Where, when and why are there elephant poaching hotspots in Kenya?

Edward Opiyo Ouko
March, 2013

Course Tittle: Geo-Information Science and Earth Observation for Environmental Modelling and Management

Level: Master of Science (MSc)

Course Duration: September 2011 – March 2013

Consortium Partners: Lund University (Sweden)
University of Twente, ITC (the Netherlands)

Where, when and why are there elephant poaching hotspots in Kenya?

by

Edward Opiyo Ouko

Thesis submitted to the to the faculty of Geo-information Science and Earth Observation, University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation, Specialisation: (fill in the name of the specialisation)

Thesis Assessment Board

Internal Examiner (Chair): Dr. Albertus G. Toxopeus
External Examiner: Dr. Joost F. Duivenvoorden
Primary supervisor: Dr. Tiejun Wang
Secondary supervisor: Prof. Dr. Andrew Skidmore

UNIVERSITY OF TWENTE.
ITC
FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

Disclaimer

Abstract

Poaching for elephant tusks is a major short-run threat to the African elephant with land fragmentation a threat in the longer run. Due to difficulties in distinguishing poached ivory and ivory purchased from legal sources, the Kenyan government decided not to trade in ivory confiscated from poachers. This decision was announced to the world on 18th July 1989. Kenya burned 2,000 confiscated elephant tusks to show its effort and commitment to saving the elephant from eminent extinction. This study identifies the spatial and temporal clusters of elephant poaching incidences in Kenya and the associated biophysical and human factors using geographical information systems, spatial scan statistic-SaTScan, and boosted regression trees. The spatial scan statistic detected most likely significant clusters (hotspots) for time window of 1, 6, and 12 months. Similarly, significant secondary clusters were also simulated from the analysis. More elephant poaching crimes were confirmed to be repeated next to the protected areas boundaries, at lowlands and at mean altitude of 1300 meters above sea level. Areas closer to roads and rivers contributed more to poaching cases. High income regions recorded more elephant related crimes. Regions dominated by kaolin clay soils, bush-lands, forests, plantations and grasslands are main targets of the poachers. This study provides evidence of the existence of statistically significant poaching hotspots/clusters in Kenya and also identifies the associated factors explaining such patterns. The applied methods demonstrated their relevance and applicability in analysing elephant crime data to identify hotspots.

Keywords: SaTScan, spatial and temporal clusters, boosted regression trees, most likely clusters, secondary clusters, variables.

Acknowledgements

The successful completion of my thesis work is due to a host of input and support of many people. I hereby, would want to sincerely take this opportunity to thank some of them. My sincere thanks to the Nuffic scholarship programme for awarding me the scholarship to pursue this competitive and noble graduate programme. I 'am equally heartily thankful to my course lecturers in Lund and Twente (ITC) universities for their great professionalism and commitment during the entire course.

I want to convey a deep gratitude to my primary supervisor, Dr. Tiejun Wang, for recommending this interesting research topic, the appropriate methods and also for providing the requisite guidance throughout the whole period of this research. His constant support and encouragements, I admit has resulted into a productive research outcome. I will always remember his concern of a quality presentation and orderly writing.

It is my sincere appreciation to my secondary supervisor, Prof. Andrew Skidmore, and the GEM course coordinator Dr. Michael Weir for their expert guidance and moral support. I 'am grateful to Dr.Shadrack Ngene for his help in correcting my thesis. I would like to thank, Kenya Wildlife Service for allowing me use the mortality data in writing my thesis.

To my fellow inaugural Erasmus GEM students, you have been true source of inspiration throughout this long journey. Your unwavering friendship through both the challenging and exciting periods will; indeed be memorable. I would like to acknowledge and my heartfelt appreciation to: my family and friends in Kenya for their love, encouragements and support throughout the entire period of

academic separation. Your kind hearted spirit gave me the will to accomplish my studies.

Finally, my utmost thanks to Almighty God for his eternal grace, love and mercy and without whom all these would not be surmountable.

Table of Contents

List of figures

List of tables

Chapter 1. Introduction

1.1 General background

The primary threats to biodiversity conservation in Africa are habitat loss and fragmentation as well as exploitation such as hunting and commercial trade (Grey-Ross et al., 2010). African governments, hoping to save species by protecting their habitats have established national parks, national reserves, community conservancies and sanctuaries. Close to 400 protected areas covering 1.2 million km^2 are spread across sub-Saharan Africa. Countries for instance, Kenya, Botswana, Malawi, Zimbabwe, and Tanzania, have 8% of their land-masses or even more set aside for wildlife conservation (Western, 1987).

The African elephant (*Loxodonta africana*) and the Asian elephant (*Elephas maximus*) are the surviving species in the order proboscidae (Ngene et al., 2010). Both genera originated in Sub-Saharan Africa in the early Pleistocene. The African elephant remained in Africa while Asian elephant moved into Asia during the late Pleistocene. Two sub-species of the African elephant are recognized: *Loxodonta africana cyclotis* (the forest elephant) and *Loxodonta africana africana* (the savannahh elephant) (Ngene et al., 2010).

The Convention on International Trade in Endangered Species (CITES) voted to place the African elephant (*Loxodonta africana*) on Appendix I of endangered species in the last two and half decades. This was followed with a ban on commercial trade in all elephant related products between the signatories of the treaty (Burton, 1999).

Poaching for elephant tusks is a major short-run threat to the African elephant with land fragmentation a threat in the longer run. Following intense pressure from opponents of the ban, limited export quota was allowed in 1997 enabling Botswana, Zimbabwe and Namibia to sell 50 tons of stock-piled raw ivory to Japan traders. South Africa, Botswana, Zimbabwe and Namibia proposed an annual export quotas which would allow them to export certain limited amounts of elephant tusks and hides, while Kenya and India, amongst others, were opposed to it (Heltberg, 2001).

The ban on trading in ivory at the international market was intended to reverse an acute decline in the African elephant population, as a result of the widespread poaching for ivory in the previous years. Though the continent's overall population of elephants was reported to have increased after the ban, an analysis of elephant population data from 1979 to 2007 revealed that some of the 37 states in Africa continued to lose substantial numbers of elephants. The pattern was attributed to unregulated domestic ivory markets in and near countries experiencing declines in elephant populations (Lemieux et al., 2009).

Due to difficulties in distinguishing poached ivory and ivory purchased from legal sources, the Kenyan government decided not to trade in ivory confiscated from poachers. This decision was announced to the world on 18th July 1989. Kenya burned 2,000 confiscated elephant tusks. This was to affirm the commitment to save the elephant from the eminent extinction (Lemieux et al., 2009).

Identified as a keynote species and grouped as vulnerable, elephants are under threat in most parts of Africa from poaching and human disturbance on their habitats (van Kooten, 2008). In contrast, the numbers of elephants have been reported to increase within the

confined parks of South Africa, this is associated to increased artificial water sources, man-made fences, restricting the natural movement of the elephants outside such areas and lastly, due to protection from poachers (Thomas and Minot, 2012). Confinement however, has been attributed to alteration of forests into savannah by elephants debarking and knocking trees. It also leads to inbreeding and loss off genetic diversity (Ngene et al., 2010). Human induced predation and injury majorly affect adult elephants attributing to the reduced numbers of mature elephants in the 1970's and 1980's especially in East Africa (Kyale et al., 2011).

Naturally induced mortality in large, well-established free-ranging elephant populations is age-dependent, with the youngest being vulnerable to drought. Other least common causes of elephant mortality include disease, injury, and predation by lions (Woolley et al., 2008). According to a recent study, the decline in the population of the African elephant from 1.3 million to 600,000 between 1979 and 1987 has been attributed especially to indiscriminate poaching for ivory (Maingi et al., 2012).

Despite great local and international conservation efforts, Kenya has lost some 44% of its large mammal fauna (elephants and rhinos) in the last 17 years (Norton-Griffiths, 2000). This has been blamed on a mixture of policy issues and availability of ivory market. Adult elephant mortality and human-induced injuries of elephants is closely correlated with indices of economic conditions in nomadic pastoral communities. Human mostly target adult elephant due to the large size of the tusks and associated weight (meat) (Wittemyer, 2011).

Kenya realized a population decrease of approximately 140,000 elephants in a span of 16 years (1973-1989). This resulted in a price increase of ivory in Kenya; with the price of a kilogram of

un-carved ivory worth approximately $5.50 in 1969, $75 in 1978 and $198 in 1989 (Maingi et al., 2012). Poaching of the African elephant for ivory had been on gradual increase since 1997 when the Convention on International Trade in Endangered Species allowed a one off sale of ivory by most of the Southern African states. Specifically, the recent reports produced by MIKE (Monitoring of Illegally Killed Elephants) and other conservation bodies in Kenya indicate an increase in illegal killing of elephants (Maingi et al., 2012). Quantitatively, according to a recent study conducted in Samburu-Laikipia region of Kenya, most of the areas experiencing highest numbers of poached elephants are relatively inaccessible with substantial species diversity, though not patrolled due to lack of roads and inadequate resources (finance and aircrafts). On the other hand, regions that are well managed, protected and of structured law enforcement for instance, national reserves and ranches encounter lower events of illegal killings (Kahindi et al., 2010).

1.1.1 Elephant conservation in Kenya

Kenya covers an area of 584,000 km^2, of which 7.5% is under conservation protection. These include: National Parks (NPs), where activities are limited to tourism, National Reserves (NRs) where some limited human activities apart from tourism are allowed. NPs are government owned and managed by the Kenya Wildlife Service (KWS), while the NRs are under the ownership and management of local district councils. Kenya has 21 terrestrial NPs and 23 terrestrial NRs. The other remaining areas have been proposed for conservation areas which will increase the areas under wildlife to 8% (KWS, 1990).

Conservation of wildlife in Kenya has been hindered by land use conflicts in wildlife areas and poaching amongst other factors and hence developing and managing the wildlife sector needs the

strengthening institutions and processes which involve the needs of the local societies and the wildlife to avoid the perennial land use problems (Lado, 1992).

KWS has developed elephant conservation policies to govern the wildlife management, with an aim of optimizing the returns from elephants' resources. These key policies include: (1) International trade in ivory – Kenya will continue to support the ban on commercial trade in ivory as well as cooperate with members of the treaty, (2) Poaching and illegal trade – KWS shall cooperate with other countries in gathering intelligence information concerning illegal elephant killing, (3), Monitoring status and trends – KWS will continue to continue monitor the status and trends of elephant populations, specifically those that have been identified as priority populations and not involve other stakeholders in the conservation and scientific sector as much as possible, (4) Compression and habitat destruction in small enclosed regions, (5) Prevention of crop damage – KWS will reduce damage to life and property through control shooting, (6) Stimulating tourism – Some elephant projects shall be focused in protected areas which are meant for tourism development (KWS, 2012).

1.1.2 Elephant numbers, mortality and threats

Elephant estimates are usually used to compare population and their status within the ranges in a country, regions and across the continent. Such estimates are as well vital in determining the population trends. Increases in population have been realized in: Coast, Tsavo, Southern, and Central Rift conservation areas of Kenya. In the year 2006, these regional sums composed 3%, 35%, 5%, and 12% respectively of the estimated elephant national total (i.e., 35,201). These four major regions contribute to 55% of the elephants

in Kenya. On the contrary, there are no clear trends in the elephant's population totals of Northern, Mountain, and Western regions. In 2006, the counts in these areas were as follows: 2%, 41%, and 2% respectively of the estimated national elephant numbers. These three regions accounts for 45% country elephant populations (KWS, 2012). The Tsavo ecosystem is the largest elephant sanctuary in Kenya (Ottichilo, 1987). The small population of Meru was estimated to have grown by 4.3% while those of Masai Mara have recorded an average increase of 2.4% and Samburu/Laikipia population increased by an average of 6.25%; these figures were arrived after an aerial survey between 1990 and 2006 (KWS, 2012).

Due to importance of wildlife in tourism, the wildlife authorities in Kenya have intensified surveillance and the elephant populations have increased in Samburu and Buffalo springs national reserves, though other MIKE sites in Samburu/Laikipia areas experience less patrolling hence, vulnerable to elephant criminal activities (Wittemyer et al., 2005). The decrease in elephant numbers between 1975 and 1980 was attributed to poaching and drought (Ottichilo, 1987). The elephant population is threatened by land use pressure, habitat loss, human elephant conflict, and illegal killing for meat and ivory while global warming causing unprecedented erratic climatic fluctuations is another latest victim (CITES, 2012).

KWS is charged with the responsibility of conserving and managing all the protected areas in Kenya. These also include all wildlife recourses in both NRs and private lands, this is because 70% of Kenya's large mammal species occur in both private and trust lands (KWS, 1994). Several action plans and policies have been produced in a bid to help in the conservation of elephants. These include: Law enforcement to minimize poaching, establishment of an elephant population dynamics database, investigating human wildlife

conflicts cases and implementing the relevant mitigation measures (KWS, 1991 a & b).

The recent development of long distance movements of elephants in Samburu is attributed to the change in the dominant vegetation from grassland to bushlands and decrease in the number of permanent water sources, together with an upsurge in poaching. Other factors include competition for water due to increased human population (Thouless, 1995).

According to a recent report by KWS, the number of elephants in Tsavo – Mkomazi ecosystem increased at a declining rate of about 2%. The 1988 counts showed a 75% decrease in elephant numbers within the protected areas and a further 87% in the adjacent non-protected areas since the 1972 total counts. This was attributed to two major factors: reductions in the carrying capacity of Africa for elephants, as a result of habitat change, and hunting for ivory (KWS, 2011). Despite the reductions in population sizes, the Tsavo ecosystem is habitat to Kenya's largest population with a population of 35,000 animals in 1974 and about 11,733 in 2008. The elephant population in Samburu-Laikipia-Marsabit declined by 14 % between 2008 and 2012 (KWS, 2011).

The other threats to elephant survival across Africa include: land use pressure, habitat loss, human elephant conflict, and illegal killing for both meat and ivory (CITES, 2012). The increase in levels of illegal killing has encompassed not only small and fragmented elephant populations that are faced with eminent extirpation; but also the previously secure large populations (Beyers et al., 2011).

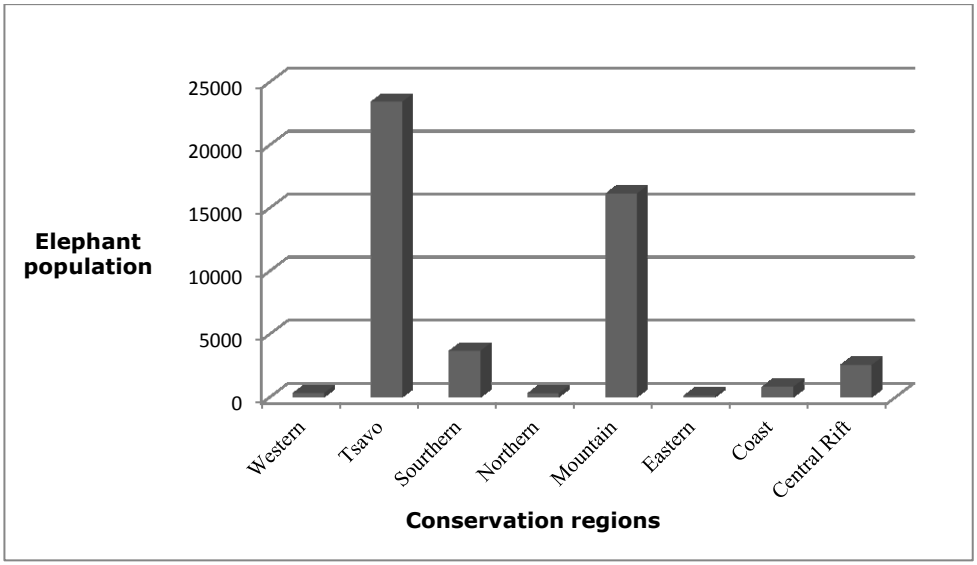


Figure 1: Elephant population and conservation regions in Kenya, 1997 to 2010. Source: (KWS, 2002)

Figure 2: Herd of elephants in Amboseli National Reserve. Photographed by Dancan Ouko

Figure 3: Kenyan government burning confiscated ivory from poachers

1.2 Problem statement

There is inadequate updated and precise information on the status of African elephants and the poaching intensity after the CITES ban went into effect from 1989 (Lemieux et al., 2009). Though, there are special elephant monitoring systems being developed today, no results has been achieved yet. Currently, there is no evidence which associates changes in poaching with CITES decision to uplift trade in ivory. Ivory trade has remained active during the ban period in a number of South-East Asia countries (Heltberg, 2001).

The knowledge on distribution of biodiversity and the threats that face them makes it easier to assimilate the general ecological requirements and security measures needed to safeguard their populations. This provides an opportunity to assess disturbances that hold the species away from a region and thus help design appropriate conservation measures. "In Northern Kenya there is a very positive energy between multiple different stakeholders working towards the same conservation goals and this is already showing dividends in the

increase in game populations in the newly formed conservation areas" (Gross, 2008).

Poaching disrupts elephants' social relations. Evidence suggests that an elimination of kin as a social partner has negative consequences on some elephants. Despite the free ranging nature of elephant societies, elephants are capable of maintaining ties with kin, especially in populations not heavily affected by poaching (Archie et al., 2012).

Despite the ecological significance of elephants in Kenyan ecosystem, most risk areas of this species in relation to anthropogenic factors, management practices and biophysical factors are not yet identified. This key issue has not received adequate attention in most of the existing studies and surveys. The identification of conflict hotspots has consequences on academic and practical level: for instance, understanding the reason why conflicts are clustered in a certain area, as well as location specific tools concerning relationships. Also, its usefulness is realized by the apparent increase in conflict related to natural resources including wildlife management (Mola-Yudego et al., 2010). Anti-poaching patrols are particularly challenging in Kenya due to limited resources, and the large areas of the parks which limits the effectiveness of patrols by park rangers (Maingi et al., 2012).

The results from space-time cluster analysis would be valuable for KWS in making sure that both financial and human resources are allocated as effectively as possible, at the right places at the right times. The information will be core in forming a basis for decision making in conservation and provide a basis for policy and decision making, advocacy and awareness creation. These would integrate future national and regional management programs in order to minimize further human induced deaths.

1.3 Research objectives

The main objective of this study was to determine whether the observed patterns in elephant poaching incidences are simply random or clustered in space and time. The study also aimed at identifying factors influencing space-time elephant poaching patterns such as biophysical and anthropogenic factors between 2002 and 2012. The specific objectives were:

- To examine space-time patterns of elephant poaching incidences in Kenya.
- To identify the biophysical and anthropogenic factors which contribute to the observed elephant poaching patterns in Kenya.

1.4 Research questions

- What are the space-time patterns of elephant poaching in Kenya?
- What predictor variables determine the observed patterns of elephant poaching incidences in Kenya?

1.5 Research hypothesis

- **H_o:** Space-time patterns of elephant poaching incidences in Kenya are random.
- **H_1:** Space-time patterns of elephant poaching incidences in Kenya are non-random.

1.6 Outline of the thesis and research approach

This thesis is organized into five chapters: Introduction, materials and methods, results, discussion, conclusion and recommendations.

Chapter 1 presents the background study of African elephants mainly focusing on their conservation status, population size over-time and the threats to their survival. Subsequently, the study area, the research objectives, questions, and hypotheses are highlighted.

Chapter 2 organizes and presents the research data (i.e., the elephant data and anthropogenic and, the environmental modeling data) and preprocessing procedures involved. Quantitatively, the methods used to model the space-time clustering of elephant poaching incidences and their association with biophysical and human factors that explain the patterns of the hotspots.

Chapter 3 explains the results by validating and analysing the results produced by SaTScan space-time permutation model specifying the location, the time frame and spatial extents of the clusters. The most important interaction factors explaining the spatial temporal elephant poaching events are presented including the maps and partial dependence plots.

Chapter 4 discusses in detail the significance of the whole approach to model the elephant poaching events through the use of scan statistic and boosted regression trees. Besides, the possible factors that contribute to repeat elephant poaching events and the specific areas prone to such incidences are identified. The strengths and weaknesses of such methods are elucidated as well.

Chapter 5 summarizes the general overview of the results. The management modalities and strategies required to curb the elephant crimes in future are also recommended.

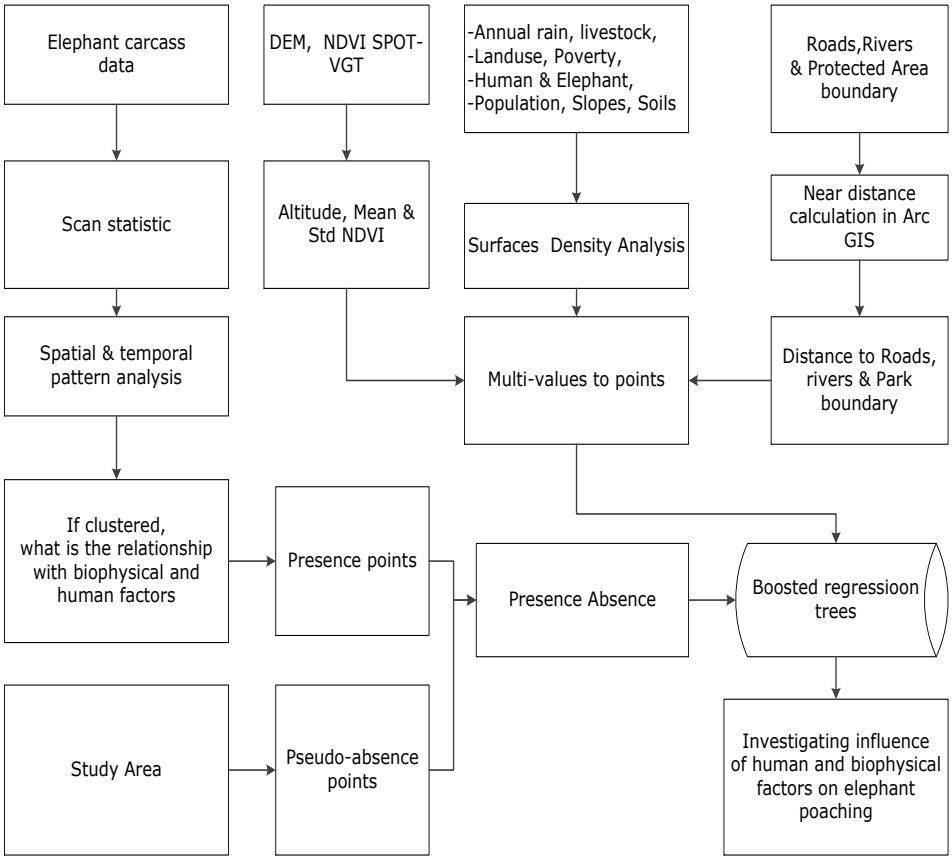


Figure 4: Research framework

Chapter 2. Methods and materials

2.1 Study area

The republic of Kenya lies along the equator in East Africa and is bounded by 5°30' N and 4° 30' S latitude and 34° E and 42° E longitude. It covers 582, 646 km^2. It is composed of four major relief zones: coastal and eastern plains, the central and western highlands, the Rift Valley Basin and the lake Victoria Basin. The country shows a wide range of natural regions, varying from hot arid lowlands; with various soils types. The altitude gradually increases from 0 m above mean sea level near the Indian Ocean to between 2000 m and 3400 m in the highlands. Kenya has several mountain ridges with elevations 3000 m, including Mount Elgon (4,375 m) and Mount Kenya (5,199 m). Many regions of Kenya experience wet seasons from March through May and the short rains from October to November. The dry seasons extend from January to February and from June to September in most years (Batjes, 2004).

The mean annual air temperature is highly related to elevation. It decreases from about 27° C near the sea level, to 17° C in Nairobi in the central highlands, to less than 10° C above 3000 m. The average annual rainfall ranges from 150 to 500 mm in the arid east and northeast of the country, from 500 to 1000 mm in the semi-arid regions and 1000 to 2500 mm in the more humid areas in the central highlands and near Lake Victoria. Kenya is divided into seven agro-climatic zones based on the ratio of annual rainfall over average potential evaporation (r/E_o). This varies from < 0.15 in the very arid regions up to > 0.8 in the humid zones (Batjes, 2004). Kenya has thirteen National Parks and twenty-five reserves that occupy ten percent of the country (Burnett et al., 1990). The main areas of

contiguous elephant range are: the Northern coast, the Tsavo-Chyulu-Amboseli-Kilimanjaro complex, the Aberdare-Mt.Kenya-Laikipia-Samburu-Northern Area complex, the Nguruman-Mara-Serengeti complex and Nasolot-Romoi-Kerio Valley (KWS, 2012).

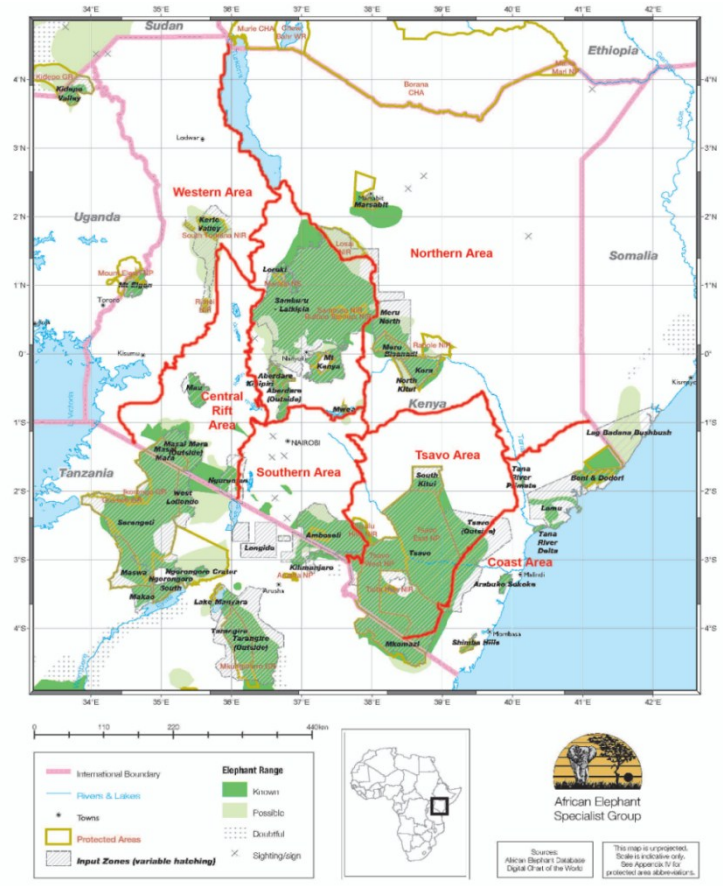


Figure 5: A Map showing the range of elephants in Kenya. Source: (Blanc et al., 2007)

2.2 Elephant poaching data

Elephant-mortality data covered the period from January 2002 to August 2012. The data set included, geographic coordinates of poaching events, names of the locations where the carcasses were

found, date and cause of the elephant mortality. The excel spreadsheet from KWS contained a total of 5,052 elephant mortality incidences due to: natural deaths (1,515), poaching (1,612), problem animal control (261), accidents (97), unknown deaths (1,045) and as a result of human wildlife conflicts (522).

Some of the historic data-sets were collected in military grids and hence were converted to decimals degrees through the use of 1:250,000 geo-referenced topographic maps obtained from Department of Surveys, Kenya. The points that lacked geographic coordinates and those which fell out of Kenyan boundaries were excluded from further analyses. The geographical coordinates collected in UTM (Universal Transverse Mercator) were converted to decimal degrees for harmonization of all data-sets into a single coordinate system. A total of 1,006 poaching events were extracted from the main excel-sheet. To facilitate iteration in SaTScan statistic software, the data was categorized into date, region, and reason for the elephant mortality and finally into coordinates of the points where carcasses were found.

Poaching activities in Kenya occurs in most of the conservation regions including: Tsavo, Mountain, Central Rift, Coast, Southern, Northern, Western and Eastern. With Mt.Kenya-Laikipia-Samburu and Tsavo conservation areas accounting for the highest concentration of the poaching related elephant crimes. The poaching crime occurs in both protected and un-protected conservation areas (figure 6 below).

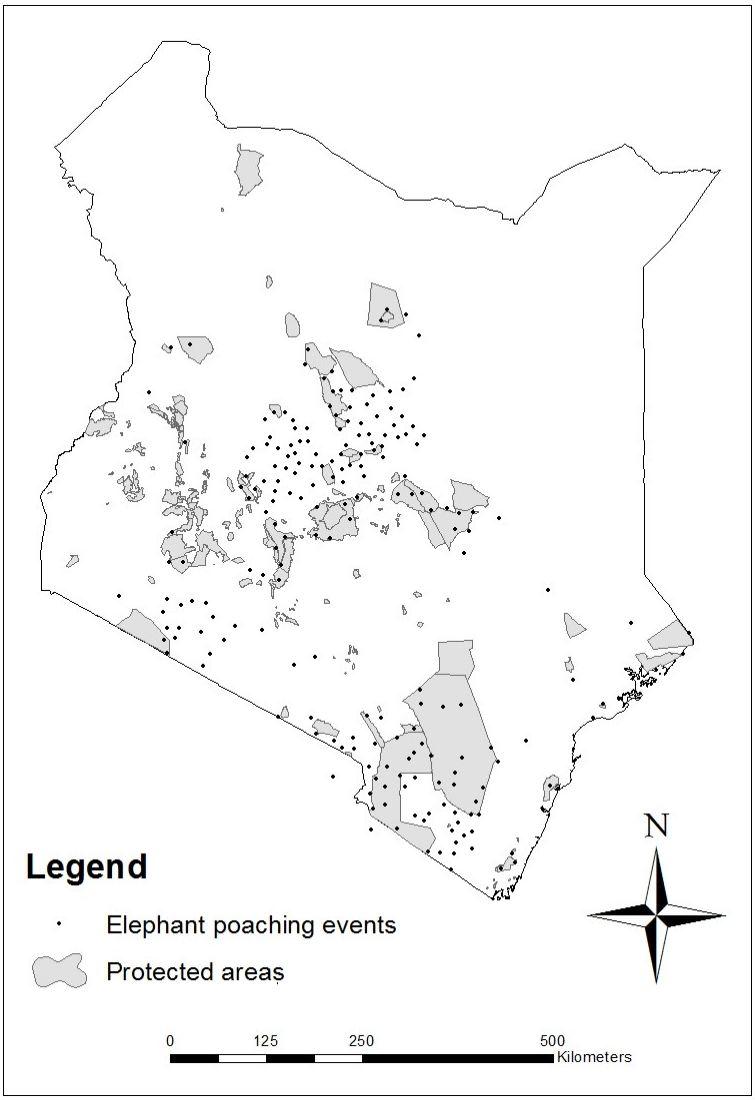


Figure 6: Decadal distribution of elephant poaching incidences in Kenya (2002 - 2012). Source: (KWS security database, 2002).

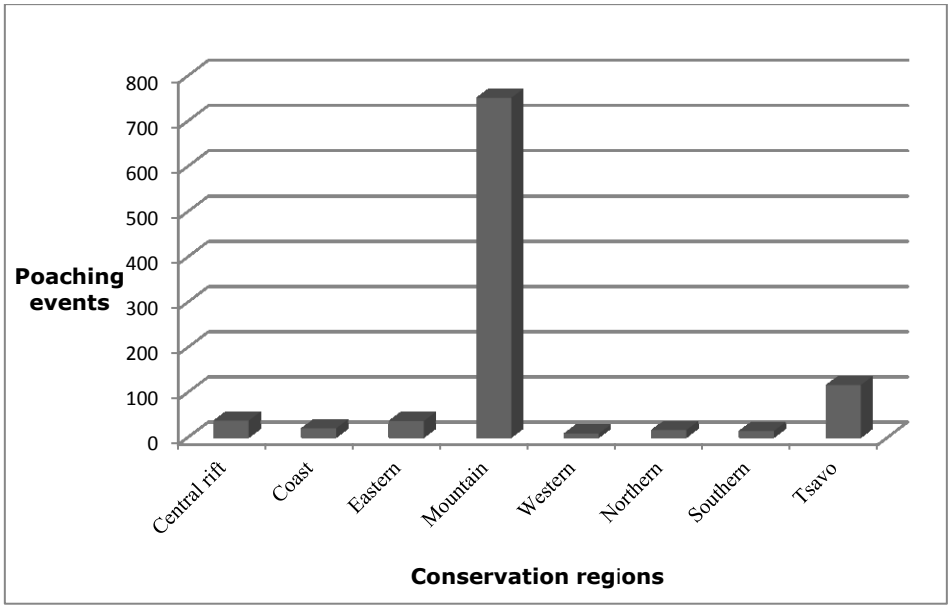


Figure 7: The number of poaching events per conservation areas (2002 - 2012). Source: (KWS security database, 2002).

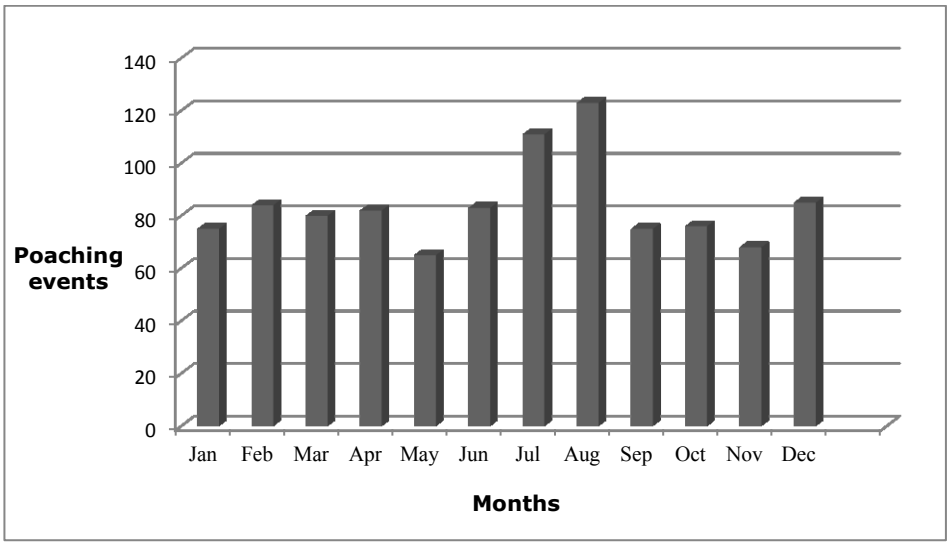


Figure 8: Monthly poaching trends (2002 - 2012). Source: (KWS security database, 2002).

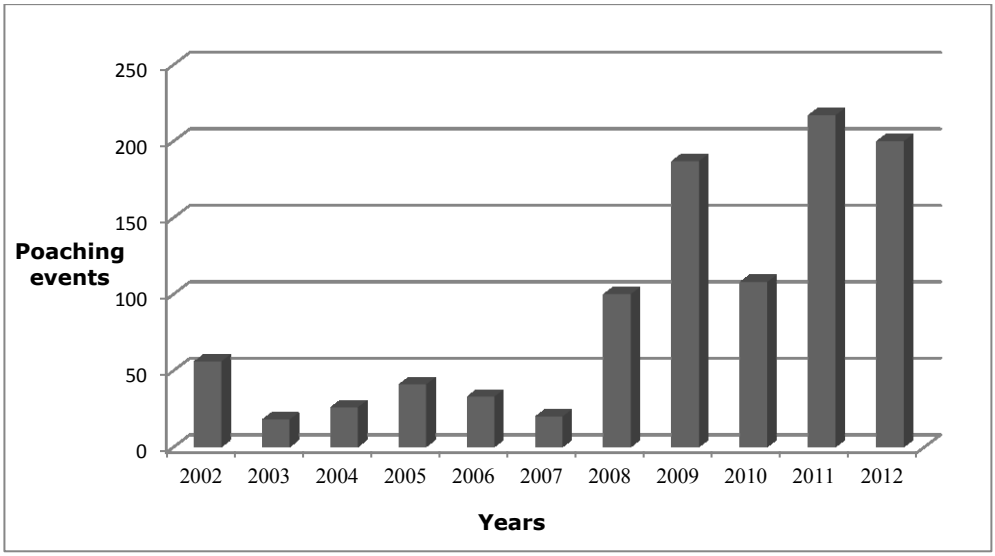


Figure 9: Yearly elephant poaching cases (2002 - 2012). Source: (KWS security database, 2002)

2.3 Biophysical and anthropogenic factors

The association between elephant poaching patterns and sets of anthropogenic (human factors) and biophysical (environmental) factors were determined through boosted regression trees in R-Studio. Another factor is functional - ecological factors such as competition and predation. Extensive prior knowledge of the study area was available from previous studies. Hence the prior information informed the basis of choice of the input variables that were used in the subsequent simulations in R-Studio. The description of each layer is as follows: Distance to roads: Euclidian distance in kilometres to the main and secondary roads, (2) Distance to towns: Towns were represented by large and small village towns, the distance to towns raster was created by calculating the Euclidian distance in kilometres to each town, (3) Distance to rivers: Euclidian distance in kilometres to the primary and secondary rivers, (4) Distance to protected area boundaries: Represented by the Euclidian distance in kilometres to

the boundaries of protected areas, (5) Altitude extracted from the 90 m DEM, (6) Mean annual precipitation derived from ILRI website, (7) Soil types and slopes (%): Obtained from the Soil and Terrain Database for Kenya (KENSOTER), at a scale of 1:1,000,000 compiled by Kenya soil survey, (8) Poverty density: Represented the density of poor people per square kilometre, (9) Land use: Coverage showing general land use classes derived from Africover Kenya Multipurpose Landcover Database full resolution (FAO-Africover, 2003), through re-classification into vegetation types, (10) Livestock density representing the number of domesticated animals per square kilometre, (11) Mean NDVI and (12) Standard deviation NDVI both derived from SPOT-VGT as a stack of 361 images of 10 days' time series temporal resolution. In order to reduce potential noise of cloudiness but also keep the high fidelity of the data, SPOT-VGT were cleaned and smoothened using an adaptive Savitzky-Golay filter in TIME-SAT program (Per et al.), (13) Elephant population, and finally, (14) Population density representing the total number of elephants and people in a square kilometre respectively. Euclidean distances were calculated using ArcGIS Spatial Analyst Tools. The polygons of poverty index, land use, soil types, livestock density, population density and slope were rasterised through polygon to raster function in ArcGIS Spatial Analyst, creating uniform density surfaces; with the input field determining the type of output raster.

The relationship between elephant poaching events and the explanatory variables (biophysical and human factors) was examined through boosted regression trees. 205 presence points' points were randomly generated within the 17 one month time precision clusters generated by SaTScan program. Similarly, an equal number of random pseudo-absence points (points assumed not to experience the occurrences of poaching cases within the study area); were

randomly generated all over Kenya. The two shape-files were merged, and subsequently used to extract multi-values to points from all the density surfaces of the environmental and human variables. The Spatial Analyst Tool in Arc GIS 10.1 was used in the generation and extraction of values to points as mentioned above.

Table 1: Source, data type and units of the independent variables

Variable	Source	Unit/Data types
DEM (90 m)	(SRTM) [a]	Meters
Altitude/Elevations	(SRTM) [a]	Meters
Slope	(KENSOTER scale 1:1,000,000) [b]	%
Landcover	(FAO-Africover, 2003) [c]	Categorical
NDVI SPOT-VGT	(SPOT-VGT) [d]	Categorical
Soil Types	(KENSOTER scale 1:1,000,000) [c]	Categorical
Park boundaries, rivers, Roads, towns, human population density, Poverty density, elephant population density	(WRI & ILRI) [e]	Shape files
Elephant poaching data	(KWS) [f]	Lat/Long

[a] SRTM- Shuttle Radar Topography Mission.
[b] KENSOTER - Soil and Terrain Database for Kenya.
[c] FAO - Food and Agricultural Organization.
[c] KENSOTER - Soil and Terrain Database for Kenya.
[d] NDVI SPOT-VGT - Normalized vegetation index - Spot Vegetation.
[e] WRI & ILRI - World Resources Institute and International Livestock Research Institute.
[f] KWS - Kenya Wildlife Service.

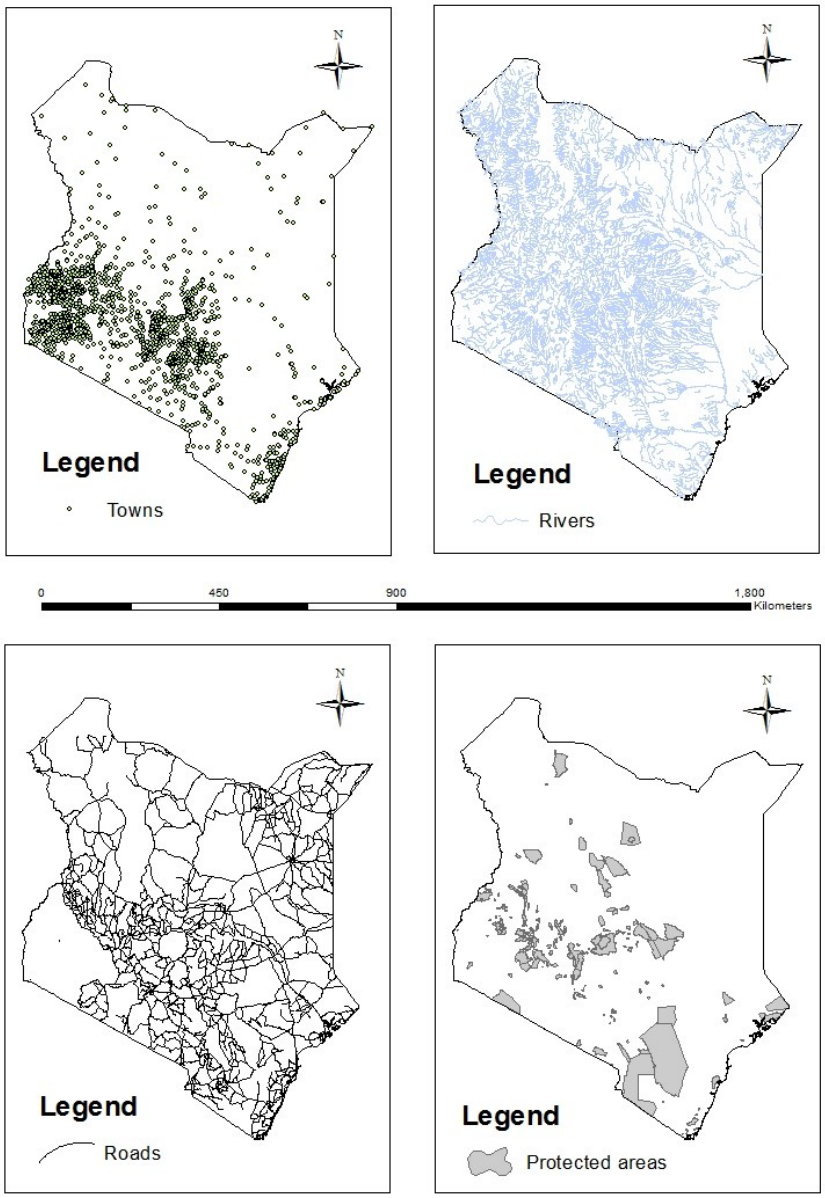


Figure 10: Distribution of towns, rivers, roads, and protected areas in Kenya

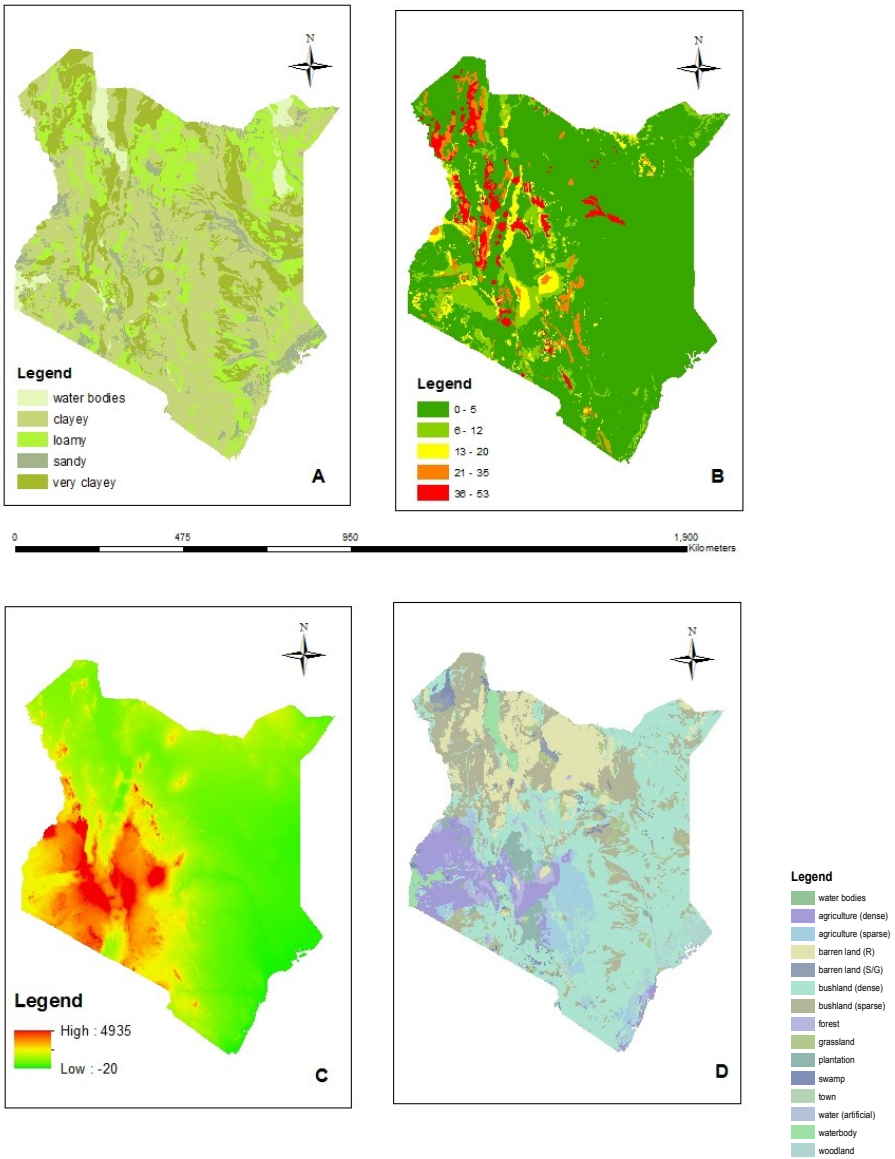


Figure 11: Soil types (A), % slopes (B), elevations (C), and land use (D).

2.4 Multi-collinearity and variance inflation factor tests

The parameter estimates for most spatial modeling are not strongly biased, only in the situations of autocovariate models. In the implementation of such models, the predictor variables are consistently underestimated (Dormann, 2009). The author continues to explain that autocovariate approaches in logistic regression models applied for binomially occurring data would be biased and unreliable.

Also known as co-dependence, multi-collinearity occurs in a multiple regression when many predictors (regressors) are highly correlated hence the inflation of regression parameter estimates for instance variance (Fox, 1997). Variance Inflation Factor is a common way used to detect multi-collinearity (Montgomery et al., 1982), and is denoted by the following mathematical expression;

$$VIF=\sqrt{1/(1-R2)} \quad (1)$$

VIF represent the inflation that each regression coefficient experiences if the correlation matrix were an identity matrix i.e. if multi-collinearity was not present in the data (Owen, 1988). The correlation between the variables might lead over-representation of the response variable (i.e., poaching events). The explanatory variables were tested for multi-collinearity in R Studio prior to performing boosted regression trees. There are various rules of thumb regarding the variance inflation factors cut off values, the rule of 4, and the rule of 10 amongst others. When VIF exceeds these set values, these rules often are interpreted as casting doubts on the results of the regression analysis. High values of VIF leads to inflation of standard errors of coefficients of variables (Craney et al., 2002).

All the predictor variables had VIF of less than 4 and 10 respectively compared to the cut off rules (Table 2). Hence the exclusion of independent variables (predictor) was based on expert knowledge and the values of correlation coefficients (i.e., a pairwise correlation more than 0.5 or less than – 0.5 was a concern). The annual rainfall was correlated to altitude; mean NDVI was correlated to land use types and population density was correlated to poverty density. Hence annual rainfall, mean NDVI and population density were not used in fitting the model in boosted regression trees.

Table 2: Multi-collinearity assessment

Variables	**VIF's**	**R square**
Altitude	2.30	0.56
Annual rainfall	3.40	0.67
Distance from roads	1.15	0.13
Distance from rivers	1.11	0.01
Distance from park boundaries	1.24	0.19
Soil types	1.06	0.058
Slope	1.27	0.22
Poverty density	1.70	0.41
Land use	1.87	0.46
Population density	1.85	0.46
Livestock density	1.70	0.41
Mean NDVI SPOT VGT	3.10	0.67
STD NDVI SPOT VGT	1.25	0.10
Elephant population	1.02	0.02

2.5 Space-time statistic analysis

Kulldorff's scan statistic is a spatial scan statistics method for detecting and evaluating statistically-significant spatial clusters (e.g. disease, crime amongst others). This method and its associated software implementation – SaTScan; is used widely in a wide array of fields for instance epidemiology and other research fields (Chen et al., 2008). The space-time scan statistic (Kulldorff, 2010) was used in searching, testing for significance and locating approximate locations of space-time clusters. The search is done using cylindrical moving windows of variable sizes which moved in both space and time across the study area. Space-time permutation scan statistic involves the probability model since the population at risk data is not known, the expected values are calculated using cases only (Kulldorff et al., 2005).

The space-time permutation model automatically adjusts for both purely spatial and purely temporal clusters. Hence there are no purely temporal or purely spatial versions of this model. Space –time permutation model is only used when only case data is available, and when one wants to adjust for purely spatial and purely temporal clusters (Kulldorff, 2010).

The spatial dimension is represented by the circular base of the window, this varied from zero up until a specified maximum value allowing the inclusion of 50% of the total number of incidences in the study region - representing the geographical area of the potential poaching events. The height of the moving cylinder reflected the time period of potential clusters, up to 50% of the research period with a time precision of 1, 6 and 12 months respectively (Kulldorff, 2010).

The space-time analysis was conducted using a maximum spatial cluster size of 50% of the population at risk because a larger

cluster size would indicate areas of extremely low rates outside the circle rather than an area of exceptionally high-rates. To facilitate arriving at core clusters and avoid likely misleading clusters, it is of great importance to avoid the selection of an excessive maximum-size value (Chen et al., 2008). The selection of the maximum spatial cluster size of 50% of the population at risk as described in the SaTScan User Guide is meant to avoid pre-selection bias.

The maximum size parameter sets an upper bound on the circle radius in one the following two ways: (1) by determining the maximum percentage of the total population at risk within a circle or (2) by specifying the geographic extent of the circle (Chen et al., 2008). The maximum size parameter (maximum radius) was set upon a circle with 100 km and 80 km radius representing dry and wet seasons migratory routes of elephants in Kenya (Thouless, 1995). The program scans for clusters of geographic size between zero and some limit defined by the user (i.e., in this case 100 km as the upper limit).

With an assumption that within each window the incidences follow a binomial distribution, space-time clustering was examined by comparing the proportion of observed cases in a cluster to what would have been expected if the spatial and temporal locations of all events were randomly distributed in space and time (i.e., so that there exist no space-time interaction). The null hypothesis is that, the number of poaching events is the same all over the study region and the alternative hypothesis is that the proportion of poaching incidences within the cylinder is higher than outside the cylinder (Abatih et al., 2009). SaTScan program uses computer simulations to generate a number of random replications of the data set under the null hypothesis. If the maximum likelihood ratio calculated for the most likely cluster in the real data set is high compared to maximum

likelihood ratio calculated from the most likely clusters in the random data sets that is evidence against the null hypothesis and for existence of clusters.

There is a cluster in geographical area if, during a specific time period, that area has a higher proportion of its cases in that time period compared to the remaining geographical areas. For space-time analyses, case and coordinate data files were used as inputs in SaTScan. The space-time scan statistic can be applied for either a single retrospective analysis, using historic data, or for time-periodic prospective surveillance, where the analysis is repeated every day, week, month and year (Kulldorff, 2010).

A retrospective space-time based model was used, where the number of events in an area is Poisson distributed according to a known underlying population at risk. Retrospective analysis scans for both historic and active space-time clusters. For criteria of reporting secondary clusters No Geographical Overlap – Secondary clusters will only be reported if they do not overlap with a previously reported cluster and they may not have any location IDs in common. Only the general location and size of a cluster was considered not its exact boundaries. Hence no overlapping clusters will be reported, presenting the fewest and distinct numbers of clusters (Kulldorff, 2010).

SaTScan detects potential clusters by calculating a likelihood ratio for each circle; which is proportional to the Equation 2:

$$\left(\frac{c}{e}\right)\left(\frac{C-c}{C-e}\right)^{C-c} I\ () \qquad (2)$$

Figure 12: Illustrations of how space-time permutation model functions in SaTScan

Where C is the total number of cases, c is the observed number of cases within a circle and e is the adjusted expected number of cases within the circle. I () is a binary indicator that facilitates the identification of high-risk clusters (hot spots), low risk clusters (cold spots), and both. If SaTScan is set to scan for high-risk clusters, I ( ) is equal to "1" when c > e and equal to "0" otherwise; for low-risk clusters, the ">" would change to "<"; and for both, I () = 1. The circle with the maximum likelihood ratio among all radius sizes at all likely point locations is regarded as the most likely cluster (the primary cluster). SaTScan also identifies secondary clusters which have significantly large likelihood ratio but are not the most likely clusters (Chen et al., 2008).

Several secondary clusters are more similar to primary clusters in geographic position and extent; they are used as estimates of location and sizes of detected clusters. The secondary clusters mainly occur due to slight alteration to the circle radius or relocation of the circle to a different nearby point location changes the likelihood ratio slightly, mainly when the newly included or removed locations have a small population at risk (Chen et al., 2008).

The significance of identified space-time cluster is tested through the likelihood ratio test statistic and p-values of test are

obtained through Monte Carlo simulations. The p-value is given by R/(#SIM + 1) where R is the rank of the test statistic from real data among all data sets and #SIM the number of simulated data sets. To achieve excellent power for all datasets 999 simulations were used (Abatih et al., 2009). The statistical significance of the secondary clusters was determined by comparing and ranking its likelihood ratio value with the Monte Carlo maximum likelihood ratios. This test procedure is deemed conservative in that, a secondary cluster from the data set is compared with most likely cluster from the simulations (Kulldorff, 2010).

SaTScan will evaluate very small and very large clusters, and everything in-between. For all the analyses, the most likely and secondary clusters with statistical significance of $p < 0.05$ were considered based on comparing the size of the log likelihood ratio against a null distribution obtained from Monte Carlo 999 replications. SaTScan does not assume that there is no spatial auto-correlation in the data. It tests whether there is spatial auto-correlation or other divergences from the null hypothesis (Kulldorff, 2010).

2.6 Species distribution models

Species distribution models (SDMs) are numerical tools which involves observations of species occurrence or abundance with environmental estimates. They are applied in ecological and evolutionary studies to gain insights and to predict distributions across landscapes, in some cases requiring extrapolation in space and time within marine, terrestrial and freshwater ecosystems (Leathwick, 2009).

For the last two decades, many types of statistical models have been used in ecological modeling. Though, the earlier linear regression models were very simplistic in analyzing real life

situations. In recent time, generalized linear models and generalized additive models have increased the capacity to analyze data with non-normally distributed errors (presence-absence and count data), and to model nonlinear relationships (Creek, 1990). In addition, a wide variety of algorithms have been applied in ecological predictions, for example, neural network, decision trees and support vector machines – the machine learning methods. These are less often used in ecology than regression methods; this is attributed to their complexity and hence liable to critics (Praagman, 1985).

In most cases models are used to detect and describe patterns, or to predict to new situations. Regression models are usually used as tools for quantifying the relationship between one variable and others on which it depends. Models can be used to identify factors with the most explanatory power, indicate optimal conditions and predict to new cases. For instance, in analyzing vegetation type in relation to aspect, rainfall, and soil nutrients (Elith et al., 2008).

We demonstrate the use of BRT using data describing the distribution of, and environments occupied by, the carcass of poached African elephants in Kenya. The data used in BRT does not need prior transformation or elimination of outliers, BRT can fit complex nonlinear relationships, and automatically performs interaction effects between predictors (Elith et al., 2008).

The produced model identifies major environmental and human factors of elephant poaching cases. The model is a form of logistic regression modeling the probability that poaching occurs, y= 1, at a point with covariates **X**, P(y= 1|**X**).The probability is modeled via a logit:

$$\text{logit } P(y=1|\mathbf{X}) = f(\mathbf{X}) \quad \text{.....(3)}$$

A boosted regression trees (BRT) is a technique which aims to improve performance of a single model by fitting many models and combining them for prediction. It applies two algorithms: regression trees are from the classification and regression tree (decision) group of models, and boosting builds and combines a group of models. It is an additive regression model in which individual terms are simple trees, fitted in a forward, stage-wise fashion. The process is stage-wise (stepwise), meaning that existing trees are left unchanged as the model is enlarged, only the fitted value for each observation is re-calculated at every step to reflect the contribution of the newly added tree. The fitted values in outcome model are derived as the sum of all trees multiplied by the learning rate, which are more stable and accurate than those from a single decision tree model (Elith et al., 2008).

The model was fitted in R version 2.15.2, using gbm package version 1.5-7 (Ridgeway, 2006) plus custom code written by J.L and J.E (Elith et al., 2008). Generally, BRT regularization involves jointly optimizing the number of trees (nt), learning rate (lr), and tree complexity (Elith et al., 2008). A bag fraction was used to introduce randomness in the model hence reducing over-fitting of the model to the data, hence 0.5 was used, meaning that at each iteration, 50% of the of the data are drawn randomly, without replacement from the training data-set. Tree complexity of 5 was used to determine how many splits for each tree, more complex trees reduces error in predictive deviance. A slower learning rate of 0.005 was used to achieve at-least 1000 trees. Cross validation was applied in portioning the data into test and training sets. Data was divided to n-fold (n splits) .The model runs n-times, each time one fold is used as

test and the other (n-1) folds are used as training data. Therefore, at the end, all n folds are used as test data.

Chapter 3. Results

3.1 Space-time patterns of elephant poaching in Kenya

Using a spatial cluster size of 50% of the population with a circle radius of 100 km and minimum temporal cluster size of 50% with time precision of 1, 6, and 12 months. The most likely statistically significant clusters consisted of 14 repeat elephant poaching incidences with 11 observed cases compared to 0.195 expected cases from 1st of April 2008 to 30th April 2008 (Radius, 6.97 km), 40 repeat events with 18 observed cases compared to 2.17 expected cases from 1st of August 2008 to 28th of February 2009 (Radius, 15.84 km) and 43 coincidence poaching events with 15 observed cases compared to 1.19 expected cases from 1st of January 2004 to 31st of December 2004 (Radius, 14.45 km) respectively (Table 3). Each cluster had a specific time period of poaching and individual extent (i.e., the radius). All primary clusters are spatially and temporally different; though marginally different in size (Figures 14 and 15). Another 16, 9 and 7 statistically significant secondary clusters were identified for time precision of 1, 6 and 12 months respectively, each of which occurring at differing time frames and locations (Table 3 and Figure 15 **A**, **B** and **C**).

Most of the space-time clusters appeared at the mountain conservation region (Samburu, Isiolo, and Laikipia) and in Tsavo – Coast conservation areas. The numbers of clusters reduce as time precision increases. The Tsavo ecosystem experienced a consistent

cluster both in size (approximately 90 km radius) and location. The most likely hotspots (i.e., that least occur by chance) were found to occur in the Mountain conservation area (Figure 15 **A**, **B**, and **C**).

Table 3: Space-time elephant poaching incidences in Kenya using maximum spatial cluster of 50% of the cases, 100 km circle radius at varying temporal windows

Clusters No.	No of PEVs	Radius (Km)	No. Obs.	No of Expe.	Time – Frame
1 month					
1*	14	7	11	0.19	2008/4/1-2008/4/30
2	3	19	7	0.5	2008/12/1-2008/12/31
3	6	1	8	0.11	2008/11/1-2008/11/30
4	10	15	7	0.15	2009/4/1-2009/4/30
5	6	10	5	0.042	2002/9/1-2002/9/30
6	11	15	7	0.27	2012/8/1-2012/8/31
7	11	19	6	0.19	2008/8/1-2008/8/31
8	25	11	6	0.20	2006/10/1-2006/10/31
9	2	0.1	5	0.12	2011/7/1-2011/7/31
10	7	6	4	0.06	2002/4/1-2002/4/30
11	71	8	12	1.72	2012/7/1-2012/7/31
12	4	2	5	0.14	2012/1/1-2012/1/31
13	7	17	4	0.06	2010/6/1-2010/6/30
14	4	5	12	0.08	2011/11/1-2011/11/30
15	7	7	5	0.18	2011/1/1-2011/1/31
16	10	12	4	0.09	2008/7/1-2008/7/31
17	7	11	4	0.09	2010/12/1-2010/12/31
6 months					
1*	40	15	18	2.12	2008/9/1-2009/2/28
2	15	8	12	0.82	2008/3/1-2008/8/31
3	25	11	11	0.65	2006/9/1-2007/2/28
4	44	26	12	0.93	2004/9/1-2005/2/28
5	34	31	16	2.02	2008/3/1-2008/8/31
6	122	91	44	17	2012/3/1-2012/8/31
7	6	10	5	0.09	2002/9/1-2003/2/28
8	37	14	10	1	2005/3/1-2005/8/31

9	20	56	10	1.5	2010/9/1-2011/2/28
10	22	50	13	2.7	2011/3/1-2011/8/31
12 months					
1*	43	14.41	15	1.2	2004/1/1-2004/12/31
2	25	10.54	13	0.9	2006/1/1-2006/12/31
3	40	58.86	24	4.7	2008/1/1-2008/12/31
4	91	92.86	48	17.3	2009/1/1- 2009/12/31
5	16	7.24	13	1.6	2008/1/1-2008/12/31
6	126	91.3	55	25.25	2012/1/1-2012/12/31
7	37	13.46	11	1.51	2005/1/1-2005/12/31
8	22	50.27	17	4.4	2012/1/1-2012/12/31

Clusters numbers, No.of PEV's, number of poaching events, Radius; the extent of cluster in km, No. Obs, number of observed cases, No. Expec; number of expected cases,Time-Frame, time period of cluster occurrence.

*Most likely cluster.

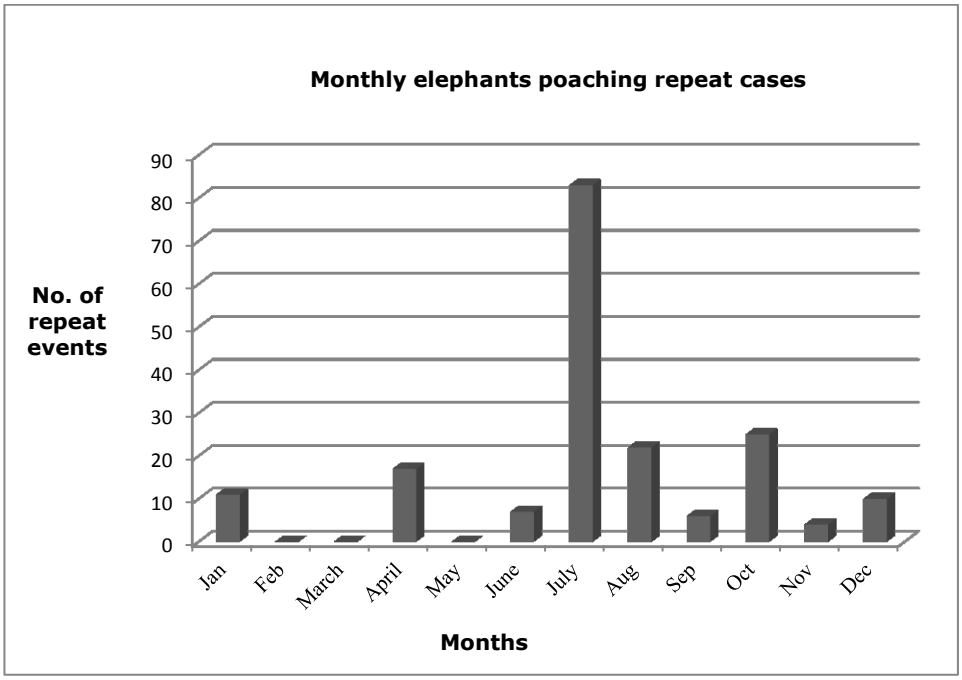


Figure 13: Monthly elephant poaching repeat events.

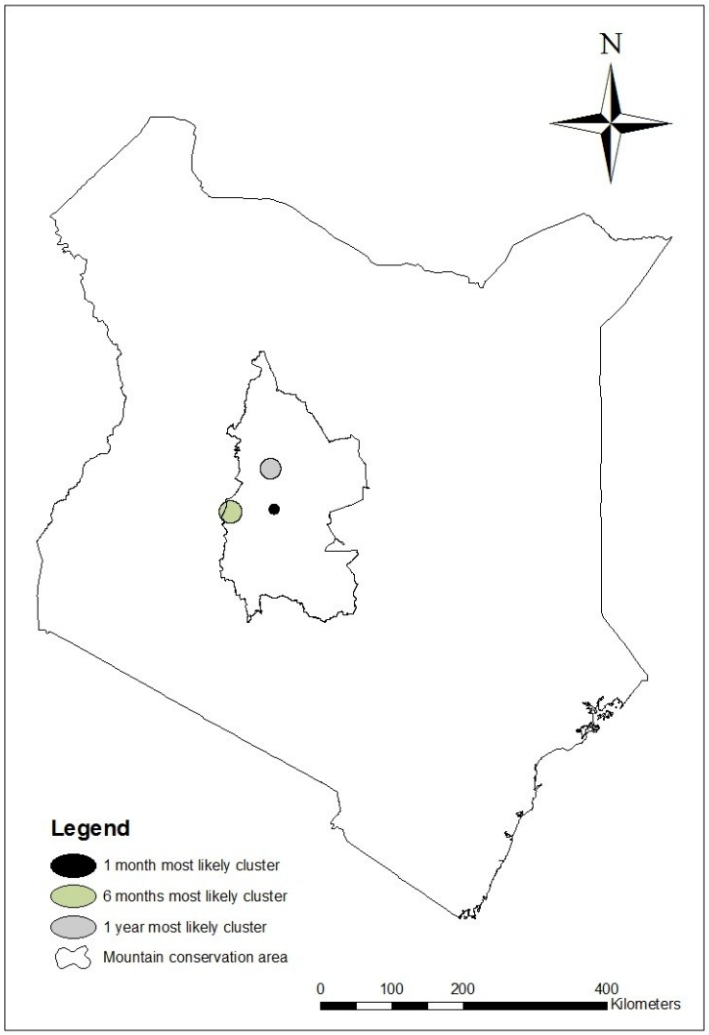


Figure 14: A Map showing the locations of primary hotspots for 1 month, 6 months and 1 year time windows

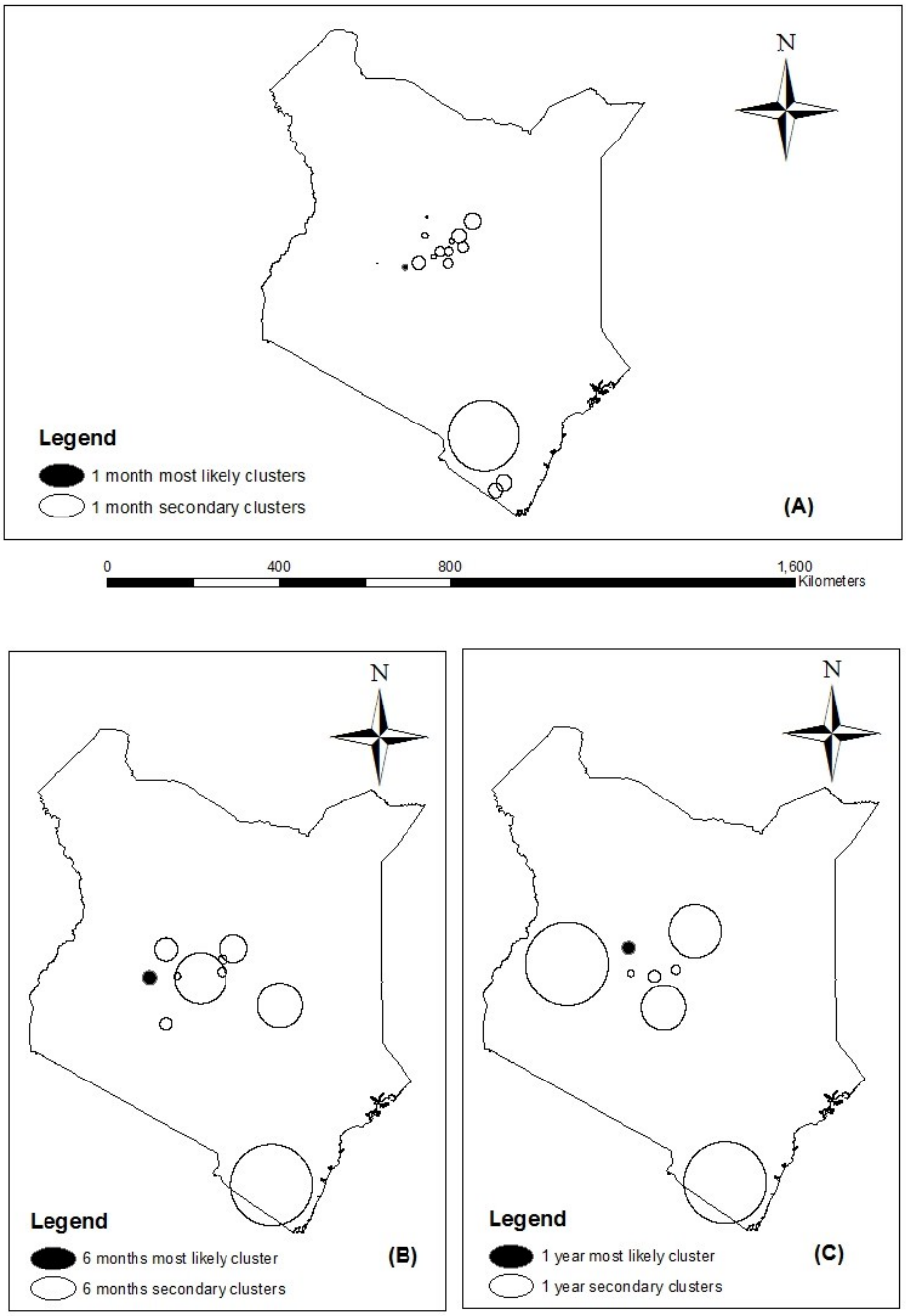


Figure 15: Mostly likely and secondary clusters: (A), (B), (C) for 1, 6 and 12 months temporal windows respectively

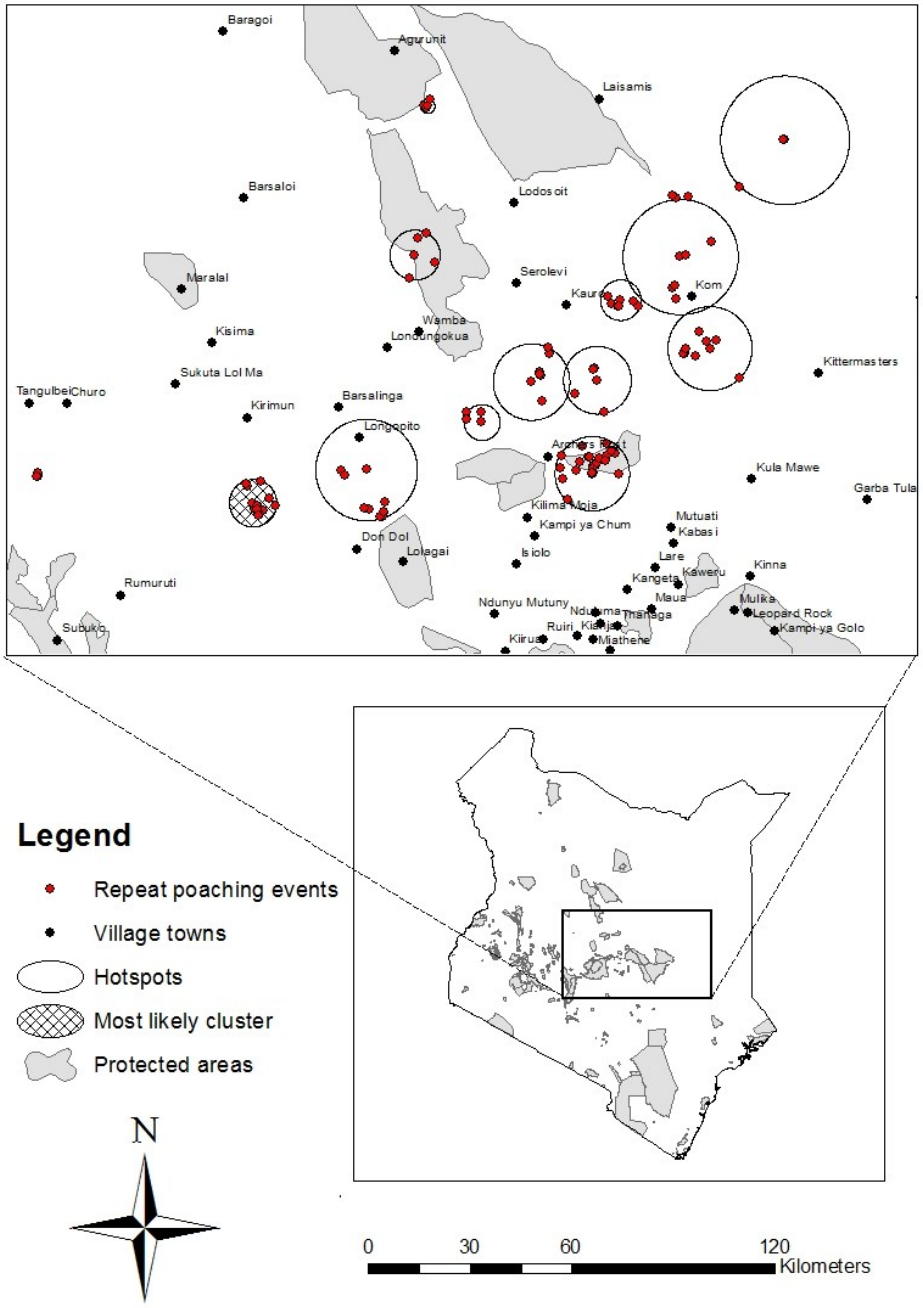


Figure 16: Laikipia - Samburu elephant poaching prone areas

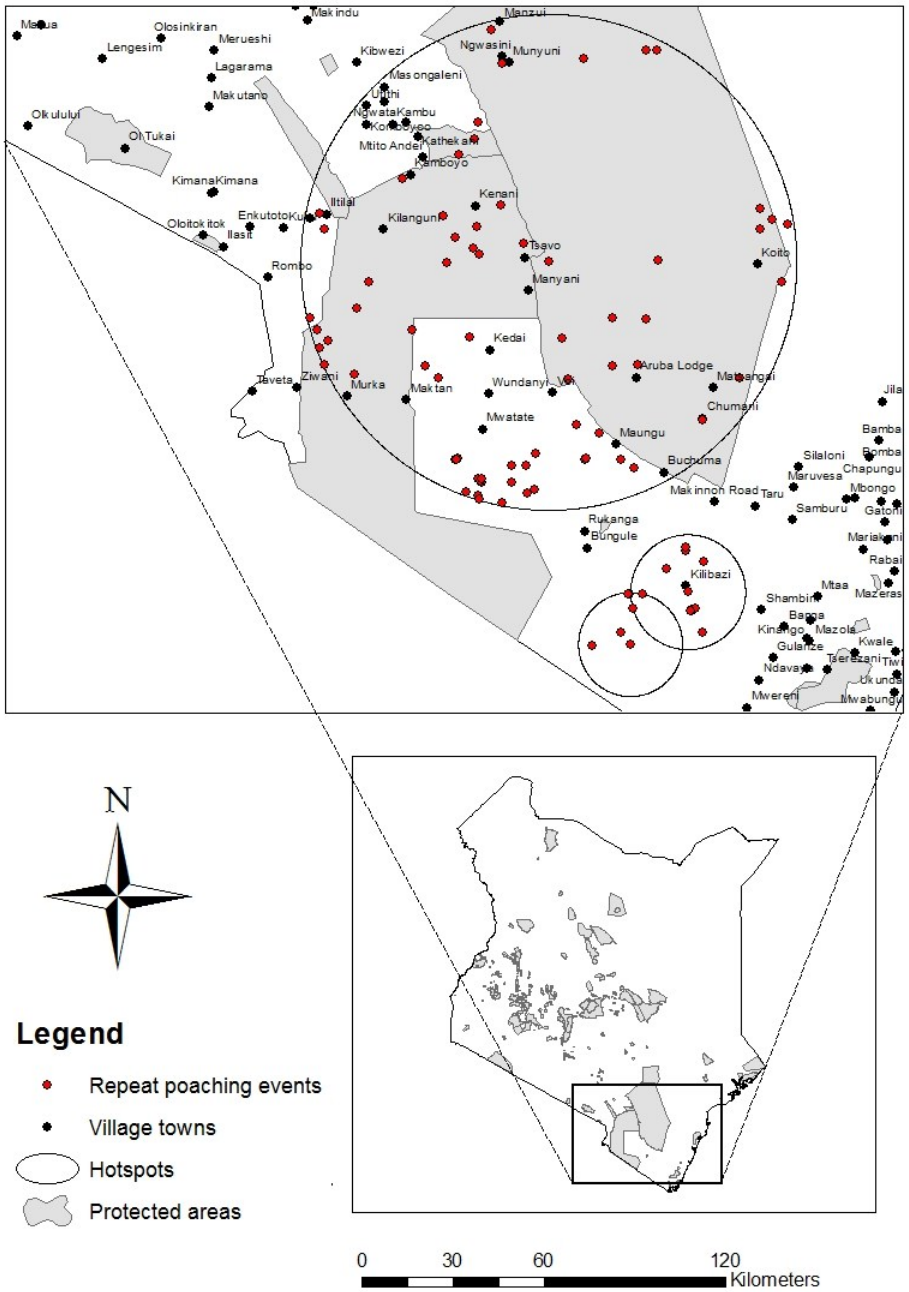


Figure 17: Tsavo ecosystem elephant poaching prone areas

3.2 Predictor variables for elephant poaching

Table 4: Summary of weighted means and relative contributions (%) of predictor variables for boosted regression tree model with cross validation on data from 410 sites using tree complexity of 5 and learning rate of 0.005.

Predictor	**Relative Contribution (%)**	**Weighted Mean of non-factor variables**
Distance to park boundaries (km)	21.5	15
Altitude (m)	18.3	1309
Poverty density (no. of poor people sq. km)	15.5	9
Land use	10.7	Categorical
Vegetation heterogeneity (Index)	7.6	0.0063
Distance to towns (km)	6.5	15
Soils	6.1	Categorical
Slopes	6	8.4
Distance to roads	3	3.5
Livestock density (no. of livestock per sq. km)	2.6	7
Distance to rivers (km)	2.2	2
Elephant population (No. of elephants per sq.km)	0.5	37

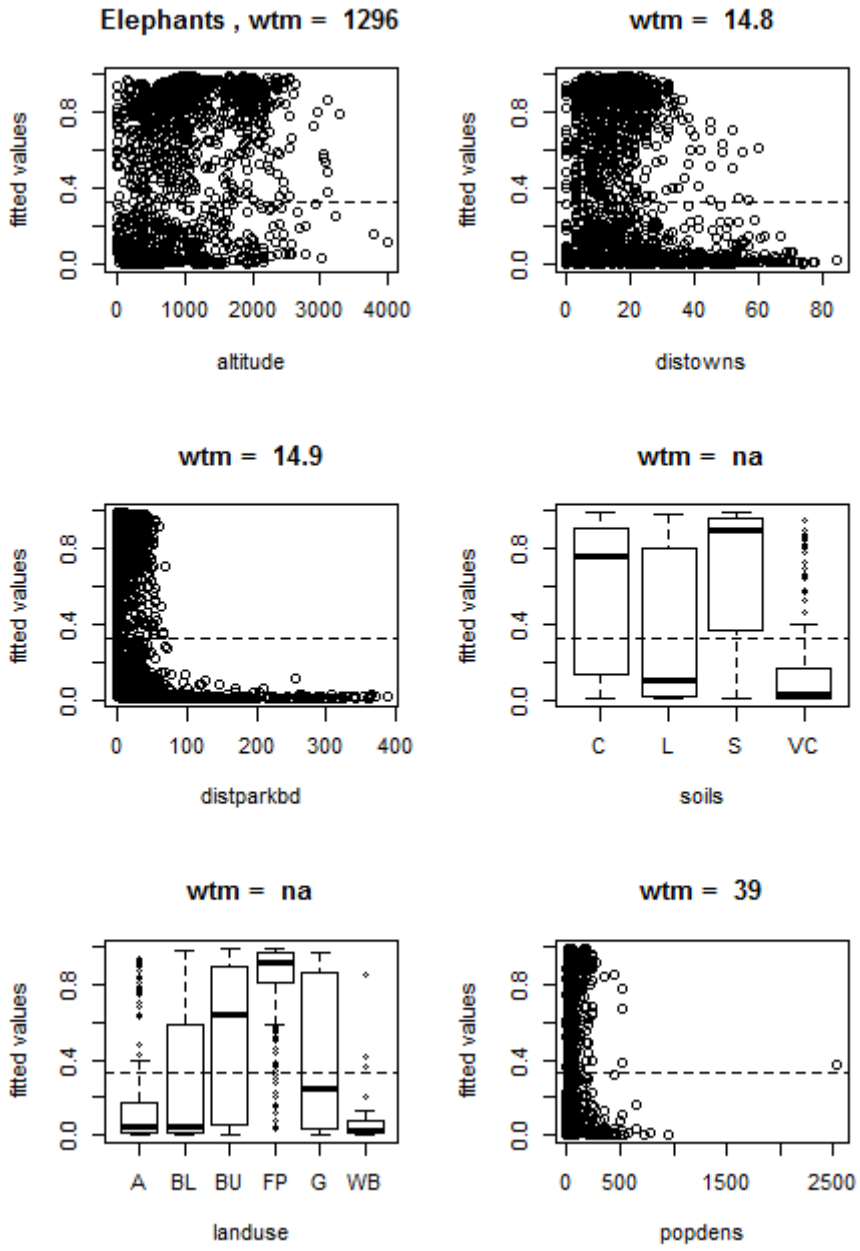


Figure 18: Each graph indicates the weighted mean of fitted values in relation to each non-factor predictor. ***wtm(weighted mean)**

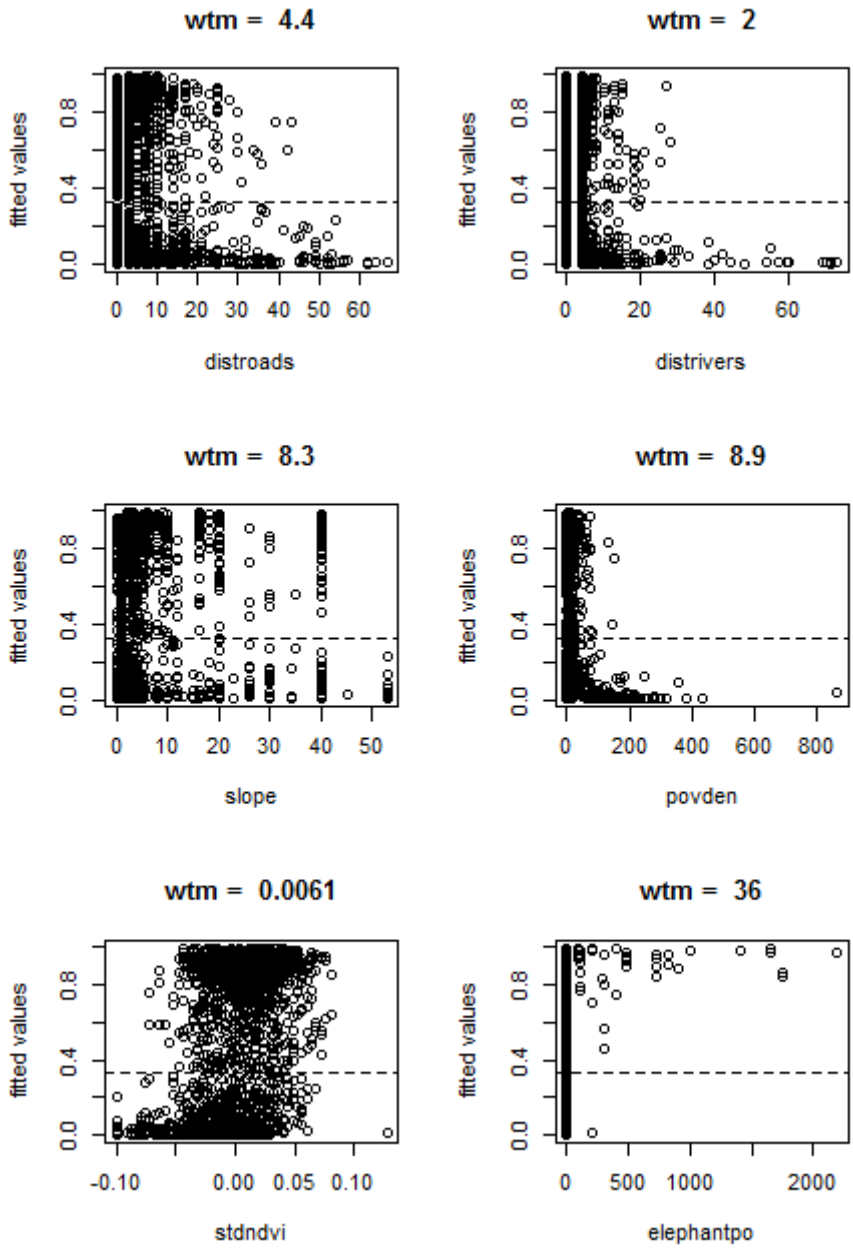


Figure 19: Each graph indicates the weighted mean of fitted values in relation to each non-factor predictor. ***wtm(weighted mean)**

The measures of relative importance of variables are based on the frequency of which a variable is selected for splitting, weighted by the squared improvement to the fitted model due to the result of each split, averaged over all trees. The relative influence of each tree is scaled to sum to 100, with higher numbers indicating stronger contribution (Elith et al., 2008).

For the model build for elephants poaching incidences on 410 sites through Cross validation, the six most important variables that explain the poaching events include: Distance to park boundaries, altitude, poverty density, land use types, vegetation heterogeneity and distance to towns (Table 4 and figure 20 and 21).

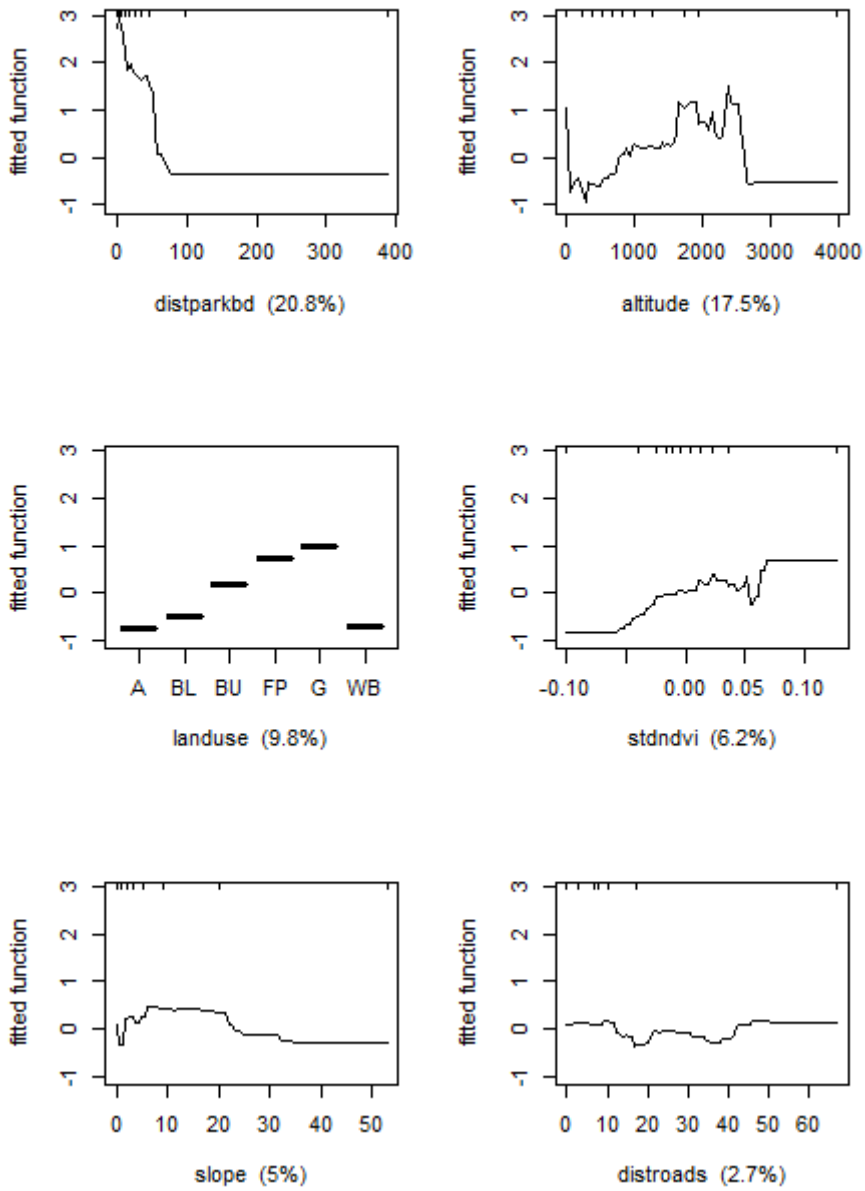


Figure 20: Partial dependence plots for 6 variables in the model for elephant poaching. Y axes are on the logit scale and centred to have zero mean over the data distribution. The rug plots on inside top plots representing distributions of sites across that variable.

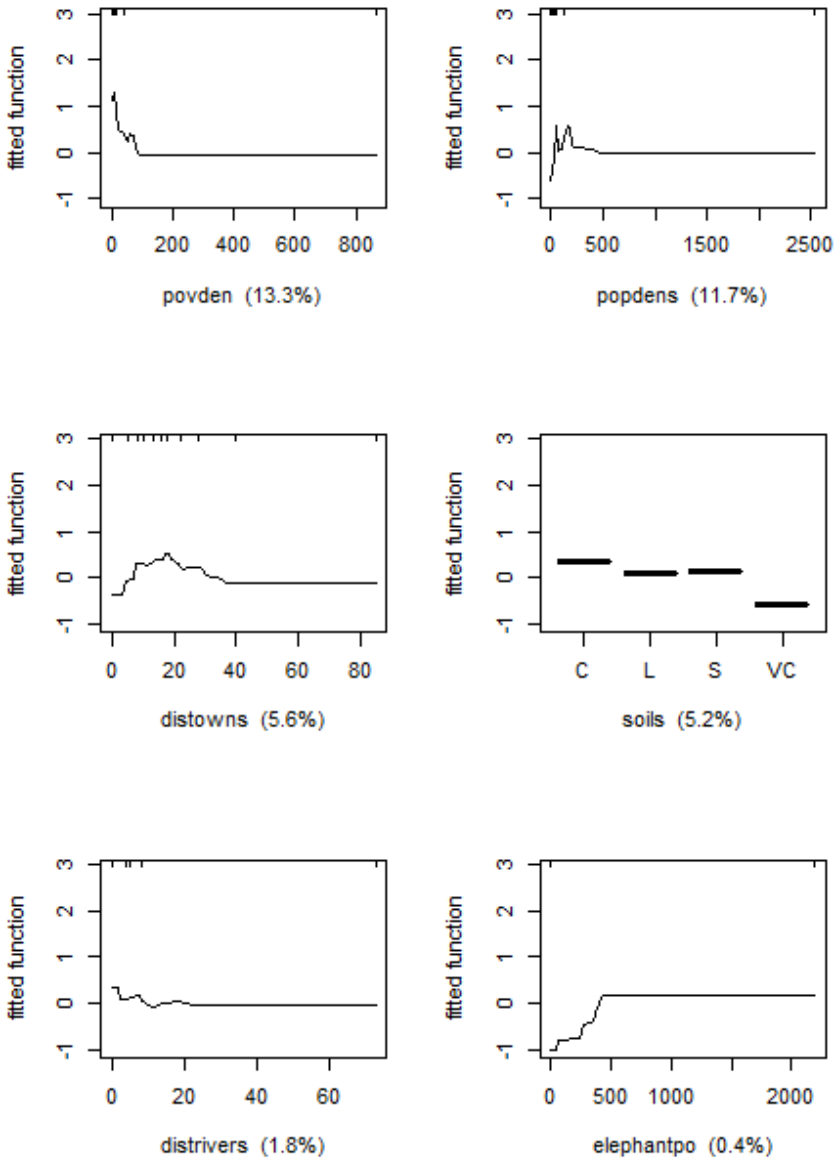


Figure 21: Partial dependence plots for 6 variables in the model for elephant poaching. Y axes are on the logit scale and centred to have zero mean over the data distribution. The rug plots on inside top plots representing distributions of sites across that variable

**distparkbd(distance to park boundaries), stdndv (standard deviation NDVI), distowns (distance to towns), distroads (distance to roads), livestockden (livestock density), A (Agriculture Lands), BL (Bare land), BU (Bush lands), FP (Forest &Plantations), G (Grassland) and WB (Water bodies). Popden (human population density) povden (poverty density),distrivers(distance to rivers), elephantpo(elephant population), C (Clayey), L (Loamy), S(Sandy), VC(Very clayey).

For easier visualization of fitted functions (i.e., the fitted values from the final tree, on the response scale) in a BRT model, partial dependence functions are vital. These show the effect of variable on response (i.e.; in this study, the elephant poaching cases) after accounting for the average effects of all other variables used to fit the model (Elith et al., 2008). The partial responses for elephant repeat poaching incidences described by most influential variables; indicates poaching mainly occurs near the park boundaries at mean distance of 15 km, at mean altitude of 1309 m above sea level, high income individuals contribute to elephant poaching crimes, and bush-lands (BU), plantations and forests (P) and grasslands (G), clay soils, dense settlements, flat regions experiencing more poaching crimes. Proximity to roads and rivers explains poaching cases, with more incidences occurring next to the roads and rivers, with mean distances to roads and rivers being 4, and 2 kilometers respectively. Livestock density does not clearly explain poaching. There is a direct relationship between the poaching cases and the elephant population. This indicates that the poachers target mainly regions of higher density of elephant populations (Figures 20 and 21).

Chapter 4. Discussion

4.1 Space-time clusters of elephant poaching

SaTScan evaluated very small and very large clusters, and everything in-between. In all the analyses, the most likely and secondary clusters of statistical significance of $p < 0.05$ were considered based on comparing the size of the log likelihood ratio against a null distribution obtained from Monte Carlo 999 replications.

The one month time precision window in the SaTScan program (Figure 15 **A**), shows clusters of varying extents with the Samburu-Laikipia area experiencing relatively smaller clusters. The space-time most likely clusters were significant for the years 2008, 2009 and 2004 respectively (Table 3). In contrast, Tsavo ecosystem experienced a much larger and consistent cluster. The home ranges of the Samburu-Laikipia elephants are relatively small less than 100 km^2 compared with other populations in Africa. In Tsavo National Park the greatest recorded individual home range is 3,744 km^2. By the year 1993, Samburu-Laikipia region held the largest population (i.e., 3000) of elephants outside of the protected areas in Kenya. This population covers a wide range of habitats including mountain forests, arid and semi-arid bushlands, with its range comprising forests and wildlife reserves, areas dominated by nomadic pastoralists, private ranches and small-scale agricultural settlements. A section of the population moves seasonally between ranchlands in Laikipia and pastoralists areas in Samburu (Thouless, 1995). The varied land use types, land fragmentation and pastoral communities reduce the home range of the elephants minimizing elephant movements across the ecosystem. "Even though the elephants' habitat is highly fragmented now, the animals have developed a

highly sophisticated pattern of moving between different safe areas" (Gross, 2008). This reason, we propose to explain smaller clusters extent in Laikipia-Samburu region as opposed to Tsavo ecosystem, which has a relatively homogeneous land use type (largely protected area) allowing free movements of the elephants; and hence the larger cluster and home range sizes respectively.

Clusters are non-existence in western conservation area and parts of central-rift in finer time precision (i.e., 1 and 6 months) (Figure 15 **A** and **B**). This could be explained probably by the fewer number of elephants populations in such areas (Figure 1), and hence not one of the main targets of poachers. Another reason could be that these areas receive better anti-poaching surveillances. Though, such regions still experience elephant poaching incidences in yearly space-time pattern (Figure 15 **C**).

The elephant population of Laikipia-Samburu ecosystem is on the increase despite drawbacks due to poaching and drought; resulting in high elephant mortality numbers. Encroachment of corridors increases the incidences of human wildlife conflicts. Hence, there is need for high level sustenance of security, joint planning initiatives on land use involving all conservation stakeholders and local communities (KWS, 2008). Areas with higher degrees of insecurity or political instability experience some levels of poaching, likewise protected areas surrounded by insecure regions for instance Shaba National Reserve are also a target. Though, regions bordering secure areas are record fewer cases of illegal elephant's killings (CITES, 2010).

The elephant populations in Coast, Tsavo and Southern conservation areas are known to be well protected than those in Mt. Elgon, Central Rift, Meru, Northern, Samburu-Laikipia and Mountain regions, which are known to be inadequately protected. With the

main cause of threat to elephants in the northern conservation area believed to be the proliferation of small firearms from politically unstable Somalia to the hands of the local communities (KWS, 2012).

The spatial and temporal analyses of poaching cases are helpful in managing the elephant crime incidences by determining where and when limited conservation resources should be allocated. For instance, this research reveals two major zones that require urgent surveillance; these include majorly the Tsavo conservation area with most repeat poaching cases reported around: Klibazi, Shambini, Murka, Kedal, Ziwani, Kilanguni, Aruba lodge, Kamboyo, Ngwashi, Munyuni, Kuku, Enkutoto, Matsangal, Koito and Rukanaga. Besides, Samburu-Laikipia region also encounter poaching crimes mostly in these specific regions: Kiri-mum, Kula-mawe, Kauro, kom, Doldol, Wamba, Baragoi, and Ratat. Refer to figures 16 and 17 respectively for detailed descriptions and illustrations.

Most of the clusters occur outside the protected areas, especially in Samburu-Laikipia area, with all the most likely clusters appearing on the private lands (Figure 6). This is consistent to the findings of KWS which indicates that 70% of Kenya's large mammal species occur in both private and trust lands and therefore are prone to human induced activities like poaching (KWS, 1994).

4.1 The relationship between elephant poaching patterns and biophysical and anthropogenic factors

Repeat poaching events has higher probability of existence near the park boundaries, at a mean distance of 15 km from the park boundaries. According to a study (Ottichilo, 1987), poaching was reported to be intense in areas bordering Tsavo national park in the years 1976 to 1978. At land-scape level, the probability of elephant

occurrence declines with increasing distance from the boundary of the protected areas. Protected areas and availability of drinking water are major predictors of elephant presence at landscape scale in both dry and wet seasons (Pittiglio et al., 2012). This otherwise reveals that the poachers would wait for the elephants to stray outside protected areas and kill them; hence security is a major factor determining frequency of poaching.

The relationship between the incidences of poaching and level of income is directly proportional (i.e., higher income higher poaching events). On the contrary, infant elephant mortality in and around MIKE's (Monitoring of illegally Killed Elephants) sites, is used as a proxy for poverty at site level. This has continued to be the strongest site-level predictor of PIKE (Proportion of illegally Killed Elephants), with sites suffering from higher levels of poverty experiencing elephant poaching (CITES, 2012). According to a study carried out in Samburu, Kenya. An adult elephant mortality and poaching are closely correlated with indices of economic conditions in local pastoral communities (Wittemyer, 2011). The increased conflict between humans and wild animals in times of economic recessions could as well be a force behind poaching. Human generally target adult animals as a function of the size of their tusks and their weight (meat). The livelihoods of the local people may be more sensitive to conflict with wild animals or tolerance of wild animals may decline as a function of stress on local communities caused by economic hardships (Wittemyer, 2011).

The association between poverty, food security and proportion of illegally killed elephants shows a close linkage between the well-being of local communities and the health of elephant's population. Hence, the local communities could engage in illegal

killing of elephants in exchange with incentives, mainly in areas where livelihoods are insecure (CITES, 2012).

Information concerning economic fluctuations can be utilized to focus on management activities for instance, enhancing job creation during economic downturns in response to or anticipation of economically associated changes in natural resource reliance (Wittemyer, 2011). Poaching at a local level is affected by local socio-economic factors and incentives for example, the opportunity costs of poaching and penalties (Poudyal et al., 2009).

The parameters used to determine the level of poverty in the pastoral communities is still debatable. Some reports for instance the CITES report, uses types of housing units, level of income, and living standards to determine the levels of poverty of the local communities. The economic value of herds of the livestock owned by the local communities living close to the protected areas have been ignored in estimating the poverty levels of communities living within and around protected areas. This has been sighted as a possible source of bias in concluding that the local communities are poor; therefore engage in poaching related activities.

There is a higher likelihood of an elephant being poached in bushlands, forests, and plantations and in the grasslands than in agricultural land, swampy and less vegetative regions. In regard to CITES report, a strong relationship exists between vegetation density and proportion of illegally killed elephants. Animals stay at feeding sites for a specific time period and move on once the advantage of remaining there diminishes (Ngene et al., 2010). Densely vegetated areas provide favourable sites for poachers to hide hence this variable is interpreted as an indicator with which poaching can be conducted. Land cover is a moderate predictor of elephant presence. During the dry spells, closed woody vegetation and closed shrubs are

strongly associated with elephant presence (Pittiglio et al., 2012). On the contrary, the recent increase in poaching cases across both savannah and forests has decreased the importance of this variable to a level of becoming statistically not correlated to the proportion of illegally killed elephants (CITES, 2012).

Closer proximity to both permanent and seasonal rivers explains greater elephant poaching incidences with a mean distance of 2 km to both permanent and seasonal rivers (Table 4). Elephants need drinking water at least in a day or two and hence elephants like being next to drinking water points. Though in wet season, the elephants prefer seasonal rivers in lowlands. Water therefore, is a major determinant of the distribution of elephants across an ecosystem (Ngene et al., 2010). The period of intense poaching in Tsavo National park is in the dry months of June to October, with peaks in June and July, hence this could also explain the aggregation of elephants near permanent rivers hence an appropriate target to poachers (Maingi et al., 2012). Elephants may live at densities as low as 0.024 km^2 or as high as 5 per km^2 (Douglas Hamilton, 1972).

More elephants are poached near the built up areas at a mean distance of 15 km from settlements. A recent similar but a regional study, reveals that Marsabit elephants aggregate next to human settlements with a mean distance of 3 km. This is as a result of humans sharing the same water points with the elephants (Ngene et al., 2010). Distance from settlements has a positive correlation with elephant presence in both seasons, with the settlements close to permanent water sources experiencing greater local elephant probability density even in drier periods (Ngene et al., 2010).

The mean distance from the locations of elephants poaching to both major and minor roads is 3.5 km (Table 4). The Marsabit elephants prefer being next to both minor and major roads; this is

because they cross the roads in search of water and lush pasture (Ngene et al., 2010). Therefore, the poachers may be waylaying the elephants when crossing the insecure and un-patrolled roads. This is contrasted by a recent study which reveals that the elephants avoid areas with high road densities relating to high human pressure implying that elephants avoid human encroachment. Hence, the elephants tend to move away from human-dominated areas and move into more vegetated areas (Rood et al., 2010).

In reference to the results of our study in (Table 4), poaching crimes occur at mean altitude and slopes of 1309 m above sea level, and 8.4 % respectively. The elephants use areas between 1600 to 2200 m above sea level; though the elephants might prefer flatter, lowland area, this does not imply that they are non-existence in mountainous areas with steep slopes that could limit their movements. Elephants have a strong preference for forests with a high productivity located within a valley. This pattern is attributed to the fact that landscape depressions also act as waterways providing a source of water and natural routes for crossing. Similarly, elephants like lowland forest habitats in which nutritious foliage is abundant (Rood et al., 2010).

During the dry season, the drinking water points in the lowlands dry-up and the foliage plants drop their leaves, thereby becoming unsuitable forage areas for the elephants. Alternatively, during the wet season food and water are not limiting factors and hence the elephants move from slippery clayey highlands to lowlands. Also, the elephants avoid the high elevations due to the tall trees (over 20 m) that supress the under-growth resulting in less shrubby patches. Marsabit elephants avoid high elevations due to their steepness, which poses higher chances of injuries due to the slippery soils. High elevations are utilized by elephants during the dry

seasons. There is absolute avoidance of cliffs by elephants in all seasons since they cannot move with ease in such conditions (Ngene et al., 2010). Steep slopes are a constrain to elephant movement (Lin et al., 2008).

The results of our study show that poaching tends to occur in areas with some level variability in vegetation cover (higher standard deviation of SPOT-VGT NDVI). Elephant prefers an environment of high variability of vegetation species cover; they can as well stay for more than 5 hours in natural vegetation patches of about 0.25 km^2 (Murwira et al., 2005). Murwira further claims that elephants do respond to changes in spatial heterogeneity over time. For instance, a decrease in elephant presence is correlated with: (1) a decline in the intensity of spatial heterogeneity which simultaneously occurs with an increase in dominant scale of heterogeneity in cultivated areas and (2) a decrease in both the intensity of spatial heterogeneity and consistent scale of spatial heterogeneity revealed that elephants move away when the small vegetation cover patch composed the cultivated landscape. In contrast, elephants persisted in environments with constant levels of spatial heterogeneity (Murwira et al., 2005).

In figure 21 above, the areas dominated with clay soils contribute to greater incidences of human induced elephant deaths as opposed to the sandy, very clayey and loamy soils. Very clayey soils tend to be slippery in wet seasons, hence could be avoided by elephants due to the associated injury risks. To contest our findings, African elephants regularly eat soils for instance, some sites like Mount Elgon on the Kenya-Uganda border, experience extensive caves excavated by elephants due to their quarrying activities (Houston et al., 2001). Elephants select soil with a high level of sodium and other mineral nutrients to supplement the deficiency in

their diet (Ruggiero, 1992). The pharmaceutical properties of kaolin are well known; it adsorbs toxic substances from the alimentary tract and increases the bulk of faeces. Kaolin clays are utilized in animal feeds, partly as binding agents, but also because of their effect in preventing diarrhoea and mycotoxicosis (Houston et al., 2001).

Forest elephants that usually have access to kaolin soils are able to feed on a wider range of plant species than animals limited to such access, and may as well have an enhanced digestive efficiency (Klaus et al., 1998). Animals may select regions where they have access to kaolin soils during periods of low food quality, and hence the distribution of clay soils could influence animal distribution and movements (Houston et al., 2001).

Conclusively, poaching incidences are positively correlated to elephant population. Though up-to some level over which it becomes indifferent (Figure 21). This shows that potential poachers mainly target areas of higher population of elephants like Tsavo and mountain conservation regions (Figure 1).

Chapter 5. Conclusion and recommendations

5.1 General conclusion

The first objective aimed at determining whether the observed patterns of elephant poaching incidences are simply random or clustered in space and time. Our results ascertained that once the null hypothesis was rejected, the clusters were formed hence the detected repeat poaching cases were significantly different from other regions not experiencing hotspots. Generally, this study provides a first attempt to visually and quantitatively describe the geographic and temporal features of elephant poaching in Kenya; thus demonstrating the utility of space-time scan statistics in demonstrating elephant crimes. The precise spatial and temporal SaTScan's analysis of poaching events could be applied in managing the frequent elephant associated criminal activities around the country by focusing on where and when the scarce resources need to be concentrated. Results from boosted regression trees identify what combination factors provide convenient regions where poachers would carry out their criminal acts. The KWS management, hence can easily site where to construct rangers post for an effective elephant poaching surveillance.

Quantitatively, the second objective aimed at identifying both biophysical and human factors which explain the clustering patterns of the poaching incidences. From the results of boosted regression trees, such factors as: Distance to protected areas boundaries, altitude, poverty density and land use contributed more in explaining the repeat elephant poaching incidences at 22%, 18%, 15% and 11% respectively. Though, the elephant population density contributed the

least at 0.5%. The scan statistic method had two major strengths: (1) The spatial scan statistic identifies single or multiple clusters over space and time and (2) the availability of fairly large sample of poaching cases over 10 years facilitated reliable, dynamic identification of clusters and ensured a sufficiently high statistical power.

Besides, there were limitations of Kulldorff spatial scan statistic and SatScan: First, SaTScan has no visual interface to explore the cluster features for instance, the cluster radius, the centre location and other data entities associated with the cluster. Instead, these are only available in text format, and therefore to visualize the geographic extent and size of the cluster, the user has to process textual output in GIS software (e.g., ArcGIS). This is time consuming and renders inadequacy in data exploration. Second, it is challenging to determine an optimal setting for SaTScan scaling parameters. The determination of the most appropriate maximum-size parameter is difficult; too large maximum size can hide small, homogeneous clusters within larger, heterogeneous ones, and too small of maximum size can fail to detect significant, local level clusters.

5.2 Recommendations for future management actions

The identified poaching hotspots mostly occur within the mountain conservation region and more specifically within the Samburu-Laikipia area. Majority of the poaching hotspots appeared in the un-protected and private lands. This area is prone to the proliferation of small arms from Somali into the hands of criminals. This calls for bottom-up approach, where the local communities,

private land owners, and governmental agencies are jointly engaged in conservation and security issues.

Kenya Wildlife Service should intensify patrolling both in Samburu-Laikipia and Tsavo regions mainly in areas such as Klibazi, Shambini, Murka, Kedal, Ziwani, Kilanguni, Aruba lodge, Kamboyo, Ngwashi, Munyuni, Kuku, Enkutoto, Matsangal, Koito and Rukanaga in the larger Tsavo conservation area. Besides, areas including: Kirimum, Kula-mawe, Kauro.Kom, Doldol, Wamba, Baragoi, and Ratat in Samburu/Laikipia region are worst affected.

KWS should carry out frequent surveillance within and around the elephant poaching hotspots especially targeting the major permanent rivers during the dry seasons, and around the protected areas boundaries in all seasons.

The regions, roads in the proximity of the neighbourhoods where the elephants share water points with communities should be provided with security due to greater poaching risks near the human settlements. Probably, the local people are secretly being involved in poaching.

Some of the elephant poaching hotspots are relatively inaccessible thereby hindering effective elephant crime surveillance especially in Samburu-Laikipia and a few areas around Banga and Kwale regions. Hence a need to construct roads to ease security patrols within such mentioned areas.

References

Abatih, E. N., Ersbøll, A. K., Lo Fo Wong, A., D. M., and, & Emborg, H. D. (2009). Space–time clustering of ampicillin resistant Escherichia coli isolated from Danish pigs at slaughter between 1997 and 2005. *Preventive Veterinary Medicine, 89*(1–2), 90-101.

Archie, E. A., & Chiyo, P. I. (2012). Elephant behaviour and conservation: social relationships, the effects of poaching, and genetic tools for management. *Molecular Ecology, 21*(3), 765-778.

Batjes, N. H. (2004). Soil carbon stocks and projected changes according to land use and management: a case study for Kenya. *Soil Use and Management, 20*(3), 350-356.

Beyers, R. L., Hart, J. A., Sinclair, A. R. E., Grossmann, F., Klinkenberg, B., & Dino, S. (2011). Resource Wars and Conflict Ivory: The Impact of Civil Conflict on Elephants in the Democratic Republic of Congo - The Case of the Okapi Reserve. [Article]. *PLoS ONE, 6*(11), 1-13.

Burnett, G. W., & Rowntree, K. M. (1990). Agriculture, research and tourism in the landscape of Lake Baringo, Kenya. *Landscape and Urban Planning, 19*(2), 159-172.

Burton, M. (1999). An assessment of alternative methods of estimating the effect of the ivory trade ban on poaching effort. *Ecological Economics, 30*(1), 93-106.

Chen, J., Roth, R. E., Naito, A. T., Lengerich, E. J., and, & MacEachren, A. M. (2008). Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. *International Journal of Health Geographics*.

CITES. (2010). LEVELS OF ILLEGAL KILLINGS OF ELEPHANTS IN THE LAIKIPIA-SAMBURU MIKE SITE.

CITES. (2012). ELEPHANT CONSERVATION, ILLEGAL KILLING AND IVORY TRADE .

Craney, T. A., & Surles, J. G. (2002). Model-Dependent Variance Inflation Factor Cutoff Values. [Article]. *Quality Engineering, 14*(3), 391.

Creek, P. J. (1990). Generalized Linear Models (Book). [Book Review]. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 39*(3), 385.

Dormann, C. F. (2009). Response to Comment on "Methods to account for spatial autocorrelation in the analysis of species distributional data: a review". [Article]. *Ecography, 32*(3), 379-381.

Douglas Hamilton, I. (1972). On the ecology and behaviour of the African

Elephant. [D. Phil. Thesis. University of Oxford, U.K.].
Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. [Article]. *Journal of Animal Ecology, 77*(4), 802-813. Fox, J. (1997). Applied regression analysis, linear models and related methods:California: Sage Publications.
Grey-Ross, R., Downs, C. T., & Kirkman, K. (2010). An assessment of illegal hunting on farmland in KwaZulu-Natal, South Africa: implications for oribi (Ourebia ourebi) conservation. [Article]. *South African Journal of Wildlife Research, 40*(1), 43-52.
Gross, M. (2008). Kenya's conservation challenge. *Current Biology, 18*(14), R576-R577.
Heltberg, R. (2001). Impact of the ivory trade ban on poaching incentives: a numerical example. *Ecological Economics, 36*(2), 189-195.
Houston, D. C., Gilardi, J. D., & Hall, A. J. (2001). Soil consumption by Elephants might help to minimize the toxic effects of plant secondary compounds in forest browse. *Mammal Review, 31*(3-4), 249-254. doi: 10.1111/j.1365-2907.2001.00091.x
Kahindi, O., Wittemyer, G., King, J., Ihwagi, F., Omondi, P., & Douglas-Hamilton, I. (2010). Employing participatory surveys to monitor the illegal killing of elephants across diverse land uses in Laikipia-Samburu, Kenya Onesmas Kahindi et al. Approaches for monitoring illegal killing of elephants. [Article]. *African Journal of Ecology, 48*(4), 972-983.
Klaus, G., Klaus-Hugi, C., & Schmid, B. (1998). Geophagy by Large Mammals at Natural Licks in the Rain Forest of the Dzanga National Park, Central African Republic. *Journal of Tropical Ecology, 14*(6), 829-839.
Kulldorff, M. (2010). SaTScan User Guide for version 9.0.
Kulldorff, M., Heffernan, R., Jessica, H., Assunc¸a˜, R., and, & Mostashari, F. (2005). A Space–Time Permutation Scan Statistic for Disease Outbreak Detection

PLoS MEDICINE, 2(3). doi: 10.1371/journal.pmed.0020059
KWS. (1990). A policy Framework and Development Programmme 1991- 1996. Kenya Wildlife, Nairobi.
KWS. (1991 a & b). A policy frame work and development programme 1991-1996: Annex6.Community conservation and wildlife management outside parks and reserves. Kenya Wildlife Service, Nairobi.
KWS. (1994). Wildlife - human conflicts in Kenya. Report of the the Five-Person Review Group. Kenya Wildlife Service, Nairobi.
KWS. (2008). Total Aerial Count of Elephants in Laikipia-Samburu Ecosystem in November 2008. Elephant program.

KWS. (2011). TOTAL AERIAL CENSUS OF ELEPHANTS AND OTHER LARGE MAMMALS IN THE TSAVO-MKOMAZI ECOSYSTEM.

KWS. (2012). Conservation and Management Strategy for the Elephants in Kenya. Kenya Wildlife Service, Nairobi.

Kyale, D. K., Ngene, S., & Maingi, J. (2011). Biophysical and human factors determine the distribution of poached elephants in Tsavo East Nationak Park, Kenya. *Pathyderm*(49).

Lado, C. (1992). Problems of wildlife management and land use in Kenya. *Land Use Policy, 9*(3), 169-184.

Leathwick, J. E. a. J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Ecology, Evolution, and Systematics, Vol. 40: 677-697*. doi: 10.1146/annurev.ecolsys.110308.120159

Lemieux , A., & Clarke, R. (2009). The international ban on ivory sales and its effects on elephant poaching in africa

British journal of criminology, 49(4), 451-471.

Lin, L., Feng, L., Pan, W., Guo, X., Zhao, J., Luo, A., & Zhang, L. (2008). Habitat selection and the change in distribution of asian elephants in Mengyang Protected Area, Yunnan, China. *Acta Theriologica, 53*(4), 365-374.

Maingi, J. K., Mukeka, J. M., Kyale, D. M., & Muasya, R. M. (2012). Spatiotemporal patterns of elephant poaching in south-eastern Kenya. [Article]. *Wildlife Research, 39*(3), 234-249.

Mola-Yudego, B., & Gritten, D. (2010). Determining forest conflict hotspots according to academic and environmental groups. *Forest Policy and Economics, 12*(8), 575-580.

Montgomery, D. C., & , & Peck, E. A. (1982). Introduction to linear regression analysis. New York.

Murwira, A., & Skidmore, A. K. (2005). The response of elephants to the spatial heterogeneity of vegetation in a Southern African agricultural landscape. *Landscape Ecology, 20*(2), 217-234.

Ngene, S. M., Skidmore, A. K., Prins, H. H. T., & van Gils, H. A. M. J. (2010). *Why elephant roam.* (171), University of Twente Faculty of Geo-Information and Earth Observation ITC, Enschede. Retrieved from http://www.itc.nl/library/papers_2010/phd/ngene.pdf

Norton-Griffiths, M. (2000). wildlife losses in kenya: an analysis of conservation policy *Natural Resource Modeling, 13*(1), 13-34.

Ottichilo, W. K. (1987). The causes of the recent heavy elephant mortality in the Tsavo ecosystem, Kenya, 1975–1980. *Biological Conservation, 41*(4), 279-289. Owen, J. G. (1988). On Productivity as a Predictor of Rodent and Carnivore Diversity. *Ecology, 69*(4), 1161-1165.

Per, J., & Lars, E. TIMESAT—a program for analyzing time-series of satellite sensor data. [Article]. *Computers and Geosciences, 30*, 833-845. doi: 10.1016/j.cageo.2004.05.006

Pittiglio, C., Skidmore, A. K., van Gils, H. A. M. J., & Prins, H. H. T. (2012). Identifying transit corridors for elephant using a long time-series. *International Journal of Applied Earth Observation and Geoinformation, 14*(1), 61-72.

Poudyal, M., Rothley, K., & Knowler, D. (2009). Ecological and Economic Analysis of Poaching of the Greater One-Horned Rhinoceros (Rhinoceros unicornis) in Nepal. *Ecological Applications, 19*(7), 1693-1707. doi: 10.2307/40346282

Praagman, J. (1985). Classification and Regression Trees (Book). [Book Review]. *European Journal of Operational Research, 19*(1), 144-144.

Rood, E., Ganie, A. A., & Nijman, V. (2010). Using presence-only modelling to predict Asian elephant habitat use in a tropical forest landscape: implications for conservation E. Rood et al. Sumatran elephant habitat use in Indonesia. [Article]. *Diversity & Distributions, 16*(6), 975-984.

Ruggiero, R. G. (1992). Seasonal forage utilization by elephants in central Africa. *African Journal of Ecology, 30*(2), 137-148.

Thouless, C. R. (1995). Long distance movements of elephants in northern Kenya. *African Journal of Ecology, 33*(4), 321-334.

van Kooten, G. C. (2008). Protecting the African elephant: A dynamic bioeconomic model of ivory trade. *Biological Conservation, 141*(8), 2012-2022. doi: 10.1016/j.biocon.2008.05.016

Western, D. (1987). Africa's elephants and rhinos: Flagships in crisis. *Trends in Ecology & Evolution, 2*(11), 343-346. doi: 10.1016/0169-5347(87)90112-1

Wittemyer, G. (2011). Effects of Economic Downturns on Mortality of Wild African Elephants *Conservation Biology, 25*(5), 1002-1009.

Wittemyer, G., Daballen, D., Rasmussen, H., Kahindi, O., & Douglas-Hamilton, I. (2005). Demographic status of elephants in the Samburu and Buffalo Springs National Reserves, Kenya. [Article]. *African Journal of Ecology, 43*(1), 44-47.

Woolley, L.-A., Mackey, R. L., Page, B. R., & Slotow, R. (2008). Modelling the effect of age-specific mortality on elephant <italic>Loxodonta africana</italic> populations: can natural mortality provide regulation? *Oryx, 42*(01), 49-57.

Appendices

Appendix 1. Statistically significant clusters

Statistically significant clusters at 0.005 simulated by SaTScan space-time permutation model using maximum spatial cluster size of 50% of the cases, maximum temporal cluster size of 50% of study period, and circle radius of 100 km.

```
                    ____________________________
                          SaTScan v9.1.1
                    ____________________________
Program run on: Wed Nov 07 11:53:41 2012
Retrospective Space-Time analysis
scanning for clusters with high rates
using the Space-Time Permutation model.
_____________________________________________________________________________
SUMMARY OF DATA
Study period..................: 2002/1/1 to 2012/8/31
Number of locations...........: 969
Total number of cases.........: 1006
_____________________________________________________________________________
MOST LIKELY CLUSTER
1.Location IDs included.: 1624, 1608, 1631, 1620, 1635,
1625,
                     1617, 1623, 1910, 1622, 1616,
1630,
                     1609, 1619
  Coordinates / radius..: (0.518600 N, 36.886870 E) /
6.97 km
  Time frame............: 2008/4/1 to 2008/4/30
  Number of cases.......: 11
  Expected cases........: 0.19
  Observed / expected...: 56.46
  Test statistic........: 33.621776
  Monte Carlo rank......: 1/1000
  P-value...............: 0.001
SECONDARY CLUSTERS
2.Location IDs included.: 1706, 1707, 1708, 1709, 1710,
1711,
                                  1705
  Coordinates / radius..: (1.489100 N, 38.296770 E) /
19.37 km
  Time frame............: 2008/12/1 to 2008/12/31
  Number of cases.......: 7
```

```
  Expected cases........: 0.049
  Observed / expected...: 143.71
  Test statistic........: 27.847571
  Monte Carlo rank......: 1/1000
  P-value...............: 0.001
3.Location IDs included.: 1696, 1697, 1698, 1700, 1701,
1699,
                    1034, 1702
  Coordinates / radius..: (0.597500 N, 36.317900 E) /
0.94 km
  Time frame............: 2008/11/1 to 2008/11/30
  Number of cases.......: 8
  Expected cases........: 0.11
  Observed / expected...: 71.86
  Test statistic........: 26.339787
  Monte Carlo rank......: 1/1000
  P-value...............: 0.001
4.Location IDs included.: 1500, 1804, 1797, 1798, 1803,
1785,
                    1801, 1544, 1799, 1723
  Coordinates / radius..: (0.607400 N, 37.187570 E) /
14.77 km
  Time frame............: 2009/4/1 to 2009/4/30
  Number of cases.......: 7
  Expected cases........: 0.15
  Observed / expected...: 46.95
  Test statistic........: 20.115575
  Monte Carlo rank......: 1/1000
  P-value...............: 0.001
5.Location IDs included.: 1488, 1490, 1489, 1972, 1487,
1486
  Coordinates / radius..: (0.847300 N, 37.799380 E) /
9.80 km
  Time frame............: 2002/9/1 to 2002/9/30
  Number of cases.......: 5
  Expected cases........: 0.042
  Observed / expected...: 119.76
  Test statistic........: 18.981517
  Monte Carlo rank......: 1/1000
  P-value...............: 0.001
6.Location IDs included.: 1415, 1440, 1441, 1439, 1444,
1400,
                    1403, 1430, 1219, 1450,
1451
  Coordinates / radius..: (3.990776 S, 38.952500 E) /
14.95 km
  Time frame............: 2012/8/1 to 2012/8/31
  Number of cases.......: 7
  Expected cases........: 0.27
  Observed / expected...: 25.61
```

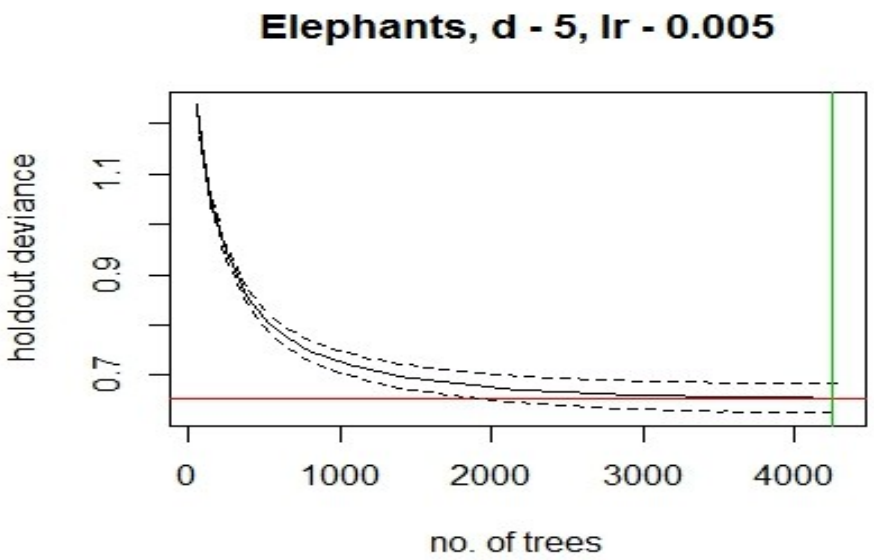
```
  Test statistic........: 15.996040
  Monte Carlo rank......: 1/1000
  P-value...............: 0.001
7.Location IDs included.: 1654, 1746, 1939, 1866, 1572,
1676,
                    1677, 1680, 1704, 1032, 1679,
1033
  Coordinates / radius..: (1.176500 N, 38.019990 E) /
18.51 km
  Time frame............: 2008/8/1 to 2008/8/31
  Number of cases.......: 6
  Expected cases........: 0.19
  Observed / expected...: 31.44
  Test statistic........: 14.895672
  Monte Carlo rank......: 2/1000
  P-value...............: 0.002
8.Location IDs included.: 1583, 1896, 1923, 1605, 1602,
1585,
                    1591, 1584, 1603, 1579, 1734,
1732,
                    1733, 1581, 1735, 1542, 1736,
1965,
                    1580, 1826, 1690, 1582, 1472,
1570,
                    1453, 1590
  Coordinates / radius..: (0.597600 N, 37.785980 E) /
10.54 km
  Time frame............: 2006/10/1 to 2006/10/31
  Number of cases.......: 6
  Expected cases........: 0.21
  Observed / expected...: 29.02
  Test statistic........: 14.431228
  Monte Carlo rank......: 2/1000
  P-value...............: 0.002
9.Location IDs included.: 1189, 1190, 1192, 1194, 1193
  Coordinates / radius..: (0.307300 S, 38.942727 E) /
0.068 km
  Time frame............: 2011/7/1 to 2011/7/31
  Number of cases.......: 5
  Expected cases........: 0.12
  Observed / expected...: 41.92
  Test statistic........: 13.809562
  Monte Carlo rank......: 5/1000
  P-value...............: 0.005
10.Location IDs included.: 1460, 1461, 1992, 1991, 1462,
1531,
                     1988, 1463
   Coordinates / radius..: (1.061300 N, 37.861280 E) /
5.67 km
   Time frame............: 2002/4/1 to 2002/4/30
```

```
   Number of cases.......: 4
   Expected cases........: 0.056
   Observed / expected...: 71.86
   Test statistic........: 13.162129
   Monte Carlo rank......: 5/1000
   P-value...............: 0.005
11.Location IDs included.: 1007, 1022, 1144, 1177, 1224,
1218,
                    1020, 1008, 1117, 1010, 1419,
1019,
                    1397, 1101, 1122, 1121, 1445,
1002,
                    1016, 1208, 1406, 1377, 1105,
1106,
                    1413, 1235, 1412, 1149, 1228,
1089,
                    1068, 1414, 1365, 1234, 1223,
1055,
                    1216, 1017, 1374, 1138, 1373,
1423,
                    1227, 1196, 1384, 1362, 1417,
1248,
                    1338, 1083, 1203, 1188, 1289,
1202,
                    1187, 1283, 1350, 1401, 1207,
1446,
                    1421, 1422, 1410, 1004, 1176,
1009,
                    1171, 1353, 1352, 1290, 1418,
1404
   Coordinates / radius..: (3.004000 S, 38.538000 E) /
81.99 km
   Time frame............: 2012/7/1 to 2012/7/31
   Number of cases.......: 12
   Expected cases........: 1.72
   Observed / expected...: 6.99
   Test statistic........: 13.097601
   Monte Carlo rank......: 5/1000
   P-value...............: 0.005
12.Location IDs included.: 1277, 1276, 1275, 1278, 1280
   Coordinates / radius..: (1.579386 N, 37.349200 E) /
2.10 km
   Time frame............: 2012/1/1 to 2012/1/31
   Number of cases.......: 5
   Expected cases........: 0.14
   Observed / expected...: 34.69
   Test statistic........: 12.888083
   Monte Carlo rank......: 6/1000
   P-value...............: 0.006
13.Location IDs included.: 1107, 1247, 1236, 1095, 1110,
```

```
1111,
                                        1113
  Coordinates / radius..: (4.147300 S, 38.780837 E) /
17.16 km
  Time frame............: 2010/6/1 to 2010/6/30
  Number of cases.......: 4
  Expected cases........: 0.063
  Observed / expected...: 63.87
  Test statistic........: 12.697928
  Monte Carlo rank......: 7/1000
  P-value...............: 0.007
14.Location IDs included.: 2004, 2005, 2006, 1971
  Coordinates / radius..: (0.734400 N, 37.493840 E) /
5.25 km
  Time frame............: 2011/11/1 to 2011/11/30
  Number of cases.......: 4
  Expected cases........: 0.076
  Observed / expected...: 52.95
  Test statistic........: 11.960405
  Monte Carlo rank......: 9/1000
  P-value...............: 0.009
15.Location IDs included.: 1783, 1808, 1904, 1903, 1905,
1742,
                    1906, 1907
  Coordinates / radius..: (1.181700 N, 37.315850 E) /
7.43 km
  Time frame............: 2011/1/1 to 2011/1/31
  Number of cases.......: 5
  Expected cases........: 0.18
  Observed / expected...: 27.34
  Test statistic........: 11.735654
  Monte Carlo rank......: 14/1000
  P-value...............: 0.014
16.Location IDs included.: 1865, 1661, 1662, 1652, 1663,
1665,
                    1745, 1664, 1744, 1650,
1028
  Coordinates / radius..: (0.932000 N, 38.097870 E) /
12.40 km
  Time frame............: 2008/7/1 to 2008/7/31
  Number of cases.......: 4
  Expected cases........: 0.087
  Observed / expected...: 45.73
  Test statistic........: 11.385874
  Monte Carlo rank......: 28/1000
  P-value...............: 0.028
17.Location IDs included.: 1577, 1889, 1885, 1848, 1467,
1890,
                                        1891
  Coordinates / radius..: (0.842300 N, 37.624920 E) /
```

```
11.28 km
  Time frame............: 2010/12/1 to 2010/12/31
  Number of cases.......: 4
  Expected cases........: 0.090
  Observed / expected...: 44.22
  Test statistic........: 11.254753
  Monte Carlo rank......: 35/1000
  P-value...............: 0.035
```

Appendix 2. Boosted Regression Trees model produced by tree complexity of 5 and a learning rate of 0.005



*This model was built with the default 10-fold cross-validation (CV) the solid black curve is the mean, and the dotted curves ± 1 standard error, for the changes in predictive deviance (ie as measured on the excluded folds of the CV). The red line shows the minimum of the mean, and the green line the number of trees at which that occurs. The final model that is returned in the model object is built on the full data set, using the number of trees identified as optimal.

Appendix 3. Summary of the ranked list of the 10 most important pairwise interactions

Rank list	var1.index	var1.names	var2.index	var2.names	int.size
1	11	meanndvi	1	altitude	44.8
2	11	meanndvi	8	povden	42.52
3	11	meanndvi	9	Land use	29.18
4	9	Land use	5	distparkbd	22.64
5	9	Land use	6	soils	17.62
6	12	stdndvi	1	altitude	15.48
7	3	distroads	1	altitude	12.13
8	11	meanndvi	5	distparkbd	11.25
9	5	distparkbd	1	altitude	9.81
10	12	stdndvi	9	Land use	8.18

Appendix 4. Collinearity and Variance Inflation Tests

```
 ##COLLINEARITY ANALYSIS IN R
source("brt.functions.R")
setwd("H:\\BRT_tutorial")
library(gbm)
model.data <- read.csv("H:\\BRT_tutorial\\collinearity.csv")
##model.data <-
read.csv("H:\\BRT_tutorial\\1006_1006_multicollinearty.csv")
model.data[1:3,]
## A first indication of collinearity would be to
## take a look at plots that map out
## each variable against the other variables.
## in R you can very easily plot out many variables at
## once with the plot() function
##plot(data.frame(model.data$distowns,model.data$distroads,model
.data$distrivers,model.data$distparks,model.data$soils,model.data$p
```

```
overtyindex,model.data$landuse,model.data$altitude,model.data$ann
ualrain))
##plot(data.frame(model.data$altitude,model.data$distowns,model.
data$distroads,model.data$distrivers,model.data$distparkbd,model.d
ata$soils,model.data$slope,model.data$popden,model.data$landuse,
model.data$popdens,model.data$meanndvi,model.data$stdndvi,mod
el.data$livstockden,model.data$elephantpo))
plot(data.frame(model.data$altitude,model.data$annurain,model.dat
a$distowns,model.data$distroads,model.data$distrivers,model.data$
distparkbd,model.data$soils,model.data$slope,model.data$popden,m
odel.data$landuse,model.data$popden,model.data$livestock,model.d
ata$meanndvi,model.data$stdndvi,model.data$elephantpo))
## the data.frame() command inside the plot() function
## creates a new data frame of the selected variables
## Create a matrix of scatter plots with variables that you
plot(data.frame(model.data$altitude,model.data$annurain))
plot(data.frame(model.data$landuse,model.data$meanndvi))
plot(data.frame(model.data$popden,model.data$povden))
## Are there any pairs of variables that you suspect on the basis of
their scatterplots?
## A second step in collinearity analysis is to check
## the correlation coefficient between pairs of variables
## for this you can use the cor() command.
##cor(d$tmed_ju,d$tmean)
cor(model.data$altitude,model.data$annurain)
cor(model.data$landuse,model.data$meanndvi)
cor(model.data$popden,model.data$povden)
##cor(model.data$tmed_ju,model.data$tmean)
###FACTORS
```

```
model.data$soils <- factor(model.data$soils, levels =
levels(model.data$soils))
model.data$landuse <- factor(model.data$landuse, levels =
levels(model.data$landuse))
model.data$slope <- factor(model.data$slope, levels =
levels(model.data$slope))
model.data$altitude <- factor(model.data$slope, levels =
levels(model.data$altitude))
model.data$slope <- factor(model.data$annualrain, levels =
levels(model.data$annualrain))
## Calculate correlation coefficients for each suspected pair.
## Are there any pairs that you suspect? Consider which one you
would eliminate from the database.
## What would be your conclusion on the basis of this first analysis?
## To continue we should calculate VIF values.
## to do so we should first fit a linear model that explains
## variation in one explanatory variable as a function of ALL other
## you can fit models using the lm() function
## the you can create a linear model for tmean as follows
model.altitude <-lm(altitude~annurain+ distroads+ distrivers+
distparkbd+soils+slope+povden+landuse
+popden+livestock+meanndvi+stdndvi+elephantpo,
data=model.data)
model.meanndvi <-lm(meanndvi~annurain+ distroads+ distrivers+
distparkbd+soils+slope+povden+landuse
+popden+livestock+altitude+stdndvi+elephantpo, data=model.data)
model.landuse <-lm(landuse~annurain+ distroads+ distrivers+
distparkbd+soils+slope+povden+meanndvi
+popden+livestock+altitude+stdndvi+elephantpo, data=model.data)
```

```
model.annurain <-lm(annurain~landuse+ distroads+ distrivers+
distparkbd+soils+slope+povden+meanndvi
+popden+livestock+altitude+stdndvi+elephantpo, data=model.data)
model.popden <-lm(popden~landuse+ distroads+ distrivers+
distparkbd+soils+slope+povden+meanndvi
+annurain+livestock+altitude+stdndvi+elephantpo,
data=model.data)
## as you can see, the "d$" part for addressing the variables is not
used this time.
## to see the specifics of the fitted regression model
## including R2, you can use the summary() function
## (the same you used for the data exploration)
##summary(model.tmean)
summary(model.altitude)
summary(model.meanndvi)
summary(model.landuse)
summary(model.annurain)
summary(model.popden)
## you can extract values from this summary by using the $ sign
## similar to how you used it for the extraction of variables
## from the data frame called "d"
##summary(model.tmean)$r.squared
summary(model.altitude)$r.squared
summary(model.meanndvi)$r.squared
summary(model.landuse)$r.squared
summary(model.annurain)$r.squared
summary(model.popden)$r.squared
## instead of just calling the value for R2
## you can assign it to a variable that can be used to calculate
further
```

```
##r2.tmean<-summary(model.tmean)$r.squared
r2.altitude<-summary(model.altitude)$r.squared
r2.meanndvi<-summary(model.meanndvi)$r.squared
r2.landuse<-summary(model.landuse)$r.squared
r2.annurain<-summary(model.annurain)$r.squared
r2.popden<-summary(model.popden)$r.squared
## This value can then be used to calculate the
## VIF value for this selection of variables, and this specific
##VIF.tmean<-1/(1-r2.tmean)
VIF.altitude<-1/(1-r2.altitude)
VIF.meanndvi<-1/(1-r2.meanndvi)
VIF.landuse<-1/(1-r2.landuse)
VIF.annurain<-1/(1-r2.annurain)
VIF.popden<-1/(1-r2.popden)
```

Appendix 5. Scripts for Boosted Regression Trees in R

```
source("brt.functions.R")
setwd("H:\\BRT_tutorial")
# This assumes that the brt functions file is located in the working
directory.
model.data <-
read.csv("H:\\BRT_tutorial\\1006_1006merge_21st_used.csv")
## Fitting a model
## These data have 2012 sites, comprising 1002 presence records
for the elephants (the ##command sum(model.data$elephants) will
give you the total number of presences).
angaus.tc5.lr01 <- gbm.step(data=model.data,
   gbm.x = c(2,3,4,5,6,7,8,9,10,11,12,13,14,15),
   gbm.y = 1,
   family = "bernoulli",
```

```
   tree.complexity = 5,
   learning.rate = 0.005,
   bag.fraction = 0.5)
# There will also be a graph..
# This model was built with the default 10-fold cross-validation (CV)
– the solid black curve is the mean, and the dotted curves ± 1
standard error, for the changes in predictive deviance (ie as
measured on the excluded folds of the CV). The red line shows the
minimum of the mean, and the green line the number of trees at
which that occurs. The final model that is returned in the model
object is built on the full data set, using the number of trees
identified as optimal.
 par(mfrow=c(3,4))
##gbm.plot(angaus.tc5.lr005, n.plots=12, write.title = F)
gbm.plot(angaus.tc5.lr01, n.plots=16, write.title = F)
# #Depending on the distribution of observations within the
environmental space, fitted functions can give a misleading indication
about the distribution of the fitted values in relation to each predictor.
The function gbm.plot.fits has been provided to plot the fitted values
in relation to each of the predictors used in the model.
gbm.plot.fits(angaus.tc5.lr01)
# #This has options that allow for the plotting of all fitted values or of
fitted values only for positive observations, or the plotting of fitted
values in factor type graphs that are much quicker to print. Values
above each graph indicate the weighted mean of fitted values in
relation to each non-factor predictor.
## Interrogate and plot the interactions
## This code assesses the extent to which pairwise interactions exist
in the data.
find.int <- gbm.interactions(angaus.tc5.lr01)
```

The returned object, here named test.int, is a list. The first 2 components summarise the results, first as a ranked list of the 10 most important pairwise interactions, and the second tabulating all pairwise interactions. The variable index numbers in $rank.list can be used for plotting.