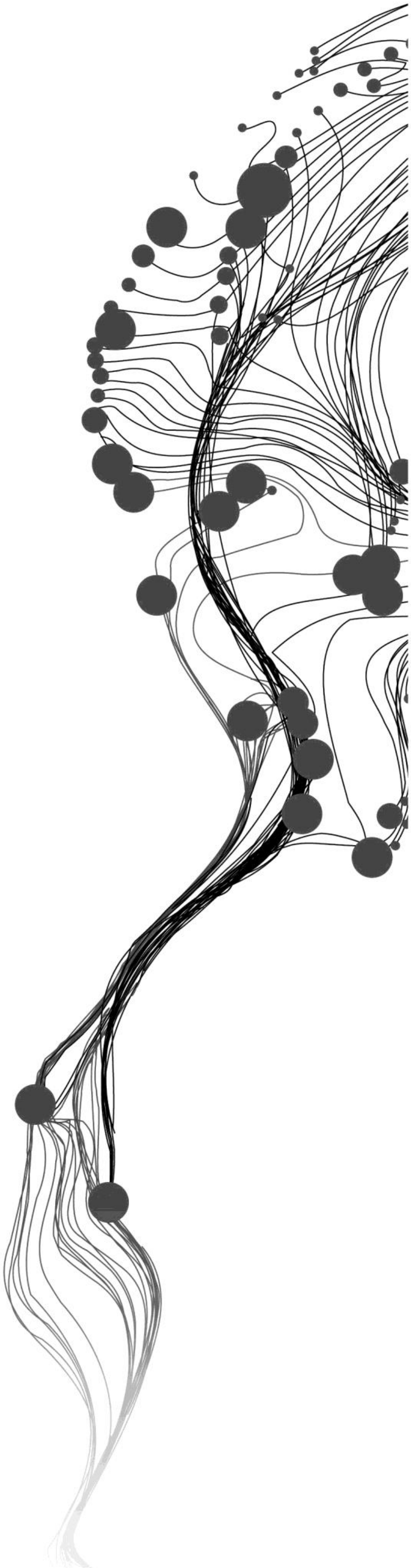
Assessment of the herbaceous above ground biomass gain on the fringe of Lake Naivasha, Kenya

BESTER TAWONA MUDERERI
February, 2012

SUPERVISORS:
Dr. Ir. T.A Groen
Dr. A. Voinov



Assessment of the herbaceous above ground biomass gain on the fringe of Lake Naivasha, Kenya

BESTER TAWONA MUDERERI

Enschede, The Netherlands, February, 2012

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. T.A Groen
Dr. A. Voinov

THESIS ASSESSMENT BOARD:
Dr. A.G. Toxopeus (Chair)
Drs. R. Becht - ITC (External Examiner)

Abstract

The herbaceous vegetation in the fringe zone of Lake Naivasha is vital for providing ecosystem services to both humans and animals inhabiting the fringe. Various fauna depend on the herbaceous vegetation for forage and shelter. The fringe zone area has declined significantly in the last 30 years mainly attributed to increases in agriculture, declines in the ground and lake water levels. These water level declines have been attributed to changes in the rainfall patterns in the upper catchment and increased abstraction for irrigation of farms surrounding the lake. The increase in the number of herbivores and loss of land to agriculture has increased susceptibility of the remaining herbaceous patches to overgrazing thus causing a reduction in productivity. Though there has been studies conducted on the fringe zone vegetation, most of them have focused on papyrus productivity. Measurement of productivity is critical as it indicates the net carbon assimilation from the atmosphere. Productivity can also explain land surface area conditions like change in water and nutrient availability among many other ecological processes. This research aimed to determine how the biomass gains varied spatially as influenced by changes in the fringe zone including water levels. To achieve this, an ASTER image was classified using a combination of automated isodata clustering and expert knowledge to separate the herbaceous classes from other natural and semi natural vegetation classes. The disk pasture meter was used to estimate the standing biomass for both June and September sampling periods. Biomass gain was the difference between the total biomass accumulated between the successive periods. This was compared between different herbaceous communities and inherent characteristics of the dominant species. The effect of lake level was analysed by establishing the effects of frequency of inundation and the ground water level on the biomass gained. A least squares regression analysis was conducted to determine if there were interaction effects in the explanatory variables for biomass gain. We found that 17 classes comprising three separate herbaceous classes could be discriminated successfully at 90 % overall accuracy and a kappa coefficient of 0.89. Expert knowledge and use of ancillary data in classification improves classification of satellite images as compared to using only pixel based approaches. The overall standing biomass in the fringe zone was 381 g/m^2 and an average decrease in biomass gain of 3.7g/m^2. The average grass height in the fringe zone of Lake Naivasha was 5.5cm with *Cynodon dactylon* and Kikuyu grass (*Pennicetum clandestinum*) dominating in most of the area especially in the north. There was no significant difference ($p>0.05$; CI=0.95) in the medians of the standing biomass in June and September using a Wilcoxon rank test, which could have been a result of season position of the study period. Although descriptive statistics showed variations among the different herbaceous groups, a one way analysis of variance established that there was no significant difference ($p>0.05$; CI=0.95) in biomass gain among the groups. Frequency of inundation has influence on biomass gain as evidenced by the one way ANOVA significant differences ($p<0.05$ among the frequency of inundation groups). A Tukey Kramer post hoc test established that patches inundated for 20-60% times a year and those less than 20% have significantly different productivity levels. The lake level and ground water are highly correlated for an estimated distance of 9.5 km from the lake centre. Ground water has no significant influence ($p>0.05$) on the biomass gain between the period June to September. Results of the test for interaction effects using regression analysis established that the grassland type, frequency of inundation bulk density and elevation are the best explanatory variables for biomass gain yielding an average R^2 of 0.4. The model can be improved in successive studies by increasing the accuracy of productivity estimation and include other variables like precipitation and grazing that seem to have very profound effects on herbaceous biomass gain. We conclude that lake water level has influence on the biomass gain of grasslands directly on the margin of the lake influenced by frequency of inundation. The fluctuations in water levels affected the herbaceous biomass gain however, further study on how level and period of inundation influence biomass gain could be further explored.

Acknowledgements

My sincere gratitude goes to the trinity God the Father, the son Jesus Christ and the Holy Spirit for awarding me this opportunity, for the gift of life, wisdom, endurance and comfort that saw me through the 18 months.

I am profoundly indebted to the Netherlands government through the NUFFIC program for granting me the scholarship to study in Europe without which the desire to obtain an MSc would have remained a dream.

My heartfelt gratitude goes to my supervisors Dr. Ir. Thomas Groen and Dr. Alexey Voinov for their support, ideas and critical reviews that provoked my intellectual ability and left me a better researcher. I would like to thank Mr Francis Muthoni for the unwavering support during fieldwork and brainstorming during data analysis and thesis writing. I also would like to thank Mr Vincent Odongo for the help with atmospheric correction and image classification. I extend the same gratitude to the Earth observation and integrated assessment approach for the governance of Lake (EOIA) project team for tolerating my presence and numerous questions. I remain indebted to the EOIA for the lake level, ground water and bathymetry data used in this research. I also thank the entire ITC staff and in particular the Natural Resource Management (NRM) department staff for setting out the premise that facilitated the success of this research.

To the great Zimbabwe community at ITC, I say thank you for friendship and unwavering support. May God bless you and provide for your future endeavours.

Last but not least, to the whole Mudereri family I am immensely and forever grateful for the emotional support and encouragement to soldier on, without which I couldn't have coped with the huge workload. Thank you for your prayers and love that kept me strong and for giving me the inspiration to climb the academic ladder. I appreciate your love.

TABLE OF CONTENTS

List of figures

List of tables

List of Acronyms

ANPP	Annual Net Primary productivity
ANOVA	Analysis Of Variance
ASTER	Advanced Space borne Thermal Emission and Reflection Radiometer
ASAR	Advanced Synthetic Aperture Radar
ATCOR	Atmospheric correction
CI	Confidence interval
DEM	Digital Elevation Model
DPM	Disk Pasture Meter
DNPP	Daily Net Primary Productivity
DN	Digital number
ETM	Enhanced Thematic Mapper
EOIA	Earth observation and integrated assessment approach for the governance of Lake Naivasha
GIS	Geographical Information System
GPS	Geographical Positioning System
IBP	International Biological Program
IF	Inundation Frequency
ITC	International institute of Geoinformation Science and Earth observation
ITCZ	Inter Tropical Convergence Zone
IUCN	International Union for the Conservation of Nature
KWS	Kenya Wildlife Services
LIDAR	Light Imaging Detection and Ranging
LNROA	Lake Naivasha Riparian Owners Association
LNRA	Lake Naivasha Riparian Association
NDVI	Normalised Difference Vegetation Index
m.a.s.l	meters above sea level
RMSE	Root Mean Square Error
R^2	R- Squared
SAR	Synthetic Aperture Radar
TOA	Top Of Atmosphere
UNEP	United Nations Environmental Program
VIF	Variable Inflation Factor
VNIR	Visible Near Infrared

1. INTRODUCTION

1.1. Rationale

In arid and semi-arid regions, grasslands play a crucial role in providing forage for domestic and wild herbivores. These grasslands display great spatial and temporal variability in net primary production mainly attributed to variations in light interception, water and nutrient availability (Brinkmann *et al.*, 2011; McNaughton, 1993). The herbaceous biomass produced within these rangelands influences the density, diversity and distribution of herbivores based on their feeding behaviours (Mutanga and Rugege, 2006). Measurements of herbaceous biomass and its production provide clear understanding for wildlife ranchers and managers on suitable habitats so as to calculate and implement sustainable carrying capacities to achieve optimal grazing on their farms. Throughout this thesis the term "herbaceous" shall refer to both grasses and forbs as they form the herbaceous layer in vegetation communities.

In Africa, the community to which the herbaceous unit primarily belongs to is known as the Savannah. The African savannah grasslands cover approximately half of the continent surface area (Groen *et al.*, 2011). These savannahs' are characterised by rolling grasslands with scattered trees and shrubs. They form the principal food base for the livestock and wildlife industry which brings foreign currency and food in most African countries. This vegetation community is also crucial for humans as it provides services like fibre, medicines, meat wool and milk. Apart from these, the savannah grasslands provides ecosystem services such as contributing to the genetic library, microclimate controls, nutrient cycling, organic matter production, soil conservation, and atmospheric carbon sequestration (Groen *et al.*, 2011; McNaughton *et al.*, 1988; Tainton, 1999).

Although most of these services do not have a market value, they are crucial for the day to day living of most African local human communities. However, these savannahs' are highly dynamic alternating between woodland and grassland following succession and influence of human activities (Roques *et al.*, 2001). This often poses a threat to the viability of these grassland communities in providing ecosystem services. The greatest threat to most grassland in Africa is the loss of surface coverage to agriculture. Due to increasing world population and demand for food most of the pristine grasslands have been degraded or fragmented for croplands. Other natural succession changes like bush encroachment pose also as major threats to the naturalness of the herbaceous unit (Roques *et al.*, 2001). There are also changes that are caused by water stress, nutrient suppression and over utilisation, such is the case in the fringe zone of Lake Naivasha (Bemigisha, 1998).

The fringe of Lake Naivasha is surrounded by private wildlife ranches/sanctuaries inhabited by various free ranging large and small herbivores. Tourists visit these ranches to view the animals bringing in foreign currency that plays a crucial role in sustaining the economy of Kenya. During the dry season, pastoral Maasaii migrate into the fringe zone of Lake Naivasha to graze their livestock in these areas as their traditional grazing areas are dry (Becht and Harper, 2002). However, the fringe zone area has been degraded significantly in the last 30 years mainly attributed to the dwindling lake levels, decline in ground water, overgrazing and agricultural expansion (Morrison and Harper, 2009). Increase in the number of flower farms coupled with increasing herbivore densities has propelled overgrazing in the remaining grasslands.

The concept of proper use emphasises on guarding against overgrazing (Keya, 1998). Overgrazing is critical in semi-arid climates like Lake Naivasha which experience fluctuating climatic conditions. The fringe zone of lake Naivasha, like most African rangelands is characterised by complex vegetation stands in terms of structure and species diversity. Overgrazing may lead to decline in species diversity while also promoting proliferation of unpalatable species at the cost of nutritive species as animals perform selective grazing (Metera *et al.*, 2010). If an area experiences continuous overgrazing it may lose its resilience and never recover from resultant species loss and reduced total productivity. As a result, this reduces the available forage for herbivores relying on the fringe. There is an information gap of how much biomass is available for consumption by these large herbivores within the fringe zone of Lake Naivasha. Furthermore, it is also not clear how much biomass is available to each different group of herbivore species based on their fodder preference and foraging behaviour. As a result, it is unknown which of the fringe areas would be more susceptible to overgrazing than others hence priority areas to protect from overgrazing are unknown.

Within most arid regions, fires are often rampant and continue to play a pivotal role in the conservation and management of these savannah grasslands (Zambatis *et al.*, 2006). If not properly monitored they have the potential to destroy the forage base for most wildlife and livestock. However, they can positively be used in the control of bush encroachment, tick control and destroying the accumulated dead material known as moribund which hinders production of new shoots (Roques *et al.*, 2001). Around lake Naivasha, random fires are often experienced in the fringe zone and herbaceous component acts as the primary fuel load (Bemigisha, 1998). In these rangelands the herbaceous component influences the occurrence, severity and magnitude of most of the fires. Thus a reliable and accurate measurement of the herbaceous biomass not only provides information on available forage quantity but also the fuel load. Thus this information is essential in designing effective grazing, fire combating and prescribed burning management plans (Groen *et al.*, 2011; Mutanga and Rugege, 2006).

Herbaceous vegetation productivity is largely driven by climate, availability of water and nutrients and herbivory (McNaughton, 1993). Other external factors such as species diversity and grassland management system also play key roles in determining productivity of vegetation communities (Keya, 1998; McNaughton *et al.*, 1988). This is more critical in dynamic ecosystems experiencing high temporal fluxes in resource availability such as water and different management structures as evidenced in the fringe of Lake Naivasha. Therefore, understanding all the vital factors that are limiting the fringe ecosystem primary productivity is paramount to the sustainable and proper monitoring and management. Other ecosystem components such as the vegetation structure species diversity and ultimately grazing resource for the benefit of the entire lake ecosystem in particular the herbivores inhabiting the fringe zone will sustainably be maintained. The spatial variations in vegetation productivity are not clearly known in the fringe zone. Most of the studies performed on vegetation productivity within the fringe zone were mainly centred on the productivity of papyrus (Boar, 2006; Boar *et al.*, 1999; Muthuri *et al.*, 1989). Thus within the herbaceous unit the areas of high, medium or low productivity remains obscure.

Monitoring and conservation of the herbaceous unit does not only benefit the herbivores but biodiversity and the entire ecosystem at large. Various avifauna utilise the lake as breeding and foraging habitat. The fringe vegetation also filters the inflowing and lake water thus improving water quality and reducing eutrophication. Excess nutrients from surrounding agricultural activities and pollutants from the local town cause algal blooms and aquatic growth. These are reduced from entering the lake waters through increased infiltration rates by shoreline plants (Becht *et al.*, 2005). Because of the perpetual decline of the fringe vegetation, especially papyrus over the past 30 years, there has been a significant increase in algal bloom resulting in the reduction of the euphotic depth of lake waters (Majozi, 2011). Vegetation especially the herbaceous component within the fringe controls erosion levels especially the shoreline erosion as the rooting system of the vegetation stabilises the soils (Metera *et al.*, 2010).

Lake water use has been at the centre of discussion in Naivasha since 1929 when the Lake Naivasha Riparian Owners Association (LNROA) was formed (Becht *et al.*, 2005) which was later on changed to Lake Naivasha Riparian Association (LNRA). The complexity of lake water management increased in the early 1980s when normal irrigation of fodder crops was converted to flower farming which demands more water (Becht and Harper, 2002; Becht *et al.*, 2005). These developments resulted in complex relationships between resource use and the resource availability. Most of the farms abstract groundwater, direct lake water and river water. Since the late 1970s, the lake levels tend to fluctuate in different periods and seasons causing shoreline shifts that varies from few meters to several kilometres (Becht and Harper, 2002). These fluctuations have been attributed to the variation in precipitation levels in the upper catchment and the level of water use with the lake area (Becht *et al.*, 2005). However, the understanding of effects of these dynamic changes in lake levels on the riparian herbaceous vegetation productivity remains vague. This dynamic fluctuation of lake levels in between season and due to abstraction results in different flooding effects. The herbaceous plants on the margins of the lake shore experience dynamic periods of flooding and dryness as a consequence of increasing and decreasing lake level. How the herbaceous vegetation productivity is influenced by these frequencies of inundation remains a question.

Various studies have attributed vegetation primary productivity to precipitation and vegetation dynamics like community species diversity among others (Brinkmann *et al.*, 2011; Stachová and Lepš, 2010). However, there are various other factors that can affect vegetation e.g. geographic location, plant herbivore interaction or topographical position (Keya, 1998; McNaughton *et al.*, 1988).The northern fringe of Lake Naivasha is flat compared to the southern and western sides that are steep and hilly. Different vegetation types occur in these locations mainly due to soil moisture content and topography (SPARVS Agency, 2008). This has effects on the vertical distance that the roots have to penetrate to reach the water table or soil moisture content within the rooting zone. On the contrary, the north shore is perhaps more susceptible to flooding hence influencing the productivity during the periods of inundation. It was therefore paramount to investigate the effect of water level as represented by vertical distance to the water table and frequency of inundation on the total productivity.

As established in literature, the lake water replenishes the ground water through seepage (Abdulahi, 1999; Morrison and Harper, 2009). When a clear relationship between lake level and ground water is established, conclusions can be drawn about how lake fluctuations may impact the herbaceous components on different sides of the lake at different topographic positions. A relationship between lake level and ground water was therefore sought in this research in order to investigate the influence of the lake water on the performance of the fringe herbaceous layer as the lake levels continue to dwindle causing the lake surface area to recede in the last 30 years.

In-situ measurement and monitoring of biomass and/or primary productivity is complex, expensive and has very low spatial and temporal coverage often resulting in inadequate and inconsistent estimates and measurements. The most common in-situ measurements of herbaceous biomass include the destructive sampling (clipping and weighing) or non-destructive sampling that use the Disk Pasture Meter (DPM) (Zambatis *et al.*, 2006). These methods are costly, labour intensive and area specific which limits their usability to limited areas. However these field based methods are very useful in producing accurate results which can be extrapolated to other areas with similar conditions using statistical models and remote sensing techniques. Combining field based estimates and remotely sensed parameters has been used successfully in herbaceous biomass estimation (Brinkmann *et al.*, 2011; Mutanga and Rugege, 2006).

Remote sensing broadly refers to indirect measurements of emitted electromagnetic energy from various earth materials using sensors or cameras (ITC, 2010). These earth material observables make it easier to deduce parameters from which we can make inference to draw conclusions about phenomena of different earth systems. The potential of remote sensing to capture high spatial and high temporal resolutions makes the technique even more attractive to use in scientific research. With this background, remote sensing facilitates for monitoring of broader landscapes than would field based methods. The frequency of monitoring plant communities is increased at low cost. Thus changes and effects of disturbance can be ameliorated quickly before they reach irreversible levels. This is key for the herbaceous vegetation which experiences high turnovers of succession.

A combination of field based estimates, remote sensing data GIS and statistical models therefore pose as a practical, less costly and more accurate way to monitor vegetation productivity and its dynamics. Remote sensing helps in facilitating identifying and characterisation of spatial positions and their trends. Satellite images are a series of digital numbers or reflectance values represented by singe pixels. Unless these individual pixels are correctly classified to a known class on the ground the images remain less useful. Thus, image classification stems as a basis for all analysis that can be conducted using satellite derived images. In many cases the key concern when using thematic maps derived from satellite imagery is its quality in operational applications (Foody, 2002). Therefore efforts have to be put to achieve the highest possible classification accuracy to deem the produced map a useful benchmark reference.

This thesis describes the use of high resolution freely available ASTER (Advanced Space borne Thermal Emission and Reflection Radiometer) 15 m resolution remote sensing imagery to accurately classify the riparian vegetation. Land cover classification stems as the primary step to analyse spatial distribution of land cover classes as observed from satellite imagery. Several natural vegetation classes exist in the fringe zone off Lake Naivasha from natural, semi natural and artificial surfaces. In order to achieve our broad objective classification was key in discriminating the key classes of interest which were within one subunit of the herbaceous vegetation. Other classes were classified for archiving although some also assisted with the associations with the herbaceous unit.

As a guideline to achieve the broad objective, research questions where drawn from the gaps that were obscure in knowledge of the system. These questions were used as a guideline to the drawing of achievable objectives using practical and reviewed methods. Geographical Information Systems (GIS), statistical regression models, interpolation models and sampling were used to cost effectively discover how the water level in Lake Naivasha influences the herbaceous vegetation productivity.

This research was part of a comprehensive project titled "Earth observation and integrated assessment approach for the governance of Lake Naivasha (EOIA)". The objectives and results obtained by this analysis are baseline to some of the broad objectives of the entire project.

1.2. Research Objectives

1.2.1. Broad objective

To investigate the role of hydrology on the herbaceous vegetation productivity in a tropical riparian zone.

1.2.2. Specific Objectives

- To identify and map fringe vegetation types along the fringe of lake Naivasha from satellite imagery
- To analyse standing above ground herbaceous biomass and biomass gain for a period between June to September 2011.
- To compare the effect of lake levels on above ground herbaceous biomass gain
- To find the best explanatory variables for biomass gain

1.3. Research questions

- How separable are herbaceous vegetation types/ classes when mapping using ASTER imagery in a riparian ecosystem?
- Are there significant differences (CI=95%) in herbaceous productivity over different vegetation classes?
- How does ground water influence the productivity of the herbaceous vegetation?
- How does frequency of inundation influence productivity of the herbaceous vegetation?
- What threshold distance does proximity to the lake influence herbaceous vegetation productivity?
- What are the major drivers of primary productivity and how do they influence the productivity?

1.4. Conceptual framework and system analysis

The diagram below depicts a simplification of the lake Naivasha fringe ecosystem. The boxes represent a particular stock, forcing function or phenomenon and the arrows indicate the direction of flow exchange of material or process. (Figure 1)

The precipitation in the upper catchment of the lake has the greatest influence to the lake water level. The precipitation directly falling on the lake area has very little influence on the lake water level because of its amount and evaporation rate hence the lake is very much dependent on the upper catchment activities and rainfall (Becht *et al.*, 2005; Bergner *et al.*, 2003). The water flows into the Lake mainly through the perennial river Malewa and the Lake in turn recharges the ground water aquifers. The flower farms draw water either directly from the rivers/ lake or pump from ground water (Figure 9). Depending on the levels there is a biflow of water either from the ground water to the lake or the opposite. The ground water mimics the lake level with a slight delay showing the correlation interflow of water between them (Abdulahi, 1999).

Soil moisture benefits from direct precipitation or by capillary action from ground water (Ahamed *et al.*, 2011). As the ground water table rises the soil moisture content is replenished depending on how far the ground water table is away from the surface. The opposite is true in the dry season or drought periods when the water levels are low the root zone soil moisture content for the herbaceous vegetation component will be very low. The soil moisture supplies water to herbaceous plants (Stachová and Lepš, 2010). When soil moisture is high these plants tend to draw more water hence they become more productive than in moisture suppressed environments or moisture suppressed moments such as the dry season or drought periods. The productivity of a particular plant species or a community is one of the important factors that determine the standing biomass in that area measured at any one time (Whitley *et* *al.*, 2011). There are other factors however, that would regulate or enhance these processes such as: nutrients, plant species or plant species composition and herbivory among others (Figure 1).

All above processes ultimately influences the quantity and quality of forage available for the herbivores. The species and number of herbivores will also have feedback on the NPP and available forage. They are however other factors that are not outlined in the diagram as they are beyond the focus of this research. Review of the interactions between the different flows are detailed in Chapter 2 of this document

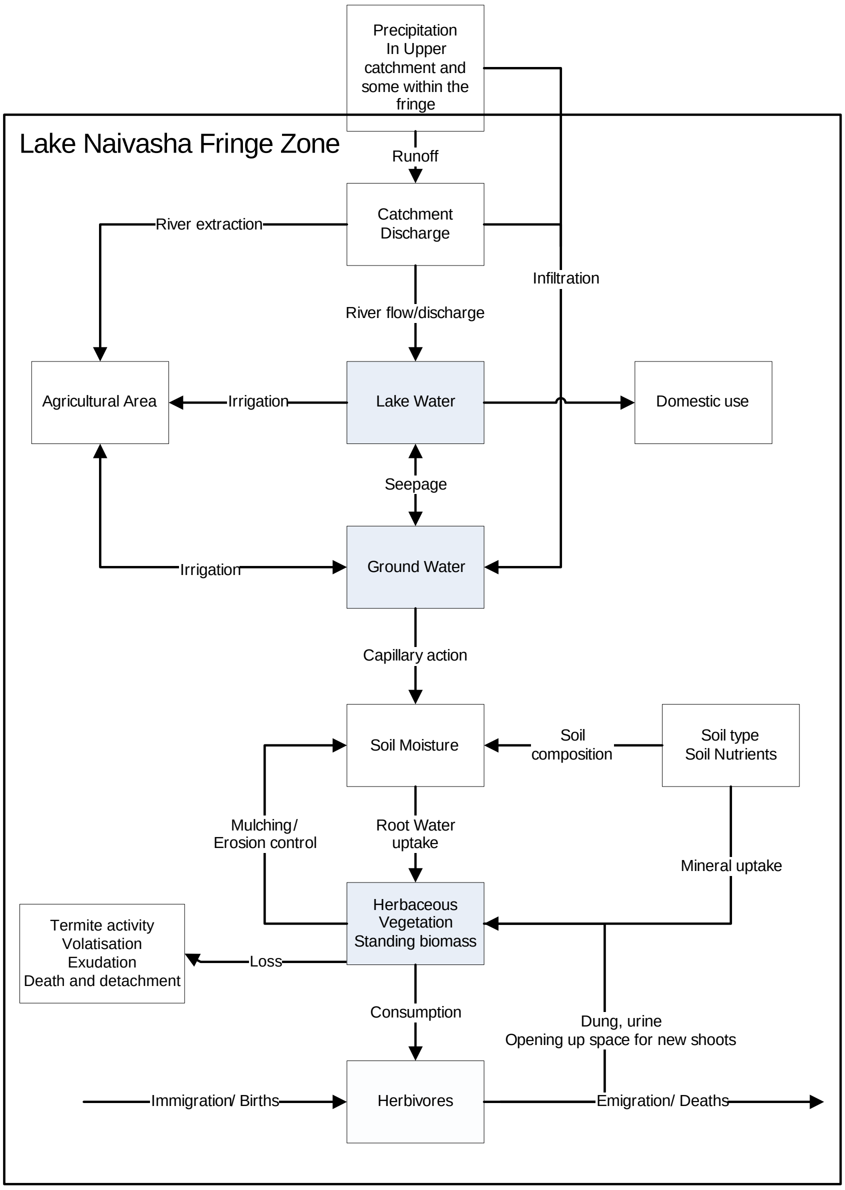


Figure 1: Conceptual diagram of the terrestrial biomass exchange flow within the study area ecosystem filled boxes are the areas addressed by this research

2. LITERATURE REVIEW

2.1. Land cover mapping and image classification

Identification of landscape characteristics facilitates effective monitoring of target components on the earth's surface. Spatially, land cover mapping initiates the identification of the land components. The Africover project by FAO established vegetation coverage using agreed mapping standards for the whole of Africa providing baseline for effective monitoring of vegetation and loss of biodiversity in Africa (FAO, 2003). This approach was largely based on the use of GIS and remote sensing to enumerate process and publish the data.

Remote sensing hinges on the principles of the abilities of earth bodies to absorb reflect and transmit at different wavelengths. In vegetation, chlorophyll plays a key role as it is the primary biophysical variable useful for biogeographical and ecological investigations. Within the mapping concept the reflections of different substances like water, soil and vegetation helps in the discernment of the different features from satellite imagery. For a very long time now use of remote sensing has proved effective to produce thematic land cover maps using different image classification techniques (ITC, 2010).

Classification of remotely sensed images remains at the centre of many scientists research as it forms the bases for environmental and socioeconomic applications. A variety of efforts have been channelled towards improving the accuracy of classification to create thematic maps from remote sensing data with relatively high accuracy (Foody, 2002; Zalazar, 2006). The overall accuracy of a classification is influenced by a variety of factors including the type of sensor, spatial resolution, atmospheric condition, availability of appropriate software and the nature of classifier among other numerous factors (Lu and Weng, 2007). In general classification can be grouped into supervised and unsupervised, parametric and nonparametric and pixel based or object oriented classifications.

There is a wide range of pixel classification methods, either per pixel based, sub pixel and object oriented but the commonly used are the traditional per pixel based which classify each individual pixel based on its inherent information. Per pixel based classification combines the spectral set of all the signatures in a training set. The popular traditional classifiers include Isodata clustering, K-means data clustering, mean distance and maximum likelihood. The widely used Maximum Likelihood classifier tool considers both the variances and covariance of the class signatures of a pixel to assign it to one class represented in the signature file (Schowengerdt, 1983). A class can be characterized by the mean vector and the covariance matrix assuming normality of the class sample

With these two qualities for each cell value, the statistical probability is used to determine membership of the cells to the class which it has the highest probability of belonging. This classification can also be considered as a parametric classification as it is based on the assumption of data normality. However per pixel based classifications does not take into consideration the heterogeneity of materials within a single pixel. Sub pixel classification in this instance comes in handy as it quantifies the variations in percentage content of a single pixel. Both sub-pixel and per pixel classification generate a lot of salt and pepper as the contextual position of a single pixel is most times not considered. Therefore the use of non-parametric classifiers becomes paramount. For fine resolution images per field classification can successfully be used

Previous research has shown that non parametric classification produces better results as compared to parametric classifiers as the contextual component of a pixel is considered (Lu and Weng, 2007). These classifiers like neural networks, support vector machines and decision trees do not base allocation of a pixel to a class on any statistical assumption. These methods have also been widely used and are promising in producing accurate results. A combination of classifiers has also been proven to improve classification results. These non-parametric classifiers include also the contextual based classifiers like object oriented, Markov random fields which address the intra class variations. These classifiers exploit the spectral information of a pixel in combination with the neighbouring pixels to improve the accuracy.

Although important results have been achieved with these automated classifiers they still cannot compare to human interpretation(Zhang and Zhu, 2011). These parametric and non-parametric classifiers have their strengths and weaknesses. Human interpretations do not only consider the contextual position of the object but would include the shape and the spatial relations between the regions. Use of multi-source classification has been also used in classifications with the goal of achieving high accuracy. The combination of pixel based, parametric or non-parametric information on spectra with ancillary data such as digital elevation models, topographic and geological maps and in some cases vegetation indices have been used successfully in post classification to improve the classification results of the automated systems(Lu and Weng, 2007). Using human interpretation alone is tedious, cumbersome, prone to bias from interpreter and sometimes impractical if done regional or continental scale. However a combination of expert knowledge with the automated systems outputs have produced successful results(Su *et al.*, 2011; Zhang and Zhu, 2011) Rule sets are produced which illustrate what can and what cannot be classified into a particular class using expert knowledge. Although the approach of using multiple sources produces highly accurate results there is a trade off with computing, time consumption and availability of data and software.

Following the review of classification methods by Lu and Weng, (2007) this research used a combination of automated unsupervised Isodata clustering, ancillary data and expert knowledge classification since the study area was small and there was an enormous bank of knowledge of the area. In addition, combining 2 classifiers and ancillary data is known to improve classification results. Apart from these facts, the study area comprised fragmented landscapes with a mixture of agriculture, semi natural and natural vegetation, the spectral based classifiers alone are insufficient (Jianwen and Bagan, 2005).

2.1.1. Validation of the results

The actual quality of the classification process must be checked. There is a key concern as land cover derived thematic maps are often judged to be of poor quality by users in the field (Foody, 2002). The most widely used method to assess the quality of the mapping output is use if the error matrix (Foody, 2002; ITC, 2010). An error matrix is computed to portray the misclassified pixels in categorical classes (Lu and Weng, 2007). Two errors are detected either error of commission or error of omission. The error of commission refers to the incorrectly classified pixels also referred to as user accuracy whereas the error of omission also referred to as producer accuracy is those points that are omitted in the interpretation result (ITC, 2010). From the error matrix an overall accuracy value and kappa statistics can be computed.

The overall accuracy is an addition of all the correctly classified divided by the total number of pixels that will have been classified. This gives an overall view of how well the classifier will have performed. However this value on its own does not show much. The Kappa statistic shows how much a classifier would have performed as compared to a random classification (Cohen, 1960). Kappa is the proportion of agreements after chance agreement has been removed. Thus in classification Kappa is a measure of overall agreement of the classified pixels tested using those that could have been correctly classified by a hypothetical probability classifier. The equation for Kappa is as given in Equation 1.

$$K = \frac{P_o - P_c}{1 - P_c} \qquad \text{Equation 1}$$

Source (Cohen, 1960)

Where K = Kappa coefficient, P_o = Observed P_c = hypothetical probability of chance

Therefore the two measures of accuracy overall accuracy and Kappa were used to validate the classification output

2.1.2. Aster Imagery

The Advanced Space borne Thermal Emission and Reflectance Radiometer (ASTER) is one of the remote sensing sensors generating satellite imagery at high spatial and spectral resolution. ASTER was launched on 18th December 1999 aboard Terra as a collaboration project between NASA and Japan's Ministry of International Trade and Industry. It operates on 14 channels producing bands that can be categorised into three spectral regions namely Visible and Near infrared (VNIR), shortwave infrared (SWIR) and Thermal Infrared (TIR). It is sun synchronous sensor at 750 Km altitude orbiting at 98.3 degrees from the equator crossing the equator at 10:30am (north and south) with a 16 day revisit cycle. The sensor is commonly used for monitoring vegetation patterns, land use, land surface temperature, cloud cover, sea ice, glaciers and snow cover. Lately the SWIR bands have malfunctioned and data can only be obtained from VNIR and TIR bands. Table 1 summarises the specifications of the ASTER sensor (SIC, 2012).

Table 1: ASTER Sensor specifications characteristics: source (SIC, 2012)

Instrument	**VNIR**	**SWIR**	**TIR**
Bands	1-3	4-9	10-14
Spatial Resolution	15m	30m	90m
Swath Width	60km	60km	60km
Cross Track Pointing	±318km(±24 deg)	±116km(±8.55 deg)	±116km(±8.55 deg)
Quantisation (bits)	8	8	12

2.1.3. Applications of ASTER imagery

Lu and Weng (2007) emphasise the need to take caution when choosing a sensor if good image classification results are to be achieved. The application of ASTER is reviewed here to motivate the rationale for its use in image classification for this research.

ASTER operations differ with objective and the channel under use. In their research to monitor wetlands along the western-Greek bird migration route Bortels *et al*., (2011) confirmed the usefulness of ASTER imagery to monitor small wetlands of approximately 0.5 ha. Another major use of ASTER imagery is the generation the DEM using the backward looking 3b band of 30m resolution. Fujisada *et al.,* (2011) used ASTER data to improve DEM generation. They reiterated the need for a sensor image to be useful in correctly identifying water bodies. In their research they concluded that the ASTER imagery can detect water bodies as small as 0.2 km^2.

Some of the studies done using ASTER imagery have been in terrestrial environmental management using the VNIR channels for image analysis. Zalazar, (2006) used ASTER imagery to compare per pixel based and object oriented classification using Welkoposka region in Poland as the study site. The results showed that an overall accuracy between 80-90% could be achieved when classifying using ASTER imagery.

Similar results were obtained in a research done in China using neural networks and ASTER imagery (Jianwen and Bagan, 2005) obtaining 95% overall land cover land use classification accuracy. These successes motivated the choice of an ASTER image for use in this study.

The facts that ASTER imagery was available and was acquired during the fieldwork period were the major motivating factors for its choice in this research. ASTER imagery is available free upon request from Global land Facility FTP website. What further stimulated its choice was the high spatial resolution 15m in the VNIR channel. This was sufficient to map land cover at a more fine resolution discriminating herbaceous classes which have similar spectral signatures. Since the study area was also small fine detail had to be enumerated baring in mind that the vegetation types were very similar and varied with few hundreds of meters

Furthermore, the sensor has a revisit time of 16 days therefore can facilitate comparisons between periods, which this facilitates the usability of the findings of this research. However the SWIR channel bands have malfunctioned which limits the use of ASTER to only VNIR and MIR channels.

2.2. Field based herbaceous biomass estimation methods

2.2.1. Use of the quadrants

A quadrant is a square, circle or rectangle made of either metal, plastic or wood with a known area. There are no clear cut advantages of a particular shape over the other but care has to be taken to be consistent in the use for the same research. The most commonly used quadrates are square for the purpose of isolating samples mostly in ecology and geography. Quadrants are usually $0.25m^2$ to $1m^2$ depending on purpose, sample size and the variability of the organisms being sampled. For long term studies it is required to use the same quadrate for comparability of results. The most common uses of quadrants are for plant species abundance, species distribution, assessments of succession research and in clipping experiments (Adams *et al.*, 2002).

To estimate herbaceous biomass quadrants can be used in combination with clipping and weighing (Zambatis *et al.*, 2006). All the herbaceous components within the quadrate are clipped with shears and weighed to obtain biomass. This is a laborious process that would be best used where data quality is more important than data quantity.

2.2.2. The disk pasture meter (DPM)

The disk pasture meter (DPM) is less labour intensive and more practical as compared to the quadrants clipping and weighing method. It is a very useful tool in the measurement of herbaceous standing crop in large areas. It has been in use since its early development in New Zealand (Phillips and Clarke, 1971). Its use has been extensive in South Africa mostly for veld management and research (Bransby and Tainton, 1977). The instrument is made of aluminium which comprises a bar with graduations fitted in a tube that is attached to a disk (Figure 2). It has an average mass of 1,5kg and uses the concept of compression by a disk falling from the same height. The bar is marked at 1.0 and 0.5 cm intervals with a maximum possible height of 60cm. The height readings are read from the top of the tube and should read 0 cm when placed on a flat surface. The common idea behind the DPM is the greater the standing height of the crop under the disk the greater the amount of biomass we would expect. However in some cases such as communities with tough culms or woody herbaceous components this hypothesis is defeated. To perform a measurement the tube is raised to the top of the bar and released to compress the sward. The measurement is taken from the top of the tube to the nearest 1 or 0.5cm. (Zambatis *et al.*, 2006)

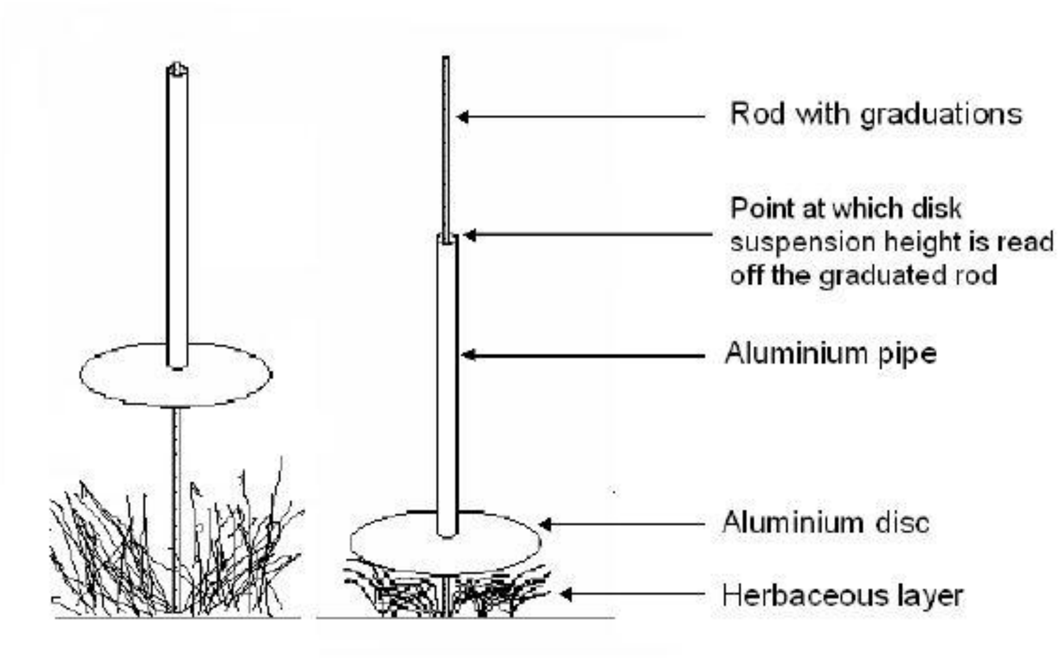


Figure 2: Disk pasture meter structure and components Source: (Dwyler, 2011)

Before any measurements are taken and related to biomass, the instrument must be calibrated first for each specific conditions in which it is to be used (Bransby and Tainton, 1977). Two methods for the calibration of the DPM are encountered in literature. The first and standard method (Bransby and Tainton, 1977) involves the use of a sleeve with a slightly larger diameter than the disk which is about 10cm high. Several measurements are taken by dropping the disk and placing the sleeve to enclose the compressed sward. All the grass within the sleeve is clipped and weighed to a height not greater than 3cm from the ground. The other method involves taking several samples with the DPM in a in a 4 X 4 m quadrate. The whole quadrate is clipped and the dry weight related to the average height. In both methods several DPM measurements are taken and usually linear regression is performed between the settling height and the dry weight although nonlinear relationships are also possible (Bransby and Tainton, 1977; Zambatis *et al.*, 2006).

Most of the DPM calibrations have been done in South Africa (Brockett, 1996; Trollope and Potgieter, 1986). These calibrated equations have been used by various scientists in the specific areas they have been calibrated for (Mutanga and Rugege, 2006). The accuracy of the equations is most likely to decrease when estimating outside the calibrated area. This is mainly due to variations in herbaceous structure, vegetation condition and species composition. This follows also different grass heights. Whilst calibrating for Kruger national park Zambatis *et. al.*, (2006) did not employ a single equation for the entire area but split the calibration to match grass heights ≤26cm and >26cm. This improves the estimation results as compared to when a single equation is used for the whole area in heterogeneous grassland communities.

2.3. Net primary productivity and Biomass gain

Net primary productivity (NPP) is the net flux of carbon into green plants from the atmosphere. It can be simply be defined as the net photosynthetic yield of vegetation per unit area for a specified time period (Scurlock *et al.*, 2002). NPP is a fundamental ecological variable as all life forms depend on the primary production. Within the ecosystems concepts primary production is the core energy source for all life forms on earth. The process is also essential as it indicates the net carbon assimilation from the atmosphere and has the potential to explain land surface area conditions like change in water and nutrient availability among other multitude ecological processes.

Although NPP is critical, it is a process which in itself can not directly be measured from the field like for example precipitation or temperature. For that reason surrogates like above and below ground biomass are measurements in combination with calculating algorithms are used (Scurlock *et al.*, 2002). Different methods for different vegetation types are encountered in literature. Proxies have been set that are used to estimate field NPP by use of different equations and algorithms to account for "True NPP". For grasslands, there are various methods suggested in literature but scientists have not come to an agreement on the standard procedure to use since there are lots of uncertainties with the estimates produced by all the methods (Lauenroth *et al.*, 2006). Most agree on the use of more complex equations to reduce underestimation. These uncertainties are mainly brought about by differences in grazing regimes, seasonality and geographic location thus influencing the phenology of the grasses. The most agreed general equation is given by Lauenroth *et al.,* (2006) as in below. This equation accounts for all the other seven methods discussed in the next paragraph.

$$NPP = \Delta B + \Delta H + \Delta E + \Delta D + \Delta V \qquad \text{Equation 2}$$

Where ΔB is change in biomass

ΔH is the quantity consumed by herbivores

ΔE is amount lost to exudation, sloughing and transfers to symbiosis and parasites

ΔD is amount lost to death and detachment

ΔV is the volatile losses of organic compounds, all these changes occurring between interval t =1 and t = 2

There are seven methods encountered in literature used to calculate in-situ estimates of NPP. The methods are summarised by Scurlock, *et al.*, (2002). These authors list the seven methods in order from simplest to complex. This same order was referred to in this document as Methods 1 to Method 7. These include in their order Peak live biomass thus only the peak biomass is regarded as NPP and does not account for losses. This method has the limitation of not accounting for simultaneous growth and death between two successive periods (Long *et al.*, 1989). This method is normally used where one or two measurements are available. Method 2 is similar to Method 1 as it integrates peak standing crop and dead material to the peak biomass measured at that sampling period. This method has the same disadvantages as Method 1 as they do not account for below ground biomass which is necessary since some of the production materials are transferred to the roots for storage and root growth.

Method 3 is one of the most widely used which is max-min live biomass thus the lowest biomass in a season deducted from peak biomass. Like its formerly mentioned methods it does not account for below ground biomass and is useful when only two measurements are available between two successive sampling periods. Method 4 was adopted and used by the International Biological Program (IBP) and now is accepted by United Nations Environmental program (UNEP) (Lauenroth *et al.*, 2006). It involves three intertwined assumptions. The first assumption is that all production is positive and if negative or zero increments are obtained there is no production. Thus the method also assumes that all increments in biomass equal production. The third assumption made by the IBP standard method is that death does not occur during productivity period so any decrease in measurements between sampling periods is random(Long *et al.*, 1989). Therefore only positive increments between successive sampling periods to obtain the annual increment are considered. Besides adding up positive increments of live standing biomass coinciding dead standing increments are also added. Although widely used this method was criticised for not considering relocation to below ground increments. When the products are transferred to below ground there is a possibility of double measurement without accounting for the transfer therefore may not be true increments but simply relocation (Long *et al.*, 1989). In most grasslands death of grasses and simultaneous growth occurs therefore although the method accounts for dead standing material it underestimates losses to decomposition (Scurlock *et al.*, 2002).

This method was later modified to also include all dead standing material and litter to yield Method 5 and 6. These are more complex methods as they compute monthly increments accounting for decomposition. The monthly losses are determined by the change in the dead material and the estimated disappearance of the dead matter assumed to have been lost to decomposition. Method 7 accounts for all the losses as in Equation 2 and is computed on a monthly bases. Method 7 however has limitations as the data required to determine NPP requires systematic measuring throughout the entire period which maybe a logistical challenge in most research.

These methods were further reviewed for their challenge and uncertainties by Lauenroth *et al.,* (2006). Their results showed that Method 3, 4 and 5 had the greatest uncertainties whereas 1 and 2 had the lowest and method 6 was intermediate. In addition using different methods with the same data produced different estimation results and the different methods had advantages and disadvantages. They thus concluded that the use of these methods depended solely on the geographical area and objective of research.

In most cases the geographical location and seasonality determines the choice of method. In most tropical climates like Kenya which experiences bimodal rainfall, there is high chances of under estimation of NPP if only the standing crop is considered or only the peak biomass as there are various changes that occur within a year. Lake Naivasha experiences bimodal seasonality and methods accounting for monthly computations would be the most relevant to account for the monthly increments. However within the context of this research only part of season was accounted for. This period was neither the peak lowest nor the peak highest biomass period. A surrogate quantity here termed "Biomass gain" was used to account for the in season productivity between the June and September 2011 period. Biomass gain shall be defined as the change in standing crop between two successive periods following Method 3. Only ΔB from Equation 2 shall be analysed. Thus no conclusions about Annual Net Primary Productivity (ANPP) or Below Ground Net Primary Productivity (BNPP) can be drawn from this research. However, in the broad EOIA project Mr Francis Muthoni shall look into ΔH and ΔD making estimation of ANPP more likely.

2.4. Factors explaining variations in herbaceous biomass/Net primary productivity

2.4.1. Soil type

There are a variety of soil characteristics that influence the presence and productivity of vegetation that are rooted on them. These characteristics include soil depth, soil pH, water holding capacity, soil nutrients and soil development. These characteristics influence the distribution of and type of vegetation growing on them. Most soils of volcanic origin like in parts of Naivasha can support deep rooted plants while some water dependent plants like papyrus exist on mud hydrosoils (Gaudet, 1977) Fine grained soils store more water for longer periods than course grained soils (Bemigisha, 2000) . Clay content was also reported to have a positive effect on the distribution of vegetation (Groen *et al.*, 2008). This influences the water holding capacity of an area hence influencing the soil moisture content. Most of the well-drained sandy soils occurring on slopes and those occurring on shallow water tables are normally water logged and are said to be poorly drained. This influences the aeration capability of a soil patch hence may have negative effects to the performance of vegetation existing on them. However vegetation types have adapted to the soil types hence a high correlation between soil type and vegetation type is expected.

The major influential property of a soil patch to NPP is the soil nutrients. The availability of the basic nutrients Nitrogen (N), phosphorus (P) and potassium (K) influences the productivity of the plants as they form the basic structure of plant tissue and are involved in crude protein synthesis (Whitley *et al.*, 2011). Soils derived from weathering of Cambrian rock or sedimentary rock are often nutrient suppressed (McNaughton *et al.*, 1988). The opposite is true of soils derived from alluvial deposits, volcanic ash and basic rock. The amount of salts and pH in a soil patch may influence how plants absorb nutrients. Sodic soils hinder some species of mineral uptake and growth. In this research the soil type data was coarse but was used to differentiate the soils based on amount of bulk density, clay content and soil cation exchange capacity as the major determinants of nutrients and water holding capacity with the notion that clay soils hold water for longer periods as compared to sandy soils.

2.4.2. Elevation, topographic position and slope

The elevation and topographic position and slope in them do not directly influence NPP. In a research done in Mongolia slope and aspect were important variables in explaining grass land productivity (Gong *et al.*, 2008). The authors found that the variation in water stress levels and N-P-K nutrients as influenced by the slope brought variations in productivity levels in different aspects and topographic position. They also found out that in C4 plants there was evidence of water stresses influenced by the slope level.

2.4.3. Vegetation type

The vegetation species dominating the community determines the overall productivity of a community. This phenomenon depends on the ability of a species to capture sunlight and convert the energy through photosynthesis. The major determinant is whether a plant follows the C_3 or C_4 photosynthetic pathway. Most of the grasses fall within the category C_4 and approximately 40% of the Gross Primary Production (GPP) comes from C_4 grasses (Whitley *et al.*, 2011). These plants photosynthesise faster in high temperature and light. The higher photosynthetic rate of C_4 plants plant species results in more dry matter production per unit time. In their research in Australia Whitley *et al.*, (2011) concluded that light interception rather than water availability had a greater influence to productivity.

Other factors include the resilience of a plant species to disturbance. The plant species that exhibit compensatory photosynthesis tend to produce more after herbivory which leads to the concept of increaser and decreaser plant species discussed in section 2.5.7 below on grazing. Ultimately it is the physiology of the dominant plant species that determines its productivity.

In other research on community productivity, papyrus showed that these species are highly productive species showing on average a productivity rate of 6000 gm^2/year (Boar, 2006; Muthuri *et al.*, 1989). In a research evaluating herbaceous productivity by Keya, (1998) in some grasslands in Kenya reports peak biomass standing of an average 1 500 kgDM/ha in ungrazed areas which translates to 150 gDM/m^2 which is far much less than that of papyrus. Therefore the vegetation type plays a role in the productivity of a community.

2.4.4. Water inundation

Different vegetation species respond differently to level, frequency and period of flooding in which the patch will be exposed. Gaudet, (1977), Boar, (2006) and Bemigisha, (2000) concurred that along the banks of Lake Naivasha, within 100 m from the water line, in years of low water level most of the dormant papyrus seeds thrived and germinated in the mud but decayed as water level increased. This has an effect on the productivity of the papyrus during the different periods. During flooding, flood resistant shrubs like *Conyza floribunda* have a higher chance of survival than all the other vegetation types (Bemigisha, 2000). Grasses also die during periods of flooding but would thrive due as the water recedes because of the high water table level. As the total surface area of the lake is increased by the water level increase the adjacent grassland area is reduced. This ultimately would influence ecosystem total productivity depending on lakeshore boundary.

Most research on inundation has been done to reduce the risk of natural disaster areas prone to flooding. The methods used have evolved significantly over time. Most of the inundation maps have been deduced from cross sectional 1D and 2D mesh resolution models. The incorporation of the topographic data and bathymetry improves results of such models. LIDAR data, SAR and DEM have been used in many inundation mapping approaches, however most topographic datasets are found lacking when it comes to the lake or river bathymetry (Cook and Merwade, 2009; Zhao and Li, 2012). Zhao and Li, (2012) used Advanced Synthetic Aperture Radar (ASAR) and Enhanced Thematic Mapper (ETM) images if two period including water sensitive vegetation indices to map flooding in Poyong lake in China.

Based on the available data, within the context of this research the lake boundary as derived from the lake bathymetry was used to assess the frequency of inundation within the lake fringe.

2.4.5. Lake water level

The lake water level in Naivasha depends on the upper catchment rainfall which is brought in mainly through river Malewa. Part of the water is diverted from the river for irrigation in some farms. The larger portion of the water in the lake is drawn through ground water extraction for irrigation as well. These parameters ultimately influence the amount of water in the lake. The lake water recharges the underground aquifers hence the higher the lake level, the higher the ground water table. Consequently this influences the soil moisture and in the end the standing biomass. Water abstraction has been reported to have negative consequences especially in dry years to the whole fringe ecosystem and a safe abstraction has to reached (Becht and Harper, 2002; Gaudet, 1977)

2.4.6. Precipitation

Lake Naivasha is located in a water scarce zone in a semi-arid climatic region. The mean annual precipitation is 600mm with two peak rainfall seasons in April and September due to the seasonal migration of the Inter Tropical convergence Zone (ITCZ) (Bergner *et al.*, 2003; Vincent *et al.*, 1979). However, Boar, (2006) suggested that significant changes in water level followed between-years variations corresponding to the El Nino-La Nina cycle instead of seasonal patterns. Bergner *et al.*, (2003) suggested that understanding of precipitation regimes helps our understanding of vegetation parameters. The root zone of the herbaceous layer is shallow to directly access water from the water table. Therefore the herbaceous component relies on the soil moisture zone for water extraction. Precipitation replenishes the top soil layer most often than capillary action from underground. Therefore precipitation and ground water are some of the key variables that were expected to influence the herbaceous biomass gain.

2.4.7. Grazing

Wildlife and livestock graze in the fringe of Lake Naivasha. Large mammals in Africa have long been known to influence the dynamics and productivity of the savannah grasslands (McNaughton *et al.*, 1988; Metera *et al.*, 2010). The effects of herbivore grazing on grasslands are highly debatable as they may be positive or negative. If overgrazed, areas are highly likely to lose species diversity and promote growth of unpalatable species (Milchunas and Lauenroth, 1993). On the other hand research shows a dynamic relationship between productivity and species diversity. In some cases there is a negative diversity to productivity relationships in areas where competition for resources is high. However in some cases the species pool size does not affect community productivity (Stachová and Lepš, 2010). Diversity often affects communities where the annual grasses are more than the perennials which are more resilient.

Grasses also adapt to overgrazing by changing the growth form from being tufted to rhizomatous to avoid being grazed. This normally converts tall grasslands to short grazing lawns (McNaughton, 1993). Ultimately, productivity is reduced as the grasses channel some of their productivity to production of anti-nutritional factors and some mechanical defoliation avoidance mechanisms.

Grass species type also influences how resilient grassland would be. Increaser grass species are species that increase with increased grazing pressure e.g. *Themeda triandra*. On the contrary decreaser species like *Sporobolus fimbriatus* decrease with increased grazing intensity (Danckwerts and Stuart-Hill, 1987; Van Oudtshoorn, 2004). In Southern Africa grasslands are sometimes broadly classified as sourveld and sweetveld species. A sourveld is a veld that is palatable for a short period of time and becomes unpalatable for the entire season due to accumulation of crude fibre whereas a sweetveld is palatable throughout the year (Tainton, 1999; Van Oudtshoorn, 2004). Thus, the sourveld species are grazed less than sweetveld. Although these terms may not hold in East African grasslands, it is paramount to note that the type of grassland community influences the productivity and or the standing biomass of grasslands and hence the grazing or vice versa.

However, as pointed out by Milchunas and Lauenroth, (1993) grazing may also increases the ANPP of a grassland. This concurred with the findings of McNaughton *et al.,* (1988) who found out that where animals defoliated they opened up for new shoots which increased production in their research in South Africa. The excretions and urine also increase mineral cycling as the correlation between productivity and dung quantity increases with seasonal progression (McNaughton *et al.*, 1988; Metera *et al.*, 2010).

McNaughton *et al.*,(1988) suggested that the influence of large herbivores cannot be ignored in grassland productivity research analysis. In the fringe of Lake Naivasha the most common grazers include the bulk feeders' e.g. Hippopotamus (*Hippopotamus amphibius*), Buffalos (*Syncerus caffer*), Zebras (*Equus burchelli*) and Wildebeest (*Connochaetes gnou*). Several methods to account for animal grazing have been suggested in literature. These include the use of cage and plot comparison or comparison of animal weight before and after grazing or stem count techniques, dung counts or transects animal counts (Keya, 1998). All these methods require long term measurements which were not possible considering the fieldwork period for this study. Apart from the time constraint the objectives of this study focused on the influence of water factors on productivity. Enclosures to account for the effect of herbivory have been set up in the study area however they did not form part of the scope of this study.

The other key variable is the effect of termite activity on the herbaceous unit. In some cases termites consume more standing crop than large herbivores. Their contribution to mineral cycling and removal of plant products can not be ignored (Okullo and Moe, 2012).

2.4.8. Management structures

Kenya Wildlife Services (KWS) is the body with the legal mandate to protect all wild species in Kenya. It works as a member of the Lake Naivasha Riparian Association (LNRA) which is a stakeholder association responsible for the sustainable management and resource utilisation monitoring in the fringe of Lake Naivasha. Other organisations and stakeholders which a part of LNRA include the International Union for the Conservation of Nature (IUCN), the fisherman's cooperatives, the local flower growers, Kenya Power Generating company (KenGen) and various other government departments (Becht *et al.*, 2005).

There are various grassland management structures that are implemented at different levels depending on objective, expertise and availability of resources. In the fringe of Lake Naivasha management is mainly follows the land use structures in the area. The main land uses that may have an impact on the productivity of the grasslands include crop farming, livestock ranching and wildlife ranching. Some farms may be a combination of all these land uses. These land uses may be privately or publicly owned properties. Most of the privately owned farms are fenced hence either wildlife or livestock are fenced out or fenced in. On the other hand publicly owned grasslands are unfenced.

Restriction of grazers with fences normally transforms African grasslands producing changes in both species composition and growth form both inside and outside the fenced area. Within the fenced area where animals are fenced out grasses channel most of their production in building strong and erect stem tissue. Whereas in unfenced grazed areas the grasses are more dwarfed with low growing internodes (McNaughton *et al.*, 1988). These grasses invest heavily in replacing their defoliated leaf tissue at the expense of other organs. The rate of productivity turnover therefore becomes high in unfenced areas as the grazers open up spaces for new shoots and plants develop compensatory behaviors in response to defoliation (Metera *et al.*, 2010).

2.4.9. Ground water Level

The ground water influences the soil moisture within the plants rooting zone hence may explain the species dominance and production as water is a key component in photosynthesis. Ground water depth was essential to assess the distance from the root zone to the water table. The assumption is that the ground water level would also influence the soil moisture content within the reach of the herbaceous vegetation root zone. The average rooting zone for the savannah herbaceous rooting depth is 0.5±0.1m and for the temperate grasslands the average rooting depth is 3.7±0.5m (Canadell *et al.*, 1996). This depth is shallow for the herbaceous vegetation to directly draw water from ground water table.

Research has been conducted in Lake Naivasha and models were developed for ground water (Abdulahi, 1999; Legese Reta, 2011). The lake has no surface outlet hence part of the water is lost through seepage and evaporation. The lake water feeds into the shallow water aquifer. It is then drained into the deep aquifer where it is thought to flow from Lake Naivasha to the terminal lakes Magadi and Elementeita taking over millions of years to reach (Becht *et al.*, 2005). This ground water seepage was measured by Åse *et al.,* (1986)to be 45-50 X 10^6 m3/month and Gitonga, (1999) long term water balance outflow of 4.6 X 10^6 m3/month.

These outflows are now also greatly influenced by the amount of water abstraction. Water abstraction in Naivasha is one of the least documented variables within the available data for the lake (Legese Reta, 2011). Much of the ground water level in the North has been attributed to abstraction more than the level of the lake. Becht and Harper, (2002) estimated an annual abstraction rate of 60 X 10^6m3/year between the period 1983 – 1998. This has since changed as the number of flower farms has significantly increased as of the year 2000.

The distance threshold in which water influences the ground water table and ultimately the standing biomass was assessed. Using this variable, two types of phenomenon were derived. The first was to establish the threshold distance in which the lake level has on the ground water and the second was the interpolation of the ground water depth

2.5. Ground water interpolation

To evaluate the effect of ground water a simple model had to be developed. This model would use prehistoric lake level and ground water level data of the sampled points to interpolate for the whole study area. This would facilitate the modelling of the distance that the herbaceous plants' roots would have to penetrate to reach for water.

There are several interpolation techniques encountered in literature. In general interpolation techniques are models that estimate values where no samples were collected (ESRI, 2011). Most of these techniques require that certain conditions be met before interpolation can be conducted. Kriging poses as a more flexible approach with little or no assumptions or no requirement of pre-known parameters. However a thorough knowledge of the nature of the data and how it was sampled is key in the choice of kriging parameters (Isaaks and Mohan Srivastava, 1989). The underlying principles and review of the model application are detailed in the sections that follow.

2.5.1. Kriging

Kriging is an interpolator that predicts exact or smoothened surfaces depending on the calculated statistical error models. It is a technique that posits a statistical model given a response variable at a given location to estimate an underlying surface of unmeasured positions (Isaaks and Mohan Srivastava, 1989). All interpolation methods achieve this by assigning unknown locations values based on known surrounding values. This allows predictions to be modelled based on the distance between two known measured variables which are spatially auto correlated as shown in Equation 3 which shows the basic kriging equation

$$\hat{Z}(S_0) = \sum_{i=1}^{N} \lambda_i \, Z(S_i) \qquad \text{Equation 3}$$

Where $Z(S_i)$ = measured value at the i^{th} location
λ = an unknown weight for the measured value at the i^{th} location
S_0 = the predicted location
N = number of measured values

Source (ESRI, 2011)

Kriging relies on the notion of spatial autocorrelation in contrast to classic statistics which assumes independence of individual points (ESRI, 2011).Spatial autocorrelation is the tendency of two variables to be related on assumption that things more close together are more related than those far away or further apart (ESRI, 2011; Isaaks and Mohan Srivastava, 1989; Naimi *et al.*, 2011; Odland, 1988). Thus geostatistics plays a role in modelling correlation as a function of distance as distances can be computed and correlation decay modelled as a function of distance. Thus the basic understanding of the technique is that it is regression using the spatial coordinates as the explanatory variables (ESRI, 2011).

The weightings assigned to unknown areas for the better approximation of the nature of spatial autocorrelation are done through the use of semi-variance. This is calculated based on the separation distance between two paired points as the squared difference between the values at those two points. One single point distance is calculated to each measured point but for computational purposes bins or lags are assumed to ease computation. These values in each bin are plotted against distance resulting in a semi variogram. Figure 3 shows a typical variogram showing the nugget which is the position where the model intercepts y and the range distance where the model flattens out depicting end of significant autocorrelation and the sill which is the position that the model attains its range on the y-axis. The partial sill is the distance between the sill and the nugget.

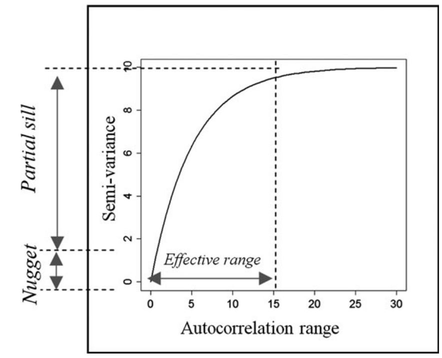


Figure 3: The exponential variogram model for range = 15, nugget = 0 and partial sill = 10. Source: (Naimi, *et al.*, 2011)

This method has been used in many interpolation research including species distribution modelling (Naimi *et al.*, 2011), in interpolation of above ground biomass (Mutanga and Rugege, 2006) interpolation of pollution and heavy metals and soil properties (Guo *et al.*, 2007; Li *et al.*, 2006; Shad *et al.*, 2009; Xavier, 2006). Within this research context kriging was used for interpolate ground water level from known borehole measurements. Research has been done in the study area and models to predict ground water have also been developed

2.5.2. Ordinary kriging versus Universal kriging for ground water interpolation

Ordinary kriging is the commonly used interpolator which assumes that a constant mean is unknown as compared to universal kriging which assumes an underlying trend in data being interpolated (Isaaks and Mohan Srivastava, 1989). It therefore requires thorough understanding of the data and the sampling design used to be able to use universal kriging. Ordinary kriging was used to interpolate the average difference between lake level and water table on the assumption of an unknown mean. The ability of a cross validation tool to assess the performance of the model further motivated the use of the ordinary kriging tool in ArcGIS geostatistical analyst tool.

2.6. Statistical Methods to be used for data analysis.

2.6.1. Normality test

Data is said to follow a normal distribution if it follows a Gaussian function which is bell shaped given by the following function:

$$f(x) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \qquad \text{Equation 4}$$

Where μ = mean
σ^2 = variance

Source: (Moore and McCabe, 1998)

The standard normal distribution has μ =0 and σ^2 = 1 where μ = mean and σ^2 is variance measuring the spread of data of variation from the mean. Data which follows a normal distribution is symmetrical around the mean. Although this is the case in reality some data may be far from the mean due to experimental error, thus the normal distribution is not robust to outliers (Moore, 2001).

The normal distribution is useful in a variety of applications and fields including natural and social sciences. Many statistical parameters, statistics and estimators base their assumption on data that would not violate the normality assumption hence the need to always test data for normality as one of the steps in choosing the proper analysis statistic. Statistics that work on data with certain distribution assumptions are generally called parametric and those that do not assume any distribution are called nonparametric.

There are various methods in literature to test data for normality and the numbers can be overwhelming. The common ones test data for skewness and kurtosis, distribution or function and the linear relationship between the variable and the standard normal Z-test. One of the commonest and simple to understand test is the Kolmogorov-Smirnov test. It compares data to the expected normal distribution basing the p-value on the largest discrepancy. This test has been reviewed by various authors as too simple and weak in discriminating (Henderson, 2006). The Shapiro-Wilk and the D'Agostino-Pearson Omnibus test all test the null hypothesis that the data is drawn from a normal distribution. However the Shapiro-Wilk test does not work properly if the data contains multiple similar values. This is not so with the D'Agostino-Pearson Omnibus test. The D'Agostino-Pearson Omnibus test initially analyses the data for skewness and kurtosis and then calculates the deviations from the expected Gaussian distribution. The p-value is then computed from the discrepancies sum of squares. As with many other tests it is recommended not to use this test for samples below 20.

In this research the Shapiro-Wilk was used as the data does not contain any similar numbers and the sample size is greater than 20. In a review by Henderson, (2006) of normality testing methods the Shapiro Wilk test was one of the best among other normality tests methods. Visual analysis was done using a normal quartile plot also known as QQ plot. A QQ plot is a graphical representation for a comparison of two probability distribution functions by plotting their quartiles against each other. The data is said to be normal if the points lie on or very close to the straight line y=x. The opposite is true for non-normal data that is if the points show a systematic deviation from the straight line (Moore and McCabe, 1998). Thus this provides a more robust way of showing a goodness of fit as compared to the numerical summary or histogram. The QQ plot would also show clearly the outliers and where they lie within the distribution which appear as points far away from the overall plot pattern.

2.6.2. Analysis of Variance (ANOVA) and *t*- test

Analysis of variance (ANOVA) in its simplest form computes the differences between means of two or more groups or populations (Moore and McCabe, 1998). It assesses whether these means are statistically significantly different. A two sample *t*-test on the other hand compares significant differences between only two samples thus a one way ANOVA is a multiple two sample *t*-test. It uses an F statistic to test the null hypothesis that the means are equal. If the p-value is not significant ($p< \alpha$) then the null hypothesis is rejected. A post hoc test to see which of the means a significantly different from the other groups or which pair or groups are a significantly different would have to be performed (Moore and McCabe, 1998). Normally a Tukey Kramer range test follows a one way ANOVA. It simultaneously performs multiple *t*-tests which compare every group mean and to other group means controlling chances of committing a Type I error which would be inevitable if normal *t*-testing is performed.

2.7. Regression analysis

Regression analysis is a modelling and analysing technique which normally gives the conditional expectation of a dependent variable given the independent variable (Snee, 1977). It can be used in predicting and forecasting the values of y given the values of x. The measure widely used in regression analysis is the coefficient of variation (R^2). The R^2 is the "fraction of the variation of the values of y that is explained by the least squares regression of y on x"(Moore and McCabe, 1998). It is given by dividing the variances of the predicted values by the values observed. Thus the R^2 value tells how much the predicted values are explained by the regression model. This value differs from the correlation coefficient "r" in that "r" just tells the strength and direction of relationship between x and y.

For regression there has to be a predefined independent variable (x) which explains the dependent variable (y). The regression line produced within a scatter plot of these two variables can be used to predict the values of y given x. However the values of y may lie far from the prediction line hence the need to minimise the prediction error using the least squares regression approach (Moore and McCabe, 1998). Residuals are a computation of the difference between the observed measurements and the predicted. These residuals have a mean value of zero. When residuals are plotted against the explanatory variable they can tell the overall pattern of the data or the relationship of x and y. This may be useful in determining the fact that the data or relationship is non-linear. The simple least square regression equation is given below

$$\hat{y} = a + bx \qquad \text{Equation 5}$$

With slope

$$b = r\,\frac{Sy}{Sx} \qquad \text{Equation 6}$$

And intercept

$$a = \bar{y} - b\bar{x} \qquad \text{Equation 7}$$

Source (Moore and McCabe, 1998)

Where $\hat{y}$ is the predicted value, a is the y intercept, b is the constant, Sy is the variance in the predicted, Sx is the variance in the observed, r =correlation between x and y and $\bar{y}$ and $\bar{x}$ are the means of y and x respectively

In some cases many variables (x_1, x_2…..x_n) are used to explain a single dependent variable (y). This is normally called multiple regression. They follow the same principles however the basic equation changes with addition of several values of x with varying coefficients as given in the Equation 10. Care is taken to avoid collinearity by measuring the variable inflation factors (VIFs) and or modelling regression of pairs of x variables and assessing their R^2 value to find correlation. These variables may become redundant if they are used in the same multiple regression or may cause over parameterisation thereby inflating the R^2 of the regression.

$$y = b_0 + b_1x_1 + b_2x_2 \dots\dots + b_nx_n \qquad \text{Equation 8}$$

Where: b_0 = y-intercept

$b_1 \dots \dots b_n$ = coefficients of the variables included in the multiple regression

This method was used in finding the best explanatory variable to explain the variable "biomass gain". The explanatory variables included in the regression equation were ground water level, frequency of inundation, elevation, slope, bulk density, soil cation exchange capacity and the herbaceous land cover type.

3. STUDY AREA AND DATASETS

3.1. Study Area

3.1.1. Location and extent

The study area is one of Africa's important fresh water ecosystems lying in a dry water-scarce zone (Harper *et al.*, 2002). The study was carried out within the 10km fringe radius from the shoreline of Lake Naivasha in Kenya. The lake (0^0 45'S, 36^0, 26'E) is located about 80km Northwest of Nairobi in the Eastern Rift valley at an altitude of 1890 m.a.s.l. It is the second largest lake in Kenya after Lake Victoria its surface area fluctuating between 120 and 180km^2. It receives water from a 30km^2 basin and contains approximately 0.85km^3 of water (Bergner *et al.*, 2003). It is shallow water basin Lake with an average depth of 5m with maximum depth at Hippo point being 7m. The surrounding island lakes Oloiden and Crater Lake exceed 20m depth. The northern shore is the shallowest, most colonised by vegetation and all the rivers enter in its direction. The lake is bound to south east by the Olkaria and Longonot volcanic mountains to the south and to the east by the Kinangop Plateau the west by the Mau Escarpment and to the north and part of north east by the Eburu volcanic pile

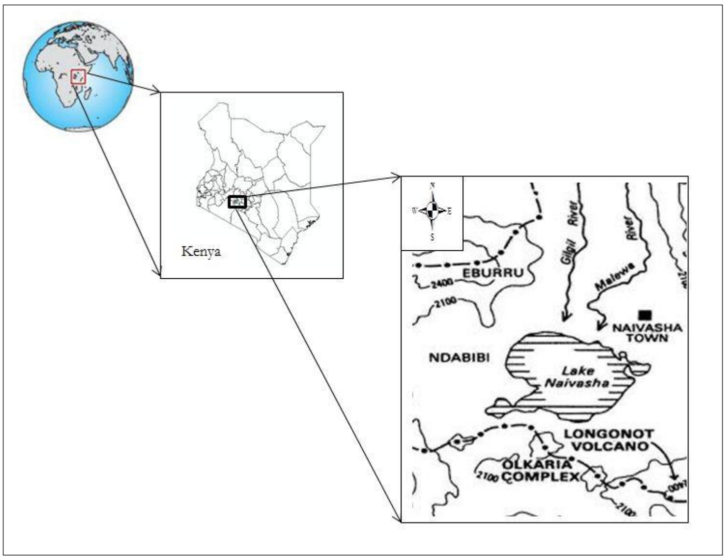


Figure 4: Study area, Lake Naivasha Kenya modified from SPARVS Agency, (2008)

3.1.2. Climatic conditions

Lake Naivasha is located in a semi-arid climatic region. The mean annual precipitation is 600mm with two peak rainfall seasons in April (highest) and September due to the seasonal migration of the ITCZ. It experiences short rains in the period March to May and long rains from September to December. The mean annual temperature is approximately 25^0 C and the maximum 30^0 C. The period, December to March is the hottest and July the coldest month (23^0 C).

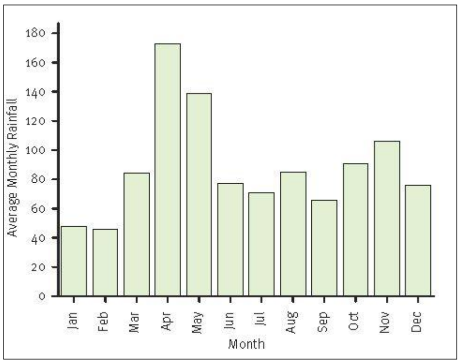


Figure 5: Average monthly rainfall for the study area: Source (Becht *et al.*, 2005)

3.1.3. Flora and fauna in the fringe of Lake Naivasha

There are various biodiversity components relying on the fringe vegetation for forage and shelter. The ecosystem around the fringe of Lake Naivasha consists mostly of highly productive emergent macrophytes (Muthuri *et al.*, 1989), submerged or floating. Since 1988, Water Hyacinth (*Eichhornia crasspies*) has dominated other species forming dense mats in the water. Various other species and vegetation was documented by Gaudet (1977). The total number of species as recorded by (Gaudet, 1977) was 108 plant species in a primary succession sequence from lake edge to dry land . This number has since gone down as reported by Adams et al., (2002). Figure 6 shows a general view of the vegetation communities present within the fringe from the lake to dry land.

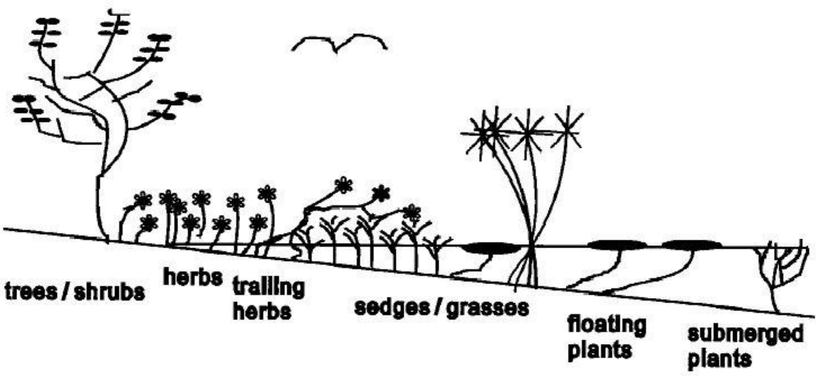


Figure 6: Naivasha vegetation gradient from lake shore to dry land source: (Adams *et al.*, 2002)

The mono-specific stands of giant Papyrus dominate the fringe. These papyrus fringes are filled with Hippopotamus (*Hippopotamus amphibius*), and hundreds of bird species foraging and sheltering in their cover. In addition, the riparian grasslands are grazed by Buffalos (*Syncerus caffer*), Zebras (*Equus burchelli*), Wildebeest (*Connochaetes gnou*) and other various antelopes throughout the year but mostly during the dry season. Apart from these, the ecosystem also consists of dense riparian Acacia forest mainly *Acacia xanthophloea* characterised with various browsers such as Giraffes (*Giraffa camelopardalis*) and Bush bucks (*Tragelaphus sylvaticus*) (Gaudet, 1977). There are also vast stands of most common shrub in the fringe zone commonly known as Leleshwa (*Tarchonathus camphorutus*). Concurrently, during the dry season the surrounding hills and valley bottom produce grass mostly *Cynodon* species and Kikuyu grass (*Pennicetum clandestinum)*. The pastoral Massai bring their herds within the fringe zone for dry season grazing. This increases demand for forage on the riparian vegetation (Harper and Mavuti, 2004; Mavuti and Harper, 2006).

3.1.4. Lake level

Considerable amounts of the water in the lake are used for irrigation of flowers, horticulture and hydroelectric power generation (Becht and Harper, 2002). These are highly water demanding industries which play a key role in the socioeconomic development of Naivasha and Kenya at large. The flower industry began in the early 1980s and has grown significantly over the years. This has improved the standard of living of locals and the economy of Kenya. On the other hand, the increase in demand for water has complicated the enforcement to stay under the safe abstraction threshold due to the perpetual increase in the number of farms.

The current annual abstraction is six times more than the safe threshold determined by Becht and Harper, (2002). The water levels in Naivasha can change by several meters in just a few months causing a shift in the shoreline of several kilometres subsequently reducing fringe ground water level (Becht and Harper, 2002).

3.1.5. Geology and soils

The study area lies within the “Gregory valley” which is part of the Great Rift Valley. The geology is mainly volcanic rocks with lacustrine deposits that have undergone several tectonic processes (SPARVS Agency, 2008).

3.2. Primary and Secondary data

A field work campaign was conducted in the period 13 September 2011 to 03 October 2011 and the following primary and secondary data was collected.

3.2.1. Ground truth data

A total of 133 ground truth points were sampled and saved in an excel database. The data collected includes the vegetation types observed in the field and their estimated canopy coverage in percentage. The data also contains elevation, photo numbers corresponding to field plot surveyed, the final classification class assigned to each vegetation combination and the final class code used for accuracy assessment. Details of how the data was collected are as compiled in the Methodology chapter.

3.2.2. Herbaceous biomass measurements

Between June to September 2011, a total of 47 plots were sampled using the DPM. Of these 47 plots 38 plots had been sampled earlier in June in an on-going EOIA study. Thus, 9 more plots were added in September to achieve a better representation of riparian vegetation. Due to the increase in the water level some of the plots sampled in June were not accessible in September as they were submerged in water. In addition some of the newly proposed plots were inaccessible due to security restrictions. Therefore, a total of 34 plots with both June and September measurements were used to calculate the biomass gain. Figure 7 shows how the DPM was used to enumerate the herbaceous standing biomass in the field

Figure 7: Use of the DPM to measure herbaceous height in field

3.2.3. Lake water level and Ground water level data

Naivasha daily lake water level data was used for the lake level information. The levels were based on the laws and Flintoft system done circa 1950. This is the official system which is the new government gauge at Yatch club read at a zero value of 1885.26 m.a.s.l (Vaughan, 1998). These daily measurements were available for the period February 1997 to November 2011. This data was sufficient to cater and equate with the ground water data that was available within the ITC database that corresponded with the lake level at that instance. The data was averaged to obtain monthly lake levels which were equated to corresponding groundwater data for analysis and modelling of ground water.

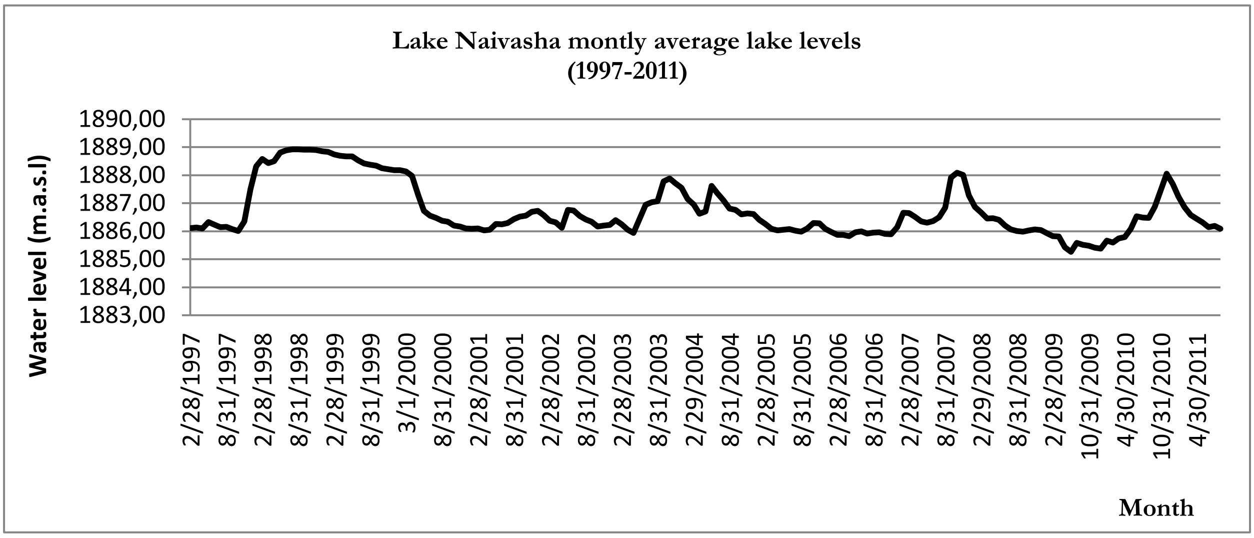


Figure 8: Historical lake level monthly averages derived from daily lake level measurements for the period February 1997 and September 2011

The ground water data was mainly obtained from data collected by ITC MSc students from previous years (Abdulahi, 1999; Kibona, 2000; Legese Reta, 2011; Morgan, 1998; Nabide, 2002; Ramirez Hernandez, 1999; Yihdego, 2005). The ground water data used was from 156 boreholes distributed around the lake around the lake. The ground water data was checked for spatial accuracy and all the boreholes that fell in the water or outside the study area were not considered in the analysis. The data contained the measured elevation at borehole and the measured water table depth for each corresponding date. The boreholes were sampled at different periods, hence there was no consistency in the measuring periods and some boreholes had more records than others.

Figure 9 shows the distribution of the boreholes that were used in the analysis. It also shows the sources of water for irrigation. Most of the northern farms use underground water whereas those on the south draw water from the lake. Becht, *et al.,* (2005) cited that the soils in the south increase in alkalinity in the ground water due to the volcanic nature of the geology hence the limitation of using ground water effectively in the south.

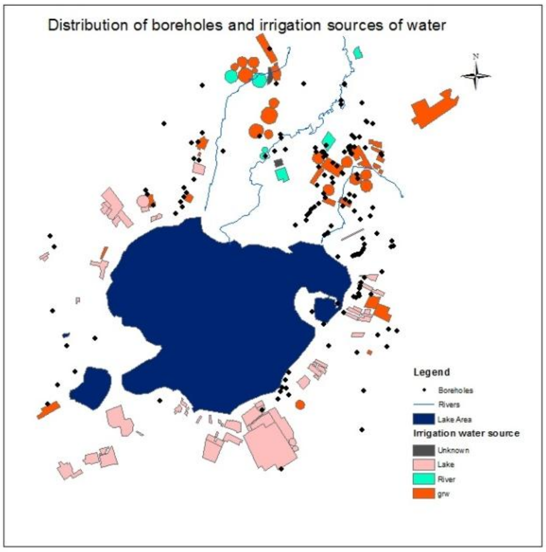


Figure 9: Distribution of boreholes used for analysis in this research and the sources of water for farms as of the year 2006

3.2.4. Soil type data

A coarse resolution map of 1:100 000 for whole of Kenya was used for the Naivasha soils data. The map was obtained freely from http://www.reading.ac.uk/GEFSOC/ (Batjes and Gicheru, 2004). The data used comprised different soil parameters including soil constituencies like clay, sand and silt content. This data was extracted per every borehole point and was used in the multiple regression analysis.

3.2.5. Digital Elevation model (DEM)

A 30m resolution Digital elevation model was derived from the 3b backward looking band of the ASTER image of 22 September 2011. This DEM was used to derive the slope elevation and aspect data for the study area. The DEM was tested for accuracy by correlating with the field measured elevations. The DEM was correlated with the ground truth measured elevations with a Garmin GPS with ~3m accuracy and then also correlated with the borehole measured elevations.

The measured elevations during the September 2011 ground truth and biomass gain survey plots were highly correlated with the DEM elevations (R^2 =0.97). For all elevation analysis for the biomass gain the DEM was not adjusted as it was highly correlated to the measured values. The DEM was used to derive the plot elevation and slope for use in the regression analysis to establish the best explanatory variables for biomass gain.

For the ground water analysis, the borehole elevation measurements were initially correlated with the DEM elevations and were also highly correlated (R^2 =0.91) was obtained. The difference between the DEM elevation and the measured borehole elevations was computed. The sensor derived DEM elevation were lower than the field measured elevation by a mean difference of 8 m and a mode of 10 m as obtained from descriptive statistics of all the 156 boreholes used in the analysis. A correction of 9 m was done to the DEM to improve the correlation. All the boreholes that had elevation difference of more than 20 m were eliminated from the analysis. It was assumed that the differences were brought about by field measurement errors since the data were collected by different individuals in different years. The other possible explanation to these differences could have been as a result of low accuracy GPS devices used to collect the data. After correction a mean difference of 0.78m, a mode of 1 m a standard error of 0.45 m and an R^2 of 0.95 was obtained. The correlations are as shown Figure 10.

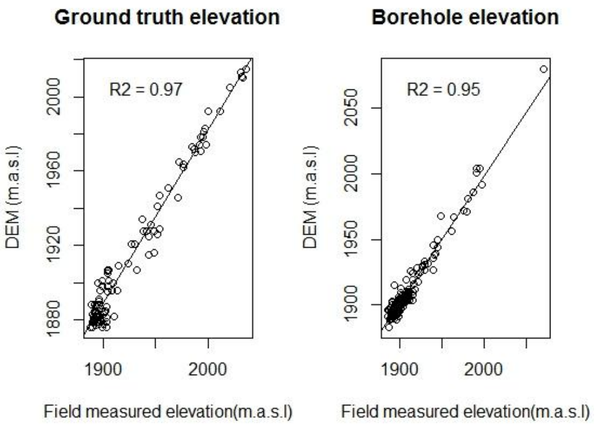


Figure 10: Correlation of DEM to field measured elevation

3.2.6. Bathymetry data

The 1957 Lake Naivasha bathymetry was used to derive the boundaries of the lake shoreline at different lake levels. The data was available as isolines. These isolines were converted to polygons which were used as the lake perimeter in all the analysis. The lake perimeter isolines available were for the lake levels 1884, 1887, 1888 and 1890 m.a.s.l. These sufficed the purpose as the lake levels period 1997-2011 (period used in the study) were within the range 1884-1890 m.a.s.l. These were used to derive the distance from the lake shoreline and were also used in determining the frequency of inundation.

3.2.7. Image processing

An ASTER image of the 22nd of September 2011 was acquired at level L1B. The VNIR bands were stacked and imported into ERDAS Imagine format. The images were then clipped to the area of interest within the radius of the lake. The subset image was then georeferenced using topographic maps of Naivasha. The image was then further corrected for atmospheric noise using ATCOR extention in Erdas. The DN values were converted to spectral radiances and then the spectral radiances were finally converted to Top Of Atmosphere (TOA) reflectances following the procedure by Yüksel *et al.*, (2008). The parameters used for atmospheric correction were obtained from the ASTER header file.

4. METHODOLOGY

4.1. Steps and activities followed in this research

This research was conducted in three phases: pre-fieldwork, fieldwork and post-fieldwork or analysis phase. The pre-fieldwork phase comprised of proposal writing and literature review which was conducted from August to September. This phase was aimed at evaluating the rationale to conduct this research, and to have a thorough review of the methods encountered in literature. After the proposal stage a fieldwork campaign was conducted for three weeks in Lake Naivasha, Kenya from the 13th of September to the 3rd of October. The data collected during field work facilitated the post-fieldwork phase. The post fieldwork phase comprised of a continuation of literature review, data analysis and thesis writing which was the final phase for the research. An overview of the step by procedure followed is as given in Figure 11 .

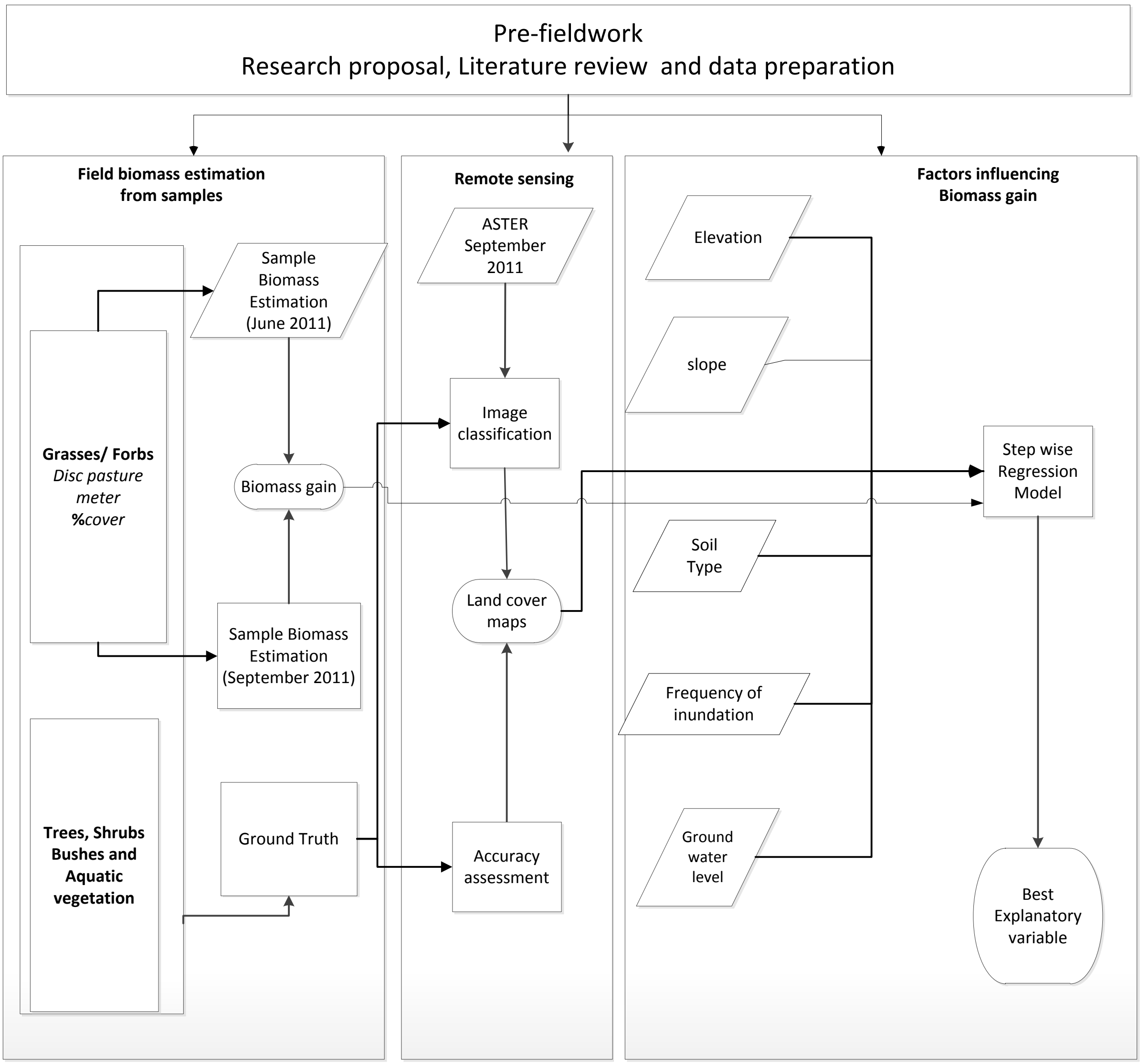


Figure 11: Overview of the complete steps and activities followed in this research

4.2. Pre-fieldwork Preparation

The initial stages included a preliminary literature review leading to a research proposal. This facilitated the preparatory work for sampling and choice of equipment for fieldwork. A sampling design was established with precise daily activity work plan. The predefined mapping and survey points were mapped out on the ASTER image of 14th of March 2011 and the 2009 aerial photos by Ramani (25 cm resolution) to assist in field navigation and plot identification. The list of materials used in the field is as in Appendix 1

4.3. Field data collection procedure

4.3.1. Sampling Design

Stratified representative sampling design was used. The strata were based on vegetation cover/land use types and distance from the lake which was at 1 km interval from each successive strata was used to collect ground truth data for image classification and accuracy assessment. A total of 4 strata equating to 4km from the lake shore were used. The 1 km distance was chosen as this was the anticipated distance where a completely different vegetation community would be found. This method was preferred as it is more time efficient and less time consuming as compared to the random or systematic sampling as you would not need to have equal number of sampling points in small strata (ITC, 2010). For each colour on the satellite image representative samples points were assigned. Sampling sites were located on the four sides of the Lake (north, south, east and west) were used for sampling. In each site, 500 m long line transects were established running perpendicular to the lake. This was done to capture the heterogeneity of the vegetation as we move away from the lake. Vegetation structure and type changes with distance from the lake shoreline (Gaudet, 1977). Four plots measuring 30 x 30 m were established along each transect at a constant interval of 150 m. All plant species for the five plants structural formations (trees, shrubs, bushes, grasses and aquatic vegetation) were identified and recorded together with their percentage coverage within the plot. The vegetation species were identified with the help of Mr Francis Muthoni (a PhD student at ITC) and field guide books.

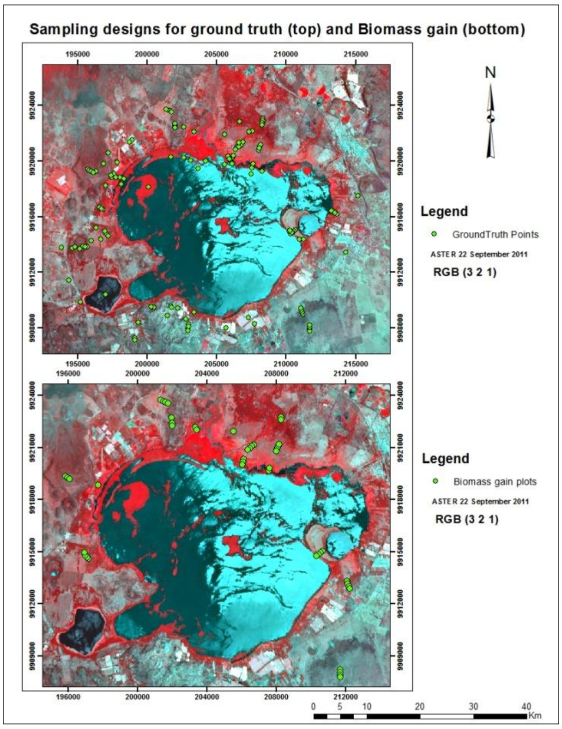


Figure 12: Maps showing sampling design (a) Biomass gain transects and plots used (b) ground truth points sampled

4.3.2. Ground truth data for mapping

To produce a good classification thematic map from remotely sensed data a good sampling design for ground truth is critical (Lu and Weng, 2007). The vegetation in Lake Naivasha follows a certain sequence with increased distance from the lake shore to dry land (Gaudet, 1977). Using this sequence, stratification was used because of the homogeneity of the different vegetation communities based on the distance from the lake. It was structured in a systematic way so as to reduce the cost of travel between plots.

A preliminary legend was made based on an arbitrary 30 classes of an unsupervised classification of the ASTER image. These classes determined the interpretation units. The 30 classes were used based on the idea that more than 30 colours were observed on the image. Points were then randomly generated in the interpretation units which were then used for ground truth. An iPAQ G.P.S and a Garmin G.P.S 360Cx with 3m accuracy were used for navigation to locate the points in the field. At each sample point, a plot of 30m X 30m was marked using the Garmin G.P.S. The 30 X 30m plots were developed a-priori and were navigated to using the centre points. A single data sheet was filled for each individual sampling plot. The precise coordinates of the centre of the plot were recorded along with the slope and elevation. Photos were taken from the four corners of each plot using a digital camera for further reference and assistance in classification. These sampling points obtained from the field were stored in a Microsoft Excel database and were used for accuracy assessment of the classified image.

Field classification was based on the level of dominance and cover percentage of the vegetation class. The highest layer in combination with the most dominant lower layer was used to characterise the vegetation classes. Five major layers were used to guide classification trees, shrubs, bushes, herbaceous and aquatic vegetation following Gaudet, (1977) who studied the vegetation successional patterns on the fringe of Lake Naivasha.

In the context of this research, a tree was defined as any woody perennial with a single stem or several stems but having more or less definite crown and height greater than breast height (5 m). Shrubs were defined as woody vegetation taller than 1.5 m but less than 5m with multiple stems and bushy appearance (FGDC, 1997). Bushes were defined as any woody vegetation between 0.5m – 1.5m with single or multiple stems. The herbaceous component comprised of all the grasses and forbs. Aquatic vegetation was defined as any vegetation existing on or below the water surface which cannot withstand excessive dryness. Aquatic vegetation comprised mostly Papyrus (*Cyperus papyrus*) and Water hyacinth (*Eichhornia crasspies*) (Adams *et al.*, 2002).

Table 2: Definitions of Vegetation cover types modified from Bemigisha, (1998) and FGDC, (1997)

Percentage cover per layer					Vegetation cover class
Trees (%)	**Shrubs (%)**	**Bushes (%)**	**Herbaceous (%)**	**Aquatic Vegetation (%)**	
≥60					Forest
20-60					Woodland
<20	≥55				Dense Shrubland
<20	<55				Open Shrubland
		≥60			Dense bushland
		<60			Open bushland
			≥60		Dense Grassland
			<60		Open grassland
			<10	>10	Aquatic Vegetation
			<10	<10	Bare

4.3.3. Herbaceous biomass estimation

Calculation of biomass gain was conducted for the plots that were sampled both in June and September 2011 to allow calculation of biomass accumulation between the two periods.

The available herbaceous biomass within each sampling plot was measured using a DPM (Bransby and Tainton, 1977). The DPM was used to estimate the above ground herbaceous biomass of grass. To calibrate the DPM, each 30x 30 m plot was subdivided into 1x1 m spacing grids. The nested 1 m spacing grids were coded before using a random number generator to select 30 random grids for sampling. The nearest grass patch to the selected grids was used to measure biomass using the DPM. The settling height (cm) of disc pasture meter for each measurement was recorded. A total of 30 readings were done in 30x30 m plot. After every 5th reading, the grass layer under the plate of the disc pasture meter was cut at ground level, oven dried at 70^{0}C to constant weight and weighed to obtain dry weight (g/m^2).

Linear regression was conducted between the recorded disc settling heights (cm) and measured above ground biomass (g/m^2) (in an area covering the dimension of the disc plate). The linear regression was performed on untransformed disc heights (independent variable) and above ground biomass (dependent variable) per square. In addition, regression analyses was conducted with different transformed disc heights (independent variables) i.e. (1) square (x^2), (2) square root ($\sqrt{x}$), (3) reciprocal ($1/x$), (4) natural log ($\ln x$), and (5) log x. All these transformations were conducted to find one that would best correlate with the above ground biomass (g/m^2). The regression analysis having the best fit as determined by the coefficient of determination (R^2) was used for estimating the biomass from the disc readings.

During the September field work campaign the same plots were used and nine more plots were added. A total of 30 random points per plot were generated in ArcGIS and the coordinates loaded into the GPS.

These navigation waypoints were the same points used in June. This was an attempt to measure precisely the same position measured in June to increase estimation accuracy and reduce measuring bias. n. When using the DPM there is tendency of observers to be drawn to measure the tall grass thereby introducing bias in the data (Zambatis *et al.*, 2006). These GPS waypoints were used to navigate within each of the individual 30mX30m plots to obtain an average height per plot. The GPS waypoint ID, the dominant grass species and the settling height (cm) of disc pasture meter for each measurement was recorded. A total of 30 readings per plot were recorded on a single sheet. The average height per plot was computed and was used as the "x" variable in Equation 9 to obtain estimated standing biomass (gm^2) for September

The linear regression contains a dummy variable to correct for the differences in mean height in the different vegetation types. This was in reference to plots occurring in closed forest where the herbaceous vegetation were tall, shade tolerant and highly lignified herbs with low biomass i.e. *Hypoestes forskahlii* and *Archyranthes aspera*. A dummy variable to reduce the overestimation in these plots was introduced in the equation depicting the vegetation type. This was based on the findings of Zambatis *et al.*, (2006) who concluded that categorising grasses based on their heights before calibrating the DPM could increase accuracy of derived calibration equation. Equation 12 was used to compute the average per plot standing herbaceous biomass from average plot height. This equation was obtained from the June calibration of Mr Francis Muthoni. Since the surface area of the disc was 0.16 m^2, the estimated dry weights from this equation were converted to (g/m^2) by multiplying by 6.25. The equation had an R^2 of 0.83 and at (p-value = 0.05; CI =0.95).

$$\mathrm{DM} = 19.828 + 10.461\mathrm{x} - 7.588\mathrm{Dveg} \qquad \text{Equation 9}$$

Where: DM is the estimated dry weight
x is the per plot measured average DPM height
Dveg is the dummy variable representing the interaction between herbaceous biomass and vegetation type.

4.4. Post-Field work

4.4.1. Image classification

In order to answer the research question "How separable are herbaceous vegetation types/ classes when mapping using ASTER imagery in a riparian ecosystem?" three ASTER image bands green, red and NIR bands (VNIR) were used for image classification. Following the review of classification methods by Lu and Weng, (2007) this research used a combination of automated unsupervised Isodata clustering, ancillary data and knowledge based classification since the study area was small and there was a vast bank of knowledge of the area. Combining 2 classifiers and ancillary data is known to improve classification results (Jianwen and Bagan, 2005).

Built up areas, green houses and croplands irrigated by pivots were first masked out from the image by manual digitising. The areas were clearly visible from the image and were also evaluated and verified on the ground and on high resolution aerial photos. All the positions with the aforementioned were converted to "no data" values as to reduce the confusion within the classification algorithm because of the heterogeneity of these classes.

The masked image was initially classified using the unsupervised Isodata clustering algorithm. However in remotely sensed images it is difficult to separate all the features using the spectral properties only (Zhang and Zhu, 2011). Hence expert knowledge was utilised as an additional step to improve the classification of the classes whose spectral separability was very low.

Isodata clustering algorithm in ENVI using 30 classes and a minimum mapping unit of 45m X 45m thus 9 pixels was used on the masked image. This was aimed at separating the aquatic vegetation from the rest of the terrestrial classes and also discriminating classes within herbaceous grassland communities. Some of the aquatic portions were small patches hence the smallest mapping unit was set to cater for those small units. The polygon shape file produced was overlaid on the ASTER image, the aerial photo, the DEM and NDVI map.

4.4.2. Establishment of knowledge rules

Expert knowledge following Su *et al.*,(2011), using the "if and then" method was adopted to come up with rule sets that distinguished the classes as given in Table 3.

Table 3: Rule sets used in the manual assignment of the isodata generated polygons for the expert knowledge classification

Class of Interest	**Set rule(s)**
Temporarily flooded grasslands	- contiguous to ≤100m from the lake shoreline - little to no bare spaces in-between - high NDVI value
Dense sward of medium height	>100m from the lake shoreline - little to no bare spaces in-between - moderate NDVI value
Short-sparse grasslands	>100m from the lake shoreline - large to 80% bare spaces in-between - low NDVI value
Bushed tall tufted grasslands on coarse bare soils	- 60-80 %bare - low NDVI value
Aquatic vegetation	- contiguous to or in water -space in-between covered by water -smooth texture
Open bushland	- low NDVI value - large to 80% bare spaces in-between
Forests	- can never be in water - coarse texture - closed canopy
Woodland	- can never be in water - coarse texture - open canopy
Euphobia woodland	- can never be in water - slope > 20 % - high elevation
Tarchonathus camporatus shrublands	- can never be in water
Senna didymobotrya	≤300 m from lake shoreline
Croplands	- clear-cut/regular boundaries
Water	≤ 1890 m.a.sl NDVI = 0 - 0.3 - Lake is circular and river is linear
Built up, Green houses,	- regular shape

This included consultation from experts who have been to the study area. The knowledge rules were based on spectral, visual interpretation of proximity, shape, texture and the topographic information drawn from the digital elevation model. The key classes of interest were the herbaceous units which are represented as the first four classes in Table 3. Manual assignment of polygons to class was done for the whole study area following the set rules.

4.4.3. Accuracy assessment

Evaluation of the accuracy of the output classification was done using an "Accuracy assessment" tool in ERDAS imagine 10 software. The ground truth points were used as the reference points. The output of the comparison of the classified to the reference ground points were recorded on an accuracy table. The error matrix produced from the ERDAS accuracy assessment was used to evaluate the miss classified classes. The overall accuracy, Kappa statistics User's and producer' accuracy were used to evaluate the classified map. A summary of the whole classification procedure is given in Figure 13.

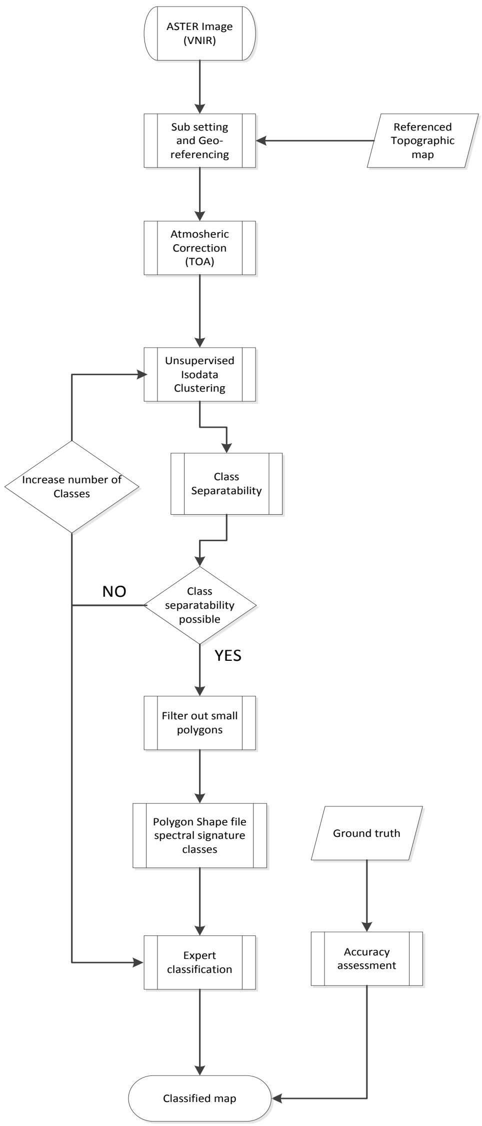


Figure 13: Flow chart for image classification

4.5. Analysis methods for biomass gain.

To answer the research question that sought to establish if the herbaceous vegetation component could have different productivity classes, purely statistical methods were adopted. Before any analysis was conducted, the data was explored using a histogram, a QQ plot and Shapiro Wilk normality test following Henderson, (2006). Descriptive statistics using box plots were also explored to assess the behaviour of the data and the spread (Moore, 2001)

To analyse the standing biomass between June and September box plots were used to show the spread of data. A Wilcoxon signed rank test was used for analysis of statistical differences between the June and September standing crop. A Wilcoxon signed rank test is a non-parametric analogue that compares significance of differences in medians. It is a parallel to the parametric two sample pair wise t test used when data is not normal. The null hypothesis for the Wilcoxon signed rank test states that the median distance between two pairs is zero(Moore and McCabe, 1998).

The "True NPP" of the area could not be established since the study period was not the period in which we would expect peak standing biomass, since this was the middle of the wet season. However, increase in standing biomass was expected within the period June to September. Biomass gain was used as a proxy to assess the general productivity for the period June to September which is in the middle of the growing season for most of the grasses. The assumption was that biomass gain would to some extent explain the level of productivity in the area. Equation 10 is the equation used to estimate the biomass gained between the period June to September 2011.

$$B_{gain} = SC_{t+1} - SC_t \qquad \text{Equation 10}$$

Where B_{gain} = Biomass gain
SC_{t+1} = maximum standing crop in September
SC_t = maximum standing crop in June

There were 34 plots that were used for biomass gain analysis. The standing biomass estimated from the June survey was subtracted from standing biomass measured in September following "Method 3" of Scurlock, *et al.*,(2002). The resulting difference was the assumed biomass gained within the period.

4.5.1. Analysis of Biomass gain based on herbaceous vegetation community

In the initial analysis, the 34 plots were grouped into four classes. This grouping was according to herbaceous vegetation types derived from the classified ASTER image and the knowledge from the field (Table 4). The four classes include temporarily flooded grasslands. These are grasslands that are frequently inundated with water as the lake level increases and recedes. Therefore, these grasslands were mostly within 100m from the lake shoreline. The second class was named "Dense sward of medium height" which was a combination of the short-sparse class with the dense sward of medium height derived from the classified image. The third and fourth classes were grasslands that existed in combination with other vegetation types thus shrublands and forests respectively. The total number of plots in each group is as shown in
Table 4

Table 4: Herbaceous vegetation groups and the number of plots for biomass gain analysis

Herbaceous vegetation class/group	Number of plots
1. Temporarily flooded grassland	6
2. Dense sward of medium height	9
3. Herbaceous vegetation in shrublands	13
4. Herbaceous vegetation under forest canopies	6
Total	**34**

Following Moore and McCabe, (1998), a one way ANOVA test was done to test if there were significant differences in the biomass gain of these four classes (CI= 95%, p= 0.05). The analysis was conducted using R-software. The null hypothesis was to be rejected if the p-value was below the critical value α ($p<0.05$) and not rejected if otherwise. The null (H_0) and alternative hypothesis (H_1) were stated as below:

$$H_0: \mu_1=\mu_2=\mu_3=\mu_4 \qquad \text{Equation 11}$$
$$H_1: \mu_1\neq\mu_2\neq\mu_3\neq\mu_4 \qquad \text{Equation 12}$$

The groups were further reclassified into open and closed herbaceous plots based on percentage cover of the over-story vegetation. Temporarily flooded grasslands and dense sward of medium height classes were combined into open grasslands. These were grasslands with no/very low (<10%) trees and shrubs canopy cover. The closed group comprised herbaceous vegetation under forests and shrubs.

A Welch two sample t-test was used to assess if there were differences between the means in the two groups (CI= 95%, p= 0.05) (Moore and McCabe, 1998). This test was chosen as it does not require that the two samples be of equal size or that the variances be equal. The test was also conducted in R-Software. The null hypothesis was to be rejected if the p-value was below the critical value α ($p<0.05$) and not rejected if otherwise. The null (H_0) and alternative hypothesis (H_1) were stated as below:

$$H_0: \mu_1-\mu_2= 0 \qquad \text{Equation 13}$$

$$H_1: \mu_1-\mu_2\neq 0 \qquad \text{Equation 14}$$

4.6. Analysis of Lake level effect on biomass gain

From literature, the ground water level around lake Naivasha mimics the changes in lake level (Abdulahi, 1999). The relationship between lake level and biomass gain is not direct therefore we sought to derive the influence that lake level has on ground water hence soil moisture content.

To compare the effect of lake level change on above ground herbaceous biomass gain, analysis of how the lake level influences the ground water was conducted. This analysis assisted in answering two research questions "What is the threshold influence distance that the distance from the lake shoreline has an on ground water and ultimately herbaceous vegetation productivity?" and "How does ground water influence the productivity of the herbaceous vegetation?"

An algorithm was developed in R using the historical lake level and ground water level data available from the ITC database. The data had already been organised as stated in Chapter 3 of this thesis. Two more variables "distance to the lake centre" and "distance to the lake shoreline" were computed for each borehole and each corresponding measurement. The distance to the lake centre did not differ among successive measurements per individual borehole but distance to the shoreline varied as it was derived based on the lake level at that particular period.

The bathymetrical contour lines of 1957 available from the ITC database were used to construct the lake boundaries at different lake levels as shown in Figure 14. Data was only available for 1884 m, 1887 m, 1888 m and 1890m therefore distances for lake levels between 1885 and 1886 up to 1887 were calculated to 1887 lake boundary. All distances for lake levels within the range 1887 and 1888 were calculated to the 1888 boundary. The distances for lake levels beyond 1888 up to 1890 were all calculated to the 1890 boundary.

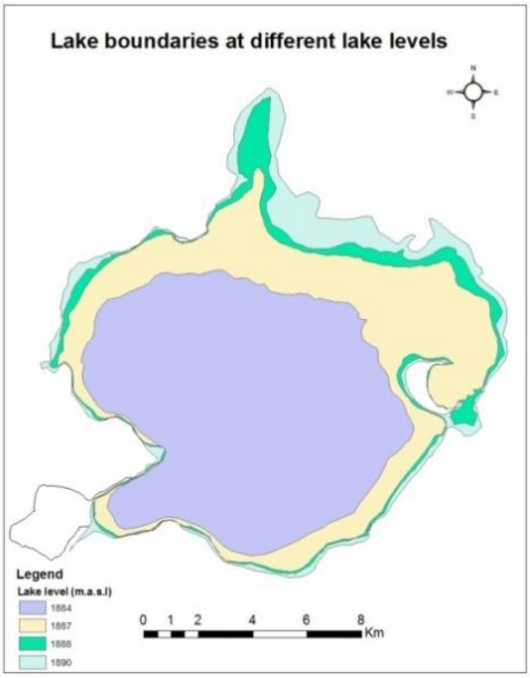


Figure 14: Lake boundaries at different lake levels constructed from Lake bathymetry of 1957

4.6.1. Threshold distance model

The available ground water dataset spanned from the year 1997 to 2010. The number of successive measurements per borehole was not consistent. Some boreholes were measured more times than others hence the measurements of ground water level per borehole and their corresponding lake level measurement were averaged. The output computation of this averaging process was a single average ground water level per borehole and a corresponding lake level average for all the measurements of that individual borehole. The averages per borehole were used to calculate the deviations of either in lake level or in ground water by subtracting the average per borehole from the actual measured value at time *t*. These deviation were calculated for each measurement and two columns were added to the table containing “deviations in lake level” and “deviation in ground water level”

Because ground water and lake level are measured at different locations fitting their correlation directly would not be logical. However, a change in lake level was expected to result in a change in ground water level hence correlating the deviations was more realistic. A model between the ground water deviations and the corresponding lake level deviations was developed to find if the changes in lake level were correlated to the changes in ground water level.

The model was developed following step wise sequence in which for first run (run 1) the model would include only the first 15 boreholes closest to the lake and fit the model. The R^2, the slope coefficient of the model and the average distance to the lake centre for those 15 boreholes were computed and stored in additional columns. In run 2 the model would eliminate the first borehole closest to the lake and would select the borehole with next closest distance from those that will have been excluded from the first run (thus 14 boreholes that were in run 1 were included). Again the R^2, the slope coefficient and the average distance of those 15 boreholes in run 2 would be computed. The process was iterative in the same order until all the boreholes were considered. After every quartile the model was tasked to plot the average distance to the lake centre for viewing and analysis of model performance. The same was done using the distance to the shoreline.

To find the threshold distance where the correlation between ground water and lake level would break down an additional script was developed in R. Simultaneously, the R^2 and the slope coefficients generated from the previous values from the linear model between lake level deviations and ground water deviations were plotted against the average distance of each corresponding run. The threshold distance was determined by the first run that would have a low R^2 value below an arbitrary value of 0.4. The R-script is as shown in Appendix 2

4.6.2. Ground water depth interpolation

A simple model using historical lake level and ground water measurements together with ordinary kriging technique was used to come up with the ground water depth model. The difference between the average lake level and average groundwater table measured in the field was computed. Therefore for every borehole there was an average difference between the level of the lake and the corresponding ground water table depth measured. The lake level was subtracted from the ground water level so that a negative value would represent ground water lower than the lake level and a zero value would be ground water level equals lake level. A positive difference would mean the ground water level was higher than the lake level.

Additional fictional boreholes were added around the perimeter of the lake using the 2011 boundary prior to ordinary kriging. This boundary was used as it was assumed to be the most appropriate since it was derived from the boundary of the image taken during the field work period. These fictitious boreholes were placed at an arbitrary 50 m interval from each other and were given the difference value of zero so as to mimic the a “no difference” scenario in which the lake water level is the same as the ground water level. A step by step procedure is as given in Figure 15

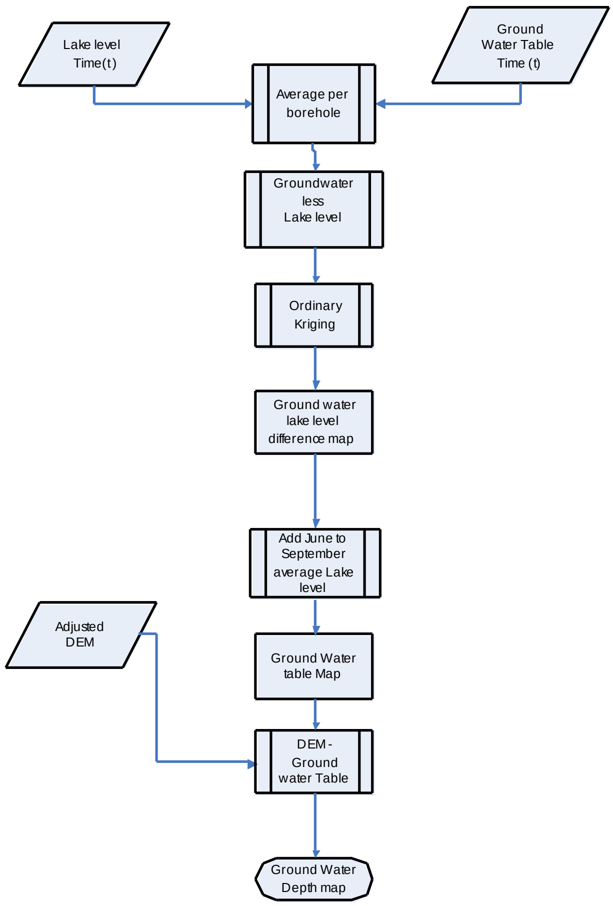


Figure 15: Procedure followed to interpolate ground water depth

Using the ordinary kriging interpolator tool in ArcGIS geostatistical analyst extension, the average differences between the lake level and ground water level were predicted for the whole area. Three stages were followed in the interpolation procedure. The data was first explored using the explanatory tool in the Geostatistical analyst tool box. The data was assessed for normality by viewing a QQ plot and the spatial autocorrelation by using a semi-variogram. The second stage was variogram modelling using a lag size of 850 m and range of 10 000 m. These parameters were derived from the fact that since lake level influenced the ground water up to a threshold of 9 500 m then spatial dependence was expected in the same range. The 850 m lag size was arrived at by calculating the average distance between two successive points within the dataset. The model was used to predict for the unmeasured areas and a difference map was obtained. Using the cross validation tool imbedded in the kriging toolbox, the model was validated using the Standard Error (SE) and RMSE.

After the modelling of the difference map, average lake level for the period June to September was added to the interpolated ground water difference map using ArcGIS 10 raster calculator in spatial analyst tools. The result was an estimated average ground water table map given for the period. This output was subtracted from the adjusted DEM using the same tool in ArcGIS 10 spatial analyst tools raster calculator. The resulting output was an estimated average ground water depth from June to September 2011 map interpolated for the whole study area.

As established in literature the savannah grasslands root depth goes down to half a meter below ground (Canadell *et al.*, 1996). The data from ground water was split into two groups (i) shallow (≤5m)and (ii) deep ground water (>5m) based on the potential of the ground water to influence the herbaceous rooting zone. The data was analysed using a Welch t-test to compare if there were differences in the means of the biomass gain between the two groups.

This output was used in the determination analysis to evaluate the effect of ground water depth on herbaceous biomass gain. The "Extract values to points" tool in ArcGIS spatial analyst was used to extract the ground water depth per plot which was added as a column on the table which had other parameters' values. This output was used in the multiple regression analysis to establish the best explanatory variable for biomass gain.

4.6.3. Procedure to determine frequency of patch inundation

Most studies on frequency of inundation use topographic, LIDAR and SAR remote sensing to map flooding areas (Cook and Merwade, 2009; Zhao and Li, 2012). Since there was no SAR or LIDAR data a simple method was adopted to establish frequency of inundation with the available data.

To analyse for the frequency of flooding per plot, inundation images were generated by combining lake bathymetries with the historical lake level information. The images were designed in such a way that the area enclosed within the lake boundary as determined by the lake level and subsequent bathymetry would be given a pixel value of 1 and all the outside pixels were given the value 0. Four template images were made from the four bathymetrical boundaries thus 1884, 1887, 1888 and 1890. The value 1 depicted that the area was inundated and 0 the opposite. Using ArcGIS 10 raster calculator in spatial analyst tools a total of 174 images were added to each other for lake level monthly average within the dataset (February 1997-September 2011). To find the average period of inundation the output image was divided by 174 to find the average. The area completely covered for the whole period was of the value "1" and the areas that were never covered by water remained "0" for the whole period. The values were changed to percentage to facilitate calculation of the average number of days a patch is inundated in any given year or period.

Four groups were established from the four categories established from the frequency of inundation to analyse biomass gain. These four groups include Almost always inundated – these are areas that are covered by water 60-100% times during the year hereby referred in the analyses as inundation frequency 1(IF1). The second category was areas covered by water 20-60% (IF2), the third was those covered less than 20% of the time (IF3) and the fourth was for those that are never flooded (IF4). Box plots were drawn for descriptive analysis. Following Moore and McCabe, (1998), a one way ANOVA test was done to test if there were significant differences in the biomass gain of these four classes (CI= 95%, p= 0.05). The analysis was conducted using R-software. The null hypothesis was to be rejected if the p-value was below the critical value α ($p<0.05$) and not rejected if otherwise. The null (H_0) and alternative hypothesis (H_1) were stated as below:

$$H_0: \mu_1=\mu_2=\mu_3=\mu_4 \qquad \text{Equation 15}$$

$$H_1: \mu_1\neq\mu_2\neq\mu_3\neq\mu_4 \qquad \text{Equation 16}$$

A post hoc Turkey Kramer method analysis to test to check which of the groups were significantly different was used to establish the groups that had their means significantly different. This data was further used in the multiple linear regression.

4.7. Assessment of the explanatory variables for biomass gain

After the analysis of the individual variables effect to biomass gain it was paramount to check for the interaction effect and to determine the best variables to predict biomass gain and ultimately productivity. Eleven variables were reviewed from literature (Chapter 2) and were considered as potential explanatory variables for biomass gain. Seven variables were used in a stepwise regression because of the availability of data. These were initially tested for collinearity using their pairwise R^2 values, and were used in the step wise regression model as shown in Figure 16. The elevation data was as obtained from the DEM in m.a.s.l, Slope was a continuous variable as percentage, and the land cover classes were categorical as derived from the classified map from the four herbaceous types. The two soil type parameters included in the regression were bulk density and cation exchange capacity. These were important parameters determining water and nutrient retention. The depth to ground water (distance from the surface) variable was included in the regression as a continuous variable as derived from the modelled ground water depth. Lastly the frequency of inundation was included as a percentage as derived

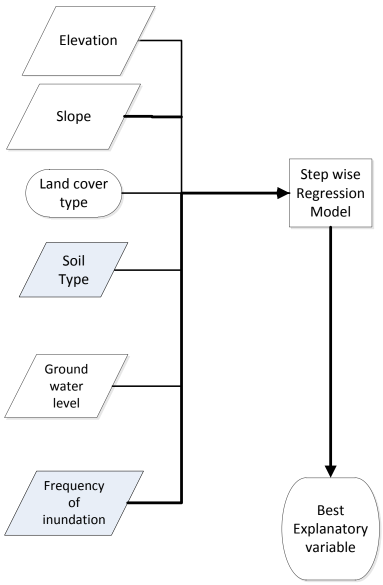


Figure 16: Flowchart of the multiple regression analysis for the explanatory variables for biomass gain

A step wise linear multiple regression model was developed in R-software (Appendix 3). The variables were considered valid based on the level of significance in the model ($p = 0.05$; CI = 95%).

To validate, the model was designed to draw samples from data split into train (70%) and test (30%) for each successive run. Stochastic results were obtained for individual runs as it was programmed to draw random samples for each run. The split was done following Snee, (1977). The model was run 50 times to obtain the best model coefficients. For each successive run, model graphs of measured against predicted and residuals using the test data were plotted and the R^2 observed. The R^2 for all the 50 models were averaged to find the average performance of the model. The average and the highest R^2 obtained from the validation were used to evaluate the model performance. In addition the randomness and homoscedasticity of the residuals were tested using a QQ normal to assess the goodness of fit of the predicted values to the model.

An equation with the variables included in the final model of the step wise process was considered the best to predict biomass gain.

5. RESULTS AND DISCUSSION

5.1. Vegetation mapping using ASTER imagery

Seventeen classes were identified using a combination of the automated and knowledge based methods. One class of combined aquatic vegetation, three classes for the herbaceous, two for the bushland, three for shrubland and three for trees were established for the natural vegetation. Five other classes were also identified (Figure 17).

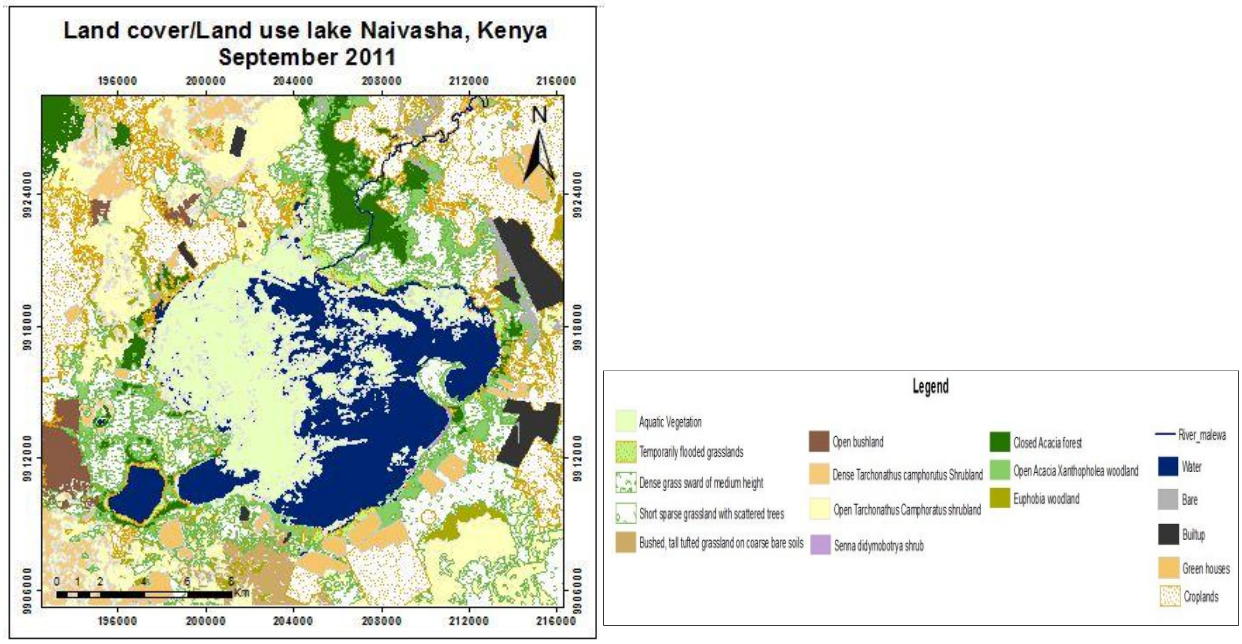


Figure 17: Land cover/land use for Lake Naivasha September 2011 using a combination of automated unsupervised isodata classification and expert knowledge

The isodata algorithm was utilised to classify the vegetation based on their spectral reflectance. The method was able to clearly discriminate different herbaceous classes successfully from an ASTER image. The overall classification accuracy of the ASTER image was 90% and had a kappa coefficient of 0.89. This accuracy satisfies the standard requirements for practical applications. This accuracy was within the range produced by similar studies using similar approaches (Zalazar, 2006; Zhang and Zhu, 2011). Most of the classes had both 100% user and producer accuracies and this could be attributed to the high spatial resolution of ASTER which facilitated the discrimination of some objects.

The lowest producer accuracy was in open bushland and the lowest user accuracy was in dense Tarchonathus Camporutus shrubland (Table 5). These results could have been because of the low number of samples within these classes. The miss-classifications were mostly between classes that had similar spectral signatures and those that were in close proximity to each other. Miss-classification emanating from the spectral signature was mainly between open and closed acacia forests and aquatic vegetation. This is because they exhibit high NDVI values through the year. The classes misclassified because of proximity could have been influenced by the majority filtering that smoothed the image causing homogeneity of pixels close together.

Although there were these challenges, this confirmed the potential and strength of combining multi source data in classification with expert knowledge to produce high accuracies (Foody, 2002; Lu and Weng, 2007). However the computational time required is high as most remote sensing images have numerous pixels.

Table 5: Confusion matrix from the accuracy assessment of the classified ASTER image of the 22nd of September 2011 showing the correctly classified and misclassified pixels.

Class	Water	Open Tarcho	Euphobia Wood	Dense Tarcho	Open Acacia	Closed Acacia	Temporarily flooded	Aquatic vegetation	Dense Sward	Bushed -tufted	Short sparse	Open bush	Crop lands	Senna	Bare	Green houses	Built -up	Producer %	User's%	Kappa
Water	**6**	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	100	85.7	0.85
Open Tarchonathus	0	**15**	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	88.24	93.7	0.93
Euphobia Woodland	0	0	**9**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	100	1.0
Dense Tarchonathus	0	2	0	**3**	0	0	0	0	0	0	0	0	0	0	0	0	0	100	60	0.6
Open Acacia	0	0	0	0	**13**	0	0	1	2	0	0	0	0	0	0	0	0	86.67	81.2	0.79
Closed Acacia	0	0	0	0	2	**15**	0	1	0	0	0	1	0	0	0	0	0	100	78.9	0.76
Temporarily flooded	0	0	0	0	0	0	**8**	0	0	0	0	0	0	0	0	0	0	88.89	100	1.0
Aquatic vegetation	0	0	0	0	0	0	0	**11**	0	0	0	0	0	0	0	0	0	78.57	100	1.0
Dense Sward	0	0	0	0	0	0	0	1	**14**	0	0	0	0	0	0	0	0	82.35	93.3	0.92
Bushed-tufted	0	0	0	0	0	0	0	0	0	**3**	0	0	0	0	0	0	0	75	100	1.0
Short sparse	0	0	0	0	0	0	0	0	0	1	**6**	0	0	0	0	0	0	100	85.7	0.85
Open bushland	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	0	0	0	66.67	100	1.0
Croplands	0	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	100	100	1.0
Senna	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	0	0	0	100	100	1.0
Bare	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	100	100	1.0
Green houses	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	0	100	100	1.0
Built-up	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**3**	100	100	1.0

5.2. Analysis of standing biomass and biomass gain for June to September period

5.2.1. Data normality test

Testing for normality is critical for choice of suitable statistics as most analytical statistics assume a normal distribution in the data. Most of the data points lay on the y=x model of the QQ plot which showed that the data was normally distributed. There were however outliers which showed that the data was skewed to the left of the normal distribution curve. The Shapiro-Wilk normality one sample test null hypothesis is that the data is normally distributed. The data is normally distributed if the p-value is greater than chosen alpha or reject null if W is too small. In this case $\alpha = 0.05$ was used and the results revealed that p-value was greater than α ($p>0.05$) (Table 6). Therefore we do not reject the null hypothesis and conclude that the data is normally distributed for biomass gain. The opposite is true for the standing biomass for both June and September which all had p-values < 0.05. Because the data for biomass gain was normally distributed parametric statistical methods were used for all the analysis and non-parametric for standing crop biomass. The summary of the normality test is as shown in Figure 18:

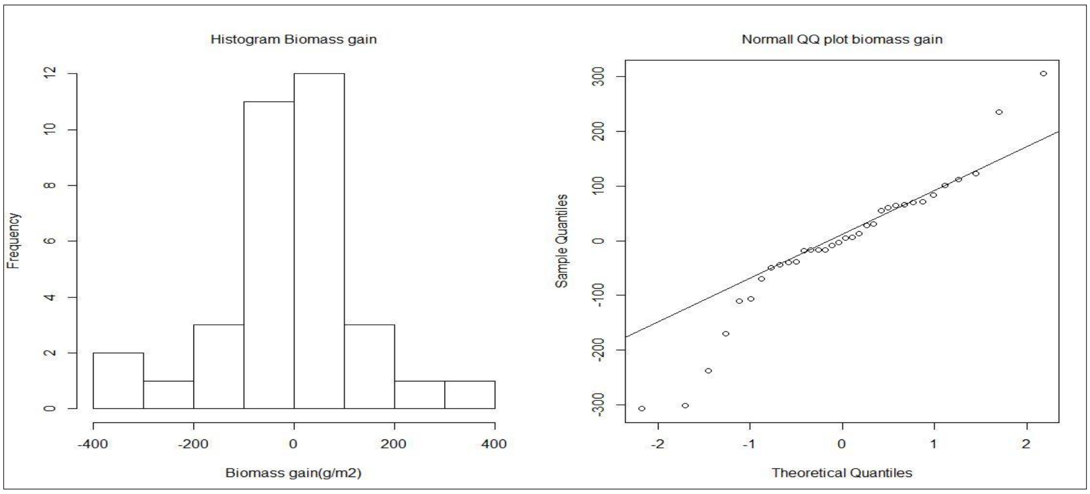


Figure 18: Normality testing using histogram and normal QQ plot for biomass gain between June and September 2011

5.2.2. Analysis of standing biomass and biomass gain for the whole study area

The average grass height in the fringe zone of Lake Naivasha was 5.5cm dominated by *Cynodon dactylon* and Kikuyu grass (*Pennicetum clandestinum*). The estimated average standing biomass as derived from the DPM compressed height was 381 g/m². The resulting biomass gain is within the range reported in similar studies in Kenya. Keya, (1998) reported a maximum of 432 g/m2 though in a non-grazed semi arid region. Considering the fact that the 381 g/m² estimated for Lake Naivasha which is heavily grazed the results show that the lake Naivasha herbaceous grasslands fair in productivity. This could be attributed to the proximity to water and the ability of optimum grazing to stimulate growth as compared to the study area of Keya, (1998)

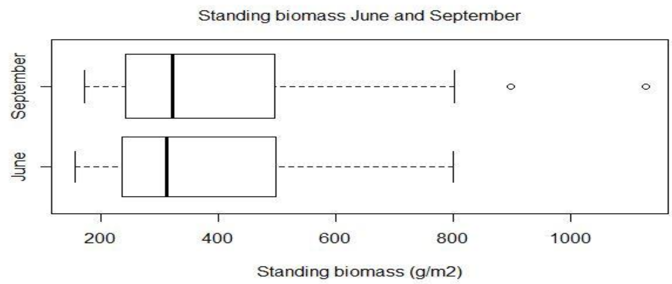


Figure 19: Standing above ground herbaceous biomass as derived from the field using a disk pasture meter

Comparing the results locally between the two sampling periods, the standing biomass for the two sampling periods did not differ significantly (Table 6). There was an average drop in the above ground standing biomass of 3.7g/m² with the June standing biomass being more than the September standing biomass. This was a result from the average of the paired differences of the June and September standing biomass. This result could signify that the phenology of the herbaceous biomass in all the years could have this drop in productivity in September as opposed to the assumption that the productivity increases.

This result could be attributed to the bimodal seasonality of precipitation in Naivasha. Kenya has two rainfall peak seasons, one commonly referred to as the long rains period (March to May) and short rains (October to December). April receives the highest rainfall which could have attributed to the high standing biomass in June which is the peak of the long season (Becht and Harper, 2002). September is at the beginning of the short rains hence the anticipated result was the standing biomass would have gone down. The standing biomass in September was a bit higher than expected. This anomaly could be attributed to the fact that Naivasha received more rainfall than normal in September 2011. The September period is normally very dry (Bergner *et al.*, 2003). The continuous rainfall could have prompted the increased productivity during the period.

Table 6: Descriptive statistics for biomass gain, June Standing and September Standing crop (n = 34) and the results of the Shapiro-Wilk normality test

No	Parameter	*Biomass gain value (g/m^2/3 months)*	*June Standing crop (g/m2)*	*September Standing crop (g/m2)*
1	Mean	-3.7	382.8	379.1
2	Median	0.9	327.5	333.1
3	Maximum	305.1	800.0	800.6
4	Minimum	-307.3	166.8	173.6
5	Standard deviation	126.3	166.6	209.2
6	Shapiro-Wilk normality test	W = 0.9382, p-value = 0.05443	W=0.8923, p-value = 0.00289	W=0.9189 p-value = 0.01509

The maximum standing biomass for June and September was observed in KARI (Kenya Agriculture and Research Institute) farm in the North of the study area. The north is flat and the water table is closer to the ground surface as compared to all the other sides of the lake. The high water table and gentle slope promotes higher soil moisture availability that could have stimulated the high level of productivity that yielded the high standing biomass in the area. Moreover, historically the northern area was part of the lake before the 1970s. When the lake receded the nutrients that were in the lake bed sediments remained in the soil where these grasslands exist today. The high soil nutrient content is also evidenced by the presence of mole-hill like clumps that were formerly papyrus stands. As observed in the field, grasses occupying these former papyrus clumps tend to grow taller than other areas. This could be evidence of high nutrient content left when the lake receded. This assumption can only hold in the former north swamp area hence measurement of soil nutrients could be a reasonable explanatory variable to pursue.

In addition these areas are heavily grazed as evidenced by the presence of large herbivores in the area. These areas could be exhibiting higher compensatory growth due to high grazing intensity as compared to all the other areas (Milchunas and Lauenroth, 1993). The nutrients recycled by the grazers and browsers through their droppings and urine could have also influenced the quantity of biomass (McNaughton *et al.*, 1988)

Although the two highest biomass stand plots were in different transects they were in a similar vegetation community with similar dominant grass species *Cynodon dactylon* and *Pennicetum clandestinum*. These species are of high grazing preference and highly palatable producing dense swards. Because of their adaption to grazing by transforming themselves to rhizomatous growth forms in heavily grazed areas they often tend to avoid being grazed (Danckwerts and Stuart-Hill, 1987; Tainton, 1999). This could have aided the resilience of these grasslands occurring in the northern side of the lake. Thus the inherent characteristics of the grasses could have played a role although detailed analyses to test for the variations caused by these inherent characteristics was analysed and discussed in sections that follow. Statistical analysis showed no significant differences ($p>0.05$; CI = 95%) between increaser and decreaser or between high preference grazing and low preference grazing which is contrary to some studies which report differences in productivity in such species (Danckwerts and Stuart-Hill, 1987). It might be that differences exist but could not be detected with the current data.

This is different from the areas that had the lowest standing biomass in June and September. These grasslands are generally dry and one of them was at Crescent Island and the other one in a shrubland where there are steep slopes and high level of grazing. Although the species were almost similar to those with high standing biomass and also the fact that these areas are heavily grazed, the slope and water availability caused the low standing biomass. The plots at Crescent island are overgrazed hence we would not expect significant increases in biomass gain as any shoot is quickly defoliated when it comes out.

A Wilcoxon signed rank test was conducted to confirm if statistically the June and September standing biomass had significant differences. This statistic was chosen as standing biomass data was not normally distributed (Table 6). With W=567.5, p-value = 0.9024 the Wilcoxon test suggests that there is no statistical difference (CI=95%, $p>0.05$) between the standing biomass in June and September which confirms the issues discussed above.

The dispersion of biomass gain values in the study area are represented in a box plot in Figure 20. The overall biomass gain, the median and mean did not vary much from zero, and the mean was in the negative (Table 6). As highlighted above the September biomass was at the onset of the short wet season and the biomass was also low as compared to the June biomass. This research did not account for the litter fall and grazing. It is assumed that much of the biomass generated in June was lost to litter fall and grazing during the three months interval. This result shows that the grazing and growth could be in equilibrium or that compensatory growth is high. This result also confirms the claim by (McNaughton *et al.*, 1988) that large mammals have a major effect on grasslands by stimulating productivity.

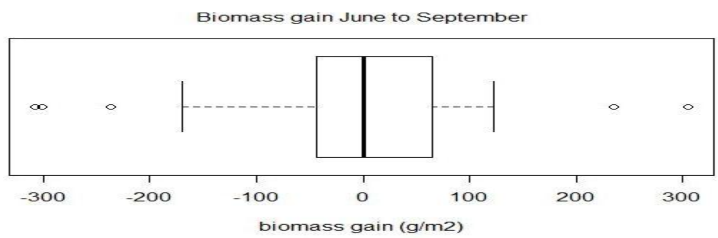


Figure 20: Box plot showing the dispersion of the biomass gain data between the period June to September 2011

As demonstrated by (Lauenroth *et al.*, 2006) to quantify the total productivity, all the biomass fluxes need to be incorporated i.e. biomass gain, decomposition, exudation, consumption and volatilisation (Equation 2). In this research only the change in biomass (ΔB) due to time limitation was measured and analysed. This is a start of a long process as this research was part of an EOIA long term research set up to quantify all the other subcomponents in this equation to produce the "Total NPP" for the fringe zone.

The approach taken to quantify only "ΔB" has been shown by Long *et al.*,(1989) to underestimate the productivity of herbaceous communities. This is so as some of the production lost between sampling periods is not accounted for. Although this is the case this research was able to establish the variations in biomass gain in the different sides of the lake and an overview how grasslands perform in the fringe zone, which shall be discussed in sections below.

5.2.3. Biomass gain in different herbaceous communities

The differences in biomass gain as depicted by the herbaceous community are as shown in Figure 21. The highest biomass gain was in inundated grasslands and the lowest was for grasslands in shrubland which had a mean loss of -37g/m². Shrubland grasslands and dense sward grasslands both had negative gains whereas inundated and forested grasslands had positive gains. The negatives in the shrublands maybe due to the shading effect and competition for water and nutrients while in dense sward grasslands it may be as a result of high grazing intensity. Most plots in associated shrublands were in Kedong farm on the east of the lake. In most of these plots competition for water and nutrients plus the shading from the thick *Tarchonathus camphorutus* shrubs could have contributed to the reduction in biomass gains.

These plots were also in a relatively dry and sloppy area because of the high elevation levels. The low soil moisture content in this area could have been one of the factors apart from the fact that probably within these area grazing and production was in equilibrium. Although the same amount of grazing intensity was exerted to inundated grasslands they had a positive gain which signified higher productivity as compared to other communities this could be because of the water availability. The other expectation was that the productivity would drop as the water levels rise in September covering the majority of the inundated grassland plots. However this was not the case probably because the grasses within these plots are already adapted to such heavy flooding.

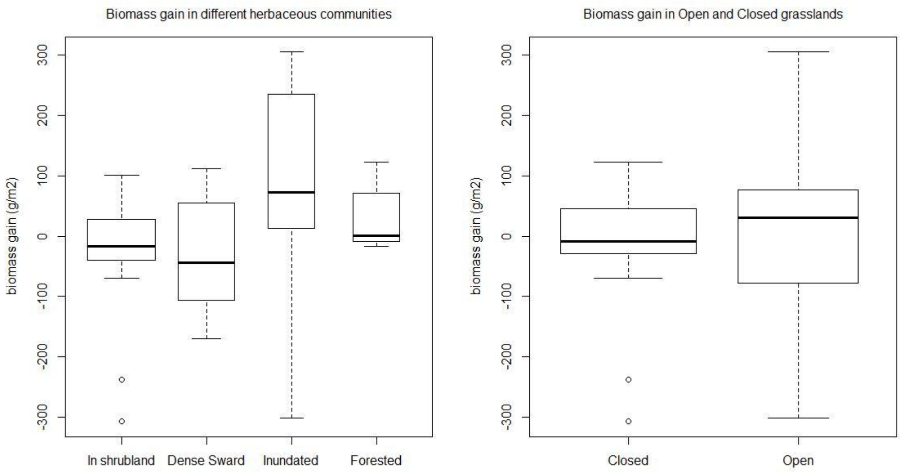


Figure 21: Summary descriptive statistics for the biomass gain grouped according to the herbaceous community (i) Herbaceous in shrubland (ii) Dense sward (iii) Temporarily inundated and (iv) Herbaceous in forests (v) Closed herbaceous (vi) Open herbaceous

Most of the forested grasslands were mostly in the area where animals are fenced out and were mostly associated with unpalatable herbs like *Achyranthus aspera* and *Hypoestes forkeolii*. This could be the possible explanation to the reason the range for forested areas is very small (Table 7). The plants within the forested areas are quite tall (ranging from 10 to 30 cm). The DPM is therefore not the best suitable instrument for this vegetation type since they do not match up with the compression assumption that accompanies the DPM. These herbaceous plants underneath large canopies of forests grow very tall and thin because of light competition but the amount of biomass accumulated is low. These plants also grow tall mostly under *Acacia xanthophloea* because of the high moisture content underneath these forests as observed in the field which is contrary to the dense sward which takes the horizontal growth form and accumulate high above ground biomass.

Table 7: Summary descriptive statistics for the biomass gain grouped according to the herbaceous community (i) Herbaceous in Shrubland (ii) Dense sward (iii) Temporarily inundated and (iv) Herbaceous in forests (v) Closed herbaceous (vi) Open herbaceous

Parameter	*Shrub*	*Dense sward*	*inundated*	*Forested*	*Closed*	*Open*
Mean	-37.0	-23.5	66.1	28.4	-16.3	12.3
Standard Error	32.2	31.8	86.3	22.8	23.8	39.3
Median	-17.4	-43.6	71.9	0.9	-8.7	31.2
Standard Deviation	115.9	95.4	211.4	55.9	103.9	152.4
Range	408.0	281.6	605.9	139.0	430.0	605.9
Minimum	-307.3	-170.0	-300.8	-16.3	-307.3	-300.8
Maximum	100.7	111.6	305.1	122.7	122.7	305.1
n	13	9	6	6	19	15

5.2.3.1. One way analysis of variance for different herbaceous vegetation types

Although the descriptive statistics and the visual analysis of the box plot showed that there were differences, a one way ANOVA test for significant difference in means among groups showed that there was no significant difference ($p>0.05$) in the biomass gain in the four herbaceous communities (Table 8). This result assisted in answering the research question that sought to find if there were significant differences in biomass gain in different herbaceous vegetation classes. However this result cannot be generalised for the whole year as the study was conducted for the June to September period. What can be drawn from these results is that between the June to September period we cannot clearly distinguish productivity classes by splitting the herbaceous vegetation into the four classes using this data. Maybe if the sample sizes had increased, the results could have been different. The number of plots sampled was low as compared to most similar studies that used more than 100 samples (Mutanga and Rugege, 2006; Zambatis *et al.*, 2006). However the 34 samples used for analysis met the minimum 30 samples standard for most statistical analysis.

Other options would be to do monthly assessments of productivity following Method 5, 6 or 7 of Scurlock, *et al.,* (2002) if "Total productivity" is to be established since Naivasha experiences a bimodal rainfall seasons. The other consideration could be probably to use species diversity in these vegetation communities as some studies have shown there is a relationship between species diversity and productivity (Stachová and Lepš, 2010). This relationship varies in different communities depending on the species present. The diversity to productivity relationship could be positive in communities where there is low interspecific competition. In such communities, productivity increases with increase in species diversity. Where the diversity pool is large and competition is also high chances are high that productivity would go down (Stachová and Lepš, 2010). This could have been a fundamental variable however it requires more time to measure species diversity and evaluate than was available.

Table 8: Summary of a One way analysis of variance for the four groups of herbaceous communities (i) Herbaceous in shrubland (ii) Dense sward (iii) Temporarily inundated and (iv) Herbaceous in forests

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	53324.93	3	17774.98	1.126901	0.353869495	2.922277191
Within Groups	473199.6	30	15773.32			
Total	526524.5	33				

Welch Two Sample t-test for comparison of open grasslands and grasslands under closed canopy
When comparison was conducted between open and closed communities, the results displayed a difference in the means of 28.6 g/m². Open grasslands had a higher biomass gain than closed grasslands. This could be credited to the fact earlier mentioned that grasses that grow in forested areas often grow tall with low biomass levels. Most of the herbivores within the fringe of Lake Naivasha are open plain grazers like wildebeests, zebras, and buffalos. These animals graze on the open plains impacting either positively or negatively on the behaviour of the herbaceous layer of areas (Milchunas and Lauenroth, 1993). Although this assumption cannot be verified from only these two sampling periods, the high biomass gain could have been the effect of grazing with animals opening up new spaces for shoots to develop.

In an attempt to find the possibility of classifying these two grasslands into either high or low at 95 % confidence interval a Welch two sample t test for comparing differences in means was used (Table 9). With a p-value of 0.5392 the results indicate that there were no significant differences ($p>0.05$) between biomass gain in open or closed herbaceous communities.

Table 9: Summary of the Welch two sample t-test analyses for the comparison of difference in means for open and closed canopy grasslands.

Mean (Open) g/m^2	*Mean (Closed) g/m^2*	*t-calculated*	*df*	*p-value*	*CI (%)*
12.3	-16.3	- 0.623	23	0.5392	95

These results could have been obtained because of the number of plots which could have been small to detect differences. On the contrary the open grasslands may have had more productivity than closed grasslands but the productivity could have been lost to grazing and termite activity. Within the closed forest the productivity could have been low and little grazing to promote growth. Some of the closed sample plots were in fenced areas devoid of grazing. This could have resulted in the low productivity levels hence a no significant difference result. However this study produced irreplaceable results for archiving regarding the means of the different vegetation communities.

5.2.4. Biomass gain as determined by dominant species inherent characteristics

Some interesting results came out after the analysis of how the biomass gain is determined by the inherent characteristics of the dominant species. Figure 22 shows the box plots of the differences of biomass gain in plots with highly preferred grasses and those with low palatability. The right side of Figure 22 shows biomass gain depending on whether dominant species increase or decrease with increased grazing intensity.

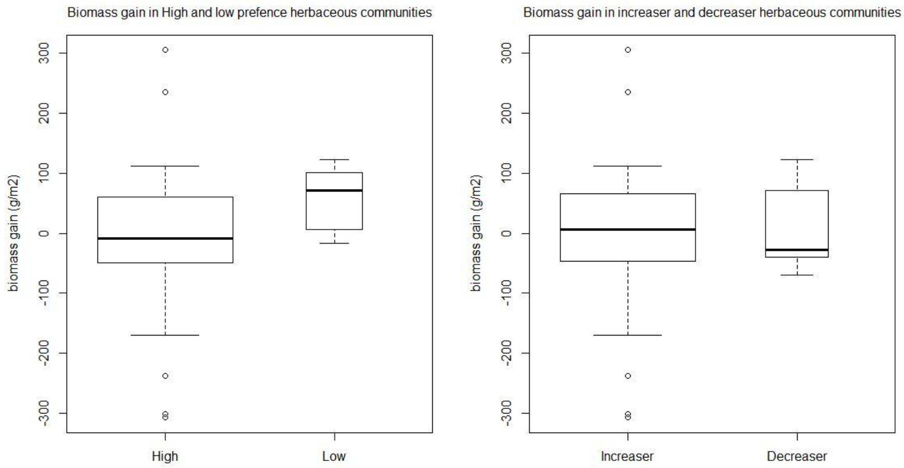


Figure 22: Box plots for biomass gain as determined by the inherent characteristics of the herbaceous dominant species (i) High-High preference grazing, (ii) Low-Low preference grazing (iii) Increaser-Increaser grass species dominant (iv) Decreaser-Decreaser grass species dominant

It was observed that the mean of the plots that are of low grazing preference like *Hypoestes forskahlii* had a higher mean biomass gain as compared to the plots with high preference grazing species like *cynodon* species and *Themeda trianda*. This result was interesting as it depicted the typical environmental scenario whereby the species that are most preferred tend to be grazed more than those that are less preferred by grazing animals. In such cases monitoring of species diversity becomes critical as the preferred species are likely to become extinct in those areas if overgrazed. On the other hand, the decreaser like *Themeda Triandra* species displayed a higher biomass gain as compared to increaser dominated plots like *Cynodon* species, however the large range showed how resilient increaser species are (Table 10). This concurred with the findings of Danckwerts and Stuart-Hill, (1987) who found that increaser species are adapted to grazing and in frequent fire whereas decreaser species are adapted to moderate grazing and infrequent fire.

Table 10: Summary descriptive statistics for biomass gain as determined by the inherent characteristics of the herbaceous dominant species (i) High-High preference grazing, (ii) Low-Low preference grazing (iii) Increaser-Increaser grass species dominant (iv) Decreaser-Decreaser grass dominant

Parameter	*High pref*	*Low Pref*	*Increaser*	*Decreaser*
Mean	-14.1	56.7	-5.5	4.8
Standard Error	24.6	26.9	25.7	30.6
Median	-8.7	70.9	5.7	-27.2
Standard Deviation	132.4	60.3	135.8	75.0
Range	612.4	140.1	612.4	192.4
Minimum	-307.3	-17.4	-307.3	-69.7
Maximum	305.1	122.7	305.1	122.7
n	29	5	28	6

Although for High versus Low preference, the P value was not that high (0.08), it might be that there is a difference, but that it cannot be detected with the current data (Table 11). The unequal sample sizes could have influenced the general output of the results. If the samples had been equal maybe the result could have been significant. Although this is the case descriptive statistics shown above shows variability in the groups which can be attributed to the inherent characteristics of the dominant grass species in that community.

Table 11 : Summary of the Welch two sample t-test analyses for the comparison of difference in means (i) High – Low preference grazing and (ii) Increaser – Decreaser species dominance

Compared means	*t-calculated*	*df*	*p-value*	*CI (%)*
(i) High – Low preference grazing	-1.9392	12	0.08	95
(ii) Increaser – Decreaser species	-0.2589	13	0.78	95

5.2.5. Biomass gain as determined by Management structure

Within the bounds of this study, it was difficult to discriminate plots according to management structures as the fenced areas were too few to draw meaningful conclusions about differences when compared to the unfenced. The other limitation was that most of the private farms are fenced mainly to limit movement of livestock belonging to different farms hence also making it difficult to separate the plots.

5.2.6. Biomass gain as determined by level of inundation.

Approximately 13879 ha are often covered by water (60-100%), 1178 ha (20-60%), 1630 ha (<20%) and 54198 ha of 54198 ha of the study area are never covered in water. For writing convenience in this analysis these zones shall be shall be termed Inundation Frequency 1 up to 4 (IF1, IF2, IF3, IF4) respectively.

Figure 23 gives the spatial location and coverage of these areas. The terrain and low elevation makes clear why the most of the plots in the north are frequently inundated than other plots elsewhere.

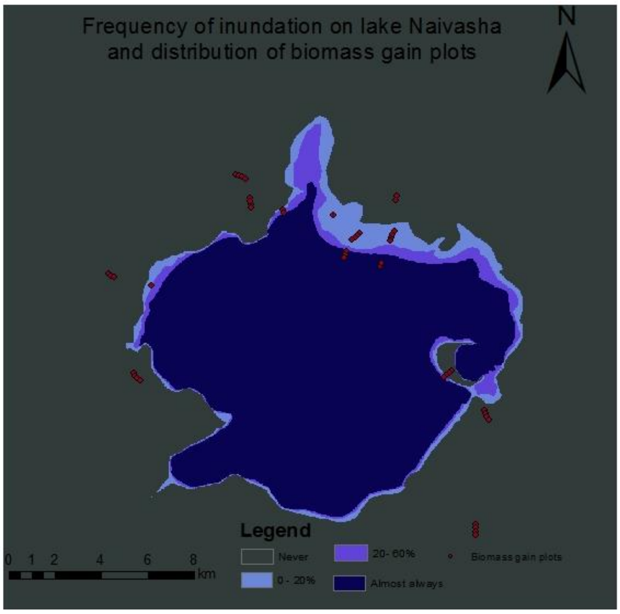


Figure 23: Inundation zones for the main lake

When analysis was conducted in-between these four classes, the descriptive statistics showed that there were differences in the biomass gained in different inundation zones. IF2 had the highest biomass gain approximately 160 g/m² and IF3 had the lowest (Table 12). The marginal effects of the water are clear from the box plots and these results show that the herbaceous layer is dependent on water. IF1 had a lower value as compared to IF2 probably because the high flooding frequency which causes reduction in productivity for a period. In addition within this zone it was expected that flooding in IF1 would be intense and covers the leaves of the grasses and may lead to death. This is in contrast to the other zones where the plants partially covered and able to photosynthesise. Much of the grazing occurs within zone IF1, IF2 and IF3 as compared to IF4. This reason could have caused the variations in the biomass gain within the zones. The average height of grasses in the fringe was 5.5 cm but increases in the temporarily flooded grasslands to 20cm. It would have been better to see how the productivity would be in areas where the grasses are fully covered (level of inundation) and those that are partially covered.

Table 12: Summary descriptive statistics for biomass gain as determined by frequency (i) IF1 (60-100%) (ii) IF2 (20-60%) (iii) IF3 (<20%) (iv) IF4 (0%)

Parameter	*IF1*	*IF2*	*IF3*	*IF4*
Mean	19.6	159.1	-95.3	12.1
Standard Error	113.6	146.0	47.7	14.0
Median	71.9	159.1	-74.6	0.9
Standard Deviation	227.2	206.5	135.0	62.4
Range	536.1	292.0	372.7	232.8
Minimum	-300.8	13.1	-307.3	-110.1
Maximum	235.4	305.1	65.4	122.7
n	4	2	8	20

However these results make it possible to classify these zones into high, medium and low productivity areas based on their productivity levels. The results show that productivity is highest in IF2 followed by IF1, IF4 and lastly IF3. Thus since IF1 and IF2 had positive biomass gains they could be classified as high productivity zones and IF3 and IF4 as low productivity zones. This therefore confirms the importance of lake level on determining productivity. However with the current data we cannot confirm whether productivity goes up or down with increase in lake level

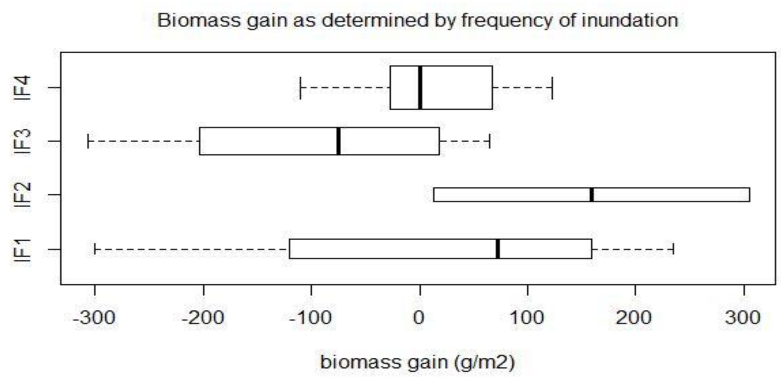


Figure 24: Boxplots showing the distribution of biomass gain as influenced by frequency of inundation

A one way analysis of variance to compare differences in the means between the zones showed that there were significant differences (CI = 95%) between the groups ($p<0.05$) (Table 13). This confirmed the visual results obtained in descriptive statistics as shown by the box plots in figure above. To view which groups were significantly different from each other a post hoc method following a one way ANOVA was used.

Table 13: Summary of a One way analysis of variance for the four groups as determined by frequency (i) IF1 (60-100%) (ii) IF2 (20-60%) (iii) IF3 (<20%) (iv) IF4 (0%)

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	127313.1	3	42437.69	3.189113	0.037784	2.922277
Within Groups	399211.5	30	13307.05			
Total	526524.5	33				

5.2.6.1. Turkey Kramer Post-hoc analysis

Tukey-Kramer Method for unequal sample sizes was used as a follow up method to the statistically significant one way ANOVA to check the groups that were significantly different. The results show that IF3 and IF2 are significantly different from each other ($p<0.05$) whereas all the other pairs are not significantly different (Table 14)

Table 14: Turkey's post-hoc multiple means analysis to find the significantly different means.

Means(μ_1-μ_2)	*p-value*
IF2 - IF1 == 0	0.4960
IF3 - IF1 == 0	0.3644
IF4 - IF1 == 0	0.9993
IF3 - IF2 == 0	0.0396
IF4 - IF2 == 0	0.3179
IF4 - IF3 == 0	0.1311

The difference in the groups IF3 and IF2 were not the hypothesised to have the most significant difference. It was expected that the plots that are never inundated would suffer water stresses more often that all the other plots. These results could be because of the other interacting factors that influence biomass gain like management structures. The conclusions that can be drawn from these results are that the zone IF1 and IF2 are key productivity area and need proper management to maintain their viability. on the contrary zone IF3 and IF4 need caution when formulating grazing regimes.

5.3. Effects of lake level on above ground biomass gain

5.3.1. Threshold distance analysis

To assess the effect of lake level change on the herbaceous biomass gain, understanding of the relationship between lake level and ground water was sought. Figure 25 shows that the lake level and ground water are highly correlated for an estimated distance of 9.5 km from the lake centre. It was clear from this result that when a direct relationship is sought using the exact same date and correlating the ground water measurement and lake level of that date there is no relationship. This is paramount to note as these measurements are taken at different geographical positions. Therefore correlating change in ground water to change in lake level by using the deviations produced better results.

Furthermore, there is a time delay in recharge of the ground waster aquifers after the lake level rises. (Abdulahi, 1999). Therefore a relationship can clearly be viewed when changes in lake level are correlated with changes in ground water level. Figure 25 shows that this relationship breaks down at approximately 9.5 km when the R^2 values from the model of deviations of lake level fit to the model of ground water level deviations. The coefficients plotted against distance give a graph that has an almost constant coefficient value 1. This is the gradient of the model which depicts a constant slope of high correlation. This distance is as measured from the centre of the lake in all directions. This distance is however not a true reflection of all the directions but gives a confident overview as to the threshold distance. This distance is likely to be increased in the North where it is flat. This is evidenced by the noise produced by the model after the 9.5 km threshold

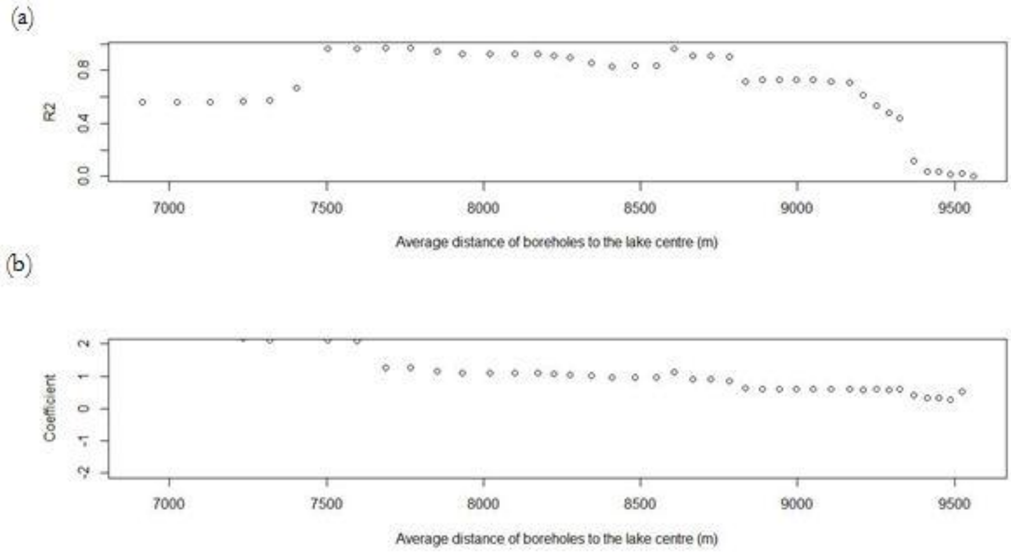


Figure 25: Relationship between deviations in lake level against deviations in ground water level plotted against distance from the lake (a) The coefficient of determination (R^2) against distance from the center of the lake and (b) Model coefficient plotted against distance from the center of the lake

5.3.2. Biomass gain as determined by ground water level

The median and mean of the biomass gain as determined by ground water shows a difference (
Figure 26). The mean for plots on shallow ground water tables (closer to the ground surface) had a higher mean than plots on deep water tables. This was the expected result as it was anticipated that the herbaceous vegetation would respond to the differences in ground water that is within the rooting zone.

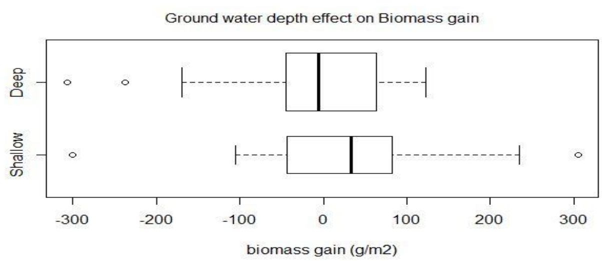


Figure 26: Box plots for biomass gain as determined by ground water depth shallow ground water depth (≤5m deep) and Deep water depth (> 5m deep)

Although this was the case a Welch t- test (Table 15) showed that there were no significant differences ($p>0.05$; CI = 95%). This variation could be attributed to the fact that the sampling size was small hence could have influenced the result. Apart from this fact sampling was conducted in a wet season and the September 2011 rainfall was more than the normal. This could have probably drifted the plants dependence on ground water to precipitation unlike if the test was conducted in the dry season. On the other hand since the rooting depth of the herbaceous vegetation is shallow, the direct influence of the ground water on the herbaceous unit could be limited. Ground water influences soil moisture quantity hence maybe the direct relationship between biomass gain and soil moisture could have yielded a better result. This was evidenced by the importance of bulk density as a variable in the multiple linear regression analysis. Bulk density is inversely related to porosity hence influences the soil moisture content.

The other option that could have improved the analysis of this assessment would have been to measure the evapotranspiration of these herbaceous vegetation units. The evapotranspiration would assist in understanding how the plants utilise the underground water for both the dry and wet seasons.

Table 15: Summary of the Welch two sample t-test analyses for the comparison of difference in means of biomass gain as influenced by shallow and deep ground water tables.

Mean (Shallow) g/m²	*Mean (Deep) g/m²*	*t-calculated*	*def*	*p-value*	*CI (%)*
-17.27	-16.3	0.8013	12	0.4386	95

5.4. Multiple Regression analysis of the biomass gain explanatory variables

The best regression equation ($R^2 = 0.51$) was obtained from a stepwise multiple linear regression model between land cover type, frequency of inundation elevation and bulk density. Three variables, soil cation exchange capacity, slope and ground water depth were eliminated from the model. The variables which remained in the model are as shown in Table 16 with very high level of significance and their respective model coefficients. These variables can be used as productivity indicators.

Table 16: The remaining variables from the stepwise showing their level of significance and an overall R^2

	Model Coefficients	*S.E*	*t-value*	*p-value (p = 0.01)*
Intercept	-673.82	1018.0179	-3.233	0.0044
Land cover(LC)	24.40	17.8729	3.448	0.0027
Inundation frequency (IF)	0.12	0.7052	3.026	0.0070
DEM (DEM)	0.35	0.5316	3.243	0.0043
Bulk density (BD)	-41.07	57.0078	-2.933	0.0085

Overall result: $R^2 = 0.51$; p-value = 0.0065

The variables that remained in the model were valid for explaining biomass gain and productivity as already reviewed in "Section 2.5" of this thesis. The land cover type influences biomass gain because of the variations in productivity among different species thus the dominant species would ultimately influence the productivity of that area. The dominant species and species diversity would influence the ultimate productivity (Stachová and Lepš, 2010). Frequency of inundation was also a critical variable as the amount of water that can be tolerated varies per species. The results of this research would have been more interesting if the height of inundation and the period of inundation was also established

In this model slope and ground water could have been excluded because of the relative homogeneity of these variables within the measured plots. The slopes within the the fringe zone are generally gentle on the north and steeper on the south and west where few samples were taken. This could have influenced the results of this regression. The same could also be said of ground water but measuring soil moisture content at plot level could have helped establish the influence of the ground water at plot level hence extrapolate to the whole fringe zone.

Validation using 50 models fitting the measured and the predicted had a highest R^2 of 0.7 and an average R^2 value of 0.4 was obtained. Detailed results of the validation are given in in Appendix 4 The visual summary of the validation results are given in Figure 27

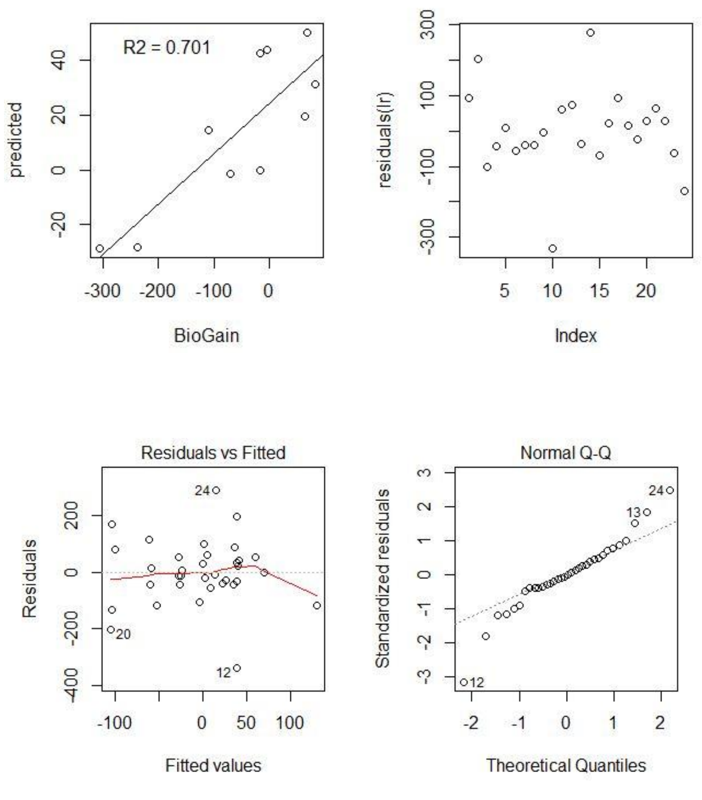


Figure 27: The validation graphs top – Highest R^2 of 50 models fit of the correlation between the predicted and the measured biomass gain using the 30% independent testing dataset. The top right and bottom plots show the plot of the residuals of the regression analysis between land cover type, inundation, elevation and bulk density.

The graphs in Figure 27 show goodness of fit of the measured to the predicted. The model fits very well with the data used for testing. There are a few outliers as shown by the fitted residuals and the normal QQ plot which validates the model as a relatively good model. However the average R2 (0.4) from the 50 runs fitting Training and Test data at random using the 70-30% ratio respectively was low. Although this is the case the model gave several other high values. This model can be conservatively be used to predict biomass gain. The results could have improved if the sampling size had been larger so that the model would have a larger training set. From the coefficients obtained in the model as shown in Table 16 we draw an equation that can be used to predict biomass gain and ultimately productivity. The equation is as given in

$$B_{gain} = 24.4\ LC + 0.12IF + 0.35DEM - 41.07BD - 673.82 \qquad \text{Equation 17}$$

Where: B_{gain} = Biomass gain, LC = Land cover type, IF= inundation frequency DEM = Elevation and BD = bulk density

As established from the validation results the equation is highly likely to yield results with the same average R^2 of 0.4. This equation can be used but would yield less accurate results. However the analysis highlighted the importance of these four variables as biomass gain indicators as well as productivity to estimate change in standing biomass. Thus the model can be improved by successive studies by increasing the accuracy of productivity estimation and include other variables like precipitation and grazing that seem to have very profound effects on herbaceous biomass gain.

The equation can be utilised to produce spatially explicit biomass gain and productivity maps in the fringe zone. Although in a study done by Mutanga and Rugege , (2006), they concluded that using Cokriging yielded better results than regression models, this equation is a first step towards improving the estimation of the herbaceous unit in the fringe zone of Lake Naivasha.

The results of the maps produced using this equation can be used in management plans with regard management strategies and animal usage of the area. This is crucial as this equation requires data that is readily available and can be used for estimation.

6. CONCLUSION AND RECOMMENDATIONS

6.1. Conclusions

The main aim of this research was to assess how the hydrology around Lake Naivasha influences the herbaceous productivity measured as biomass gain between June and September. To achieve this following research questions were answered. Because of the seasonal variability some of the conclusions from this study cannot be generalised for all the time periods of the year but can only be applicable to the June to September periods.

How separable are herbaceous vegetation types/ classes when mapping using ASTER imagery in a riparian ecosystem?

This research confirmed the strength of ASTER imagery to map riparian vegetation with high accuracy. The overall accuracy of the classification accuracy was 90% and a kappa coefficient of 0.89. This accuracy satisfies the standard requirements for practical applications. The produced map can be used to assess any vegetation changes either due to succession or disturbance in the future since it was produced at dominant species level. Several other succeeding and on-going studies can utilise this map output for their analysis.

It was possible to discriminate different herbaceous communities using expert knowledge and unsupervised classification. The herbaceous classes that can be discriminated using this method include Temporarily flooded grasslands, Dense sward of medium height, Short sparse grass and Bushed tufted grasslands. Although this is the case this method will not be possible if it were to be applied in discriminating the herbaceous unit at species level. This is mainly due to the similarity of grass characteristics and also the time resource required by this method.

With regard other land cover classes, this method is expected to be applicable in a wide range of remotely sensed images including very high resolution imagery. This approach limits the tendency of the automated classifiers to produce the so called “salt and pepper”. It can also be concluded that expert knowledge and ancillary data in classification improves classification of satellite images as compared to using only pixel based approaches. Due to the fragmentation of most landscapes with heterogeneous mixtures of agriculture, semi natural and natural vegetation covers, spectral based classifiers alone are not adequate. However within the study area built up areas and greenhouses can easily be manually digitised with high accuracy.

Are there significant differences (CI=95%) in herbaceous productivity over different vegetation classes?

Several attempts were made to classify the herbaceous vegetation into different classes. It can be concluded that between the period June and September it is not possible to statistically discriminate vegetation types into productivity classes using this data. Maybe if the sample sizes had increased, the results could have been different Biomass gain can be used as a proxy for productivity conservatively as it does not account for losses that occur in-between two successive periods. It was observed that there were no differences in the biomass gains in neither the four herbaceous classes nor when they were split into open and closed communities.

The inherent characteristics cannot be used to categorise the herbaceous vegetation into productivity classes. The tested characteristics include whether the dominant species are increaser or decreaser species and whether they are highly preferred or not. Maybe if the inherent characteristics of several species within a plot are considered instead of only the dominant species maybe the results can differ.

Within the bounds of this study, it was difficult to discriminate plots according to management structures as the fenced areas were too few to draw meaningful conclusions about differences when compared to the unfenced. The other limitation was that most of the private farms are in themselves fenced to limit mix of animals belonging to different farms hence also making it difficult to separate the plots. Therefore when comparing productivity using management structures it may be paramount to compare and report on the differences caused by farms themselves rather than the management structure itself. However this may have implications since it may become personal and cause conflict if negative results are obtained.

How does ground water influence the productivity of the herbaceous vegetation?

From the analysis conducted between shallow (closer to the surface) and deep water tables, ground water showed little direct influence on biomass gain. From the analysis between shallow and deep water tables, the herbaceous vegetation performed. The study period could have been too short to notice significant variations. On the other hand since the rooting depth of the herbaceous vegetation is shallow, the direct influence of the ground water on the herbaceous unit could be limited. Ground water and precipitation influence soil moisture quantity hence maybe the direct relationship between biomass gain and soil moisture could have yielded a better result. Therefore it can be concluded that ground water as a variable cannot be used to explain biomass gain however its counterparts, precipitation and soil moisture could be better explanatory variables

How does frequency of inundation influence productivity of the herbaceous vegetation?

Frequency of inundation has significant influence on the productivity. The results from the analysis of biomass gain as influenced by inundation showed that the patches that are often inundated (20-60%) are more productive than most of the patches in the study area. The patches that are 60-100% inundated have high biomass gain but less than the 20-60% patches. Those that are 20% to never inundated have the least biomass gain. Since the differences between the means of these groups were statistically significant, it can be concluded that these can be split productivity classes. Thus two classes are evident high productivity 20-100% and Low productivity (less than 20 to never). From this result, it can also be concluded that the grasslands that are immediate to the lake shore are highly productive as compared to grasses that are far from the shore. Future endeavors can utilise this result for analysis and comparison and to check for changes in productivities using these classes as bases for differences. Farmers can also relate this for farm animal usage. Care needs to be taken when assigning carrying capacities in never inundated areas as chances of overgrazing are high as compared to frequently flooded areas.

What threshold distance does nearness to the lake influence herbaceous vegetation productivity?

Overall y, within 9.5 Km from the lake centre there is high correlation between lake level and ground water. In general this value translates to approximately ±5 km from the lake. This distance increases in the North where the terrain is flat. However this same distance is not the same distance that the lake level influences the biomass gain. It also emerged from this analysis that using the same dates of lake level and ground water measurements in a regression produced very low R2 because these are measurements at two mutually exclusive geographic locations. Therefore it can be concluded that if such analysis is to be conducted using historical lake level and ground water levels the change in ground water to change in lake level by using their deviations produces better results.

What are the major drivers of primary productivity and how do they influence the productivity?

Four variables herbaceous vegetation community type, frequency of inundation elevation and bulk density were the best explanatory variables for biomass gain in the fringe zone of Lake Naivasha. These variables exhibit interaction effect and as established from the validation results the equation is highly likely to yield results with the same average R^2 of 0.4. This equation can be used but would yield less accurate results. Thus the model can be improved by successive studies by increasing the accuracy of productivity estimation and include other variables like precipitation and grazing that seem to have very profound effects on herbaceous biomass gain. The equation can be utilised to produce spatially explicit biomass gain and productivity maps in the fringe zone.

The results from the maps produced using this equation can be used in management strategies with regard animal usage of the area.

6.2. Recommendations

- Since classification was not the major focus of this study, a more detailed analysis of incorporating the expert knowledge into an expert system may be paramount for concrete reproducible output. Furthermore, method to standardise the quality and adequacy of the expert knowledge can further be explored.
- Productivity analysis can be conducted on monthly bases over a full year and taking the losses into consideration and compared the same variables over a long period using multi temporal images.
- Use of precipitation, evapotranspiration and soil moisture content as measurable determinants of how the herbaceous vegetation utilises water can further be explored. These variables can be evaluated in place of using ground water as an explanatory variable.
- Further in-depth study may be conducted on the relationship between species diversity and productivity
- It emerged from this study that frequency of inundation is one of the significant variables explaining biomass gain. Further analysis can be conducted using the period of inundation (for how long it is covered) and the level of inundation (whether the herbaceous vegetation is fully covered or partially covered by water).

List of references

Abdulahi, B. H. (1999). *Surface water - groundwater interaction, near lake Naivasha, Kenya.* ITC, Enschede.

Adams, C. S., Boar, R. R., Hubble, D. S., Gikungu, M., Harper, D. M., Hickley, P., andTarras-Wahlberg, N. (2002). The dynamics and ecology of exotic tropical species in floating plant mats: Lake Naivasha, Kenya. *Hydrobiologia, 488*(1-3
), 115-122.

Ahamed, T., Tian, L., Zhang, Y., andTing, K. C. (2011). A review of remote sensing methods for biomass feedstock production. *Biomass and Bioenergy, In Press, Corrected Proof.*

Åse, L. E., Sernbo, K., andSyrén, P. (1986). Studies of Lake Naivasha, Kenya, and its drainage area. . *Naturgeografiska Institutionen Stockholm, Sweden,, Stockholms Universitet in Stockholm.*

Batjes, N. H., andGicheru, P. (2004). Soil data derived from SOTER for studies of carbon stocks and change in Kenya (ver. 1.0; GEFSOC Project). ISRIC - World Soil Information,Report 2004/01,Wageningen

Becht, R., andHarper, D. M. (2002). Towards an understanding of human impact upon the hydrology of Lake Naivasha, Kenya. . *Hydrobiologia, 488(1)*, 1-11.

Becht, R., Odada, E. O., andHiggins, S. (2005). Lake Naivasha : experience and lessons learned brief. *In: Lake basin management initiative : Experience and lessons learned briefs. including the final report: Managing lakes and basins for sustainable use, a report for lake basin managers and stakeholders. Kusatsu : International Lake Environment Committe Foundation (ILEC), 2005. pp. 277-298.*

Bemigisha, J. R. (1998). *An assessment of the spatial and temporal distribution of a papyrus swamp : the application of remote sensing, GIS and a decision support system for the monitoring and management of the wetland vegetation of lake Naivasha, kenya* Unpublished MSc Thesis, International Institution of Geo-information Science and Earth observation (ITC), Netherlands, Enschede.

Bemigisha, J. R. (2000). Landcover mapping and multicriteria modelling for explaining the spatial and temporal distribution of a papyrus swamp at Lake Naivasha, Kenya. *International Archives of Photogrammetry and Remote Sensing, XXXIII*(Part B7), 165-174.

Bergner, A. G. N., Trauth, M. H., andBookhagen, B. (2003). Paleoprecipitation estimates for the Lake Naivasha basin (Kenya) during the last 175 k.y. using a lake-balance model. *Global and Planetary Change, 36*(1-2), 117-136.

Boar, R. R. (2006). Responses of a fringing Cyperus papyrus L. swamp to changes in water level. *Aquatic Botany, 84*(2), 85-92.

Boar, R. R., Harper, D. M., andAdams, C. S. (1999). Biomass allocation in Cyperus papyrus in a tropical wetland, Lake Naivasha, Kenya. *Biotropica, 31*(3
), 411-421.

Bortels, L., Chan, J. C. W., Merken, R., andKoedam, N. (2011). Long-term monitoring of wetlands along the Western-Greek Bird Migration Route using Landsat and ASTER satellite images: Amvrakikos Gulf (Greece). [Article]. *Journal for Nature Conservation, 19*(4), 215-223.

Bransby, D. I., andTainton, N. M. (1977). The disc pasture meter : Possible applications in grazing management. *Proceedings of the Annual Congresses of the Grassland Society of Southern Africa, 12*(1), 115-118.

Brinkmann, K., Dickhoefer, U., Schlecht, E., andBuerkert, A. (2011). Quantification of aboveground rangeland productivity and anthropogenic degradation on the Arabian Peninsula using Landsat imagery and field inventory data. *Remote Sensing of Environment, 115*(2), 465-474.

Brockett, B. H. (1996). Research note: Calibrating a disc pasture meter to estimate grass fuel loads on the Zululand coastal plain. *African Journal of Range & Forage Science, 13*(1), 39-41.

Canadell, J., Jackson, R. B., Ehleringer, J. R., Mooney, H. A., Sala, O. E., andSchulze, E. D. (1996). Maximum rooting depth of vegetation types at the global scale. *Oecologia, 108*(4), 583-595.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement. *20(1)*, 37-46.

Cook, A., andMerwade, V. (2009). Effect of topographic data, geometric configuration and modeling approach on flood inundation mapping. *Journal of Hydrology, 377*(1–2), 131-142.

Danckwerts, J. E., andStuart-Hill, G. C. (1987). Adaptation of a decreaser and an increaser grass species to defoliation in semi-arid grassveld. *Journal of the Grassland Society of Southern Africa, 4*(2), 68-73.

Dwyler, P. C. (2011). *Spartial estimation of herbaceous biomass using remote sensing in Southern African Savannas.* University of Witwarsrand, Johannesburg.
ESRI. (2011). ArcGIS10 Help. Retrieved 16 January, 2012, from http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html
FAO. (2003). Africover, LCCS : flexibility and standardization in land cover classification
Retrieved 18/11, 2011, from http://www.africover.org/LCCS.htm
FGDC. (1997). Vegetation classification standards. Retrieved 18/11, 2011, from http://www.fgdc.gov/
Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment, 80*(1), 185-201.
Fujisada, H., Urai, M., andIwasaki, A. (2011). Advanced Methodology for ASTER DEM Generation. [Article]. *Ieee Transactions on Geoscience and Remote Sensing, 49*(12), 5080-5091.
Gaudet, J. J. (1977). Natural drawdown on Lake Naivasha, Kenya, and the formation of papyrus swamps. *Aquatic Botany, 3*, 1-47.
Gitonga, M. S. (1999). *Study of long term waterbalance of lake Naivasha, Kenya.* ITC, Enschede.
Gong, X., Brueck, H., Giese, K. M., Zhang, L., Sattelmacher, B., andLin, S. (2008). Slope aspect has effects on productivity and species composition of hilly grassland in the Xilin River Basin, Inner Mongolia, China. *Journal of Arid Environments, 72*(4), 483-493.
Groen, T. A., van Langevelde, F., van de Vijver, C. A. D. M., de Raad, A. L., de Leeuw e, J., andPrins, H. H. T. (2011). Continental analysis of correlations between tree patterns in African savannas and human and environmental variables. *Journal of arid environments, 75*(8), 724-733.
Groen, T. A., Van Langevelde, F., Van de Vijver, C. A. D. M., Govender, N., andPrins, H. H. T. (2008). Soil clay content and fire frequency affect clustering in trees in South African savannas. *In: Journal of Tropical Ecology, 24(2008)3, pp. 269-279.*
Guo, D., Guo, R., andThiart, C. (2007). Predicting air pollution using fuzzy membership grade Kriging. *Computers, Environment and Urban Systems, 31*(1), 33-51.
Harper, D., andMavuti, K. (2004). Lake Naivasha, Kenya: Ecohydrology to guide the management of a tropical protected area. *Ecohydrology and Hydrobiology, 4*(GEOBASE), 287-305.
Harper, D. M., Harper, M. M., Virani, M. A., Smart, A., Childress, R. B., Adatia, R., Henderson, I., andChege, B. (2002). Population fluctuations and their causes in the African Fish Eagle, (Haliaeetus vocifer (Daudin)) at Lake Naivasha, Kenya. [Article]. *Hydrobiologia, 488*(1-3), 171-180.
Henderson, A. R. (2006). Testing experimental data for univariate normality. *Clinica Chimica Acta, 366*(1–2), 112-129.
Isaaks, E. H., andMohan Srivastava, R. (1989). *introduction to applied geostatistics*. New York etc.: Oxford University Press (OUP).
ITC. (2010). *GI science and earth observation : a process - based approach : also as e-book*. Enschede: University of Twente Faculty of Geo-Information and Earth Observation ITC.
Jianwen, M., andBagan, H. (2005). Land-use classification using ASTER data and self-organized neutral networks. *International Journal of Applied Earth Observation and Geoinformation, 7*(3), 183-188.
Keya, G. A. (1998). Herbaceous layer production and utilization by herbivores under different ecological conditions in an arid savanna of Kenya. *Agriculture, Ecosystems & Environment, 69*(1), 55-67.
Kibona, S. R. U. (2000). *Temporal and spatial varation of groundwater level north of Lake Naivasha, Kenya (Analysed using Modflow).* . Unpublished MSc thesis, ITC, Enschede, The Netherlands.
Lauenroth, W., Wade, A., Williamson, M., Ross, B., Kumar, S., andCariveau, D. (2006). Uncertainty in Calculations of Net Primary Production for Grasslands. *Ecosystems, 9*(5), 843-851.
Legese Reta, G. (2011). *Groundwater and lake water balance of lake Naivasha using 3 - D transient groundwater model.* University of Twente Faculty of Geo-Information and Earth Observation ITC, Enschede.
Li, B. G., Cao, J., Liu, W. X., Shen, W. R., Wang, X. J., andTao, S. (2006). Geostatistical analysis and kriging of Hexachlorocyclohexane residues in topsoil from Tianjin, China. *Environmental Pollution, 142*(3), 567-575.
Long, S., Garcia Moya, E., Imbamba, S., Kamnalrut, A., Piedade, M., Scurlock, J., Shen, Y., andHall, D. (1989). Primary productivity of natural grass ecosystems of the tropics: A reappraisal. *Plant and Soil, 115*(2), 155-166.
Lu, D., andWeng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing, 28*(5), 823-870.
Majozi, N. P. (2011). *Remote sensing of euphotic depth in lake Naivasha.* University of Twente Faculty of Geo-Information and Earth Observation ITC, Enschede.

Mavuti, K. M., andHarper, D. M. (2006). The ecological state of Lake Naivasha, Kenya, 2005: Turning 25 years research into an effective Ramsar monitoring programme. University of Nairobi.
McNaughton, S. J. (1993). Grasses and Grazers, Science and Management. *Ecological Applications, 3*(1), 17-20.
McNaughton, S. J., Ruess, R. W., andSeagle, S. W. (1988). LARGE MAMMALS AND PROCESS DYNAMICS IN AFRICAN ECOSYSTEMS. *Bioscience, 38*(11), 794-800.
Metera, E., Sakowski, T., Sloniewski, K., andRomanowicz, B. (2010). Grazing as a tool to maintain biodiversity of grassland - a review. [Review]. *Animal Science Papers and Reports, 28*(4), 315-334.
Milchunas, D. G., andLauenroth, W. K. (1993). Quantitative Effects of Grazing on Vegetation and Soils Over a Global Range of Environments. *Ecological Monographs, 63*(4), 327-366.
Moore, D. S. (2001). *Statistics : concepts and controversies* (Fifth edition ed.). New York: W.H. Freeman.
Moore, D. S., andMcCabe, G. P. (1998). *Introduction to the practice of statistics* (Third edition ed.). New York: Freeman.
Morgan, N. E. (1998). *Groundwater chemistry and quality assessment of the Lake Naivasha area, Kenya.* ITC, Enschede.
Morrison, E., andHarper, D. (2009). Ecohydrological principles to underpin the restoration of Cyperus papyrus at Lake Naivasha, Kenya. [10.2478/v10104-009-0036-6]. *Ecohydrology and Hydrobiology, 9*(1), 83-97.
Mutanga, O., andRugege, D. (2006). Integrating remote sensing and spatial statistics to model herbaceous biomass distribution in a tropical savanna. *International Journal of Remote Sensing 27*(16), 3499-3514.
Muthuri, F. M., Jones, M. B., andImbamba, S. K. (1989). Primary productivity of papyrus (Cyperus-papyrus) in a tropical swamp - Lake Naivasha, Kenya. *Biomass, 18* (1
), 1-14.
Nabide, I. K. (2002). *Development of 3 - D conceptual hydrogeological model for Lake naivasha area : based on the integration of geology, hydrochemistry, isotopic analysis and boundary conditions.* ITC, Enschede.
Naimi, B., Skidmore, A. K., Groen, T. A., andHamm, N. A. S. (2011). Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. *In: Journal of biogeography, 38(2011)8, pp. 1497-1509.*
Odland, J. (1988). *Spatial autocorrelation* (Vol. 9). Newbury Park etc.: Sage.
Okullo, P., andMoe, S. R. (2012). Termite activity, not grazing, is the main determinant of spatial variation in savanna herbaceous vegetation. [Article]. *Journal of Ecology, 100*(1), 232-241.
Phillips, D. S. M., andClarke, S. E. (1971). The calibration of the weighted disk against pasture dry matter yield. *Ibid, 33*, 68-75.
Ramirez Hernandez, R. (1999). *Groundwater flow modeling of Naivasha basin, Kenya.* ITC, Enschede.
Roques, K. G., O'Connor, T. G., andWatkinson, A. R. (2001). Dynamics of shrub encroachment in an African savanna : relative influences of fire, herbivory, rainfall and density dependence. *In: Journal of Applied Ecology, 38(2001), pp. 268-280.*
Schowengerdt, R. A. (1983). *Techniques for image processing and classification in remote sensing.* New York etc.: Academic Press.
Scurlock, J. M. O., Johnson, K., andOlson, R. J. (2002). Estimating net primary productivity from grassland biomass dynamics measurements. *Global Change Biology, 8*(8), 736-753.
Shad, R., Mesgari, M. S., abkar, A., andShad, A. (2009). Predicting air pollution using fuzzy genetic linear membership kriging in GIS. *Computers, Environment and Urban Systems, 33*(6), 472-481.
SIC. (2012). ASTER Satellite Imagery. Retrieved 01/01, 2012, from http://www.satimagingcorp.com/satellite-sensors/aster.html
Snee, R. D. (1977). Validation of Regression Models: Methods and Examples. *Technometrics, 19*(4), 415-428.
SPARVS Agency. (2008). *Final report on threat reduction assessment.* LIMURU, Kenya: USAID/PACT and NWC.
Stachová, T., andLepš, J. (2010). Species pool size and realized species richness affect productivity differently: A modeling study. *Acta Oecologica, 36*(6), 578-586.
Su, M.-C., Huang, D.-Y., Chen, J.-H., Lu, W.-Z., Tsai, L.-C., andLin, J.-Z. (2011). Mapping multi-spectral remote sensing images using rule extraction approach. *Expert Syst. Appl., 38*(10), 12917-12922.
Tainton, N. (1999). *Veld management in South Africa.* Pietermaritzburg: University of Natal.
Trollope, W. S. W., andPotgieter, A. L. F. (1986). Estimating grass fuel loads with a disc pasture meter in the Kruger National Park. . *Journal of the Grassland Society of Southern Africa*
4, 148-152.
Van Oudtshoorn, F. (2004). *Guide Grasses of Southern Africa.* Pretoria: Briza publications.

Vaughan, J. S. (1998). *Naivasha Lake levels done for the Lake Naivasha Riparian Association*. Naivasha, Kenya.

Vincent, C. E., Davies, T. D., andBeresford, A. K. C. (1979). Recent changes in the level of Lake Naivasha, Kenya, as an indicator of equatorial westerlies over East Africa. *Climatic Change, 2*(2), 175-189.

Whitley, R. J., Macinnis-Ng, C. M. O., Hutley, L. B., Beringer, J., Zeppel, M., Williams, M., Taylor, D., andEamus, D. (2011). Is productivity of mesic savannas light limited or water limited? Results of a simulation study. *Global Change Biology, 17*(10), 3130-3149.

Xavier, E. (2006). Ordinary multigaussian kriging for mapping conditional probabilities of soil properties. *Geoderma, 132*(1–2), 75-88.

Yihdego, Y. (2005). *three dimensional ground water model of the aquifer around lake Naivasha area, Kenya.* ITC, Enschede.

Yüksel, A., Akay, A., andGundogan, R. (2008). Using ASTER Imagery in Land Use/cover Classification of Eastern Mediterranean Landscapes According to CORINE Land Cover Project. *Sensors, 8*(2), 1237-1251.

Zalazar, L. V. (2006). *Comparison of classification approaches for land cover mapping in the Wielkoposka region, Poland.* ITC, Enschede.

Zambatis, N., Zacharias, P. J. K., Morris, C. D., andDerry, J. F. (2006). Re-evaluation of the disc pasture meter calibration for the Kruger National Park, South Africa. *African Journal of Range & Forage Science, 23*(2), 85-97.

Zhang, R., andZhu, D. (2011). Study of land cover classification based on knowledge rules using high-resolution remote sensing images. *Expert Syst. Appl., 38*(4), 3647-3652.

Zhao, X., andLi, S. (2012) Flood mapping using ASAR and landsat data in Poyang Lake. *Vol. 145* (pp. 114-118).

Appendix 1: List of materials used in the field

No	Type of material
1	Aerial photo (25 cm resolution) with plots printed on top
2	Laptop and Digital camera
3	Disk pasture meter
4	Garmin GPS 3m accuracy
5	iPAQ GPS
6	ASTER image 14 March 2011 (RGB: 342) with plots printed on top
7	Data collection sheets
8	Field guide book
9	GIS shape files

Appendix 2: R script for modelling the threshold distance that lake level has on ground water.

```
## SET THE WORKING DIRECTORY
setwd("B:/Scripts")

## LOAD THE DATA AND ORGANIZE IT
d <- read.csv("Boreholedata2.csv",sep=";",header=T)
names(d)[40:41] = c("dist2cntr","dist2shore")
names(d)[11:12] = c("watertable","lakelevel")
#==================================================================================
## SOME ADDITIONAL RE-ORGANIZING TO CORRELATE DEVATIONS FROM NORMAL IN THE GROUND
WATER TABLE  AGAINST DEVIATIONS FROM NORMAL IN THE LAKE LEVEL
##calculate the average water levels for the period in which station data is
available
d$avWatertable<-1
d$avLakelevel<-1
for(i in 1:length(levels(d$station)))
{
  ID=levels(d$station)[i]
  d[d$station == ID,names(d)]$avWatertable<-mean(d[d$station ==
ID,names(d)]$watertable)
  d[d$station == ID,names(d)]$avLakelevel<-mean(d[d$station ==
ID,names(d)]$lakelevel)
}
#==================================================================================
## calculating the deviations of the monthly observations from the averages
d$devLakelevel <- d$avLakelevel-d$lakelevel
d$devWatertable <- d$avWatertable-d$watertable
#==================================================================================

## KEEPING THE NUMBER OF RECORDS PER SUBSET FIXED, WORKING FROM THE NEAREST LOCATION
OF THE LAKE OUTWARDS

## SORTING AND ORGANIZING THE DATA INCLUDING SELECTING A SPECIFIC YEAR
d.sort<-d[sort(d$dist2shore,index.return=T)$ix,]
table(d$year)
d.sort<-d.sort[d.sort$year== 1999,names(d.sort)]

## SET SOME GENERAL PARAMETERS
sub.size <- 15
l <- length(d.sort$UTM_X)
#==================================================================================

## KEEPING THE NUMBER OF RECORDS PER SUBSET FIXED, WORKING FROM THE NEAREST LOCATION
OF THE LAKE OUTWARDS WITH THE DEVIATION DATA

## GROUNDWATER TABLE AS A FUNCTION OF LAKE LEVEL AT DIFERENT DISTANCES TO CENTER OF
THE LAKE

## SORTING AND ORGANIZING THE DATA INCLUDING SELECTING A SPECIFIC YEAR
d.sort<-d[sort(d$dist2cntr,index.return=T)$ix,]
table(d$year)
d.sort<-d.sort[d.sort$year!= 1999,names(d.sort)]

## SET SOME GENERAL PARAMETERS
sub.size <- 15
l <- length(d.sort$UTM_X)

## CREATE RELEVANT OBJECTS TO COLLECT PARAMETERS
## FOR EACH FITTED MODEL
RsqList<-NULL # R2 OF EACH MODEL
CoefList<-NULL # SLOPE COEFFICIENT OF EACH MODEL
DistList<-NULL # AVERAGE DISTANCE OF THE BOREHOLES IN EACH MODEL
```

```
## THE FITTING ALGORITHM
par(mfrow=c(2,2))
for(i in 1:(l-(sub.size-1)))
{
  d.subset<-d.sort[i:(i+(sub.size-1)),]
  model<-lm(d.subset$devWatertable~d.subset$devLakelevel)
  if((((i/l) %% 0.25) < 0.01) || (i == 1))
  {
    plot(  d.subset$devWatertable~d.subset$devLakelevel,
           main=paste(floor(mean(d.subset$dist2cntr))," m from the centre"),
           xlab="deviation in lake level (m)",
           ylab="deviation in water table in borehole (m)")
    abline(model,col="red")
    abline(a=0,b=1, lty=2)
  }
  RsqList<-rbind(RsqList,summary(model)$r.squared)
  CoefList<-rbind(CoefList,model$coefficients[2])
  DistList<-rbind(DistList,mean(d.subset$dist2cntr))
}
par(mfrow=c(1,1))

## PLOTTING THE RESULT AGAINST DISTANCE TO THE LAKE centre
par(mfrow=c(2,1))
plot(DistList,RsqList,xlab="Average distance of boreholes to the lake centre (m)",
ylab="R2")
plot(DistList,CoefList,xlab="Average distance of boreholes to the lake centre (m)",
ylab="Coefficient", ylim=c(-2,2))
par(mfrow=c(1,1)
#==================================================================================
```

Appendix 3: R Script for the multiple linear regression model

```
##Set working directory and organise data
setwd("B:/Scripts")
b<-read.csv("Regression.csv")
d1<-data.frame(b)
#===================================================================================
##Split the data into 70% Training and 30% Test and the drawing of samples to be
random
n <- round(nrow(d1) * 0.7) # n is the number of records for 70% of data. round is
used to avoid decimal in n!
n #Sample function can be used to random selection from a set of data.
s <- sample(1:nrow(d1), n) #s includes n randomly selected numbers from 1:nrow(dfn).
So it represents the row numbers that should be kept to have randomly selected rows.
Train <- d1[s,]
Test<-d1[-s,]
#===================================================================================
##Fit the multiple linear regression model
lr<-lm(BioGain~Grassland+Inundation+DEM+slope+GWD+BULK+CECS,data=Train)#linear model
lr1<-step(lr)# Stepwise regression
summary(lr1)
#===================================================================================
##Validation of the model by correlating the predicted by the measured
lr.p<-predict(lr1,newdata=Test,type=("response")) #fitting the test data
df.eval<-data.frame(lr.p,Test$BioGain)
colnames(df.eval)<-c("predicted","BioGain")
validation <- lm(BioGain ~ predicted , data = df.eval)# validation of the developed
model
summary(validation)
(rmse <- sqrt(mean(residuals(lr1)^2))) calculating RMSE
par(mfrow=c(1,2))
plot(predicted ~ BioGain ,data = df.eval)
abline(fit <- lm(predicted ~ BioGain, data=df.eval))
legend("topleft", bty="n", legend=paste("R2 =", format(summary(fit)$r.squared,
digits=3))) # pasting the R2 value to the graph
plot(residuals(lr1))
#===================================================================================
##Further evaluation of the model by fitting the obtained explanatory variables to
the full dataset
modelfit<-lm(BioGain ~Grassland+Inundation+DEM+BULK, data=d1)
modelfit
summary(modelfit)
plot(modelfit)#testing for heteroscedasticity
#===================================================================================
```

Appendix 4: Result of the multiple linear regression analysis

```
> lr<-lm(BioGain ~Grassland+Inundation+DEM+BULK, data=Train)# linear model
> lr
=============================================================================
Call:
lm(formula = BioGain ~ Grassland + Inundation + DEM + BULK, data = Train)

Coefficients:
(Intercept)    Grassland   Inundation          DEM         BULK
  -673.8199      24.4029       0.1233       0.3545     -41.0700
==============================================================================
> lr.p<-predict(lr,newdata=Test,type=("response")) #fitting the test data
> df.eval<-data.frame(lr.p,Test$BioGain)
> colnames(df.eval)<-c("predicted","BioGain")
> validation <- lm(BioGain ~ predicted , data = df.eval)
> summary(validation)

Call:
lm(formula = BioGain ~ predicted, data = df.eval)

Residuals:
   Min     1Q Median     3Q    Max
-88.18 -60.35 -16.47  65.05  99.70
==============================================================================
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -109.5356    27.2030  -4.027  0.00381 **
predicted      3.8203     0.8812   4.335  0.00249 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 76 on 8 degrees of freedom
Multiple R-squared: 0.7014,    Adjusted R-squared: 0.6641
F-statistic:  18.8 on 1 and 8 DF,  p-value: 0.002494
```