



UNIVERSITY OF NAIROBI

MASTERS THESIS

---

# Estimating Large Carnivore population using Mixture Distributions

---

*Author:*

Vincent MULEKA

*Supervisor:*

Dr. Nelson ONYANGO

*A thesis submitted in fulfillment of the requirements  
for the degree of Masters of Science in Biometry*

*School of Mathematics*

July, 2016

**Estimating Large Carnivore population using  
Mixture Distributions**

by

Muleka Onyango Levis Vincent

A Thesis Submitted to **School of Mathematics,**  
**University of Nairobi** in Partial Fulfillment of  
the Requirements for the Degree of Masters of  
Science in Biometry.

July, 2016

Copyright

Muleka Onyango Levis Vincent

# Declaration of Authorship

I, Vincent MULEKA, declare that this thesis titled, “Estimating Large Carnivore population using Mixture Distributions” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

---

Date:

---

I, Dr. Nelson ONYANGO, declare that this thesis titled, “Estimating Large Carnivore population using Mixture Distributions” and the work presented in it have been done under my supervision. I confirm that:

- This work was done wholly or mainly while the author candidate for Masters of science degree at this University.
- I have guided the candidate throughout the research process and the writing of this thesis.
- Where quotation from the work of others has been done, the candidate has appropriately acknowledged them. With these exception, the thesis is entirely the candidate’s work.

Signed:

---

Date:

---

*“Thanks to almighty God, my family especially my wife Sarah and my daughter Tinashe my solid academic training through my very able teachers, my long list of friends especially my colleagues at work and all who impacted my life in all ways as i did journey towards this degree, today I can write this research project to the uttermost completion.”*

Vincent Onyango Muleka



UNIVERSITY OF NAIROBI

# *Abstract*

College of Biological and Physical Science

School of Mathematics

Masters of Science in Biometry

**Estimating Large Carnivore population using Mixture  
Distributions**

by Vincent MULEKA

**Context.** In order to manage any animal species well, the knowledge and understanding of how abundant they are in their habitats and their spread around the same habitat ought to be as accurate as possible. However, this is not easy given the nature of large carnivores being nocturnal and their poor relationship with human beings.

**Aims.** The main objective of the project is to estimate accurately the current population density and total of Tsavo National Park in terms of Lions and other large carnivores using playback recordings as the main method of luring them to be counted.

**Methods.** Applying mixing distributions to construct a distribution for the number of counts. Justify the use of Bayesian methods to bypass mathematical intractability of many mixtures arising from natural and biological processes. Construct an algorithm for estimating the various parameters of the model(s) by MCMC process in WinBUGS. Compare lion densities and total counts by habitat.

**Key results.** A model for estimating the number of lions in Tsavo was constructed as an MCMC algorithm in WinBUGS after it was evident that constructing a mathematically tractable equations following the conditions available was not possible. This served to both help estimate the lion density and justify the use of Bayesian methods. The model produced consistent results for parameter estimates making it the best alternative for evaluating the population density under the prevailing conditions.

**Conclusions.** Bayesian methods as implemented with MCMC algorithm provides the best alternative approach to parameter estimation especially in cases where the exact distributions are not known but their general characteristic behaviors are available as was the case here.

**Implications.** Provided the general characteristics of a phenomenon is known, MCMC sampling can be effectively used to estimate the best possible parameters. However, this needs to be checked if it is in agreement with the nearest available mathematically tractable model equations and be used as reference points...

**Key words.** Bayesian inference, MCMC sampling, WinBUGS, population density, carnivore, lion.



## *Acknowledgements*

I would like to acknowledge the following for facilitating my studies

1. First, is the almighty God for giving me understanding and guiding my step all the way.
2. Dr. Nelson Owuor Onyango for giving me careful guidance throughout the lifespan of my studies and project work.
3. HELB for facilitating my fee payment in the second year of my studies.
4. University of Nairobi through Gandhi Smarak Nidhi Fund Scholarship for the offer of scholarship during my second year of studies.
5. KWS for offering their data to facilitate my research project.
6. TSC for maintaining me and offering me a study leave.
7. Last but not least is all family and friends for standing with during the whole study period.



# Contents

<b>Declaration of Authorship</b>	<b>iii</b>
<b>Abstract</b>	<b>viii</b>
<b>Acknowledgements</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Statement of the Problem . . . . .	3
1.3 Significance of the study/justification . . . . .	4
1.4 Assumptions . . . . .	4
1.5 Objectives . . . . .	5
1.6 Research Questions/Hypotheses . . . . .	5
<b>2 Literature Reviews</b>	<b>7</b>
2.1 Introduction to Literature . . . . .	7
2.2 Method Literature . . . . .	7
2.3 Mathematical Literature . . . . .	10
2.4 Algorithm Literature . . . . .	12
<b>3 Materials and Methods</b>	<b>13</b>
3.1 Description of Survey . . . . .	13
3.2 Response Range and Probability . . . . .	14
3.3 Data . . . . .	16
3.4 Model for Count Data . . . . .	16
3.5 Development OF MCMC algorithm . . . . .	21
<b>4 DATA ANALYSIS</b>	<b>31</b>
<b>5 Conclusions and Recommendations</b>	<b>39</b>
5.1 Conclusions and Discussions . . . . .	39
5.2 Recommendations . . . . .	42
<b>Bibliography</b>	<b>43</b>
<b>6 Appendices</b>	<b>49</b>
6.1 Appendix A . . . . .	49
6.2 Appendix B . . . . .	57



# List of Figures

3.1	Map of Tsavo	15
4.1	iteration history comparison	33
4.2	kernel density comparison	38



# List of Tables

3.1	Summary of Habitat Totals . . . . .	16
4.1	Species Totals . . . . .	31
4.2	Total counts per Habitat . . . . .	31
4.3	Habitat Means and Density . . . . .	32
4.4	Summary of Lion Data . . . . .	34
4.5	Summary of Hyena Data . . . . .	34
4.6	Model 3 . . . . .	35
4.7	Model 7 . . . . .	35
4.8	Model 9 . . . . .	35
4.9	R-L Test Model 3 . . . . .	36
4.10	R-L Test Model 7 . . . . .	36
4.11	R-L Test Model 9 . . . . .	37
5.1	lion analysis . . . . .	39
5.2	Hyena analysis . . . . .	40
6.1	Model 1 . . . . .	49
6.2	Model 2 . . . . .	50
6.3	Model 3 . . . . .	51
6.4	Model 4 . . . . .	52
6.5	Model 5 . . . . .	53
6.6	Model 6 . . . . .	54
6.7	Model 7 . . . . .	55
6.8	Model 8 . . . . .	56
6.9	Model 9 . . . . .	56
6.10	Model comparison . . . . .	58





# List of Abbreviations

<b>MCMC</b>	<b>Markov Chain Monte Carlo</b>
<b>WinBUGS</b>	<b>Windows based Analysis Using Gibbs Sampling</b>
<b>TNP</b>	<b>Tsavo National Park</b>
<b>TSC</b>	<b>Teachers Service Commission</b>
<b>KWS</b>	<b>Kenya Wildlife Service</b>
<b>HELB</b>	<b>Higher Education Loans Board</b>



*My beautiful daughter Tinashe Liana for  
gracing my life at the opportune time. May  
God guide you in everything you do in your  
life...*



# Chapter 1

## Introduction

### 1.1 Background

Ecological dynamics are a great influence on the population of any species. That explains why some animals have carved an ecological niche for themselves. The lion is the king of the jungle for instance. There are other ecosystems where there is no clear king but many of the animals in those niches share the habitat with each carving its own domain and in return respecting the other's domains. Dorazio and Connor (2014) correctly asserts that the desire to understand pattern variation in species abundance in any habitat has always been the driving force behind ecological research. A clear understanding of the interaction between various occupants of any given ecosystem breeds a balance and ensures that harmony prevails. Take the food chain for instance, large carnivores are basically at the top. To make it more interesting, there are several of them. By being many of them in the same category means they share the available resources. There can only be so much prey in any given habitat. This means that the number of the predators must be kept in check to prevent a scenario where they have fed on all available prey and remain to starve to death. Yirga et al. (2014) suggest that the population of large carnivores worldwide has been on the decline due to a conglomerate of factors among them action of human 'persecution' and degradation of habitats. The situation is doubly tricky for circumstances where the habitat is limited by other factors. Limited habitats include national parks and game reserves even though the whole earth itself is a limited resource. On the other hand, the human factor in these dynamics is not to be assumed or underrated. Dolrenry et al. (2014)

states that the population of African lion in Kenya and Tanzania is shrinking because of encroachment that human beings are doing on their habitats. It is only when these human wildlife conflicts are reduced to the bare minimum that we can hope to conserve the quickly dwindling large carnivore species Thorn et al. (2013). It's at this point that we focus more on this bound habitats. Being resources that are closely monitored and under careful management, those tasked with such are to be utterly aware of the dynamics of population change among the animals they are managing in those ecosystems. Large carnivores among them lions, hyenas, leopards and cheetahs being at the top of the food chain become candidates for great interest by all and sundry in ecology and animal conservation studies Young-Overtton, Funston, and Ferreira (2014). Because many tourist are attracted to viewing large carnivores in their free ranging habitats Borg et al. (2016), those kinds of habitats must be conserved by any country and state interested in attracting such caliber of tourists. And in most African countries, tourism is still a major determinant of economic growth and development. The national parks and game reserves have been source of income for many African nations for a long time. Tourists from both within the country/continent and outside visit the parks to enjoy a view of the charismatic animals Winterbach, Whitesell, and Somers (2015). They go ahead to argues that the tourism aspect acts as an incentive to incite the conservation of the wildlife. Others also visit for trophy hunting. Studies may focus on abundance of prey which helps manage endangered species Ariefiandy et al. (2013). Other studies just focus on estimating the abundance of the species of interest without much reference to the prey factor. Irrespective of the approach, the management of the national parks and game reserves have to be able to estimate the number of animals in those ecosystems they are in charge of. In order to manage any animal species well, the knowledge and understanding of how abundant they are in their habitats and their spread around the same habitat ought to be as accurate as possible. This is quite an uphill task unless substantial research work is done to come up with some of the effective methods of estimating population size which in effect is just the culmination of abundance and distribution. Cozzi et al. (2013) argues that despite the importance of getting accurate

estimates especially of the large carnivores for effective habitat management, it is challenging to obtain. Ariefiandy et al. (2013) tries to find environmental drivers of the variations in population of lions and hyenas in Kruger National Park. Their focus is the effects of rainfall on behavior of the underscored carnivores.

## 1.2 Statement of the Problem

The concern is how best to determine the population of large carnivores in the national parks and game reserves. These being national resources under management especially for purposes of keeping the ecosystem and for financial gains as tourist attraction centers. Of course tourism has crawled into a major foreign exchange earner in Kenya Revord Larkin (2014). Its importance cannot be overemphasized. Revord Larkin (2014) argues that the conflict between human and wild animals can only make the industry grow from bad to worse. Even though most of the conflict arise when the large carnivores encroach on livestock, Osano et al. (2013) shows that the tourism income equals and to some extent exceeds income from the said livestock. The large carnivores though have been shown to encroach on livestock only when the wild prey in those parks have reduced to the utter minimum Khorozyan et al. (2015). This explain the importance of knowing the numbers so as to prevent the conflict that will arise due to lions and hyenas going out of their way to attach livestock because of famine in the parks. With such facts at hand there is need to know exactly how many of the species there are so as to put in place proper strategies for taking care of them and remain in business as a tourist attraction center. There has been a consistent decline in the large carnivore population as was shown by Lamprey and Reid (2004). With all that said and known, it is therefore important for appropriate ways to properly estimate the number of large carnivores.

### 1.3 Significance of the study/justification

Calling animals to call station by playback recordings is among the most effective methods of assessing density and distribution of the elusive carnivores Cozzi et al. (2013). This is due to the fact that these carnivores shy away from human beings who have been known to persecute them Borg et al. (2016) for one reason or another but mostly through conflict arising from the carnivores encroaching on livestock Khorozyan et al. (2015) or through the famous trophy hunting Cozzi et al. (2013) which has been shown to negatively impact the population of wild life especially the lions. Well known methods of estimating wildlife population have been in use. Methods like capture-mark recapture, roar counts, aerial counts, individual identification for total counts, setting up camera traps etc. have been in use for a very long time. However, they exhibit a lot of limitations including but not limited to underestimating the abundance of carnivores, low encounter rates in woodlands, time consuming for large areas to be covered, unusually expensive, carnivores like lions having no defined routes for cameras to capture them etc. Kiffner et al. (2008), Mills, Juritz, and Zucchini (2001), and Ogotu and Dublin (1998). The researches above also showed that playback sounds are quick and cheap to carry out, they do not need intensive and expertise training, are less invasive on the animals' environment and most importantly can allow for determination of habitat specific population densities.

### 1.4 Assumptions

Even though research has shown that environmental confounders affect the behavioral responses of lions to any spotting, Young-Overton, Funston, and Ferreira (2014) makes the assumption that any environmental variations present during the time of experiment are not so extreme as to alter the expected responses. Although not all lions in the response area approximated to be circle of radius 2.5km which was the lower bound for the calibration experiment in Kruger National Park by Ferreira and Funston (2010) responded to calls due to various reasons including but not limited to those already having a



carcass, mothers nursing their cubs, sick etc. the response probability of the same animals ought to take care of such non-response cases and reasons Mills, Juritz, and Zucchini (2001).

## 1.5 Objectives

The main objective of the project is to estimate accurately the current population density and total of Tsavo National Park in terms of Lions and other large carnivores like Hyenas, jackals, cheetahs, using playback recordings as the main method of luring them to be counted. This will be achieved by;

- Applying mixing distributions to construct a distribution for the number of counts.
- Justify the use of Bayesian methods to bypass mathematical intractability of many mixtures arising from natural and biological processes.
- Construct an algorithm for estimating the various parameters of the model(s) by MCMC process in WinBUGS.
- Compare lion densities and total counts by habitat.

## 1.6 Research Questions/Hypotheses

Considering the task of estimating the number of lions in Tsavo National Park, there are many questions that need to be answered by this project. In conjunction with our objectives, the following research questions will be assessed;

- Does the parameter  $\phi$ - the response probability for lions to any calling station- from MCMC process agree with that from calibration experiment Ferreira and Funston (2010) which is our reference at 5% confidence?
- Does the response probability vary from lions to hyenas as suggested by the works of Young-Overton, Funston, and Ferreira

(2014), Mills, Juritz, and Zucchini (2001), Ogutu and Dublin (1998), and Ferreira and Funston (2010)?

- Are the mixing distributions used in constructing the model the only available options or can there be other alternatives?
- Is the algorithm constructed the best alternative for deriving the parameters of the models?
- Does the response probability remain the same or vary by habitat?
- In which habitat is the population density highest?

Following the objectives and the research questions stated above, the main hypotheses of the project are;

- Due to their biological and social structure differences, lions respond differently to call stations on hearing the sounds than any other animals. This implies that the response probability  $\phi$  will be different for each and every different carnivore.
- The response probability is the same for all habitats in agreement with other studies.
- The algorithm arrived at is the best alternative out of many others.
- Population density varies by habitat since lions prefer some ecological environments to others.

# Chapter 2

## Literature Reviews

### 2.1 Introduction to Literature

Literature review was looked at in three different categories to effectively assess the available material to necessitate the construction of the models in this paper and effectively justify the choice

### 2.2 Method Literature

A lot has been done in terms of using sound playback to attract carnivores to a call in station so that the abundance of the species can be estimated. Ogutu and Dublin (1998) estimated the population of lions in Masai Mara national reserve-Kenya. The main focus was what other factors may have affected the response of the lions. The response probability from the calibration experiment only showed about 25 percent of lions responding. They used sampling design approach and not the model approach.

Mills, Juritz, and Zucchini (2001) applied the same approach to lure spotted Hyenas in Kruger national park-South Africa. Using a model approach instead of sampling design approach. They employed mixture modelling using the gamma distribution. Making assumption of uniformity in habitats, mathematical tractability was achieved. Unlike Ogutu and Dublin (1998), they did calibration experiment for the response range but employed maximum likelihood profiling to estimate the response probability for the Hyenas.

Kiffner et al. (2008) focused on what experimental principles to adhere to in order to achieve the best results with sound playback as a

method for estimating population density and size for spotted hyenas and lions. The study is carried out in Katavi national park- Tanzania. The focus of the study was investigating the effect of time of playback, phase of the moon and presence or absence of other species on the response of spotted hyenas and lions. They also sought to test the assumptions that the animals were lured from a circular plot and that all the lured animals are observed. They found that time of night and phase of the moon did not sufficiently affect the luring of lions. Because of the choice of open places for stations, all animals lured must have been detected and the model based approach satisfied most if not all carnivore characteristics hence was proved to be efficient.

In a slightly different perspective, Ferreira and Funston (2010) evaluates the population variables on the lion population in Kruger national park- South Africa and especially the perceived effects of bovine Tuberculosis. The actual aims were to calibrate call-up stations and define survey effort needed to get estimates with known precision, and extract from them age structures and survival rates estimates, to estimate number of lions and then to evaluate impact of bovine tuberculosis (bTB) on lion density and survival estimates. The approach used is using call-up stations to estimate selected population variables and evaluated the perceived threat of bTB in landscapes of varying prey biomass in the Kruger National Park. The main results were as follows;

- The size of the lion population was stable, although long intervals between present and historical estimates limit this conclusion.
- Density and survival rates associated positively with prey biomass, and a positive association was detected between the survival rate and bTB prevalence, with survival being higher in areas that had high prevalence of bTB.
- Male survival was lower than female survival, disregarding the effects of prey biomass or bTB prevalence.
- Body condition of lions was high, with scores lower at low to medium prey density.

Similarly, Bauer (2007) on determining the status of lions population

in Bouba Ndjida National Park, Cameroon uses sound playback to lure the lions from their dens. From the hypothesis that the region being studied has been affected by infiltration of varied human activities, the results confirm the same as only nine lions are observed. Because of the extreme low biomass density, the calibration experiment could not be done despite the recommendation. The result was that the response range and probability had to be assumed from previous works by Mills, Juritz, and Zucchini (2001) and Ogutu and Dublin (1998) that effective response range is 2.5km and 3.2km respectively with the response probability of 26% for lions. To complete their work, they had to postulate a response probability of 75% and response range of 2.5km.

Also the works of Ogutu, Bholá, and Reid (2005) on how pastoralism in protected areas affect the population density and distribution of large carnivores and their prey use sound playback to attract the animals for counts. The hypothesis of pastoralism and protection influence carnivore density and distribution in the Mara ecosystem of Kenya was tested by:

1. estimating the density and size of lions, spotted hyenas and jackals;
2. characterizing their spatial distribution inside and outside the reserve and
3. establishing the extend to which land management influences carnivore population levels and spatial distribution by controlling for the effects of grass height, elevation and prey biomass density

The result of the study was as follows;

- Lions never responded to playbacks in the ranches, so the potential shift in lion behavioural response for different land use zones is another potential explanation for the patterns found here.
- A real shift in lion populations is a better explanation than a behavioural change in relation to playbacks based on additional

data from independent systematic and intensive censuses and playback surveys conducted in the ranches.

- Lion populations in the pastoral ranches seemed headed for extinction, probably owing to conflicts with pastoralists, necessitating urgent conservation interventions that integrate pastoral economic welfare with large carnivore conservation goals to foster long-term viability of lion populations in the pastoral systems.

## 2.3 Mathematical Literature

Mixture modelling provides a more flexible way of modelling data that is heterogeneous Jasra, Holmes, and Stephens (2005) and because they can be used to approximate large class of functions, they have been employed to describe nonstandard distributions. They go ahead to argue that mixture modelling in the frequentist approach of maximum likelihood has the disadvantage of remaining unbounded for location scale component models. However, the advent of MCMC processes introduced in the Bayesian approach helps bypass that impasse using what is known as the reversible jump MCMC algorithm. Mixture modelling has had many applications, for instance Kupzyk and Unit (2011) in his lectures uses mixture modelling to assess link types in trajectories and in fact determine homogeneity in developmental arrays from data collected over time. Some of his conclusions are;

- Can help identify at-risk individuals.
- Flexible (can use categorical or continuous outcome and predictor variables; model cross-sectional or longitudinal data.
- Useful for condensing a large amount of information in order to see patterns in data.

- Useful for when groups are unknown.
- Avoids some of the problems of traditional clustering methods.

Similarly, Sawatzky and Richardson (2009) uses mixture models to explore the density of a sample utilizing information of latent variables. They concluded that most of the variation in the latent variable remained unexplained. Also categorization of the latent variable was consistent in all the models.

Theobald, Chatterjee, and Horgan (2012) in constructing a model for analyzing dietary records uses mixture modelling. First, they considered how finite mixture models could be used in cases where data appears as repeated records, and then apply a Bayesian version of one such extension to data on the consumption. They also illustrate how factors such as sex and age may be included in the model. The approach of mixture-modelling provided better estimates as compared to the alternative methods of the probability distributions of daily consumptions and of maximum consumption over a number of days.

In estimating the abundance of carnivores with remote camera traps-a mark re-sight approach- Alonso et al. (2015) uses mixture modelling to compare the traditional method mark recapture with the new method mark re-sight. This led to the conclusion that the new estimator may be particularly useful for estimating abundance of uniquely identifiable species that are difficult to sample using camera traps alone.

On the same note, Buckland et al. (2014) on determining the effects of the invasion of the giant Madagascar day gecko on the four endemic Mauritian day geckos, uses binomial mixtures in their species distribution models. An ensemble forecasting approach was used to predict the spatial distribution and hierarchical binomial mixture models with repeated visual estimate surveys to calculate the abundance. Their conclusion was that species distribution modelling, together with the breadth metrics, predicted that Madagascar day gecko can partly share the equivalent niche with endemic species and survive in a range of environmental conditions. We provide strong evidence that smaller endemic geckos are unlikely to survive in sympatry with

Madagascar day gecko.

Abas (2013) On determining efficient finite mixture models with compact and essential components for clustering data proposes an algorithm that can be used effectively to evaluate and determine the best finite mixture model for clustered data. His results is that the proposed algorithm EMCE is superior to all other algorithms in the literature, especially with small data sets that are sparsely distributed or generated from overlapping clusters.

## 2.4 Algorithm Literature

Fryback, Stout, and Rosenberg (2001) does an elementary introduction to Bayesian computing using WinBUGS. They aimed to provide an elementary tutorial of WinBUGS to enable performance of Bayesian Statistical analysis. They also provided a background on the computational methods used by the software. They scripted simple algorithms that can be used by novices as they start out creating their own. They also provided few working examples which could be implemented by a learner. They were able to provide a schematic overview of how the program works.

Sturtz, Ligges, Gelman, et al. (2005) introduced an R package for running WinBUGS. They assists in giving directions of how to write R scripts in a language that the WinBUGS program can understand. Their argument of implementing in R is justified by the fact that R-GUI provides the necessary tools to do further analysis on the output from WinBUGS



# Chapter 3

## Materials and Methods

### 3.1 Description of Survey

The description below follows in the footsteps of the work of Mills, Juritz, and Zucchini (2001). A 6 min long tape of sounds known to attract lions was made. Sounds used included the bleating of a wildebeest (*Connochaetes taurinus*) calf, hyenas mobbing lions, and lions competing on a kill. This combination of sounds was used in order to attract lions under a range of conditions; i.e., not only when they might be hungry, but also when, even satiated, social circumstances would be expected to attract them. The sound was played through a 4000 Report-L tape recorder attached to a 12 V mobile amplifier and two 45.7 cm, 8 ohm, horn speakers, connected in series and pointing in opposite directions. One survey was conducted, between X and Y dates. On the basis of the landscapes, the TNP was consolidated into six habitats. Each habitat was sampled roughly in proportion to its area. Prior to the first survey, randomly chosen routes along the extensive road system in the TNP were drawn onto a map, so that the entire 20,000 km<sup>2</sup> TNP was sampled. Each night of the survey a route of approximately 100 km was driven. At approximately 10 km intervals along the route the vehicle was stopped at a suitable point and the tape was played. We selected as open an area as possible so as to obtain the best visibility. In some habitats this was not always easy, but crossroads which gave good four-way visibility as well as slightly elevated stretches of road often helped. Thirty minutes was spent at each station. After 3 min of playing, the speakers were turned through 90°. Approximately 5 min after the end of the first playing the tape was replayed and again the speakers were

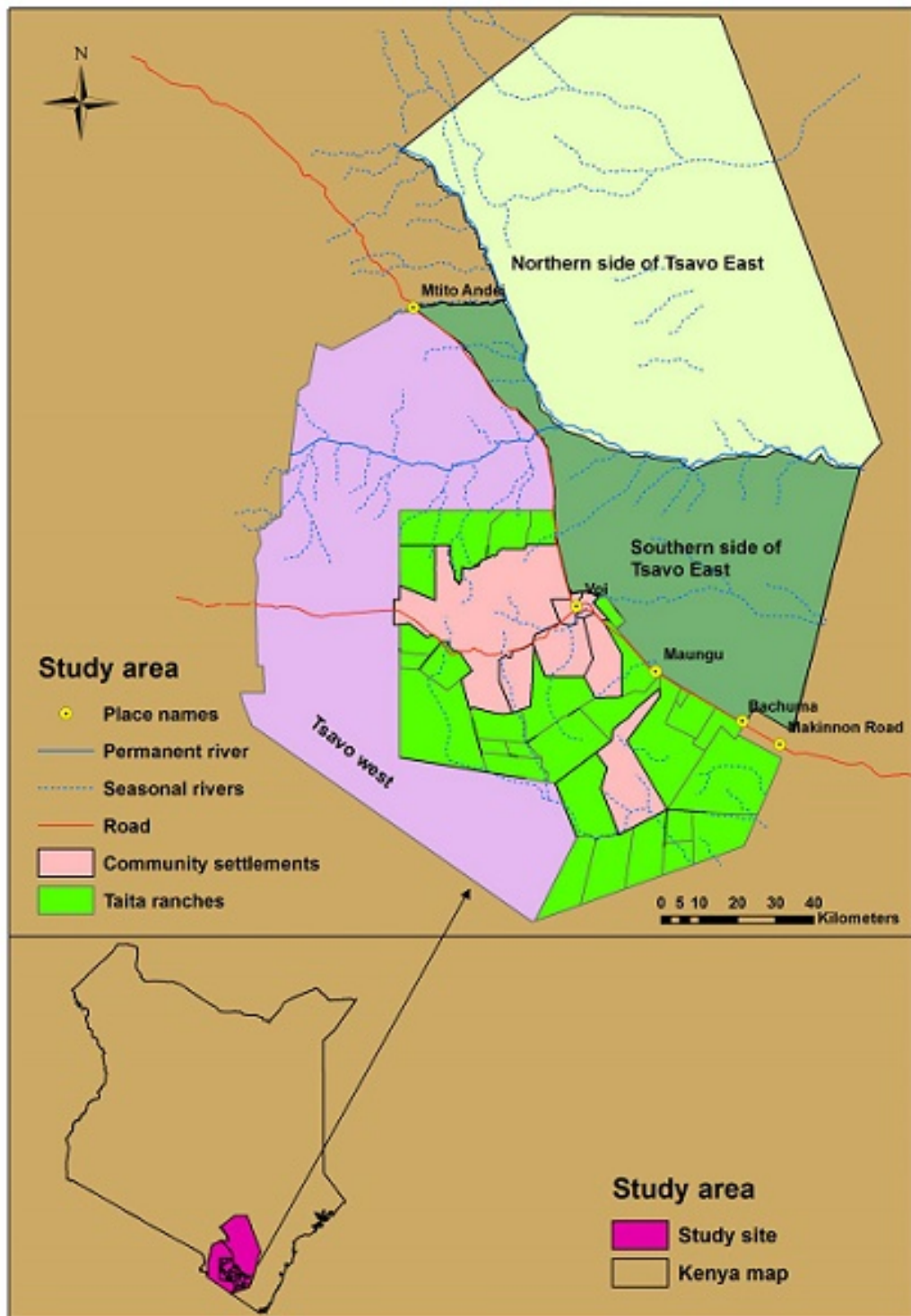
turned 90 after 3 min. If lions were heard in the vicinity, but did not appear, the tape was played a third time.

Two or three observers, each armed with a spotlight, stood on the open back of the vehicle, a small truck, so that eye height was about 3 m above the ground. From time to time the observers scanned the vicinity with the spotlights. Lions are easy to see with a spotlight as the light is reflected in their eyes and their characteristic gait can quickly be identified. Any lions or other carnivores that were located were carefully monitored so as to prevent double counting. Environmental conditions were held as constant as possible, particularly with regard to wind, and call-ups were not done if the wind exceeded 2 on the Beaufort wind scale. On each night a different and non-overlapping area was sampled. Consecutive calling stations were at least 10 km apart and the drive between them took at most 40 min. These facts make the chances of double counting extremely remote.

## 3.2 Response Range and Probability

Many studies have been done to establish the most appropriate response range for lions and hyenas to calling stations using playback sound recordings. In this study, the results of Ferreira and Funston (2010) will be used. They found the optimal response range as 4.2km. However, the lower bound of 2.5km will be most appropriate to avoid the likelihood of the model overestimating the population size of the lions and hyenas in Tsavo national park. However, for the response probability, the use of independent experiments will not be possible hence the MCMC algorithm constructed will be used to evaluate best approximation of this parameter  $\phi$ , the probability of spotting an animal.

FIGURE 3.1: Map of Tsavo



### 3.3 Data

There were 116 calling stations spread out in the vast areas subdivided into 6 different habitats. Out of all that lions turned up at 22 stations (18.96%) and hyenas turned up at 70 stations (60.34%). The number at each station ranged from 0- 5 and 0- 16 for lions and hyenas respectively. Hyenas turned at 13 out of 22 stations (59.09%) were lions turned up and they were mostly equal in number or slightly more than the lions.

TABLE 3.1: Summary of Habitat Totals

Habitats	Min of Area (KM2)	Sum of Lion	Sum of Hyena	Count of Call_stn
CDF	89.87	0	10	2
CG	5913.15	12	65	24
CL	3349.79	2	34	19
DSST	1145.1	0	25	7
OG	1137.85	0	12	7
OGSS	8117.81	31	116	57
Grand Total	89.87	45	262	116

### 3.4 Model for Count Data

The number of lions and hyenas that responded to the different calling stations varied widely. There may have been many reasons for that but chief among them will be distribution of animals in their respective habitats Mills, Juritz, and Zucchini (2001). As they suggested and has been the case, this is made lighter by setting up stations in specific habitats so some kind of strata are created. The model therefore is fitted and tested per habitat. At any given random calling station, there may have been  $Y$  lions who were called. But due to many reasons mentioned early, not all of them responds. Therefore, let's say only  $X$  lions respond. The interest is determining how the lions are distributed in the entire park and therefore the distribution of  $Y$  and especially its mean will be key to this. Of course only the values of  $X$  are observed and not the values of  $Y$ .

From Unwin (2012) who defines Bayes' Rule as a consequence of conditional probability as follows;

Assuming that both  $P(A) > 0$  and  $P(B) > 0$  we can rewrite the definition of conditional probability as

$$P(A \cap B) = P(A/B)P(B)$$

or equivalently as

$$P(A \cap B) = P(B/A)P(A)$$

Setting the right-hand sides of the above equations equal to each other we have

$$P(B/A)P(A) = P(A/B)P(B),$$

which yields the formula originally discovered by Bayes

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)}$$

In the same accord therefore,

$$P(X = x) = \sum_{y=x}^{\infty} P(X = x/Y = y)P(Y = y) \quad (3.1)$$

Moving forward, there is need to define appropriate distributions for the total population Y and also for the conditional of X on Y. Considerations here are whether or not mathematically tractable results is possible and more importantly is the physical circumstances of the study. The two are sufficient to guide proper models choices. From the data, both lions and hyenas are turning up in varying numbers at the calling stations and different habitats. Therefore, the thought of the animals being uniformly distributed in the entire Tsavo national park may be inappropriate. The number of lions near any calling stations chosen at random has to have a distribution of sorts. Just like Mills, Juritz, and Zucchini (2001) the choice taken here is Poisson distribution with mean lambda.

$$P(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!} \text{ Where } y = 0 \ 1 \ 2 \ \dots \quad \lambda > 0 \quad (3.2)$$

Since as observed from the data, the mean and variance  $\lambda$  seems to vary from station to station and habitat to habitat, there is need to allow the  $\lambda$  to have its own distribution. Finding an appropriate mixing distribution for the Poisson is therefore necessary. As Suk-saengrakcharoen and Bodhisuwan (2014), states, the mixture distribution is one of the most important ways to obtain new probability distributions in applied probability and several research areas. Considering the properties and characteristics of  $\lambda$ , another of candidate distributions exist;

1. *Gamma*( $\alpha, \beta$ ) as in Mills, Juritz, and Zucchini (2001)
2. *Exponential*( $\beta$ ) which will be considered in this paper given its close relationship with the Gamma.
3. The Log-Normal distribution.

$$f(\lambda) = \beta e^{-(\beta\lambda)}, \text{ where } \lambda, \beta > 0 \quad (3.3)$$

This means that,

Given that the mixing distribution is continuous, the distribution of  $Y$  is therefore given by

$$\begin{aligned} P(Y = y) &= \int_0^{\infty} P(Y/\lambda)P(\lambda)d\lambda \\ &= \int_0^{\infty} \frac{e^{-\lambda}\lambda^y}{y!}\beta e^{-\lambda\beta}d\lambda \\ &= \frac{\beta}{y!} \int_0^{\infty} e^{-\lambda(1+\beta)}\lambda^y d\lambda \end{aligned}$$

Letting  $R = (1 + \beta)\lambda$  and hence  $d\lambda = \frac{1}{1+\beta}R$

$$\begin{aligned} P(Y = y) &= \frac{\beta}{y!} \int_0^{\infty} e^{-R}R^y \left(\frac{1}{1+\beta}\right)^{(y+1)} dR \\ &= \frac{\beta}{1+\beta} \frac{1}{y!} \int_0^{\infty} e^{-R}R^{(y+1)-1} dR \\ &= \left(\frac{\beta}{1+\beta}\right) \left(\frac{1}{1+\beta}\right)^y \frac{\Gamma(y+1)}{y!} \end{aligned}$$

Whereby from the definition of the  $\Gamma$  function,

$$\Gamma(\alpha) = \int_0^{\infty} e^{-x} x^{(\alpha-1)} dx$$

Therefore,

$$P(Y = y) = \left( \frac{\beta}{1 + \beta} \right) \left( \frac{1}{1 + \beta} \right)^y \quad (3.4)$$

Which means that

$$Y \sim \text{geom} \left( \frac{\beta}{1 + \beta} \right)$$

Going back to equation [1.1], which is the main focus here and having analytically acquired a distribution for Y, then we need to make an assumption concerning the distribution of X conditioned on Y to be binomial since X is capped by the Y i.e the number of lions actually observed is only a fraction of the actual number of lion within any given calling station. Also we are going to take the probability of success in this binomial to be the response probability ( $\pi$ ).

Therefore, we have,

$$P(X = x) = \sum_i^n \binom{y_i}{x_i} \pi^{x_i} (1 - \pi)^{y_i - x_i} \left( \frac{\beta}{1 + \beta} \right) \left( \frac{1}{1 + \beta} \right)^{y_i} \quad (3.5)$$

On simplification, we have

$$P(X = x) = \left( \frac{\beta}{1 + \beta} \right) \sum_{i=1}^n \binom{y_i}{x_i} \left( \frac{1 - \pi}{1 + \beta} \right)^{y_i - x_i} \left( \frac{\beta}{1 + \beta} \right)^{x_i} \quad (3.6)$$

At this point, the assumption of the parameter  $\beta$  is assumed to be sufficiently small to allow the binomial-like component of the equation [6] to be equivalent to 1 making.

$$P(X = x) = \frac{\beta}{1 + \beta}, \forall \beta \neq 1, x = 0, 1, 2 \dots \quad (3.7)$$

So that  $X \sim \text{unif} \left( \frac{1 + \beta}{\beta} \right)$

With the interest to determine the  $E(X)$ , which basically should be the same as the  $E(E(X/Y))$ . In this case however, the assumption causes the two expressions to be different.

$$E(X) \neq E(E(X/Y))$$

Since the expression of equation [1.6] generally appears like a binomial with the exact parameters unknown then its necessary to find a way around determining the most appropriate parameters of this particular binomial. This can be best achieved by employing MCMC processes. From the work of Mills, Juritz, and Zucchini (2001) whereby they used a distribution for  $\lambda$  in equation [1.3] as

$$f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \quad (3.8)$$

And using it as a mixing distribution for  $P(Y = y/\lambda)$ . This enabled the distribution of Y to be derived as;

$$P(Y = y) = \int_0^\infty \frac{e^{-\lambda} \lambda^y}{y!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda \quad (3.9)$$

Which on simplification became;

$$\begin{aligned} P(Y = y) &= \frac{\beta^\alpha}{y! \Gamma(\alpha)} \int_0^\infty e^{-\lambda(1+\beta)} \lambda^{(y+\alpha-1)} d\lambda \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(y + \alpha)}{(1 + \beta)^{y+\alpha}} \end{aligned} \quad (3.10)$$

By the definition of the gamma function and appropriate substitution. The outcome was that Y was found to be negative binomially distributed with parameters  $\alpha$  and  $\beta$  Using equation [1.2] and equation [1.10] as the mixing distribution with the same assumptions that X conditioned on Y is binomially distributed with  $\pi$  as the response probability. This enabled them to get that;

$$P(X = x) = \sum_{i=1}^n \binom{y_i}{x_i} \pi^{x_i} (1 - \pi)^{y_i - x_i} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(y_i + \alpha)}{(1 + \beta)^{y_i + \alpha}} \quad (3.11)$$

Despite the original requirement that  $\lambda$  has a distribution and is not rather constant, they had to assume that  $\pi = 1$  which implies that  $y_i = x_i$

Which simplified the distribution of X to;

$$P(X = x) = \sum_{i=1}^n \pi^{x_i} \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(x_i + \alpha)}{(1 + \beta)^{x_i + \alpha}} \quad (3.12)$$



And simplifying the sums yielded

$$P(X = x) = \binom{y_i}{x_i} \left( \frac{\beta\pi}{1 + \beta\pi} \right)^\alpha \left( \frac{1}{1 + \beta\pi} \right)^{x_i} \quad (3.13)$$

Therefore the distribution of counts was found to be

$$X \sim \text{negbin} \left( \alpha, \frac{\beta\pi}{1 + \beta\pi} \right)$$

Considering the unrealistic nature of the assumptions made by Mills and the ones made in this project in an attempt to make an analytic evaluation of the mixture of distributions. This still shows how important it is to revert to MCMC sampling processes to evaluate the parameters for the appropriate distribution of the observed number of lions.

### 3.5 Development OF MCMC algorithm

Both approaches outlined and summarized into equations [1.6] and [10] point to a binomial-like structure in the distribution of the number of observed counts. Fryback, Stout, and Rosenberg (2001) gives an outline on model specification in WinBUGS. Following the cue therefore, we propose that the distribution of  $X$  be binomial but with both parameters unknown which will incite a reason to find an appropriate distribution for both both  $p$  and  $k$  in the  $X \sim \text{bin}(p, k)$ . The probability of success  $p$ , is the measure of response rate for lions to the calling stations. Since it is a continuous variable between 0 and 1, then we suggest that to use a better distribution hence  $p \sim \text{dunif}(\pi, 1)$  where  $\pi$  is the actual response rate that are interested in. However, there is a possibility of modeling  $p$  as a beta distributed random variable but the challenge is that the beta distribution does not provide us with an opportunity to insert the response probability  $\pi$  in its specification as readily as the uniform distribution does. On the other hand the appropriate distribution for  $k$  should consider its discrete nature and since the previous definitions as outlined in equation [2] above allows two routes that it can take. Then we will

implement both routes and find the most appropriate. First, if the response probability was assumed uniform then;  $k \sim dpois(\lambda)$  would be the choice of its distribution. However, if the response probability is not constant, then the derivations of equations [4] and [9] suggest geometric and negative binomial respectively. Assuming that  $k \sim dgeom(\alpha)$ , no algorithm acceptably runs. For the case of  $k \sim dnegbin(\alpha, \beta)$ , the choice of the appropriate prior distributions for both  $\alpha$  and  $\beta$  is very important to enable the model to compile and run.

#### Case 1

Let the case involve  $X$  binomially distributed with parameters  $p$  and  $k$ . And followed by  $p$  being uniformly distributed containing the response probability  $\pi$  and  $k$  Poisson distributed with parameter  $\lambda$ , then the following WinBUGS codes for the models are used.

```
for i in 1:N {
    x(i) ~ dbin(p(i), k(i))
    p(i) ~ dbeta(delta, theta)
    k(i) ~ dnegbin(alpha, beta)
}
```

The prior distributions for  $\lambda$  and  $\pi$  are such that they are non-informative. Candidate distributions for  $\pi$  are continuous uniform (0, 1) or beta, whereas those for  $\lambda$  are gamma or Log-normal. The choice of the distribution for the parameters  $\pi$  and  $\lambda$  and the subsequent initial values assigned to it will determine the compilation of the distribution of  $X$ . More so, the model must have the appropriate environment for initiation of the MCMC process. That explains why some distributions cannot be candidates for the various parameters. Similarly, the distributions of the various parameters can also be informed by whether the expected behavior of the parameter. For instance,  $\pi$  is expected to be continuous ranging from 0 to 1 since it is a probability. This leaves only uniform and beta distributions as the only candidates. On the other hand, the parameter  $\lambda$  is continuous ranging from 0 to  $\infty$ . This leaves only gamma, exponential, Weibull and log-normal as the candidates. The model now needs the specifications of the priors to be complete. For the above case we have,

```

{
    p(i) ~ dunif(0.0, 1.0)
    k(i) ~ dgamma(10, 10)
}

```

The final step would be to write down the data in a form that WinBUGS do recognize. The simplest being writing as a list. The same goes for the appropriate initial values that will enable the MCMC process to start. The complete model statement therefore appears as

Model

```

{
  for i in 1:N {
    x(i) ~ dbin(p(i), k(i))
    p(i) ~ dbeta(delta, theta)
    k(i) ~ dnegbin(alpha, beta)
  }

  pi ~ dunif(0.0, 1.0)
  lambda ~ dgamma(10, 10)
}

```

On running the model, its goodness is measured by a number of things namely;

- Checking its density graph shows how well the simulation assimilates the actual distribution of the parameter in question.
- The estimate of the parameter is provided together with its standard deviation. In addition, the best way to check for the goodness of fit of the estimated value of the parameter is the MCMC error which requires to be as small as possible.
- As the process iterates, the simulation plots a curve. Observing if the curve shows a sign of convergence provides another check on how well the estimate fits the distribution for the parameter in question.

Provided below are an alternative model that were created to test Model

```
{
for i in 1:N {

    x(i) ~ dbin(p(i), k(i))
    p(i) ~ dbeta(δ, θ)
    k(i) ~ dnegbin(α, β)

}

λ ~ dgamma(10, 10)
π ~ dbeta(100, 100)

}
```

### Case 2

Let the case involve  $X$  having binomial distribution with parameters  $p$  and  $k$ . And followed by  $p$  being uniformly distributed containing the response probability  $\pi$  and  $k$  distributed as negative binomial with parameters  $\alpha$  and  $\beta$ , then the following winBUGS codes for the models are used.

Model

```
{
for i in 1:N {

    x(i) ~ dbin(p(i), k(i))
    p(i) ~ dbeta(δ, θ)
    k(i) ~ dnegbin(α, β)

}

}
```

The same procedure as above is followed and since the parameters  $\pi$  and  $\alpha$  are measures of probability, their non-informative priors can only take on the candidature of uniform and beta distributions. However, the parameter  $\beta$  could take on the candidatures of uniform, Poisson and gamma distributions with proper specifications. Below are the possible complete model specifications that were created in an attempt to get the best

## MODEL 1

Model

```

{
for i in 1:N {

    x(i) ~ dbin(p(i), k(i))
    p(i) ~ dunif(π, 1)
    k(i) ~ dnegbin(α, β)

}

α ~ dunif(0.0, 1.0)
β ~ dgamma(10, 10)
π ~ dbeta(100, 100)

}

```

## MODEL 2

Model

```

{
for i in 1:N {

    x(i) ~ dbin(p(i), k(i))
    p(i) ~ dunif(π, 1)
    k(i) ~ dnegbin(α, β)

}

α ~ dunif(0.0, 1.0)
β ~ dgamma(100, 100)
π ~ dunif(0.0, 1.0)

}

```

## MODEL 3

## Model

```

{
for i in 1:N {
    x(i) ~ dbin(p(i), k(i))
    p(i) ~ dbeta(pi, 1)
    k(i) ~ dnegbin(alpha, beta)
}

alpha ~ dbeta(100, 100)
beta ~ dunif(5, 25)
pi ~ dbeta(100, 100)
}

```

## MODEL 4

## Model

```

{
for i in 1:N {
    x(i) ~ dbin(p(i), k(i))
    p(i) ~ dunif(pi, 1)
    k(i) ~ dnegbin(alpha, beta)
}

alpha ~ dunif(0.3, 1.0)
beta ~ dunif(7, 22)
pi ~ dbeta(3.0, 3.0)
}

```

## MODEL 5

Model

```

{
for i in 1:N {
    x(i) ~ dbin(p(i), k(i))
    p(i) ~ dunif(π, 1)
    k(i) ~ dnegbin(α, β)
}

α ~ dunif(0.0, 1.0)
β ~ dgamma(100, 100)
π ~ dbeta(100, 100)
}

```

## MODEL 6

Model

```

{
for i in 1:N {
    x(i) ~ dbin(p(i), k(i))
    p(i) ~ dunif(π, 1)
    k(i) ~ dnegbin(α, β)
}

α ~ dunif(0.0, 1.0)
β ~ dpois(20)
π ~ dbeta(2.0, 2.0)
}

```

## MODEL 7

Model

```

{
for i in 1:N {

    x(i) ~ dbin(p(i), k(i))
    p(i) ~ dunif(π, 1)
    k(i) ~ dnegbin(α, β)

}

α ~ dbeta(100.0, 100.0)
β ~ dpois(100)
π ~ dbeta(100, 100)

}

```

## MODEL 8

Model

```

{
for i in 1:N {

    x(i) ~ dbin(p(i), k(i))
    p(i) ~ dunif(π, 1)
    k(i) ~ dnegbin(α, β)

}

α ~ dunif(0.0, 1.0)
β ~ dgamma(10, 10)
π ~ dbeta(100, 100)

}

```



Case 3

There is this option of  $X$  having a beta-binomial distribution whereby the distribution of lions and hyenas in the TNP is generally binomially distributed but the probability of encountering an animal is unknown and hence takes on the beta distribution. The result of the mixture is beta-binomial and its model is expressed as follows

MODEL 9

Model

```

{
for i in 1 : N {
    x(i) ~ dbin(p(i), k(i))
    p(i) ~ dbeta(δ, θ)
    k(i) ~ dnegbin(α, β)
}

α ~ dbeta(100.0, 100.0)
β ~ dgamma(10, 10)
δ ~ dgamma(100, 100)
θ ~ dgamma(100, 100)
π = δ / (δ + θ)
}

```

All the models apart from model 9 helps to directly estimate the parameter  $\pi$  the response rate of the lion or hyena. However, in model 9 the estimate of  $\pi$  is indirectly estimated as a proportion using the parameters  $\theta$  and  $\delta$  respectively as

$$\pi = \frac{\delta}{\delta + \theta}.$$

In either case, the objective is determining  $E(X) = \mu \pi$  whereby  $\mu = E(Y)$  and  $\pi$ - the response rate. Therefore the estimate of  $\mu$  can be obtained for each habitat or be generalized for the whole park depending on whether or not the estimates for  $\pi$  can be taken to be constant for all habitats.



## Chapter 4

# DATA ANALYSIS

Using the data below, each of the models are tested to find the most appropriate one. The data was collected over the whole of Tsavo national park which is divided into East and West. The number of carnivores observed during data collection were as follows:

TABLE 4.1: Species Totals

Species	Number observed
Cheetah	6
Genet	3
Hyena	262
Jackal	7
Lion	45
Leopard	0
Grand Total	283

The distribution of the animals per each habitat was as follows

It appears that because the open grassland with sparse shrubs has

TABLE 4.2: Total counts per Habitat

Habitats	Area (KM2)	Sum of Lion	Sum of Hyena	Call_stn
CDF	89.87	0	10	2
CG	5913.15	12	65	24
CL	3349.79	2	34	19
DSST	1145.1	0	25	7
OG	1137.85	0	12	7
OGSS	8117.81	31	116	57
Grand Total	19753.57	45	262	116

the largest area, it explains why the largest population of observed lions and hyenas are found there. However, the closed deciduous

forest has the least area yet as many animals were observed there as they were observed in open grassland despite open grassland having over 12 times as much area. This means the size does not directly imply numbers even though the OGSS is the most expanse and populous. The mean per habitat is follows;

Each of the nine models were ran and the estimates for the param-

TABLE 4.3: Habitat Means and Density

Habitat	Sum of lion	Mean of lion	Sum of hyena	Mean of hyena	Total area
CDF	0	0	10	5	89.87
CG	12	0.5	65	2.7083	5913.15
CL(>50%)	2	0.1053	34	1.3158	3349.79
DSST	0	0	25	3.5714	1145.1
OG	0	0	12	1.7143	1137.85
OGSS	31	0.5439	116	2.0351	8117.81
Grand Total	45	0.3879	262	2.2586	19753.57

eters were compared based on least standard deviations and least MC errors to determine which estimate to be used. Considering the conditions for the best estimates, model 9 seems to have provided the best set of estimates which are very consistent throughout the changes in the value of N. The model also provides the smallest and most consistent std deviations and MC errors. This will therefore be used to estimate the population of lions and hyenas as

$$\hat{\mu} = \frac{\hat{x}}{\hat{\pi}}$$

Because of the inflation of zeros in the data caused by some habitats recording no observation in all the call stations, the overall estimate will be used for lions. However, for hyenas the individual habitat estimates will be focused on.

The table values in Appendix B indicate that models 3,7 and 9 have the best estimates of  $\pi$  across the habitats. They are also supported by very small values of standard deviations and the minimal MC errors. However, unlike for the other parameters  $\alpha$  and  $\beta$ , the parameter  $\pi$  does not show convergence for models 3 and 7. The graphs below show the history extract from WinBUGS

FIGURE 4.1: iteration history comparison

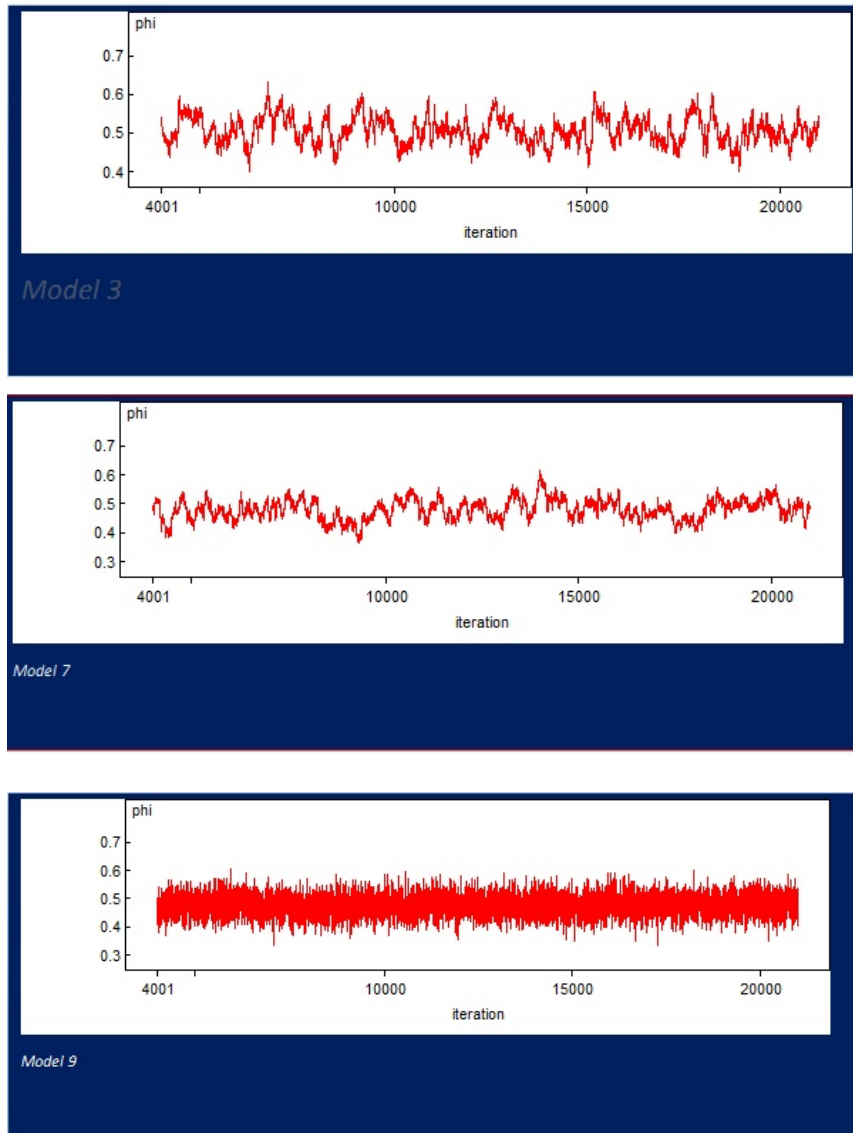


TABLE 4.4: Summary of Lion Data

Habitat	Mean of lion	$\hat{\mu}$ for lion	Density R=2.5 km	Total pop.
CDF	0.00	0.00	0.00	0.00
CG	0.50	1.01	0.05	304.50
CL( $\geq 50\%$ )	0.11	0.22	0.01	36.80
DSST	0.00	0.00	0.00	0.00
OG	0.00	0.00	0.00	0.00
OGSS	0.54	1.09	0.06	448.90
Grand Total	0.39	0.82	0.04	820.76

TABLE 4.5: Summary of Hyena Data

Habitat	Mean of hyena	$\hat{\mu}$ of hyena	Density of Hyena	Tot. pop
CDF	5.00	9.90	0.31	27.66
CG	2.71	5.05	0.16	928.42
CL(50%)	1.32	2.59	0.08	269.17
DSST	3.57	6.91	0.21	246.03
OG	1.71	3.40	0.11	120.22
OGSS	2.04	3.64	0.11	918.26
Grand Total	2.26	3.67	0.11	2252.28

In addition, since each of the models are compiled as single chains, using test statistics like Geweke, Raftery or Heidelberger for single chains and Gelman for multiple chains Raftery and Lewis (1992); Dong (2015), Smith (2007), Sahlin (2011), Cowles and Carlin (1996). It is proposed that a convergence diagnostic for Markov chains be such that the first and last part of the Markov chain have equal means (by default the first 10% and the last 50%). Drawing samples with the two means equal shows they are from the stationary distribution of the chain, and Geweke's statistic has an asymptotically standard normal distribution Cowles and Carlin (1996) and Sahlin (2011). The implication is that the value should be as close to zero as possible.

TABLE 4.6: Model 3

MODEL 3			
Fraction in 1st window = 0.1			
Fraction in 2nd window = 0.5			
alpha	beta	deviance	phi
-0.6482	0.1718	0.764	-1.3788

TABLE 4.7: Model 7

MODEL 7			
Fraction in 1st window = 0.1			
Fraction in 2nd window = 0.5			
alpha	beta	deviance	phi
-0.08045	1.68164	0.80875	-1.36732

TABLE 4.8: Model 9

MODEL 9			
Fraction in 1st window = 0.1			
Fraction in 2nd window = 0.5			
alpha	beta	deviance	phi
-1.36028	-0.27076	2.3717	-0.08052

Model 9 indicates the best convergence of the three model alternatives. The test statistic can be looked at as Two-sample ( $X_1$  and  $X_2$ ) T-test of mean (unequal variance)

$$T = \frac{(\bar{X}_1 + \bar{X}_2)}{\sqrt{\left(\frac{s_1^2}{n} + \frac{s_2^2}{m}\right)}}$$

The D.F for  $T$  can be treated roughly as  $\min(n - 1, m - 1)$ . When  $n, m \rightarrow \infty$ ,  $T$  can be approximated using the standard normal  $Z$ . However, there is need to adjust the sample variances since they are not independent.

Geweke's diagnostics estimate for the sample variances use spectral densities Cowles and Carlin (1996) and Sahlin (2011).

Similarly, the Raftery-Lewis test approves that model 9 outperforms all the others as shown below

TABLE 4.9: R-L Test Model 3

		Quantile (q) = 0.025		
		Accuracy (r) = +/- 0.005		
		Probability (s) = 0.95		
	Burn-in(M)	Total(N)	Lower bound (Nmin )	Dependence factor(I)
alpha	6	11379	3746	3.04
beta	6	14079	3746	3.76
deviance	6	11598	3746	3.1
phi	132	155892	3746	41.6

TABLE 4.10: R-L Test Model 7

		Quantile (q) = 0.025		
		Accuracy (r) = +/- 0.005		
		Probability (s) = 0.95		
	Burn-in(M)	Total(N)	Lower bound (Nmin )	Dependence factor(I)
alpha	6	4500	3746	1.2
beta	6	6405	3746	1.71
deviance	12	15429	3746	4.12
phi	93	101697	3746	27.1



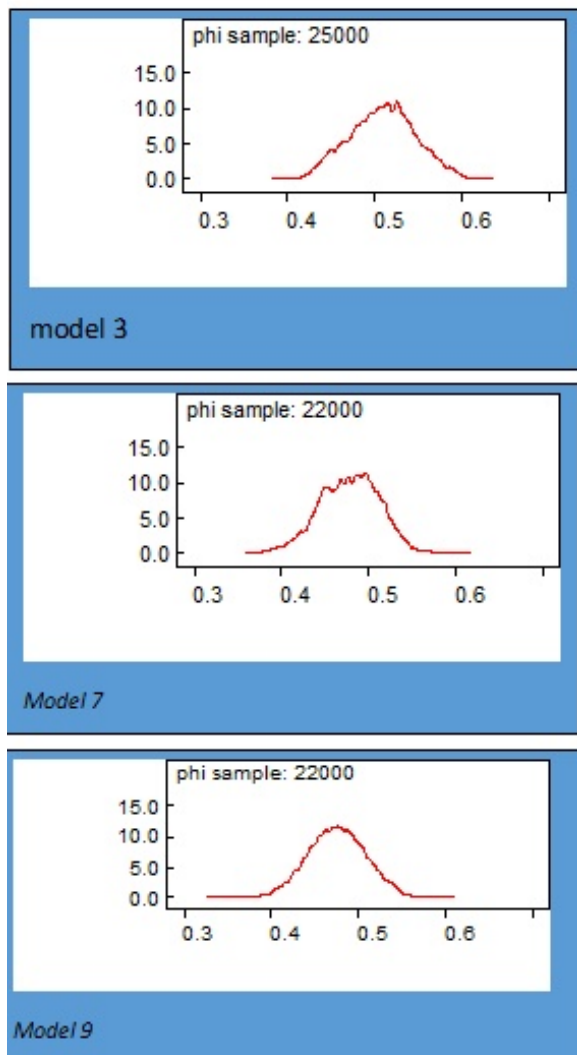
TABLE 4.11: R-L Test Model 9

		Quantile (q) = 0.025		
		Accuracy (r) = +/- 0.005		
		Probability (s) = 0.95		
	Burn-in(M)	Total(N)	Lower bound (Nmin )	Dependence factor(I)
alpha	6	7656	3746	2.04
beta	8	9526	3746	2.54
deviance	8	8452	3746	2.26
phi	5	6184	3746	1.65

The best model should have the Raftery-Lewis statistic as close to 1 as possible. This gives model 9 an edge over the others since its  $\pi$  estimate has R-L value of 1.65 as compared to 27.1 and 41.6 for the others. Indeed, the MCMC process has totally converged for the model 9. Even though this R-L statistic is not exactly 1, its close enough since it is not possible to exactly pin point convergence as correctly stated in Cowles and Carlin (1996) The approach is based on two-state Markov chain theory, with binomial variance of standard samples. Since N is based on binomial variance leads to the negative result that more iterations are required for estimating quantiles near the median than extreme quantiles to obtain the same degree of accuracy. This approach may be applied to the output of any MCMC algorithm as stated by Cowles and Carlin (1996) & Sahlin (2011)

In addition, the total iterations N needed for the convergence to take place helps to suggest how close the MCMC process is to convergence. When  $N \gg A$  where A= the actual total iterations ran, then the process has not reached convergence otherwise it most likely has converged. On that note therefore, model 9 above has converged while models 3 and 7 have not. It's worth noting that the kernel density does not give a clear distinction between the models as can be seen below. They all appear equally good because of the large number of iterations involved.

FIGURE 4.2: kernel density comparison



# Chapter 5

## Conclusions and Recommendations

### 5.1 Conclusions and Discussions

The estimate of the response probability of lions  $\pi = 0.4754 \pm 0.0341$  (95%CI) which is key in estimating population mean  $\mu$  both per habitat and generally was determined by using the algorithm in model 9 which happened to be the best alternative out of all the possible candidates especially for the sake of the other parameters of the model.

However, this estimate seems to be the same for both models 3 and 7 with the difference being that the estimate varies across habitats for models 3 and 7 and only settles at 0.4754 for the whole study areas whereas its virtually the same all-through for all habitats when it comes to model 9. This estimate is quite in agreement with the reference calibration experiment by Ferreira and Funston (2010) The estimate for the response of Hyenas is  $\pi = 0.616 \pm 0.029$  (95%CI) also falls within the reference calibration experiment by Mills, Juritz, and Zucchini (2001) and Ogutu and Dublin (1998) There is a significant difference between the response rate of lions and hyenas in agreement with calibration experiments of Ferreira and Funston (2010)

The table above shows clearly that the lions and even hyenas are

Habitat	Counts	Mean of Lion	Mu for Lion	Density R=2.5	Total Pop	Total Area
CDF	0	0	0	0	0	89.87
C G	12	0.5	1.012	0.0515	305	5913.15
CL(>50%)	2	0.1053	0.2157	0.01099	37	3349.79
DSST	0	0	0	0	0	1145.1
O G	0	0	0	0	0	1137.85
OGSS	31	0.5439	1.086	0.0553	449	8117.81
Grand Total	45	0.3879	$0.8159 \pm 4.4E - 3$	$0.04155 \pm 2.1E - 4$	$821 \pm 4$	19753.57

TABLE 5.1: lion analysis

distributed differently per habitat. The lions are found almost as many in the open grassland with sparse shrubs as they are in the closed grasslands. While the absence of observed numbers in the open grasslands, closed deciduous forests and deciduous shrubland with sparse trees gives an impression that lions are not found in those habitats, which may not be absolutely true. This informed the use of the overall density for the whole study zone of TNP instead of the individual habitat densities. This is because out of the total area of 2372.82sq. km of land, only 514.72 sq. km was under observation constituting a mere 22%. There is no sufficient evidence to dismiss the whole of this area as having no lions at all. In order to make use of this area therefore the whole area density of 4 lions in every 100sq. km of land was used. All this was on the assumption of a response radius of 2.5 km which was the minimum response radius in the work of Mills, Juritz, and Zucchini (2001) and later confirmed by Ferreira and Funston (2010)

Concerning Hyenas, their distribution also varies by habitat. Ignoring closed deciduous forests for the limited size and calling stations used, deciduous shrubland with sparse trees in the most populous at about 21 hyenas in every 100sq. km and cropland being the least populous with just about 8 hyenas in every 100sq. km. With or without including the closed deciduous forests, this brings the total population of Hyenas to about 2500 in the expanse TNP. For the case of hyenas there is evident various between using overall data for the whole study region of TNP and individual habitat estimates. This is due to the fact that hyenas are found in all habitats with varying density and each habitat has different size. The individual habitat estimates are as shown It is important to note that the

Habitat	pihat (95% CI)	Mean of Hyena	Muhat for hyena (95% CI)	Density R=3.2(95% CI)	Total Pop (95% CI)	Total Area
CDF	0.505±0.0351	5	9.901±0.048	0.3078±0.0015	28±0	89.87
CG	0.5362±0.0335	2.7083	5.051±0.020	0.1570±0.0006	928±4	5913.15
CL	0.5091±0.0347	1.3158	2.585±0.012	0.0804±0.0004	269±1	3349.79
DSST	0.5167±0.0347	3.5714	6.912±0.031	0.2149±0.0009	246±1	1145.1
OG	0.5043±0.0348	1.7143	3.399±0.017	0.1057±0.0004	120±1	1137.85
OGSS	0.5592±0.0322	2.0351	3.639±0.012	0.1131±0.0004	918±3	8117.81
Total				0.1270±0.0005	2509±11	19753.57

TABLE 5.2: Hyena analysis

whole idea of using MCMC sampling to estimate the parameters  $\pi, \alpha$  and  $\beta$  arise due to the fact that it was not possible to analytically determine them. The approaches used by Mills Mills, Juritz, and

Zucchini (2001) ended up being mathematically intractable just like the one used in this paper save for the assumptions by Mills that  $\pi \approx 1$  which brought a virtual tractability owing to the fact that the need for employing mixture distributions arose from the fact that the response probability was not a constant across all the habitats. Similarly, the alternative approach used in this paper ends up being intractable save for the assumption that  $\beta \rightarrow 0$  that brings forth a kind of tractability even though most of the original assumptions are violated by this new assumption. The actual states of equations [6] and [11] cannot be simplified, implying that the equations [7] and [13] cannot be justified in a strict sense. At the same time, it could be noted that equation [6] and even [11] has a binomial like structure. This prompted the ideas that it was likely that the distribution of the animals in the study area had a kind of binomial structure with the parameters  $p$  and  $n$  unknown. Evidently, the only feasible way to try estimate those parameters was to scramble Bayesian methods and simulate samples that could efficiently help estimate them. This justified the employment of MCMC sampling as implemented in WinBUGS. However, to enable further analysis of the samples the MCMC generates, the use of R was imperative hence the whole MC was implemented in R-GUI using the package R2WinBUGS. In addition, under R-GUI, it's much feasible to determine if estimates from the models of choice has converge using the Geweke, Raftery-Lewis and Heidelberger statistics for all single chain simulations and Gelman statistic for multiple chain simulations. With all diagnostics confirming convergence of the estimate of  $\pi$ , which indeed agrees with calibration experiments, it is found to be a prudent approach and a robust one since it is capable of producing very good estimates without the need to carry out the experiments which are an extra financial burden to the research process.

## 5.2 Recommendations

The MC algorithm used in this project is robust enough to be applied to any similar situations where the density of rare species like the lion and other carnivores is required. It is worth noting that the prevailing conditions are generally the same in all protected areas like Parks and game reserves. This means that the application of this approach will give rise to desired results. This research also recommends looking into the spatial distribution of lions and hyenas in the TNP if anything to complement the claimed distribution per habitat. The same should be extended to study the trends in population declined as asserted in most lion studies like most recently Goldman, Roque De Pinho, and Perry (2010), Nelson, Lindsey, and Balme (2013), Riggio et al. (2013), and Schuette, Creel, and Christianson (2013)

On mixture distributions, this project has used the exponential distribution as the mixing distribution for the Poisson and Mills employed the gamma distribution. It may be important to try out the Log-Normal as a third alternative in an attempt to find the distribution for the count data that is mathematically tractable. Concerning models for data analysis, it was noted that the lion data was full of zeros. There were habitats where there were totally no observed lions like closed deciduous forests among others. This paper proposes further work in the same field to find alternative approaches of handling data with unusually many zeros especially using Zero-inflated Poisson or Zero-inflated Negative Binomial. A good alternative which this research intends to look into in the future is small area estimation as works of Rao and Molina (2015), Chambers and Tzavidis (2006), and Jiang, Lahiri, and Nguyen (2016) & Jiang and Lahiri (2006).

# Bibliography

- Abas, Ahmed R (2013). "On determining efficient finite mixture models with compact and essential components for clustering data". In: *Egyptian Informatics Journal* 14.1, pp. 79–88.
- Alonso, Robert S. et al. (2015). "Mark-Recapture and Mark-Resight Methods for Estimating Abundance with Remote Cameras: A Carnivore Case Study". In: *PLoS ONE* 10.3, e0123032.
- Ariefiandy, Achmad et al. (2013). "Monitoring the ungulate prey of the Komodo dragon *Varanus komodoensis*: distance sampling or faecal counts?" In: *Wildlife Biology* 19.2, pp. 126–137.
- Bauer, Hans (2007). "Status of large carnivores in Bouba Ndjida National Park, Cameroon". In: *African Journal of Ecology* 45.3, p. 448.
- Borg, Bridget L. et al. (2016). "Implications of Harvest on the Boundaries of Protected Areas for Large Carnivore Viewing Opportunities". In: *PLoS ONE* 11.4.
- Buckland, Steeves et al. (2014). "Ecological effects of the invasive giant madagascar day gecko on endemic mauritian geckos: Applications of binomial-mixture and species distribution models". In: *PloS one* 9.4, e88798.
- Chambers, Ray and Nikos Tzavidis (2006). "M-quantile models for small area estimation". In: *Biometrika* 93.2, pp. 255–268.
- Cowles, Mary Kathryn and Bradley P Carlin (1996). "Markov chain Monte Carlo convergence diagnostics: a comparative review". In: *Journal of the American Statistical Association* 91.434, pp. 883–904. ISSN: 0162-1459.
- Cozzi, G. et al. (2013). "Density and habitat use of lions and spotted hyenas in northern Botswana and the influence of survey and ecological variables on call-in survey estimation". In: *Biodiversity and Conservation* 22.12, pp. 2937–2956. ISSN: 1572-9710. DOI: [10.1007/s10531-013-0564-7](https://doi.org/10.1007/s10531-013-0564-7). URL: <http://dx.doi.org/10.1007/s10531-013-0564-7>.

- Dolrenry, Stephanie et al. (2014). "A metapopulation approach to African lion (*Panthera leo*) conservation". In: *PloS one* 9.2, e88081.
- Dong, Yan (2015). "Nonparametric Bayes approach for a semi-mechanistic pharmacokinetic and pharmacodynamic model". In:
- Dorazio, Robert M. and Edward F. Connor (2014). "Estimating Abundances of Interacting Species Using Morphological Traits, Foraging Guilds, and Habitat". In: *PLoS ONE* 9.4, e94323. DOI: [10.1371/journal.pone.0094323](https://doi.org/10.1371/journal.pone.0094323). URL: <http://dx.doi.org/10.1371%2Fjournal.pone.0094323>.
- Ferreira, Sam M and Paul J Funston (2010). "Estimating lion population variables: prey and disease effects in Kruger National Park, South Africa". In: *Wildlife Research* 37.3, pp. 194–206. ISSN: 1448-5494.
- Fryback, Dennis G, Natasha K Stout, and Marjorie A Rosenberg (2001). "An elementary introduction to Bayesian computing using WinBUGS". In: *International Journal of Technology Assessment in Health Care* 17.01, pp. 98–113.
- Goldman, Mara J, Joana Roque De Pinho, and Jennifer Perry (2010). "Maintaining complex relations with large cats: Maasai and lions in Kenya and Tanzania". In: *Human Dimensions of Wildlife* 15.5, pp. 332–346.
- Jasra, Ajay, CC Holmes, and DA Stephens (2005). "Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling". In: *Statistical Science*, pp. 50–67. ISSN: 0883-4237.
- Jiang, Jiming and P Lahiri (2006). "Mixed model prediction and small area estimation". In: *Test* 15.1, pp. 1–96.
- Jiang, Jiming, P Lahiri, and Thuan Nguyen (2016). "A Unified Monte-Carlo Jackknife for Small Area Estimation after Model Selection". In: *arXiv preprint arXiv:1602.05238*.
- Khorozyan, Igor et al. (2015). "Big cats kill more livestock when wild prey reaches a minimum threshold". In: *Biological Conservation* 192, pp. 268–275.
- Kiffner, Christian et al. (2008). "Response of lions (*Panthera leo* LINNAEUS 1758) and spotted hyaenas (*Crocuta crocuta* ERXLEBEN 1777) to sound playbacks". In: *African Journal of Ecology* 46.2, pp. 223–226. ISSN: 1365-2028. DOI: [10.1111/j.1365-2028.2007.](https://doi.org/10.1111/j.1365-2028.2007.)



00813.x. URL: <http://dx.doi.org/10.1111/j.1365-2028.2007.00813.x>.

Kupzyk, Kevin A and CYFS SRM Unit (2011). *Introduction to mixture modeling*. Generic.

Lamprey, Richard H. and Robin S. Reid (2004). "Expansion of human settlement in Kenya's Maasai Mara: what future for pastoralism and wildlife?" In: *Journal of Biogeography* 31.6, pp. 997–1032. ISSN: 1365-2699. DOI: [10.1111/j.1365-2699.2004.01062.x](https://doi.org/10.1111/j.1365-2699.2004.01062.x). URL: <http://dx.doi.org/10.1111/j.1365-2699.2004.01062.x>.

Mills, M. G. L., J. M. Juritz, and W. Zucchini (2001). "Estimating the size of spotted hyaena (*Crocuta crocuta*) populations through playback recordings allowing for non-response". In: *Animal Conservation* 4.4, pp. 335–343. ISSN: 1469-1795. DOI: [10.1017/S1367943001001391](https://doi.org/10.1017/S1367943001001391). URL: <http://dx.doi.org/10.1017/S1367943001001391>.

Nelson, Fred, Peter Lindsey, and Guy Balme (2013). "Trophy hunting and lion conservation: a question of governance?" In: *Oryx* 47.04, pp. 501–509.

Ogutu, JO and HT Dublin (1998). "The response of lions and spotted hyaenas to sound playbacks as a technique for estimating population size". In: *African Journal of Ecology* 36.1, pp. 83–95. ISSN: 1365-2028.

Ogutu, Joseph O, Nina Bhola, and Robin Reid (2005). "The effects of pastoralism and protection on the density and distribution of carnivores and their prey in the Mara ecosystem of Kenya". In: *Journal of Zoology* 265.3, pp. 281–293.

Osano, Philip M et al. (2013). *Why keep lions instead of livestock? Assessing wildlife tourism-based payment for ecosystem services involving herders in the Maasai Mara, Kenya*. Wiley Online Library.

Raftery, Adrian E and Steven M Lewis (1992). "[Practical Markov Chain Monte Carlo]: comment: one long run with diagnostics: implementation strategies for Markov Chain Monte Carlo". In: *Statistical science* 7.4, pp. 493–497.

Rao, John NK and Isabel Molina (2015). *Small area estimation*. John Wiley & Sons.

- Revord Larkin, Lexie (2014). "Human–Wildlife Conflict in Maasai Mara National Reserve: Demonstrating the Need for an International Framework on Wildlife-Based Tourism". In: *Journal of International Wildlife Law & Policy* 17.1-2, pp. 1–31.
- Riggio, Jason et al. (2013). "The size of savannah Africa: a lion's (Panthera leo) view". In: *Biodiversity and Conservation* 22.1, pp. 17–35.
- Sahlin, Kristoffer (2011). "Estimating convergence of Markov chain Monte Carlo simulations". In: *Stockholm University, Master Thesis*.
- Sawatzky, Rick and Chris Richardson (2009). "Using latent variable mixture modelling to explore sample diversity". In:
- Schuette, Paul, Scott Creel, and David Christianson (2013). "Coexistence of African lions, livestock, and people in a landscape with variable human land use and seasonal movements". In: *Biological Conservation* 157, pp. 148–154.
- Smith, Brian J (2007). "boa: an R package for MCMC output convergence assessment and posterior inference". In: *Journal of Statistical Software* 21.11, pp. 1–37.
- Sturtz, Sibylle, Uwe Ligges, Andrew Gelman, et al. (2005). "R2WinBUGS: a package for running WinBUGS from R". In: *Journal of Statistical software* 12.3, pp. 1–16.
- Suksaengrakcharoen, Satsayamon and Winai Bodhisuwan (2014). "A NEW FAMILY OF GENERALIZED GAMMA DISTRIBUTION AND ITS APPLICATION". In: *Journal of Mathematics and Statistics* 10.2, pp. 211–220. ISSN: 1549-3644.
- Theobald, Chris, Ayona Chatterjee, and Graham Horgan (2012). "A hierarchical Bayesian mixture model for repeated dietary records". In: *Food and chemical toxicology* 50.2, pp. 320–327.
- Thorn, Michelle et al. (2013). "Characteristics and determinants of human-carnivore conflict in South African farmland". In: *Biodiversity and conservation* 22.8, pp. 1715–1730.
- Unwin, Antony (2012). "Biostatistics: A Computing Approach by Stewart Anderson". In: *International Statistical Review* 80.2, pp. 335–336.
- Winterbach, Christiaan W, Carolyn Whitesell, and Michael J Somers (2015). "Wildlife abundance and diversity as indicators of tourism potential in northern Botswana". In: *PloS one* 10.8, e0135595.
- Yirga, Gidey et al. (2014). "Status of Lion (Panthera leo) and Spotted Hyena (Crocuta crocuta) in Nechisar National Park, Ethiopia". In: *Momona Ethiopian Journal of Science* 6.2, pp. 127–137.

- Young-Overton, Kim D., Paul J. Funston, and Sam M. Ferreira (2014).  
“Rainfall driven changes in behavioural responses confound measuring trends in lion population size”. In: *Wildlife Biology* 20.6, pp. 344–355. ISSN: 0909-6396. DOI: [10 . 2981 / wlb . 00015](https://doi.org/10.2981/wlb.00015). URL: [http : // dx . doi . org / 10 . 2981 / wlb . 00015](http://dx.doi.org/10.2981/wlb.00015).



# Chapter 6

## Appendices

### 6.1 Appendix A

Model 1

N=2								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.7142	0.2133	0.002285	0.2254	0.7571	0.9902	4006	16995
beta	0.9476	0.3078	0.002506	0.4467	0.9161	1.635	4006	16995
phi	0.4951	0.1102	0.001282	0.2829	0.4953	0.7118	4006	16995
N=7								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.8457	0.1364	0.001857	0.4953	0.8842	0.9958	4001	17000
beta	0.9198	0.2995	0.002521	0.4278	0.8883	1.587	4001	17000
phi	0.4963	0.1083	0.003102	0.2824	0.4971	0.7045	4001	17000
N=19								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.8016	0.1115	0.00171	0.5345	0.8218	0.96	4001	17000
beta	0.895	0.301	0.003006	0.4107	0.8578	1.571	4001	17000
phi	0.4844	0.1166	0.007461	0.2675	0.4829	0.7084	4001	17000
N=24								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.494	0.1234	0.002586	0.2498	0.4966	0.7257	4001	17000
beta	0.7607	0.2589	0.003617	0.3517	0.7286	1.35	4001	17000
phi	0.46	0.1038	0.006438	0.262	0.4583	0.6606	4001	17000
N=57								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5224	0.098	0.002715	0.322	0.5265	0.7015	4001	17000
beta	0.8786	0.2603	0.004566	0.4617	0.8454	1.468	4001	17000
phi	0.4833	0.1266	0.01064	0.2152	0.4923	0.7403	4001	17000
N=116								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.4731	0.09131	0.002154	0.2971	0.4722	0.6524	4001	17000
beta	0.69	0.1969	0.00418	0.3792	0.6645	1.146	4001	17000
phi	0.2709	0.0787	0.006559	0.1393	0.2606	0.4285	4001	17000

TABLE 6.1: Model 1

## Model 2

N=2								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.719	0.2073	0.002546	0.2487	0.759	0.9897	4001	17000
beta	0.9943	0.09968	7.67E-04	0.8094	0.9907	1.202	4001	17000
phi	0.4714	0.2921	0.009555	0.02046	0.4566	0.9708	4001	17000
N=7								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.8581	0.1234	0.002205	0.5408	0.8917	0.996	4001	17000
beta	0.9928	0.09948	7.77E-04	0.8065	0.989	1.2	4001	17000
phi	0.399	0.2564	0.01485	0.01856	0.3693	0.9014	4001	17000
N=19								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.8465	0.08444	0.002503	0.6411	0.861	0.9666	4001	17000
beta	0.9873	0.09778	8.26E-04	0.8051	0.9835	1.191	4001	17000
phi	0.7157	0.3097	0.02652	0.03631	0.8615	0.9897	4001	17000
N=24								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5495	0.09801	0.003393	0.3553	0.5508	0.7358	4001	17000
beta	0.973	0.09978	9.13E-04	0.7895	0.9691	1.179	4001	17000
phi	0.2658	0.2194	0.0173	0.006239	0.2087	0.7646	4001	17000
N=57								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5358	0.07261	0.003421	0.3915	0.5375	0.6724	4001	17000
beta	0.9854	0.09742	0.001056	0.8005	0.9826	1.182	4001	17000
phi	0.2977	0.2171	0.01854	0.009612	0.2493	0.7057	4001	17000
N=116								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5621	0.05669	0.00196	0.4486	0.5632	0.6699	4001	17000
beta	0.9493	0.09664	0.001199	0.7699	0.9462	1.149	4001	17000
phi	0.09851	0.08627	0.007118	0.002501	0.07251	0.3087	4001	17000

TABLE 6.2: Model 2

Model 3

N=2								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5272	0.03447	2.53E-04	0.4598	0.5272	0.5942	4053	16948
beta	5.821	0.8221	0.01061	5.021	5.57	8.029	4053	16948
phi	0.4999	0.03509	3.27E-04	0.4308	0.4999	0.5681	4053	16948
N=7								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5755	0.03214	2.56E-04	0.5114	0.576	0.6381	4001	17000
beta	5.276	0.2833	0.004146	5.007	5.187	6.033	4001	17000
phi	0.4986	0.03512	4.84E-04	0.43	0.4986	0.5681	4001	17000
N=19								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.6549	0.02751	2.51E-04	0.5996	0.6552	0.7078	4001	17000
beta	5.136	0.1367	0.001896	5.003	5.093	5.511	4001	17000
phi	0.4968	0.03505	7.52E-04	0.4289	0.4966	0.5655	4001	17000
N=24								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.6597	0.02601	2.33E-04	0.6074	0.6599	0.7101	4001	17000
beta	5.133	0.132	0.001779	5.004	5.091	5.488	4001	17000
phi	0.5012	0.03462	9.49E-04	0.4341	0.5009	0.5693	4001	17000
N=57								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.741	0.01939	1.70E-04	0.7023	0.7412	0.7781	4001	17000
beta	5.095	0.09618	0.001302	5.002	5.065	5.356	4001	17000
phi	0.5117	0.03469	0.001632	0.4443	0.512	0.5798	4001	17000
N=116								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.8192	0.0138	1.11E-04	0.7913	0.8195	0.8452	4005	16996
beta	5.071	0.07227	8.89E-04	5.002	5.049	5.265	4005	16996
phi	0.5059	0.03539	0.002627	0.4422	0.5044	0.579	4005	16996

TABLE 6.3: Model 3

## Model 4

N=2								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.9603	0.04119	3.89E-04	0.8493	0.9737	0.999	4002	16999
beta	13.15	4.301	0.03917	7.207	12.52	21.36	4002	16999
phi	0.4987	0.1923	0.003249	0.1402	0.4992	0.853	4002	16999
N=7								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.9882	0.0129	1.08E-04	0.9526	0.9922	0.9997	4001	17000
beta	13.1	4.272	0.03931	7.208	12.44	21.43	4001	17000
phi	0.5078	0.1926	0.009071	0.1517	0.5059	0.8681	4001	17000
N=19								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.9865	0.009196	9.23E-05	0.9622	0.9887	0.9977	4001	17000
beta	13.03	4.275	0.04332	7.201	12.29	21.34	4001	17000
phi	0.5434	0.2056	0.0161	0.1751	0.5278	0.9524	4001	17000
N=24								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.9522	0.01967	3.01E-04	0.9072	0.9551	0.9814	4001	17000
beta	12.14	4.141	0.06412	7.123	11.05	21.11	4001	17000
phi	0.6408	0.1474	0.01163	0.317	0.651	0.8841	4001	17000
N=57								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.9499	0.0173	3.92E-04	0.9142	0.9514	0.977	4001	17000
beta	11.92	4.099	0.0954	7.121	10.75	20.98	4001	17000
phi	0.8337	0.06387	0.005441	0.672	0.8474	0.9191	4001	17000
N=116								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.9598	0.01183	2.80E-04	0.9374	0.9594	0.9813	4001	17000
beta	10.43	3.501	0.08826	7.062	9.15	19.95	4001	17000
phi	0.8469	0.05586	0.00486	0.7126	0.8585	0.9257	4001	17000

TABLE 6.4: Model 4



Model 5

N=2								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.7231	0.205	0.002057	0.2561	0.7639	0.9897	4001	17000
beta	0.9944	0.0994	8.84E-04	0.8098	0.9908	1.197	4001	17000
phi	0.4968	0.1094	0.001253	0.2865	0.4961	0.71	4001	17000
N=7								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.8671	0.1147	0.001314	0.5734	0.8985	0.9957	4001	17000
beta	0.9927	0.1009	8.22E-04	0.8062	0.9891	1.199	4001	17000
phi	0.4934	0.106	0.002833	0.2885	0.4936	0.6976	4001	17000
N=19								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.8332	0.08599	0.001308	0.6346	0.846	0.9626	4004	16997
beta	0.9882	0.09944	8.48E-04	0.8041	0.9851	1.192	4004	16997
phi	0.4857	0.1122	0.006505	0.2642	0.4863	0.6949	4004	16997
N=24								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5849	0.08883	0.001738	0.4057	0.5872	0.7523	4002	16999
beta	0.9698	0.09685	8.42E-04	0.7901	0.9675	1.165	4002	16999
phi	0.4812	0.1148	0.00761	0.2608	0.4778	0.7052	4002	16999
N=57								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5707	0.06225	0.001402	0.4456	0.5716	0.6889	4001	17000
beta	0.9789	0.09743	8.90E-04	0.7993	0.9756	1.178	4001	17000
phi	0.5032	0.09154	0.007349	0.3172	0.5068	0.6713	4001	17000
N=116								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.6183	0.05157	0.001432	0.5133	0.6201	0.7142	4001	17000
beta	0.9382	0.09613	0.001074	0.7602	0.9345	1.136	4001	17000
phi	0.3801	0.09345	0.007939	0.2215	0.3774	0.6273	4001	17000

TABLE 6.5: Model 5

## Model 6

N=2								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.9732	0.02758	2.08E-04	0.8976	0.9817	0.9994	4001	17000
beta	18.93	4.462	0.03157	10.69	18.73	28.24	4001	17000
phi	0.4891	0.2264	0.004346	0.0878	0.4863	0.8993	4001	17000
N=7								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.992	0.008485	6.96E-05	0.9699	0.9947	0.9998	4001	17000
beta	18.96	4.48	0.03521	10.67	18.81	28.16	4001	17000
phi	0.5052	0.2243	0.01194	0.09759	0.5095	0.9089	4001	17000
N=19								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.9912	0.005836	5.01E-05	0.9764	0.9924	0.9984	4001	17000
beta	18.92	4.436	0.03963	10.65	18.77	28.06	4001	17000
phi	0.577	0.2332	0.01863	0.1343	0.5809	0.9425	4001	17000
N=24								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.9687	0.01289	1.81E-04	0.9367	0.9711	0.9861	4001	17000
beta	18.22	4.64	0.0575	9.644	18.07	27.74	4001	17000
phi	0.6525	0.1984	0.01632	0.2211	0.6893	0.9383	4001	17000
N=57								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.9681	0.01071	1.90E-04	0.9413	0.9701	0.9828	4001	17000
beta	18.25	4.528	0.07655	9.778	18.08	27.53	4001	17000
phi	0.9181	0.03061	0.002603	0.8488	0.9227	0.9661	4001	17000
N=116								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.8913	0.189	0.01555	0.3513	0.9747	0.9865	4001	17000
beta	14.08	7.582	0.5513	0.3202	15.53	26.27	4001	17000
phi	0.7262	0.2115	0.01849	0.1188	0.7916	0.932	4001	17000

TABLE 6.6: Model 6

Model 7

N=2								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.6469	0.03425	3.45E-04	0.5779	0.648	0.7109	4001	17000
beta	42.21	7.781	0.08393	27.8	41.99	58.28	4001	17000
phi	0.5	0.03529	3.58E-04	0.4314	0.5	0.5699	4001	17000
N=7								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5361	0.04057	4.93E-04	0.4582	0.5356	0.6166	4018	16983
beta	2.443	1.629	0.02784	0.2152	2.102	6.429	4018	16983
phi	0.497	0.03524	4.65E-04	0.4263	0.4975	0.5658	4018	16983
N=19								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5191	0.03667	3.12E-04	0.4465	0.5193	0.5902	4002	16999
beta	0.7166	0.3702	0.005453	0.09232	0.6779	1.581	4002	16999
phi	0.4942	0.03513	8.44E-04	0.4253	0.4941	0.5638	4002	16999
N=24								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5139	0.03643	3.66E-04	0.4433	0.5136	0.586	4002	16999
beta	0.9879	0.3789	0.00505	0.4792	0.9256	1.894	4002	16999
phi	0.4977	0.0356	9.66E-04	0.4283	0.4974	0.5685	4002	16999
N=57								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5162	0.03594	4.02E-04	0.4464	0.5159	0.5873	4001	17000
beta	0.9433	0.2403	0.003485	0.5544	0.9176	1.495	4001	17000
phi	0.4919	0.03514	0.001999	0.4194	0.4931	0.5576	4001	17000
N=116								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5056	0.03345	4.14E-04	0.4412	0.5056	0.5709	4001	17000
beta	0.6388	0.1217	0.001931	0.4308	0.6193	0.9247	4001	17000
phi	0.4789	0.03533	0.002681	0.4105	0.4796	0.5435	4001	17000

TABLE 6.7: Model 7

## Model 8

N=2								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.7231	0.205	0.002057	0.2561	0.7639	0.9897	4001	17000
beta	0.9944	0.0994	8.84E-04	0.8098	0.9908	1.197	4001	17000
phi	0.4968	0.1094	0.001253	0.2865	0.4961	0.71	4001	17000
N=7								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.8671	0.1147	0.001314	0.5734	0.8985	0.9957	4001	17000
beta	0.9927	0.1009	8.22E-04	0.8062	0.9891	1.199	4001	17000
phi	0.4934	0.106	0.002833	0.2885	0.4936	0.6976	4001	17000
N=19								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.8332	0.08599	0.001308	0.6346	0.846	0.9626	4004	16997
beta	0.9882	0.09944	8.48E-04	0.8041	0.9851	1.192	4004	16997
phi	0.4857	0.1122	0.006505	0.2642	0.4863	0.6949	4004	16997
N=24								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5849	0.08883	0.001738	0.4057	0.5872	0.7523	4002	16999
beta	0.9698	0.09685	8.42E-04	0.7901	0.9675	1.165	4002	16999
phi	0.4812	0.1148	0.00761	0.2608	0.4778	0.7052	4002	16999
N=57								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5707	0.06225	0.001402	0.4456	0.5716	0.6889	4001	17000
beta	0.9789	0.09743	8.90E-04	0.7993	0.9756	1.178	4001	17000
phi	0.5032	0.09154	0.007349	0.3172	0.5068	0.6713	4001	17000
N=116								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.6183	0.05157	0.001432	0.5133	0.6201	0.7142	4001	17000
beta	0.9382	0.09613	0.001074	0.7602	0.9345	1.136	4001	17000
phi	0.3801	0.09345	0.007939	0.2215	0.3774	0.6273	4001	17000

TABLE 6.8: Model 8

## Model 9

N=2CDF								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5034	0.03474	3.18E-04	0.4352	0.5033	0.5717	4001	17000
beta	0.9928	0.09884	8.72E-04	0.8063	0.9899	1.197	4001	17000
phi	0.4984	0.03534	3.21E-04	0.4297	0.4985	0.5676	4001	17000
N=7DSST/OG								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.509	0.03494	2.71E-04	0.4408	0.5088	0.577	4001	17000
beta	0.9777	0.09832	7.64E-04	0.7922	0.9748	1.179	4001	17000
phi	0.494	0.03545	3.09E-04	0.4251	0.4938	0.5635	4001	17000
N=19 CL								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.5167	0.03429	3.96E-04	0.449	0.517	0.5835	4001	17000
beta	0.9577	0.09569	8.65E-04	0.7788	0.9549	1.153	4001	17000
phi	0.4881	0.03508	3.09E-04	0.4197	0.488	0.5564	4001	17000
N=24 CG								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.4989	0.03327	3.84E-04	0.4349	0.4985	0.5649	4001	17000
beta	0.9828	0.09746	9.19E-04	0.802	0.9792	1.181	4001	17000
phi	0.4941	0.03479	3.60E-04	0.4261	0.4942	0.5612	4001	17000
N=57 OGSS								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.4952	0.03127	4.54E-04	0.433	0.4952	0.5561	4001	17000
beta	1.002	0.09533	9.97E-04	0.8236	1.001	1.196	4001	17000
phi	0.5008	0.03418	4.77E-04	0.4332	0.5011	0.5668	4001	17000
N=116								
node	mean	sd	MC error	2.50%	median	97.50%	start	sample
alpha	0.509	0.0306	4.76E-04	0.4483	0.5091	0.5682	4001	17000
beta	0.9453	0.08865	0.001006	0.7784	0.9419	1.127	4001	17000
phi	0.4754	0.03413	6.21E-04	0.4085	0.4754	0.542	4001	17000

TABLE 6.9: Model 9

## **6.2 Appendix B**

model comparison									
	N	mean	sd	MC error	2.50%	median	97.50%	start	sample
model 1	2	0.4951	0.1102	0.001282	0.2829	0.4953	0.7118	4006	16995
	7	0.4963	0.1083	0.003102	0.2824	0.4971	0.7045	4001	17000
	19	0.4844	0.1166	0.007461	0.2675	0.4829	0.7084	4001	17000
	24	0.46	0.1038	0.006438	0.262	0.4583	0.6606	4001	17000
	57	0.4833	0.1266	0.01064	0.2152	0.4923	0.7403	4001	17000
	116	0.2709	0.0787	0.006559	0.1393	0.2606	0.4285	4001	17000
model 2	N	mean	sd	MC error	2.50%	median	97.50%	start	sample
	2	0.4714	0.2921	0.009555	0.02046	0.4566	0.9708	4001	17000
	7	0.399	0.2564	0.01485	0.01856	0.3693	0.9014	4001	17000
	19	0.7157	0.3097	0.02652	0.03631	0.8615	0.9897	4001	17000
	24	0.2658	0.2194	0.0173	0.006239	0.2087	0.7646	4001	17000
	57	0.2977	0.2171	0.01854	0.009612	0.2493	0.7057	4001	17000
116	0.09851	0.08627	0.007118	0.002501	0.07251	0.3087	4001	17000	
model 3	N	mean	sd	MC error	2.50%	median	97.50%	start	sample
	2	0.4999	0.03509	3.27E-04	0.4308	0.4999	0.5681	4053	16948
	7	0.4986	0.03512	4.84E-04	0.43	0.4986	0.5681	4001	17000
	19	0.4968	0.03505	7.52E-04	0.4289	0.4966	0.5655	4001	17000
	24	0.5012	0.03462	9.49E-04	0.4341	0.5009	0.5693	4001	17000
	57	0.5117	0.03469	0.001632	0.4443	0.512	0.5798	4001	17000
116	0.5059	0.03539	0.002627	0.4422	0.5044	0.579	4005	16996	
model 4	N	mean	sd	MC error	2.50%	median	97.50%	start	sample
	2	0.4987	0.1923	0.003249	0.1402	0.4992	0.853	4002	16999
	7	0.5078	0.1926	0.009071	0.1517	0.5059	0.8681	4001	17000
	19	0.5434	0.2056	0.0161	0.1751	0.5278	0.9524	4001	17000
	24	0.6408	0.1474	0.01163	0.317	0.651	0.8841	4001	17000
	57	0.8337	0.06387	0.005441	0.672	0.8474	0.9191	4001	17000
116	0.8469	0.05586	0.00486	0.7126	0.8585	0.9257	4001	17000	
model 5	N	mean	sd	MC error	2.50%	median	97.50%	start	sample
	2	0.4968	0.1094	0.001253	0.2865	0.4961	0.71	4001	17000
	7	0.4934	0.106	0.002833	0.2885	0.4936	0.6976	4001	17000
	19	0.4857	0.1122	0.006505	0.2642	0.4863	0.6949	4004	16997
	24	0.4812	0.1148	0.00761	0.2608	0.4778	0.7052	4002	16999
	57	0.5032	0.09154	0.007349	0.3172	0.5068	0.6713	4001	17000
116	0.3801	0.09345	0.007939	0.2215	0.3774	0.6273	4001	17000	
model 6	N	mean	sd	MC error	2.50%	median	97.50%	start	sample
	2	0.4891	0.2264	0.004346	0.0878	0.4863	0.8993	4001	17000
	7	0.5052	0.2243	0.01194	0.09759	0.5095	0.9089	4001	17000
	19	0.577	0.2332	0.01863	0.1343	0.5809	0.9425	4001	17000
	24	0.6525	0.1984	0.01632	0.2211	0.6893	0.9383	4001	17000
	57	0.9181	0.03061	0.002603	0.8488	0.9227	0.9661	4001	17000
116	0.7262	0.2115	0.01849	0.1188	0.7916	0.932	4001	17000	
model 7	N	mean	sd	MC error	2.50%	median	97.50%	start	sample
	2	0.5	0.03529	3.58E-04	0.4314	0.5	0.5699	4001	17000
	7	0.497	0.03524	4.65E-04	0.4263	0.4975	0.5658	4018	16983
	19	0.4942	0.03513	8.44E-04	0.4253	0.4941	0.5638	4002	16999
	24	0.4977	0.0356	9.66E-04	0.4283	0.4974	0.5685	4002	16999
	57	0.4919	0.03514	0.001999	0.4194	0.4931	0.5576	4001	17000
116	0.4789	0.03533	0.002681	0.4105	0.4796	0.5435	4001	17000	
model 8	N	mean	sd	MC error	2.50%	median	97.50%	start	sample
	2	0.4968	0.1094	0.001253	0.2865	0.4961	0.71	4001	17000
	7	0.4934	0.106	0.002833	0.2885	0.4936	0.6976	4001	17000
	19	0.4857	0.1122	0.006505	0.2642	0.4863	0.6949	4004	16997
	24	0.4812	0.1148	0.00761	0.2608	0.4778	0.7052	4002	16999
	57	0.5032	0.09154	0.007349	0.3172	0.5068	0.6713	4001	17000
116	0.3801	0.09345	0.007939	0.2215	0.3774	0.6273	4001	17000	
model 9	N	mean	sd	MC error	2.50%	median	97.50%	start	sample
	2	0.4984	0.03534	3.21E-04	0.4297	0.4985	0.5676	4001	17000
	7	0.494	0.03545	3.09E-04	0.4251	0.4938	0.5635	4001	17000
	19	0.4881	0.03508	3.09E-04	0.4197	0.488	0.5564	4001	17000
	24	0.4941	0.03479	3.60E-04	0.4261	0.4942	0.5612	4001	17000
	57	0.5008	0.03418	4.77E-04	0.4332	0.5011	0.5668	4001	17000
116	0.4754	0.03413	6.21E-04	0.4085	0.4754	0.542	4001	17000	

TABLE 6.10: Model comparison