UNIVERSITY OF NAIROBI

SCHOOL OF MATHEMATICS

Species Distribution Model:

A Case Study of *Prunus Africana* in Kakamega Forest

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

Murungi Titus Murithi

©December 2012.

DECLARATION

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

Murungi Titus Murithi   Reg. No: I56/65124/2010.

Signature: _____   Date: 01 / 12 /2012

This project has been submitted for examination with my approval as a University supervisor.

John M. Ndiritu

Signature: _____   Date: 01 / 12 /2012

DEDICATION

I dedicate this report to my mother Mrs.Rael Murungi. 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. Rael Murungi you are indeed the title of a woman's supremacy dignity", May you live to see the sweet fruits.

Thanks to my fiancée Doris Kagendo for moral support during the academic journey.

ACKNOWLEDGEMENT

I acknowledge and salute all those who contributed to success of my research project and subsequent preparation of this report.

I would like to thank my supervisor Mr. Ndiritu for willfully guiding me all the way through this research. Fondly cherished is he for working round the clock to fix whatever problem I encountered in the course of the work. I bear testimony to the struggle he made in guiding me throughout this period. His affluent knowledge and experience with species modeling has been of great help. As far as I am concerned he is a gold standard when it comes to expertise and devotion. Indeed his unreserved support and invaluable guidance is an inspiration.

My appreciation also goes to my family for giving me company and a word of encouragement during the normal ups and downs during academic work and during the entire research period.

I am particularly grateful to all my fellow colleagues in the M.SC (Social Statistics) class. Our endurance and support to each was not in vain.

In a special regard, I extend my appreciation to the University of Nairobi and particularly the school of Mathematics for shaping up my statistical skills. I am now more enlightened in the field of Statistics.

Above all, I thank the Almighty God for being with me throughout my study period

TABLE OF CONTENTS

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 *Prunus Africana* (also known as *Pygeum africanum*, or *African Stinkwood*) in Kenya with spatially explicit environmental factors to understand the *Prunus Africana* resource selection and habitat preference using a logistic model as developed from the generalized linear model (GLM). Resource selection among the *Prunus Africana* was similar in that across all sub-regions they selected an area which has similar related factors in terms of the surrounding.

Prunus africana is a large evergreen tree from the Rosaceaea family. It is found in the montaine forests of Kenya. It is an upper canopy tree that can grow as high as thirty meters. Prunus bark contains three active constituents including *phytosterols*, which are anti-inflammatory agents. It takes fifteen years for the bark to develop the active ingredients necessary for medicinal use. Traditional medicinal uses of the bark include the treatment of stomach aches, urinary and bladder infections, chest pain, malaria, and kidney disease.

In addition, a GIS was used to construct an expert-based habitat suitability map for the *Prunus africana* across Kenya using a spatial model approach. The map integrates spatial information on biological habitat requirements of the species. This can serve as a useful tool for determining the future of the tree in our forests and the prevention prioritization in Kenya.

LIST OF ABBREVIATIONS AND ACRONYMS

S.D.M - Species Distribution model

G.L.M – Generalized linear models

G.A.M - Generalized Additive Model

G.I.S - Geographic Information System

KEFRI - Kenya Forestry Research Institute

ICIPE - International Centre of Insect Physiology and Ecology

KWS - Kenya Wildlife Service

FD - Forest Department

DRSRS - Department of Remote Sensing and Resource Survey

DEFINITION OF TERMS

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

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.

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.

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).

Spatial interpolation

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

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.

Spatial interaction

A spatial interaction or gravity model is the estimate of flow of people, material or information between locations in geographic space.

CHAPTER ONE

INTRODUCTION

1.0: Introduction

The distribution of different animal and plant species on earth and in space has been a major concern of many people and especially, the scientists and hence has a long history. As a result these people, especially the scientists have sought explanations on this issue of distribution for a very long time. The major challenge to their answers being the complexity of the environmental factors and the ecological processes, that is, dispersal, reproduction, competition, and the dynamics of large and small populations and also the challenge of insufficient data on the species distributions.

With the raise of new powerful statistical techniques and the Geographic Information Systems (GIS) tools there has been increased developments of predictive habitant distribution models in ecology. This has also led to increased analysis of the spatial data by many scientists.GIS are an innovative and powerful tool that helps analysts and decision makers organize, visualize, analyze, present and understand complex layers of data. The key to spatial analysis is that most data contain a geographic component that can be tied to a specific location, such as a county, zip code, census block or single address. The geographic analysis therefore enables users to explore and overlay data by location, revealing hidden trends that are not readily apparent in the traditional spreadsheets and statistical packages. A Species Distribution Model(s) (SDM) can be defined as statistical and / or analytical algorithm that predicts distribution of species, given field observations and auxiliary maps or more specifically, species distribution models can also be looked at as numerical tools that combine observations of species occurrence or abundance with environmental estimates. With the

SDMs we can gain ecological and evolutionary insights and be able to predict distributions across landscapes, sometimes requiring extrapolation in space and time. In general, these models seek to provide the users with the statistical relationship between the response and a series of environmental variables for use in predicting the species distributions and estimating the number of organisms in new previously unsampled geographic locations.

Currently and in previous years, many scientists have increasingly carried out predictive modeling of species distributions in order to address a number of issues that are involved in ecology, biogeography, biology and climate change. In the early, 19th century, climate was recognized has one of the key geographical elements that explained the distribution of plant and animals species across the world. There are quite a number of modeling techniques that are available to assess the relationship between the distribution of species and the sets of predictor variables (ecological gradients). There are three types of ecological gradients, these include; Resource, direct and indirect gradients (Austin et.al, 1984). It is important to note that the distribution patterns of species and ecosystems are not only a function of post-glacial migrations and human influences, but also present day response to ecological gradients. Ecological gradients are measures of the physical environment that explain the distribution of organisms and ecosystems in terms of environmental tolerances.

Prunus Africana in Kakamega forest

Kakamega rainforest is the only tropical rain forest in Kenya of the Guineo-congolian type. It once stretched across the Central Africa to East Africa. Kakamega forest is located in Western Kenya, Kakamega East District (Kakamega County).

Kakamega forest is one region where high population growth has resulted in deforestation. The Kakamega forest is an afromontaine forest in western Kenya. It is a small fragment of

the Guineo-Congolian rainforest that once spanned the equatorial region as well as parts of the western coast of Africa (Earlham University, 1999). Within the forest, land cover is very heterogeneous, resulting in many different forest types. As a result, the forest is an area of high biodiversity. Though the forest is small in area, it is species-rich, with over 350 bird species, and over 380 plant species identified (Earlham University, 1999; ICIPE, 2005). In fact, around 20% of all Kenyan plant and animal species are endemic to the forest, including up to 75% of all butterfly species in the country (KIFCON, 1994).

The Kakamega forest has a very complex land use history, and has been highly influenced by the growth and development of the surrounding communities. The dominant tribe surrounding the forest is the Luhya, one of Kenya's most populous tribes, extending across Kenya's western border into Uganda (KIFCON, 1994). Most of the communities survive by practicing subsistence agriculture, growing mostly maize with a mix of other food crops including bananas, tomatoes, sukuma wiki (collard greens) and sweet potatoes (KIFCON, 1994; ICIPE, 2005). The soil in and around the forest is a moderately fertile clay-loam mixture, and the river system within the forest makes fresh water easily attainable, making agricultural initiatives much more successful than in other regions of the country (Earlham University, 1999).

The promise of agricultural success has long resulted in dense settlement surrounding the Forest, making this area have one of the fastest growing populations in the country, with a 2.8% annual growth rate (Earlham University, 1999). With the steady increase in human population has come a corresponding increase in demands on the forest, Local populations rely on forest products for religious, medicinal, building, grazing, fuelwood, charcoal production, and water needs. Where once these extractive uses did not interfere with the

3

natural regeneration of the forest, they now result in widespread species endangerment, as well as loss of ecosystem function.

Some NGOs have intervened to try and merge conservation issues with income generation and environmental education for local communities. International Centre of Insect Physiology and Ecology (ICIPE) and the Kenya Forestry Research Institute (KEFRI) have collaborated to initiate agro-forestry projects that emphasize the cultivation of medicinal plants that have been traditionally gathered from the forest (ICIPE, 2005). These plants are then processed by ICIPE into powder, salves and other forms, and sold in markets around the forest as well as in neighboring towns and as far away as Nairobi.

Two medicinal plants have been integrated into such programs with great success: *Ocimum kilamscharium* and *Mondia whytei*. Many farmers around the Kakamega forest are raising *Ocimum*. ICIPE purchases the raw plant material from farmers and processes it into a medicinal salve, Naturub, which is then sold at local and regional markets. *Prunus africana* is another important medicinal species that inhabits the forest. Attempts to incorporate this tree into afro-forestry projects have failed thus far (Cunningham, 1993)

In the Kakamega Forest, Western Kenya, human encroachment on the forest for agricultural land and forest products has resulted in wide spread deforestation, with the result that only a small fragment of intact forest remains. One species suffering from this increased pressure in Kakamega forest is *Prunus africana*. One method of combating these detrimental effects is through agro forestry, which seeks to mirror forest structure by establishing multiple canopies of growth including timber species, medicinal and food crops in one integrated system. Figure 1.2 shows the map of Africa, with locations of Kenya and Kakamega County.

4

Figure 1.2: Map of Africa with Location of Kenya and Kakamega (Glenday, 2004)

Important is that, Kakamega Forest has other species of plants and animals which include; (Wagner et al. 2008) over 330 species of birds, 380 species of plants, 400 species of butterflies, 7 species of primates, other animals include chameleons, skinks and lizards. other mammals includes ,squirrels, Bush bucks, Aardvarks, porcupines, giant forest hogs and many others. The Rainforest is penetrated by a network of walking/hiking trails silent with only melody of singing birds, whispering trees, rasp of butterflies as they fly on by, the croaking frogs, chattering of monkeys as they move from one tree to another and gurgling streams nearby. Table 1.1 summarizes the taxa in the Kakamega forest (Wagner et al. 2008).

Table 1.1: Biodiversity by taxa in Kakamega Forest (Wagner et al. 2008)

GROUP /TAXA	ESTIMATED NUMBER OF SPECIES
Plants	380
Trees/shrubs/vines	150
Orchids	60
Lepidoptera (Butterflies/moths)	400
Odonata (Dragonflies)	72
Reptiles	58
Snakes	36
Lizards	21
Turtles	1
Birds	330
Mammals	50
Primates	7

Prunus Africana

Prunus africana (Hook f.), also known as *Pygeum africanum*, or *African Stinkwood*, is a large evergreen tree from the Rosaceaea family. It is found in the montaine forests of Kenya. It is an upper canopy tree that can grow as high as thirty meters (Hall et al, 2000). Prunus bark contains three active constituents including *phytosterols*, which are anti-inflammatory agents (Hall et al, 2000; Longo, 1981). It takes fifteen years for the bark to develop the active ingredients necessary for medicinal use (Longo, 1981). Traditional medicinal uses of the bark include the treatment of stomach aches, urinary and bladder infections, chest pain, malaria, and kidney disease. Figure 1.1(a) and (b) shows a photograph of the *Prunus Africana* tree;

Figure 1.1 (a); A branch of the *Prunus Africana* tree

Figure 1.1 (b); *Prunus Africana* trees in Kakamega forest

1.1: Statement of the problem

Models of the geographic distributions of species have wide application in ecology. But the nonspatial, single-level, regression models that ecologists have often employed do not deal with problems of irregular sampling intensity/irregularity or spatial dependence, and do not adequately quantify uncertainty. In this study, we focus on how to build statistical models that can handle the features of spatial prediction and provide richer, more powerful inference about species niche relations, distributions, and the effects of human disturbance. We therefore want to extend the generalized linear model (GLM) to be explicit in their use of the spatial information.

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 (especially the rare ones) 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. As a result, this research examines the environmental factors contributing to species distributions with a key focus to the *prunus africana* , in order to make use of the information on the species survey data and hence we predict and map the distribution of species for the purpose of planning.

1.2: Objectives of the study

The objectives of this study were divided into two:General objective and the specific objectives

1.2.1: General objective

The overall objective of this study is to develop a species distribution model to be used for describing the *Prunus Africana* patterns as well as make predictions and map the species distributions. Species Distribution models assist in characterising the natural distribution of species. In addition,when a 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.

1.2.2: Specific Objectives

The research aims at three specific objectives;

a) **To study Generalized Linear Models (GLMs) that relates the distribution of *Prunus Africana* data to environmental factors;**

The spatial area covered by the species distributions is larger than the sampled area; also, the spatial unit of prediction is usually larger than the sites sampled on the ground. Furthermore, there is the challenge that relates to the heterogeneity of the sampling intensity: that is, while large parts of the domain may be unsampled, there are other areas that may have been heavily sampled. Finally, the environmental data for the region of interest are typically available at a much coarser spatial resolution than the scale at which species distribution data may be collected. Relating the species distribution data to the environmental data and to the region of where prediction is sought thus often presents problems of spatial misalignment since at

times one may only have point data, data in form of regular grid or irregular polygons, and hence the different kinds of data sources don't lineup.

This study aimed at looking into those factors that make *Prunus Africana* to inhibit a specific area, that is, choice of those species in relation to the environment.The major environmental concerns are rainfall,temperature and altitude.Other relating factors that the research establish are human population and market distribution for the *Prunus Africana* products.

b) **To measure the degree of the spatial dependence or spatial autocorrelation on the distribution of *Prunus Africana* species**

In ecological data, pairs of observations that are closer together often tend to be more similar than pairs of observations that are further apart. This is known to produce positive autocorrelation. Using models that ignore the dependence can lead to inaccurate parameter estimates and inadequate quantification of uncertainty. In addition, if spatial dependence is ignored, then much of the meaningful information will be left out and hence the need for the determination of the degree of the spatial autocorrelation.

This research aims at establishing a set of scale –dependent predictors to represent factors affecting the distribution of *Prunus Africana* 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 locality. Geographic Information System (GIS) is a useful tool for analyzing resource selection of species and predicting and mapping habitat suitabilitywhen the degree of correlation among observations depends on their relative locations.

c) **To determine the quantification of uncertainty in spatial models and perform predictions for *Prunus Africana***

Predictions involve extrapolation to unobserved parts of the study region and the large scale areal units. Thus, assessment of uncertainty in such predictions is crucial when they are used

to set conservation policy or to evaluate the impact of climate change on species. It is important to note that ecological distribution models that focus exclusively on the mean (average) yield false confidence in the predictions made (Ver Hoef et al. 2001)

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, in particular for new combinations of predictor values or for predictor values outside the given data. This prediction to new environments is generally termed forecasting or extrapolation (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(e.g. objective (b)).The research will show that there is a strong case for making a species model spatially explicit.

1.3: Justification of the study

Species distribution maps are important to the conservation managers, planners and ecological researchers wishing to generate species red lists, measure rate of decline, identify the areas of endesm and locate 'hotspot' of diversity Gaston and Blackburn (2000). In the previous studies this distribution maps have represented the probabilities of occurrence on the sampled locations only. Thus, with recent advances in species distribution models one is able to potentially forecast anthropogenic effects on patterns of biodiversity at different spatial scales, for instance, Araujo and Williams (2000) suggest that probabilities of occurrence could be used as attributes for species-based selection. These probabilities would be regarded as estimates of the likelihood that a species might occur at a given unsampled location.

CHAPTER TWO

LITERATURE REVIEW

This chapter reviews existing literature on the variables of the study. The literature has been derived from various sources such as past studies, documented reports, researches and secondary available information in support of the study.

2.1: Introduction

Ecologists increasingly use species distribution models to address theoretical and practical issues, including predicting the response of species to climate change (Midgley et al. 2002), identifying and managing conservation areas (Austin and Meyers 1996). In addition ecologists are also interested in finding additional populations of known species or closely related sibling species (Raxworthy et al. 2003), and they are also interested in seeking evidence of competition among species (Leathwick 2002). In all of these applications, the core problem is to use information about where a species occurs (and where it does not) and about the associated environment to predict how likely the species is to be present or absent in unsampled locations.

Spatial prediction of species distributions is directly related to the concept of the environmental niche, a specification of a species' response to suite of environmental factors (MacArthur et al. 1966, Austin et al 1980, Brown et al. 1995). However, the environmental factors alone are not sufficient to account for the species distributions. According to (Gaston 2003) there are other ecological processes that may also affect the spatial arrangement of species distributions which include; dispersal, reproduction, competition and the dynamics of large and small populations.

There are various statistical techniques that have been used to model species probabilities of occurrence in response to the environmental variables. These techniques differ in their ability to summarize useful relationships between the response and the predictor variables. The models can be grouped into three categories as follows;

Analytical or mathematical models: They focus on generality and precision and are designed to predict accurate response within a limited or simplified reality

Mechanistic, physiological, causal or process models: They are designed to be realistic and general. The base predictions on real cause-effect relationships

Empirical/ statistical/ phenomenological models: They sacrifices generality for precision and reality.

Gaisan and Zimmermann (2000) reviewed various modeling techniques.

2.2: Generalized regressions

Regressions relate response variable to a single (simple regression) or a combination (multiple regression) of environmental predictors (explanatory variables). An important statistical development has been the advance in regression analysis provided by generalized linear models (GLM) and generalized additive models (GAM). GLM s are mathematical 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 relationship (called the link function) between the mean of the response variable and the linear combination of the explanatory variables. The GLMs are more flexible and better suited for analyzing ecological relationships, which can be poorly represented in classical Gaussian distributions (Austin 1987).

13

Generalized Additive Models (GAMs), are semi-parametric extensions of GLMs; the only underlying assumption made is that the functions are additive and that the components are smooth (Hastie & Tibshirani 1986, 1990). A GAM also uses the link function just like the GLM in order to establish a relationship between the mean of the response variable and a 'smoothed' function of the explanatory variable(s). GAMs are simply data driven rather than model driven, that is, they allow the data to determine the nature of the relationship between the response and the set of the explanatory variables rather than assuming some form of parametric relationship (Yee & Mitchell 1991).

2.3 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 2.1.

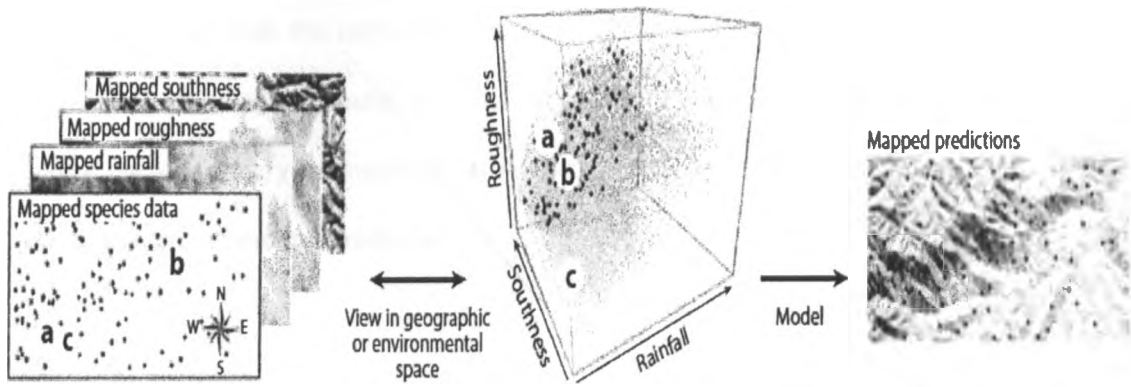Figure 2.1: Spatial Autocorrelation of the Environment

Figure 2.1 :(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)

2.4: Classification techniques

The techniques include the methods such as classification and regression tree (CART), rule-based classification and the maximum likelihood classifications. The classification techniques assign a class of binary or multinomial response variable to each combination (nominal or continuous) environmental predictor

2.5: Ordination techniques

Most models use the ordination techniques to predict the distribution of species. They are based on canonical correspondence analysis (CCA). It is a direct gradient analysis where the principal ordination axes are constrained to be a linear combination of environmental descriptors.

2.6: Bayesian Technique

Bayesian modeling has been rarely applied to ecological spatial prediction. A few early applications have used Bayes' Theorem to combine relationships between observed data and individual predictive factors with prior probabilities of presence to produce probability surfaces for species (Aspinall 1992; Aspinall and Veitch 1993; Royle et al. 2002) or vegetation types (Fischer 1990). Since these approaches use a contingency table approach and carry over only point estimates from the data stage to the generation of predictions, they are not directly comparable to hierarchical regression models. Only very recently have Wikle and collaborators presented examples of full Bayesian hierarchical 50 Explaining Species Distribution Patterns modeling applied to individual plant or bird species(Wikle 2002, 2003; Wikle and Royle 2002). See also Clark et al. (2003) in this regard. A Bayesian model also

16

allows the introduction of spatial dependence naturally into the model through random effects that capture spatial association not contained in the other covariates. Through marginalization, the spatial random effects are incorporated directly into the model likelihood, and are fitted simultaneously with the other model parameters. They are introduced into the mean (on a transformed scale) and, controlling for the other covariates in the mean, encourage mean behavior to be similar when cells are close to each other. Like random effects in general, they soak up the lack of explanation of the fixed component of the mean but in a spatial fashion. Thus, they account for omitted or unmeasured explanatory variables having spatial content. Like other parameters, the spatial random effects have fully specified probability distributions, providing information about both their magnitude and uncertainty. Their effect in explaining potential presence is explicitly specified, so their contribution to the model and to prediction may be rigorously investigated. Random effects may be introduced in different ways, e.g., through a conditional auto-regressive (CAR) model as implemented here (Besag et al., 1974) or via a matrix of spectral functions (Hooten et al. 2003). Finally, through the implicit dependence structure, spatial modeling for random effects allows learning about their contribution even for cells where there has been no sampling, accommodating gaps in sampling and irregular intensity in sampling.

2.7: Prunus Africana

Prunus africana (Hook f.), also known as Pygeum africanum, or African Stinkwood, is a large evergreen tree from the Rosaceaea family. It is found in the montaine forests of Kenya, Cameroon, Zaire, and Madagascar (Figure 2.2)

Figure 2.2: Locations of Prunus Africana in Africa
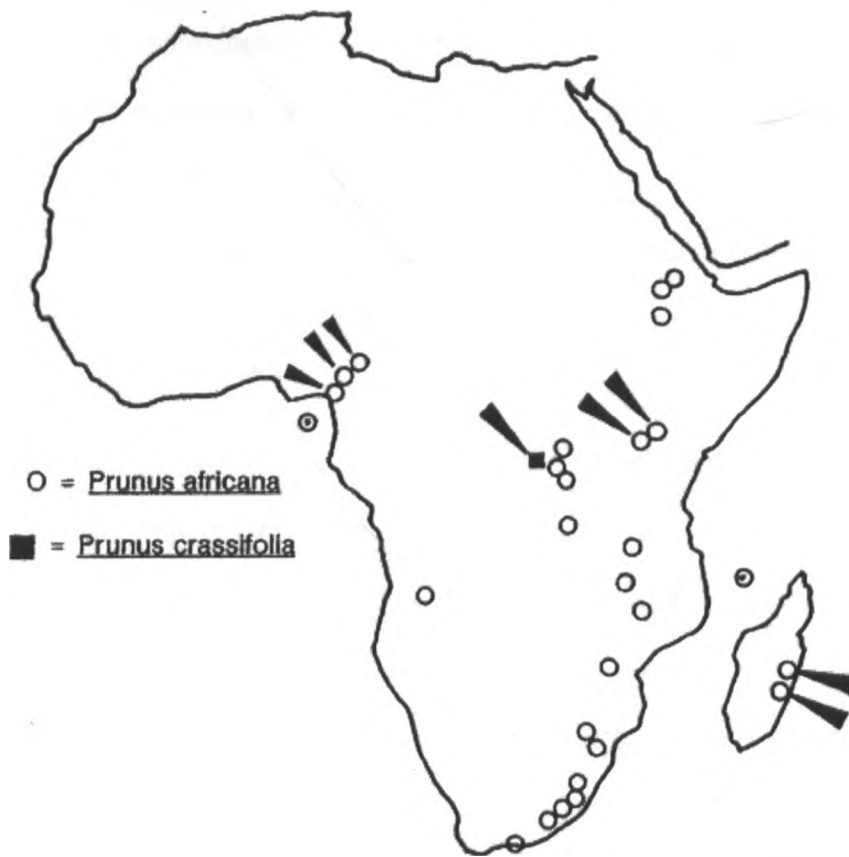
O = Prunus africana
■ = Prunus crassifolia

It is an upper canopy tree that can grow as high as thirty meters (Hall et al, 2000). Prunus bark contains three active constituents including phytosterols, which are anti-inflammatory agents (Hall et al, 2000; Longo, 1981). It takes fifteen years for the bark to develop the active ingredients necessary for medicinal use (Longo, 1981). Traditional medicinal uses of the bark include the treatment of stomach aches, urinary and bladder infections, chest pain, malaria, and kidney disease.

The international market for *Prunus bark* is focused on its use as a treatment of prostate gland hypertrophy and the closely related but more serious condition benign prostatic hyperplasia (Hall et al, 2000; Cunningham & Mbenkum, 1993). Prostate enlargement currently affects r >50% of men over the age of 50. The market for prostrate treatments, especially herbal

18

remedies, is growing every year. Global demand rose from 2,800 Mg (tonne) in 1995 to 3,100 Mg in 1997 (Anonymous, 2000). The over-the-counter value of retail trade in the United States alone is approximately $220 million a year (Cunningham et al, 2002). The bark of Prunus africana is exported from Africa to Europe, mainly France, where it is processed and then distributed globally (Cunningham & Mbenkum, 1993). The demand of the market and the limited geographic distribution of Prunus africana has resulted in widespread endangerment due to over-harvesting Cunningham & Mbenkum, 1993; Hall et al, 2000).

In response to the growing scarcity, *Prunus africana* was declared endangered under CITES II (the convention on trade in endangered species) (Cunningham et al, 1997), and listed as vulnerable by the IUCN (IUCN, 2002). As a result, all imports and exports of Prunus must be declared. In addition, exporting countries must demonstrate that the source of the *Prunus africana* was harvested in a "sustainable" manner. Unfortunately, monitoring trade is difficult because *Prunus africana* is exported in many forms, including bark, bark extract, capsules, and tonic (Hall et al, 2000). In addition, a lack of regulatory infrastructure within forests where *Prunus africana* thrives lessens the efficacy of harvesting limitations.

Within the Kakamega forest, no large scale bark harvesting is currently occurring. In fact, there is no knowledge of the potential for a widespread *Prunus africana* bark market (Fashing, 2004). *Prunus africana* has been found to play a key role in forest ecology. It is the primary food source for Columbus monkeys, the dominant primate species in the forest. Studies have shown that the tree is scarce throughout the forest (Fashing, 2004; Tsingalia, 1989). Attempts to propagate *Prunus africana* seedlings in nurseries as well as attempts to integrate them into agroforestry projects have largely failed.

2.8: 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/or irregular grid,some are available at point locations(point data) thus bringing inconsistencies,b iasness(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 problem.

20

CHAPTER THREE

DATA AND METHODOLOGY

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 nowadays 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 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: The study Area

The Kakamega National Forest is located in the Western Province (0.28° N and 34.8° E) of Kenya. It is a montaine forest, located approximately 1600m above sea level, has an average annual rainfall of 2,000mm, and temperature ranges between 11° and 26° C in both the rainy (April-November) and the dry season (December-March). The forest lies in the Lake Victoria Basin, and is an important watershed for rivers that flow into Lake Victoria such as the Isiukhu and Yala (Figure 3.1).

The forest is governed by two separate management bodies: the Forest Department (FD) and the Kenya Wildlife Service (KWS). The majority of the forest (approximately 200 KM^2) is managed by the Forest Department. This area was established as a National Forest in 1933, and affords free entrance to the public. Within this area limited extractive use is permitted, including deadwood and grass collection and licensed cattle grazing. In the past, some licensed commercial logging of plantations was also permitted, though currently there is a country-wide logging ban which prohibits such activity. Kenya Wildlife Service (KWS) governs only approximately 40KM^2 of the northern area of the forest, which was designated a National Reserve in 1986. There is only one legal entrance into this area, through the forest station. Within this portion of the forest there is a policy of allowing no extraction of natural resources or biota. *Prunus Africana* is one of the trees found in the forest and hence a suitable region to study the distribution of the species. Figure 3.1 shows the distribution of Prunus Africana in Kakamega forest of Western Kenya.

Figure 3.1: Prunus Africana Distribution map in Kakamega Forest

6000   0   6000   12000   Meters
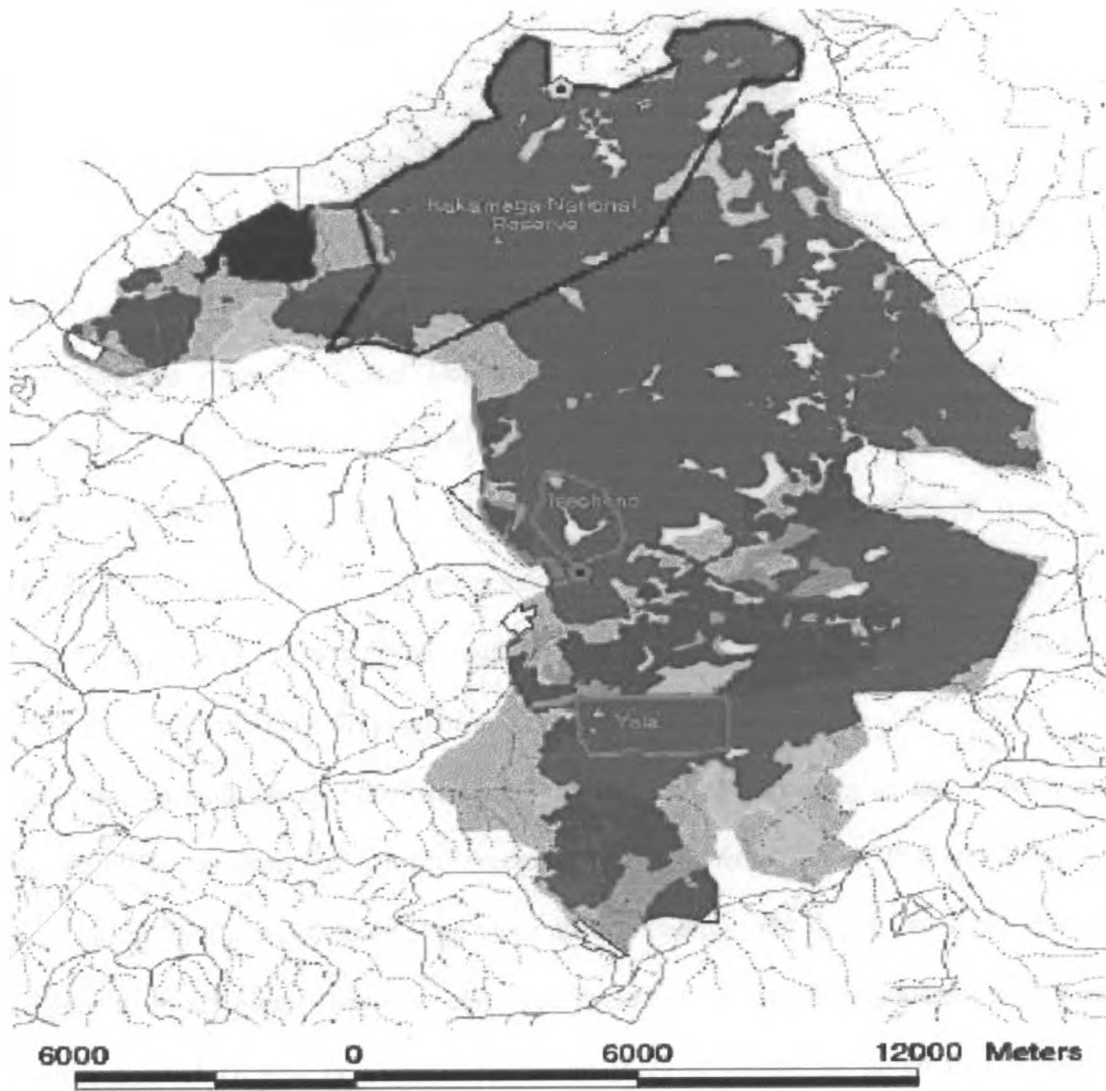
Key

Parts of the Forest with the high numbers of Prunus Africana

Forest cover

3.2: 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.

23'

Further Kenya Forest Department is a credited institution and the economy depends on its data for research and development not only in Kenya but also internationally.

3.3: Data Collection

The land cover of the Kakamega forest is very heterogeneous. Plots were stratified based on land cover type, age, and managing agency. This was done using land cover maps from 1975-2000 obtained from the Kenya Department of Remote Sensing and Resource Survey (DRSRS). These maps were created by DRSRS using aerial photography as well as Landsat TM satellite images (Rogo et al., 2003). The maps identified three types of tree cover: indigenous forest, hardwood timber plantation, and softwood timber plantation. Field observations as well as Forest Department registry records resulted in the addition of a fourth forest class: mixed indigenous plantation.

In addition to land-use cover, plots were further stratified according to age. 'Young plots were identified as areas that had regenerated or been planted between 1986 and 2000 (< 14 years old in 2000), while 'old' plots were present before 1986 (> 14 years old in 2000). Indigenous forest plots were also classified according to the management body: Kenya Wildlife Service (KWS) or Forest Department (FD). It should be noted that within the indigenous forest classification an 'old' plot would contain vegetation of many ages, due to forest successional patterns as well as local disturbances. In addition, both 'old' and 'young' plots represent a range of stand ages. For example, registry data indicates that while some hardwood plantations were established 50 years ago, others may be only 25 years old (Kakamega Forest Department, 2003, Glenday, 2004).

A total of 95 plots were sampled. The number of plots sampled in each sampling class was roughly proportional to the amount of forest area covered by each class. In addition,

24

heterogeneity of indigenous forest structure (as opposed to plantations) was also considered (table 1).

Table 3.1: Plot stratification Strata

Plot stratification Strata	Number of 'old' plots	Number of 'young' plots	Total number of plots
Forest Department Indigenous Forest (IF-FD)	32	11	43
Kenya Wildlife Service Indigenous Forest (IF-KWS)	14	5	19
Mixed Indigenous Plantation (MI)	4	6	10
Softwood Timber Plantation (SW)	5	5	10
Hardwood Timber Plantation (HW)	6	7	13

3.4: Modeling of Species Distributions

3.4.1: Linear regression

In modeling statistical data, the classical linear model involves a relationship of the form:

$$Data = Pattern + \text{Re}\,sidual \qquad\qquad 3.1$$

Where;

Data is the response (dependent) variable

Pattern is the set of predictor variables. It is the part of the model that explains the variability in the data

Residual represents measurement error, as well as any variation unexplained by the linear model (i.e. it is stochastic part and is the unexplained variability in the data). This model assumes a Gaussian (normal) distribution for the response variable and an identity link. The linear regression is limited by three basic assumptions namely;

  i.   The errors are assumed to be identically and independently distributed; this includes the assumption that the variance of Y is a constant across observations.

  ii. The errors are assumed to follow a normal (Gaussian) distribution

  iii. The regression function is linear in the predictors

3.4.2: Generalized linear models

The generalized linear model is a mathematical extension of linear model that allows for non-linearity of non-constant variance structures in the data. In order to build a generalized linear model three decisions have to be put into consideration, these are;

a. What is the distribution of the data (for fixed values of the predictors and possibly after a transformation)?

b. What function of the mean will be modeled as linear in the predictors?

c. What will the predictors be?

This model is defined in terms of a set of independent random variables; $Y_1, Y_2, Y_3, ..., Y_N$ each with a distribution from the exponential family (Binomial, Poisson, Gamma, Negative Binomial and Normal) and has the following properties;

i. The distribution of each Y_i has the conical form and depends on the single parameter θ_i (the θ_i's do not all have to be the same), thus

$$f(y_i; \theta_i) = \exp[y_i b_i(\theta_i) + c_i(\theta_i) + d_i(y_i)] \qquad 3.2$$

ii. The distribution of all the y_i's are of the same form so that the subscripts of b, c and d are all needed. Thus the joint probability density function of $Y_1, Y_2, Y_3 ..., Y_N$ is;

$$f(y_1, ..., y_N; \theta_1, ..., \theta_N) = \prod_{i=1}^{N} \exp[y_i b(\theta_i) + c(\theta_i) + d(y_i)]$$

$$= \exp[\sum_{i=1}^{N} y_i b(\theta_i) + \sum_{i=1}^{N} c(\theta_i) + \sum_{i=1}^{N} d(y_i)] \qquad 3.3$$

iii. The linear relationship between predictors X and the response variate $\underline{y}$ is given by;

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \qquad 3.4$$

Where; $E[y] = \mu$, $V(y) = \sigma^2 I_N$ and $\underline{\mu} = X\underline{\beta}$

The vector $\underline{\varepsilon}$ is the error term that measures the discrepancy of the fitted model and the observed data. The error terms are assumed to be normally distributed with zero mean and

unit variance and $\underline{\beta}$ is a set of unknown regression coefficients. Equation 3.4 is often referred to as the generalized linear model. This model has three main components namely;

i.   Response component – the response variables $Y_1, Y_2, Y_3,..., Y_N$ are assumed to share the same distribution from the exponential family

ii. The systematic component ; a set of parameters $\underline{\beta}$ and explanatory variables X produce a linear predictor η given by;

$$\eta = \sum_{j=1}^{N} X_j \beta_j \qquad\qquad 3.5$$

iii. The monotone differentiable link function g that gives the relationship between the mean response variable and the linear combination of the explanatory variables, such that;

$$g(\underline{\mu}_i) = \underline{X}_i^T \underline{\beta} \qquad\qquad 3.6$$

Where $\mu_i = E(\underline{y}_i)$, $\underline{X}_i$ is a $PX1$ vector explaining of explanatory variables (dummy variables for factor levels and measured values for covariates); $\underline{X}_i = \begin{bmatrix} x_{i1} \\ . \\ . \\ . \\ x_{ip} \end{bmatrix}$, so that

$\underline{X}_i^T = [x_{i1},..., x_{ip}]$; and $\underline{\beta}$ is the $PX1$ vector of parameters; $\underline{\beta} = \begin{bmatrix} \beta_1 \\ . \\ . \\ . \\ \beta_p \end{bmatrix}$

It is important to mention that the generalized linear models (G.L.Ms) are more flexible and better suited for analyzing ecological relationships which can be poorly represented by classical Gaussian distributions.

3.4.3: Logistic regression and Binomial models

Logistic regression is widely used for analyzing data involving binary or binomial responses and several explanatory variables. In this study we use the Binomial GLM with a logit link to model the probability binomial observations of the form $\frac{Y_i}{n_i}$, $i = 1,2,...,n$ such that

$$E(Y_i) = n_i p_i \qquad\qquad 3.7$$

Where p_i is the probability of species being present and n_i is the number of observations made in site i

The linear logistic model relating p_i to the k explanatory variables; $X_{1i},..., X_{ki}$ associated with i^{th} observations is;

$$\log it(p_i) = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{Ki} \qquad\qquad 3.8$$

Where $\log it(p_i) = \log\left[\dfrac{p_i}{1-p_i}\right]$

This is the logistic transformation of the probability of presence of species. To obtain the expected values we compute the inverse link function.

Let;

$$\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{Ki} = n_i \qquad\qquad 3.9$$

This implies that;

$$\log\left[\frac{p_i}{1-p_i}\right] = i^{ni} \qquad\qquad 3.10$$

This implies that; $p_i = (1-p_i)e^{ni}$

Therefore the logit inverse is:

$$p_i = \frac{e^{ni}}{1-e^{ni}} \qquad\qquad 3.11$$

In order to fit a linear logistic model to given set of data, the $k+1$ unknown parameters $\beta_0, \beta_1, ..., \beta_k$ are estimated using the method of maximum likelihood estimation. The likelihood function is given by;

$$L = \prod_{i=1}^{n} \binom{n_i}{y_i} p_i^{y_i} (1-p_i)^{n_i-y_i} \qquad 3.12$$

The log-likelihood function is given by

$$\ln L = \sum_{i=1}^{n} \left[\ln \binom{n_i}{y_i} + y_i \ln p_i + (n_i - y_i) \ln(1-p_i) \right]$$

$$= \sum_{i=1}^{n} \left[\ln \binom{n_i}{y_i} + y_i \ln \left(\frac{p_i}{1-p_i} \right) + n_i \ln(1-p_i) \right]$$

$$= \sum_{i=1}^{n} \left[\ln \binom{n_i}{y_i} + y_i n_i - n_i \ln(1-e^{n_i}) \right] \qquad 3.13$$

We then differentiate the log-likelihood with respect to the unknown $k+1$ parameters and equate to zero to get the maximum likelihood estimates of β

3.4.4: 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 $X_1, X_2, ..., X_k$. The class of generalized additive models replaces the linear form $\sum \beta_j X_j$ by a sum of smooth functions $\sum s_j(x_j)$. The $s_j()s$ are unspecified functions that are estimated using a scatter plot smoother, in an iterative procedure. The linear predictor $\sum \beta_j x_j$ is replaced by the additive predictor $\sum s_j(x_j)$. 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

30

distribution) and a link function g relating the expected value of the distribution to the m predictor variables, and attempts to fit functions $f_i(x_i)$ to satisfy:

$$g(\in (Y)) = \beta_0 + f_1(x_1) + \dots + f_m(x_m)$$ 3.14 (a)

or generally

$$\in \left(Y / X = s_0 + \sum_{j=1}^{p} s_j x_j \right)$$ 3.14 (b)

Where $\in s_j(x_j) = 0$ for every j

The functions $f_i(x_i)$ 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 $f_1(x_1)$ and then use a factor model for $f_2(x_2)$. 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.

Over fitting 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.

3.4.5: Getting spatially explicit model: A 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 $(x^T\beta)$ which is catered for by the logistic $(Logit)$ function.

Considering equation; $Y_i + \sim Binomial(n_i p_i)$; meaning that the probability p_i that a species occurs in cell i is related to the environmental variable) 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.

$$\frac{p}{1-p} = \exp(\beta_0 + \beta_1 x_1 +,..., \beta_p x_p) + \theta_i \qquad\qquad 3.15$$

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 $X_1, X_2,..., X_n$ be a finite collection of random variables which are associated with sites labelled 1,2 ... n,respectively.For each site, $P(x_i/x_1,..., x_{i-1}, x_{i+1},..., x_n)$ 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.

CHAPTER FOUR

RESULTS

4.0: Introduction

This chapter discusses the research data obtained. The relationships of *Prunus Africana* 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' characterization of the distributions of species. The research now turns to a more thorough evaluation of the model output, including the estimates for the environmental coefficients, the spatial random effect variables, and the uncertainty associated with model parameters. To simplify the analayis,the reseach started by giving the various factors variable names as follows in relation to the model

$$g(u_i) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 W \qquad 4.1$$

Where; $g(u_i)$ is the Prunus Africana density (PA)

X_1 is the population density (P) in the area; X_2 is market distribution (M); X_3 is the average ranifall expericed in the given area(R), and X_4 is the temperature in the area (T)

4.1.1: Summary of the data

The section gives the summary of the data considered. Continuous variables include population density (P), rainfall (R) and temperature(T) while the outcome, *prunus africana* occurrence and market occurence are binary(0/1) response.

4.1.2: Continuos variables summary

Population density	Rainfall	Temperature
Min. : 0.00	Min. : 400	Min. : 9.00
1st Qu.: 42.18	1st Qu.: 875	1st Qu.:15.00
Median : 71.12	Median :1000	Median :17.50
Mean : 66.46	Mean :1060	Mean :18.27

4.1.3: Binary variables summary

PA represent *Prunus Africana* and M reprents market

table(PA,M)

	M	
PA	0	1
0	1	1
1	0	38

4.1.4: Using the logit Model

Here, a logistic regression model is estimated using the GLM (generalised linear model) function.The use of 'as.factor (M)' indicates that the market should be treated as a factor (i.e categorical) variable. Having given the model name (mylogit), R will not produce any output from the regression but instead the research used summary commands.

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.587	-0.7962	-0.6072	1.1462	2.421

Coefficients:

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
Intercept	*0.864*	*1.628*	*2.04*	*0.000524*
P	-1.0063	0.003142	-2.1	0.04056
R	0.24	0.332973	2.627	0.02681
T	-0.2378	0.31254	-2.133	0.032356
As.factor(M)1	1.8645	0.215201	2.976	0.000345

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 48.084 on 39 degrees of freedom

Residual deviance: 37.27 on 35 degrees of freedom

AIC: 10

In the output above, the first thing is the call reminding r about the model we ran and the options specified. Next are the deviance residuals,which 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.

The following can be noted:

i. For every unit change in population density, the log odds of *prunus africana* density (versus none) decreaeses by *1.0063*.

ii. For every unit change in in rainfall,the log odds *prunus africana* density (versus none) increases by 0.2

iii. For every unit change in temperature,the log odds of *prunus africana* density (versus none) decreases by 0.2378

iv. Having the market for *Prunus Africana* in the region increases the log odds of *prunus africana* density by *1.8645*. This can be interprated to mean there is an effect of *prunus africana* density which can be associated with the market for the tree products.

The model can thus be written as:

$$\log\left(\frac{p}{1-p}\right) = 0.864 - 1.0063X_1 + 0.24X_2 - 0.2378X_3 + 1.8645X_4 \qquad 4.2$$

The cefficient estimates for the confidence intervals were also confirmed using R software and the results were tabulated as follows.

	2.5%	97.5%
Intercept	-2.36402	-0.94831
P	0.006285	0.007834
R	0.150346	1.4100695
T	-1.28901	0.412901
As.factor(M)x	4.3984	NA

4.1.5: Exponeniating the coefficients

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

Intercept	P	R	T	As.factor(M)x
1.189028	2.802643	1.20341	4.23108	7.58201

This means that for one unit increase in market demand, the odds of having *Prunus Africana* in the area decreases by a factor of 7.58201

4.1.6: Predicted probabilities

This was carried out for both categorical and continuous predictor variables. To carryout this, a new data frame was created with the values whose independent variables were to be taken to create the predictors.

The research started by calculating the predicted probability of *Prunus Africana* at each value of the market demand, holding population density, rainfall and temperature constant. An object was created In R with the observations of presence and absence (1/0). The second, third and fourth objects of P, T and R were created that only took on one value each. In this

36

case, the variables were set to their means using the R commands, 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 data frame called *newdata* and the final code asked R to display the data frame *newdata*

4.1.7: New data Results.

M		P	R	T
1	0	56.3490	1280	20.675
2	1	56.3490	1280	20.675

This data frame was then used to calculate the predicted probabilities through R.

M		P	R	T	M*x*
1	0	56.3490	1280	20.675	0.74
2	1	56.3490	1280	20.675	0.13

The output confirms that the predicted probability of having the *Prunus Africana* is 0.74 when there is no market demand for its products and 0.13 when there is market demand for *Prunus Africana.*

4.1.8: Model fit

This is useful when comparing competing models. The output obtained from the summary(mylogit) 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 confirms 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) we used the R commands and the following result was obtained.

[1] 15.88122

The degrees of freedom for the difference between the two models are equal to the number of predictor variables in the model, and was also obtained using R. The result obtained was;

[1] 2

Finally, the p-value was as well obtained from R, and the value obtained was;

[1] 6.456192

Interpretations

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 residualis -2*log likelihood)

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

Further the correlation matrix was obtained through R as follows;

	PD	P	M	R	T
PD	1.0000000	0.3748285	0.6979824	0.4016462	-0.3206608
P	0.3748285	1.0000000	0.3066419	0.7686704	-0.5443849
M	0.6979824	0.3066419	1.0000000	0.2573632	-0.2570967
R	0.4016462	0.7686704	0.2573632	1.0000000	-0.4945105
T	-0.3206608	-0.5443849	-0.2570967	-0.4945105	1.0000000

Notable is that all immediate factors that the researcher put into focus were all positively correlated with the *Prunus Africana* density, that is, Population density in the area, Rainfall

and market demand. Temperature showed negative association perhaps due to the selective nature of this *Prunus Africana* habitant determined by the other variables like the altitude.

4.2.2: Covariance

In addition, the covariance matrix was obtained using R and tabulated as follows;

	PD	P	M	R	T
PD	0.04871795	2.907855	0.02435897	31.28205	-0.3448718
P	2.90785474	1235.353546	1.70410814	9533.29097	-93.2328613
M	0.02435897	1.704108	0.02500000	14.35897	-0.1980769
R	31.28205128	9533.290974	14.35897436	124512.82051	-850.256413
T	-0.34487179	-93.232861	-0.19807692	-850.25641	23.7429487

The covariance between Prunus Africana density and 3 of the researched factors was observed to be positive. This is an emphasis to the fact that the *Prunus Africana* occurrence is determined by these factors.

4.3: G.I.S Mapping

An expert-oriented map of *Prunus Africana* presence in Kakamega was created by combining spatially explicit information on the *Prunus Africana* habitat and environmental factors which favour the *prunus Africana* occurence.

4.3.1: *Prunus Africana* Density prediction model.

Using ArcGIS 10, the four variables, rainfall, temperature, market distribution and population densities within the study area were modeled together using the geo-processing tool in Arc-GIS software. The *Prunus Africana* distribution modeling tool was developed and spatial join 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 *Prunus Africana* occurrence noted 1 and

the areas with the very low chance of *Prunus Africana* 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 *Prunus Africana* occurrence, 2-high; 3-average; 4-low and 5-very low. This was extrapolated to produce *Prunus Africana* density prediction map for the entire study area.

4.4: Assumption

The assumptions in the model was that temperature, rainfall, Population density and Market demand are the main factors that influence *Prunus Africana* occurrence 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 forest department (FD) which collects and compile datasets on trees 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 *Prunus Africana* distribution density, and this may be a limitation since there are many other variables which have an influence on *Prunus Africana* distribution. In addition, 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 FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.0: Introduction

This chapter presents the discussion of key data findings, conclusion drawn from the findings highlighted and recommendation made there-to. The conclusions and recommendations drawn were focused on addressing the purpose of this study which was; to develop a species distribution model to be used for describing the *Prunus Africana* 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 *Prunus Africana* 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 a species includes more information than can be explained through just the mean effect of a suite of environmental variables, that is, the temperature, rainfall, population density and further addition of the market demand presence like cutting trees for timber, firewood and so on. There are two main explanations for this. One is that biological processes tend to generate spatial pattern. In the case of the *Prunus Africana,* the probability that a site contains a species depends on 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 (for instance, effects of population desisty on the *Prunus Africana* occurence, 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 individual species distributions, as well as interspecific interactions and thus community composition and potentially ecosystem processes.

The second explanation for autocorrelation is 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 geographical range to another (Gaston 2003). 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 bedramatically underestimated and poorly characterized. One effect of this is that a model will identify more explanatory variables as significantly related to species presence/absence (Ver Hoef et al. 2001).

Additional studies which seek to understand resource selection of the *Prunus Africana* at fine scales are crucial to developing a foundation for determining appropriate management, prevention and conservation strategies for this species. In the subregion studied, the *Prunus Africana* preferred proximity to one or many ecological features present in these regions including population density, temperature scales and rainfall .We note here that these factors are some of those that influence the occurrence of the *Prunus Africana* and have been noted to positively influence this existence except the the population density and the market demand. It is upon the government and all stake hokders concerned to minimise or control completely the destruction of the trees from their inhabitants. As it has been statistically noted, all these factors are interelated and influnce this distribution of *Prunus Africana*. From the prediction by GIS and that by modeling, it has been noted that the *Prunus Africana* is set to decrease with time. This is a big threat to the Kenyan economy.

5.2: Recommendations

The following are the research recommendations

5.2.1: Resource selection of the *Prunus Africana*

GIS proved to be an extremely useful tool for elucidating relationships between *Prunus Africana* 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, especially at a 1kilometer resolution. The presence of this *Prunus Africana* 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 *Prunus Africana* 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 that 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 aparticular challenge – for example, *Prunus Africana*

grow in montane areas (High altitude areas). There is no comprehensive dataset of locations of the *Prunus Africana* for the entire region so the model excluded this important influence on *Prunus Africana* habitat.

As with all expert models, the model depends on multiple iterations of revisions and "fine-tuning" based on

i.   trial and error and;

ii. expert opinion by biologists and wildlife ecology experts who are familiar with the biological habitat selection of the *Prnus Africana*.

It is recommended that both the Biological 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.

5.2.2: Stakeholders

This research recomends to stakeholders to speed up measures for the control of deforestation and especially to the destruction of *Prunus Africana* as it poses as a major threat to the Kenyan economy. The Kenya Forest Department can use such a research background to determmine best areas to grow more trees and also to encourage farmers to carry out agroforestry for not only increased production of tree products but also to prevent desertification in the country. Kenya Wildlife Service and the National Museums of Kenya can use it for conservation purposes,management and control mechanisms.

REFERENCES

1. Ahmadi-Nedushan B, St-Hilaire A, Berube M, Robichaud E, Thiemonge N, et al. 2006. A review of statisticalmethods for the evaluation of aquatic habitat suitability for instream flow assessment. River Research andApplications 22: 503-23

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

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

4. Beissinger SR, Walters JR, Catanzaro DG, Smith KG, Dunning JB, et al. 2006. Modeling approaches in avian conservation and the role of field biologists. Auk 123: 1-56

5. Best BD, Halpin PN, Fujioka E, Read AJ, Qian SS, et al. 2007. Geospatial web services within a scientific workflow: Predicting marine mammal habitats in a dynamic environment. *Ecological Informatics* 2: 210-23

6. Block WM, Brennan LA. 1993. The habitat concept in ornithology: theory and applications. Current Ornithology11: 35-91

7. Turner W, Spector S, Gardiner N, Fladeland M, Sterling E, et al. 2003. Remote sensing for biodiversity science and conservation. Trends in Ecology and Evolution 18: 306-14

8. 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

9. Buse J, Schroder B, Assmann T. 2007. Modelling habitat and spatial distribution of an endangered longhorn beetle - A case study for saproxylic insect conservation. *Biological Conservation* 137: 372-81

10. Cumming GS. 2000. Using habitat models to map diversity: pan-African species richness of ticks (*Acari: Ixodida*). *Journal of Biogeography* 27: 425-40

11. Dedecker AP, Goethals PLM, Gabriels W, De Pauw N. 2004. Optimization of Artificial Neural Network (ANN)model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders, Belgium). *EcologicalModelling* 174: 161-73

12. 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

13. Garza-Perez JR, Lehmann A, Arias-Gonzalez JE. 2004. Spatial prediction of coral reef habitats: integrating ecology with spatial modeling and remote sensing. *Marine Ecology-Progress Series* 269: 141-52

14. Hortal J, Garcia-Pereira P, Garcia-Barros E. 2004. Butterfly species richness in mainland Portugal: predictive models of geographic distribution patterns. *Ecography* 27: 68-82

15. Anadon JD, Gimenez A, Martinez M, Palazon JA, Esteve MA. 2007. Assessing changes in habitat quality due to land use changes in the spur-thighed tortoise Testudo graeca using hierarchical predictive habitat models.*Diversity and Distributions* 13: 324-31

16. Beever EA, Swihart RK, Bestelmeyer BT. 2006. Linking the concept of scale to studies of biological diversity: evolving approaches and tools. *Diversity and Distributions* 12: 229-35

17. Boyce MS. 2006. Scale for resource selection functions. *Diversity and Distributions* 12: 269-76

18. Bradshaw GA, McIntosh BA. 1994. Detecting climate-induced patterns using wavelet analysis. *Environmental Pollution* 83: 135-42

19. Brotons L, Herrando S, Pla M. 2007. Updating bird species distribution at large spatial scales: applications of habitat modelling to data from long-term monitoring programs. *Diversity and Distributions* 13: 276-88

20. 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

21. Dollar ESJ, James . CS, Rogers KH, Thoms MC. 2007. A framework for interdisciplinary understanding of rivers as ecosystems. *Geomorphology* 89: 147-62

48

22. Edwards TC, Moisen GG, Frescino TS, Lawler JJ. 2002. Modeling multiple ecological scales to link landscape theory to wildlife conservation. In *Landscape Ecology and Resource Management: Linking Theory withPractice*, ed. JA Bissonette, I Storch, pp. 153-72. Washington D.C.: Island Press

23. Enoki T. 2003. Microtopography and distribution of canopy trees in a subtropical evergreen broad-leaved forest in the northern part of Okinawa Island, Japan. *Ecological Research* 18: 103-13

24. Feagin RA. 2005. Heterogeneity versus homogeneity: A conceptual and mathematical theory in terms of scaleinvariant and scale-covariant distributions. *Ecological Complexity* 2: 339-56

25. 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

26. 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

27. 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

28. 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

29. 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

30. Keitt TH, Urban DL. 2005. Scale-specific inference using wavelets. *Ecology* 86: 2497-504

31. Krawchuk MA, Taylor PD. 2003. Changing importance of habitat structure across multiple spatial scales for three species of insects. *Oikos* 103: 153-61

32. Luoto M, Virkkala R, Heikkinen RK. 2007. The role of land cover in bioclimatic models depends on spatial resolution. *Global Ecology and Biogeography* 16: 34-42

33. Luoto M, Hjort J. 2008. Downscaling of coarse-grained geomorphological data. *Earth Surface Processes and Landforms* 33: 75-89

34. McPherson JM, Jetz W, Rogers DJ. 2006. Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions-possibilities and limitations. *Ecological Modelling* 192: 499-522

35. Meyer CB, Thuiller W. 2006. Accuracy of resource selection functions across spatial scales. *Diversity and Distributions* 12: 288-97

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

37. Bahn V, McGill BJ. 2007. Can niche-based distribution models outperform spatial interpolation? *GlobalEcology and Biogeography* 16: 733-42

38. Bio AMF, De Becker P, De Bie E, Huybrechts W, Wassen M. 2002. Prediction of plant species distribution in lowland river valleys in Belgium: modelling species response to site conditions. *Biodiversity andConservation* 11: 2189-216

39. Bonte D, Baert L, Lens L, Maelfait JP. 2004. Effects of aerial dispersal, habitat specialisation, and landscape structure on spider distribution across fragmented grey dunes. *Ecography* 27: 343-49

40. Borcard D, Legendre P. 1994. Environmental control and spatial structure in ecological communities: an example using oribatid mites (Acari, Oribatei). *Environmental and Ecological Statistics* 1: 37-61

41. Carl G, Kühn I. 2008. Analyzing spatial ecological data using linear regression and wavelet analysis. *Journal for Stochastic Environmental Research and Risk Assessment* 22: 315-24

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

43. 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

44. Austin MP. 1998. An ecological perspective on biodiversity investigations: Examples from Australian eucalypt forests. *Annals of the Missouri Botanical Garden* 85: 2-17

45. Capen DE, Fenwick JW, Inkley DB, Boynton AC. 1986. Multivariate models of songbird habitat in New England forests. In *Wildlife 2000: Modeling Habitat Relationships of Terrestrial Vertebrates. Based on anInternational Symposium held at Stanford Sierra Camp, Fallen Leaf Lake, California*, ed.J Verner, ML Morrison, CJ Ralph, pp. 171-75. Wisconsin, USA: The University of Wisconsin Press

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

47. Heegaard E. 1997. Ecology of *Andreaea* in western Norway. *Journal of Bryology* 19: 527-636

48. Iverson LR, Prasad AM. 1998. Predicting Abundance of 80 Tree Species Following Climate Change in the Eastern United States. *Ecological Monographs* 68: 465-85

49. 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* 22: 967-1001

50. 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

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

52. Maravelias CD, Reid DG. 1997. Identifying the effects of oceanographic features and zooplankton on prespawning herring abundance using generalised additive models. *Marine Ecology Progress Series* 147: 1-9

53. Moore ID, Grayson RB, Ladson AR. 1991. Digital terrain modelling: a review of hydrological,

54. geomorphological and biological applications. *Hydrological Processes* 5: 3-30

55. Prentice IC, Cramer W, Harrison SP, Leemans R, Monserud RA, et al. 1992. A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography* 19

56. Wilson JP, Gallant JC, eds. 2000. *Terrain Analysis: Principles and Applications*. New York: John Wiley and Sons

57. Brown DG, Aspinall RJ, Bennett DA. 2006. Landscape models and explanation in landscape ecology; a space for generative landscape science? *The Professional Geographer* 58: 369-82

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

59. Iverson LR, Prasad AM, Bossenbroek J, Sydnor D, Schwartz MW. 2009. Modeling potential movements of an ash threat: the emerald ash borer. In *Advances in threat assessment and their application to forest andrangeland management.* *www.threats.forestencyclopedia.net*, ed. J Pye, M Raucher

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

61. Morin X, Augspurger C, Chuine I. 2007. Process-based modeling of species' distributions: What limits temperate tree species' range boundaries? *Ecology* 88: 2280-91

62. 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

63. 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.

64. 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

65. Beven K. 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources* 16Burgman M. 2005. *Risks and Decisions for Conservation and Environmental Management*. Cambridge, UK:Cambridge University Press. 488 pp.

66. Dormann CF, Purschke O, García-Marquez J, Lautenbach S, Schröder B. 2008 accepted. Components of uncertainty in species distribution analysis: A case study of the Great Grey Shrike *Lanius excubitor. Ecology*

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

68. Fitzpatrick MC, Weltzin JF, Sanders NJ, Dunn RR. 2007. The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography* 16: 24-33

69. 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

70. 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

71. 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* 15: 361-87

72. 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

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

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

75. 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

76. 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

Appendix 1: R commands that were 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(PD,T)
mylogit<-glm(PD~P+R+T+as.factor(M),family=binomial(link="logit"),na.action=na.pass)
mylogit
summary(mylogit)
confint(mylogit)
M<-c(0,1)
P<-c(mean(data$P))
R<-c(mean(data$R))
T-c(mean(data$T))
newdata<-data.frame(M,P,R,T)
newdata
newdata$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)
exp(mylogit$coefficients) is used to do so.
exp(mylogit$coefficients)
newdata$ecf1<-predict(mylogit,newdata=newdata,type="response")
newdata
M<-c(0,1)
P<-c(mean(data$P))
R<-c(mean(data$R))
```

T<-c(mean(data$T))

newdata<-data.frame(M,P,R,T)

newdata

glm(PA~P+R+T+as.factor(M),family=binomial(link="logit"),na.action=na.pass)

call:

glm(formula = PA ~ P + R + T + as.factor(M), family = binomial(link = "logit"),

 na.action = na.pass)

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)

'log Lik.' -2.021658 (df=5)

cor(data)

 cov(data)