that the independent variables statistically significantly predict biomass in grams per meter square. The obtained model unstandardized coefficients, B_1, (table) indicate how much biomass value vary with each of the predictor variables when all the others are held constant. The unstandardized coefficients as statistically significant ($t > 2$; $p < 0.05$). The model coefficients results shown support statistically (t-value and p-value) the rejection of the null hypothesis that either of the model parameters equals to zero. The alternative hypothesis that at least one of the model parameters is not equal to zero is upheld.

Model Assumptions Test Results

The assumption of linear regression analysis that the residuals are normally distributed was checked graphically using Q-Q plot and Histogram as well as non-graphical method, Shapiro-Wilks test. The Q-Q plot (Appendix 8) show that the points do not fall perfectly well on the diagonal line, more so, on the lower-left and upper-right. This is not considered as severe outliers to sufficiently lead to rejection of statistical normality at 0.05 significance level. Shapiro-Wilk W test tested for the hypothesis that the residuals of the model have a normal distribution and the result of this test (W = 0.958, p-value = 0.126) reveal that the true distribution of the errors is normal.

Serial correlation or autocorrelation was checked and tested using residual autocorrelation plots and Durbin-Watson statistic. The plot show that there is no autocorrelation of the residuals and this is confirmed through Durbin-Watson test of statistical independence in errors results, DW = 1.97, p-value = 0.8812. This means that there is no major structural problems in the regression model created.

4.1.6. Validation and Biomass Prediction

The Leave One Out Cross Validation (LOOCV) method that was used to perform a cross validation of residual sums of squares, that is, a corrected meaure of prediction error averaged across all folds produced an RMSE value of 50. 86. This can be compared to the original model which had residual standard error of 48.1 and this difference in terms of RMSE between validation model and the original model is 5.7%. There is a difference but it is not much to say one model is better or poorer than the other.

The regression model generated was used to estimate biomass in grams per square meter. The variables relationship explained by the model ($R^2 = 0.653$) enabled quantification of grass biomass by applying the coefficients generated to each image pixel relating to the predictor variable; SD, NDVI and distance to Bomas. The linear regression model generated that was used in a biomass distribution mapping equation is as shown in the formula below.

$$\text{Biomass (g/m}^2\text{)} = -303.17 + 6.484*\text{SD} + 1.289*\text{Oct NDVI} + 0.00447*\text{Dist. Bomas}$$

The result of the predicted biomass distribution in the study area show that biomass production is averagely lower than 100 grams per square meter in the Mara and sample portions of Loita and Siana range areas as at October 2014. Masai Mara National Reserve, one of the protected areas in Narok County, had on average more than 150 grams per square meter of grass biomass. In Figure 21, a map of biomass distribution is presented with a legend not in equal interval, showing biomass ranges in grams per square meter.

The model was applied in the same way as for October 2014 for the past 14 year for the same period in time. The prediction maps differ from one year to the other for the month of October indicating that measurable grass biomass vary every year (Appendix 9).

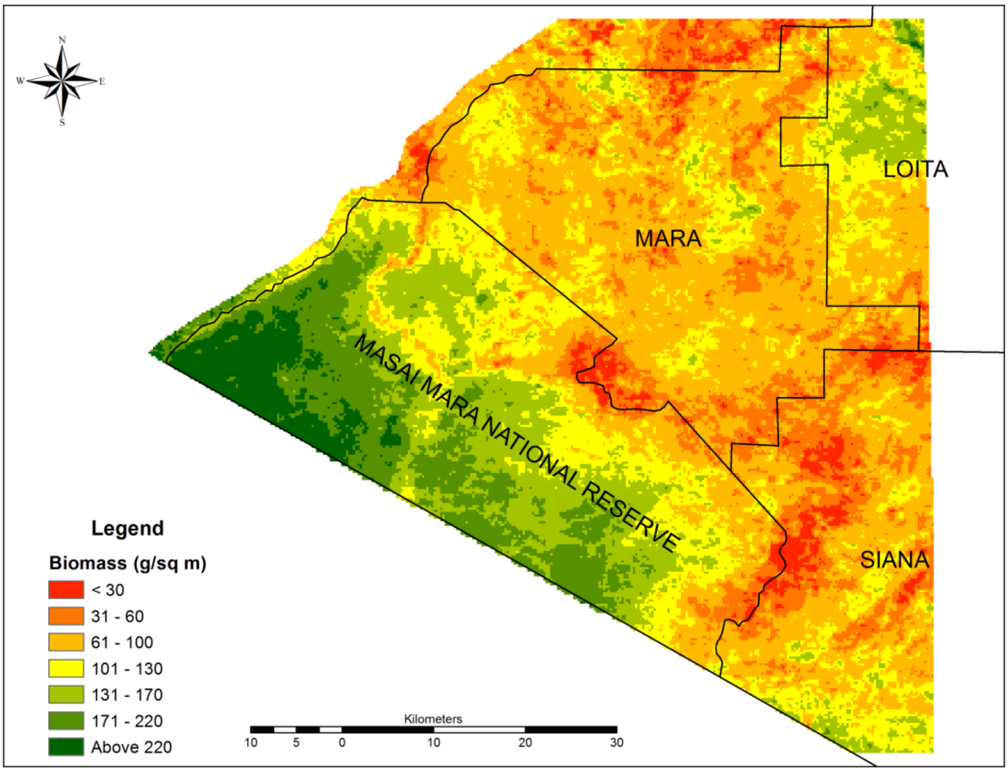


Figure 21: Biomass distribution as at October 2014

4.1.7. Comparisons between model estimates and historical biomass data

A Wilcoxon signed rank test with continuity correction test was conducted to test the null hypothesis that the two biomass samples; measurements collected by ILRI see spatial distribution in Figure 23 and the estimations from the model generated are identical. The matched sample distributions are shown in the boxplots of Figure 22 indicating that measured biomass from ILRI are right skewed while sample estimates of 2006 follow a normal distribution.

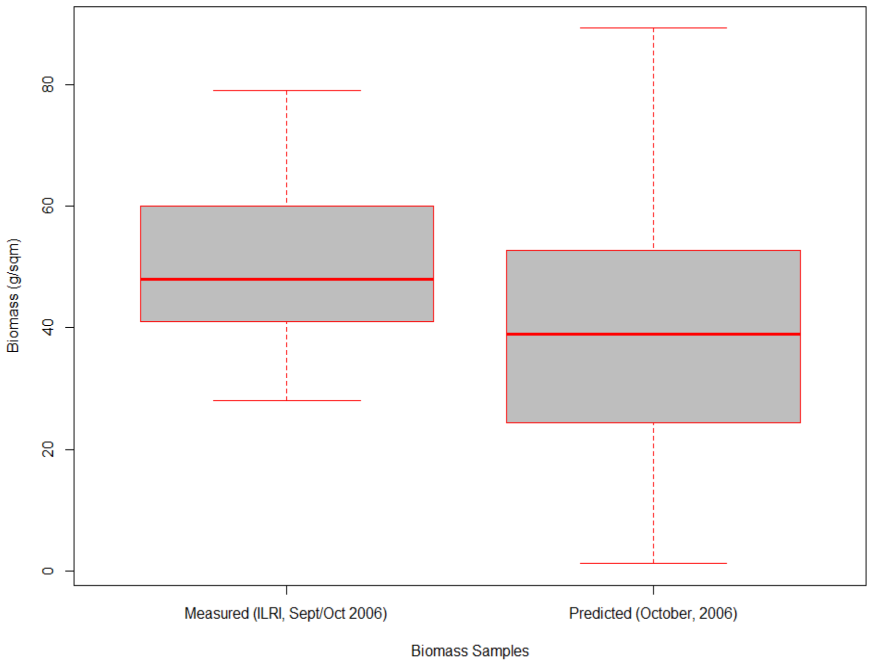


Figure 22: Boxplots of historical actual biomass and predicted biomass

A Wilcoxon signed-rank test showed that the two samples of grass biomass are not identical at a significance level 0.05 as the resulting p-value (Z= -7.238314; p-value = 4.543e-13) is less than 0.05.

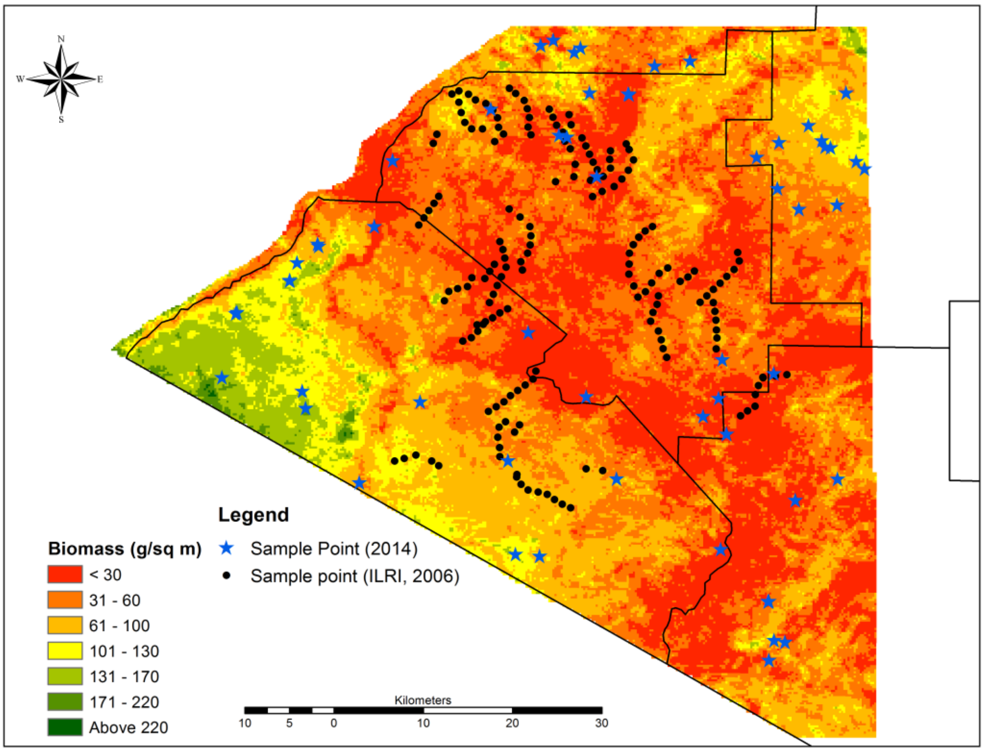


Figure 23: Distribution of Sample points by ILRI on top of the 2006 Sep/Oct biomass estimates

The year 2006 map show that there was on overage very low measurable grass biomass for the period Oct. on average, that period the predicated average measurable grass biomass was dominantly lees that 30 grams per square meter. ILRI 2006 sample points were more distributed in Mara and Masai Mara range areas. Loita and Siana were not targeted in their sampling.

4.2. Discussion

4.2.1. Analysing Hyper-temporal Modis NDVI Data for Vegetation Cover Mapping

Definitions for grasslands vary in many ways from one researcher to another depending on the scope at which they look at it. Some prefer defining grasslands based on vegetation while others distinguish them based on climate and soils. Definitions can range from global to local scale. According to White, Murray, and Rohweder, (2000), grassland is defined as vegetated areas predominantly covered by grasses, with little or no tree cover. A definition by FAO (Reid, Serneels, Nyabenge, & Hanson, 2005) refer to grasslands as "land covered with herbaceous plants with less than 10 percent tree and shrub cover". White et al., (2000) categorically highlights definitions of grassland based on classification of vegetation whereby forests constitute complete tree canopy cover with three or more overlapping vegetation groups, woodlands as areas with 50 to 100 percent tree canopy cover and a graminaceous layer, savannahs are defined as those areas composed of 10 to 15 percent cover by woody plants and finally, well-developed grass and grasslands are those areas with less than 10 percent tree cover.

This study presents an approach of discerning vegetation cover based on analysis of remote sensing NDVI data. The results of this study show that grasslands can be distinguished and mapped from remote sensing data following the vegetation classification scheme as highlighted by White et al., (2000). The mapping result show that the characteristics of the vegetation observed in the field can be categorised into 6 cover units using Modis NDVI data. Mapping units coded A, B, C, D, E and F are complexes of vegetation types differentiable through remote sensing data. Observed percent cover can be grouped into two; living and non-living. This means that the legend of the map produced is interpreted in a way such that percent cover for trees and high shrub, low shrub, herbs and grass constitute the living while on the other hand, litter, bare and stony represent non-living. The percent cover in each cover unit sums up to more than 100% because of overlaps between ground layer cover and tree and high shrub layer. Naturally, trees and high shrubs provide an overlap in cover percentage whereby under the canopy of trees and high shrub, low shrubs, herbs and grass are found depending on environmental conditions supporting vegetation growth.

The map in Figure 17 reveal that mapping units A, B, C, D and E can be classified as grasslands and/or savannah since tree and high shrub cover an area less than 15%. The differences in these classes are also indicated in the non-living part which in most cases, the bare portions. Bare percent indicates to a greater percentage the status of the vegetation in a grassland cover type; either degrading or degraded. The high percent cover of tree and high shrub in map unit F means that the unit is under the vegetation class woodlands.

Studies have used Remote Sensing NDVI data before to distinguish vegetation based on phonological variations of different cover types (Skidmore et al., 2003; de Bie et al., 2008; Ali et al., 2012 and Jiang et al., 2013). This study did not stop at using NDVI but moved further to classify hyper-temporal NDVI data from Modis Terra into statistics Median, SD and Trend. Median NDVI statistics distinguishes vegetation based on the optimum measurable NDVI over a long period of time. Median is used because it is less sensitive to fluctuations or extreme values of NDVI that result from highly variable rainfall in the area as well as human effects to vegetation. Variability in measureable NDVI is detected and mapped through SD and it shows how, in long term, vegetation in the area deviates from the mean NDVI. When the SD gets large it means that the vegetation cover is not in a stable state over that period of time, it fluctuates more every time. However, it is important to note that the strength of SD is to map variability in vegetation and insinuates status of vegetation.

The vegetation dynamics in this ecosystem dates back to early 60s when human settlement started to rapidly increase in the ecosystem. Dublin, Sinclair, & McGlade, (1990) notes that the woodlands in Mara-Serengeti rapidly begun to decline in 1960s and greater parts fully got converted to grasslands by 1980. Land conversion to different uses has not stopped in this area yet. Therefore, mapping of vegetation cover units using Median and SD captures the vegetation dynamics resulting from human activities or heavy grazing by wildlife. Trees and high shrub and woody areas have less variations in measured NDVI (low SD) and this is depicted by the map where the low values of SD combine with high Median as seen in Unit F.

Mapping areas with very low median and high SD relate to very unstable vegetation. For instance, Map Unit A and F depict a scenario where Unit A have low median and relatively high SD combinations with highest percent in bare meaning it is a grassland that is degraded or degrading. On the other hand Unit F is a combination of high median with relatively low SD dominated by trees and high shrub, an indication of non-varying vegetation. It was possible also to depict the peaks of vegetation growing season through spectral profiles of long term NDVI (Appendix 2). The profiles show that the mapped vegetation are different but follow same annual pattern. This means that the differences mapped are due to plant species and the effects of weather and human activities cause variation. The mapping approach could not discern some observed percentage cover properly however, they were kept as separate mapping units because they were distinct in their annual spectral curves of the NDVI data.

The mapping approach used in this study performed averaged well in depicting the differences in vegetation cover observed in the field through the analysis and application of NDVI Median, SD and as well as NDVI Trend. All the mapping units from A – E have a trend in vegetation phenology for at least some parts meaning that the conditions of vegetation in the units are either improving or deteriorating. The bare conditions observed could be attributed to degradation of vegetation except for the cover Unit F which has only positive trend. Degradation in this case refer to change from the original vegetation cover to bare see Figure 24 (b). Bare can be as a result of overgrazing or land being converted to other uses such as crop farming, road construction etc. Cover Unit F also has high bare percent as depicted in the main map which can be attributed to high trees that do not allow growth of low shrubs, herbs or grass under their canopy.

(a) (b)

Figure 24: Cover Units as observed in the field; (a) high bush with bare, (b) human activity in the rangeland-new roads are constructed to facilitate tourism activities

4.2.2. Differences in measured grass biomass by grassland cover type

Measured grass biomass as grouped according to vegetation classes mapped through the process discussed earlier in section 4.2.1. Results of Kruskall-Wallis test carried out to check if the means of measured grass biomass were different by cover type indicate that the means are different from one cover type to the other. A post hoc analysis revealed that measured grass biomass are significantly different in mapping units C – E, D – E, and E – F.

The difference between Unit E and C, D and F could be as a result of different land management practices. Cover unit E is dominantly found in Masai Mara National reserve while the others are dominant in the surrounding conservancies and community land around. As said earlier, human pressure is a key driver of change observed in this ecosystem. Measurable grass biomass in this case could be an indicator of the differences that result from the human influence to the ecosystem. Grazing and farming activities could be rampant in the cover units C, D and E causing the significant difference in the measured biomass. Measured grass biomass could also be different in different cover units due to different grass species in each of these mapping units. Grasses species and palatability can also make grazers select where to graze and live. Another reason to key differences is through the distribution of grazers across the ecosystem. In some parts of the ecosystem you get wildlife dominate over the livestock. This is because of the restriction from the reserve management that grazing of livestock is only allowed in the park during dry season (KWS interview).

4.2.3. Assessing the main factors associated with measurable grass biomass in Masai Mara

i) Correlation analysis

Linear relationships between dry weights of grass biomass and variables SD, Median, Trend, NDVI, distance to Bomas and grazer's density were analysed for the entire ecosystem. The correlation coefficient results demonstrated that the strength of linear relationships between biomass weights and the variables SD, distance to Bomas and grazers density were significant at alpha 0.05. The relationships were not significant with variables Median, Trend and NDVI by Oct (see Table 5). The relationship was strong with SD (rho = 0.76) and moderate with the distance to Bomas (rho = 0.59). The relationship was positive between biomass and SD, Median, NDVI by Oct and distances to Bomas. On the other hand, it was negative with Trend and grazer's density.

Correlation coefficient significance tests between biomass measurements and the variables were to show if there exist a linear relationship between them which has a nonzero slope. The purpose of this test was not to tell about causality but association between the variables. It will be wrong to interpret the results of this analysis by saying, for instance, that large deviations from mean NDVI lead to high measurable grass biomass. This result can be deduced to mean what having high SD implies to grass biomass. If correlations coefficient get close to or is zero, it implies that there is no linear relationship between the variables (Taylor, 1990). However, this does not mean that Biomass and the variables are not related, they might be related through nonlinear relationships.

The relationship between biomass and SD, which is a strong positive linear correlation is not surprising and is in support of the idea that low SD value means least or non-varying vegetation characteristic, that is, a more stable vegetation over time. In this ecosystem, there is intense grazing from the large numbers of wildlife and livestock hence getting grass with near zero variation is not real. It is also possible to get very low long term NDVI standard deviation values in areas covered by trees, high shrubs and nonpalatable vegetation. Non-grazed areas or areas with little or non- palatable type of grass species may also record low SD and low measurable grass biomass. However, this ecosystem is a free-range dominated by grass species *Themeda triandra* which are very much palatable (Stelfox et al., 1986). Thus, variation in grasses is much more in the ecosystem in this current situation where modifications are more so caused by human activities than the animals themselves. Annual movements of the wildebeest, zebras and elands, back and forth into the ecosystem from Serengeti national park in Tanzania intensively graze on the grass rich ecosystem to near completion before they migrate back. This means that those grasslands that record high measurable grass biomass can also record high standard deviations because of the left-overs at the end of dry season and re-growth after the migratory animals have left.

The results of this study also indicate that there is no significant relationships between measurable grass biomass and NDVI value measured by the time of this study (October 2014) and the long term Median. The analysis reveal that there is a positive linear relationship between biomass and the two variables though not significant. This can be alluded to the fact that field sampling took place at the beginning of the short rains, just at the end of the dry period. Grass re-growth at the beginning of short rains influenced the satellite measured NDVI of the month of October. Previous studies (McNaughton, 1985; Stelfox et al., 1986) have found a strong relationship between biomass and rainfall and wildlife distribution. Stelfox et al., (1986) describes the period after September as the period of stabilization of the non- migratory wildlife in this ecosystem. Therefore, having the factors, beginning of short rains and emigration of the wildlife, effects of re-growth explains the relationship of the measured biomass to the NDVI as at October and Median NDVI at this particular point in time.

Wildlife distributions was found in this study to have a significant negative correlation. A part from this being as a result of the large migratory wildlife and non-migratory animals concentrating on the re-growths, the other reason why biomass is negatively correlated to animal densities as McNaughton, (1985) explains is about the diets of grazing animals. Some areas are rich in some minerals required by grazers thus attracting them. Areas rich in nitrogen have fast re-growth and attracts animals. In this ecosystem, wildlife and livestock coexist outside the national reserve since it is not fenced. Traditionally livestock are corralled at night in Bomas by the Masai pastoralists. After Bomas being abandoned, they develop into a nutrient-rich, treeless glades that can persist for centuries (Donihue, Porensky, Foufopoulos, Riginos, & Pringle, 2013). The glades attract both native and domestic large herbivores thus the negative linear relationship between measurable biomass to the distance from Bomas. Bomas are also put up not far from water points. This means that Bomas attract animals for two things; nutrient-rich grass and water.

(a) (b)

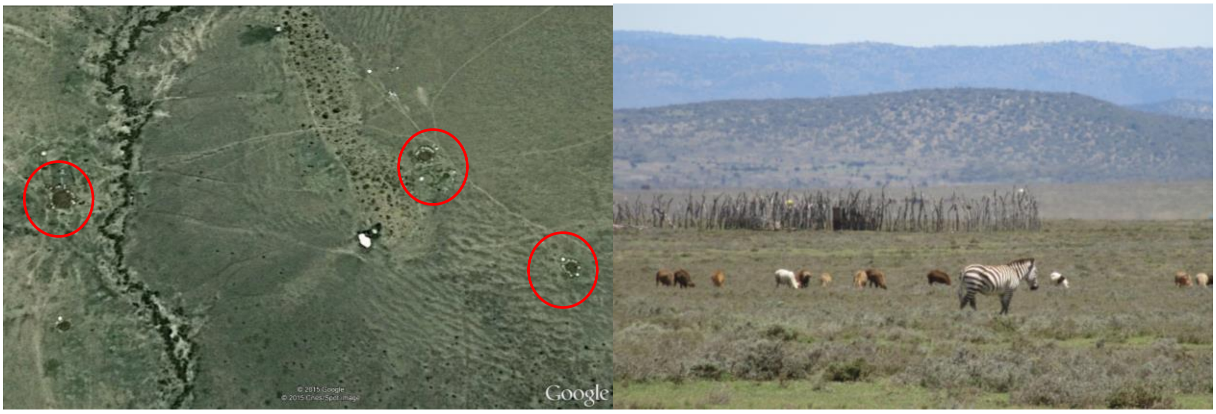


Figure 25: Bomas as seen from Google Earth image (a); Wildlife and Livestock graze together near an existing Bomas (b)

ii) Regression Analysis and Spatial Prediction for Grass Biomass

Biomass measurements collected from the field were used as training dataset for biomass distribution model. The patterns of the existing relationships between measured above ground grass biomass and the associated variables were investigated through a regression analysis. The actual measurements of the relationships were indicated by the coefficient of determination R^2. The scatterplots in Figure 20 indicate that SD, NDVI in Oct, distance to Bomas and grazers' density can significantly explain, to some percentage, the variations in the measureable above ground grass biomass in the ecosystem. They can be used to fit a model for predicting grass biomass in the Masai Mara ecosystem. Stepwise linear regression was used to evaluate the relationships and establish the function for estimating biomass.

A multiple regression was run to generate a prediction formula for grass biomass using SD, NDVI as at Oct 2014 and Distance to Bomas as these were found through a stepwise procedure to be the most significant variable to use. These variables statistically significantly predicted biomass, $F(3, 38) = 23.8$, $p < 0.0005$ and $R^2 = 0.653$. The selected variables added statistically significantly to the prediction, $p < 0.05$, where their relative contribution to the prediction are in the order; SD, followed by NDVI for Oct 2014 and Distance to Bomas. In this modelling procedure, some highly predictive variables were left out for the sake or a good model. For instance, the variable Grazers density was left out not because it was a poor predictor but due to its moderate negative correlation with the variables distance to Bomas see Table 5 and Table 6, the strength of its association with grass biomass was assessed in comparison to that of Distance to Bomas then a decision was made.

The results of this regression analysis demonstrate that remote sensing data can be used to predict an estimate of grass biomass in Masai Mara. The reason being, SD, NDVI and distance to Bomas are variables that can be acquired through remote sensing at a reasonable cost for large scale applications. NDVI data used in this study is a freely available data and Bomas data are also visible and easy to collect through free google earth images. This saves the cost of data collection and still get results with an accuracy of more than 65%. The large population of wild herbivores and livestock who have different nutrient requirements that graze in this ecosystem (Bhola et al., 2012) coupled with ever increasing human activities in the ecosystem necessitates a constant assessment of biomass availability which is possible through application of this model.

Spatial variation and distribution map for measurable grass biomass was produced through application of the coefficients of the model variables to make a biomass map for the ecosystem (Figure 26). Masai Mara National Reserve had on average the highest measurable grass biomass by October 2014 predicted by the model. This can be attributed to the management at the reserve especially in the Mara triangle to the west of the reserve. The lowest measurable grass biomass estimate were found just right at the edges of the reserve, especially to the Eastern/north eastern side of the reserve. This can be linked to the influence of human population, that is, the pastoral community grazing large numbers of cattle from the edges towards the reserve. Ottichilo, Leeuw, & Prins, (2001), attributes the decline of resident wildebeests in Masai Mara to habitat loss from human encroachment at the periphery of the national reserve boundary. This is in support to why there is little biomass estimates more so outside the park.

In Figure 26 below, it shows spatial biomass predictions overlaid with conservancies boundaries. The area called Talek is not a conservancy but an urban centre developing just at the edge of the Masai Mara National reserve. In general, measurable grass biomass is low in the conservancies than within the national reserve.

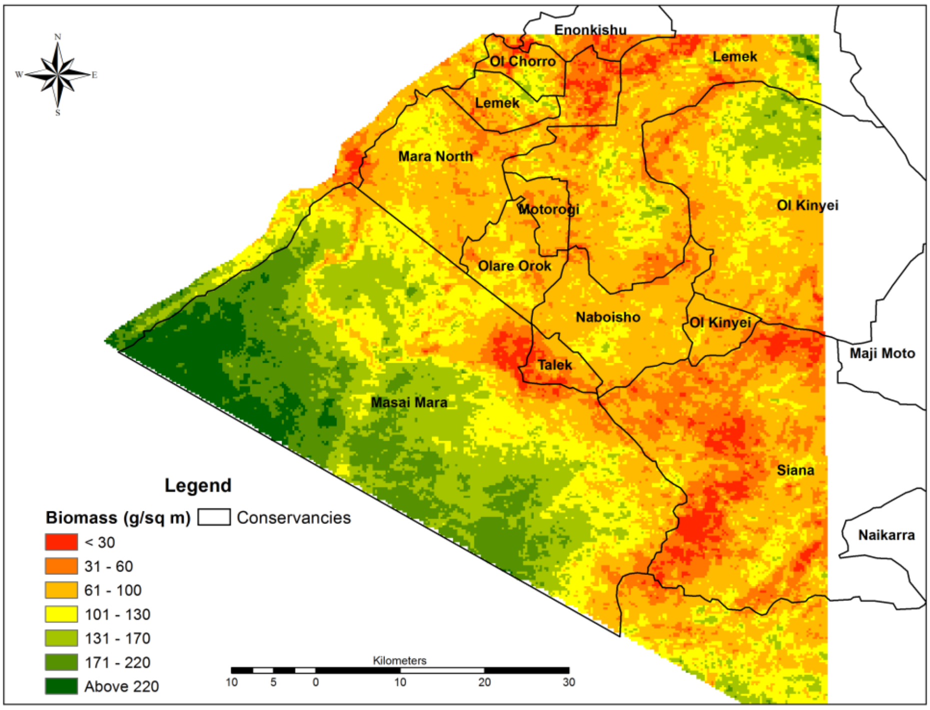


Figure 26: Biomass distribution within conservancies in Mara

In monitoring grass biomass, it is important to consider biomass produced in every growing season. The growing peaks of grass in between the months April/May just before the great wildebeest and zebra's migration in the ecosystem which occur around June-July. This is the period that grazing is intense in this ecosystem and it ends around Aug-Sept, see Appendix 2. However, biomass vary from year to year as could be seen in predicted maps of same period of this study for the past 14 years. Appendix 9 show maps of predicted biomass to differ from year to year. This means that a part from intense grazing in the ecosystem, there are other factors which influence measurable grass biomass.

4.2.4. Comparison between estimated grass biomass and actual historical biomass

A pairwise comparison test was done to compare estimated grass biomass with biomass samples acquired from ILRI in 2006. A prediction samples of Oct 2006 map were paired to the ILRI data for comparison. The results of Wilcoxon signed rank test reveal that the two samples collected by ILRI, Kenya and model estimates for 2006 are not identical at a significance level 0.05 as the resulting (Z= -7.238314; p-value = 4.543e-13) p-value is less than 0.05.

The difference can be attributed to the different approaches to biomass measurements. ILRI data was collected using 8 quadrats (0.25m by 0.5m) in transects and then averaged comparable to what this study used, 1x1m quadrat. The results are not comparable because of the sampling design that was used. ILRI used systematic sampling method while this study used random stratified sampling.

The non-identical nature of these two samples can also be attributed to the timings of the two studies. ILRI's data of more than 170 samples collected in Sept-Oct 2006 in comparison to 3 weeks data collection exercise in Oct 2014 that yielded 42 samples that this study used. A lot might have changed in this ecosystem that in continuous state of modifications which affect measurable grass biomass. Rainfall is also highly variable and unreliable in this ecosystem as mentioned earlier in this study.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusion

This study aimed at advancing research in the analysis and applications of hyper-temporal Modis terra NDVI data to estimate and assess key factors associated with grass biomass. The design and approach of this study was to first map the vegetation cover units in the ecosystem and then provide techniques for estimating grass biomass for the grasslands. Using GIS and Remote sensing techniques, a sample scheme for vegetation survey and biomass sampling was designed and used to collect data for vegetation mapping and biomass estimation modelling. The study was to find an approach to science that will give the most accurate estimates of grass biomass explained by the characteristics of vegetation cover mapped.

The mapping approach followed in this study succeeded in mapping out the differences in vegetation cover types through analysis of hyper-temporal remote sensing data. Long term standard deviation of NDVI and Median captured well the dynamics of vegetation in a rapid and constantly changing environment. The difference in degrading, degraded and stable vegetated area was mapped through this technique. However, due to overlaps between cover percentages observed, it is not easy to differentiate properly the mapped classes although they are distinct through remote sensing data. Unsupervised classification of hyper-temporal Modis NDVI data is a powerful technique and reliable in mapping vegetation since it is climate dependent and not weather dependent.

Measured grass biomass is linearly correlated to remote sensing data; long term SD, Median and NDVI. Measurable grass biomass is significantly explained through the outputs of analysis of hyper-temporal NDVI data. Statistics of standard deviation and median of NDVI explains to a greater extent the amount of measurable grass biomass because as deviations from satellite-measured NDVI reduces or increases, it indicates that grass is in a stable condition. This study have further shown that animal density, presence of Bomas and human activities in the ecosystem affect the amount of measurable grass biomass negatively. As the amount of human activities such as increased animal keeping, farming and development of tourist facility increase in this ecosystem, significant variations in measurable grass biomass are reported. Therefore, the study concludes that SD, Bomas and NDVI measurements are key factors associated to measurable grass biomass.

The prediction of grass biomass using field measurements and remote sensing data significantly indicate the spatial and temporal characteristics of grasslands productivity. This is useful in monitoring grass biomass in space and time. The correlation coefficients between biomass and remote sensing data showed differences. This means therefore that the purpose of which to use each of the remote sensing products has to be clear right from the beginning.

In comparison to historical dataset and methods, the biomass technique used in this study is no comparable to the approach followed by ILRI in 2006. However, this study had a better approach that used few biomass samples and achieved a significant biomass prediction with accuracy of 65%. The methods are not comparable with methods applied by previous studies since it combines traditional clipping method with a remote-sensing based sampling approach to get as much as representative samples as possible. This improves the level of prediction and mean of estimates higher than that from the method applied in previous sampling.

5.2. Limitations

Key limitations to this study were that limited samples could not allow for clear differentiation between different grassland types in terms of percentage cover. The overlaps in estimated cover percentages could not be discerned through remote sensing properly. Secondly, the study lacked enough data to use for grass biomass model validation.

5.3. Reccommendations

This research was able to use effectively remote sensing data to design a sample scheme for vegetation surveying and biomass sampling. Field sampling data that was collected provided the basis for a better vegetation cover map and estimated grass biomass. The study recommends the following for future research.

- Future mapping projects should consider long term Standard Deviation of NDVI since it captures weather-based vegetation variability better than just one time stamp satellite imagery.
- Biomass estimation through remote sensing and statistical models can be improved using data collected for several successive years or seasons to enhance the stability of estimates and compare the variability in estimates over seasons.
- It will also be useful to include soil data and herbivore grazing intensity in future research in regard to grassland vegetation mapping and biomass estimation for this ecosystem.

LIST OF REFERENCES

Akter, J. (2014). Bootstrapped Durbin – Watson Autocorrelation for Small Samples Test of. *ABC Journal of Advanced Research*, *3*(2), 68–73. doi:prefix 10.15590/abcjar

Ali, A., de Bie, C. A. J. M., Scarrott, R. G., Thi, N., Ha, T., & Skidmore, A. K. (2012). Comparative Performance Analysis of a Hyper-Temporal NDVI Analysis Approach and a Landscape-Ecological Mapping Approach. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Vol. I, pp. 105–110). doi:10.5194/isprsannals-I-7-105-2012

Arai, K. (2007). ISODATA clustering with parameter (threshold for merge and split) estimation based on GA : Genetic Algorithm, *36*(1), 17–23.

Beltran-abaunza, J. M. (2009). *Method Development to Process Hyper-Temporal Remote Sensing (RS) Images for Change Mapping*. ITC, University of Twente. Retrieved from http://www.itc.nl/library/papers_2009/msc/gem/beltran-abaunza.pdf

Bhola, N., Ogutu, J. O., Piepho, H.-P., Said, M. Y., Reid, R. S., Hobbs, N. T., & Olff, H. (2012). Comparative changes in density and demography of large herbivores in the Masai Mara Reserve and its surrounding human-dominated pastoral ranches in Kenya. *Biodiversity and Conservation*, *21*(6), 1509–1530. doi:10.1007/s10531-012-0261-y

Biro, K., Pradhan, B., Buchroithner, M., & Makeschin, F. (2013). Landuse/Landcover Change Analysis and its Impact on Soil Properties in the Northern Part of Gadarif Region, Sudan. *Land Degradation & Development*, *24*(1), 90–102. doi:10.1002/ldr.1116

Boorman, L. (1997). Climate change 1995—Impacts, adaptations and mitigation of climate change: Scientific-technical analyses: Contribution of working group II to the second assessment report of the intergovernmental panel on climate change. In *Biological Conservation* (Vol. 81, p. 301). doi:10.1016/S0006-3207(97)83751-2

Boutton, T., & Tieszen, L. (1983). Estimation of Plant Biomass by Spectral Reflectance in an East African Grassland. *Journal of Range Management*, *36*, 213. doi:10.2307/3898165

Boutton, T., Tieszen, L., & Imbamba, S. (1988). Biomass dynamics of grassland vegetation in Kenya. *Ajr. J. Ecol*, *26*(2), 89–101. doi:10.1111/j.1365-2028.1988.tb00960.x

Boyd, D., & Danson, M. (2006). Hyper-Temporal earth observation theme, Earth Observation Technology Cluster. Retrieved January 04, 2015, from http://www.nottingham.ac.uk/eotechcluster/themes/hypertemporal.aspx

Cho, M. A., Skidmore, A., Corsi, F., van Wieren, S. E., & Sobhan, I. (2007). Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation*, *9*(4), 414–424. doi:10.1016/j.jag.2007.02.001

De Bie, C. A., Mobushir, R. K., Toxopeus, A. ., Venus, V., & Skidmore, A. K. (2008). Hypertemporal Image Analysis for Crop Mapping and Change Detection. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Vol. 37, pp. 803–814).

Delaney, H. D., & Vargha, A. (1998). The Kruskal-Wallis Test and Stochastic Homogeneity. *Journal of Educational and Behavioral Statistics*, *23*(2), 170–192. Retrieved from http://www.jstor.org/stable/1165320

Donihue, C. M., Porensky, L. M., Foufopoulos, J., Riginos, C., & Pringle, R. M. (2013). Glade cascades: indirect legacy effects of pastoralism enhance the abundance and spatial structuring of arboreal fauna. *Ecology*, *94*(4). Retrieved from http://www.esajournals.org/doi/abs/10.1890/12-0856.1

Dublin, H. T., Sinclair, a R. E., & McGlade, J. (1990). Elephants and fire as causes of multiple stable states in the Serengeti-Mara-Woodlands. *J Anim Ecol*, *32*(3), 674–682. Retrieved from http://www.jstor.org/stable/5037

Elmore, A. J., Mustard, J. F., Manning, S. J., & Lobell, D. B. (2000). Quantifying Vegetation Change in Semiarid Environments. *Remote Sensing of Environment*, *73*(1), 87–102. doi:10.1016/S0034-4257(00)00100-0

ESRI. (n.d.). ArcGIS Help 10.1. Retrieved January 16, 2015, from http://resources.arcgis.com/en/help/main/10.1/index.html#//00s50000001r000000

Fay, M. P., & Proschan, M. A. (2010). Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics Surveys*, *4*, 1–39. doi:10.1214/09-SS051

Field, A. (2009). *Discovering Statistics Using SPSS* (Third Edit., p. 857). London: SAGE Publications Ltd.

Gumma, M., Pyla, K., Thenkabail, P., Reddi, V., Naresh, G., Mohammed, I., & Rafi, I. (2014). Crop Dominance Mapping with IRS-P6 and MODIS 250-m Time Series Data. *Agriculture*, *4*(2), 113–131. doi:10.3390/agriculture4020113

Homewood, K., Lambin, E. F., Coast, E., Kariuki, a, Kikula, I., Kivelia, J., … Thompson, M. (2001). Long-term changes in Serengeti-Mara wildebeest and land cover: pastoralism, population, or policies? *Proceedings of the National Academy of Sciences of the United States of America*, *98*(22), 12544–9. doi:10.1073/pnas.221053998

Jakubauskas, M. E., & Legates, D. R. (2000). Harmonic Analysis of Time-Series AVHRR NDVI Data for Characterizing US Great Plains Land Use/Land Cover. In *Kansas Applied Remote Sensing* (Vol. XXXIII, pp. 384–389). Amsterdam: International Society for Photogrammetry and Remote Sensing.

Jakubauskas, M. E., Legates, D. R., & Kastens, J. H. (2001). Harmonic Analysis of Time-Series AVHRR NDVI Data. *Photographic Engineering and Remote Sensing*, *67*(4), 461–470.

Jakubauskas, M. E., Legates, D. R., & Kastens, J. H. (2002). Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and Electronics in Agriculture*, *37*(1-3), 127–139. doi:10.1016/S0168-1699(02)00116-3

Jiang, Y., de Bie, C. A. J. M., Wang, T., Skidmore, A. K., Liu, X., Song, S., & Shao, X. (2013). Hyper-temporal remote sensing helps in relating epiphyllous liverworts and evergreen forests. *Journal of Vegetation Science*, *24*(2), 214–226. doi:10.1111/j.1654-1103.2012.01453.x

Jin, Y., Yang, X., Qiu, J., Li, J., Gao, T., Wu, Q., … Xu, B. (2014). Remote Sensing-Based Biomass Estimation and Its Spatio-Temporal Variations in Temperate Grassland, Northern China. *Remote Sensing*, *6*(2), 1496–1513. doi:10.3390/rs6021496

Justice, C. O., Townshend, J. R. G., Holben, B. N., & Tucker, C. J. (2007). Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing*, *6*(8), 1271–1318. doi:10.1080/01431168508948281

Laliberte, A. S., Winters, C., & Rango, A. (2011). UAS remote sensing missions for rangeland applications. *Geocarto International*, *26*(2), 141–156. doi:10.1080/10106049.2010.534557

Lamprey, R. H., & Reid, R. S. (2004). Expansion of human settlement in Kenya ' s Maasai Mara : what future for pastoralism and wildlife? *Journal of Biogeography*, *31*(6), 997–1032. doi:10.1111/j.1365-2699.2004.01062.x

Lobell, D. B., & Asner, G. P. (2004). Cropland distributions from temporal unmixing of MODIS data. *Remote Sensing of Environment*, *93*(3), 412–422. doi:10.1016/j.rse.2004.08.002

Marill, K. A. (2004). Advanced Statistics : Linear Regression , Part I : Simple Linear Regression, *11*(1). doi:10.1197/S1069-6563(03)00600-6

McNaughton, S. J. (1985). Ecology of a Grazing Ecosystem : The Serengeti. *Ecological Monographs*, *55*(3), 259–294. Retrieved from http://www.jstor.org/stable/1942578

Menke, J., & Eric Bradford, G. (1992). Rangelands. *Agriculture, Ecosystems & Environment*, *42*(1-2), 141–163. doi:10.1016/0167-8809(92)90024-6

Mundia, C. N., & Murayama, Y. (2009). Analysis of land use/cover changes and animal population dynamics in a wildlife sanctuary in East Africa. *Remote Sensing*, *1*(4), 952–970. doi:10.3390/rs1040952

Ogutu, J. O., Piepho, H. P., Dublin, H. T., Bhola, N., & Reid, R. S. (2008). Rainfall influences on ungulate population abundance in the Mara-Serengeti ecosystem. *Journal of Animal Ecology*, *77*, 814–829. doi:10.1111/j.1365-2656.2008.01392.x

Oindo, B. O. (2008). Predicting Mammal Species Richness from Remotely Sensed Data at Different Spatial Scales. *The Open Remote Sensing Journal*, *1*, 7–16. doi:10.2174/1875413900100101007

Osborne, J., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research and Evaluation*, *8*(2), 1. Retrieved from http://pareonline.net/getvn.asp?v=8&n=2

Ottichilo, W. K. (2000). *Wildlife dynamics −an analysis of change in the Masai Mara ecosystem of Kenya*. Wageningen University.

Ottichilo, W. K., Leeuw, J. De, & Prins, H. H. T. (2001). Population trends of resident wildebeest [Connochaetes taurinus hecki (Neumann)] and factors influencing them in the Masai Mara ecosystem , Kenya. *Biological Conservation*, *97*(3), 271–282. doi:10.1016/S0006-3207(00)00090-2

Peterson, D. L., Price, K. P., & Martinko, E. A. (1998). Investigating Grazing Intensity and Range Condition of Grasslands in Northeastern Kansas Using Landsat Thematic Mapper Data. *Geocarto International*. doi:10.1080/10106040208542250

Puth, M.-T., Neuhäuser, M., & Ruxton, G. D. (2015). Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. *Animal Behaviour*, *102*, 77–84. doi:10.1016/j.anbehav.2015.01.010

Ran, G., Xiao-ke, W., Zhi-yun, O., & Ying-nian, L. (2006). Spatial and temporal relationships between precipitation and ANPP of four types of grasslands in northern China. *Journal of Environmental Sciences*, *18*(5), 1024–1030. doi:10.1016/S1001-0742(06)60033-8

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk , Kolmogorov-Smirnov , Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21–33.

Redjadj, C., Duparc, a., Lavorel, S., Grigulis, K., Bonenfant, C., Maillard, D., … Loison, a. (2012). Estimating herbaceous plant biomass in mountain grasslands: a comparative study using three different methods. *Alpine Botany*, *122*(1), 57–63. doi:10.1007/s00035-012-0100-5

Reid, R. S., Serneels, S., Nyabenge, M., & Hanson, J. (2005). The changing face of pastoral systems in grass-dominated ecosystems of eastern Africa. In J. M. Suttie, S. G. Reynolds, & C. Batello (Eds.), *Grasslands of the World* (p. 535). Rome: FAO. Retrieved from ftp://ftp.fao.org/docrep/fao/008/y8344e/y8344e00.pdf

Ruxton, G. D., & Beauchamp, G. (2008). Some suggestions about appropriate use of the Kruskal-Wallis test. *Animal Behaviour*, *76*, 1083–1087. doi:10.1016/j.anbehav.2008.04.011

Said, M. Y., Skidmore, A. K., Leeuw, J. De, & Prins, H. H. T. (2003). *Declining population of wild ungulates in the Masai Mara ecosystem : a sign of resource competition. Multiscale perspectives of species richness in East Africa*. Wageningen University, Wageningen.

Sala, O., Deregibus, A., Schlichter, T., & Alipe, H. (1981). Productivity Dynamics of a Native Temperate Grassland in Argentina. *Journal of Range Management*, *34*(1), 48–51. Retrieved from http://digitalcommons.library.arizona.edu/objectviewer?o=http://jrm.library.arizona.edu/Volume34/Number1/azu_jrm_v34_n1_48_51_m.pdf

Sala, O. E., Parton, W. J., Joyce, L. A., & Lauenroth, W. K. (1988). Primary production of the central grassland region of the united states. *Ecological Society of America*, *69*(1), 40–45. Retrieved from http://dx.doi.org/10.2307/1943158

Schucknecht, A., Erasmi, S., Niemeyer, I., & Matschullat, J. (2013). Assessing vegetation variability and trends in north-eastern Brazil using AVHRR and MODIS NDVI time series. *European Journal of Remote Sensing*, *46*, 40–59. doi:10.5721/EuJRS20134603

Sha, Z., Bai, Y., Xie, Y., Yu, M., & Zhang, L. (2008). Using a hybrid fuzzy classifier (HFC) to map typical grassland vegetation in Xilin River Basin, Inner Mongolia, China. *International Journal of Remote Sensing*, *29*(8), 2317–2337. doi:10.1080/01431160701408436

Skidmore, A. K., Toxopeus, A. G., Bie, K. De, Corsi, F., Venus, V., Omolo, D. P., … Real, R. (2003). Herpetological Species Mapping for the Meditteranean. In *Commission VII*.

Stelfox, J. G., Peden, D. G., Epp, H., Hudson, R. J., Mbugua, S. W., Agatsiva, J. L., & Amuyunzi, C. L. (1986). Herbivore Dynamics in Southern Narok, Kenya. *Journal of Wildlife Management*, *50*(2), 339–347. Retrieved from http://www.jstor.org/stable/3801925

Taylor, R. (1990). Interpretation of the Correlation Coefficient: A Basic Review. *Journal of Diagnostic Medical Sonography*, *6*(6), 35–39. doi:10.1177/875647939000600106

Terry, W. S., Hunter, D. H., & Swindel, B. F. (1981). Herbage Capacitance Meter: An Evaluation of Its Accuracy in Florida Rangelands. *Journal of Range Management*, *34*(3), 240. doi:10.2307/3898053

Tucker, C. J. (1979). Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. doi:10.1016/0034-4257(79)90013-0

Waite, R. B. (1994). The application of visual estimation procedures for monitoring pasture yield and composition in exclosures and small plots. *Tropical Grasslands*, *28*, 38–42.

Wang, X., Ge, L., & Li, X. (2013). Pasture Monitoring Using SAR with COSMO-SkyMed, ENVISAT ASAR, and ALOS PALSAR in Otway, Australia. *Remote Sensing*, *5*(7), 3611–3636. doi:10.3390/rs5073611

White, R., Murray, S., & Rohweder, M. (2000). *Pilot analysis of global ecosystems: Grassland ecosystems*. *World* (p. 69). Washington DC: World Resources Institute.

Wylie, B. K., Pieper, R. D., & Southward, G. M. (1992). Estimating Herbage Standing Crop from Rainfall Data in Niger. *Journal of Range Management*, *45*(3), 277. doi:10.2307/4002978

Xie, Y., Sha, Z., & Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology*, *1*(1), 9–23. doi:10.1093/jpe/rtm005

Yang, X., & Guo, X. (2011). Investigating vegetation biophysical and spectral parameters for detecting light to moderate grazing effects: a case study in mixed grass prairie. *Central European Journal of Geosciences*, *3*(3), 336–348. doi:10.2478/s13533-011-0032-4

Yeganeh, H., Khajedein, S. J., Amiri, F., & Shariff, A. R. B. M. (2012). Monitoring rangeland ground cover vegetation using multitemporal MODIS data. *Arabian Journal of Geosciences*, *7*(1), 287–298. doi:10.1007/s12517-012-0733-0

Yu, L., Zhou, L., Liu, W., Zhou, H.-K., Li, Z., Liu, W., & Zhou, H.-K. (2010). Using Remote Sensing and GIS Technologies to Estimate Grass Yield and Livestock Carrying Capacity of Alpine Grasslands. *Pedosphere*, *20*(3), 342–351. doi:10.1016/S1002-0160(10)60023-9

Zhang, X., Friedl, M. a., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., … Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, *84*(3), 471–475. doi:10.1016/S0034-4257(02)00135-9

Zhao, F., Xu, B., Yang, X., Jin, Y., Li, J., Xia, L., … Ma, H. (2014). Remote Sensing Estimates of Grassland Aboveground Biomass Based on MODIS Net Primary Productivity (NPP): A Case Study in the Xilingol Grassland of Northern China. *Remote Sensing*, *6*(6), 5368–5386. doi:10.3390/rs6065368

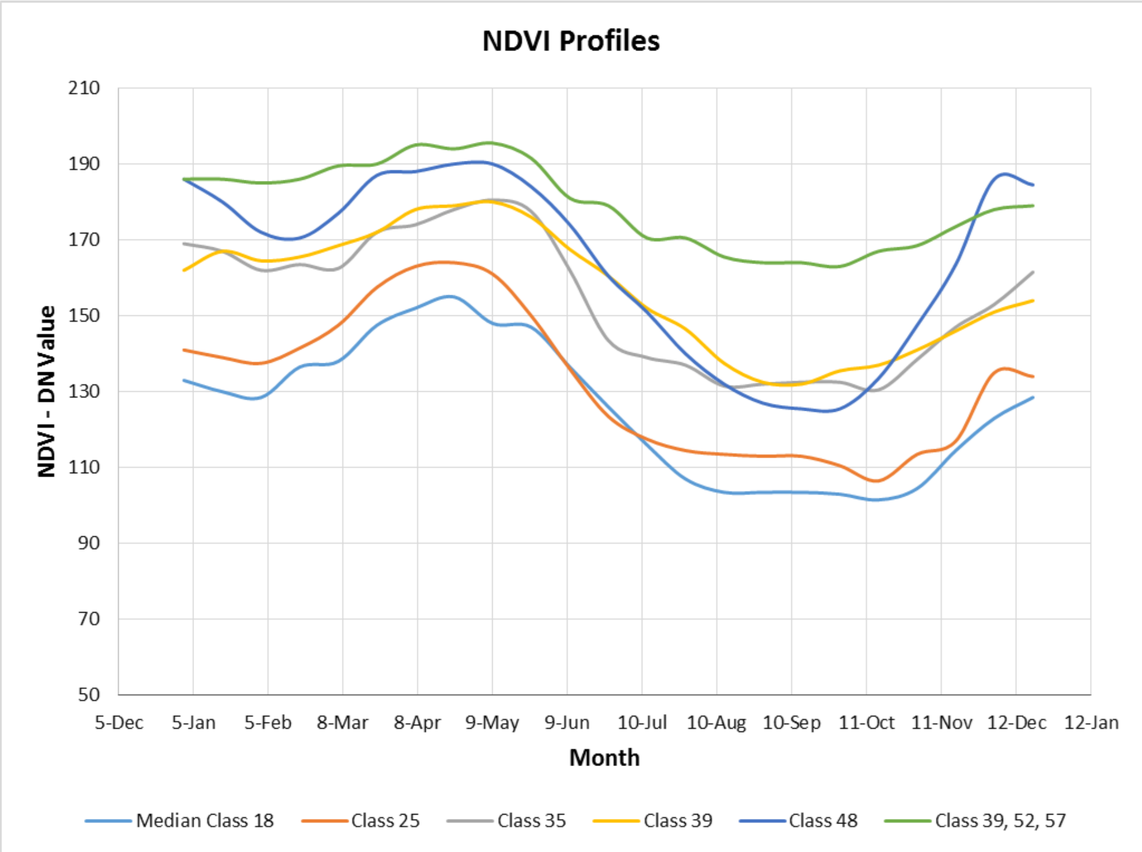
APPENDICES

Appendix 1: Data Table for the Map Legend

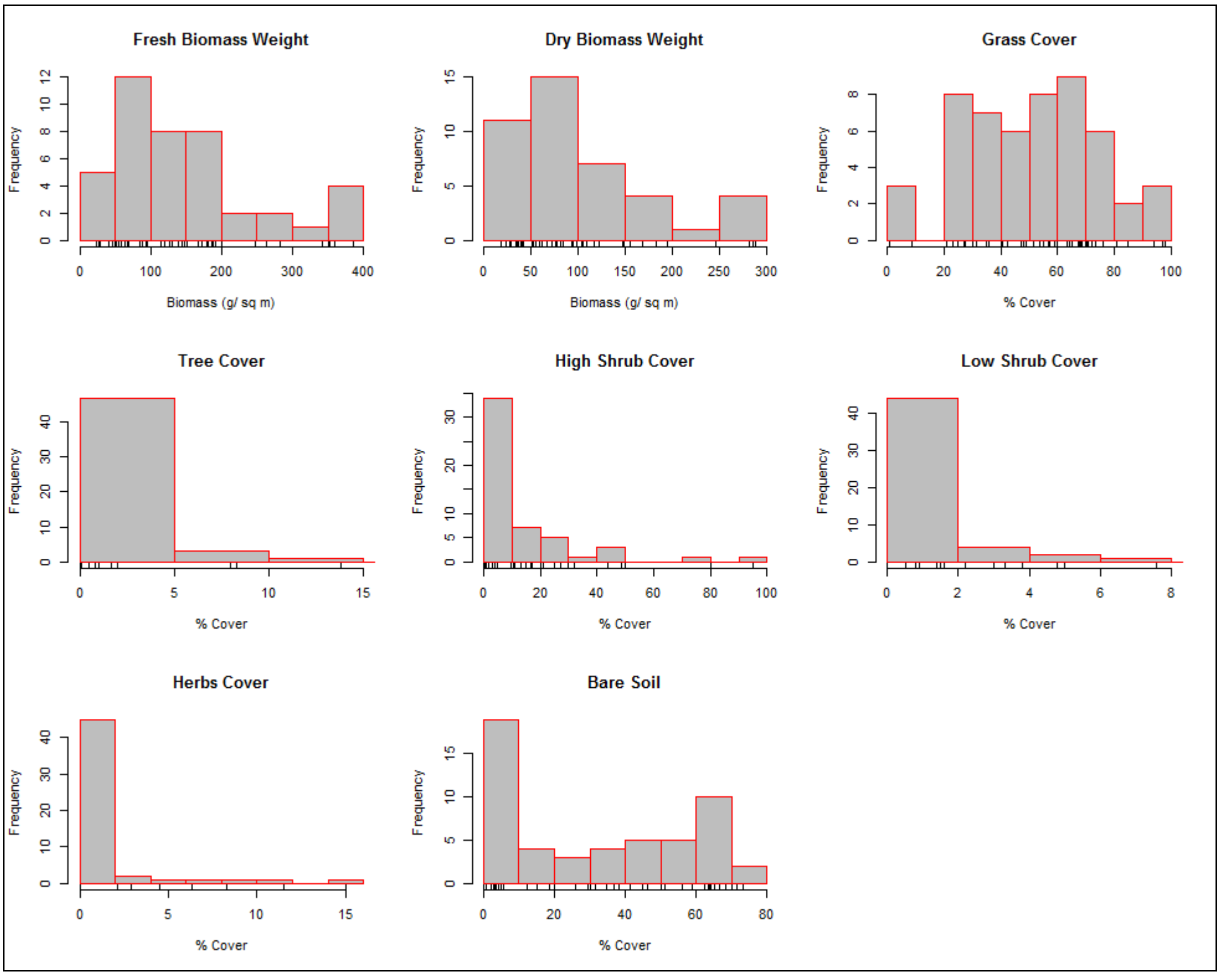
NDVI Class	Median Class	SD Class	Trend Class	Tree & High Shrub	Low Shrub	Grass	Litter	Bare	Stony
18	10	5	3	4.80	1.37	27.34	0.00	71.29	0.00
18	10	3	4	0.90	3.31	30.13	0.00	66.56	0.00
18	10	4	3	10.80	0.00	35.84	0.00	64.16	0.00
25	11	2	4	17.60	0.00	26.94	0.00	73.06	0.00
25	11	3	4	25.00	7.61	27.17	0.00	65.22	0.00
25	11	4	3	0.00	0.79	35.64	0.00	63.56	0.00
25	11	3	3	0.00	0.89	40.10	0.00	59.01	0.00
25	12	5	4	18.40	1.65	47.11	0.00	51.24	0.00
25	12	4	3	0.00	0.00	50.00	0.00	50.00	0.00
25	11	4	4	3.00	0.00	53.61	0.00	46.39	0.00
25	11	4	4	3.00	0.00	58.76	0.00	41.24	0.00
25	11	5	4	5.00	0.00	68.42	0.00	31.58	0.00
25	12	4	4	0.50	0.00	70.00	0.00	30.00	0.00
25	11	5	3	17.15	0.00	70.56	0.00	29.44	0.00
25	12	5	3	14.70	0.00	80.82	0.47	18.71	0.00
25	12	3	4	0.50	0.00	85.00	0.00	15.00	0.00
25	11	4	3	0.00	2.17	97.83	0.00	0.00	0.00
35	14	2	4	25.25	0.00	20.69	6.00	69.97	3.33
35	13	3	3	22.20	3.75	20.79	0.00	63.26	12.20
35	13	3	4	5.50	0.00	31.58	0.00	68.42	0.00
35	13	4	4	3.00	0.00	33.33	0.00	66.67	0.00
35	12	3	3	0.50	0.00	35.00	0.00	65.00	0.00
35	13	4	3	0.00	0.45	40.72	0.00	58.82	0.00
35	13	5	4	6.00	0.00	58.82	0.00	41.18	0.00
35	12	4	3	5.00	0.00	63.16	0.00	36.84	0.00
39	15	2	4	50.00	0.00	20.83	0.00	75.00	4.17
39	14	3	4	20.50	0.00	37.50	0.00	62.50	0.00
39	14	3	3	25.00	1.52	42.30	0.00	56.19	0.00
39	13	5	3	2.50	0.00	55.00	0.00	45.00	0.00
39	15	4	4	48.70	0.00	55.95	0.00	44.05	0.00
39	14	4	3	17.00	4.76	57.14	0.00	38.10	0.00
39	14	4	4	1.00	0.00	64.36	0.00	34.65	0.99
48	12	3	4	0.00	0.00	49.00	0.00	51.00	0.00
48	14	6	4	0.00	3.00	57.00	35.00	5.00	0.00
48	14	5	4	0.00	0.50	59.70	34.83	4.98	0.00
48	15	6	4	2.00	0.00	65.31	30.61	4.08	0.00
48	15	5	4	0.00	0.00	67.39	28.99	3.62	0.00
48	15	4	4	0.00	0.00	67.86	28.57	3.57	0.00
48	15	5	4	1.00	0.00	68.35	28.78	2.88	0.00
48	15	4	3	1.00	0.00	69.50	28.37	2.13	0.00
48	15	4	3	0.50	0.00	70.37	29.63	0.00	0.00
48	16	5	3	1.00	0.00	70.37	29.63	0.00	0.00

48	14	5	4	0.00	5.00	72.00	3.00	20.00	0.00
48	14	4	4	0.00	0.00	73.40	0.60	26.00	0.00
48	15	5	3	0.00	0.52	94.24	5.24	0.00	0.00
48	16	5	4	2.00	0.00	96.94	0.00	3.06	0.00
52	17	4	4	81.00	0.00	0.58	58.48	35.09	5.85
52	15	3	4	57.50	0.00	8.46	0.82	89.91	0.82
52	16	2	4	32.00	0.00	25.71	0.00	74.29	0.00
52	16	3	4	30.95	0.00	87.46	0.00	12.54	0.00
57	17	3	4	96.00	0.00	0.62	37.27	49.69	12.42
57	18	2	4	75.30	19.50	48.00	7.50	15.00	0.00

Appendix 2: Profiles of Map NDVI Classes; depicting vegetation growing seasons



Appendix 3: Histograms of field data



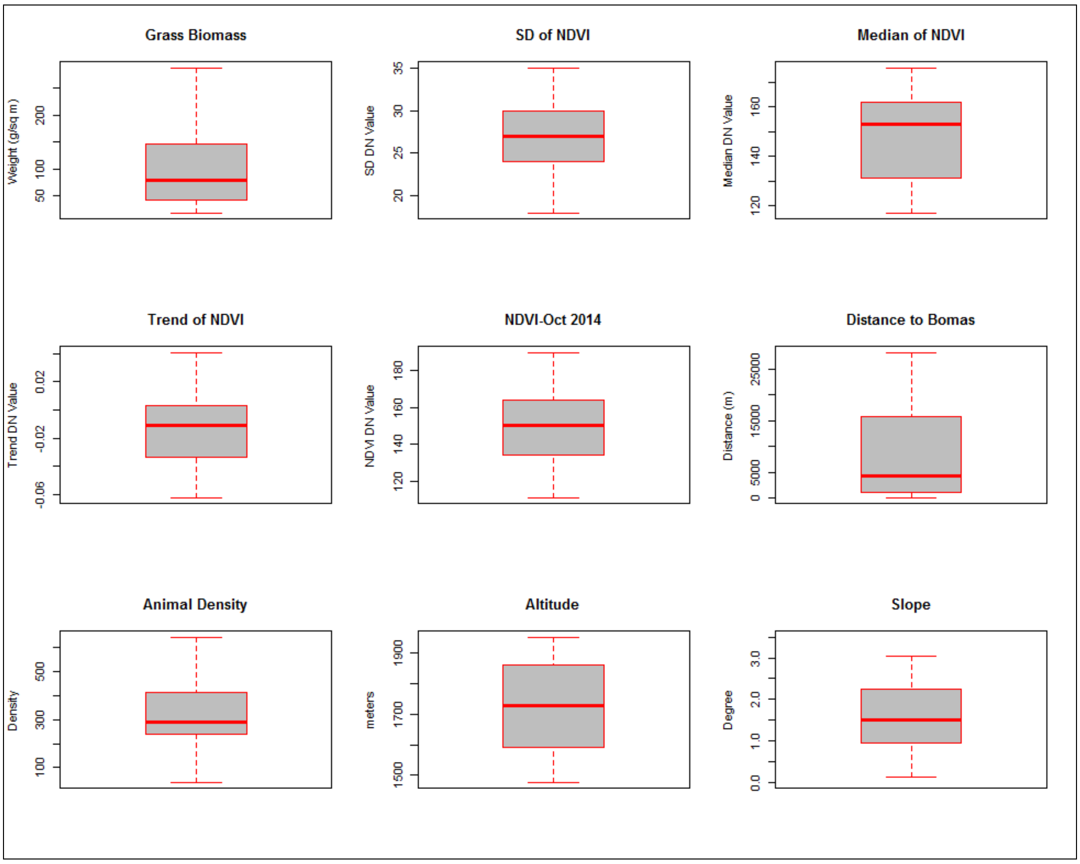
Appendix 4: Data Table for Biomass and Predictor Variables

Fresh Weight (g)	Dry Weight (g)	Median NDVI	SD NDVI	TREND NDVI	Oct NDVI (2014)	Altitude (m)	Density Grazers	Slope (degrees)	Dist. To Bomas
88.69	62.07	121	23	-0.0114	153	1935	298.88	1.18	13260.8
200.49	168.47	129	31	-0.0336	126	1953	262.04	1.56	19743.6
180.92	108.73	128	31	-0.0139	120	1946	262	0.94	20321.4
22.87	19.09	135	24	-0.0374	182	1757	288.88	1.87	424.264
57.8	39.92	159	25	-0.0134	134	1586	506	1.25	100
130.71	71.91	135	27	-0.0482	138	1691	250.24	1.64	316.228
179.48	93.47	137	27	-0.0055	164	1730	409.28	1.49	1208.3
190.7	147.62	131	30	-0.0461	122	1949	259.32	0.79	17489.7
84.92	59.37	117	27	-0.0330	111	1907	281.24	2.25	19376.3
69.76	51.57	147	24	-0.0512	144	1736	261.44	3.03	1118.03
50.05	29.43	146	23	-0.0043	151	1687		0.73	3962.32
139.31	103.97	145	27	-0.0623	134	1604	336.04	1.47	1769.18
114.11	81.88	149	30	-0.0043	140	1724	328.68	1.49	3080.58
46.05	42.25	161	20	-0.0042	153	1842	642.04	2.96	905.539
26.8	23.51	158	23	-0.0439	147	1835	441	2.1	824.621
55.64	35.26	157	25	-0.0048	144	1858		2.27	316.228
147.23	61.44	157	28	-0.0057	141	1719	505.88	0.76	400
264.18	195.07	160	32	0.0123	157	1521	505.84	2.63	13928.4
247.36	154.71	159	34	0.0032	155	1617	95.28	1.91	18011.4
29.57	23.76	162	21	0.0020	153	1874		2.03	4328.97
64.31	36.97	146	28	-0.0108	172	1725	569.28	1.47	984.886
41.76	28.19	170	22	0.0080	173	1821	218	2.3	4500
359.81	285.08	169	29	-0.0129	188	1567	204.48	1.12	15715.6
281.94	182.71	173	31	-0.0304	185	1567	426.92	0.11	22574.3
119.42	76.03	118	30	-0.0456	111	1912	228.32	0.94	19997
350.35	281.08	168	30	-0.0415	190	1586	40	0.88	10948.1
342.95	245.9	167	31	-0.0148	190	1591	204.48	1.65	14058.4
187.3	147.88	162	35	0.0063	159	1589	95.28	0.73	18313.4
94.68	40.71	175	18	-0.0085	172	1692		0.56	2502
151.77	83.82	176	25	0.0335	138	1921	288.88	3.04	1341.64
83.94	52.61	165	29	0.0057	177	1743		3.05	1216.55
386.32	288.24	176	31	0.0095	182	1505	113.6	1.25	28212.4
51.57	33.88	129	18	0.0408	150	1477		1.8	4285.4
351.94	287.85	159	29	0.0095	157	1591	379.96	1.12	23272.5
143.87	95.02	126	25	-0.0219	113	1863	442.48	1.75	15009.7
171.93	122.19	125	27	-0.0432	141	1726	164.52	1.03	583.095
94.33	66.53	127	28	0.0099	118	1911	305.4	0.59	14042.8
166.6	98.07	157	30	-0.0052	161	1592	415.4	2.39	9200.54
67.88	55.61	157	27	-0.0335	163	1830	248.96	2.64	1077.03
126.32	77.15	129	28	-0.0145	124	1903	305.4	0.39	13877.3
139.13	105.16	139	27	-0.0457	132	1696	250.24	2.7	316.228
185.73	117.34	138	24	0.0034	139	1748	336.36	0.15	2408.32

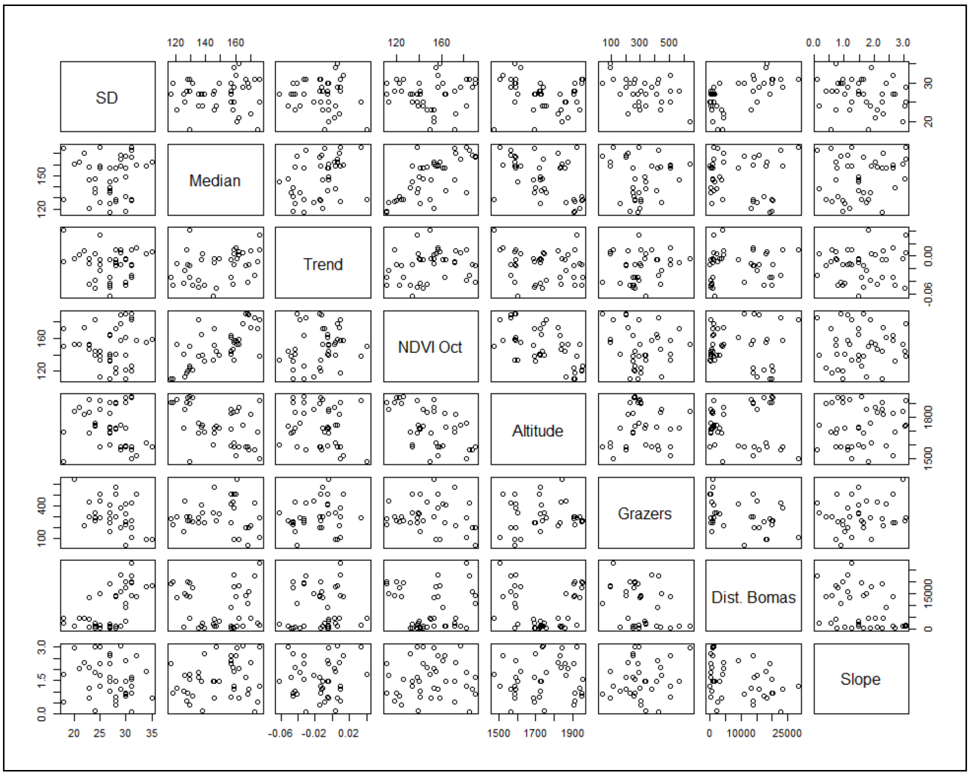
Appendix 5: Section of Historical biomass (2006) data and Estimated biomass (2006)

Sample	Lat	Long	DryWeight 2006	BiomassEst 2006
1	-1.395583	35.132985	148	35.21
2	-1.348995	35.110244	99	49.99
3	-1.358459	35.105883	103	52.09
4	-1.483486	35.181460	69	62.00
5	-1.490800	35.176321	73	49.78
6	-1.443350	35.187065	139	47.52
7	-1.448846	35.179621	73	51.68
8	-1.453997	35.172160	154	65.37
9	-1.460353	35.165648	138	52.44
10	-1.465188	35.157890	74	44.97
11	-1.470025	35.150164	44	65.12
12	-1.298566	35.161693	44	21.96
13	-1.306575	35.164768	35	29.49
14	-1.334155	35.163119	78	16.36
15	-1.343038	35.158964	114	22.51
16	-1.351686	35.155108	49	41.26
17	-1.200144	35.288954	44	42.51
18	-1.205723	35.291603	41	33.17
19	-1.216074	35.296199	36	58.84
20	-1.226626	35.290913	42	34.41
21	-1.235453	35.284393	68	31.37
22	-1.243175	35.276586	42	19.78
23	-1.248090	35.264501	48	8.71
24	-1.253099	35.255035	42	25.39
25	-1.226131	35.259661	47	38.00
26	-1.218441	35.255530	43	3.51
27	-1.210414	35.251804	47	35.64
28	-1.202231	35.248053	28	6.00
29	-1.194073	35.244364	40	38.88
30	-1.180682	35.233336	49	50.32

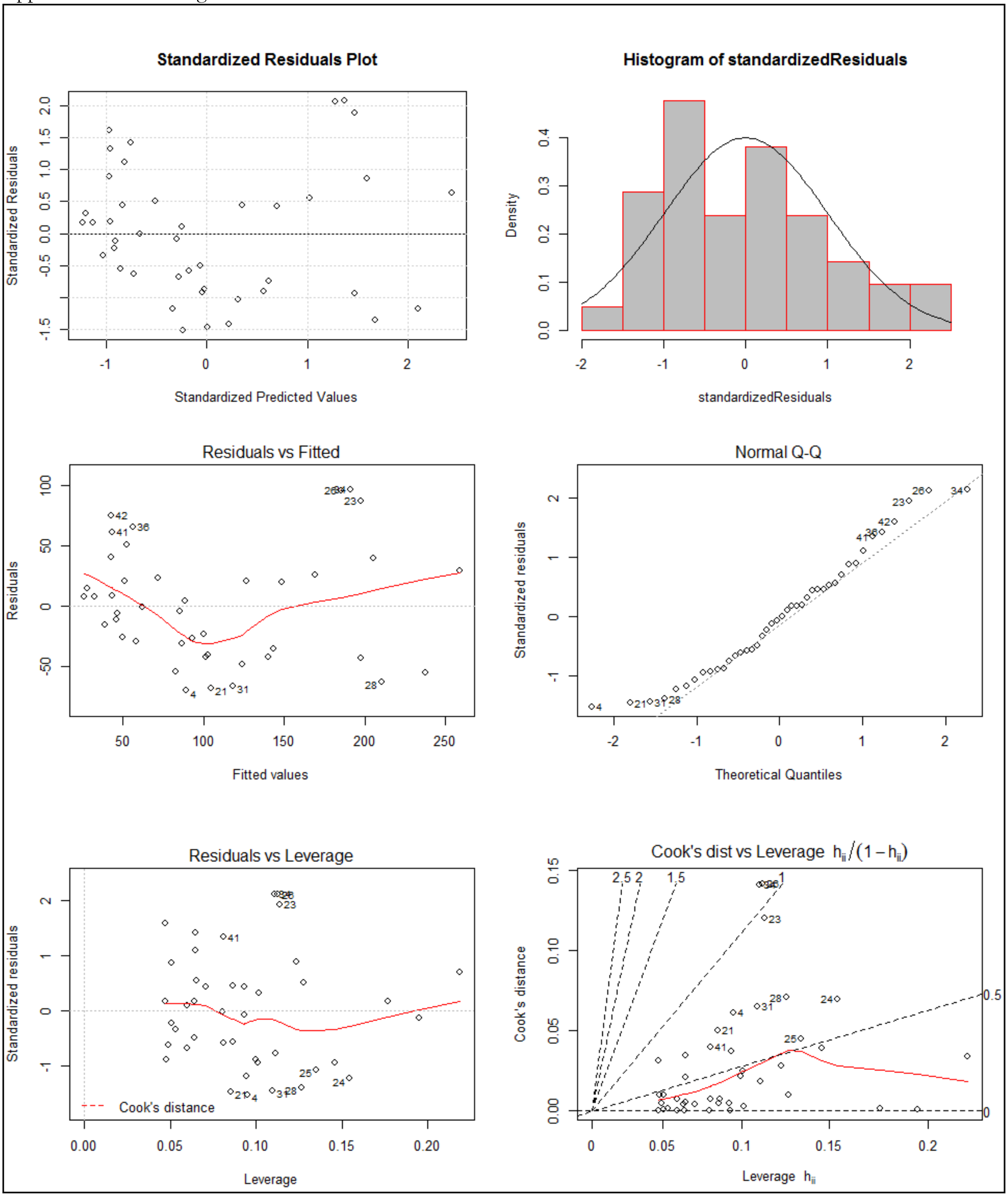
Appendix 6: Boxplot of Biomass weight and all the predictors



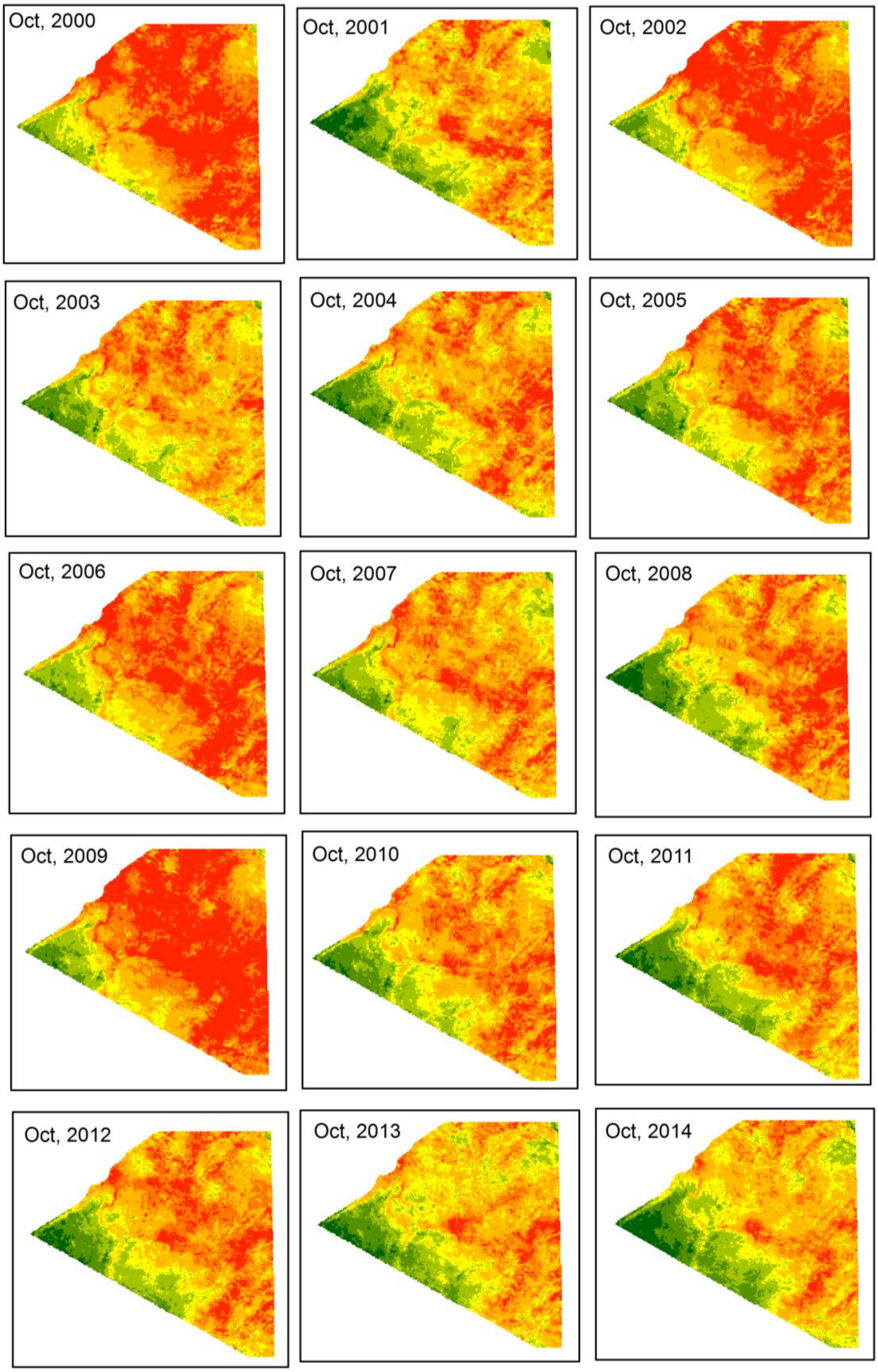
Appendix 7: Scatterplots of all the predictor variables used



Appendix 8: Model diagnostics



Appendix 9: Biomass Predictions in grams per square meter for the past 15 years



Appendix 10: Field-Data-Sheet-land-cover-use-Mara-rangelands 1

DATA COLLECTION SHEET

Date:	**Sample CODE:**		**Observers:**				
Photo No.:			**GPS:** S			**Altitude:**	
			E			**Plot size:**	
SLOPE		**STEEPILY DISSECTED**		**ALMOST FLAT**		**HILLY**	
STEEPY		**FLAT**		**UNDULATING**		**RIVERBED**	
MOUNTAINOUS		**VERY FLAT**		**ROLLING**		**OTHER (specify)**	

A. LAND COVER/USE DATA – (semi-)natural or planted

SAMPLE PLOT				POLYGON		
STRATA	**HEIGHT**	**COVER %**	**DORMINANT SPECIES (for details p.to.)**	*General cover/use type (if complex, estimate % cover of each type)*		
				TYPE	**DESCRIPTION**	**%AGE**
Trees						
Shrubs						
Herbs						
Grass						
Litter cover						
Bare soil						
Stones/rocky boulders						
ACTUAL LAND COVER				**ACTUAL LAND USE**		

OBSERVATIONS:

Appendix 11: Field-Data-Sheet-land-cover-use-Mara-rangelands 2

B. PLANT SPECIES LIST						SCALE FOR COVER ESTIMATION
PLANT CODE	SPECIES	COVER %AGE (see scale)	PLANT CODE	SPECIES	COVER %AGE (see scale)	R = rare P= few a = abundant m= many i = invasive* — < 5%
						1 = 5-15% 6 = 55-65% 2 = 15-25% 7 = 65-75% 3 = 25-35% 8 = 75-85% 4 = 35-45% 9 = 85-95% 5 = 45- 55% 10 = 95 - 100%
						• *Datura stramonium* • *Parthenium hysterophorus* • *Lantana camara* • *Tithonia diversifolia*
						REMARKS:

C. OBSERVATIONS/INTERVIEWS ON VEGETATION, CROPS, ANIMALS AND MANAGEMENT ASPECTS (SAMPLE POLYGON)				
1. Range condition: Annuals grasses (list them if Y) Y N		2. Burning Y N	%age burn:	<25% | 25-50% | 50-75% | >75%
			frequency in a year	
Perennial grass/weeds (list them if Y) Y N		3. Termite molds		
4. Water: Source:	Quality (*1poor to 5 high*): 1 2 3 4 5	5. Rangeland modifications Y N	Grazing induced	Stocking rate
Distance (*in km*)	Accessible Y N | Operational: Y N			
Seasonal | Permanent	Use: Domestic			
All use *Remarks:*		Others (specify)		
6. Droppings/footmarks/track Y N	Domestic | Wild	7. Mechanized Agriculture	Size of land (in ha):____________ Crops grown: ____________ Purpose:	

Appendix 12: Field-Data-Sheet-land-cover-use-Mara-rangelands 3

8. Fencing	Y	N	Distance	Purpose		9. <u>Smallholder Agriculture</u> Size of land (in ha): ____________ Main Crops grown: ____________ Purpose

10. Fuel wood collection	Y	N	Frequency	Daily	Weekly	Monthly	11. Livestock herd/wildlife: Count ____________ Movement: Free ________ Restricted ________ Interaction: YES/NO Frequency: Dry Season ________ Wet season ________ Mixed animals: YES/NO Wildlife alone: YES/NO Domestic animals alone: YES/NO
			Amount (bundles)				

12. Car tracks	Y	N	Frequency of use,	regularly	SIZE	Remarks
				seasonal		

D. SOIL DATA							
EROSION:						AREA AFFECTED	
TYPE: None☐	Sheet☐	Rill☐	Gully☐	Eolic☐	Pedestals☐	< 25% ☐	20-50%☐
RATE: Very low☐	Low☐	Moderate☐	Strong☐		Severe☐	50-75%☐	>75%☐
PROPERTIES (*AT 30CMS*)	TEXTURE		COLOUR			pH	
REMARKS							

E. BIOMASS – DRY WEIGHT QUADRAT CLIPPING									
PASTURE SPECIE	RANK	CANOPY HEIGHT	PASTURE WEIGHT		RANK TALLY			WEIGHTED RANKS	% COMPOSITION
			FRESH WEIGHT (g)	DRY WEIGHT (g)	1	2	3		
GRAZING INTENSITY									

Very low grazing	<20%		GRAZING INDICATORS:
Low grazing	20-40%		
Moderate grazing	40-60%		
High grazing	60-80%		COMMENTS:
Severe grazing	>80%		