Master’s Thesis
Regional Studies
Development Geography

SPECIES DISTRIBUTION MODELS EXPLAINING HUMAN-WILDLIFE CONFLICTS IN TAITA TAVETA COUNTY, KENYA

Sakari Äärilä
2017

Supervisors:
Mika Siljander, Tino Johansson, Petri Pellikka

UNIVERSITY OF HELSINKI
FACULTY OF SCIENCE
DEPARTMENT OF GEOSCIENCES AND GEOGRAPHY
DIVISION OF GEOGRAPHY

P.O. 64 (Gustaf Hällströmin katu 2)
FI-00014 University of Helsinki
Finland


Avainsanat – Nyckelord – Keywords
Villieläinkonflikti, Taita Taveta County, lajilevinneisyyssmallinnus, SDM, elefantti, leijona, leopardi, gepardi, hyeena

Säilytyspaikka – Förvaringställe – Where deposited
Helsingin Yliopisto, Kumpulan Tiedekirjasto

Muita tietoja – Övriga uppgifter – Additional information
Species distribution models explaining human-wildlife conflicts in Taita Taveta County, Kenya

In this thesis, species distribution modelling (SDM) approaches are used to examine and predict HWCs in Taita Taveta County in Southern Kenya. Data for SDM was collected in February-September 2016 from compensation forms filed by victims of HWC and it consists of all the reported conflicts in 2014 and 2015. Temporal distribution of HWCs is also examined through analysis made with another dataset obtained from Kenya Wildlife Service including reported HWCs between 1990–2016. The study focuses on conflicts involving elephants, lions, leopards, hyenas and cheetahs.

Species distribution modelling combines ecological theory and mathematical methods basing their results on the relationship between species and environment. This study uses biomod2 package in R, which includes 10 state-of-the-art modelling techniques: Generalized linear models (GLM), Generalized Additive Models (GAM), Generalized Boosted Models (GAM), Random Forest (RF), MaxEnt, Multiple Adaptive Regression Splines (MARS), Artificial Neural Networks (ANN), Classification Tree Analysis (CTA), Flexible Discriminant Analysis (FDA), Surface Range Envelope (SRE). Model prediction accuracy was estimated with True Skill Statistic (TSS) and Area Under the Curve (AUC). Most of the models reached moderate to good accuracy. Only human-cheetah conflicts were left out of the final analysis due to poor prediction accuracy.

Explanatory variables used in the final models were: distance to protected area, annual average precipitation, NDVI, population density, distance to water point, distance to river, distance to house and distance to road. HWCs involving different species were seen to be driven by different factors. Overall, distance to protected area, annual average precipitation and population density were selected as the most important variables determining the distribution of conflicts. The models were seen to be accurate and realistic in most cases. However, the models’ ability to be generalized in different areas is debatable and the models must be tuned for distinct regions separately.

Keywords: Human-wildlife conflict, Taita Taveta County, species distribution modelling, SDM, elephant, lion, leopard, hyena, cheetah
Contents
List of Abbreviations .................................................................................................................. I
List of Figures ............................................................................................................................. II
List of Tables ................................................................................................................................ III
1. Introduction ........................................................................................................................... 1
2. Research Questions .............................................................................................................. 3
3. Theoretical framework .......................................................................................................... 4
   3.1 Human-wildlife conflicts ................................................................................................. 4
      3.1.1 Human-Elephant Conflict ....................................................................................... 6
      3.1.2 Human-predator conflicts ....................................................................................... 9
      3.1.3 People affected by HWC ........................................................................................ 11
      3.1.4 HWC compensation schemes in Kenya .................................................................... 12
4. Methodological Framework ................................................................................................... 13
   4.1 Species distribution modelling (SDM) ............................................................................. 13
   4.3 SDM in HWC studies ..................................................................................................... 17
5. Study area .............................................................................................................................. 18
   5.1 Protected areas .............................................................................................................. 21
   5.2 Future prospects of Taita Taveta County ....................................................................... 22
6. Data ......................................................................................................................................... 23
   6.1 Human-wildlife conflict data ....................................................................................... 23
   6.2 Data for temporal distribution analysis ......................................................................... 24
   6.3 Environmental variables ............................................................................................... 24
7. SDM methods ......................................................................................................................... 27
   7.1 Modelling algorithms .................................................................................................... 27
      7.1.1 Generalized linear models (GLM) ............................................................................ 27
      7.1.2 Generalized additive model (GAM) .......................................................................... 27
      7.1.3 Multiple Adaptive Regression Splines (MARS) ........................................................ 28
      7.1.4 Flexible Discriminant Analysis (FDA) ..................................................................... 28
      7.1.5 Surface Range Envelope (SRE) ................................................................................. 28
      7.1.6 Classification Tree Analysis (CTA) ........................................................................... 29
      7.1.7 Generalized Boosted Model (GBM) .......................................................................... 29
      7.1.8 Random Forest (RF) .................................................................................................. 29
      7.1.9 Maximum Entropy (MaxEnt) ................................................................................. 30
      7.1.10 Artificial Neural Networks (ANN) .......................................................................... 30
   7.2 Model selection ................................................................................................................ 30
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Networks</td>
</tr>
<tr>
<td>AUC</td>
<td>Area Under the Curve</td>
</tr>
<tr>
<td>CTA</td>
<td>Classification Tree Analysis</td>
</tr>
<tr>
<td>FDA</td>
<td>Flexible Discriminant analysis</td>
</tr>
<tr>
<td>FEWS</td>
<td>Famine Early Warning System</td>
</tr>
<tr>
<td>GAM</td>
<td>Generalized Additive Models</td>
</tr>
<tr>
<td>GBM</td>
<td>Generalized Boosted Models</td>
</tr>
<tr>
<td>GLM</td>
<td>Generalized Linear Models</td>
</tr>
<tr>
<td>HWC</td>
<td>Human-Wildlife Conflict</td>
</tr>
<tr>
<td>IUCN</td>
<td>International Union for Conservation of Nature</td>
</tr>
<tr>
<td>KWS</td>
<td>Kenya Wildlife Service</td>
</tr>
<tr>
<td>LUMO</td>
<td>LUMO Community Wildlife Sanctuary</td>
</tr>
<tr>
<td>MARS</td>
<td>Multivariate Adaptive Regression Splines</td>
</tr>
<tr>
<td>MaxEnt</td>
<td>Maximum Entropy</td>
</tr>
<tr>
<td>NDVI</td>
<td>Normalized Difference Vegetation Index</td>
</tr>
<tr>
<td>NP</td>
<td>National Park</td>
</tr>
<tr>
<td>PA</td>
<td>Protected area</td>
</tr>
<tr>
<td>RF</td>
<td>Random Forest</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiving Operating Characteristic</td>
</tr>
<tr>
<td>SDM</td>
<td>Species Distribution Modelling</td>
</tr>
<tr>
<td>SRE</td>
<td>Surface Range Envelope</td>
</tr>
<tr>
<td>TSS</td>
<td>True Skill Statistic</td>
</tr>
<tr>
<td>THWS</td>
<td>Taita Hills Wildlife Sanctuary</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.</td>
<td>Elephants inside LUMO Community Wildlife Sanctuary</td>
<td>7</td>
</tr>
<tr>
<td>Figure 2.</td>
<td>Trees Gnawed by elephant</td>
<td>9</td>
</tr>
<tr>
<td>Figure 3.</td>
<td>Lions inside Taita Hills Wildlife Sanctuary</td>
<td>10</td>
</tr>
<tr>
<td>Figure 4.</td>
<td>Examples of model response curve, ecological niche, and geographical niche</td>
<td>14</td>
</tr>
<tr>
<td>Figure 5.</td>
<td>Different types of models</td>
<td>16</td>
</tr>
<tr>
<td>Figure 6.</td>
<td>Species distribution modelling step by step</td>
<td>17</td>
</tr>
<tr>
<td>Figure 7.</td>
<td>Agriculture in Taita Hills</td>
<td>19</td>
</tr>
<tr>
<td>Figure 8.</td>
<td>The study area</td>
<td>20</td>
</tr>
<tr>
<td>Figure 9.</td>
<td>Environmental variables used in the models</td>
<td>26</td>
</tr>
<tr>
<td>Figure 10.</td>
<td>Example of a receiver operating characteristic (ROC) plot</td>
<td>33</td>
</tr>
<tr>
<td>Figure 11.</td>
<td>Average variable importance for elephant crop raiding models</td>
<td>36</td>
</tr>
<tr>
<td>Figure 12.</td>
<td>GBM response curves for elephant crop raiding</td>
<td>37</td>
</tr>
<tr>
<td>Figure 13.</td>
<td>Majority vote prediction of elephant crop raiding</td>
<td>38</td>
</tr>
<tr>
<td>Figure 14.</td>
<td>Seasonal elephant crop raiding kernel densities</td>
<td>38</td>
</tr>
<tr>
<td>Figure 15.</td>
<td>Averaged variable importance for all human-lion conflict models</td>
<td>40</td>
</tr>
<tr>
<td>Figure 16.</td>
<td>GBM response curves for human-lion conflicts</td>
<td>41</td>
</tr>
<tr>
<td>Figure 17.</td>
<td>Majority vote prediction of human-lion conflict</td>
<td>42</td>
</tr>
<tr>
<td>Figure 18.</td>
<td>Seasonal kernel density maps of human-lion conflict</td>
<td>42</td>
</tr>
<tr>
<td>Figure 19.</td>
<td>Averaged variable importance of all human-hyena conflict models</td>
<td>44</td>
</tr>
<tr>
<td>Figure 20.</td>
<td>GBM response curves for human-hyena conflicts</td>
<td>45</td>
</tr>
<tr>
<td>Figure 21.</td>
<td>Majority vote prediction of human-hyena conflict</td>
<td>46</td>
</tr>
<tr>
<td>Figure 22.</td>
<td>Seasonal kernel density maps of human-hyena conflicts</td>
<td>46</td>
</tr>
<tr>
<td>Figure 23.</td>
<td>Averaged variable importance of all human-leopard conflict models</td>
<td>48</td>
</tr>
<tr>
<td>Figure 24.</td>
<td>GBM response curves for human-leopard conflicts</td>
<td>49</td>
</tr>
<tr>
<td>Figure 25.</td>
<td>Majority vote prediction of human-leopard conflict</td>
<td>50</td>
</tr>
<tr>
<td>Figure 26.</td>
<td>Seasonal kernel density maps of human-hyena conflict</td>
<td>50</td>
</tr>
<tr>
<td>Figure 27.</td>
<td>Monthly distribution of HWC in Taita Taveta County 1990–2016</td>
<td>51</td>
</tr>
<tr>
<td>Figure 28.</td>
<td>Agricultural calendar of of Taita Taveta County</td>
<td>51</td>
</tr>
</tbody>
</table>
List of Tables

Table 1. Confusion matrix 32
Table 2. Averaged AUC and TSS values of all elephant models 35
Table 3. Statistical characteristics of environmental variables used in the human-elephant conflict models 36
Table 4. Averaged variable importance for all elephant crop raiding models 36
Table 5. AUC and TSS scores of human-lion conflict models 39
Table 6. Statistical characteristic of environmental variables used in the human-lion conflict models 40
Table 7. Variable importance for all human-lion conflict models 40
Table 8. AUC and TSS scores of human-hyena conflict models 43
Table 9. Statistical characteristic of environmental variables used in human-hyena conflict models 44
Table 10. Variable importance for all human-hyena conflict models 44
Table 11. AUC and TSS scores of human-leopard conflict models 47
Table 12. Statistical characteristic of environmental variables used in the human-leopard conflict models 48
Table 13. Variable importance for all human-leopard conflict models 48
1. Introduction

International Union for the Conservation of Nature (IUCN) defines human-wildlife conflict (HWC) as an incidence where wildlife’s needs become incompatible with those of human populations, with costs both to humans and wild animals (IUCN 2005). HWCs include crop and livestock damage, human injury and death, as well as the spread of disease from wildlife to livestock and humans. HWCs can also be manifested in the form of infrastructure damages and negative interaction with other valuable species (Lamarque et al. 2009). As such, HWCs are a global phenomenon as old as human history.

While HWCs occurs on all continents, the vulnerability to the incidents varies across different areas. For example, communities relying on agriculture and livestock husbandry in the developing countries are much more vulnerable to HWC than more wealthy people in the Global North. While material damages are felt stronger in the developing countries, the threat to human lives is not to be underestimated anywhere in the world.

Invertebrates and small animals like rodents cause more damage on agriculture on a national scale (Lamarque et al. 2009) but the HWC discussion is usually concentrated around large mammals. This is likely due to large mammals being able to cause significant threat to humans and severe stress on food security on a household level, causing major damage over short periods of time.

The problem of HWC is also linked to wildlife conservation. Many times, these large wildlife species like elephants, feline predators or wolves are also protected by legislation. In rural Africa, the species that are held in high respect globally can cause significant damage locally. Across Africa the discontent seems to culminate in elephants (see Barnes 1996, Wunder 1997, Woodroffe et al. 2005, Lamarque et al. 2009) but also the large predators, such as lions, are widely disliked for livestock depredation. The HWC damage can lead to retaliatory killings (Kissui 2008), which contradict with the aims of wildlife conservation. One of the biggest challenges for conservationists is indeed, how to protect species that are highly valuable globally but have negative impacts locally. Transforming the global value of nature conservation into local benefits and tolerance against HWC is essential for successful conservation.

As noted, the forms of HWC are diverse. This thesis focuses on crop raiding by elephants and livestock depredation and human damage by large predators. In Eastern Africa, these incidents are seasonal, partly driven by the availability of food and water. According to the optimal foraging theory (Krebs & Davies 1991), animals will maximize the quality of their nutrient intake when
possible. During the dry season, the natural forage of elephants is worsening in quality and quantity, while cultivated crops are ripening. Being a generalist feeder elephants will eat whatever is available (Osborn 2004). Therefore, it is not surprising that elephants are raiding agricultural crops. Livestock depredation is seen to increase during seasonal rains (Patterson et al. 2004, Lamarque et al. 2009), when natural prey is more difficult to find. Scarce water points serve as good hunting grounds during dry season but with the rains, the prey animals are scattered more widely across the area. Livestock in Kenya is often either herded in large quantities or left unattended in small herds tied to trees, which makes these animals an easy prey for predators.

As wide ranging and dire phenomenon, HWC has been of interest to scientific research for decades. In recent years, the advancements in computational power and remote sensing methods have allowed new kinds of research methods to be utilized in spatial studies about HWC, as data is more abundant and precise and the means to process it are better. The advancements in technology have also increased the use of species distribution modelling (SDM) approaches in conservation studies and ecology. SDM is commonly used to determine species ranges, potentially extrapolating SDM data in space and time (see Franklin & Miller 2010). However, the use of SDM in HWC studies is still relatively rare, but it offers a potentially powerful tool for research. One of the objectives of this study is to assess how species distribution modelling approach can be used in studying HWCs.

In this thesis, I attempt to form models of HWC from countywide data from Taita Taveta County, Kenya, including all reported HWC incidents between 2014–2015. The models can be then used to predict the most HWC prone areas in the county and to find the importance of different anthropogenic and environmental factors impacting the distribution of HWC. Rather than traditionally determining the species range, this thesis attempts to predict landscapes where human and wildlife are most likely to come into conflict. Different species have different kinds of foraging and predation behaviors. Consequently, models had to be made separately for each, to determine which factors are central to different kinds of HWCs.

Human communities encroaching traditional wildlife habitats, fragmenting and degrading wildlife species range create the basis for the HWCs. In Africa, this is intensified by growing population changing the landscape and population pressure of wildlife in some areas (Lamarque et al. 2009). In the study area, Taita Taveta County of Kenya, the national parks (NP) cover over 62% of the land (Taita Taveta County Government 2013), acting as safe zones for wildlife. Wildlife roams also outside of the parks, sometimes causing dire impacts on nearby communities. Research on and definitions of this kind of human–wildlife interfaces is central to the mitigation of HWC to strengthen the food security, safety, and conservation efforts globally, regionally and locally.
2. Research Questions

1) Can species distribution modelling methods be used to model and predict HWC incidents in the research area?

This thesis focuses on modelling HWCs involving six different species: Elephants (*Loxodonta africana*), lions (*Panthera leo*), cheetahs (*Acinonyx jubatus*), leopards (*Panthera pardus*) and hyenas (*Crocuta crocuta, Hyaena hyaena & Hyaena brunnea*). The choices in SDM varies between the species and some species are more challenging to model than others. Hypothesis for this question is that HWCs involving most of the studied species can be modelled accurately with choice of variables based on earlier literature.

2) How are the human-wildlife conflicts distributed spatially in the research area?

Determining the importance of environmental and anthropogenic factors in the distribution of HWCs is central to mitigating the phenomenon. This way the root causes of HWC could be exposed to minimize the impacts these conflicts impose on local communities. This could increase food security and gain more local support for conservation work. Hypotheses for this question are that the HWCs are concentrated near borders of protected areas (PA) in areas with relatively abundant vegetation. Human-elephant conflicts are also expected to intensify in areas with higher annual precipitation, close to rivers and water points. Human-predator conflicts are also expected to concentrate in areas with higher precipitation but further away from the water points and rivers. The conflict risk for all species is likely to be smaller in areas with denser human populations.

3) How are the human-wildlife conflicts distributed temporally in the research area?

Seasonal changes in the study area have a possible impact on the distribution of the conflicts. The human-elephant conflicts are expected to concentrate in times of ripening crops in the shifts from rainy to dry season. Human-predator conflicts are expected to increase during rainy season because natural prey is not as easily available around water points.
3. Theoretical framework

This study draws from the tradition of cultural animal geography, a term coined by Charles F. Bennet Jr. (1960). Animal geography examines the present and historical distributions of animals and tries to understand the environmental dynamics behind them. Cultural animal geography focuses on the interactions of animals and human cultures. Humans have domesticated, killed and transported animal species for its entire history impacting the species distributions globally. On the other hand, animals have contested with humans for resources and space, eating livestock and human cultivated crops as well as causing human fatalities and spread of diseases. The role of animals in evolution of space, region and landscape is also central to cultural animal geography, as well as questions on how the animals impact the human opportunities and risks.

3.1 Human-wildlife conflicts

Defined by the International Union for the Conservation of Nature (IUCN, 2005) human-wildlife conflict is an event where wildlife requirements encroach on those of human populations, with costs to both, humans and wild animals. HWCs include human injuries and deaths, crop damages, predation of livestock, transmission of diseases and damage on infrastructure (Lamarque et al. 2009). Also, adverse interaction across different wildlife species resulting in loss of valuable or endangered species can be seen as a form of HWC. HWCs have existed as long as humans have walked the earth and it is not seen as a phenomenon that could be eradicated. This calls for peaceful coexistence between human populations and wildlife. Emphasis is put on the solution of increasing the local tolerance against HWCs (Naughton-Treves 1998). This is done, for example, in form of compensation schemes, new preventive measures, or wildlife awareness projects. However, human populations are growing and land is getting scarcer in many places leading to continuous encroachment of human settlement into areas previously inhabited by wildlife. The increasing demand for land and resources are fragmenting and degrading wildlife habitats. The competition for same resources and space, leading to increased human-wildlife interface, is indeed seen as one of the main factors behind HWCs (Lamarque et al. 2009). HWCs are present in all kinds of environments, on land, in water, in rural settings and in cities. However, HWCs pose a significant threat to food security in areas that depend on rain-fed subsistence agriculture. It is stated that the conflicts are also more intense in areas where people rely on agriculture and owning livestock.
Generally, HWCs are more common inside and around protected areas, where wildlife is more abundant (Distefano 2005).

In Akama et al. (1995) study conducted near Tsavo and Nairobi National Parks in Kenya, nearly all of the respondents had experienced crop raiding by wildlife. Mackenzie & Ahabyona (2012) studied the costs of crop raiding to farmers in Uganda and presented the average financial losses of farmers over six-month period being 74 USD. The respondents of the study estimated losing 30% of staple crop yield to crop raiding. The damage was seen highest to be within 500 meters from protected area boundaries and affecting the household food security significantly. While average losses are at best just estimates of the costs of HWC, the blight of a farmer who loses all of the season’s crops over a single night, is not exposed from the numbers. Also, fatalities in families caused by HWCs can lead to significant changes in household dynamics, resulting in increased debt and poverty.

HWCs are also a major threat to the success of wildlife conservation, turning local communities against it. Anthony (2007) states that there is a paradox within wildlife conservation, where successful conservation potentially leads to increased human-wildlife conflicts, which in turn has a negative influence on local attitudes towards the animals and conservation. Kissui’s (2008) study in Tanzania, concluded that the livestock predation was provoking pastoralists into retaliation killings of lions, hyenas and leopards. Also, the wildlife authorities are sometimes forced to kill problem animals. Some of the species involved in HWCs, like elephants and big feline predators, are also seen as the flagship species of tourism and even if seen as pests locally, they are highly valued globally. This makes them understandably the focus of many HWC studies in Africa.

Another significant driver behind the HWCs is water availability and accessibility. The research area has two rainy seasons: March-June and October-December (National Drought Management Authority 2017), which have an influence on human-wildlife conflicts. During dry seasons, herbivores tend to move out of the protected areas to look for water and forage. Water sources near human settlements are usually unreachable for wildlife without coming into conflict with humans (Lamarque et al. 2009). Journeys to fetch water also put people at risk of conflict with wildlife. For livestock depredation, rainy seasons are seen as time when the predators’ natural prey is not so easily available around the water points. During rainy season water is more evenly distributed and the prey animals move in smaller herds in broader areas (Lamarque et al. 2009). Hence, the predators are drawn to closer contact with humans and their livestock to feed.

Studies of HWC have traditionally focused on perceptions of farmers, derived by interviewing the affected people (Naughton-Treves 1998). While possibly giving valuable insight, these oral
testimonies are prone to inaccuracies and exaggerations. Naughton-Treves (1998) also notes that inaccuracies are also present when trying to extrapolate observations of a single study site into larger extent of an entire protected area.

Studies (Naughton-Treves 1998, Johansson 2008) claim the proximity of forest cover to be one variable significant to the risk of HWC. Forests are seen as “refugias” for wildlife, providing resources, shelter and freedom from human disturbance. Considering the benefits provided by forest refugias, one can assume protected areas offering same kinds of advantages for wildlife. While a lot of wildlife roams outside of the protected areas, most are still concentrated inside them “leaking” to the outside territories. Consequently, farm’s or livestock herd’s proximity to protected area boundary is an obvious factor increasing the risk of HWC.

3.1.1 Human-Elephant Conflict

African savannah elephant is the largest (2200-6300 kg) living ground mammal with wide ranges around sub-Saharan Africa. Elephants are also often seen as the biggest pests when it comes to HWCs and are also expectedly the focus of many HWC studies. While rodents and invertebrate pests can cause more harm on a regional level (see Lamarque et al. 2009, Woodroffe et al. 2005), elephants can trample whole fields of agricultural products overnight, causing biggest impact on a single foray (Lamarque et al. 2009, Woodroffe et al. 2005). Consequently, elephants pose a significant threat to farming communities near the protected areas and risking food security on a household level (Chiyo et al. 2005).

Elephants can occasionally injure or kill humans. Elephants do not attack people intentionally but these incidents usually happen while people are protecting their fields from crop raiding or otherwise coming accidentally into conflict with the animals (Lamarque et al. 2009). Elephants were recorded to have killed over 200 people during 2000–2007 in Kenya (WWF 2007). However, these incidents are relatively rare compared to the frequency of crop raiding.
Figure 1. Elephants inside LUMO Community Wildlife Sanctuary (Äärilä, Sakari 2016).

Being a generalist feeders, elephants eat what is available for them. However, elephants do usually choose the food with highest rate of nutrient intake (Osborn 2004). Grass and woody browse are the main types of food for elephants (Osborn 2004). When grass matures, its nutrient and water content decreases. Elephants switch their preference to browse in dry season for higher nutritional value (Wyatt & Eltringham 1974). Also, crop raiding by elephants is seen to increase at the end of rains and in the beginning of dry season when the quality of their natural forage declines and agricultural crops are ready for harvest (Chiyo et al. 2005). These finding are in line with the optimal foraging theory, which states that animals will maximize the quality of their nutrient intake when possible (Krebs & Davies 1991). Osborn (2004) suggests that the trigger for crop raiding can indeed be predicted from rainfall patterns and rate of grass growth. The problem of human-elephant conflict is widespread, being found where the ranges of elephants and humans overlap. In addition, it does not seem to matter if the elephants are protected or not (Hoare 2000).

Crop raiding is mainly done by bull elephants after they have separated from their natal families (Chiyo & Cochcrane 2005). Crop raiding happens usually in groups of one to eight elephants and the tendency of males to raid crops more often has been linked to mating competition, where the
nutritional advantage from raiding is seen beneficial (Chiyo & Cochcrane 2005). Easy access, high nutritional value, palatability and lower secondary defenses of human grown crops are considered to draw elephants to crop raiding (Sukumar 1990). Sukumar’s study was done on Asian Elephants (*Elephas maximus*), but without doubt, the same resources attract their African counterparts near human settlements. Elephants, especially bulls, can also develop “tolerance” against human disturbance (Hoare 1999) and habitual behavior of crop raiding (Lahm 1996). Hence, crop raiding risk areas can result from few habitual raiders or number of occasional raiders. Usually elephant crop raiding takes place at night time and the gender of the individual is hard to determine in dark. Hoare (1999) also speculates that the unpredictability of elephant crop raiding results from its relativity to behavior of individual elephant bulls.

Elephant populations in Taita Taveta County have steadily increased since the 1980’s and a total of 12 573 elephants were counted during the 2011 aerial animal counts (Ngene et al. 2013). During these aerial counts, 31% of the elephants were counted outside of the national parks. Indeed, the actual living ranges of elephants are not confined by human-made boundaries and many of them reside outside the protected areas (Lamarque et al. 2009).

High density of large herbivores in a confined conservation area may cause significant impact also on vegetation cover. As noted before, during dry seasons, large herbivores migrate after water and forage, eating also much more vegetation than during rainy seasons. Fenced borders of the protected areas in Taita Taveta County can be seen from remotely sensed imagery due to reduced vegetation. Here lies a problem of fitting large ecosystems functions in a bounded space of a national park. A large forested area that was designated into a protection zone for elephants and other wildlife in 1948 turned into a deforested plain in 30 years (Robbins 2012, Ngene et al. 2011, see also Laws 1970). In the study area, a severe drought in the 1990s caused decline in the parks’ carrying capacity (Waithaka 1997), which has partly lead to wildlife migrating for crucial resources elsewhere. Wildlife and droughts are not the only factors contributing to the environmental degradation. In Taita Taveta County livestock is herded inside the NPs illegally, which increases the grazing pressure on the land and forces wildlife to seek food and water from outside the PAs. The illegal grazing inside the parks was a common topic when discussing HWC with people in Taita Taveta County but there has not been actual research on it.
3.1.2 Human-predator conflicts

Due to poaching (including illegal killing by poisoning) and other human actions degrading and fragmenting wildlife habitat, large predators are in worldwide decline (Woodroffe 2000). Large carnivores tend to have large living ranges and their predatory habits bring them frequently into conflict with humans. While human injuries and deaths are relatively rare, savannah predators cause significant harm to livestock keepers in Kenya. The damage and threat they cause to humans and livestock is also apt to complicate the conservation of these predators. Human injuries and stress caused to food security by continuous livestock raiding behavior can lead to retaliation killings (Frank et al. 2005). Therefore, it is important to understand the phenomenon and examine the ways to mitigate the damage to improve rural food security and conservation outcomes.

Predators have typically large ranges and their fragmentation caused by human action is seen to intensify the tendency to prey on livestock (Patterson et al. 2004). For livestock predation, lions are seen as the biggest concern. Patterson et al. (2004) studied livestock damages in ranches next to

Figure 2. Trees gnawed by elephants. A common sight in Taita Taveta County, Kenya (Äärilä, Sakari 2016).
Tsavo East National Park and lions were responsible for 86% of attacks while the rest of the attacks were done by hyenas and cheetahs. Their study also concluded that the attacks peaked in November and their number was high also in January, April and May. The rainy seasons in the area are March to July and October to January (National Drought Management Authority 2017). They also concluded that the livestock densities correlate with the number of predatory attacks by wildlife. As noted before, Lamarque et al. (2009) have reached similar conclusion: predation on livestock increases during rainy season when natural prey is not easily available. During dry seasons, scarce water points are good hunting grounds for predators. Livestock depredation typically happens during night time when predators are most active and an attack typically results in a single head of livestock killed. However, it is not unusual for predators to kill more than they can eat. This kind of surplus killing happens especially once a predator gets inside a livestock shelter (Patterson et al. 2004, Larmarque et al. 2009).

Figure 3. Lions inside Taita Hills Wildlife Sanctuary. (Vento, Eero 2016)

Cheetahs typically go for smaller prey such as goats, sheep or calves (Patterson et al. 2004). Leopards mainly hunt for small and medium sized natural prey (Stein et al. 2015) and this preference could be generalized to livestock predation behavior too. Predators hunting in groups,
like lions and hyenas, go normally for larger prey such as cattle and donkeys. These finding go parallel with their behavior with natural prey (Patterson et al. 2004). Losses to diseases and parasites exceed the damage caused by predators to livestock (Patterson et al. 2004), but impacts caused to a single household by livestock depredation can be overwhelming and cause severe distress in food security.

Hopcraft et al. (2005) presented two theories on the hunting ground selection by carnivores. They are also based on natural prey but can be thought to reflect on the livestock depredation behavior as well. Prey abundance hypothesis suggests that predators prefer hunting in areas with high concentrations of prey. Opposing theory is called landscape hypothesis, which proposes predators prefer hunting in areas with good cover rather than high prey density. It is probable the actual behavior is dictated by both.

Predator occurrence affects the presence of other predatory species as they tend to avoid each other. Especially presence of top predators like lions has been observed to decrease the chance of occurrence of other predatory species (e.g. hyena) in the area (Ramesh et al. 2017, Brook et al. 2012). This kind of inter-predator interaction may play a role also in HWC distributions involving different species.

Humans are also vulnerable to large carnivore attacks. Human injuries are relatively rare compared to livestock depredation but the consequences are more severe. Death or an injury to a family member may lead to increased household poverty and debt, and force children out of school to help family subsistence. Predator attacks on people can happen for example while guarding livestock, fetching water or otherwise moving outside, especially during night-time (Lamarque et al. 2009). Wildlife is seen particularly dangerous when offspring is present, when injured or if the animal is protecting its food source (Johansson 2009, Larmaque et al. 2009).

### 3.1.3 People affected by HWC

Most disadvantaged by HWCs are people with low living standards and agro-pastoralists whose income depends on their land and livestock (Lamarque et al. 2009). This is linked to the wider discussion in the field political ecology, where it is pointed out that environmental issues have heavier impacts on already disadvantaged people. This in turn can reinforce the pre-existing inequalities (Robbins 2012). In the study area of Taita Taveta County, 57.2% of the population lives
below the poverty line, earning less than 1562 KSH (14,3 EUR) per month (Taita Taveta County Government 2013).

Famine Early Warning System (FEWS) is a US founded provider of early warning and analysis on acute food security (FEWS net 2017). They have developed a five-phase scale of food security from minimal (1) to famine (5). As of April 2017, FEWS has estimated the food security situation in Taita Taveta County to be stressed (phase 2) because of the delayed long rains. The poorest households are estimated to even face food security crisis (phase 3). The situation is dictated mostly by the annual precipitation and the timing of the rains, but the HWCs do have a significant impact on already stressed food security of individual households.

Wildlife does not recognize man-made boundaries and some species have very large ranges. Proximity of protected area is an obvious factor contributing to increased risk of HWCs. It is common for the poorer people to dwell closer to the protected areas because of the scarcity of land in safer areas (Lamarque et al 2009). The study area of Taita Taveta County is surrounded by two large national parks, which act as safe zones for wildlife and their long borders make wide areas vulnerable to HWCs.

As response to HWCs, negative attitudes towards conservation arise and retaliation killings occur (Kissui 2008). Groom and Harris (2008) studied attitudes towards wildlife conservation on community owned lands in Kenya. The article mentions that the most influential negative impacts were human and livestock damages (negative perception towards carnivores), Spreading of disease, crop damage and competition for resources (negative perceptions towards herbivores). Increasing local tolerance against the HWC, especially close to the PAs, is a major key in amplifying the conservation of wildlife.

3.1.4 HWC compensation schemes in Kenya

Kenya used to have a national policy on compensation for crop and livestock damage caused by wildlife until 1989. Compensation is still paid for human injury and death but the amount is not seen as sufficient. The compensation policy for crop damage caused by wildlife was also claimed to be inefficient, corruption-prone and unworkable (Human Elephant Taskforce 2001). The IUCN African Elephant Specialist Group is against monetary compensation for elephant damage, claiming it is unable to decrease the level of the problem.
With the Wildlife Conservation and Management Act (2013), a new compensation scheme, was introduced to Kenyans. It allows one to claim compensation in cases of human damage, crop or property destruction and killed livestock. So far, this compensation scheme has not been active and no compensation has been paid since its establishment. The demands will be processed by a compensation committee but it is yet to be formed. Data used for SDM in this thesis was collected from these pending compensation claims. Damage caused by all wildlife species is not included in the new compensation scheme. For example, damages done by primates are not compensated and for this reason, the incidents are also absent from data.

4. Methodological Framework

4.1 Species distribution modelling (SDM)

Species distribution modelling is used to obtain ecological knowledge and to predict species distributions across landscapes combining ecological theory and mathematical methods. SDM techniques are based on relationship between species observations and environment (Elith & Leathwick 2009, Franklin & Miller 2010). Using different algorithms, SDMs can reveal factors affecting distribution of species and are able to extrapolate the distribution, as function of environmental surroundings of species (Guisan & Zimmerman 2000, Franklin & Miller 2010). It is widely accepted that with well-reasoned variables and a good research design SDM algorithms provide good predictive capability and understanding on species’ distributions (Elith & Leathwick, 2009). Advances in remote sensing, GIS and computer hardware have brought major improvements in the species distribution models, easing the collection and analysis of enormous amounts of data on large areas. This has contributed to increased use of SDM in ecological research in the last decades (Franklin & Miller 2010). Today, SDMs are seen as important tools in conservation biology, ecology, biogeography, evolution and climate change research (Guisan & Thuiller, 2005). The use of SDM is still relatively rare in HWC studies, although the conflicts are tightly linked into animal ranges and species conservation. By taking in consideration social aspects involving HWCs, SDMs can also provide valuable information in the field of development studies and development geography in particular.

Central concept to SDM is the species’ niche, for which there are numerous definitions and a lot of literature concentrating around the concept (see Araújo & Guisan 2006, Franklin & Miller 2010, Peterson et al. 2011). However, the most used definition for species’ niche in SDM studies is likely
from Hutchinson (1957). He defined species’ fundamental niche to be a n-dimensional hypervolume in environmental space in which the species can exist indefinitely. As such, fundamental niche is the range of environmental conditions that a species is able to tolerate. Realized niche is the segment of the fundamental niche that the species actually occupies, taking the biotic interactions (i.e. interactions between living things) into account. Each variable used in the SDM algorithms accounts for a dimension in the Hutchinson’s n-dimensional hypervolume and the response the species has on the variables, defines its fundamental niche. Niche can be also divided into ecological and geographical niches. As stated, the ecological niche is what SDM produces by defining the ecological necessities of the species (e.g. certain range of temperature and certain amount of annual rainfall) and the geographical niche is these necessities predicted into geographical space (i.e. areas that match the species’ needs). In other words, once the response of the species is obtained with a model, its potential distribution can be predicted (see fig. 4).

**Figure 4.** Examples of model response curve (A), ecological niche (B) and geographical niche (C). Also examples of different kinds of response fits and their effects on the niches are presented. (Pearson 2014).

Modelling species distributions include underlying assumptions that must be considered when doing research on topic. Fundamental assumption in SDM is presuming the modelled species (or
phenomenon in this case) being in equilibrium with its environment (Franklin & Miller 2010) as opposed to still spreading in a new habitat (invasive species). With HWCs, it can be assumed that the phenomenon has spread where it possibly can, because HWCs are as old as history and the studied species are not invasive. However, it can be also argued that the land use and climate change are simultaneously transforming the species ranges. Other basic assumption in SDM is that the habitat is actually dictated by the environmental requirements and tolerances (Franklin & Miller 2010) and the relationships are not just random correlations without causation. Most of the statistical SDM techniques assume also that the data is independent. However, spatial autocorrelation is often found in spatial phenomena.

In geography everything is related to everything else, while near things are more related than distant things (Tobler 1970). This phenomenon is called spatial autocorrelation and means that the spatially autocorrelated variable correlates with itself. Algorithms used in SDMs usually assume variables being independent, meaning not dependent on each other or correlated with itself. Spatial autocorrelation might bias the models and impede finding out variable’s importance (Franklin & Miller 2010). Taking spatial autocorrelation into account while working with SDMs is crucial, even if it would mean only pointing out its existence.

SDM methods can fit the response to the data extremely closely, i.e. overfit the model (see fig. 4). When model is overfit, the niche is defined too strictly and it can become challenging to predict the geographical niche accurately and to extrapolate the model in space and time (Franklin & Miller 2010). Evaluating the models by cross-validation tries to detect and overcome this problem. However, this works when the data points are independent from one another. The data can be spatially autocorrelated and the testing data is not independent from the calibration data. This must be kept in mind when doing SDM and the results must be read cautiously.

Environmental variables being correlated with one another is called multicollinearity. Multicollinearity of the environmental variables means that the same data is entered into the model multiple times. Multicollinearity of explanatory variables might lead to rendering of causal variables to be insignificant (Franklin & Miller 2010). Of the multicollinear variables, only the variables that correlate strongest with the response variable should be left in the models.

The data of explanatory variables used in SDM can vary from very coarse scale data to very high resolution data. The resolution of the data used should always be adjusted to the studied phenomenon or the scale of the desired prediction. According to Guisan & Thuiller (2005) the predictors are chosen to reflect the three types of influence they might have on the species: 1)
Limiting factors, for example, temperature. 2) Disturbances, meaning natural or human factors causing perturbation on the environment. These could include also competition and predation. 3) Resources, in wildlife studies usually meaning food, water, and shelter. Relationships of these different kinds of variables and the species occurrence can also be hierarchical, leading the different spatial patterns to be observed depending on the chosen scale.

Species distribution models are always simplified models of complex reality. Optimal model includes all the significant factors affecting the phenomenon, without making the model overly complex. Trade-offs between model realism, precision and generality (ability to be generalized into other areas) are made when modeling species distributions (Levins 1966). Two of these can be increased at a time, decreasing the third one (see fig. 5). Having emphasis on generality and precision, produces an analytical model. Precision and realism generate an empirical model and mechanistic model is an offspring of realism and generality (Guisan & Zimmerman, 2000).

![Figure 5](image.png)

**Figure 5.** Different types of models. After Levins (1966) and Guisan & Zimmerman (2000).
Figure 6. Species distribution modelling step by step.

When predicting the HWC prone areas, the areas that seem to be favorable for the species due to similar environment does not necessarily mean a threat of a human-wildlife conflict. As said, the models are simplifications and do not include anything but the given input variables. For example, the model does not take into account actual behavior of the species or geophysical barriers. Using SDMs to model HWCs need also a set of inputs that describe the human-actions in the area. Human-wildlife conflicts cannot occur unless the area is populated or used by humans.

4.3 SDM in HWC studies

SDM is widely used in ecological studies to explain and predict species’ distributions and impacts of environmental change on these species. Research on the methods is continuous, and new models and applications are developed constantly (see for example Guisan & Zimmerman 2000, Anderson et al. 2003, Phillips et al. 2006, Pearson et al. 2006). Species distribution modelling has been applied in HWC studies but the use is still limited. Most studies on HWC rely on oral testimony of the farmers and livestock herders (Naughton-Treves 1998) making these prone to inaccuracies. Furthermore, the studies using SDM approaches to examine human-wildlife conflicts mainly focus on large predators and studies using SDM on elephant conflicts are almost non-existent.

Zarco-González (2013) used several algorithms to model cougar (*Puma concolor*) and jaguar (*Panthera onca*) attacks on livestock in Mexico, reaching good accuracy in predictions. However,
even after making models with high predictive capability they conclude, that even though results are potentially useful, models should be tuned separately for each area they are applied to. Carvalho et al. (2015) used multiple algorithms to predict livestock damages caused by jaguars in Brazil. Seven environmental variables were used in the study: Distance to deforested areas, cattle density, distance to roads, elevation, slope, distance to forest and distance to settlement. They also reached high accuracy with the models and proceeded to make a weighted ensemble model by combining them. The algorithms used in the study were MaxEnt, Genetic Algorithm for Rule-set Production, Support Vector Machines, Environmental Distance and Ecological Niche Factor Analysis. The models reached moderate to good prediction accuracy. The impact of anthropogenic factors were underlined in the paper, seen as dynamic, changing over time, following shifts in deforestation and livestock herds. Ficetola et al. (2013) used SDM to study wild boar damages in Southern Italy. Using a single algorithm (MaxEnt), they concluded that the final model’s capability to predict high risk zones is good. The species and the environment is considerably different from this study but it implies the potential of SDM approaches in predicting crop raiding behaviour of wildlife. However, SDM studies have more credibility when multiple modelling algorithms are applied since their results can be compared to one another and multiple models can be used to form an ensemble prediction.

5. Study area

The study area of this thesis is Taita Taveta County in Southern Kenya. Taita Taveta County is located between latitude 2° 46 South and 4° 10 South and longitude 37° 36 East and 30° 14 East (Taita Taveta County 2015). The county is a part of the Tsavo-Mkomazi ecosystem, which is known for its wildlife and big national parks. There are two large national parks in the county, Tsavo East and Tsavo West National Parks. Taita Taveta County is 17,048 km² in size and 62% of its area is covered by the two national parks (Taita Taveta County Government 2013).

The people inhabiting Taita Taveta County are mostly from Taita tribe, hence the name of the county. The Taita tribe consists of three separate tribes: Wadawida, Wasaghala and Wataweta (Taita Taveta County website 2016). There are also people of other ethnic backgrounds like Somali and Maasai living in the county. Taita is also the language of the tribe, although there are several dialects of the language. Official languages are English and Swahili but people in the county use Taita for everyday communication.
The population of Taita Taveta County totaled 284,657 in the 2009 census with annual growth rate being 1.6% between 1999 and 2009 (Kenya National Bureau of Statistics 2010). According to County Annual Development Plan (Taita Taveta County 2015) 80% of rural employment is in agricultural sector, which is seen to be “predominantly small scale, rain fed and poorly mechanized”. Land scarcity is also a big problem in densely populated area surrounded by the national parks. Approximately 57.2% of the population lives below the poverty line, earning less than KSH 1562 (14,3 EUR) per month (Taita Taveta County Government 2013).

Mean annual rainfall in the Tsavo-Mkomazi ecosystem varies between 250 and 500mm, while higher altitudes in the Taita Hills can get annual rains of over 1200 mm (Taita Taveta County Government 2013). The higher annual rainfall in the higher altitudes has created a distinct kind of agroecological zone, facilitating more intensive agriculture. On the contrary, the dryer climate of lower areas has led to larger livestock herds and less intensive agriculture. Third topographical zone of Taita Taveta County is the Taveta Sub-County, which has potential underground water and it also obtains water from the rains falling down on Mt. Kilimanjaro (Taita Taveta County Government 2013). As a result, in general the agriculture around Taveta is more intense due to better availability of water compared to the low lands.

Figure 7. Agriculture in Taita Hills. (Autio, Antti. 2016)

The water availability in the area is declining and people, livestock and wildlife are competing for the precious resource. Only 35% of the county’s population has access to safe clean water (Taita Taveta County Government 2014) and locally, the lack of water is also seen as one of the main drivers of the HWCs. One of the reasons for declining and unpredictable rainfall patterns can be the
cutting of forests in the Taita Hills. Forested highlands and mountains covered in cloud forests are known locally as “water towers” for their ability to generate rainfall and water for the ecosystem from captured moisture (Hohenthal et al. 2015).

The human-wildlife conflict is a well-known and defined issue in Taita Taveta County, for example during fieldwork a stakeholder meeting was held considering the HWCs, called upon by the County Executive Secretary of Tourism, Transport, Environment & Natural Resources. Unfortunately, I was unable to take part in the meeting that included high ranking officials and different stakeholders. The issue of HWC is highly political and I was told during the fieldwork that the issue was also used to gain political leverage before the elections.

Figure 8. The study area. The two national parks cover 62% of the land in Taita Taveta County.
5.1 Protected areas

International Union for Conservation of Nature (IUCN 2015) defines national parks as: “Large natural or near natural areas set aside to protect large-scale ecological processes, along with the complement of species and ecosystems characteristic of the area, which also provide a foundation for environmentally and culturally compatible spiritual, scientific, educational, recreational and visitor opportunities.” Resource use is usually prohibited, expect sometimes for subsistence or minor recreational purposes. In the Kenya’s Wildlife Conservation and Management Act (2013), a national park is defined as: “Area of land and/or sea, especially dedicated to the protection and maintenance of biological diversity, and of natural and associated cultural resources, and managed through legal or other effective means.”

Tsavo National Park consists of two parks, Tsavo East National Park and Tsavo West National Park. The park was opened in 1948 (Tsavo National Park 2017) by the colonial government, making it one of the oldest national parks in Kenya. The laws and policies implemented before the independence denied access of local communities to land and resources that used to be theirs. Kenya gained its independence in 1963 but the laws and regulations excluding local communities from conservation persisted (Muriuki 1999). Tsavo East National Park is larger of the parks (11 747 km²) and topographically generally flat. Tsavo West National Park is more wet and mountainous and smaller of the two, 9065 km² in size. Together they cover approximately 48% of all conserved area in Kenya (Kenya Land Alliance 2002).

Kenya Wildlife Service (KWS) is a paramilitary organization managing the national parks and wildlife in Kenya. KWS was founded in 1990 and its predecessor was Wildlife Conservation and Management Department. It was seen as corrupt and underfunded and when KWS was founded, a big share of the staff was fired and better equipment was obtained for the KWS (Gibson 1999). Behind the founding of the new organization was also the desire to include local communities in conservation, and improvement of the relationship between KWS and the communities. For this reason, Community Wildlife Service was formed by KWS. Its objective was to recognize the rights of local communities and the impacts of wildlife on them. In Taita Taveta County, the Taita people carried out subsistence hunting, which was in conflict with the public authorities monitoring the country’s strict poaching laws. KWS was feared among the Taita people and negotiations between the parties were challenging to organize in the beginning. Nowadays the relationship has improved and KWS is seen as a helpful entity when it comes to human-wildlife conflicts.
KWS has built electrical fences around the national parks as a solution to the human-wildlife conflicts. The fences are built specially to stop elephants from crossing over to human inhabited areas. Building a fence is not a trouble-free project and some stakeholders state that the fences are built in wrong locations. The fences are problematic also because wildlife corridors between national parks must be maintained to allow seasonal migration of animals. Also, human-wildlife conflicts have intensified on the edges of the fences after animals have found a way around them (Kenya wildlife service 2007). KWS also called for evaluation of socioeconomic and environmental impacts of the existing fences before building new ones. However, new fences have been built. The management plan also supports the planting of live fences using plants like Mauritius thorn (Caesalpinia decapetala). In order to reduce the impacts of human-wildlife conflicts, new outposts for KWS personnel will be established and new techniques of animal control applied in the future.

There are also privately owned and managed Taita Hills Wildlife Sanctuary (THWS) and community owned LUMO Community Wildlife Sanctuary (hereafter LUMO) in the study area. THWS and LUMO are situated between the Tsavo East and Tsavo West National Parks, connected to the latter. LUMO and THWS are not national parks and they can define their relationship with human activity in the area themselves. Killing wildlife in Kenya is forbidden by law and the protected areas act as wildlife safe zones prohibiting utilization of the areas by humans. However, wildlife is seen to “leak” from the PAs to the human areas in a search for food and water increasing the risk of conflict.

5.2 Future prospects of Taita Taveta County

The whole county is in transformation due to large infrastructure development schemes. A new road connecting Voi and Taveta has been constructed and it also passes through Tsavo West National Park. It is seen to potentially boost the economy and the opportunities in the area (African Development Fund 2013). Another big construction project in action is the building of Standard Gage railway through the country, which also passes through Tsavo West National Park. Being built on a barrier several meters high, it certainly affects the migratory routes and movements of wildlife. Underpasses are being built, but these places might become hotspots for HWCs and traffic accidents. Also, increased road transportation might cause more traffic accidents involving wildlife (Egis International 2012). However, these impacts are out of scope of this research and too early to
evaluate. Studying the impacts of these big infrastructure projects would be important, for humans and wildlife alike.

6. Data

6.1 Human-wildlife conflict data

Species distribution modeling starts with mapping the known distribution of the species, collecting the data consisting of species observations. The data for this study was collected from HWC compensation request forms, which had been submitted to KWS between the years 2013 and 2016. The data collection was performed by myself and the research assistant Mwadime Mjomba during 2016. Data should not include HWCs that only caused a threat to a human but no actual damages, but there were entries in the data with damage title of “human threat”. These entries were still included in the data since compensation was demanded from the incidents. Presumably, the compensation demands are only filed for actual damages. The information derived from the compensation forms was: species involved, damage caused, location and date. No personal data was documented from these compensation forms. Another dataset was collected with help of Alfred Parsaoti from the KWS, who had notes of conflicts collected during problem animal control (PAC) unit missions. This data was combined with the one obtained from compensation forms by deleting all the incidences that did not cause damages for which compensation could have been demanded for.

The HWC occurrence data consisted only of presence points so the points of HWC involving other species were used to mark absence of the conflict for modelled species (ie. for elephants, all conflicts including other species, were marked as absences). However, if number of absences exceed the presences remarkably, it might cause bias in model evaluation. This was avoided by randomly sampling equal number of absences from the data for each species to match the number of presence points.

Since a fence had been built in 2015 to prevent animals from leaving the PAs and entering to human areas, some of the records had to be removed from the dataset. Contradictory information was obtained from the authorities of Taita Taveta County about the time when the fence was completed. Also, the fence was not completed uniformly all at once, but a segment at a time. Overall estimation
was that it was completed in the end of 2015. Therefore, all records from 2015 September onwards were removed from the data.

Human injuries totaled in 47 cases and involved, elephants, hyenas, lions, hippos, buffalos and leopards. Data included 76 cases of crop destruction, most done by elephants but also few cases involving hippos and baboons were reported. For livestock depredation, data included 166 cases, involving lions, leopards, hyenas, cheetahs, wild dogs, mongooses and baboons. The species with enough conflict records to be modelled were following: elephant (n=93), hyena (n=84), lion (n=54), leopard (n=21) and cheetah (n=9).

### 6.2 Data for temporal distribution analysis

Another HWC dataset, without coordinates of conflict, was obtained from KWS during the fieldwork. This dataset had longer temporal timeframe, ranging from 1990 to 2016. This enabled a more averaged analysis of monthly variation of the HWCs compared to the data used for SDM. It is likely that the HWC in the 1990 were less likely to be monitored as they are nowadays. However, the dataset was only cleaned up from evident faulty entries and the final dataset included 17362 incidents.

### 6.3 Environmental variables

Data for the explanatory variables was obtained from multiple sources. At first, discussions with Maktau residents and experts of wildlife and conservation gave some insight on the factors behind HWCs. Combined with literature on the topic, a sense of which variables should be used in the models was obtained. What was brought up multiple times during the fieldwork, was that the main driver of the HWCs was water, or the lack of it. Especially elephants were thought to be coming to inhabited areas for water and green vegetation. Distance to water points and rivers were chosen to represent the availability of water in the models. Normalized Difference Vegetation Index (NDVI) (Tucker 1979) and land cover were chosen as environmental variables to reflect the quality of vegetation. Also, annual average precipitation was added to the models for its part in water availability and vegetation quality. Another factor thought to affect the HWC distribution significantly was the proximity of human activity, since humans are the other component needed for HWC to happen. For this reason, population density and distance to closest house and road were
chosen as variables to be used in the models as the anthropogenic variables. Finally, distance to protected areas was chosen as explanatory variables for its logical impact in area with large national parks.

Before calibrating the models, the data was processed to be as representative and associated with the species being modeled as possible. Data was processed in ArcGis, QGIS and R. All data was also converted to have the same geographic coordinate system, the same cell size and geographic dimensions. All data was set to a resolution of 100 meters, which is a good compromise between accuracy and the time consumed by the modeling process. Also, some of the data used was not available in more precise resolution. The variables used in the beginning of the modeling process were following: distance to the closest protected area, distance to the closest water point, distance to the closest river, distance to the closest house, distance to closest the road, human population density, annual average precipitation, land cover, NDVI and elevation. Elevation and annual average precipitation were observed to have a strong correlation on one another (multicollinearity) and only annual average precipitation was chosen based on its strongest correlation with the HWC points and theory. When the model calibration proceeded, variables were dropped out of the models for their insignificance regarding the modeled species.

For NDVI, Modis NDVI band was used and the values were averaged from 2014–2015 images. For landcover layer, globcover 2009 products were used, which are based on ENVISAT’s Medium Resolution Imaging Spectrometer (MERIS) Level 1b data with spatial resolution of 300m. For elevation variable, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM) was used. Annual averaged precipitation for Taita Taveta County was obtained from AFRICLIM 3.0 dataset (see Platts et al. 2015) which was averaged for years 2014–2015.

Since there was no data for population density in Taita Taveta County, all the houses were digitized from Google satellite imagery and used to compose the population density variable. Siljander et al. (2011), suggest weighting every house for 6 people for this kind of examination in Taita Hills. There were challenges in separating residential buildings from other buildings from satellite images. Therefore, it is probable that too many houses were marked as occupied and with the weight of 6, the population of the county would have been almost double the actual size. This challenge was overcome by simply dividing the total population with the number of house points. Finally, a weight of 4 was chosen for each house point to generate the population density layer with QGIS software.
Proximity of roads, rivers, water points and PAs were calculated applying the Euclidean distance tool of ArcMap on data based on the digitized geodata of the topographic maps at 1:50 000 scale by the Survey of Kenya. Euclidean distance tool gives all cells a value according to the distance to the closest road, water point, river or PA border. Variables missing from the models, that could have had an impact on the results are following: elephant paths, distance to closest field, crops grown in the fields. These variables have been used in previous HWC research but were unavailable for this study.

Figure 9. Environmental variables used in the final models.
7. SDM methods

All of SDM was made using the R software’s (R Development Core Team 2017) biomod2 package (Thuiller et al. 2009), which allows flexible use of 10 different statistical models. Models used were: Generalized Linear Model (GLM), Generalized Additive Model (GAM), Generalized Boosted Model (GBM), Classification Tree Analysis (CTA), Artificial Neural Networks (ANN), Flexible Discriminant Analysis (FDA), Multiple Adaptive Regression Splines (MARS), Random Forest (RF), Maximum Entropy (MaxEnt) and Surface Range Envelope (SRE).

7.1 Modelling algorithms

7.1.1 Generalized linear models (GLM)

GLMs are natural generalizations of familiar classical linear models and they do not force data into unnatural scales and therefore allow non-linearity and non-constant variance structure in the data (McCullagh & Nelder, 1989). GLMs try to estimate the relationship between variables (fitting the response) by minimizing the models residual error. However, GLMs are able to work with the whole extended family of distributions, instead of only normally distributed variables. GLMs can also recognize non-linear responses using link-function, which is determined by the response variables distribution (Guisan et al. 2002). GLMs can also include interactions and polynomial terms of the explanatory variables. This may allow even better estimations of non-linear relationships between the variables (Guisan et al. 2002).

7.1.2 Generalized additive model (GAM)

GAMs (Hastie & Tibshirani 1990) also work much like GLMs, using link function to estimate the relationship between the response variable and the smoothed function of explanatory variables (Guisan et al. 2002). Using GAMs involve giving certain degrees of freedom for “smoothing” each explanatory variable, which the model uses to fit the response curves (Guisan et al. 2002). GAM is called semi-parametric extension of GLM since GAM is more data driven (as opposed to model driven) and therefore, allow more complex response shapes by using the smoother (Franklin & Miller 2010). Given too large degree of freedom, GAMs have tendency to overfit to the data. GAMs cannot use interaction terms of variables like GLMs can (Guisan et al. 2002).
7.1.3 *Multiple Adaptive Regression Splines (MARS)*

Using piecewise basis functions, MARS (Friedman 1991) defines the relationships between the response and explanatory variables. Basis functions are determined in pairs using a “knot”, a turning point in the range of the predictor. Basis functions are defined for both sides of the turning point and a linear response is modeled between each turning point. The basis function and potential knots are chosen to give the biggest decrease in residual sum of squares (Leathwick et al. 2006). MARS fits a large model, identifying most knots automatically and choosing the best basis functions and then, prunes the model, reducing its complexity by removing basis functions that have smallest contribution to the model fit. MARS is computationally fast, easing work with large datasets, when compared to other regression based models (Franklin & Miller 2010).

7.1.4 *Flexible Discriminant Analysis (FDA)*

Being a nonparametric extension of linear discriminant models, FDA (Hastie et al. 1994) is able to work with the extended family of distributions and nonlinear responses. FDA works like MARS when identifying the regressions between the variables. FDA then proceeds to classify the data based on the regressions and uses the classifications to predict, for example, the distributions of HWCs (Hastie et al. 1994).

7.1.5 *Surface Range Envelope (SRE)*

SRE is identical to the BIOCLIM algorithm (Busby 1991). It performs relatively simple analysis of the environmental factors dominating the distributions of the species and can work also with presence-only data. Once the bioclimatic envelope of the species is recognized the model can be used to predict the distributions, for example, in other areas or climatic conditions (Araújo & Peterson 2012). SRE is very much influenced by the data input and new variables may add restrictions to the models even though the variables would not be significant. This is result of SRE treating all presence points as equal, not trying to discriminate the extreme values, which it has been criticized for (Beuamont et al. 2005).
7.1.6 Classification Tree Analysis (CTA)

Sometimes called Classification and Regression Trees, CTAs (Breiman et al. 1986) repetitively split the data in two, forming a treelike structure. The model goes through each possible threshold for the split, choosing the one that generates biggest degree of homogeneity in the child nodes, or so-called branches (Vaysières et al. 2000). This binary partitioning stops when there are some defined number of observations left in the branch or when the split fails to achieve certain increase in homogeneity. If no rules are set, the final nodes include only one observation and the model is likely overfit (Franklin & Miller 2010). When the tree building has stopped, much like with MARS, the model is pruned. Tree pruning means removing splits that have overall added least homogeneity in the subgroups. Optimal pruning is usually found out by cross validation, removing the splits generating biggest prediction errors. Classification trees work well with categorical variables (such as species occurrence), hierarchical interactions, missing data and outliers.

7.1.7 Generalized Boosted Model (GBM)

Also called Boosted Regression Trees, GBMs (Friedman 2001) combine classification tree and machine learning approaches, building large number of trees and use them to produce a single model. Combining (boosting) the model from multiple classification trees, GBMs are better at predicting on new data than traditional classification tree methods because the end result is not based on a single classification tree (Elith et al 2008).

7.1.8 Random Forest (RF)

Random Forest (Breiman 2001) is a machine learning method like GBM. Much like the GBM, RF creates large amount of classification trees with different partitioning of data and averages the results into the final model. RF starts with bootstrapping the data number of times (usually 500 – 2000). The data left out of the sample is called out-of-bag observations, also known as test sample. A classification tree is made with each sample and each split in the trees is generated with a random subset of candidate predictor variables. The final predictions of the observations’ classes are decided by majority of vote from all of the trees (Cutler et al. 2007). Finally, the out-of-bag sample is used to test the model accuracy and the variable importance (Franklin & Miller 2010). Like with
GBM, the final model, averaged from multiple trees, overcomes the fundamental problem of overfitting, usually present with classification trees.

### 7.1.9 Maximum Entropy (MaxEnt)

MaxEnt (Phillips et al. 2006) is an advanced machine learning method. Maxent estimates the probability of distribution by finding the probability distribution of maximum entropy, based on the constraints defined by our observations (Thuiller et al. 2014). In other words, constraints can be specified to be the location of the observations in the n-dimensional hyperspace defined by the used explanatory variables. Maxent tries to find the largest possible uniform distribution for the response variable, based on the assumption that the sample of observations underestimates the real distribution of the phenomenon (Phillips et al. 2006, Franklin & Miller 2010). MaxEnt can also work with presence-only data and it is seen to produce accurate predictions of species distributions also with smaller datasets. It is a “black box” method, not allowing the user to examine the data processing steps more closely.

### 7.1.10 Artificial Neural Networks (ANN)

Artificial Neural Networks (Ripley 1996) create an additional hidden layer(s) of parameters based on the input explanatory variables and their interactions (Lek & Guegan 1999). From this hidden layer, ANNs are capable to model complex relationships between the variables. The information flows from input layer to output layer through the hidden layer, not allowing feedback in any phase. ANNs are seen as powerful tools for processing any kind of data. ANNs are also widely used in ecological modeling, including SDM research.

### 7.2 Model selection

To choose the best model is not simple. One may aim for a model that accurately predicts the species, or for a model that measures the variable importance correctly, or for a model that can be extrapolated to other regions. Parsimony in model building is desirable. It is a concept of simplest possible model with significant variables. Finding the parsimonious model can be automated or done manually with different techniques. (Franklin & Miller 2010). The simplest way to find
significant variables for a model is to measure the correlation between a single explanatory variable and the response variable.

In this thesis, the aim was to create models that balance between prediction accuracy and realism, trying to create accurate prediction maps and to define the underlying environmental and anthropogenic factors influencing the HWCs. As result, models’ ability to be generalized into other areas might decrease. Thus, selection of explanatory variables must be considered separately in distinct environments.

Rather than assessing the goodness-of-fit of the models (explanatory power), this thesis focuses on evaluation of the models’ accuracy (predictive power), explained in the next section. However, to formulate best possible GLMs, Akaike Information Criterion (AIC) (Akaike 1974) was used. AIC is a measure of the relative quality of a statistical model and it is regularly used in model selection. A smaller AIC (lower unexplained deviance) indicates better model since there is more deviance explained per variable (Franklin & Miller 2010).

### 7.3 Model evaluation

SDM model evaluation involves application of one or more statistical methods to calculate models’ predictive performance. This section goes through the procedures of splitting data and evaluation of prediction accuracy of models. However, predictive accuracy is not the only measure of “goodness” of the model, but other criteria, such as ecological realism and model credibility, are used as well. Overall, the validation of the model means it’s meeting specified performance requirements for its intended use.

#### 7.3.1 Data splitting

In SDM it is a common practice to split the observations into training data and testing data before model calibration. The model is calibrated using training data and its accuracy is tested against testing data (Fielding & Bell 1997). Splitting dataset creates only a semi-independent set of observations for testing, and the assessment might be biased because of the correlations between training and testing data. However, obtaining new independent dataset for testing purposes only can be a time consuming and expensive process, and it is generally seen sufficient enough to test against random subset of the original sample. There are numerous ways to split data for model calibration.
These include, for example, bootstrapping several random subsets, using leave-one-out cross-validation or k-fold cross-validation (see Fielding & Bell 1997). In this thesis, the data is simply split randomly into 70% training points and 30% testing points and altogether 20 models were calibrated and evaluated with unique data splits.

### 7.3.2 Prediction accuracy

SDMs usually work with presence/absence data. Evaluating binary prediction is done by testing how well the model predicts the actual presences (1) and absences (0). The predicted and actual instances can be positioned into a confusion matrix (table 1), showing the absolute or relative numbers of true positives, true negatives, false positive and false negatives (Fielding & Bell 1997). The methods used to measure accuracy and errors are usually based on the confusion matrix. The evaluating methods used in this thesis are area under the curve (AUC) and True skill statistics (TSS).

<table>
<thead>
<tr>
<th></th>
<th>Actual 1</th>
<th>Actual 0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predicted 1</strong></td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td><strong>Predicted 0</strong></td>
<td>False Negative</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

### 7.3.3 Area under the curve (AUC)

Area under the receiver operating characteristic curve (fig. 10) is a so-called threshold independent measure of model accuracy. Receiving operating characteristic (ROC) plot is gained by plotting true positive fraction (sensitivity) against false positive fraction (1-specificity) on all possible thresholds on the X-axis (false positive fraction). AUC describes the area under the ROC curve, providing an accuracy assessment that is not dependent on any particular threshold (threshold for probability of occurrence value to be defined as presence) (Fielding & Bell, 1997). AUC values are between 0.5 and 1, 0.5 meaning that the model is no better at predicting than random and 1 meaning perfect ability to sort out presence from absence. AUC scores of 0.5 –0.7 indicate low accuracy, 0.7–0.9 moderate accuracy and over 0.9 indicates high accuracy (Manel et al. 2001).
**Figure 10.** Example of a receiver operating characteristic (ROC) plot (Franklin & Miller 2010). AUC of this ROC plot is 0.934. The closer the curve is to the upper left corner, the higher the AUC is. The middle line represents AUC of 0.5.

### 7.3.4 True skill statistic

True skill statistic (TSS) is a threshold dependent evaluation method, using a probability threshold where the sum of sensitivity and specificity is maximized. TSS is defined as $1 - (\text{Sensitivity} + \text{Specificity})$ and the scores range from -1 to 1 (Franklin & Miller 2010). Score close to 0 means that the model is no better than random and 1 being the perfect score.

### 7.4 Majority vote predictions

Biomod2 turns probability of occurrence predictions into binary maps based on a threshold that maximizes the percentage of correctly predicted presences and absences (optimal sensitivity and specificity). Majority vote maps are produced from these binary predictions for each species.
Majority maps are made in attempt to fix possible problems of overfit and inaccuracies in the models. Majority vote map is a new raster layer with values computed from chosen binary prediction rasters. Most accurate five models per species were used, making majority of vote three. If majority of the models (three out of five) predict absence in a cell, it is simply marked with white colour, while different levels of consensus on predicted presences (3/5, 4/5 and 5/5) are presented in the maps with different colours. Thus, areas with different levels of conflict risk can be defined from the maps.

8. Results

Species distribution models were calibrated and evaluated with conflicts involving elephants, leopards, hyenas and cheetahs. Firstly, all the conflicts involving predators were tried to be predicted with same models. Results were not accurate (AUC 0.5–0.6) and new models were made for each species separately with different environmental variables to optimize model performance. Cheetah models did not reach AUC scores of 0.7 and the models were removed from further analysis. Poor prediction accuracy was likely due to the small sample size of the cheetah conflicts (n=9).

All the environmental variables used were first run through log function to make the distributions closer to Gaussian, but the process did not make the prediction accuracies any better. Thus, the original values were used for an easier interpretation of the results. Spatial autocorrelation of variables was tested with Moran’s I (Moran 1950). All the explanatory variables were spatially autocorrelated, at least at short distances (appendix 1). This and the fact that models can overfit to data means that the results must be interpreted with caution. The five most accurate models were chosen for the majority vote predictions in attempt to even out the inaccuracies of single models.

Response curves of GBM are shown for each species. GBM was chosen because it performed well across all the species and showing same algorithm’s response curves eases comparison between species. Response curves show the risk of conflict along the explanatory variable when other variables are averaged. X-axis shows the value of the explanatory variable and Y-axis presents the probability of occurrence for the response variable. The observations are shown as small pins along the X-axis. Response curves of all models can be found in the appendices. After the response curves, the majority vote predictions of models and seasonal kernel density maps of the observed conflicts are presented.
8.1 Crop damages by elephants

After the initial modelling of human-elephant conflict involving both, human damages (n=23) and crop raiding (n=70), it was noticed that better prediction accuracy was reached by modelling crop raiding separately. Models of human damage by elephants did not reach desired accuracy (AUC 0.7) and were left out of further analysis. In case of elephant crop raiding, the best model performance (AUC/TSS) was reached with GBM, RF, MaxEnt, ANN and CTA (table 2). The highest AUC scores were reached with CTA and MaxEnt. The variables used for the models were: distance to protected area, annual average precipitation, human population density, distance to river, distance to water point and NDVI. Overall, the models selected population density to be the strongest factor determining where elephant crop raiding occurs. Population density was followed by distance to PAs and NDVI in the importance of variables (fig. 11).

Table 2. Averaged AUC and TSS values of all elephant models (20 runs). Best prediction accuracies are marked with yellow.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>TSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM</td>
<td>0.67005</td>
<td>0.34445</td>
</tr>
<tr>
<td>GAM</td>
<td>0.65125</td>
<td>0.3102</td>
</tr>
<tr>
<td>GBM</td>
<td>0.7061</td>
<td>0.4239</td>
</tr>
<tr>
<td>RF</td>
<td>0.7093</td>
<td>0.46065</td>
</tr>
<tr>
<td>MaxEnt</td>
<td>0.7444</td>
<td>0.4923</td>
</tr>
<tr>
<td>MARS</td>
<td>0.6744</td>
<td>0.3944</td>
</tr>
<tr>
<td>ANN</td>
<td>0.7188</td>
<td>0.4559</td>
</tr>
<tr>
<td>FDA</td>
<td>0.61915</td>
<td>0.2579</td>
</tr>
<tr>
<td>CTA</td>
<td>0.83035</td>
<td>0.5922</td>
</tr>
<tr>
<td>SRE</td>
<td>0.62385</td>
<td>0.24735</td>
</tr>
</tbody>
</table>
Table 3. Statistical characteristics of the environmental variables used in the human-elephant conflict models.

<table>
<thead>
<tr>
<th>Variable</th>
<th>MIN</th>
<th>MEDIAN</th>
<th>MEAN</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual precipitation</td>
<td>568</td>
<td>685</td>
<td>751</td>
<td>1316</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.2525</td>
<td>0.3968</td>
<td>0.4379</td>
<td>0.7345</td>
</tr>
<tr>
<td>Distance to PA</td>
<td>0</td>
<td>6412</td>
<td>7026</td>
<td>20282</td>
</tr>
<tr>
<td>Distance to river</td>
<td>0</td>
<td>5669</td>
<td>5942</td>
<td>21674</td>
</tr>
<tr>
<td>Distance to water point</td>
<td>178.9</td>
<td>2270.6</td>
<td>2619.5</td>
<td>10050.5</td>
</tr>
<tr>
<td>Population density</td>
<td>0</td>
<td>50.64</td>
<td>77.46</td>
<td>371.06</td>
</tr>
</tbody>
</table>

Table 4. Averaged variable importance for all elephant crop raiding models (20 runs). The highest variable importance for each model is marked with yellow.

<table>
<thead>
<tr>
<th>Variable</th>
<th>GLM%</th>
<th>GAM%</th>
<th>GBM%</th>
<th>RF%</th>
<th>MaxEnt%</th>
<th>MARS%</th>
<th>ANN%</th>
<th>FDA%</th>
<th>CTA%</th>
<th>SRE%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance to PA</td>
<td>23.0%</td>
<td>17.4%</td>
<td>11.6%</td>
<td>5.4%</td>
<td>32.4%</td>
<td>11.4%</td>
<td>82.1%</td>
<td>23.3%</td>
<td>0.0%</td>
<td>27.4%</td>
</tr>
<tr>
<td>Annual precipitation</td>
<td>13.0%</td>
<td>13.4%</td>
<td>5.1%</td>
<td>8.4%</td>
<td>35.1%</td>
<td>0.0%</td>
<td>21.9%</td>
<td>14.1%</td>
<td>18.3%</td>
<td>19.0%</td>
</tr>
<tr>
<td>Population density</td>
<td>59.4%</td>
<td>30.6%</td>
<td>21.8%</td>
<td>7.0%</td>
<td>69.0%</td>
<td>43.1%</td>
<td>1.7%</td>
<td>53.0%</td>
<td>53.9%</td>
<td>27.1%</td>
</tr>
<tr>
<td>Distance to river</td>
<td>0.0%</td>
<td>14.2%</td>
<td>2.9%</td>
<td>2.9%</td>
<td>44.1%</td>
<td>11.8%</td>
<td>71.4%</td>
<td>25.4%</td>
<td>0.0%</td>
<td>6.2%</td>
</tr>
<tr>
<td>Distance to water point</td>
<td>19.9%</td>
<td>17.0%</td>
<td>10.6%</td>
<td>6.7%</td>
<td>15.3%</td>
<td>36.2%</td>
<td>30.6%</td>
<td>18.4%</td>
<td>31.9%</td>
<td>23.6%</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.0%</td>
<td>37.6%</td>
<td>14.6%</td>
<td>8.4%</td>
<td>37.6%</td>
<td>30.1%</td>
<td>0.0%</td>
<td>37.6%</td>
<td>42.3%</td>
<td>19.8%</td>
</tr>
</tbody>
</table>

Figure 11. Average variable importance of all elephant crop raiding models (A=distance to protected area, B= annual average precipitation, C= population density, D=distance to river, E=distance to water point, F=NDVI).
Figure 12. GBM response curves for elephant crop raiding (A=distance to protected area, B=annual precipitation, C=population density, D=distance to river, E=distance to water point, F=NDVI). X-axis presents the explanatory variable while Y-axis depicts the probability of occurrence of conflict. The small pins along the X-axis are point observations of the conflicts.

In most models (figure 12, appendix 2–5), response curves of distance to PA demonstrated a high risk of conflict closer to PA borders, falling drastically after 5–8 kilometres. MaxEnt fitted the response curve unexplainably at short distances, possibly implying overfit. GBM and RF responses for annual average precipitation proposed increased the risk of conflict in areas with lower annual rainfall. MaxEnt predicted high risk also in areas with very high annual average precipitation, fitting an u-shaped response to the variable. Most models predicted areas with NDVI of 0.5–0.65 having the highest risk of conflict. CTA most likely overfitted the NDVI response (Appendix 5). In most models, the elephant crop raiding risk rose in areas with relatively low population densities (< 100 people/km²). GBM, RF and CTA models fitted response showing crop raiding being more likely in areas between 1000 and 4500 meters from water points. With MaxEnt and ANN, the response was negative, predicting highest risk for conflict in close vicinity of water points, decreasing with distance. Responses to distance to river variable were contradictory between the models. While GBM and RF predicted a higher risk of conflict near the rivers, MaxEnt fitted an u-shaped response, implying higher risk also at long distances. Then, ANN estimated a lower risk of conflict near the rivers, risk chance increasing after distance of eight kilometres.
Figure 13. Majority vote prediction of elephant crop raiding. The map shows level of agreement between the five most accurate binary predictions. The models included in the prediction were GBM, RF, MaxEnt, ANN and CTA.

Figure 14. Seasonal elephant crop raiding kernel densities
8.2 Human-lion conflict

For human-lion conflicts, the best overall prediction accuracies were reached using the following variables: distance to closest PA, annual average precipitation, population density, distance to river, distance to water point and distance to house. Models chosen for the majority vote prediction were GLM, GAM, GBM, CTA and MaxEnt (table 5). Averaged AUC scores of the models varied between 0.72 (GAM) and 0.858 (CTA), implying moderate to good accuracy in human-lion conflict predictions. While also ANN and RF reached desired prediction accuracies, they were left out of the majority vote predictions in favour of more traditional parametric GLM and semi-parametric GAM. Moderate accuracy across parametric, semi-parametric and non-parametric models suggest overall solidity of the models. Also, TSS scores were moderate, reaching over 0.5 in majority of the models.

Annual average precipitation was the most significant variable for all the human-lion conflict models. It was followed by distances to PA and distance to water point (fig. 15). Population density, distances to house and distance to river were the least important variables. Distance to house had very small importance across all models (table 7). However, leaving it out of the models led to a notable decrease in the prediction accuracies.

Table 5. AUC and TSS scores of human-lion conflict models. Highest predictions accuracies are marked with yellow.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>TSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM</td>
<td>0.74125</td>
<td>0.4775</td>
</tr>
<tr>
<td>GAM</td>
<td>0.7189</td>
<td>0.4303</td>
</tr>
<tr>
<td>GBM</td>
<td>0.7454</td>
<td>0.50205</td>
</tr>
<tr>
<td>RF</td>
<td>0.7803</td>
<td>0.55385</td>
</tr>
<tr>
<td>MaxEnt</td>
<td>0.8381</td>
<td>0.60635</td>
</tr>
<tr>
<td>MARS</td>
<td>0.65995</td>
<td>0.39855</td>
</tr>
<tr>
<td>ANN</td>
<td>0.7647</td>
<td>0.53875</td>
</tr>
<tr>
<td>FDA</td>
<td>0.6873</td>
<td>0.37835</td>
</tr>
<tr>
<td>CTA</td>
<td>0.85885</td>
<td>0.64145</td>
</tr>
<tr>
<td>SRE</td>
<td>0.53335</td>
<td>0.0991</td>
</tr>
</tbody>
</table>
Table 6. Statistical characteristics of the environmental variables used in the human-lion conflict models.

<table>
<thead>
<tr>
<th></th>
<th>MIN</th>
<th>MEDIAN</th>
<th>MEAN</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual precipitation</td>
<td>568</td>
<td>685</td>
<td>751</td>
<td>1316</td>
</tr>
<tr>
<td>Distance to closest house</td>
<td>0</td>
<td>100</td>
<td>289.3</td>
<td>19181.5</td>
</tr>
<tr>
<td>Distance to PA</td>
<td>0</td>
<td>6412</td>
<td>7026</td>
<td>20282</td>
</tr>
<tr>
<td>Distance to river</td>
<td>0</td>
<td>5669</td>
<td>5942</td>
<td>21674</td>
</tr>
<tr>
<td>Distance to water point</td>
<td>178.9</td>
<td>2270.6</td>
<td>2619.5</td>
<td>10050.5</td>
</tr>
<tr>
<td>Population density</td>
<td>0</td>
<td>50.64</td>
<td>77.46</td>
<td>371.06</td>
</tr>
</tbody>
</table>

Table 7. Variable importance for all human-lion conflict models (20 runs). The highest variable importance for each model is marked with yellow.

<table>
<thead>
<tr>
<th></th>
<th>GLM</th>
<th>GAM</th>
<th>GBM</th>
<th>RF</th>
<th>MaxEnt</th>
<th>MARS</th>
<th>ANN</th>
<th>FDA</th>
<th>CTA</th>
<th>SRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance to PA</td>
<td>59.4</td>
<td>17.1</td>
<td>22.5</td>
<td>19.1</td>
<td>32.9</td>
<td>0.0</td>
<td>96.6</td>
<td>0.0</td>
<td>41.7</td>
<td>37.3</td>
</tr>
<tr>
<td>Annual precipitation</td>
<td>0.0</td>
<td>96.2</td>
<td>53.6</td>
<td>33.4</td>
<td>35.6</td>
<td>75.6</td>
<td>8.2</td>
<td>63.4</td>
<td>44.8</td>
<td>12.6</td>
</tr>
<tr>
<td>Population density</td>
<td>37.7</td>
<td>26.7</td>
<td>10.3</td>
<td>10.0</td>
<td>13.3</td>
<td>20.9</td>
<td>5.5</td>
<td>14.4</td>
<td>0.0</td>
<td>19.1</td>
</tr>
<tr>
<td>Distance to river</td>
<td>0.0</td>
<td>27.8</td>
<td>0.8</td>
<td>3.3</td>
<td>0.0</td>
<td>10.4</td>
<td>74.8</td>
<td>10.6</td>
<td>11.7</td>
<td>24.1</td>
</tr>
<tr>
<td>Distance to waterpoint</td>
<td>0.0</td>
<td>22.7</td>
<td>19.7</td>
<td>22.1</td>
<td>15.3</td>
<td>41.1</td>
<td>9.6</td>
<td>40.2</td>
<td>42.3</td>
<td>21.6</td>
</tr>
<tr>
<td>Distance to house</td>
<td>0.0</td>
<td>5.2</td>
<td>0.0</td>
<td>0.5</td>
<td>17.7</td>
<td>0.0</td>
<td>2.8</td>
<td>0.0</td>
<td>0.0</td>
<td>5.5</td>
</tr>
</tbody>
</table>

Figure 15. Averaged variable importance of all human-lion conflict models (A=distance to PA, B=annual average precipitation, C=population density, D=distance to river, E=Distance to water point, F=distance to house).
Figure 16. GBM response curves for human-lion conflicts (A=distance to PA, B=annual average precipitation, C=population density, D=distance to river, E=Distance to water point, F=distance to house).

The response curves of GBM, CTA and MaxEnt (fig. 16, appendix 8–9) were similar. Annual average precipitation had a clear effect on the predicted distribution of the human-lion conflicts. A higher risk for a conflict was predicted in areas of low annual rainfall (<600 mm/year) in models chosen for the majority vote prediction. GLM made an exception, not using the variable at all. In all the chosen models (excluding GAM), the risk of human-lion conflict was predicted to be significantly higher in the close vicinity of PAs. GBM and MaxEnt responses presented a minor risk of conflict in vicinity of water points but notably higher risk further away from them. CTA curve predicted a high risk of conflict near the water points. GAM presented slightly higher risk of conflict only at longer distances from the water points and GLM did not use the variable for the predictions. Most of the response curves for distance to river did not fluctuate much with increasing distance in the models chosen for the majority vote prediction. CTA made an exception, giving high risk of conflict in the proximity of rivers. The models that gave distance to house the highest importance (GAM and MaxEnt) seemed to have contradictory views on its impact. MaxEnt estimated a higher risk of conflict near houses (<1 km), while GAM assessed high risk after one kilometre from the closest house. The models predicted a higher human-lion conflict risk in areas of low population densities. However, GLM (appendix 6) produced slightly u-shaped response curve for population density variable, predicting a higher risk of conflict also in more densely populated areas.
Figure 17. Majority vote prediction of human-lion conflict. The models included in the prediction were GLM, GAM, GBM, CTA and MaxEnt.

Figure 18. Seasonal kernel density maps of human-lion conflict.
8.3 Human-hyena Conflict

The human-hyena conflict models used six variables: distance to PA, annual average precipitation, distance to river, distance to water point, distance to road and NDVI. The models chosen for the majority vote prediction due to highest prediction accuracies were GBM, RF, MaxEnt, MARS and CTA (table 8). The models reached AUC scores between 0.70 and 0.80 implying moderate to good accuracy.

Distance to PA was overall the most important variable in human-hyena conflict models as it was chosen by seven out of ten models to be the most influential variable (table 10). MaxEnt selected distance to road as the most important variable, while NDVI was chosen as the most important variable by CTA and SRE. The importance of most variables varied notably between the different models. Some of the variables were given very small importance but selecting them for modelling increased accuracy of the models.

Table 8. AUC and TSS scores of human-hyena conflict models. Highest predictions accuracies are marked with yellow.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>TSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM</td>
<td>0.6654</td>
<td>0.34625</td>
</tr>
<tr>
<td>GAM</td>
<td>0.66155</td>
<td>0.33085</td>
</tr>
<tr>
<td>GBM</td>
<td>0.7121</td>
<td>0.41395</td>
</tr>
<tr>
<td>RF</td>
<td>0.7075</td>
<td>0.42105</td>
</tr>
<tr>
<td>MaxEnt</td>
<td>0.75295</td>
<td>0.48995</td>
</tr>
<tr>
<td>MARS</td>
<td>0.70555</td>
<td>0.4175</td>
</tr>
<tr>
<td>ANN</td>
<td>0.68095</td>
<td>0.37275</td>
</tr>
<tr>
<td>FDA</td>
<td>0.6384</td>
<td>0.30765</td>
</tr>
<tr>
<td>CTA</td>
<td>0.80485</td>
<td>0.59945</td>
</tr>
<tr>
<td>SRE</td>
<td>0.6067</td>
<td>0.21345</td>
</tr>
</tbody>
</table>
Table 9. Statistical characteristics of the environmental variables used in human-hyena conflict models.

<table>
<thead>
<tr>
<th>Variable</th>
<th>MIN</th>
<th>MEDIAN</th>
<th>MEAN</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance to PA</td>
<td>0</td>
<td>9539</td>
<td>10175</td>
<td>20282</td>
</tr>
<tr>
<td>Annual precipitation</td>
<td>149</td>
<td>279</td>
<td>278.6</td>
<td>402</td>
</tr>
<tr>
<td>Distance to river</td>
<td>145.6</td>
<td>2600</td>
<td>5283.2</td>
<td>21673.9</td>
</tr>
<tr>
<td>Distance to water point</td>
<td>178.9</td>
<td>2004.9</td>
<td>2454.8</td>
<td>6263.4</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.28</td>
<td>0.38</td>
<td>0.39</td>
<td>0.62</td>
</tr>
<tr>
<td>Distance to road</td>
<td>20</td>
<td>859.1</td>
<td>1105.8</td>
<td>3757.7</td>
</tr>
</tbody>
</table>

Table 10. Variable importance for all human-hyena conflict models (20 runs). The highest variable importance for each model is marked with yellow.

<table>
<thead>
<tr>
<th>Variable</th>
<th>GLM</th>
<th>GAM</th>
<th>GBM</th>
<th>RF</th>
<th>MaxEnt</th>
<th>MARS</th>
<th>ANN</th>
<th>FDA</th>
<th>CTA</th>
<th>SRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance to PA</td>
<td>67.6 %</td>
<td>77.9 %</td>
<td>47.0 %</td>
<td>21.8 %</td>
<td>39.2 %</td>
<td>46.7 %</td>
<td>77.7 %</td>
<td>48.3 %</td>
<td>36.3 %</td>
<td>19.6 %</td>
</tr>
<tr>
<td>Annual precipitation</td>
<td>0.0 %</td>
<td>12.2 %</td>
<td>12.9 %</td>
<td>10.6 %</td>
<td>36.1 %</td>
<td>22.0 %</td>
<td>2.6 %</td>
<td>23.5 %</td>
<td>21.5 %</td>
<td>29.4 %</td>
</tr>
<tr>
<td>Distance to river</td>
<td>0.0 %</td>
<td>6.3 %</td>
<td>0.2 %</td>
<td>4.1 %</td>
<td>32.3 %</td>
<td>17.4 %</td>
<td>21.1 %</td>
<td>17.5 %</td>
<td>11.5 %</td>
<td>14.5 %</td>
</tr>
<tr>
<td>Distance to water point</td>
<td>0.0 %</td>
<td>10.6 %</td>
<td>0.3 %</td>
<td>4.4 %</td>
<td>36.5 %</td>
<td>20.3 %</td>
<td>27.5 %</td>
<td>22.9 %</td>
<td>25.0 %</td>
<td>18.7 %</td>
</tr>
<tr>
<td>NDVI</td>
<td>36.1 %</td>
<td>31.7 %</td>
<td>29.9 %</td>
<td>17.1 %</td>
<td>30.1 %</td>
<td>24.2 %</td>
<td>0.0 %</td>
<td>20.7 %</td>
<td>43.7 %</td>
<td>32.7 %</td>
</tr>
<tr>
<td>Distance to road</td>
<td>14.1 %</td>
<td>16.2 %</td>
<td>4.8 %</td>
<td>6.0 %</td>
<td>46.4 %</td>
<td>27.5 %</td>
<td>38.0 %</td>
<td>24.3 %</td>
<td>8.6 %</td>
<td>12.6 %</td>
</tr>
</tbody>
</table>

Figure 19. Averaged variable importance of all human-hyena conflict models (A=distance to PA, B=annual precipitation, C=distance to river, D=distance to water point, E=NDVI, F=distance to road).
Figure 20. GBM response curves for human-hyena conflicts (A=distance to PA, B=annual average precipitation, C=distance to river, D=distance to water point, E=NDVI, F=distance to road).

Response curves of all chosen models (appendix 10-13) behaved relatively similarly to the GBM responses shown in figure 20. Responses to distance to PA were very different from other species, as areas of the highest risk were not near PA borders but further away from them. For annual average precipitation responses, the risk of conflict rose sharply after 600 mm per year and stayed high after the peak. All other models agreed with GBM also about the response to NDVI, giving a high risk of conflict in areas of low NDVI (< 0.5). Fluctuation in responses to distance to road was minor in the models (excluding MaxEnt), showing only a slightly higher risk of conflict closer to roads. MaxEnt response curves also suggested slight overfitting to the data, showing high fluctuation in some of the curves. GBM responses to distance to river and water point did not fluctuate as much as responses of RF, MaxEnt and MARS. However, they were also given very small variable importance in GBM (<1%). According to RF, MaxEnt and MARS, conflicts seemed to concentrate further away from the rivers (>3km). With water points the relationship was not as simple to interpret. MARS and MaxEnt showed clear peak in conflict risk at one kilometre distance from the water points, and again, after five kilometres. Overall the models used for the majority vote prediction did not seem to have clear consensus on the relationship between the conflicts and the proximity of water points.
Figure 21. Majority vote prediction of human-hyena conflict. The models chosen for the prediction were GBM, RF, MaxEnt, MARS and CTA.

Figure 22. Seasonal kernel density maps of human-hyena conflict.
### 8.4 Human-leopard Conflict

The best models predicting human-leopard conflicts reached moderate to good accuracies with AUC scores between 0.74 and 0.82. These models were RF, MaxEnt, MARS, ANN, CTA, FDA and GBM (table 11). However, FDA and GBM were left out of the majority vote prediction for their lower TSS scores. Human-leopard conflict models selected only 4 different variables: distance to PA, annual average precipitation, distance to water point and NDVI. Models also used the environmental variables very differently (table 13). Across all the models, the most important variables impacting the distribution of human-leopard conflicts were distance to PA and annual average precipitation (fig. 23).

**Table 11.** AUC and TSS scores of the human-leopard conflict models. Highest prediction accuracies are marked with yellow.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>TSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLM</td>
<td>0.6511</td>
<td>0.37495</td>
</tr>
<tr>
<td>GAM</td>
<td>0.66005</td>
<td>0.3345</td>
</tr>
<tr>
<td>GBM</td>
<td>0.720526</td>
<td>0.487526</td>
</tr>
<tr>
<td>RF</td>
<td>0.79645</td>
<td>0.59625</td>
</tr>
<tr>
<td>MaxEnt</td>
<td>0.82735</td>
<td>0.6583</td>
</tr>
<tr>
<td>MARS</td>
<td>0.7459</td>
<td>0.5392</td>
</tr>
<tr>
<td>ANN</td>
<td>0.7428</td>
<td>0.50715</td>
</tr>
<tr>
<td>FDA</td>
<td>0.7458</td>
<td>0.4856</td>
</tr>
<tr>
<td>CTA</td>
<td>0.82495</td>
<td>0.6606</td>
</tr>
<tr>
<td>SRE</td>
<td>0.6792</td>
<td>0.36435</td>
</tr>
</tbody>
</table>
Table 12. Statistical characteristics of environmental variables used in the human-leopard conflict models.

<table>
<thead>
<tr>
<th></th>
<th>MIN</th>
<th>MEDIAN</th>
<th>MEAN</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance to PA</td>
<td>0</td>
<td>2396.3</td>
<td>4222.7</td>
<td>12029.8</td>
</tr>
<tr>
<td>Annual precipitation</td>
<td>637</td>
<td>853</td>
<td>869.6</td>
<td>1316</td>
</tr>
<tr>
<td>Distance to water point</td>
<td>782.3</td>
<td>2092.5</td>
<td>2498</td>
<td>5798.3</td>
</tr>
<tr>
<td>NDVI</td>
<td>0.36</td>
<td>0.45</td>
<td>0.48</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Table 13. Variable importance for all human-leopard conflict models (20 runs). The highest variable importance for each model is marked with yellow.

<table>
<thead>
<tr>
<th></th>
<th>GLM</th>
<th>GAM</th>
<th>GBM</th>
<th>RF</th>
<th>MaxEnt</th>
<th>MARS</th>
<th>ANN</th>
<th>FDA</th>
<th>CTA</th>
<th>SRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance to PA</td>
<td>23.2</td>
<td>22.6</td>
<td>13.4</td>
<td>18.6</td>
<td>56.1</td>
<td>60.7</td>
<td>61.9</td>
<td>83.5</td>
<td>0.0</td>
<td>42.2</td>
</tr>
<tr>
<td>Annual precipitation</td>
<td>62.1</td>
<td>45.3</td>
<td>68.6</td>
<td>49.3</td>
<td>48.6</td>
<td>0.0</td>
<td>27.2</td>
<td>0.0</td>
<td>69.9</td>
<td>29.5</td>
</tr>
<tr>
<td>Distance to water point</td>
<td>28.0</td>
<td>28.1</td>
<td>9.3</td>
<td>10.4</td>
<td>41.4</td>
<td>0.0</td>
<td>43.4</td>
<td>0.0</td>
<td>19.9</td>
<td>24.1</td>
</tr>
<tr>
<td>NDVI</td>
<td>13.0</td>
<td>17.5</td>
<td>5.8</td>
<td>11.3</td>
<td>31.2</td>
<td>51.4</td>
<td>0.0</td>
<td>37.0</td>
<td>30.5</td>
<td>35.6</td>
</tr>
</tbody>
</table>

Figure 23. Averaged variable importance of all human-leopard conflict models (A=distance to PA, B=annual average precipitation, C=distance to water point, D=NDVI).
The GBM response curves for NDVI and distance to PA did not fluctuate much (fig. 24) but other models showed more variation with changes in the variables (Appendix 14–18). Models had differing opinions about the importance and the effect of distance to PA variable. CTA, model with the best accuracy, did not use the variable at all. MARS and ANN gave it high importance and predicted a notably higher risk of conflict in vicinity of the PAs. RF and MaxEnt response curves did not show strong changes in the response with increasing distance from the PAs. Responses to NDVI are mostly positive with the risk of conflict rising after NDVI of 0.35. GBM responses to annual average precipitation and distance to water point are relatively similar with other models. Models (excluding MARS) predicted a high risk of conflict in areas closer to water points (<3 km). The same models agreed on the importance of annual average precipitation, predicting a high conflict risk in areas with higher precipitation.
Figure 25. Majority vote prediction of the human-leopard conflict. Models chosen for the prediction were RF, MaxEnt, MARS, ANN and CTA.

Figure 26. Seasonal kernel density maps of the human-leopard conflict.
8.5 Temporal distribution of the HWC

The figure 27 presents the distribution of HWC involving different species across the months of a year. Different dataset was used for the temporal analysis. The data spanned from 1990 to 2016 and included all the conflict reported to the KWS. The study area has two rainy seasons: March to July and October to January (fig. 28). Hence, the two dry seasons are: January to March and July to October. The seasons were hypothesized to dictate the temporal distribution of HWCs.

![Graphs showing monthly distribution of HWCs](image)

**Figure 27.** Monthly distribution of HWCs in Taita Taveta County 1990 – 2016. The Y-axes show the absolute number of conflicts in given month.

![Agricultural calendar of Taita Taveta County](image)

**Figure 28.** Agricultural calendar of Taita Taveta County (National Drought Management Authority 2017)
The HWCs in Taita Taveta County involving elephants peaked in May-July and in January, while other months seem to have somewhat an equal amount of human-elephant conflicts. Conflicts involving lions also peaked in January and in May-June. The human-leopard conflicts peaked in July-October and December-January. Temporal patterns in human-hyena conflicts were not as clear. A peak can be seen in September-November and few months had relatively low amount of conflicts (April and August).

9. Discussion

Overall, the HWC prediction models performed well and only human-cheetah conflict models were left out from the final analysis due to poor prediction accuracy of the models. This was likely due to small amount of data (n=9) about the conflicts. Otherwise, models had moderate to good prediction accuracy and it can be concluded that with well thought variables, SDM has potential use in HWC studies. Realism of the produced models is argued in this section by reasoning the importance of the used variables based on earlier literature and features of the study area.

This study clearly shows that the ten algorithms performed differently between the species. For each species, the best five models were chosen for the majority vote prediction. Machine learning approaches and classification trees performed best overall but satisfactory accuracies were reached also with other kinds of algorithms. Higher overall accuracy of classification trees comes as no surprise, since by their nature, they are good at processing categorical data (in this case binary presence/absence data).

Annual average precipitation and PA proximity were in most cases the strongest drivers behind the HWCs. However, it was obvious that additional explanatory variables were important since adding them to the models clearly increased prediction accuracies. Furthermore, the importance of and the responses to the variables differed substantially between the species.

It is likely that the different agroecological zones of Taita Taveta County have an impact on the distribution of HWCs in the area. Taveta Sub-County, west of the Tsavo West National Park, is ecologically very different from the other areas of Taita Taveta County. The area is more wet because of possible groundwater and runoff from Mt. Kilimanjaro. This has led to more intensive agriculture in the area (Taita Taveta Government 2013), which may attract crop raiding by elephants. Intensive agriculture might also mean less livestock in the area, which is likely the reason for the low levels of human-predator conflicts. There are also major differences in ecology and
means of livelihoods between areas of higher altitude like Taita Hills and the lowlands of Taita Taveta County. While higher elevations gain more annual rainfall, have more intensive agriculture and denser human populations, the people living in dryer lower altitudes are more dependent on raising livestock (Taita Taveta County Government 2013). Higher concentrations of livestock in turn attracts predators for livestock depredation. Higher risk of crop raiding in Taveta sub-county compared to the highlands can be explained by high degrees of human disturbance caused by denser populations in the higher altitudes.

### 9.1 Elephant crop raiding

Highest risk of elephant crop raiding was predicted in areas close to the PAs with relatively low population density and high NDVI. However, higher annual average precipitation reduced the risk of conflict, while the proximity water points and rivers gave slightly contradictory information if thought to depict the availability of water. While the set hypotheses about the impacting factors were not all proven correct, the models performed relatively well in respect of prediction accuracy. According to earlier literature (Hoare 1999) elephant crop raiding is unpredictable phenomenon because it is mostly done by individual bull elephants. As such, the moderate to good prediction accuracies of the models came as a mild surprise. Based on the results, SDM can be seen as valid approach for predicting human-elephant conflicts. However, it is possible that making models separately for dry and rainy seasons would have improved the models.

At first, all human-elephant conflicts were modelled simultaneously until it became evident that human injuries and crop damages should be modelled separately for better accuracy. This resulted in better models for elephant crop raiding, which indicates distinct features of crop raiding and conflicts resulting in human damages and implies that human injuries do not happen only when protecting fields from crop raiding elephants. However, modelling of the human-elephant conflicts resulting in human injuries or deaths did not produce accurate predictions.

The most influential variable in the elephant crop raiding models was population density. The response curves for population density showed a low risk of conflict in areas where the density is very low or zero. However, highest risk for conflicts was predicted in areas with relatively low population densities. This is logical since wildlife tends to avoid densely populated areas, but HWCs still need some human activity to occur. Population density was followed by distance to PA and NDVI in variable importance. Response to NDVI was positive and the risk of conflict increased
with NDVI. Elephants choose the food with highest rate of nutrient intake (Osborn 2004) and therefore, favour greener vegetation with higher water content. This can be argued to explain the importance of NDVI in the models. The conflicts seemed to concentrate closely around the protected areas. GBM response showed a high risk of a human-elephant conflict up to ten kilometres from the PAs, but after that the risk declined drastically. Since most of the area’s elephants reside inside the national parks (Ngene et al. 2013), the importance of distance to PA in the models can be held realistic. Further from the PA edges, either human disturbance increases excessively or animals are chased back to the parks by the KWS. The impacts of population density, distance to PA and NDVI, on distribution of human-elephant conflicts were as hypothesised.

With positive response to NDVI, the slightly negative responses to annual average precipitation came as a slight surprise. This is can be due to high population densities of areas with high annual rainfall. Densely populated area of Taita Hills for example, in addition to being further away from the PAs, receives significant amount of annual rainfall. The hypothesis of higher rainfall attracting human-elephant conflicts was not proven correct with the models.

Distance to rivers and water points give slightly contradictory information if they are thought to depict water availability (Lamarque et al. 2009). This might have to do either with the missing temporal dimension of the models, some hidden factors manifesting indirectly through them or just being randomly correlated with the conflicts. During dry season, some of the rivers and water holes become dry and the models might look very different if made separately for different seasons. The responses to distance to river variable varied notably between the models and a clear conclusion about its impact was not made. Distance to water point variable was fitted by GBM, RF and CTA to have a high risk of conflict from one up to 4,5 kilometres from water points. ANN and MaxEnt gave simpler negative response, implying a higher risk of conflict closer to the water points. While having unexplained responses to the variables, without them, the prediction accuracy of the models declined. The hypothesis of these two factors attracting crop raiding was proven correct only regarding water points.

There are few explanatory variables that could have been added to the models if they had been available. For example, it is seen that traditional migratory routes and paths of elephants are components in the distribution of human-elephant conflicts (Naughton et al. 1999). Crops, vegetables and fruits grown and proximity of fields can also impact the distribution of crop raiding. With more resources, data to cover these variables could have been obtained by digitizing the fields of Taita Taveta County from satellite imagery and in addition, conducting an extensive field survey to determine the agricultural products cultivated in the area. The resources to create these variables
were not available for this thesis and the models were made without them. Nevertheless, the models can be seen as relatively realistic and accurate with the used variables.

The graph of monthly distribution of human-elephant conflicts (figure 27) confirms the set hypothesis of the conflicts peaking in the shifts from rainy to dry seasons in January and July. This is the time of ripening crops which are seen to attract the elephants.

9.2 Human-predator conflict

Big feline predator attacks have been successfully modelled with SDM techniques before (see Zarco-González et al. 2013, Carvalho et al. 2015). In this study, conflicts involving all the species together were modelled first, because it was thought that depredation behaviour is opportunistic and manifests where prey is available. However, the models were not accurate and the results were not satisfactory, so different variables were chosen for each species. This produced more accurate predictions and suggests that SDM has potential use in examining human-predator conflicts also in Eastern Africa. The distinct predictions about conflicts involving different species suggest that the locations of the conflicts are not determined solely by the concentrations of livestock. It also reflects different kinds of livestock depredation behaviour by different species. It could be also linked to species and individual animals avoiding other predators (Ramesh et al. 2017, Brook et al. 2012). Further, this could also mean that the observations of other predatory species could be used as explanatory variables for HWC models.

There are two theories on selection of hunting grounds by carnivores. The theories are based on the natural prey, but could also give insight on livestock depredation. The two contradicting theories are the prey abundance hypothesis and the landscape hypothesis (Hopcraft et al. 2005). Hopcraft et al. (2005) suggested that lions prefer hunting grounds with good cover (landscape hypothesis) rather than high prey density. This could also explain partly the significance of NDVI and annual average precipitation for the human-predator conflicts. However, there are contradicting arguments claiming that the prey abundance is the factor determining the hunting grounds of carnivores (see Pike et al. 1999, Palomares et al. 2001, Spong 2002). This could have been tested if livestock densities were available to be used as explanatory variables. The basis for both hypotheses is to hunt in places where energy requirements can be met with least energy input and that pose least risk for the predator itself. Hence, the distribution of the predatory behavior is probably a mixture of these factors.
It was hypothesised that livestock depredation increases during the rainy seasons. This was not proven correct with the analysis carried out in this thesis. Patterson et al. (2004) stated that the correlations are found rather between the number of conflicts and the amount of rainfall in a given month, rather than the actual time of the year. This is due to annual variation in amount and timing of rainfall in the area. This was not taken into account in this study and clear patterns in the temporal distributions of the conflicts were not found.

9.2.1 Lion

Of all the studied species, the human-lion conflict models performed best in terms of prediction accuracy, reaching AUC scores between 0.718 (GAM) and 0.858 (CTA). The majority vote prediction for human-lion conflicts included also simpler parametric and semi-parametric models like GLM and GAM, which was not the case with all the species. Agreement between different kinds of algorithms refers to overall trustworthiness of the models. However, the models fitted very different looking response curves to some variables. Interestingly GLM reached moderate accuracy (AUC 0.74) using only two variables: proximity to PAs and population density.

Highest importance across most human-lion conflict models was given to annual average precipitation, while only GLM and ANN selected distance to PA as the most important explanatory variable. Risk of attacks was predicted to be highest in areas of low annual rainfall, which can be seen very clearly in the GBM response curve (fig. 16). This does not support the landscape hypothesis (Hopcraft et al. 2005) but, as noted before, these low rainfall areas are also used by livestock keepers and the prey abundance hypothesis might give better explanation for human-lion conflict distribution. As most of the lion ranges lie inside the PAs, the higher risk of conflict predicted closer to PAs comes as no surprise.

Most models fit the response to population density so that there is a risk peak in very small population densities, after which the response curve falls strictly down to very low risk. The responses to distance to house variable did not seem to fluctuate much in the models when other variables are averaged. GAM and MaxEnt make exceptions, GAM response showing the risk of conflict rising from almost zero to close to 90% after one kilometre. However, MaxEnt response curve shows smaller change but indicates that the risk chance decreases after one kilometre from low risk to very low risk. The importance given to distance to closest house variable, is however, very small across all the models. Lions avoid human disturbance, which is indicated by the negative
responses to population density variable. Hence, the GAM response to distance to house can be held more realistic than the response curve of MaxEnt.

GAM, GBM and RF responses to distance to water point variable presented a higher risk of conflict further away from the water points. With only GBM this increase in conflict risk was notable. GLM did not use the variable and MaxEnt response curve was slightly u-shaped (Appendix 9) showing increased risk also in areas closer (>1km) to them. Contradictory to all the other models, CTA predicted a high risk of conflict only in the close vicinity of water points. Of the models used for prediction, only GAM and CTA gave notable importance for distance to river variable (27.8% and 11.7%). With GAM, the response curve for distance to river variable did not fluctuate much with increasing distance. However, the response curve of CTA presented a very high risk of conflict near the rivers (<500m). Since GAM and CTA were the only models that notably used distance to river variable and they had distinct responses to it, it is likely that the effect of proximity of rivers on the majority vote prediction was minor. My hypothesis was that the conflicts are concentrated further away from water point and rivers. The hypothesis was proven correct only in the case of distance to water point.

Changes in the spatial distribution of human-lion conflicts between the seasons are clear. Unlike during rainy seasons, the kernel density maps (fig. 18) show no conflicts involving lions in the western parts of the study area during dry seasons. This could be explained by lions spreading more widely in search of prey during the rains. If SDM had been done for both seasons separately, the results would likely have been different.

9.2.2 Hyena

The HWCs involving hyenas are distributed very differently compared to rest of the predators. The human-hyena conflicts are located further away from the PA borders than conflicts involving lions or leopards. This is likely linked to locations of hyena dens and territories. However, these are unknown in the study area. Communal dens are centres of hyena social life and work also as shelters for cubs. Hyena denning behaviour is seen to be flexible and driven partly by presence of their main competitor, lion (Périquet et al. 2016). Lions are the top reason for hyena mortality and lions also engage in kleptoparasitism by stealing prey killed by hyenas (Trinkel & Kastberger 2005). As brought up before, also other studies have stated that predators are seen to avoid other
predators (Ramesh et al. 2017, Brook et al. 2012). For future SDM research on hyena ranges and HWC distributions, lion observations could be used as an explanatory variable.

Distance to road was chosen as the only anthropogenic variable impacting the distribution of human-hyena conflicts. The predicted risk of conflict was only slightly higher closer to roads in all models but MaxEnt, which showed a significant drop in the risk after three kilometres from closest road. The proximity of road depicts human activity, which logically increases the risk of HWC.

Study of Ogada et al. (2003), suggest that hyenas attack livestock more often near areas with vegetative cover. This is also the basis of landscape hypothesis presented earlier that tries to explain predation behavior (Hopcraft et al. 2005). However, the models produced in this thesis suggest a higher risk of a hyena attack in areas with lower NDVI. The models predicted higher chance of a conflict in areas of higher annual rainfall, which can be thought to be slightly contradictory to the responses to NDVI.

GBM response in figure 20 does not fluctuate much with increasing distance to a river or a water point. However, the other models suggested that human-hyena conflicts are concentrated in areas that are at distance of one to three kilometres from a water point and peaking again at distance of six kilometres. Response to proximity of rivers was more straightforward, suggesting increased risk further than three kilometres away from rivers. Both variables were weighted very differently across the models, having variable importances between zero and 36%. It is hard to verify the reasons behind the alternating responses to these variables, which evokes doubt about the realism of the models. Furthermore, locations of the human-hyena conflicts did not seem to change much with shifts in seasons (fig. 22).

Only the set presumptions about conflicts intensifying in areas of higher precipitation and further away from rivers could be unanimously proven correct with the used models. Overall, there were challenges in connecting some of the findings to previous theories and it is possible that significant explanatory variables were missing from the models. These could be the aforementioned den locations and the presence of other predators. For human-hyena conflicts, I achieved to create models with good to moderate prediction accuracies, but the realism of the models is debatable.
9.2.3 Leopard

Like with the lion models, human-leopard conflicts were predicted to be concentrated near the protected areas. Otherwise the predicted human-leopard conflict distributions were distinct from the human-lion conflicts. According to the models, human-leopard conflicts required also higher annual average precipitation and NDVI. This could be linked to leopards preferring hunting grounds where cover is available (Balme et al. 2007) and the landscape hypothesis (Hopcraft et al. 2005) in general. Indeed, leopards are less frequently met in areas of bush and grass savannahs, where annual average precipitation is low. Instead, leopards are seen to reach highest population densities in woodland savannahs (Friedmann & Traylor-Holzer 2008). Furthermore, models (excluding MARS) gave strong negative response to distance to water point variable which could be explained by the better prey availability near them. However, this is opposite to the set hypothesis of livestock depredation concentrating further away from water points.

Sunquist & Sunquist (2002), suggest that the main factors limiting the leopard distributions are the presence of competitors and humans. However, using anthropogenic variables in the models, decreased the prediction accuracy of the models so they were left out.

The kernel density maps showing seasonal changes in the spatial distribution of human-leopard conflicts (fig. 26) showed slight spatial changes with the seasons. Conflicts are more widely spread during the rainy seasons. This suggests that producing separate models for different seasons could have resulted in distinct predictions.

The final predictions reached moderate to good accuracies with following models: GBM, RF, MaxEnt, MARS, ANN, FDA and CTA. The predictions followed the presumptions of conflicts intensifying close to PAs, in areas with higher annual precipitation and NDVI. However, the higher risk of conflict near the waterpoints was not expected.

9.3 Challenges during the research process and possible sources of error

The author was not familiar with SDM approaches prior to the research and a lot of learning was made from mistakes. Multiple models were made for each species to find the optimal variables for each model. Also, the dataset had to be edited several times when mistakes and errors were realized. These errors included modelling conflicts involving all the predatory species together and not separating crop damage from human damages in case of human-elephant conflicts. New models
were made with separate data for each, which resulted in more accurate models. Removing parts of the data because of the electrical fence on the border Tsavo West National Park, was also done after first modelling with data spanning from 2014 to 2016. In the final models, only data prior to September 2015 was chosen for the models due to the fence being completed at the time. Leaving the data could have possibly biased the analysis.

Some of the algorithms used in SDM have tendency to overfit to the data and its impacts are possibly not noticed if data is spatially autocorrelated. Spatial autocorrelation is usually present in spatial data, also in this study. All the explanatory variables were spatially autocorrelated, at least at short distances. SDM is seen to overcome the problem of overfit by evaluating the results. However, when testing data is obtained by cross-validation, there is possibility that the testing data is not independent from training data due to spatial autocorrelation. This evokes healthy scepticism towards the models, but the majority vote prediction maps were made to adjust potential bias and inaccuracies of single models.

Land cover was not used as an explanatory variable for its insignificance for the models. However, the land cover data resolution was 300 meters compared to 100 meters of other explanatory variables. As it is possible that the agroecological zones of Taita Taveta County are strong determinants in the distribution of HWC, land cover most likely has an impact on the HWCs but it was not discovered by the used models. The importance of land cover and possibly NDVI-land cover interactions should be considered for further studies on the topic.

HWC is very seasonal phenomenon following shifts in rainfall. Seasons impact the HWCs in different ways as explained in previous sections. All the environmental variables used in the models are yearly averages and they do not take the seasonal changes into account. This is one of the deficiencies of the models that must be understood. Annual rainfall is not evenly distributed among the months of a year and also some of the rivers and water points are dry during the dry seasons. Hence, these explanatory variables are not completely realistic. An extension of this study could be trying to improve the models by adding temporal dimension to them. To compensate the missing temporal dimension of the models, the kernel density maps of the conflicts were made to see how much the distributions of HWCs change with the seasons.

The responses to distance to river and distance to water point variables behaved unexpectedly. There were significant differences between the species, between the models and between the two variables that should depict the same thing, water availability. These two variables were by far, the most problematic during the analysis. This was most likely due to the seasonal changes explained
before. However, if the algorithms did not reach consensus about the impact of these variables, their effect on the majority vote prediction decreased.

The differences between the models and modelling runs were significant with many species. For this reason, it can be assumed that the models have not been certain on the selection of significant variables and the splits of cross validation have a significant impact on the results. However, 20 modelling runs were made for each species and the predictions of the risk zones were done by combining the predictions of the best five models in an attempt to take also this into account.

What comes to the trustworthiness of the HWC data used in SDM, Kuronen explains in her master’s thesis (2016) that communities of Taita Hills are not eager to report all the conflicts experienced. However, during the fieldwork in the lowlands of Taita Taveta County, it was learned that all the HWC incidents experienced by the people I met were reported and compensation forms filled. Through this and the fact that the studied species do not reside in the higher elevations of Taita Hills, the data is considered as representative information about the phenomenon. Analysis of temporal distribution of HWCs (1990–2016) can be slightly biased because it is possible that the eagerness to report conflicts to KWS might have been low in the beginning of the 1990s because of bad relationship between the communities and the KWS. The relationship has improved since, as is the eagerness to report HWCs. However, this notion is impossible to know for sure and the whole data was used to estimate the temporal distribution of the HWCs by each species.

10. Conclusions

The whole study area in Taita Taveta County is affected by the human-wildlife conflicts. SDM succeeded to predict the phenomena with moderate to good accuracy in cases involving elephants, lion, hyenas and leopards. On average, the best prediction accuracies were reached with RF, MaxEnt and CTA. Conflicts involving cheetahs were left out of the final analysis since the models did not reach desired prediction accuracy (AUC 0.7). This was mostly due the small sample size of these conflicts

Furthermore, even with a single species, the responses to certain variables varied across different models and could even be completely contrary between different algorithms. When modelling HWCs with SDM approaches, it is strongly recommended to use several different models and compile an ensemble model or a majority vote prediction to overcome this challenge.
Explanatory variables used and the importance given to them varied significantly between the species. Logically, behavioural differences of the species explain this variation. Overall, the most influential explanatory variables across all HWCs were distance to PA and annual average precipitation. The impacts of explanatory variables towards the HWC distributions, were mostly in line with theory. However, the models’ responses to distance to river and distance to water point variables were in some cases left unexplained.

Models predicted increased risk of elephant crop raiding in areas close to the PAs with relatively low population density, low annual average precipitation and high NDVI. The models had contradictory views about the effects of proximity of rivers and water points. Seasonal changes did not affect the spatial distribution of elephant crop raiding incidents.

Lion conflicts were predicted to intensify close to the PAs, in areas with low population density and low annual precipitation. The impact of following variables was disputed by the different models: distance to house, distance to water point and distance to river. Lions conflicts were distributed more widely across the study area during the rainy seasons.

Human-hyena conflict models predicted a higher risk of conflict far away from the PAs, in areas with low NDVI but high annual average precipitation. Hyenas were seen to attack livestock close to roads but further away from rivers. The human-hyena conflict models did not reach a consensus about the impact of distance to water point variable. The spatial distribution of hyena conflicts does not vary between the dry and rainy season.

Leopards were seen to hunt for livestock near PAs and water points, in areas with higher annual precipitation and NDVI. Slight changes in the distribution of the conflicts were noticed with different seasons.

The results of this study conclude that with well thought variables HWCs can be modelled relatively realistically with moderate to good prediction accuracy. Having emphasis on accuracy and realism, it is likely that it is challenging to generalize the models into other areas. Hence, the selection of explanatory variables must be considered separately for each area the models are applied to.

Possible explanation for distribution of the conflicts, not directly explained through the variables stem from the different agroecological zones of Taita Taveta County. Areas of low rainfall in the lowlands are used for more intensive livestock husbandry while areas with more rainfall in the higher altitudes are used for more intensive agriculture. Logically, areas with more livestock also
attract more livestock depredation. Water availability is higher also in the sub-county of Taveta because of possible groundwater and run-off from Mt. Kilimanjaro. This has led to more intensive agriculture, higher NDVI and less livestock also in this area. This is a possible reason for the area’s increased risk of elephant crop raiding and smaller amount of human-predator conflicts. The higher altitude areas have denser human populations that might explain the smaller amounts of elephant crop raiding, although agriculture is intensive in these areas.

Producing separate models for different seasons or including the seasonal changes in the used variables would have likely resulted in different kinds of predictions. However, this would have complicated the modelling and the temporal dimension was left out of the SDM in this thesis. In addition to including temporal aspect in SDM, interesting follow-up research on the topic would be studying how the electrical fences affect the distribution of HWCs, or to examine if the inclusion of migratory routes or livestock densities as explanatory variables would improve the HWC models.
Acknowledgements

I want to thank my excellent research assistant and friend Mwadime Mjomba for his help and hard work. Without his contribution, this thesis would have never been written. I am also forever grateful for his whole family for welcoming me warmly into their home. I also want to thank the whole Taita Research Station staff for wonderful company and maintenance.

I want to thank Dr. Tino Johansson, Professor Petri Pellikka, Dr. Juha Aalto and Professor Miska Luoto for help and support during this process. A special thank you goes to Dr. Mika Siljander for his excellent advice and crucial contribution to this thesis. I also thank the University of Helsinki and the Department of Geosciences and Geography for the travel grant.

I wish to thank everyone in Taita Taveta County who contributed to this study. Special remarks to Mwaita Mwagodi, Fredrick Lala, Martha Nzisa, Stephen Kuseren, Alfred Parsaoti and Cyril Ringoma.

I also want to thank my friends and fellow students Antti Autio, Ruut Uusitalo, Eero Vento and Sanna Haapa for peer support and advice during the fieldwork and writing.

Finally, a big thank you to my family and friends for support during this process.
References


Pearson, R. (2014) Training course on Species Distribution Modelling (also called Ecological Niche Modelling). Lecture. University College London. Retrieved from https://www.youtube.com/watch?v=1cQ0-ybVM6U&index=4&list=PLKYTvTbXFuChaoF-L-1e9RzCagdLPQcCU 1.5.2017


Predicting wild boar damages to croplands in a mosaic of agricultural and natural areas. (2014). Current Zoology, 60(2), 170-179.

R Foundation for Statistical Computing, Vienna, Austria


Appendices

Appendix 1. Moran’s I of the used environmental variables. Red squares mean statistically significant spatial autocorrelation.
Appendix 2. RF response curves for human-elephant conflict

Appendix 3. MaxEnt response curves for human-elephant conflict
Appendix 4. ANN response curves for human-elephant conflicts

Appendix 5. CTA response curves for human-elephant conflict.

Appendix 7. GAM response curves for human-lion conflict.
Appendix 8. CTA response curves for human-lion conflict.

**Appendix 10.** RF response curves for human-hyena conflict.

**Appendix 11.** MaxEnt response curves for human-hyena conflict.
Appendix 12. MARS response curves for human-hyena conflict.

Appendix 14. RF response curves for human-leopard conflict

Appendix 15. MaxEnt response curves for human-leopard conflict

Appendix 17. ANN response curves for human-leopard conflict.
**Appendix 18.** CTA response curves for human-leopard conflict.
Appendix 19. Human-elephant conflict individual model binary predictions

Human-elephant conflict predictions

- **GBM**: Predicted conflict presence
- **RF**: Predicted conflict absence
- **MaxEnt**: Predicted conflict presence
- **ANN**: Predicted conflict absence
- **CTA**: Predicted conflict presence
Appendix 20. Human-elephant conflict individual model probability of occurrence predictions
Appendix 21. Human-lion conflict individual model binary predictions
Appendix 22. Human-lion conflict individual model probability of occurrence predictions.
Appendix 23. Human-hyena conflict individual model binary predictions
Appendix 24. Human-hyena conflict individual model probability of occurrence predictions
Appendix 25. Human-leopard conflict individual model binary predictions
Appendix 26. Human-leopard conflict individual model probability of occurrence predictions