UNIVERSITY OF NAIROBI

SCHOOL OF MATHEMATICS

Species Distribution Model: A Case Study of Brown Ear Tick (*Rhipicephalus appendiculatus*) in Kenya.

This research project is submitted to the School of Mathematics of the University of Nairobi in partial fulfillment of the requirement for the degree of Masters of Science in Social Statistics.
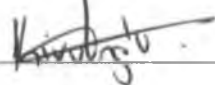
By

Kirung'o Martin Njogu.

©July 2012.

DECLARATION

This project as presented in this report is my original work and has not been presented for any other university award.

Kirung'o Martin Njogu.   **Reg. No.: I56/63752/2010.**

Signature: _____   Date: **30** / **07** /2012

This project has been submitted as a partial fulfillment of the requirements for Masters of Science in Social Statistics of the University of Nairobi with my approval as the university supervisor:

John M. Ndiritu

Signature: _____   Date: **30**/**07**/2012

DEDICATION

I dedicate this report to my mother. Through this report I have a word on your motherhood. "The greatest moral force in history is your motherhood. My childhood was directed by its force; youth kept pure and honorable by its sweet dominance; and mature age find its influence regnant, shaping character even to the end. Mrs. Kirung'o you are indeed the title of a woman's supremacy dignity", May you live to see the sweet fruits.

Thanks to my wife for moral support during the academic journey and my son Cavin for keeping me awake.

ACKNOWLEDGEMENT

ABSTRACT

A major method for analyzing resource selection of species ,predicting and mapping habitat suitability is by use of a Geographic Information System(GIS).This research combines data relating to *Rhipicephalus appendiculatus*(Brown Ear Tick) in Kenya with spatially explicit environmental factors to understand its resource selection and habitat preference using a logistic model as developed from the generalized linear model(GLM).Resource selection among the brown ear tick was similar in that across all sub-regions, they selected areas which has similar related factors in terms of the surrounding. The tick actively avoided at least one source of human threat including human settlements in all four of the studied regions and inhabited those areas that ecologically favour it.

The Brown Ear Tick *(Rhipicephalus appendiculatus)* which is a threat to livestock has not been studied until more recently in Zambia where a sudden and severe outbreak of Brown Ear Tick in the Central and Eastern areas of Zambia was reported. In Kenya its effect and spatial occurrence has not been understood. Resource selection analyses such as the one conducted in this study could help with planning of proper conservation if need be, prevention corridors for diseases control and manifestation in new areas especially as threats to livestock and wildlife increases.

In addition, a GIS was used to construct an expert-based habitat suitability map for the brown ear tick across Kenya using a spatial model approach. The map integrates spatial information on biological habitat requirements of the species with information on severity of threats to its prey. This can serve as a useful tool for determining future brown ear tick prevention prioritization in Kenya.

Table of Content

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

A.I.C-Alkaike Information Criterion.

G.A.M-Generalized Additive Model.

G.I.S-Geographic Information System.

G.L.M-Generalized linear models.

I.L.R.I -International Livestock Research Institute.

K.W.S-Kenya Wildlife Service.

K.N.M-Kenya National Museums.

S.D.M-Species Distribution model.

CHAPTER ONE:

INTRODUCTION

1.0 Introduction

Rhipicephalus appendiculatus is a hard tick found in the ears of cattle, other livestock, buffalo and antelope. This tick is considered to be a major pest in areas where it is endemic. Heavy infestations can cause anaemia, severe damage to the ears, or toxicities that result in the loss of resistance to some tick-borne infections.More than a thousand ticks have been found on some animals. *R. appendiculatus* can transmit a number of pathogens including *Theileria parva* (East Coast fever), Nairobi sheep disease virus and Thogoto virus (Arthur DR, 1961).

R. appendiculatus mainly infests cattle, buffalo and large antelope, but it can occur on other species including sheep and goats. Immature ticks may also be seen on small antelope, carnivores, hares and other species.*R.appendiculatus* prefers relatively cool, shaded, shrubby or woody savannas or woodlands with at least 24 inches of annual rainfall. This tick occurs in parts of eastern, central and southeastern Africa, and can be found from sea level to 7400 feet (2300 meters).Its distribution within this area is limited to suitable environments with appropriate hosts. It is a three-host tick. These ticks can be found on the host for several days while they feed, then they drop to the ground to develop to the next stage.

1.1 *R. appendiculatus* Control Measures

Measures used to exclude exotic ticks from a country include pre-export inspection and certification that the animals are free of ectoparasites, quarantines upon entry, and treatment with acaricides (Madder M,2005).Three-host ticks can be very difficult to eradicate once they have become established. In endemic areas, acaricides can eliminate these ticks from the animal, but do not prevent re-infestation. Three-host ticks spend at least 90% of their life cycle in the environment rather than on the host animal; ticks must also be controlled in the environment to prevent their spread. This research aims at providing the statistical ground for such control mechanisms.

In Kenya, livestock are threatened by the presence of brown ear tick .The presence of favourable environmental conditions in some parts of the country are where they inhabit causing a major threat to both livestock and tourism. This has an adverse effect to the two pillars of the Kenyan economy that is agriculture and tourism as they infest livestock and antelope. When both diminish, milk, meat and skin production goes down whereas decline of antelopes due to tick manifestation reduces wild games hence tourism loss. Stepping up efforts to control this vital species, therefore, is of greatest urgency. Increasing our knowledge of brown ear tick habitat selection and predicting their geographic distribution in Kenya using spatially explicit information and maps created by a Geographic Information System (GIS) is one step toward aiding in their control. To that aim, this analysis examines environmental factors contributing to brown ear tick habitat preference and uses a modeling approach to create a prediction map of brown ear tick suitability in other parts of Kenya. It is hoped that this research will aid in geographic targeting of control strategies for controlling brown ear tick and for biodiversity in Kenya.

1.2 Background of the Study.

Understanding how wildlife use surrounding habitats is of paramount importance to ecology and species management(McClean et al.1998,Boyce and McDonald 1999).Habitat selection studies usually compare assessments of habitat use to habitat availability,showing how animals actively select the environments where they spend most of their time (Manly et al. 1993).

Interest in species distribution models (SDMs) of plants and animals allows us to potentially forecast the effects on patterns of biodiversity at different spatial scales .The question on how plants and animals are distributed on earth is a vital one as its results helps in analyzing their presence,history,adaptations and predictions. Early identification of species that pose a significant threat of becoming invasive contributes to effective management of those same species. By this, we control the plant or animal community structures and ecosystem functions thus controlling the economic costs from the effects felt on this ecosystem (Herron et.al.2007).

Species distribution models (SDMs) are common in conservation biology, ecology and wildlife management. In this regard, the nature is studied and the status of Earth's biodiversity with the aim of protecting species, their habitats and the ecosystem from excessive rate of extinction. The dispersal,

migration,demographics,effective population size, inbreeding ,depressions and minimum population viability of rare or endangered species can be studied via the knowledge of SDMs.The phenomena that affect the maintenance, loss and restoration of biodiversity and the science of sustaining evolutionary processes is put into focus since if we don't preserve what is there, the planet can loose some of its species contributing to poverty, starvation and this can reset the course of evolution on this planet. By fitting the SDM model, we look at the pattern-recognition approaches whereby associations between geographic occurrence of a species and a set of predictor variables are explored to allow or support statements of the mechanisms governing species' distributions(Arau´jo,2006).The interlinked stages that need to be checked include: Compile data on the biodiversity of the planning region; Identify conservation goals for the planning region; Review existing conservation areas; Select additional conservation areas; Implement conservation actions and Maintain the required values of conservation areas.

The species models allows analyses of such data which is considered to be of large spatial point sets thus allowing estimation of species ecological requirements, by combining observations of species occurrence or abundance with environmental estimates(Elith et.al,2009).Constructing an SDM relies on a description of the species' relationship with its environment to depict areas within a region of interest where the species is likely to occur and thus useful tool for estimating the potential for species to occur in areas not previously surveyed (Guisan and Thuiller 2005 and Hernandez et.al,2008).The analytical part then models the species and their habitat at a spatial scale.The limitations is only that of environmental data availability and species locality which is so minimal for ideal modelling.

Some species can cause immense economic value in the countries when introduced and at the same time wreak havoc in the ecosystem ,resulting in the loss of endemic species and alter ecosystems with knock on effects for livelihood.In Africa for example some important ecosystems are under threat,consequently undermining development and livelihood opportunities, increasing human vulnerability and threatening human well-being(Chenje,2002) due to invasive species.This affects Africa's ability to meet Millenium development goals.

The challenges,however encountered while building on SDMs models cannt be over emphasized. Spatial misalignment problems are likely to occur when the environmental data is related to the species distribution data and the region where prediction is based.In addition, degree of correlation among observations depends on their relative locations and further there is always problem in quantifying uncertainity in spatial models hence the predictive capabilities

In Kenya,species distribution is also a key part of the ecosystem and the challenges mentioned above are not exception.The Strategy for Conservation and Management of Commercial Aloe Species in Kenya is an example of government attemps to focus on species.This can guide sustainable conservation and utilization of the species resource in the country. The strategy focuses on striking a balance between socio-cultural, economic and ecological needs as the core pillars of sustainable development.The aim here is to integrate interests in biodiversity conservation and economic development.The government through the National Museums has the role to correct such data involving various species in its soil.The Kenya Wildlife Service has also established a draft strategy for management of the invasive species in protected conservation areas.

It is against this background that I wish to establish the model for the species in Kenya.Specifically,this research is aimed at studying *Rhipicephalus appendiculatus* commonly called brown ear tick due to its stong influence to both agricultural and tourism aspect in the Kenyan economy.If uncontrolled,(Arthur DR,1961) *Rhipicephalus appendiculatus,* targets cattle,horses,sheep, goats,antelopes,dogs and rodents transmitting diseases like *Theileriaparva(*East coast fever*), Hepatozoon canis and Theileria mutan.*The ticks occur in the east,central and coastal areas.They prefer savannah habitat with trees. Cattle may get heavy infestations of all stages: the adults cluster in the ears where they cause the "bleeding ear" syndrome. Immature ticks are found on the head and neck. In game, especially Eland and Kudu, the ears are the most affected.

1.3 Problem Statement

Species are an important part of ecosystem.They play an important role in the survival of other species and also contibuting heavily to environmental and landscape features.
Understanding how species use surounding habitats is of paramount importance to ecology and their management.Habitat selection studies usually compare assessments of habitat use to habitat

availability, showing how animals actively select the environments where they spend most of their time (Manly et al. 1993) while species distribution models can be used to analyze intensity of resource use and predict the geographical distribution of a particular species by combining information from point occurrence data and environmental variables (Boyce et al. 2002, Graham and Hijmans 2006).

To achieve this,a statistical approach can be used to model the species distribution.For example generalised linear models have been used to model wildlife distributions (Walker 1990, Osborne and Tigar 1992,Bucklandand Elston 1993).Due to their ability to provide a solid statistical foundation for realistically modeling ecological relationships(Austin 2002),GLMs are useful for species distribution(Janet R. Nackoney).A study region is divided into grid cells and data on presence/abundance can then be determined by relating the response variable to spatially referenced covariates(Augustin et al. 1996, Boyce et al. 2002).

In Kenya,species are threatened by mounting pressure from humans including habitat fragmentation.In addition,invasive species can be a threat if not identified early enough for ease of management.The future of species in Kenya is uncertain today and thus stepping up efforts to protect,conserve,map and predict their existence is of greatest urgency.Further,the species relative location contributes on spatial autocorrelation and thus quantifying uncertainity during prediction is an important focus.The knowledge of species distribution modelling and use of maps by Geographic Information system is one step towards aiding the conservation,establishment and prediction of such species.

To that aim,this reserach examines the environmental factors contributing to species with a strong focus to the brown ear tick biologically called *Rhipicephalus appendiculatus* and uses a mathematical modelling approach to create *Rhipicephalus appendiculatus* species distribution model(SDM) that helps in prediction.It is hoped that the reseach can assist the concerned stakeholders like Kenya Government through the Museums of Kenya and International Livestock Research Institute (ILRI) in establishing and or controlling new environments for the brown ear tick,establish conservation measures and creating biodiversity. -

1.4 Objectives:

The objective of this study shall be divided into two:General and Specific Objectives

1.4.1 General Objective

The general objective of the study is:

To develop a species distribution model to be used for describing the *Rhipicephalus appendiculatus* patterns as well as making predictions.

Species Distribution models assist in characterising the natural distribution of species .In addition,when when well designed survey data and functionally relevant predictors are analysed with a good model(Elith et al,2009),the results can contribute well to this objective thus providing useful ecological insight and strong predictive capability.

1.4.2 Specific Objectives.

The research will aim at three specific objectives

i. **To establish the influences of environment on *Rhipicephalus appendiculatus* distributions:**

   The presecne/absence of a species depends on particular environmental conditions.Studies describe biological patterns in relation to environmental gradients(Grinnell 1904, Murray 1866,Schimper 1903).Sessile species have capabilities of characterising its environment while mobile species use the resources that are patchily distibuted across a landscape(Elith et al,2009).Thus this reseach aims at looking into those factors that make the brown ear tick to inhibit a specific area i.e choice of those species in relation to the environment.The major environmental concerns are rainfall,temperature and forests.Other relating factors that the research will establish are cattle density and East coast fever distribution.

ii. **Establishing the spatial dependancy of *Rhipicephalus appendiculatus:***

   Data are considered to be autocorrelated when the degree of correlation among the observations depends on their relative locations.

   i.

   The scale disparities when fitting model is an important area in species distribution.Again ,the scale of pattern in ecological data is an important aspect in any modelling as it helps in evaluating the inherent structures in data.This reseach aims at establishing a set of scale –dependent

6

predictors to represent factors affecting the distribution of the *Rhipicephalus appendiculatus* at more than one spatial scale.This objective will be enhanced further by use of GIS for mapping the species to see their distribution in those sampled localities. Geographic Information System (GIS) is a useful tool for analyzing resource selection of species and predicting and mapping habitat suitability when the degree of correlation among observations depends on their relative locations, that is the problem of spatial dependence.

iii.   **Predicting new habitats for *Rhipicephalus appendiculatus:***

The research aims at establishing a model to predict the new and unsampled domains.This will be done by assessing the environment in these new times and places ,particulary for new combinations of predictor values or for predictor values outside the given data.This prediction to new environments is generally termed extrapolation or forecasting (Arau`jo,2007, Miller et al. 2004).The model thus established will aim at capturing those features that will reflect major attiributes of the data like the spatial autocorrelation(as shown in second objective).The research will show that there is a strong case for making a species model spatially explicit.

1.5 Signifance of the study.

Understanding how wildlife use surrounding habitats is of paramount importance to ecology and species management(McClean et al. 1998, Boyce and McDonald 1999).Habitat selection studies usually compare assessments of habitat use to habitat availability,showing how animals actively select the environments where they spend most of their time (Manly et al. 1993).

Upon fulfillment of its objectives, this research will be an important tool in shaping the distribution of various species in the Kenyan context. Though there are various models that can model the species, researchers have proved GLMs as one of such tools whose results can be used to model various other species.

The results can be used to establish conservation strategies as well as in preservation of the ecosystem. By this better fruits of conservation will be seen not only via economic empowerment of the citizens but also positive concern to the environment. Further, by identifying species that pose a significant threat of becoming a threat to economy in advance will contribute to effective management of those same species. To the Kenya National Museum, it can improve domestic awareness, attracting more resources allocation from the government hence better planning via

conservation and this will be a mark of national well being.To ILRI,this can assist in putting control measures for *Rhipicephalus appendiculatus* hence assist in control of livestock related disease. To KWS, it is a great relieve management as it can assist in antelope and other wildlife conservation by reducing the threats from brown ear tick. Its products will be used by future generations to identify any areas of improvements in any future modeling of species.

The maps in this research integrates spatial information on biological habitat requirements of the species with information on severity of human threats. Increasing our knowledge of species habitat selection and predicting their geographic distribution using spatially explicit information and maps created by a Geographic Information System (GIS) is one step toward aiding in their conservation. To that aim, this analysis examines environmental factors contributing to *Rhipicephalus appendiculatus* preference and uses a modeling approach to create a prediction map of this species suitability in Kenya.It is hoped that this research will aid in geographic targeting of conservation strategies for protection and control of *Rhipicephalus appendiculatus* and for biodiversity in general.

The next chapter will now explore what others have written and researched upon regarding the species distribution and how their work will build on this research and improve on the same where need be.

CHAPTER TWO:

LITERATURE REVIEW

2.0 Introduction.

This chapter provides an overview of species distribution and global commitments towards species conservation, prevention and control. It also reviews the current efforts made in species prediction, mapping, explores the concept of conservations and discusses related research undertaken in the area of species mathematical modelling. The literature also elaborates on the concept of social threats and the theoretical underpinnings of this study.

2.1 The General Overview of Species Modeling.

In prediction of where species are likely to occur, conservationists are increasingly relying on distribution models.The knowledge is being extended to predict the response of species to climate changes, and identifying conservation areas. However,Modeling can be challenging in these cases because locality data necessary for model formation are often scarce and spatially imprecise.

Hernandez et al(2008) observes that constructing an SDM relies on a description of the species' relationship with its environment to depict areas within a region of interest where the species is likely to occur.In this regard,it is strived to utilise the information on species occurrence /non-occurrence in prediction of the likelihhod of the species presence or absence in unsampled locations(Latimer et.al,2006).In estimating the potential for species to occur in areas not previously surveyed,SDMs becomes an important tool(Guisan and Thuiller 2005).These SDMs have conservation utility as they can be used to identify high priority sites for conservation (Arau'jo and Williams 2000; Ferrier et al. 2002;Loiselle et al. 2003; Wilson et al. 2005), direct biologicalsurveys towards places where species are likely to be found (Raxworthy et al. 2003;Engleret al. 2004; Bourg et al. 2005) and provide a baseline for predicting a species' response to landscape alterations and/or climate change (Thuiller 2003; Arau'jo et al. 2006).

Predictive species distribution models are essential tools in biodiversity conservation and management (Côté & Reynolds, 2002). Fitting an SDM involves a series of steps, each requiring a

number of choices and well-justified decisions (Ferrier *et al* .,2002b;Guisan &Thuiller, 2005). Species distribution models (SDMs) can thus perform well in characterizing the natural distributions of species (within theircurrent range), particularly when well-designed survey data and functionally relevant predictorsare analyzed with an appropriately specified model(Elith et.al,2009).They provide strong capabilities to predict from the ecological insight theat tey depict.

Roots of SDMs can be traced in early studies describing biological patterns in terms of their relationships with geographical and environmental ingradients(Grinnell 1904, Murray 1866,Schimper 1903).Researches have provided strong conceptual argument for modelling species rather than communinities for not only vegetation but also birds(Whittaker 1956; and MacArthur 1958).

2.2 Development and Diversity in the Species Modeling Landscape

Applications that use inadequate data are challenging not forgetting those that model species not in equilibrium with their environment,that extrapolate in time and space.Alot of literature has explored applications that fit species which are also invasive as for example Invasive plants can alter plant community structure and ecosystem function (Vitousek *et al* ., 1987), result in large economic costsfrom lost ecosystem services (Pimentel *et al* ., 2005), and detract from an intrinsic or aesthetic value associated with native biodiversity and native plant dominance(Herron et.al,2007).SDMs can be used to provide understanding and even in prediction of species distribution across landscape.

SDMs normally take different terms depending on the context of the meaning taken and the emphases there on.They can thus be termed as bioclamatic models,climate envelopes,ecological niche models,habitat models e.tc(Elith et.al,2009)

Hernandez et al(2008) shows how species distibuition can be predicted from a poory studied landscape where comparison of the success of three algorithms(Maxent,MahalanobisTypicalities and Random Forests) at predicting distributions of eight bird and eight mammal species endemic to the eastern slopes of the central Andes was done.He observed that the three methods performed similarly for species with restricted distributions and the three methods have both strengths and weaknesses in some situations.

A biologist familiar with the species can define the species –environment relationship or analyses of the environmental conditions at points of occurrence in a statistical anlaysis can also define the same.This however will require construction of a defination of the species' relationship with the said environment.The analytical approach can be used to model specieswhose habitat requirements are poorly understood and can be developed at any spatial scale(Hernandez et al,2008).The only limitation is the data availabilty and species locality.In poor landscapes,data are often not collected systematically but gathered in an adhoc manner.Data from species can span many years.Due to lack of surveys also,the number of records available for any species is usually limited and because of species' relatively spatial distributions,the observations are not frequent.

Modeling with small numbers of spatially imprecise localities is challenging but not impossible (Pearson et al. 2007).Some methods have proven to be more effective under certain modeling conditions than others (Elith et al.2006; Hernandez et al. 2006) and according to Guisan and Thuiller 2005,they are several species distibution modelling methaods available.

Comparative analyses have investigated the efficacy of differennt methods for species distibuitions(Segurado and Arau'jo 2004; Elith et al. 2006;Hernandez et al. 2006; Tsoar et al. 2007).Some methods are effective at predictying species' distribution than others and none has been proven as best.The factors that influence model performance include quantity and quality of the species occurrence data, the accuracy and completeness (i.e.inclusion of all relevant factors contributing to the processes driving the species' distribution pattern) of the environmental data, the spatial scale (extent and size of analysisunit), and the ecological characteristics of the species being modeled(Segurado and Arau'jo2004; Elith et al. 2006; Hernandez et al. 2006; McPherson and Jetz 2007).For example in modelling species from poorly surveyed landsacpe, purpose may be to generate potential distribution maps for many species with as much confidence as possible, thereby providing baseline biological diversity information previously unavailable (Hernandez et al,2008).
Previous comparative studies demonstrated that Maxent, a statistical mechanics approach, performs very well (Elith et al.2006;Phillips et al.2006) even with small samples(Hernandez et al.2006).
Mahalanobis Typicalities, a method adopted from remote sensing analyses (IDRISI 2006), and Random Forests, a model averaging approach to the non-parametric procedure classification and

11

regression tree (CART) (Breiman 2001) also perform well. Researchers have demonstrated that both methods can produce useful results.

In plant species modelling,revealing mechanisms that allow a large number of plant species to coexist is of key interest within community ecology (Murrell et al.,2001; Loreau et al., 2001). Species coexistence is directly linked to local inter- and intra-specific interactions in a community (Durrett and Levin,1998). For example, each plant has a dependence on local growing conditions (Stoll and Weiner, 2000) and interacts mainly with its immediate neighbours, as plants are non-motile organisms. most applications of spatial point processes inecology have been either of a merely descriptive nature using second ordersummary statistics such as Ripley's K-function (Dale et al., 2002; Liebholdand Gurevitch, 2002; Wiegand and Moloney, 2004) or the models were restrictedto a very small number of species, typically not more than two orthree (Mateu et al.,1998). A traditional spatial point process analysis involvingthe inspection of first and second order summary statistics for each of thespecies as well as pairwise cross species summary statistics (Diggle,2003; Miller and Waagepetersen, 2003) becomes a very tedious task with highnumbers of species.In plant modelling,researchers have observed diffuculties of modelling spatial interactions in a plant community which requires very complexmodels with a large number of parameters.Bayesian approach is found to be more useful than the frequentist approach as it alloweda more fexible and realistic model.

2.3 Modelling across Terrestrial,Freshwater,and Marine Environments.

Across terrestial,freshwater and marine environments, SDMs have also played a role.Terrestrialplant analyses were prevalent in early years and are still common, along with studies ofterrestrial animals (including invertebrates); marine and freshwater applications were relativelyrare and soil-based organisms are still only infrequently modeled.SDMs from these diverse fields display commonalities and contrasts, with differences in mobilitybetween species prompting some major differences in modeling approach. When a speciesis sessile it is relatively easy to characterize its environment, even including the wider influence of landscape. Species which are considered mobile on the other hand use the resources which are patchily distributed across a landscape(Elith et.al,2009).Mobile species with small home ranges are often fitted using methods similar to those for sessileorganisms, perhaps with focal predictors summarizing information from the near-neighboring landscape (Ferrier et al.

2002). In contrast, models for highly mobile species (e.g., diadromous fish) need to include movement or access-related descriptors (e.g., stream-based distance to coast;Leathwick et al. 2008). In aquatic studies, observations are oftentreated as probabilities of capture and analyzed using similar methods as for sessile species, sometimesincluding temporal predictors to accommodate seasonal variation in catchability/presence(Venables&Dichmont 2004). Alternatively, specialized modeling techniques have been developedto account for imperfect detection (e.g., MacKenzie et al. 2002, Royle et al. 2004).Historic differences in the way data are collected also create different emphases across disciplines.Plant quadrats are usually regarded as statistically independent samples provided theyare sufficiently geographically separated(Alieth et al,2009).Thus,the understanding of species' geographic ranges (the areas where these species occur) is an important and classical ecological challenge(Brown *et al.* 1996; Gaston 2003).Apart from studying relationships between organisms and their environment that involve characterizing and explaining statistical patterns of abundance, distribution, and diversity"(Brown 1995, p. 10; see also Brown &Maurer 1989; Blackburn & Gaston 2003),this also helps estimate where an invasive species or disease will spread, or to predict the fate of endangered species(Jeschke et.al,2008).

2.4 Spatial Scale .

Another aspect that is importatant in distribuitions of both species and environment is the scale in terms of grain extent.The domain reflects the purpose of the analysis whrereas grain describe the properties of the data or analysis.For example, macroecological and global change studies tend to be continental to global in scope (e.g. Ara ´ ujo&New 2007), whereas studies targeting detailed ecological understanding or conservation planningtend toward local to regional extents (Fleishman et al. 2001, Ferrier et al. 2002).This shows the extent nature of species distribution models.On the other hand, grain should be consistent with the information content of the data,though in practice this is not always feasible.There is no single natural scale at which ecological patterns should be studied(Levin 1992). Rather, the appropriate scale is dictated by the study goals, the system, and available data(Jeschke et.al,2008) and several methods descibes scale of pattern in ecological data.The modern SDMs when linked with Geographic Information system(GIS) can represent convergence of site-based ecology and spatial data.

Climatic conditions determine the geographic ranges of species and this forms the basis of any boclimatic model.Researchers have have grouped these models into two.The first consideration is the use of species tolerance to heat,cold,or frost to predict the species range range (Doley 1977; Patterson *et al.* 1979; Prentice *et al.* 1992; Sykes *et al.* 1996;Kearney & Porter 2004; Hijmans &Graham2006).These are called the mechanistic models(Jeschke et.al,2008).The second approach use the knowledge of unknown physiological tolerances where it is not assumed that species geographic range is determined by tolerance but rather a number of climatic variables e.g minimum/maximum temperature,precipitation e.t.c.In this case these variables are measured for various locations and statisticall compared to the occurrence of the focal species at those locations.This gives the climatic range limits of this species distribution and helps in prediction hence their name empirical models(Jeschke et.al,2008).

It is assumed that the functional properties of species are constant over space and time.That is the phenotype and genotype do not change during the entire period during model building.During modelling, the new environment where for example range has shifted does not cause genetic or phenotypic changes.This collaborates (Peterson ,2003) that Ecological niches are often conservative.Species also occur at all locations with favourable climates i.e their dispersal is unlimited.This means there is species ability and sufficient time to populate all locations with favourable climates.However , many species lack the means to reach suitable but distant locations,and species such as trees need long time periods to extend their range even to relatively closelocations (Pearson 2006).Further,in determining the geographic ranges,biotic interactions are unimportant and they are constant over space ant time(Jeschke et.al,2008).

2.5 Conceptual and Technical Underpinnings of Species Modelling.

In species distributions,spatial prediction is an important aspect as it relates directly to the environmental niche.Factors like dispersion, reproduction competition and population dynamics also affect this kind of prediction (MacArthur et al. 1966, Gaston 2003).The assumptions that species distribuitiona are determined by the evironmental factors and that species have reached or nearly reached equilibrium with these factors are the cornerstone of the generalised linear models(GLM) and generalised additive models(GAM) that are used in the species distribuition modeling.Depending on the the evironmental changes,these assumptuons may be adequate or

inadequate(Peterson 2003).Underestimation of the degree of uncertainity in model predictions may arise if the assumptions are violated or due to inadequate data.Spatial mismatch(Agarwal et al. 2002) betwen different data sources,observer error,variable sampling intensity and gaps in sampling also arise as part of aspect of ecological issues.This reseach and modeliing here on aims at solving these issue.

Resource selection and species distribution models can be used to analyzeintensity of resource use and predict the geographical distribution of a particularspecies by combining information from point occurrence data and environmentalvariables (Boyce et al. 2002, Graham and Hijmans 2006). A variety of statisticalapproaches can be used. Many studies have utilized generalized linear models(GLMs) for modeling wildlife distributions (Walker 1990, Osborne and Tigar 1992,Buckland and Elston 1993). A GLM provides a method to estimate a function of themean response of a dependent variable as a linear combination of a set of predictors.GLMs are particularly useful for species distribution modeling because they provide a solid statistical foundation for realistically modeling ecological relationships (Austin 2002). Using this method, a study region is divided into grid cells and data on abundance or presence/absence is tabulated for each cell.The spatial distribution of a species within the grid can then be determined by relating the response variable (abundance or presence/absence of each grid cell) to spatially referenced covariates (Augustin et al. 1996, Boyce et al. 2002).For example , a poisson regression utilizes abundance or count data as the dependant variable.It belongs to a family of the GLM and was has been used in studies of elephant to examine the response of forest elephant telemetry hit abundance to a set of covariate predictors.The model is used as a means to explore relationships between telemetry fix abundance and a set of covariates.

As noted earlier,effective management of introduced species requires the early identification of those species that pose a significant threat of becoming invasive(Herron et.al,2007) as they have the ability to alter ecoystem .This is not an exceptional in Kenya where the increase in population and consequent shortable of arable land has resulted to migration into semi-arid ares with low agriculture potential(Otengi et al.,1995).This has resulted to clearance of the natural forest to provide the much needed land for cultivation(Lott et al.,200) hence meet the basic community needs.This does not mean that species of ecological use should dissapear as their habitats diminish and the community

should strive to preserve them.However with little data on them,their conservation strategy and poor predictive measures of their existence,this cant be achieved.It is against this backgound that I wish to model the species distribuition in the Kenya to solve such historic problems.

2.6 Geographic and Environmental Space.

SDMs have that element of the ability to establish distinction between geographic and environmental space.Geogragic space is characteistically defined by two-dimensional map cordinates or three-dimensional digital elevation models and on the other hand environmental space is potentially multi-dimensional,defined by some set of environmental predictors as seen in the figure1 below adapted from Elith et.al,2009.If an SDM is fitted using solely environmental predictors,it models variation in occurrence or abundance of a species in environmental space thus any predictions is also based on the species locations rather than geographic space(Elith et.al.2009).Geographic proximity is hereby ignored and hence the mapped predictions show clustering which reflects the spatial autocorrelation of the environment as in figure 1.
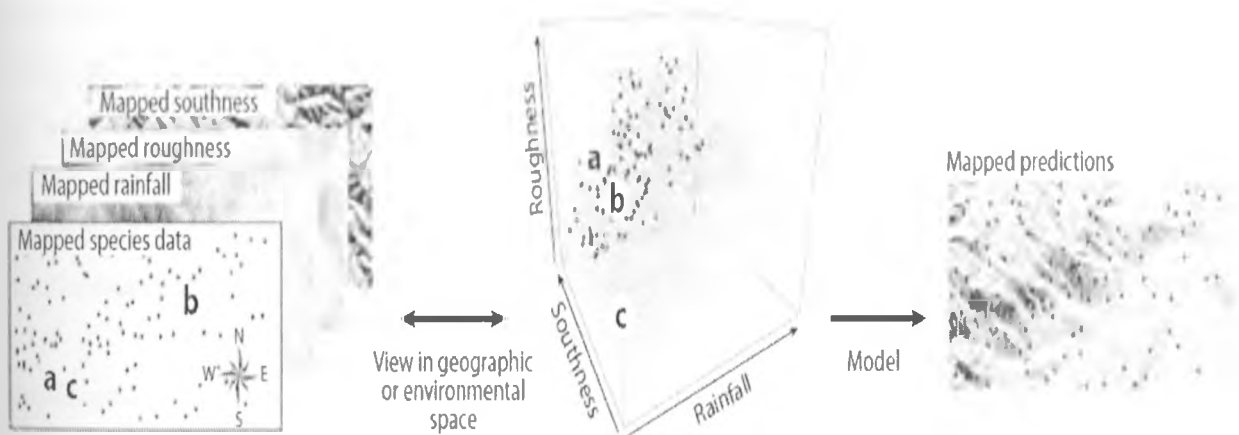
Figure 1 *:(Elith et.al,2009) The left figure shows relationship between mapped species and environment,centre is the environmetal space and in the right figure are the mapped predictions based on these environmental predictors.Points a and c are geographically close but not environmentally whereas mapped predictions shows spatial autocorrelation of environmental predictors.*

Geographic clumping of species can also result from their response to spatially autocorrelated environmental factors and/or the effects of factors operating primarily in geographic space

Strong residual geographic patterning(Elith et.al,2009) generally indicates that either key environmental predictors are missing (Leathwick & Whitehead 2001), the model is mis-specified (e.g., only linear terms where nonlinear are required), or geographic factors are influential (Dormann et al. 2007,Miller et al. 2007). The latter include glaciation, fire, contagious disease, connectivity, movement,dispersal, or biotic interactions. For these, the model might require additional relevant predictors,geographic variables and/or realistic estimates of dispersal distances or movement (Ferrier et al. 2002). Alternatively, some modelers enhance SDMs with process-based information to jointly characterize the environmental and spatial influences on distribution (e.g., Rouget & Richardson 2003, Schurr et al. 2007)

2.7 The *Rhipicephalus appendiculatus* (Brown Ear Tick)

If uncontrolled and managed properly, some species become a threat to the biodiversity and can nail down any developing economy through the sector that it touches.

It is upon this background and the literature here in that this research singles down *Rhipicephalus appendiculatus*(Brown ear tick) and seeks to model its spatial existence and prediction as it poses as a threat to the Kenyan economy.As noted earlier,the Kenyan economy relies heavily on its agriculutural capabilities and tourism sector.This tick has the behaviour tendancy of attacking these sectors and if it goes unchchecked it can contribute heavily to economic fall. *R. appendiculatus* is a member of the family Ixodidae (hard ticks). Hard ticks have a dorsal shield (scutum) and their mouthparts (capitulum) protrude forward when they are seen from above.

Rhipicephalus spp. ticks are brown ticks with short palps. The basis capitulum is usually hexagonal and generally inornate. Eyes and festoons are both present and Coxa I is deeply cleft. The spiracular plates are comma-shaped. The males of this genus have adanal shields and usually have accessory shields. The Male and Female *R. appendiculatus* are brownish, reddish-brown or very dark. This research will not look on the physical appearance of this species but goes ahead to model and providing statistical background for its control in the next chapters.

As noted, this literature concentrates on general cases of species be it animals or plants in general and its application to specific cases is minimal. In addition, GLM and its signifance influence in species modeling has not been completely utilized as well as mapping of their presence/absence using GIS.In

Kenya particularly, very few researches have been done in species modeling and where is done it is the general exploration of the species done, species influence to landscape, factors affecting the presence/absence with no keen focus on species modeling and prediction.

Further to this, the literature in the Kenyan context focuses on strategies on prevention and control and fails to identify how such species are distributed in various parts in the country and how they can inhabit other areas. For example a paper on strategy for tsetsefly and trypanosomiasis eradication in Kenya 2011-2021 explores the economic signifance of tsetsefly, control method and situational analysis.In this no statistical background and modelling to predict future occurrences has been done.The case of tick is also an isolated one as none of such nature has been done in Kenya.Tick causes anaemia, depriving animals nutrients thus lowering animals survival and their tolerance.When uncontrolled and the environmental to which ticks survive is unchecked,agriculture and tourism sector can be affected and particularly in Kenya where agriculture is backnone of the economy while Tourism sector gives the country foreign exchange.

This research therefore aims at improving on the gaps of species modeling by singling out a specific case of brown ear tick and intends to move away from general literature of species modeling to this specific case.

2.8 Concepts and Terms Definitions

i.   **Species Distribution model;**
     A model that relates species distribution data (occurrence or abundance at known locations) with information on the environmental and/or spatial characteristics of those locations.

ii. **Spatial Statistics;**
     Spatial Statistics concerns the quantitative analysis of spatial data, including their dependencies and uncertainties. Spatial-temporal statistics extend this to the spatial-temporal domain.Three major groups of data exist: lattice data that are collected on a predefined lattice, geostatistical data that represent continuous spatial variation and spatial point data that are observed at random locations. These types of data have their logical extension into the space-time domain, where the relations remain similar, but estimation may be different.

18

iii. **Spatial dependency or auto-correlation;**

Spatial dependency is the co-variation of properties within geographic space: characteristics at proximal locations appear to be correlated, either positively or negatively. Spatial dependency leads to the spatial autocorrelation problem in statistics since, like temporal autocorrelation; this violates standard statistical techniques that assume independence among observations.

iv. **Spatial sampling;**

This involves determining a limited number of locations in geographic space for faithfully measuring phenomena that are subject to dependency and heterogeneity. Dependency suggests that since one location can predict the value of another location, we do not need observations in both places. But heterogeneity suggests that this relation can change across space, and therefore we cannot trust an observed degree of dependency beyond a region that may be small. Basic spatial sampling schemes include random, clustered and systematic. These basic schemes can be applied at multiple levels in a designated spatial hierarchy (e.g., urban area, city, and neighborhood).

v. **Spatial interpolation;**

This is the method for estimating the variables at unobserved locations in geographic space based on the values at observed locations.

vi. **Spatial regression;**

Spatial regression methods capture spatial dependency in regression analysis, avoiding statistical problems such as unstable parameters and unreliable significance tests, as well as providing information on spatial relationships among the variables involved.

vii. **Spatial interaction;**

Spatial interaction or gravity models are the estimate of flow of people, material or information between locations in geographic space.

CHAPTER THREE:

RESEARCH METHODs

3.0 Introduction.

A number of plausible models in statistical applications of species has to be considered.Wide availability of advances in technology have also allowed for the collection of vast quantities of data with geo-referenced sample locations.While choosing model,greater attention is given to one that incorporates information that influences the response variable despite the fact that not everything associated with the response is known.Thus an error process should be included to account for unknowns(Hoeting et al.,2004).

Field–based ecological studies of species-habitat association is nowadyas based on regression methods that provide coherent treatments for the error distributions of presence-absence data.In particular,Generalised linear Models has enabled regression-based SDMs with their key features of non-normal error distributions,additive terms and nonlinear fitted functions(Manly et al. 2002).In Marine and terrestrial environments,remote sensing of surface conditions climatic parameters interpolation allows robust and preparation of digital models thus enhancing SDMs capabilities.

The development of Geographic Information system (GIS) assist also in storage and manipulation of both species records and environmental data. A Geographic Information System (GIS) is a useful tool for analyzing resource selection of wildlife species and predicting and mapping habitat suitability. In addition, a GIS can be used to construct an expert-based habitat suitability map for the species under study using a spatial model approach. The map can integrate spatial information on biological habitat requirements of the species with information on severity of human threats.

3.1 Problems Associated with Spatial Prediction

Spatial ecological prediction problems normally arise from those features associated with the species distribution and the data describing those distributions(Latimer et al,2006).
First,the spatial unit of prediction is usually larger than the sites sampled on the ground from the fact that the spatial area covered is large than the area sampled.This brings the issue of sampling errors since lage parts in consideration may be unsampled while others may be heavily sampled hence

heterogenity of the sampling intensity.Spatial misalignment problems are likely to occur when the environmental data is related to the species distribution data and the region where prediction is based.Whereas some data are in form of a regular and/ormirregular grid, some are available at point locations(point data) thus bringing inconsistencies,biasness(Agarwal et al. 2002) resulting to unevenly distributed sample locations with respect to the relevant characteristics of the region sampled(Mugglin et al. 2000,Gelfand et al.2002)

Secondly, when the degree of correlation among observations depends on their relative locations,the problem of spatial dependence or spatial autocorrelation arises.In ecology,observations which are close to one another tend to acquire similarities than those that are further apart resulting to positive autocorrelation.Based on the fact that process like dispersion and reproduction generate spatial autocorrelation in species occurrences,predictive models often exhibit some degree of autocorrelation.In addition, some residual autocorrelation in environmental factors do remain even when many environmental factors are included in the model.It thus calls upon any model to consider this dependency to avoid inaccurate parameter estimates and inadequate quantification of uncertainity(Ver Hoef et al. 2001).Generalised linear regression analysis and Spatial prediction solves this problem by including data that reflects neighborhhod values in model predictions though this does not quantify the strength of spatial pattern in the residualls.

Finally, the spatial domain of prediction is large relative to that in whih data is collected thus envirnmental data are not available on a scale as fine as that experienced by individual organisms.This resuts to the problem in quantifying uncertainity.Since predictions helps in extrapolating to areas not observed in the study,assessing uncertainity helps to set conservation policy and evaluating the impact of climate change on species(Thomas et al.2004).The model adapted in this reseach will assist in addressing these problems.

3.2 Data.

In this section,study area,reliability/validity of data,statistical methods used,data analaysis and mapping habita suitability will be discussed

3.2.1 Study Area.

This study adopted survey research design where GIS based secondary data was obtained directly from ILRI via *http://www.ilri.org/GIS* .The data available was based on Kenya in whole and where brown ear ticks have recently been identified. Due to the bulkiness of data, the research chose four major parts of Kenya namely: Nairobi, Rift valley, Eastern and Central .The areas were chosen as it was identified as the areas with high number of *Rhipicephalus appendiculatus* (Brown Ear Tick).In addition, these areas are considered as neighboring each other as it is the case in Kenya map locations identifications. This would help in showing any relationships and correlations in those areas infested by this species.

3.2.2 Reliability and Validity of Data.

Reliability and validity of data is an essential component in ensuring that the tools used for data collection is precise in capturing the intended information and is able to gather consistent data. The data was stored in GIS showing spatial locations thus very reliable and valid. Further ILRI is a credited institution and the economy depends on its data for research and development not only in Kenya but also internationally

3.2.3 Statistical Models

A statistical model has its main roots in the ability to provide a mathematical basis for interpretation, examining the parameters ,determining the strength of association and to ascertain the contributions and roles of the different variables.Ecologists use explanatory models to provide insights into the ecological processes that produce patterns (Austin et al. 1990) and predictive models to provide the user with a statistical relationship between the response and a series or predictor variables in predicting the probability of species occurrence or estimating numbers of an organismat new, previously unsampled locations(Guisan, et al,2002.)

3.2.4 Simple Generalized Linear Model.

Recall the general linear regression model of the form :

$$Y = \alpha + \chi^{T}\beta + \varepsilon \tag{3.1}$$

where Y denotes the response variable; α is a constant called the intercept and reflects the value of Y when $\beta = 0$; $\chi = (\chi_1, \chi_2 \ldots \chi_k)$ is a vector of k predictor variables;

$\beta = \{\beta_1, \beta_2, \ldots \beta_k\}$ is k-by-1 vector of unknown parameters (one for each predictor), and (zero-mean stochastic disturbances) is the error that represents measurement error, as well as any variation unexplained by the linear model. This model tries to minimise the unexplained variations through the least square methods. The model assumes :

i.   **Stochastic component:** the Y are usually assumed to have independent normal distributions with $\in (Y) = \mu$ with constant variance

ii. **Systematic component:** the covariates χ combine linearly with the coefficients to form the linear predictor $g = \chi\beta$

iii. **Link between the random and systematic components:** the linear predictor $\chi\beta = $ is a function of the mean parameter μ via a linkfunction,$g(\mu)$. Note that for the normallinear model, g is an identity.

Violation of assumption 1 constitutes a limitation to the application of most parametric statistical models,and is directly related to data sampling(Guisan, et al,2002).Many data in ecology are not Gaussian and do not have a constant variance and their variance is proportional to their mean (Davison 2001).Through advent of maximum likelihood GLM has been developed.

Generalized Linear Models(GLM) relates presence/absence data to environmental explanatory variables directly.They aremathematical extensions of linear models that do not force data into unnatural scales,and thereby allow for non-linearity and non-constant variance structures in the data (Hastie &Tibshirani 1990). They are based on an assumed relationshipbetween the mean of the response variable and the linear combination of the explanatoryvariables. Data may be assumed to be from several families of probability distributions,including the normal,binomial,Poisson,negative binomial,or gamma distribution, many ofwhich better fit the non-normal error structures of most ecological data.Thus,GLMs are moreflexible and better suited for analysing ecological relationships(Guisan, et al,2002.)

Here,we deal with the misalignment between the species presence /absence data which are referenced to point locations and the the environmental data which are referenced to grid cells. Guisan, et al(2002) recommends that working at the scale of the respose i.e the sample sites,and assigning to each sample site the values of the environmental factors for the grid cell in which the site falls is the best option as it results to a binomial variable reflecting the number of trials on the grid cell.

By use of GLM, the predictor variables $x_j (j = 1\,2\ldots p)$ are combined to produce a linear predictor (LP)which is related to the expected value $\mu = \in (Y)$of the response variable Ythrough a link function$g(.)$, such as

$$g\big(\in (Y)\big) = LP = \chi^T \beta + \varepsilon \tag{3.2}$$

where α, χ, β are those previously described in equation 3.1

The corresponding terms for the i^{th}observation in the sample is:

$$g(\mu) = \alpha + \beta_1 \chi_1 + \beta_2 X_2 + \cdots \beta_p X_p \tag{3.3}$$

We note that unlike the multiple linear regressin model(equation i), which assumes a normal distribution and anidentity link, the distribution of$Y$in a GLMs may be any of the exponential family distributions and the link function may be any monotonic differentiable function (like logarithm or logit).The variance of Y depends on $\mu = \in (Y)$ through the variance function$V(\mu)$,giving $Var(Y) = \emptyset V(\mu)$where$\emptyset$ is a scale (also known as a dispersion) parameter.

The main improvements of GLMs over LS regression are:
i.   The ability to handle a larger class of distributions for the response variable$Y$. Apart from the Gaussian, other distributions are the binomial, Poisson and Gamma. GLMs can also accommodate more generalqualitative (Davis & Goetz 1990) and semi-quantitative (ordinal; Guisan & Harrell 2000) response variables, usually based on a series of logistic binary GLMs.

ii.   The relationship of the response variable Y to the linear predictor($\mathcal{LP}$) through the link function $g(\in (Y))$. In addition to ensuring linearity, this is an efficient way of constraining the predictions to be within a range of possible values for the response variable (e.g., between 0 and 1 for probabilities of presence).

iii. Incorporates potential solutions (e.g quasilikelihood) to deal with overdispersion(Davison 2001).

Our data set consists of the average number of the observed species and a geo-reference for each site.It is not feasible to collect data at all possible locations, thus we are assumingthat n sites are representative of the entire area of interest.If $Y(s)$ is the presence/absence (1/0) of the particular species at sample location s in cell i,then by summing up $Y(s)$over the number of sample sites in cell(n)we get grid-cell level count Y +

Assuming the independence(as noted on GLM) for the trials and since all the sites are assigned the same levels for the environmental factors,we get a binomial distribution for Y +

$$Y_i + \sim \text{Binomial}(n_i, p_i) \tag{3.4}$$

meaning that the probability p that a species occurs in a cell is related to the environmental variable.

The distribution of these kind of observations is drawn from exponential family. The density of these distributions has the form

$$f(x) = exp\left[\frac{\theta_{..} + b(\theta)}{a(\emptyset)} + c(y, \emptyset)\right] \tag{3.5}$$

where θ is a parameter that determines the location of the distribution, $\emptyset$ is a parameter that scales the variability, and $a(.)$, $b(.)$and $c(.,.)$ are functions.The quantity$a(\emptyset)$(usually just $\emptyset$) scales the variablity up or down without changing this relationship, thus allowing the model to accommodate overdispersed or (less often) underdispersed observations.

Although any function can be used for the link, the model is simplest when the location parameter ofthe link function is the same as the location parameter $\emptyset$ of the distribution. The link function thenhas a canonical form that is determined by the distribution itself(Thomas D. Wickens,2004).With Binomial data,the canonical link function is logistic.

The generalized linear model does not have closed-form estimation equations. However, there are manyiterative algorithms by which it can be fitted. For example, maximum likelihood fits can be obtained byrepeatedly applying the weighted least squares after generating the variances from themodel. (Thomas D. Wickens,2004).

The fit of the generalized linear model is measured by a quantity known as the deviance, which is justtwice the negative log likelihood:

$$deviance = -2\log \ell \tag{3.6}$$

(where ℓ is the likelihood)

When the model allows for overdispersion or underdispersion by estimating the parameter$\emptyset$, the deviance cannot be interpreted as measuring goodness of fit—there is no independent estimate of the error as there isfor the general linear model(Thomas D. Wickens,2004)

Relation of p_i to the linear predictor $X^T\beta$ can be conviniently done using a logistic (Logit) function getting:

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1\chi_1 + \cdots + \beta_p\chi_p \tag{3.7}$$

where $\beta_0 + \beta_1\chi_1 + \cdots + \beta_p\chi_p = \chi^T\beta$ is as earlier defined.

If the cell is unsampled $(n = 0)$ there will be no contribution to the likelihood but if sampled$(n \geq 1)$ there will be a contribution to the likelihood component.

3.2.5 The Generalized Additive Model (GAM).

From the above background on GLMs,we note that likelihood-based regression models such as the normal linear regression models and the linear logistic model assume a linear (or some other parametric) form for the covariates $\chi_1, \chi_2 \ldots \chi_k$.The class of generalized additive models replaces the linear form $\sum \beta_j\chi_j$ by a sum of smooth functions $\sum s_j(\chi_j)$.The $s_j(.)'s$ are unspecified functions that are estimated using a scatter plot smoother, in an iterative procedure. The linear predictor $\sum \beta_j\chi_j$ is replaced by the additive predictor$\sum s_j(\chi_j)$. Hence the name generalized additive model.

This is a statistical model for blending properties of generalized linear models with additive models. The model specifies a distribution (such as normal distribution, or a binomial distribution) and a link function g relating the expected value of the distribution to the m predictor variables, and attempts to fit functions $f_i(\chi_i)$ to satisfy:

$$g(\in (Y)) = \beta_0 + f_1(\chi_1) + \cdots + f_m(\chi_m) \qquad (3.8)$$

The equation 3.8 above can generally be written as shown in equation 3.9 below.

$$\epsilon(Y|X) = s_0 + \sum_{j=1}^{p} s_j \chi_j \qquad (3.9)$$

$$\text{where} \in s_j(\chi_j) = 0 \text{ for every } j$$

The functions $f_i(\chi_i)$ may be fit using parametric or non-parametric means, thus providing the potential for better fits to data than other methods. The method hence is very general – a typical GAM might use a scatter plot smoothing function such as a locally weighted mean for $f_1(\chi_1)$ and then use a factor model for $f_2(\chi_2)$. By allowing nonparametric fits, well designed GAMs allow good fits to the training data with relaxed assumptions on the actual relationship, perhaps at the expense of interpretability of results.

Overfitting can be a problem with GAMs. The number of smoothing parameters can be specified, and this number should be reasonably small, certainly well under the degrees of freedom offered by the data. Cross-validation can be used to detect and/or reduce over fitting problems with GAMs (or other statistical methods). Other models such as GLMs may be preferable to GAMs unless GAMs improve predictive ability substantially for the application in question.

3.2.6 Diagnostics Involving GLMs and GAMs.

When using GLMs and GAMs, Akaike Information Criterion(AIC) can be used during the variable selection(Sakamoto et al. 1988) and Inference tests for predictors selection explain a signifanct portion of the variance. This is done especially when spatial correlation is typically

27

ignored in the selection of explanatory variables and this can influence model selection results. For example, the inclusion or exclusion of particular explanatory variables may not be apparent when spatial correlation is ignored.

In species modelling,shapes of the responses curves is given attention as well as incorporation of ecological features like dispersion and competition.Researchers including Jari Oksanen and Peter Minchin used data on vascular plant distribution along an elevation gradient to test four models for fotting such responses.On testing hierarchical(Huisman et al. (HOF)(1997); binomial GLMs(logistic link);binomial GAMs (logistic link) and beta-functions (Austin et al. 1994),they concluded that HOF are more effective and that GAMs too provided similar results.

3.2.7 Getting Spatially Explicit Model: Built up from GLM

In the problems associated with spatial prediction,it was noted that degree of correlation among observations depends on their relative locations.This is due to similarities in ecological attributes among the neighbouring cells and the dispersions associated with the populations.The products of this is the autocorrelation and the models seeks also to respond to this spatial dependence while retaining the environmental response($\chi^T \beta$) which is catered for by the logistic (Logit) function.

Considering equation iv and modelling at the grid –cell level,a spatial term θ_i associated with grid cell i is added to equation vii to get the equation below.

$$\frac{p}{1-p} = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p} + \theta_i$$

(3.10)

In this case,θ_i is a random effect associated with each grid cell and it adjusts the probability of presence of the modeled species up or down depending on the values of θ in cell i's spatial neighborhood(Latimer et al,2006).

Let $X_1, X_2, \ldots X_n$ be a finite collection of random variables which are associated with sites labelled 1,2 ... n,respectively.For each site, $P(x_i | x_1 \ldots, x_{i-1}, x_{i+1}, \ldots x_n)$,the conditional probality of X_i,given all other site values ,is specified and thus we require also the joint distibutions of the

variables(Besag 1974).Thus if $x_1, ..., x_n$,can individually occur at the sites $1,2 ... n$,respectively,then they can occur together.Statistically then if $P(x_i) > 0$ for each i,then $P(x_1, x_2, ..., x_n) > 0$.

Considering a set of sites i, j,site $j(\neq i)$ is said to be a neighbour of site i *iff* the functional form of $P(x_i | x_1 ..., x_{i-1}, x_{i+1}, ... x_n)$ is dependent upon the variable x_j.

3.3 Data Analysis

The data will be coded and then a data base will be developed in statistical software. Due to the type of analysis required, the researcher opt R-Software and Microsoft software. A form of matrix will be developed where environmental attribute were each given a variable name. The data base was first established in excel sheet and saved as CSV comma delimited. Then R Command were used to export the data from excel to R-Software where modeling and analysis was done.

The data obtained was first arranged in logical order followed by drawing tables and graphs. . Descriptive analysis was used and also an aspect of correlation to show if there is any association between various environmental factors. The data was fitted to give a logistic model with various aspects relating *Rhipicephalus appendiculatus* (Brown Ear Tick) to the environmental factors.

3.3.1 Analysis Involving Logistic Regression.

Logistic regression analysis extends the techniques of multiple regression analysis to research situations in which the outcome variable is binary. Let Y be binary outcome; then Y is coded as Y=1 if event of interest occurs and Y=0 if it does not occur. Specifically in the case of *Rhipicephalus appendiculatus*

- 1= presence of Rhipicephalus *appendiculatus*
- 0= absence of *Rhipicephalus appendiculatus*

Let Y=1 indicate that *Rhipicephalus appendiculatus* is present in the region, then statistical theory tells us that the mean of Y is a probability in this case that measures the probability of occurrence of this tick. This model is based on probabilities associated with the binary outcome as opposed to the actual outcome thus it describes probabilities as functions of explanatory variables. The explanatory variables are either continuous and/or categorical variables. The function of mean it uses is the logit function or the logarithm of the odds.

The estimated coefficients from a logistic regression fit are interpreted in terms of odds and odds ratios. The following requirements have to be specified:

- An outcome variable with two possible categorical outcomes (1=presence; 0=absence).
- A way to estimate the probability p of the occurrence of the outcome variable.
- A way of linking the outcome variable to the explanatory variables.
- A way of estimating the coefficients of the regression equation, as well as their confidence intervals.
- A way to test the goodness of fit of the regression model

3.3.2 Odds and Odds-Ratio.

By definition, if p is the probability of event occurring, then the probability of event not occurring is $1 - p$, and odds are defined as:

$$Odds = \frac{probability\ of\ event\ occurring}{probibility\ of\ event\ not\ occurring} = \frac{p}{1-p}$$ (3.11)

The odds ratio compares the odds of events of two groups. It is the ratio of two sets of odds.

$$Odds\ ratio = \frac{Odds\ of\ group\ A}{Odds\ of\ group\ B} = \frac{\frac{P_A}{1-P_A}}{\frac{P_B}{1-P_B}}$$ (3.12)

We note therefore that in logistic regression we model the natural log of the odds of event. Equation 3.7 represents such a model where $\beta_0, \beta_1 \ldots \beta_p$ are to be estimated using sample information.

3.3.3 Interpretation of Regression Coefficient.

Consider $\beta_i > 0$ where i is greater than 0,

- If $\beta_i > 0$ then there is an increase in the log odds of the event for every unit increase in explanatory variable.
- If $\beta_i < 0$ then there is a decrease in the log odds of the event for every unit increase in the explanatory variable.
- If $\beta_i = 0$, then there is no relationship between the log odds and the explanatory variable.

- If the explanatory is a continuous variable then e^{β_i} is the change in risk for every additional measure of the explanatory variable.

- If the predictor is a categorical variable then e^{β_i} is the odds ratio of one group to other; one group is taken to be the reference.

3.4 Mapping Habitat Suitability.

Spatially explicit information on the brown ear tick and factors that influence will be combined to create a map.This approach aim at providing a GIS-based method to model and predict habitat suitability for wide-ranging species. The components of the model include deriving a "Biological Landscape" layer, which takes into account biological covariates thought to influence a particular species' habitat and a "Human Landscape" layer, which takes into account human covariates which might exert significant pressure on, or pose certain threats to, the species' habitat. Both Biological and Human Landscape layer model inputs are based on expert opinion of known information about the species' biological preferences and relationships to human factors. The model approach combines both landscape layers to create a "Conservation Landscape" layer, or habitat suitability surface, for the species.

3.5 Tick Density Prediction Model.

Predictive mapping applied was based on reflecting the relationship between occurrence and covariates ad hoc. This involved application of appropriate mathematical models which permits a statistical estimation of the relationship between the occurrences of ticks.

Tick occurrence is primarily determined by macro-climatic factors hence macroclimate is commonly exploited in prediction models. In carrying out the modeling, the most frequently applied 'expert' systems; CLIMEX model (Sutherst & Maywald, 1985) procedure was used in developing the model. Such prediction model is based on the assumption that an area of occurrence is limited by a 'climatic envelope. ' The envelope is positively delimited by a 'Growth Index' which combines favorable temperature and humidity, and negatively delimited by 'Stress Indices' represented by extremes in temperature and humidity.

3.6 Kriging.

The simplest point referenced models assume that the process that generated the observed data has a spatial component, and represent this spatial component through functions of distance between

31

observation points. The best-known example of this approach is kriging, which was developed to predict locations of ore deposits from point-level drill samples(Banerjee et al. 2004). Ecological point-referenced modelling has been limited to relatively modest-sized data sets until recently (Latimer et al. 2006)

Kriging is perhaps the most familiar use of point level modeling: kriging is a way of predicting the values of a response variable (like species abundance or soil characteristics) at new, unmeasured locations (Bannerjee et al. 2003). The goal here, is not simply to predict but also to investigate the relationships between the brown ear tick distributions, environmental variables, and spatial pattern, so the modeling also includes environmental covariates as well as spatial random effects. In addition to estimating regression coefficients,a geostatistical regression model involves fitting a spatial correlation function to the regression errors. The function allows correlation between observations to decrease as separation in space increases.This is the universal kriging modeling concept.

CHAPTER FOUR:

RESULTS

This chapter discusses the research data obtained. The relationships of *Rhipicephalus appendiculatus* to various environmental factors are discussed as well as showing the maps of spatial locations, distribution and predictions by use of GIS. The various figures drawn are meant to show how the species relates with other factors independently and also dependently.

4.1 Results from the Model.

The statistical methods in use had confirmed that that adding complexity to basic generalized linear models improved the models' characterizationof the distributions of species. The research now turns to a more thorough evaluationof the model output, including the estimates for theenvironmental coefficients, the spatial random effectvariables, and the uncertainty associated with modelparameters.

For ease of analayis,the reseach started by giving the various factors variable names as follows in relation to the model $g(\mu) = \alpha + \beta_1 \chi_1 + \beta_2 \chi_2 + \cdots \beta_p \chi_p$ described in equation 3.3 where $g(\mu)$ is the *Rhipicephalus appendiculatus* density *(TD)* ; χ_1 is the cattle density in the area(CD); χ_2 is the east coast fever distribution(ECF); χ_3 is the average ranifall expericed in the given area(RF) and χ_4 is the temperature in the area(TE)

4.1.1 Summary of the data.

This section gives the summary of the data considered .Continuous variables include cattle density(CD),rainfall(RF) and temperature(TE) while the outcome,tick occurrence and east coast fever occurence are binary(0/1) response.

4.1.2 Continuos variables summary.

This shows the continuos variables namely cattle density,rainfall and temperature that affects the brown ear tick distribution.The cattle density was measured in square kilometres,rainfall in milimetres and temeperature in degree celcius.

Table1:Continuous variables used in the study.

	Cattle Density(sq km)	Rainfall(mm)	Temperature(celsius)
Minimum	0.00	400	9.00
1st Quartile	42.18	875	12.00
Median	71.12	1000	17.50
Mean	66.46	1060	18.27

4.1.3 Binary variables summary.

TD represent ticks and ECF reprents East Coast Fever.These variables are in terms of presence/absence(0/1) and 38 is the total data set involved.

Table 2:Binary variables in the study.

	East Coast Fever	
Tick Density	0	1
0	1	1
1	0	38

4.1.4 The Logit Model

This section estimates a logistic regression model using the generalised linear model (glm) function as discussed in chapter 3.East coast fever was treated as a factor (i.e categorical)variable.Having given the model name,R does not produce any output from the regression but instead the reseach used summary commands.

Model summary

The summary of the residuals and coefficients in the model is given below

Table 3:Deviance Residuals Summary

Deviance Residuals				
Minimum	1st Quartile	Median	3rd Quartile	Maximum
-1.675	-0.8552	-0.6287	1.1380	2.056

Table 4: Coefficients Summary

Coefficients	Estimate	Std.Error	Z value	Pr(>|z|)
Intercept	-1.127	1.895	-3.2	0.000435
Cattle Density	1.684	0.002126	2.1	0.034876
Rainfall	0.1	0.332973	2.627	0.01624
Temperature	-0.57243	0.31254	2.133	0.022346
East Coast Fever	1 -1.2146	0.215201	-2.976	0.000345

The dispersion parameter for binomial family taken to be 1 was also derived where null deviance was 49.998 on 39 degrees of freedom,Residual deviance: 38.41 on 35 degrees of freedom and Alkaike Information criteria (AIC) is 10

In the output above,the deviance residuals are a measure of the model fit.This part shows the distribution of the deviance for individual cases used in the model.The logistic regression gives the change in the log odds of the outcome for one unit increase in the predictor variable.

We note the following:

- For every unit change in in cattle density,the log odds of tick density(versus none) increases by 1.684.
- For every unit change in in rainfall,the log odds of tick density(versus none) increases by 0.1
- For every unit change in in temperature,the log odds of tick density(versus none) decreases by 0.57
- Having East coast fever in the region decreases the log odds of tick density by 1.2146.This can be interpreted to mean there is no effect of tick density which can be associated with East Coast fever presence or absence.

The model can thus be written as:

$$log\left(\frac{p}{1-p}\right) = -1.127 + 1.684\chi_1 + 0.1\chi_2 - 0.57245\chi_3 - 1.2146\chi_4$$

35

By definition, if p is the probability of event occurring, then the probability of event not occurring is $1 - p$. Thus, for every unit increase in cattle density and rainfall, the log of odds of having brown ear tick increases by 1.684 and 0.1 respectively.

Table 5: Obtained confidence intervals for the coefficient estimates.

	2.5 %	97.5 %
(Intercept)	-5. 17351	-1.50643
CD	-0.001346	-1.50643
RF	0.150346	1.4100695
TE	-1..63019	0.31317
(ECF)1	-1.27433	NA

4.1.5 Exponeniating the coefficients

By exponentiating the coefficients, we interplate them as odds-ratio.

Table 6: Exponeniated Coefficients

Variable	(Intercept)	Cattle Density	Rainfall	Temperature	East Coast Fever
Odds-Ratio	1.189028	1.855969	1.363994	7.005278	6.489016

Considering that tick density, temperature, rainfall and temperature are continuous variables then, there is the change in risk for every additional measure of each of these variables. On the other hand, for one unit increse in east coast fever, the odds of having brown ear tick in the area increases by a factor of 6.48. The research therefore confirms that all these factors have a positive influence in the occurrence of brown ear tick.

4.1.6 Predicted Probabilities

This was be done for both categorical and continuous predictor variables. To do this, a new data frame was created with the values whose independent variables we want to take on to create the predictors.

The research started by calculating the predicted probability of *Rhipicephalus appendiculatus* (Brown Ear Tick) at each value of east coast fever, holding cattle density, rainfall and temperature constant. An object was created in R software with the observations of presence and absence (1/0).The second, third and fourth objects of cattle density, temperature and rainfall were created that only took on one value each. In this case, the variables were set to their means, but this can also be set at any value. It is notable that these objects must have the same names as the variables in the logistic regression discussed. In the last command, the objects created were combined into a new data frame.

4.1.7 New Data Results.

This shows the data set established from the previous for prediction purpose.

Table 7: Established new data

East Coast Fever(ECF)	Cattle Density(CD)	Rainfall(RF)	Temperature(TE)
0	66.46022	1060	18.275
1	66.46022	1060	18.275

This data frame in table 7 above was then used to calculate the predicted probabilities.

Table 8: Predicted Probabilities of Occurrence.

East Coast Fever(ECF)	Cattle Density(CD)	Rainfall(RF)	Temperature(TE)	ECFp
0	66.46022	1060	18.275	0.89
1	66.46022	1060	18.275	1

The output confirms that the predicted probability of having the *Rhipicephalus appendiculatus* (Brown Ear Tick) is 0.89 when east coast fever is not observed and 1 when is observed. This confirms that the brown ear tick occurrence will also influence the presence of east coast fever in the studied areas now and in future.

4.1.8 Model Fit.

This is useful when comparing competing models. The output produced by the model summary included indices of fit (shown below the coefficients), including the null and deviance residuals and

the AIC.One measure of model fit is the significance of the overall model. This test asks whether the model with predictors fits significantly better than a model with just an intercept (i.e. null model).

The test statistic is the difference between the residual deviance for the model with predictors and the null model. The test statistic is distributed chi-squared with degrees of freedom equal to the differences in degree of freedom between the current and the null model(i.e the number of predictors variables in the model).To find the difference in deviance for the two models(i.e the test statistic), R commands are used giving 15.88122 as the result. The degrees of freedom for the difference between the two models is equal to the number of predictor variables in the model and was obtained as 2.Finally, the p-value was obtained as 6.456192.

The chi-square of 15.88122 with 4 degrees of freedom and an associated p-value of less than 0.001 tells us that the model as a whole fits significantly better than an empty model. This is sometimes called a likelihood ratio test (the deviance residual is -2*log likelihood) obtained as -2.021658 (df=5).

4.2 The Relationship between the Variables.

This sections checks the correlation and covariance exhibited by the variables in the study

4.2.1 Correlation

The correlation matrix was obtained through R as follows

$$
\begin{pmatrix}
 & TD & CD & ECF & RF & TE \\
TD & 1.0000000 & 0.3748285 & 0.6979824 & 0.4016462 & -0.3206608 \\
CD & 0.3748285 & 1.0000000 & 0.3066419 & 0.7686704 & -0.5443849 \\
ECF & 0.6979824 & 0.3066419 & 1.0000000 & 0.2573632 & -0.2570967 \\
RF & 0.4016462 & 0.7686704 & 0.2573632 & 1.0000000 & -0.4945105 \\
TE & -0.3206608 & -0.5443849 & -0.2570967 & -0.4945105 & 1.0000000
\end{pmatrix}
$$

Notable is that all immediate factors that the research put into focus were all positively correlated with the tick density i.e Cattle density in the area, Rainfall and east coast fever occurrence. Temperature showed negative association perhaps due to the selective nature of this tick's habitant determined by the other variables.

4.3 G.I.S Mapping

An expert-oriented map of brown ear tick presence in Nairobi,central.,Eastern and Riftvalley was created by combining spatially explicit information on the tick habitat and environmental factors which favour this tick.

4.3.1 Data Sources, Description and Processing

This section gives focus to the GIS layers involved in the research. This incorporates the aspect of relationships exhibited by the factors under study.

4.3.2 Temperature Layer (Figure 2)

Figure 2 shows the temperature layer coverage derived from the Exploratory Soil Survey Report number E1, Kenya Soil survey, Nairobi 1982 and shows the principle Agro-Climatic Zones of Kenya based on a combination of both moisture availability zones (I-IV) and temperature zones (1-9). Temperature covering the study areas was generated and merged to produce continuous temperature distribution within the study area and this was compiled based on temperature zones categorized on a 1-9 scale.
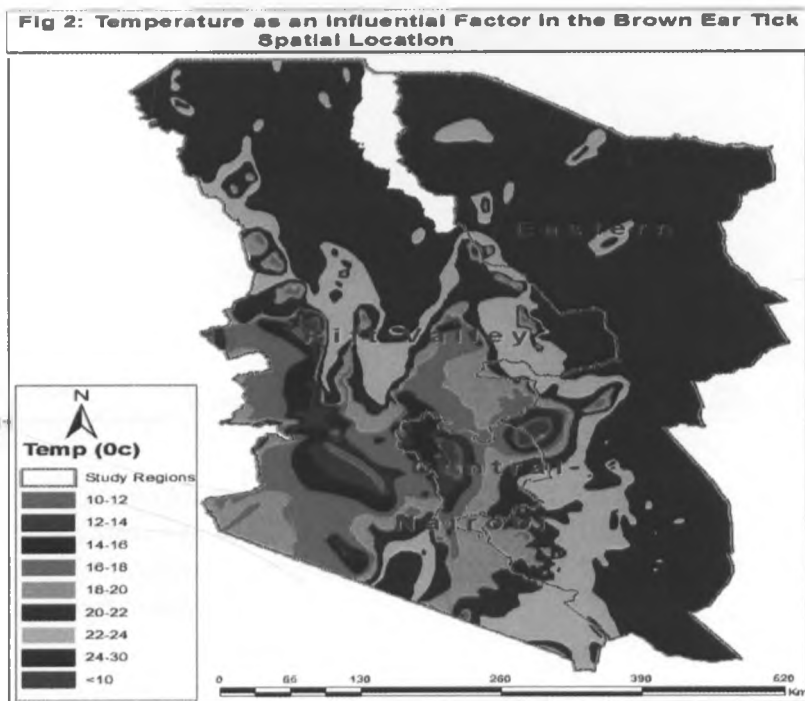

Fig.2.The Region of study(Nairobi,Central,Riftvalley and Eastern) with temperature layer overlaid. The different colouration shows the region's levels of temperature intensity and the cells used for modeling and sample locations scattered across these cells, including both sample locations at which Rhipicephalus appendiculatus was likely to occur.

4.3.3 Rainfall layer (Figure 3)

Rainfall layer obtained from ILRI GIS database shows the annual rainfall distribution in millimeters per year for Kenya. The layer was done by the Japanese International Co-operation Agency (JICA), National Water Master Plan, Kenya and compiled and stored by ILRI GIS department. The rainfall categorized on a nominal scale, (2-10) representing areas with similar rainfall amounts in millimeter per year compiled. Rainfall distribution for the specific study districts was extracted, merged with the other districts and study area rainfall distribution generated as shown in figure 3.
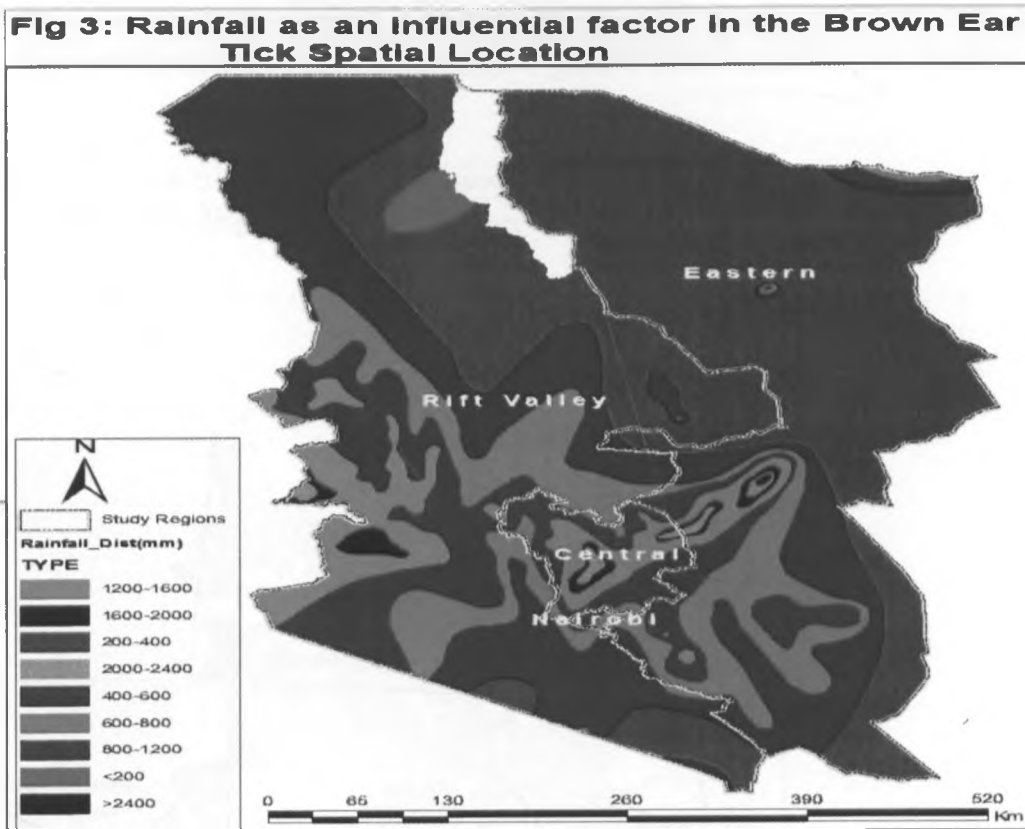

Fig.3.The Region of study(Nairobi,Central,Riftvalley and Eastern) with rainfall layer overlaid. The different colouration shows the region's levels of rainfall distribution and the cells used for modeling and sample locations scattered across these cells, including both sample locations at which Rhipicephalus appendiculatus occurs.

4.3.4 ECF Distribution layer.

The coverage showing the distribution of East Coast Fever (ECF) in Kenya was obtained from a study on distribution of *T. parva*, based on epidemiological surveys by (FAO, 1975) studies and

40

were compiled and mapped by (Lessard *et al*. 1990). The map also included 'expert opinion' based on an extensive survey of local experts. The layer was extracted for the study area and used for analysis.

4.3.5 Cattle density distribution (Figure 4)

Cattle distribution map was compiled based on 1995-1997 division level zebu and dairy animal density reports from the Ministry of Agriculture, Livestock Development and Marketing, (MALDM). The reports were based on data collected by District Livestock Officers. Cattle distribution density was extracted for the study districts and used together with the above variables to run the model which produced tick density distribution map. Figure 4 below shows cattle density in the studied region.
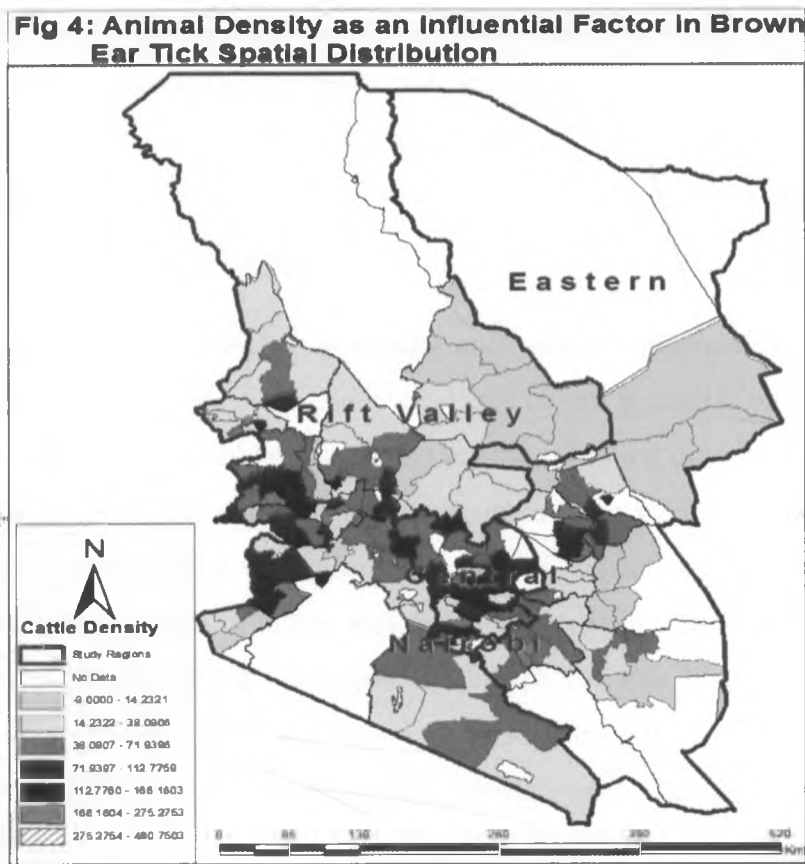

Fig. 4. The Region of study(Nairobi, Central, Riftvalley and Eastern) with cattle density sample points overlaid. The different colouration shows the region's levels of cattle density concentration and the cells used for modeling and sample locations scattered across these cells, including both sample locations at which Rhipicephalus appendiculatus occurs.

4.3.6 Tick Density Prediction.

Using ArcGIS 10, the four variables, rainfall, temperature, east coast fever distribution and cattle densities within the study area were modeled together using the geo-processing tool in Arc-GIS software. The tick distribution modeling tool was developed and spatial joint used to spatially join the four variable polygon layers and using mathematical operations available in ArcGIS, the model was subsequently run taking consideration the four variables.

Parameter containing nominal values was set for the output layer whereby the results from the model showing areas with very high degree of tick occurrence noted 1 and the areas with the very low chance of tick occurrence noted 5. The software automatically generated an output layer with an attributes field containing the nominal values assigned. This was symbolized and an additional field interested in the shape file attributes table describing the state of each field i.e 1 representing very high chances of occurrence, 2-high; 3-average; 4-low and 5-very low. This was extrapolated to produce tick density prediction map for the entire study area.

4.3.7 The Kriging Process.

Kriging technique which is a popular and widely implemented method of surface mapping in GIS packages was applied in the model so as to spatially link and determine the relationships between the four variables.The output results generated using kriging technique was combination of all the input variables to produce single output. However, this technique lacks a unique solution and needs a large amount of interactive input and subjective assessment during the processing. By this process the model was able to establish other areas of occurrence of the brown ear tick in the area of study. The relationships between the brown ear tick distributions, environmental variables, and spatial pattern are incorporated via kriging showing a high relationship of these factors. In addition to estimating regression coefficients,a geostatistical regression model involves fitting a spatial correlation function to the regression errors. This has indeed shown the tick is set to be established in the neighbourhood of the study region.Thus,kriging addresses spatial dependence while retaining the

environmental response .Figure 5 and 6 shows this further where figure 5 shows intial areas of the tick spatial location and figure 6 the predicted areas of occurrence.
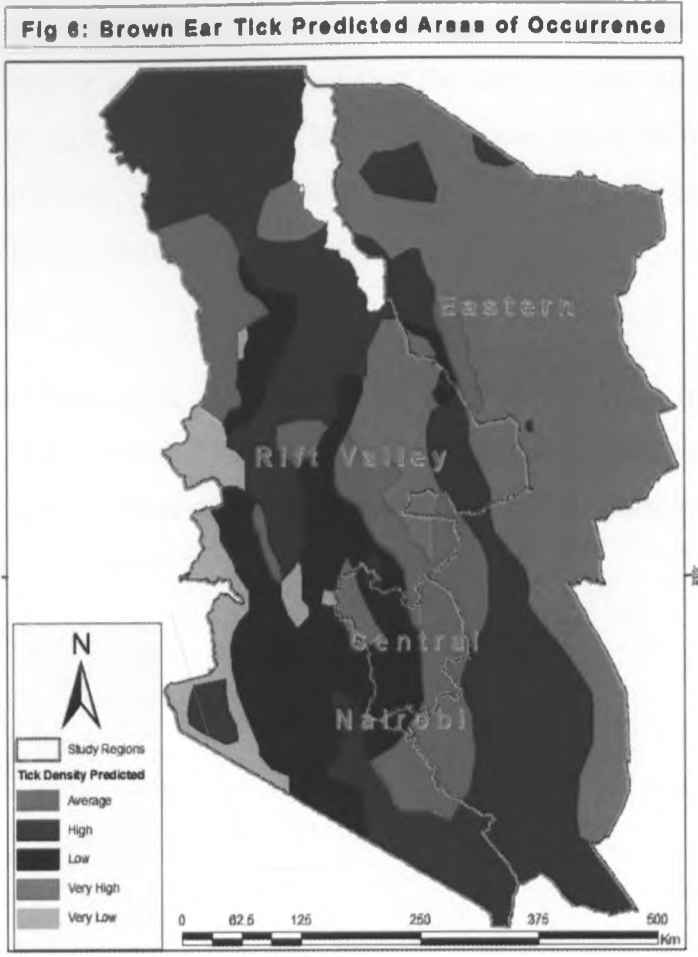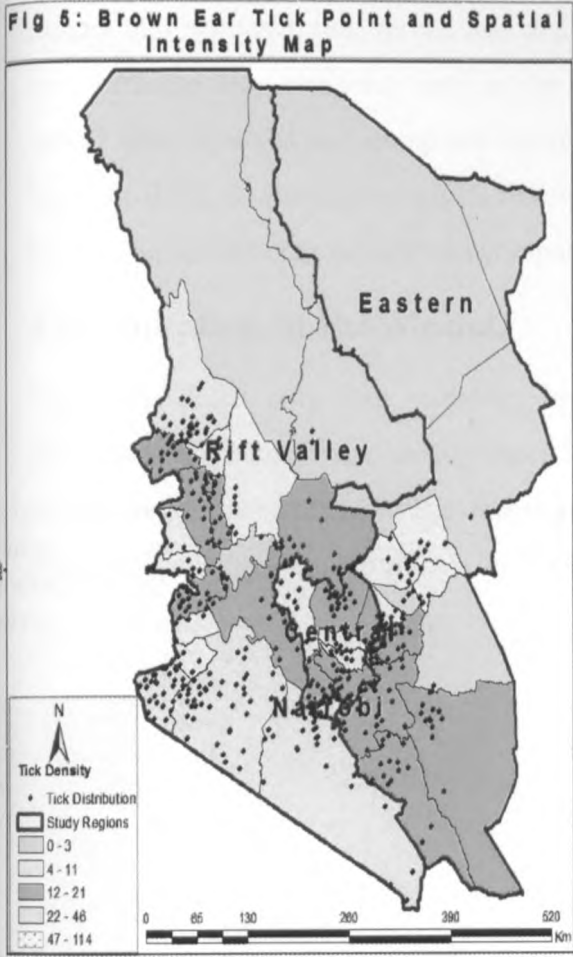

Fig.5.The Region of study(Nairobi,Central,Riftvalley and Eastern)with Rhipicephalus appendiculatus sample points overlaid. The different colouration shows the region's levels of the tick spatial locations and concentrations there on and the cells used for modeling and sample locations scattered across these cells.

Fig.6. Predicted distribution for Rhipicephalus appendiculatus.For purposes of visualization, pullout colouration is used to present portions of the predicted region. Sample locations at which the species were observed are overlaid with the other layers namely rainfall,temperature,cattle density and East Coast Fever .

43

4.4 Assumption.

The assumptions in the model was that temperature, rainfall, cattle density and rainfall are the main factors that influence tick spread and distribution within a region. This relationship was assumed as this were the only variables used in the model. It was also assumed that the datasets used in the model was collected and compiled accurately, and this was taken into consideration since datasets from an ILRI, an institution which collects and compile datasets on livestock was used for it has a higher chance of being accurate as compared to data from other sources.

4.5 Limitation of the Model.

The model used only four variables in generating tick distribution density, and this may be a limitation since there are many other variables which have an influence on tick distribution. Additionally, the accuracy of the data used could not be precisely determined, but was assumed to be accurate based on the institution which processed and compiled the datasets.

CHAPTER 5:

CONCLUSIONS AND RECOMMENDATIONS

While embarking on this researh project,one point was clear;the research was to develop a species distribution model to be used for describing the *Rhipicephalus appendiculatus* patterns as well as making predictions.This was towards one goal of biodiversity and management .

5.1 Conclusions.

The research has been able to meet its main objective and thus can conclude the various aspects of the brown ear tick as discussed here.

The basic nonspatial logistic regression model can be improved by adding features to the model that reflect major known attributes of the data, including variable sampling intensity and spatial autocorrelation.There is a strong case for making these models spatially explicit. Indeed,Gaston (2003) and others have called for the incorporation of spatially explicit methods in determining the structure and dynamics of species geographic range.Further, one of the problems in comparative biogeography is that sampling and data gathering are conducted with a multitude of different methodologies (Gaston 2003, Graham et al. 2004). The consequence is uncertainty in comparing and interpreting spatial patterns in species distributions.

This model results confirm that the spatial pattern of presence and absence of the brown ear tick includes more information than can be explained through just the mean effect of a suite of environmental variables ie the temperature ,rainfall,cattle density and further addition of the disease presence like the East coast fever which is known to be caused by this tick.There are two main explanations for this. One is that biological processes tend to generate spatial pattern. In the case of the brown ear tick, the probability that a site contains a species depends not only on its climatic and edaphic characteristics, but also on its neighborhood.Such spatial dependence can arise from biological processes at a number of levels. Processes in the life history of individual organisms, including reproduction,territoriality, and dispersal, can generate clustering or evenness in species distributions. Interactions of species with each other and with resources (e.g. effects of cattle desisty on the brown ear tick presence,the role of temperature and rainfall to determine inhabitant) can

likewise cause and perpetuate spatial association. The particular occupancy history of a site can also exert a long-term spatial influence on its neighborhood.These spatial patterns are not mere epiphenomena,but rather can strongly influence individuals species distributions, as well as interspecific interactions and thus community composition and potentially ecosystem processes.

The second explanation for autocorrelationis the influence of unobserved environmental variables, and of nonlinearities in interactions among sets of (observed and unobserved) factors, all of which may have some degree of spatial dependence and interdependence (Ver Hoef et al. 2001). Moreover,it is inevitable that the identity of critical explanatory variables may change from one part of brow ear tick geographical range to another. Since models cannot include all important variables, and may include some unimportant ones, there will usually be some degree of autocorrelation in model residuals.Critically, without spatial structure in the model, the level of uncertainty about model parameters can be dramatically underestimated and poorly characterized.One effect of this is that a model will identify more explanatory variables as significantly related to species presence/absence (VerHoefetal.2001).In this study the kriging process has eliminated uncertainity in the model by showing the spatial dependence of factors under study(Rainfall, temperature,cattle density and East coast fever)

Additional studies which seek to understand resource selection of the brown ear tick at fine scales are crucial to developing a foundation for determining appropriate management,prevention and conservation strategies for this species. In every subregion studied, the brown ear tick preferred proximity to one or many ecological features present in these regions including cattle density,temperature scales and rainfall .We note here that these factors are some of those that influence the occurrence of the brown ear tick and have all been noted to positively influence this existence except the East Coast fever.It was observed that even if ticks were absent the disease which is one of those this ticks causes still occurs in minority areas.This can be attributed to other ticks prey like antelopes.Indeed when predictions are done, the probability of the tick occurrence was 0.89 and 1 when east coast fever is not observed and 1 when is observed. This is a strong indication of the damage the tick is set to cause.It is upon all stake holders concerned to minimise or control completely the brown ear tick survival.

As it has been statistically noted,all these factors are interelated and influnce this tick occurrence.By controlling one factor perhaps can render its existence to be minimal.For example,analysis noted that when catle density increases,the likelihood of the tick occurrence increases.This can be a well utilised factor especially when these cattle can only be configured in areas they perform well where the existence of the tick will be easy to controll considering that it is as if the tick will be located in one known location.By this other measures like the use of acaricide can be used to control this tick which is a disaster to the Kenyan economy.Though,rainfall and temperature are hard variables to control,it is in view of this reseach that they can be naturally managed.For example,it is known that the forest cover attracts rainfall and by so doing controls temperatures too.Further,the data on forest cover and distribution is readily availbale to the government through the relevant sectors and concerned stake holders like ILRI,KWS and Kenya National Museums.This information can be utilsed to offset some areas considered a high risk for the disesess by increasing/decreasing the cover as per the plan in place.It is in view of the research that by for example increasing the forest cover in some predicted areas,the rainfall would increase and the temperature reduced.Statistically,it was noted these variables relate and correlate to each other and this would decrease the tick absence due to reduction of one of the related factor.

In reference to the model,when $\beta \neq 0$.The implication of this is that the explanatory variables are significant in predicting the dependent variable.Therefore using the value 1 as a reference point, then for the model with cattle density as an explanatory variable; exp(1.684)=5.3870 can be interpreted as: there is a 438% increase in odds of the brown ear tick occurrence for every unit increase in cattle density(5.3870-1=4.3870) .This is a very huge percentage increase and shows how strong the influence of cattle density is to the tick occurrence. This is not an exception to the other variables though with different percentages. For the rainfall, exp(0.1)=1.1051 which means a 10 % increase in odds of the tick occurrence and for temperature,exp(-0.57245)=0.5641 meaning 45% decrease. We note that these changes are attributed to the changes in the environmental condition as a strong influence of one factor affects the rest either positively or negatively .This kind of statistical approach can help to manage, control and preserve the brown ear tick when need be.

From the prediction by GIS and that by modelling,it has been noted that the tick is set to increase with time.This is a big threat to the Kenyan economy.Figures 2,3,4 have shown the different layers

that influences the occurrence of *Rhipicephalus appendiculatus*(Brown Ear Tick) while figure 6 shows the predicted areas. In Figure 2, the temperature layer has been shown, figure 3 the rainfall layer, figure 4 we have the cattle density, figure 5 the areas of occurrence and figure 6 the predicted areas. Notable is the relation that the maps 2,3 and 4 show when the predicted areas are mapped. Comparing figure 2 and 6, this tick is observed in the areas with different rainfall pattern and this is the case with the rest of the environmental factors. When prediction is done we see the possibility of the tick being observed in other regions due to the favorable conditions. Figure 6 generally shows these predicted areas and thus conclude that if this tick is not controlled it is bound to inhibit other areas due to shifting environmental factors. As noted earlier, these factors can be controlled thus helping to control the invasion attributed to the brown ear tick.

The spatial predictive process model described here can greatly speed computation in ecological models for point data. This approach offers a statistical method for analysing point-referenced data sets to learn about environmental relationships in the presence of spatially correlated errors. In addition to making regressions robust to spatial autocorrelation, this approach can help us learn whether any two processes are significantly associated in space, and what the scale of their spatial autocorrelation is. This is broadly applicable, as we often want to answer questions such as whether the prevalence of a particular animal behaviour is spatially associated with environmental factors, or whether trophically or competitively interacting species show residual spatial association. The simplicity, power and many important potential applications make the spatial predictive process approach a useful addition to ecologists.

From a statistical point of view, our analysis shows how the difficulties of modelling spatial interactions and modelling species especially in animal community which requires complex models with a large number of parameters can be eliminated. In this situation, it is found that GLM allows a realistic model.In addition,use of GIS mapping and kriging reduces uncertainity in a species model. Finally, from an ecological perspective, we are able both to confirm existing knowledge on species' interactions and to generate new biological questions and hypotheses on species' interactions and association with the environment.

5.2 Recommendations.

The following are the research recommendations:

5.2.1 Resource Selection of the Brown Ear Tick.

GIS proved to be an extremely useful tool for elucidating relationships between brown ear tick abundance and several biological and human threat covariates. Certain considerations should be followed for future work.One fundamental problem with the results of the analysis overall was that many covariate responses in the data are non-linear.The tests of trying to use mixed models step was eliminated in the current analysis, however, because more time must be given to assessing the ecological significance of using mixed models and polynomials for analyzing particular covariates. An alternative route might be to explore the use of generalized additive models (GAMs), which have also been found to be useful and sometimes more flexible for ecological modeling.

Scale plays a significant role in the manner in which forest elephants select their habitats and resources. Senft et al. (1987) and Boyce et al. (2003) found that because resource distribution, foraging costs and threats vary with scale, animals may pursue different resources at large and small spatial scales. This finding is certainly applicable in the case of this analysis.The presence of this tick in a certain area might be due to more of one of the environmental conditions being abundant in the area rendering the other to have strong influence to its existence.These relationships are still being understood by ecologists and this issue must be addressed for future work.In some cases, the spatial covariate data used in the analysis was lacking and could have contributed particular sources of error to the results.

5.2.2 Modeling Brown Ear Tick in the Region.

Maps showing habitat suitability and species distribution can provide a strong foundation for applied research and conservation planning (Graham and Hijmans2006). These maps are only as effective as the data and methods used to create them,however. A major challenge in undertaking any type of modeling effort exists in overcoming problems and challenges relating to the quality and type of data used for the suitability model. Overcoming data shortages and limitations proved to be a particular challenge –for example, brown ear tick prefer to spend large amounts of time with the prey .Unfortunately, there is no comprehensive dataset of locations of all types of this tick's prey for the entire region so the model excluded this important influence on brown ear tick habitat.

As with all expert models, the model depends on multiple iterations of revisions and "fine-tuning" based on a) trial and error and b) expert opinion by biologists and wildlife ecology experts who are familiar with the biological habitat selection of the brown ear tick.It is recommended that both the Biological and Human Landscape layers should be considered working hypotheses which should be improved and revised as new data layers become available and as knowledge of covariate layer relationships relating to the species' habitat suitability is enhanced.

5.2.3 Statistical Control and Conservation of Brown Ear Tick.

The research has offered not only how brown ear tick relates to its evironment but also has statistically predicted its future occurrence.Such knowledge based statistics can be used to control spread of the tick thus saving the economy from tick's threat especially the agriculture and tourism sector.In addition,conservation where need be of this tick can use such knowledge to identify where the tick survises well due to different ecologocal factors.

5.2.4 Stakeholders.

This research reccomends to stakeholders to speed up measures for the control of brown ear tick as it poses as a major threat to the Kenyan economy.Such research can be used in any other species which pose as a threat to environment and landsacpe.It is indeed a bases on which other species occurrence/absence in Kenya can be based on.

Further, various organisation can use such basis in species control and management in Kenya.For example,ILRI can use such a research background to determmine best areas for the livestock productivity to enhance development of the agriculture sector.KWS and KNM can use it for conservation purposes,management and control mechanisms.

REFERENCES

1. Agarwal, D. K., A. E. Gelfand, and J. A. Silander, Jr. 2002.Investigating tropical deforestation using two-stage spatially misaligned regression models. Journal of Agricultural Biological and Environmental Statistics

2. Arthur DR. Ticks and disease. New York: Pergamon Press; 1961. Diagnosis of *Rhipicephalus appendiculatus*; p. 70-73.

3. Augustin, N. H., M. A. Mugglestone, and S. T. Buckland. 1996. An autologistic model for the spatial distribution of wildlife. Journal of Applied Ecology **33**:339–347.

4. Austin MP, Cunningham RB, Fleming PM. 1984. New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio* 55: 11-27

5. Austin MP. 1985. Continuum concept, ordination methods and niche theory. Annual Review of Ecology and Systematics 16: 39-61

6. Austin MP. 1999. The potential contribution of vegetation ecology to biodiversity research. Ecography 22: 465-84

7. Bannerjee, S., B. P. Carlin, and A. Gelfand. 2003. Hierarchical modeling and analysis for spatial data. Chapman and Hall, Boca Raton, Florida, USA.

8. Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems (with discussion). Journal of the Royal Statistical Society Series.

9. Carroll SS, Pearson DL. 2000. Detecting and modeling spatial and temporal dependence in conservation biology.*Conservation Biology* 14: 1893-97

10. Collingham YC, Wadsworth RA, Huntley B, Hulme PE. 2000. Predicting the spatial distribution of nonindigenous riparian weeds: issues of spatial scale and extent. *Journal of Applied Ecology* 37: 13-27

11. Diggle, P.J. & Lophaven, S. (2006). Bayesian geostatistical design. Scand. J. Stat., 33, 53–64.

12. Elith J, Burgman MA. 2002. Predictions and their validation: rare plants in the Central Highlands, Victoria,Australia. In *Predicting Species Occurrences: Issues of Accuracy and Scale*, ed. JM Scott, PJ Heglund, MLMorrison, MG Raphael, WA Wall, et al, pp. 303-14. Covelo, CA: Island Press

13. Elith J, Graham C. 2009. Do they? How do they? WHY do they differ? -- on finding reasons for differing performances of species distribution models. *Ecography* 32: 66-77.

14. Araújo MB, Whittaker RJ, Ladle RJ, Erhard M. 2005. Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology & Biogeography* 14: 529-38

15. Elith J, Burgman MA, Regan HM. 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling* 157: 313-29

16. Franklin J, Simons DK, Beardsley D, Rogan JM, Gordon H. 2001. Evaluating errors in a digital vegetation map with forest inventory data and accuracy assessment using fuzzy sets. *Transactions in GIS* 5: 285-304

17. Ferrier, S., G. Watson, J. Pearce, and M. Drielsma. 2002.Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Specieslevel modeling. Biodiversity and Conservation .

18. Fleishman E, MacNally R, Fay JP, Murphy DD. 2001. Modeling and predicting species occurrence using broadscale environmental variables: an example with butterflies of the Great Basin. *Conservation Biology* 15: 1674-85

19. Graf RF, Bollmann K, Suter W, Bugmann H. 2005. The importance of spatial scale in habitat models:capercaillie in the Swiss Alps. *Landscape Ecology* 20: 703-17

20. Graham CH, Elith J, Hijmans RJ, Guisan A, Peterson AT, et al. 2008. The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology* 45: 239-47

21. Guisan, A., T. C. Edwards, Jr., and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157:89-100

22. Guisan A, Graham CH, Elith J, Huettmann F. 2007. Sensitivity of predictive species distribution models tochange in grain size. *Diversity and Distributions* 13: 332-40

23. Harrell FE, Lee KL, Mark DB. 1996. Multivariate prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* .

24. Hartley S, Harris R, Lester PJ. 2006. Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant. *Ecology Letters* 9: 1068-79

25. Heikkinen R. 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography* 30: 751-77

26. Heikkinen RK, Luoto M, Kuussaari M, Toivonen T. 2007. Modelling the spatial distribution of a threatened butterfly: Impacts of scale and statistical technique. *Landscape and Urban Planning* 79: 347-57

27. Heikkinen RK, Luoto M, Virkkala R, Pearson RG, Korber JH. 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology and Biogeography* 16: 754-63

28. Huntley B, Berry PM, Cramer W, McDonald AP. 1995. Modelling present and potential future ranges of some European higher plants using climate response surfaces. *Journal of Biogeography* .

29. Hooten MB, Wikle CK, Dorazio RM, Royle JA. 2007. Hierarchical Spatiotemporal Matrix Models for Characterizing Invasions. *Biometrics* 63: 558–67

30. Johnson C. 2004. Mapping uncertainty: sensitivity of wildlife habitat ratings to expert opinion. *Journal of Applied Ecology* 41: 1032-41

31. Johnson C. 2008. Sensitivity of species-distribution models to error, bias, and model design: An application to resource selection functions for woodland caribou. *Ecological Modelling* 213: 143-55

32. Jones MM, Tuomisto H, Olivas PC. 2008. Differences in the degree of environmental control on large and small tropical plants: just a sampling effect? *Journal of Ecology* 96: 367-77

33. Langford W, Gergel S, Dietterich T, Cohen W. 2006. Map misclassification can cause large errors in landscape pattern indices: examples from habitat fragmentation. *Ecosystems* 9: 474-88

34. Latimer AM, Wu SS, Gelfand AE, Silander JA. 2006. Building statistical models to analyze species distributions. *Ecological Applications* 16: 33-50

35. Leathwick J, Whitehead D. 2001. Soil and atmospheric water deficits and the distribution of New Zealand's indigenous tree species. *Functional Ecology* 15: 233–42

36. Leathwick JR. 1998. Are New Zealand's Nothofagus species in equilibrium with their environment? *Journal of Vegetation Science* 9: 719-32

37. Meyer CB, Thuiller W. 2006. Accuracy of resource selection functions across spatial scales. *Diversity and Distributions*

38. Meyer CB. 2007. Does scale matter in predicting species distributions? Case study with the Marbled Murrelet.*Ecological Applications* 17: 1474-83

39. Ferrier S, Smith AP. 1990. Using Geographic Information Systems for biological survey design, analysis and extrapolation. *Australian Biologist* 3: 105-16

40. Walker JB, Keirans JE, Horak I. The genus *Rhipicephalus* (Acari, Ixodidae): a guide to the brown ticks of the world. Cambridge, NY: Cambridge University Press; 2000. *Rhipicephalus appendiculatus*

41. Zimmermann N, Edwards T, Moisen G, Frescino T, Blackard J. 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology* 44: 1057-67 *http://www.ilri.org/GIS.*Thurday May 24th July 2012 at 1045hrs.

APPENDIXES

Appendix 1 :GIS Maps

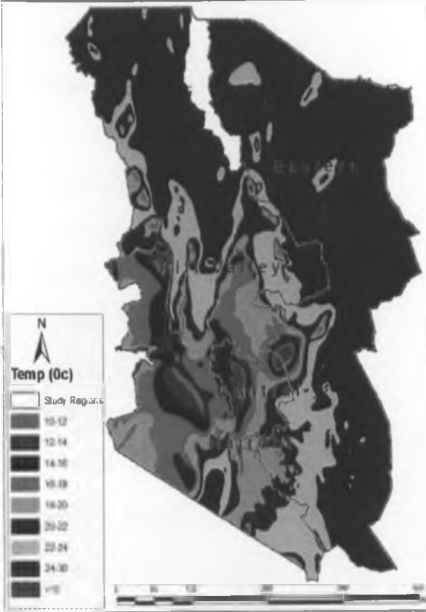Fig 2: Temperature as an Influential Factor in the Brown Ear Tick Spatial Location

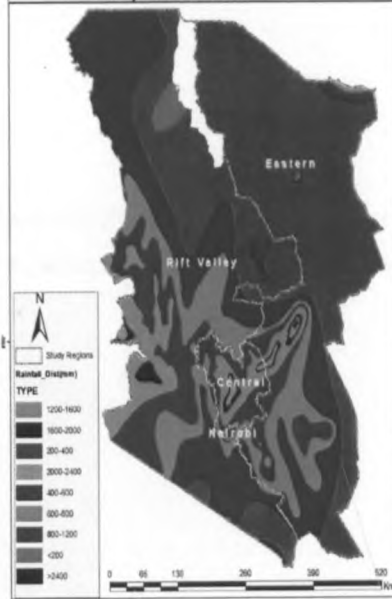Fig 3: Rainfall as an Influential factor in the Brown Ear Tick Spatial Location

Fig 4: Animal Density as an Influential Factor in Brown Ear Tick Spatial Distribution
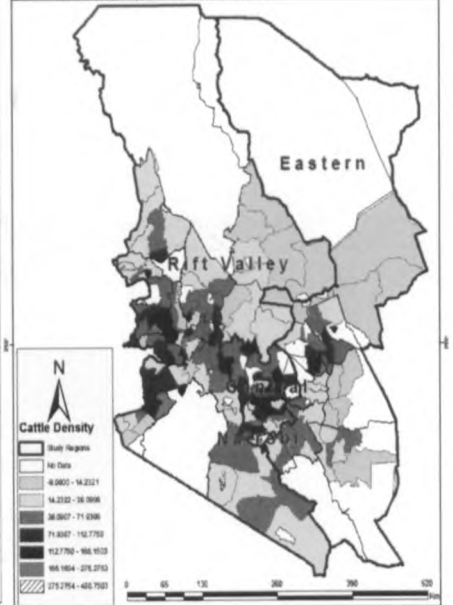

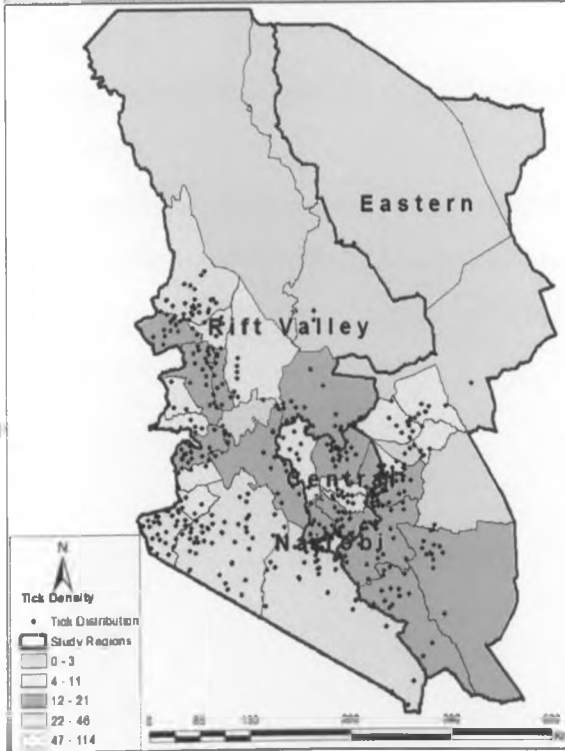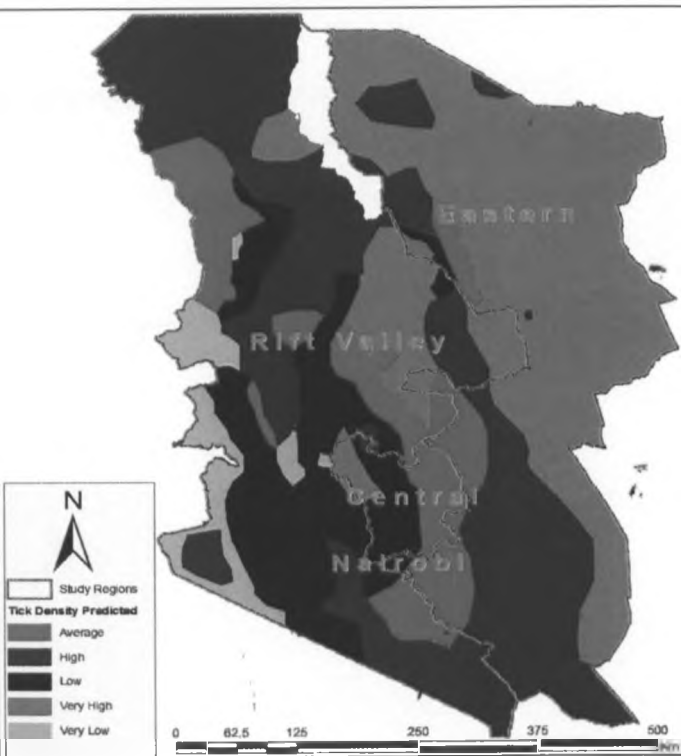Fig 5: Brown Ear Tick Point and Spatial Intensity Map

Fig 6: Brown Ear Tick Predicted Areas of Occurrence

55

Appendix 2: Programme codes

Appendices 2: R commands that were used used

In all the commands the name data stands for the name of the database that was developed in excel Microsoft software and savesd as CSV comma delimited.

```
data=read.csv(file.choose())
attach(data)
names(data)
summary(data)
table(TD,ECF)
mylogit<-glm(TD~CD+RF+TE+as.factor(ECF),family=binomial(link="logit"),na.action=na.pass)
glm(formula = TD ~ CD + RF + TE + as.factor(ECF), family = binomial(link = "logit"),
 na.action = na.pass)
mylogit
summary(mylogit)
confint(mylogit)
ECF<-c(0,1)
CD<-c(mean(data$CD))
RF<-c(mean(data$RF))
TE<-c(mean(data$TE))
newdata<-data.frame(ECF,CD,RF,TE)
newdata
newdata$ecf1<-predict(mylogit,newdata=newdata,type="response")
newdata
mylogit$null.deviance-mylogit$deviance
mylogit$df.null.deviance-mylogit$df.residual
1-pchisq(mylogit$null.deviance-mylogit$deviance,mylogit$df.null.deviance-mylogit$df.residual)
logLik(mylogit)
```