

**Development and Application of Spatial and
Temporal Statistical Methods for Improved
Sampling of Wildlife**

Wilson Mwale Khaemba



Promotors: Dr. Dieter Rasch
Professor in Mathematical Statistics
Wageningen University, Wageningen

Dr. Alfred Stein
Professor in Spatial Statistics
International Institute for Aerospace Surveys and Earth Sciences (ITC),
Enschede

Co-Promotor: Dr. Jan De Leeuw
Associate professor in Environmental Systems Assessment
International Institute for Aerospace Surveys and Earth Sciences (ITC),
Enschede

UN08201,2872

Wilson Mwale Khaemba

Development and Application of Spatial and Temporal Statistical Methods for Unbiased Wildlife Sampling

Thesis

To fulfil the requirements for the degree of doctor
on the authority of the rector magnificus
of Wageningen University,
dr. ir. L. Speelman,
to be publicly defended
on Friday 20th October 2000
at three O'clock in the auditorium of ITC, Enschede

un 984015

ITC Dissertation number 76

The research presented in this thesis was partly conducted at the International Institute for Aerospace Survey and Earth Sciences (ITC), P.O. Box 6, 7500 AA Enschede, The Netherlands



CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

© 2000 Wilson Mwale Khaemba

Printed by FEBODRUK BV, Enschede.

Cover design by Taib El Ghazi.

Development and Application of Spatial and Temporal Statistical Methods for Improved Sampling of Wildlife. Thesis, Wageningen University and ITC, With ref. Summary in Dutch

ISBN 90-5808-280-6

Propositions

Khaemba, W. M. (2000). *Development and Application of Spatial and Temporal Statistical Methods for Improved Sampling of Wildlife*. PhD Thesis, Wageningen University and ITC.

1.

Sampling of wildlife populations is improved by using appropriate sampling schemes and analysis methods, as well as modelling of space-time dependence.

This Thesis

2.

Full understanding of factors that influence wildlife abundance and distribution requires an interdisciplinary approach involving ecologists, statisticians and social scientists.

This Thesis

3.

Extension of current spatial statistical techniques is necessary for a full and thorough analysis of data collected through aerial surveys of wildlife populations.

This Thesis

4.

Consistent procedures to obtain wildlife population parameter estimates are preferable to accurate but expensive alternatives.

This Thesis

5.

Low precision and accuracy associated with parameter estimates from wildlife data can be improved by considering spatial dependence in observations.

This Thesis

6.

Statistics does not provide explanation for what has been observed, but rather induces clarity in the numerical reasoning leading to conclusions.

This Thesis

7.

Data are not just numbers, but numbers that carry information about a specific setting.

This Thesis

8.

The concept of superior sampling designs is a fallacy often propagated by inappropriate application of perfectly good designs.

9.

Discoveries are often made by not following instructions.

10.

Many researchers erroneously place more importance on statistical significance, leaving almost no room for fair reporting or what was truly observed.

11.

If one is a master of one thing and understands one thing well, one has at the same time, insight and understanding of many things.

Vincent Van Gogh

12.

He who has a why to live can bear almost any how.

Fredrich Nietzsche

13.

Blessed is he who never forgets where he comes from, for the world is round and however much you travel, just when you think you have reached, you find little difference with where you came from.

An African saying

This is for you mum... For all your faith, love and guidance.

Acknowledgements

The task of completing a PhD thesis is a very large one and in the course of carrying it out, I have had to rely on the encouragement, help and support of very many individuals and institutions. To try and name them all would fill more space than is practically possible. To all of them I say *ahsante sana*.

My coming to ITC and indeed my motivation to undertake this study could not have been possible without Alfred Stein. Throughout this period, I have fortunate to receive constructive criticism, guidance and support from him. I don't think I can find enough words to express the gratitude I feel for all that you have taught me Alfred. All I can say is that I was very lucky to have you as my supervisor and promotor, because through you I have learned to be a scientist! Thanks for translating my summary into Dutch.

I would also like to thank my first promotor Dieter Rasch. We started off well as friends in Gödölö and we will finish of well as friends in Enschede! Thank you for your support.

Many thanks go to Jan De Leeuw for being my co-promotor as well as a friend. I remember the great lengths you went through to help my family and I settle into our new house a few years ago!

I have also had the good fortune to work with Wilbur Ottichilo, Mohammed Said on several publications and I am grateful for their support. The social interaction enjoyed between us was no doubt beneficial to my stay here in the Netherlands. I also wish to thank Luc Boerboom for all his support and assistance with chapter 7 of the thesis.

Other friends that have made my stay here in the Netherlands more comfortable than it would have been and deserve special mention include Ale Raza, Apiyo, Citlali, Edwin Keizer, Farah, Ivan, Jimmy Ruto, Karin, Laurindo, Martin Blankestijn, Musyimi, Narcisso, Ogao, Rashid, Richard, Robert Albricht, Sam, Sedogo, Tiseke and Xuehua.

I would like to acknowledge the financial support extended to me by ITC, which enabled me to carry out this research. To this end, I thank Elisabeth Kusters for her support during my stay here. I can not forget the contribution of special people like Loes Colenbrander, Mireille Meester, Daniela Semeraro and Ceciel Wolters. They all offered secretarial support with a smile. Many thanks go to Roelof Schoppers and Job Duim, my favourite *rafikis*. You are great guys and treasures to ITC!

I also received considerable support from André Hintzbergen, Benno Masselink, Carla Gerritsen, Fely de Boer, Gerard Leppink, Hans Verdam, Harold Borkent, Hilde Hulsman, Jaap Duim, Hans Kuipers, Marcel Gerrits, Marga Koelen, Marion Pierik, Oscar Frowijn, Bert Riekerk, Roland Willink, Ronnie Geerdink, Tom Busscher and Wendy Stiksma.

Many thanks go all to those in the Soil Science division who tried to make me feel a little bit at home despite my being a square peg in a round hole! Special thanks to you David for all the support you have given me. I think you will take the division far and I wish you all the best. Thanks too to Abbas Farshad, Bart Krol, Rob Hennemann and Tom Loran for treating me as a friend and fellow scientist.

Many people at ITC thought that I belonged to the ACE division. And to be very honest, I have come to feel closer to that division than even the one I belong to. I owe all these to the respect and welcome that I received from Andrew Skidmore. I am grateful for inviting me to join the PhD tutorials and allowing me to share ideas with my peers.

To the indoor soccer squad: You were like family to me. Sorry my knee gave way earlier than I had hoped. Keep up the fire burning Johan Weggen! To my fellow fitness addicts: I have waited for a challenge for too long. Who will be the next Mr. ITC? Thanks for your support Ulco.

I thank my employer, Moi University, for allowing me to take study leave and pursue this research. Thanks too to DRSRS for the use of their data and the comradeship of all those who have been here at ITC.

To all the wonderful Dutch friends that I have made over the years, you are great. Keep up the warm smile and open heart. You were a great person Hermine and it is a pity that you are not here to share in my joy. But wherever you are, I am sure that you are happy. Thanks too to Miny and Erik Kortof, Whitney, Winny and Rob Krens as well as Will and Han. I will not forget you Theo and Jacqueline and your lovely children Leroy and Danistcha.

Much gratitude to members of my family for all the moral support they gave me during my studies. Thanks mum, for all the faith you have in me. Thanks to you too Jane Agusi for all the encouragement you have given me over the years. To my friends: Amu, Anne, Foxy, Gathenya, Kimeli, Sifa, Tobi and Toma. Although I have been away from you, you have kept me in your hearts.

I would finally like to thank Nancy, Elton and Alex for their love, patience and sacrifice. I can not imagine a way in which I would ever repay you. But it may help to know that you have always been a source of hope and encouragement all through the period that I have been away from you and I'll always love you for that.

Table of contents

Acknowledgements.....	vii
Table of contents.....	ix
Symbols and notations	xiii
CHAPTER 1: GENERAL INTRODUCTION.....	1
1.1. Statistics and Wildlife Management.....	2
1.2. Motivation of the study	3
1.3. Objectives of the study.....	5
1.4. Scope of the study	5
1.5. Outline of the Thesis	6
1.6. Location of study areas	7
Chapter 2: Designs for sampling wildlife populations	9
2.1. Introduction.....	11
2.2. Materials and Methods.....	12
2.2.1. Sampling wildlife.....	12
2.2.2. Sampling Designs	12
2.2.3. Jolly II procedure.....	13
2.2.4. An adaptive sampling strategy.....	14
2.3. Case Study: The Masai Mara Ecosystem of Kenya	17
2.3.1. Site description	17
2.3.2. Simulation.....	17
2.3.3. Simulated distributions	18
2.4. Discussion	23
2.5. Conclusions.....	24
Chapter 3: Modelling wildlife population using GIS	25
3.1. Introduction.....	27
3.2. Study area.....	28
3.3. Materials and Methods.....	28
3.3.1. Data.....	28
3.3.2. GIS Modelling	29
3.3.3. Distance measure of spatial correlation	31
3.3.4. Generalised Linear Models	32
3.4. Results.....	34
3.4.1. GIS modelling.....	34
3.4.2. Distance measure for spatial correlation	38
3.4.3. Generalised linear modelling	44
3.5. Discussion	46
3.6. Concluding remarks	47

Chapter 4: Modelling wildlife abundance using spatial statistics	49
4.1. Introduction.....	51
4.2. Materials and methods	52
4.2.1 Data.....	52
4.2.2 MCMC modelling.....	53
4.2.3 Space - Time modelling	54
4.2.4 Diversity measurement	56
4.3. Results and discussion.....	57
4.3.1. Survey data	57
4.3.2 MCMC modelling.....	57
4.3.3 Space-time modelling	59
4.3.4 Diversity measurement	62
4.4. Concluding remarks	64
Chapter 5: Spatial point patterns for modelling wildlife distribution	67
5.1. Introduction.....	69
5.2. Materials	70
5.3. Methods.....	71
5.4. Results.....	74
5.5. Discussion	85
5.6. Conclusions.....	87
Chapter 6: Comparison of wildlife sampling procedures	89
6.1. Introduction.....	91
6.2. Materials and methods	92
6.2.1. Study area	92
6.2.2. Data.....	93
6.2.3. Sampling designs in aerial surveys	94
6.2.4. Modelling and simulation	96
6.3. Results.....	99
6.4. Discussion	102
6.5. Conclusions	105
Chapter 7: Statistical procedures and decision making in the management of wildlife populations.....	107
7.1. Introduction.....	109
7.2. Materials and methods	110
7.2.1. Decision making process	110
7.2.2. Sampling designs	112
7.2.3. An evaluation matrix for wildlife management	112
7.3. Results.....	113
7.4. Concluding remarks	115
Chapter 8: Conclusions and Recommendations	117

8.1. Conclusions 118

8.2. Recommendations 121

References and Author bibliography..... 123

 References 124

 Author Bibliography 131

English Summary 133

Samenvatting (Dutch Summary)..... 137

Curriculum Vitae 143

Appendix 1: Completed Ph.D. studies at ITC 145

Symbols and notations

The following is a list of symbols and notations used in more than one chapter in the thesis.

N	: total number of sampling units in study region
n	: number of sampling units (sample size)
A	: size of the study area
a_i	: size of sampling unit i ($i: 1, 2, \dots, n$)
w	: width of area sampled on either side of the aircraft
D	: Population density (number per unit size)
\hat{D}	: Estimate of D
Y	: Population total (total number in study area)
\hat{Y}	: Estimate of Y
y_i	: number of observed animals in unit i
π_i	: Inclusion probability of the i th sampling unit
\bar{y}	: mean density per sampled unit
C	: Condition $\{y: y \geq c\}$ for a constant c needed to expand the sample
s_o	: sequence of labels and
S	: set of samples
β_j	: parameter corresponding to the j th variable
μ	: Mean population value
\hat{K}	: estimate of the total number of networks in the population
ξ	: A matrix of eigen vectors
z	: the standard normal variate
z_α	: the 100α per cent upper tail values for z such that $\Pr[z \geq z_\alpha] = \alpha$
\Pr	: probability
$E[y]$: the mean of a random variable y .
$\text{Var}[y]$: the variance of a random variable y
$E[x y]$: the mean of x conditional on fixed y , for random variables x and y
$\text{Var}[x y]$: the variance of x conditional on fixed y , for random variables x and y
PC	: principal component

CHAPTER 1

General Introduction

1.1. Statistics and Wildlife Management

Studies of biological populations often require estimates of population density (D) or total size (Y). These parameters vary in space, over time as well as by species, often in response to changes in environmental factors. Although ecology deals with living organisms, their habitats, modes of life and relations to the environment, it can greatly benefit from new developments in statistical procedures, particularly those used in estimation of population parameters and their precision.

The world is generally becoming more quantitative, with many professions depending more on data and numerical reasoning. As the use of data grows, so does demand for innovative statistical techniques that yield clarity in the data and help draw practical conclusions. Data are not just numbers, but numbers that carry information about a specific setting. They need to be interpreted in that particular setting.

This seemingly ecological thesis is written by a statistician in an attempt to combine statistical theory and ecological principles to provide tools for better management of wildlife populations. Although ecological principles exist to explain observed abundance and distribution of wildlife, only appropriate application of sound statistical procedures allows quantification, separation and understanding of the different processes at play in the ecosystem. Further, the modelling of variables influencing abundance and distribution of wildlife is possible only after a clarification of interactions between and within variables. This is only achieved by means of proper application of suitable statistical techniques. But why study wildlife?

Wildlife forms the backbone on which Kenya's tourism industries is based (Ottichilo, 2000). This industry has been a leading foreign exchange earner for the country for a long time, being second only to agriculture (GOK, 1994). In the recent past, the country's human population growth has far outpaced its economic growth, which has led to more people turning to subsistence agricultural production to meet an increasing demand for food (GOK, 1998). To fulfil this demand, more land traditionally reserved for wildlife and livestock grazing is being converted into agricultural land (Amuyunzu, 1984; Lamprey, 1984; Lusigi, 1986; Ottichilo, 2000). Combined with changing environmental conditions that have adversely influenced wildlife abundance and distribution, there are serious long-term implications on the survival of wildlife resources in Kenya (De Leeuw *et al.*, 1999).

A possible solution to this problem lies in formulating and implementing sustainable wildlife conservation strategies. These strategies, however, require proper understanding of factors that influence abundance and distribution of wildlife populations and how these factors interact with one another. In general, wildlife sampling is necessary for resource monitoring, conservation and proper management. It also provides useful input in policy planning. Wildlife managers require accurate, timely and up to date information on abundance and distribution of animals in their parks. Obtaining such information from wildlife surveys requires efficient sampling strategies and appropriate statistical analysis procedures. All these provide the basis and aim of this thesis.

1.2. Motivation of the study

Abundance of wildlife populations in the tropics is obtained by means of data collected by two basic counting procedures: complete enumeration or total counts (TC) and partial enumeration or sample counts (SC) (Caughley, 1977; Norton-Griffiths, 1978; Steffens, 1993). By means of TC, the whole study area is searched and all observed wildlife is counted to yield abundance. With SC, a representative sample of the whole study area is searched and all wildlife observations are recorded. Abundance is generally estimated by the following procedure:

Consider an area of known size A (km^2 , say) partitioned into N non-overlapping sampling units. Select a sample of n units out of the N and count the number of animals in each unit. If Y is the total number of animals in the whole area (which is unknown) and y_i is observed in unit i of area a_i ($i = 1, 2, \dots, N$), the mean density per sampled unit is defined for equally sized units as:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}, \quad (1.1)$$

and an estimate of Y is given by:

$$\hat{Y} = N \bar{y}. \quad (1.2)$$

It is highly unlikely that study areas will be regularly shaped to yield equally sized sampling units. A different estimation procedure must, therefore, be employed to yield \hat{Y} . One common approach is to estimate animal density in the sample and extrapolate this to the whole study area. The density D of animals in an area of known size A is estimated by:

$$\hat{D} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n a_i}, \quad (1.3)$$

which is a ratio of total sample counts to total sampled areas (Jolly, 1969a; Cochran, 1977; Thompson, 1992). An estimate of Y is then given by:

$$\hat{Y} = \hat{D} \cdot A. \quad (1.4)$$

This procedure yields correct estimates for unequally sized sampling units under the following key assumptions.

- All sampling units in the population (study area) have an equal chance of being included in the sample, which implies random sampling.
- All units in the sample are carefully searched and all animals in them are located and accurately counted.

The first assumption can be assured through application of random sampling schemes. On the other hand, the behaviour and dynamics of wildlife populations, as well as deficiencies in survey procedures, contribute to invalidation of the second assumption. This is particularly pertinent to SCs conducted through elaborate survey procedures.

Another difficulty associated with sampling wildlife populations is the definition of suitable sampling platforms. Light aircraft provide the only practical platform of sampling wildlife and are commonly employed in the tropics (Caughley, 1977; Seber, 1982; Steffens, 1993). For reasons of safety and security, such airborne platforms force observers to be separated by some distance from the sampled items. Moreover, sampling units are not physically defined, but are arbitrary regions demarcated on the ground by means of markers placed on the sampling platform (Smith, 1981; Seber, 1982; Thompson and Seber, 1996; Wint, 1998). All these introduce errors in the sampling procedure, which hamper accurate analysis of wildlife survey data.

Unlike in fields like forestry, soil science, geology, or even plant ecology, where items to be sampled are generally stationary, at least for a fixed period of time, wildlife populations are highly dynamic. A proper analysis of these dynamics must start by investigating the stochastic processes yielding observed spatial patterns. For example, failure to model the covariance structure in spatially dependent observations leads to inaccurate (biased) and imprecise (inefficient) parameter estimates (Isaaks and Shrivastava, 1989; Cressie, 1993; Legendre and Legendre, 1998). The same goes for wildlife populations, where analysis must take into account the existing spatial dependencies between observed patterns in relation to important restrictive conditions like presence of human settlements, distance to water points, distance to forests and other vegetation types, as well as other variables that affect wildlife distribution. Generally, natural plant or animal populations are rarely distributed at random but are usually clustered together (Seber, 1982; Buckland and Elston, 1993; Augustin *et al.*, 1996).

Analyses of data from wildlife surveys generally ignore the presence of this spatial dependence (Augustin, *et al.*, 1996; Grunblatt *et al.*, 1996). Use of conventional sampling designs such as simple random and systematic sampling, in which selection probabilities do not account for such clustering, results in inefficient estimates (Buckland, *et al.*, 1993; Augustin, *et al.*, 1996).

Recent advances in sampling and statistical analysis techniques provide opportunities to improve sampling of wildlife populations (Buckland, *et al.*, 1993; Thompson and Seber, 1996; Legendre and Legendre, 1998). In particular, the use of prior and expert/indigenous knowledge to improve analyses of survey data is highly desirable. This study is fortunate to have had access to data collected over a 20-year period, providing lots of prior information for future surveys. This has made it possible to apply computer intensive techniques like MCMC methods to improve estimation of population parameters (Gilks *et al.*, 1996; Brooks, 1998).

1.3. Objectives of the study

The main objective of this study is to develop appropriate statistical procedures to improve sampling for purposes of gaining greater insight in the abundance and distribution of wildlife populations. This leads to the following minor objectives:

- To investigate current wildlife sampling procedures and propose an improvement.
- To model wildlife abundance and distribution using statistical techniques and geographical information system (GIS).
- To use spatio-temporal procedures to model changes in wildlife populations in space and time.
- To model wildlife distribution using spatial point patterns
- To compare performance of proposed sampling procedures to established sampling strategies.

1.4. Scope of the study

This thesis deals exclusively with large herbivores (body weight > 20 kg) found in an African savannah and the case studies presented here focus on these. Most of the data used have been collected through aerial surveys and the proposed sampling design assumes an airborne platform. The term *sampling design* is used here to indicate a method of selecting sampling locations to be observed as opposed to the actual removal of sampling locations in the population.

A strong emphasis is placed on estimation of abundance and distribution throughout the thesis because this is the main objective of many wildlife surveys in the tropics. However, an explanation of observed distribution patterns, especially in relation to environmental factors is relevant and is also treated here. This should not be viewed as a narrow utilisation of data obtained from wildlife surveys but rather as a focus on appropriate information requirements for wildlife managers.

Since the main objective of this study is the development and application of sampling methods and other procedures, illustrations have been selected from differing case studies and species. For most illustrations, different social behaviour is distinguished and represented by three types of wildlife species. Kongoni (*Alcelaphus buselaphus*) are considered solitary, the elephant (*Loxodonta africana*), are found in small to medium herds of up to 50 animals while the wildebeest (*Connochaetes taurinus*) are mainly found in large herds of several hundred animals.

Although this may appear to limit a full discussion of results, generalisation to other species and ecosystems requires only slight modifications in most cases. Further, two of the above species are considered key species because of their influence on the ecosystem. For example, through its migration process, the wildebeest facilitates the survival of other

species in the Masai Mara ecosystem (Vasey-Fitzgerald, 1960; Bell, 1970; Prins and Olf, 1998).

1.5. Outline of the Thesis

This thesis essentially presents a collection of research papers that have been accepted for publication in or submitted to international peer-reviewed journals. Each paper has been presented as a chapter, making a partial but distinct contribution to solving the overall research problem. I have tried as much as possible, to maintain the content of each paper to reflect what was presented to the journal, however, some standardisation in the layout and symbols is necessary for consistency of the thesis. In this respect, each chapter is introduced separately, stating its contribution to the overall research. This approach is likely to lead to some gaps and overlaps, affecting the continuous flow of the thesis. It is my hope, however, that the advantages gained through a critical review process has raised these chapters to a level that renders this drawback insignificant. The following is an outline of the main chapters.

Chapter 2 looks at different sampling designs used in aerial surveys of wildlife populations. It proposes an improved sampling design based on adaptive sampling and applicable to clustered populations. Univariate and multi-variate treatments of the improved design are developed. The thesis begins with a chapter on sampling because it is my belief that estimation of wildlife population parameters will only improve with an improvement in wildlife sampling procedures.

Chapters 3 and 4 integrate the concept of spatial statistics in modelling wildlife populations. In particular, chapter 3 illustrates how generalised linear modelling combines with GIS to model the distribution of wildlife species. The chapter also introduces a distance measure that models spatial dependence and characterises clustering of wildlife species.

Chapter 4 extends the use of statistical techniques in modelling abundance and distribution of wildlife by introducing MCMC methods and space-time analysis. The chapter also explores diversity measures and develops a diversity index suitable for aerial surveys of large herbivores.

Chapters 3 and 4 work with data obtained by means of strip transects, which is a common sampling unit in wildlife sampling.

More detailed data, in which exact geographical locations of groups of animal have been recorded, are available from a second study area. These are used in chapter 5 to model the spatial point patterns of six ungulates in the area. Differences in the six wildlife species are demonstrated by relating observed patterns to environmental factors like vegetation type. Nearest neighbour distance measures like the *G*-statistic and the *K*-function are also used to classify observed patterns as clustered, regular or completely random.

Chapter 6 uses the same detailed data as in chapter 5 to compare performance of the proposed adaptive design to the conventional random and systematic designs. The chapter combines statistics with GIS in a simulation study that validates the proposed sampling design.

Chapter 7 gives a brief outline of how spatio-temporal techniques and other statistical methods introduced in earlier chapters are used in the decision making process for better wildlife management. The chapter describes different scenarios encountered when making a decision for example, to adopt one statistical sampling technique instead of another. It should be noted that treatment of decision theory in this chapter may be inadequate because that would go beyond the scope of the thesis.

Finally, chapter 8 summarises the findings of the research and concludes with recommendations for further research. A combined reference list is given at the end of thesis.

1.6. Location of study areas

Concepts and procedures introduced in the thesis are illustrated by data from two ecosystems in Kenya. Transect data is mainly obtained from the Masai Mara ecosystem in Narok district, lying between $0^{\circ} 45'$ and $2^{\circ}00'$ South, and $34^{\circ} 45'$ and $36^{\circ}00'$ East and covering an area of approximately $6,650 \text{ km}^2$. The ecosystem straddles the international border of northern Tanzania and southern Kenya and is part of the larger Serengeti-Mara ecosystem. The detailed dataset comes from Laikipia district, most of which lies to the north of the Equator between latitudes $0^{\circ} 17'S$ and $0^{\circ} 45'N$ and longitudes $36^{\circ} 15'E$ and $37^{\circ} 20'E$, covering an area of approximately $9,700 \text{ km}^2$. More details of the study areas and descriptions of the data sets used are provided in each chapter.

CHAPTER 2*

Designs for sampling wildlife populations

* This chapter is based on Khaemba, W. M. and Stein, A. (*submitted*). Improved airborne Survey Sampling and Estimation of Parameters for African Wildlife Species. *Wildlife Research*.

Abstract

Parameter estimates, obtained from airborne surveys of wildlife populations, often have large bias and contain large standard errors. Sampling error is one of the major causes of these imprecision and occurrence of several wildlife populations in herds violates common assumptions in traditional sampling designs like systematic or stratified random sampling. In this chapter, we present an adaptive sampling design that uses criteria on observed animal counts to maximise sample information, and is independent of the usual assumption of a uniform distribution for animal populations. The design is applied to data derived from a survey carried out in Masai Mara Ecosystem (Mara) in Kenya, with focus on three animal species: elephant (*Loxodonta africana*), kongoni (*Alcelaphus buselaphus*) and wildebeest (*Connochaetes taurinus*). Its more efficient estimates show an improvement to those from the conventional systematic design with a more than 10 times reduction in estimated bias and a 37% lowering of the standard error. The adaptive design, however, underestimates population totals for species in large herds, while a multivariate extension only gives marginal improvements.

KEY WORDS: Adaptive sampling; Jolly II procedure; Wildlife Survey; Bias; Population total.

2.1. Introduction

Sampling for characterising African wildlife populations is important for various reasons. First, there is an obvious ecological need to obtain information about the ecological state of the reserve. Second, an increasing economic interest exists because wildlife, through tourism, contributes to national incomes of many African countries. Third, knowledge of year-to-year variation helps to identify and explain environmental aspects influencing population sizes of various animal species.

African wildlife populations are mainly sampled by means of periodic airborne surveys (Norton-Griffiths, 1978). For example, Systematic Reconnaissance Flights (SRF) (Norton-Griffiths, 1978) are used by the Department of Resource Surveys and Remote Sensing (DRSRS) to sample all wildlife populations in the Kenyan range-lands (Grunblatt *et al.*, 1996).

While estimates of population totals constitute the most important information obtained from wildlife surveys for most managers, reported estimates are often biased and have large standard errors (Caughley, 1974; Pollock and Kendall, 1987). Bias may be caused by departures from basic design assumptions, different animal behaviour at different times of day or adverse weather conditions like cloud cover that lead to poor visibility. On the other hand, large standard errors are mainly a result of using invalid sampling designs, inefficient sampling or poor choice of estimators.

Stratified Random Sampling (SRS) and Systematic Sampling (SS) are commonly applied sampling designs for wildlife surveys (Seber, 1982), with estimation of population parameters commonly being via the Jolly II procedure (Jolly, 1969a). If accurate or "true" counts exist, Jolly (1969b) has shown how correction factors reduce bias in estimates of population parameters. Similarly, sampling error is decreased through the use of larger samples, rigid standardisation of survey methods and introduction of efficient sampling procedures (Cochran, 1977; Thompson, 1992).

Recent developments have yielded modern sampling strategies like distance sampling (Buckland, *et al.*, 1993) and adaptive sampling (Thompson and Seber, 1996). Further, Van Groenigen and Stein (1998) illustrated how prior information improves sampling for environmental variables.

This study aims to improve precision of wildlife population estimates derived from airborne survey data by extending the Jolly II estimator with an adaptive procedure. A modified sampling strategy, based on strip transects, is formulated and its performance compared using a simulation of observed data. Throughout the paper, we focus on three herbivore species having different social behaviour: elephant, kongoni (a type of antelope) and wildebeest. The elephant lives in small herds of not more than 50 animals, kongoni is generally solitary, while the wildebeest occurs in large herds of hundreds of animals. Procedures developed here, however, are well suited to other wildlife species and habitats.

2.2. Materials and Methods

2.2.1. Sampling wildlife

Sampling of wildlife populations through airborne surveys is subject to two basic assumptions: Representativeness of the population by the sample and a uniform distribution of wildlife within the study region. The first assumption is questionable as animals may hide at the sound of approaching aircraft, or may be indistinguishable from background patterns due to camouflage. Further, dynamic animals may be observed twice. Quite often, therefore, sample numbers either underestimate or overestimate population totals. The second assumption also fails because many animals occur in herds of varying sizes due to influences of local conditions like vegetation, water availability and human settlement. For such animals, the probability of observing a single member depends upon the probability of the whole herd being observed. Sampling designs that ignore such dependence introduce imprecision in the estimation procedure that can not be quantified. Moreover, straightforward application of SRS or SS leads to high sampling errors.

Most survey procedures generally partition study regions into N windows, with the i th window W_i of area a_i having y_i animals ($i = 1, 2, \dots, N$). Sampling observes the population through a reduced number of windows (sample) of size n ($n < N$). In this study we let our windows be represented by strip transects, which are of primary interest in airborne surveys of wildlife because of the ease of navigation (Caughley, 1977). Transects are defined on the ground by markers fixed on the windows and wing-struts of an aircraft. Animal counts are recorded through observations made on one or both sides of the aircraft, which flies in a straight line from one end of a study region to the other at fixed speed and height above ground. Each transect is divided into 5 km subunits to yield distributional data.

2.2.2. Sampling Designs

Both stratified random sampling (SRS) and systematic sampling (SS) provide a basis for development of sampling designs. With SRS, n out of N transects are selected without replacement to yield an inclusion probability of $\pi = n/N$ for each transect, making them all equally likely to be selected into the sample. For SS, the N transects are numbered from 1 to N and an integer r is chosen such that $N = r \cdot n$. The initial transect is randomly chosen from the first r transects in the population and thereafter, every r th transect is selected. Since N is generally not an integral multiple of r , systematic samples obtained from the same finite population may vary in size. An improvement treats the N transects as being arranged round a circle and takes r as the nearest integer to N/n (see Cochran, 1977, p. 206). The first transect is randomly identified from integers between 1 and N , after which every r th transect is included in the sample until the required sample size n is obtained. This results in an equal inclusion probability for all transects in the population.

Both the total area for the study region $|W| = \sum_{i=1}^N a_i = A$ as well as the sampled area

$|w| = \sum_{i=1}^n a_i$ must be known to estimate the population total Y . The density D in W for uniformly distributed populations, is estimated as

$$\hat{D} = \frac{\sum_{i=1}^n y_i}{|w|}. \quad (2.1)$$

Therefore, an estimate of Y is given by

$$\hat{Y} = \hat{D} \cdot A, \quad (2.2)$$

with a variance given by Jolly (1969a) and Thompson (1992, p. 60) as:

$$Var(\hat{Y}) \equiv \frac{N(N-n)}{n(N-1)} \sum_{i=1}^n (y_i - D \cdot a_i)^2, \quad (2.3)$$

which is estimated by:

$$\hat{Var}(\hat{Y}) = \frac{N(N-n)}{n(n-1)} \sum_{i=1}^n (y_i - \hat{D} \cdot a_i)^2. \quad (2.4)$$

Rasch *et al.*, (1999, p. 44) derive similar results for sampling without replacement from finite populations.

2.2.3. Jolly II procedure

Let correlated random variables Y and Z have measurements y_i and z_i for the i th sampling unit ($i = 1, 2, \dots, n$), and sample means \bar{y} and \bar{z} , respectively, and let Z be known. The ratio estimator \hat{Y}_R for the population total Y equals:

$$\hat{Y}_R = \frac{\bar{y}}{\bar{z}} \cdot Z. \quad (2.5)$$

If the relationship between Y and Z is linear through the origin and variance of Y around this line is proportional to that of Z , then \hat{Y}_R is the best among a wide class of other estimators (Cochran, 1977, p. 158). For wildlife surveys, Y and Z represent observed animals and area respectively. If y_i and a_i are observed animals and area of the i th transect, an estimate of density \hat{D} is given by equation (2.1), leading to an estimated Jolly II population total $\hat{Y}_J = \hat{D} \cdot A$ and has variance given by:

$$\widehat{Var}(\hat{Y}_j) = \frac{N(N-n)}{n(n-1)} \left(\sum_{i=1}^n y_i^2 + \hat{D}^2 \sum_{i=1}^n a_i^2 - 2\hat{D} \sum_{i=1}^n y_i \cdot a_i \right). \quad (2.6)$$

This procedure is applied for wildlife surveys regardless of whether SRS or SS is used (Smith, 1981). Although Norton-Griffiths (1981) suggests that SS estimates have conservative variances compared to SRS, a choice between the two is not entirely clear for wildlife surveys (Jolly, 1981).

2.2.4. An adaptive sampling strategy

2.2.4.1. The univariate case

Both SRS and SS are characterised by the selection of sampling units (transects) prior to the carrying out of surveys. This leads to a selection procedure that is independent of subsequent observations during surveys. In practice, this yields transect with many zero counts for sparsely or highly clumped populations, thereby resulting in imprecise estimates. We formulate an adaptive sampling procedure that improves precision.

An adaptive sampling design is a function $\psi(S|y)$ assigning conditional probability to every possible sample S in the population given observations y (Thompson, 1992). For each W_i , we define a neighbourhood H_i as a collection of windows adjacent to and including W_i , i.e. for sequentially numbered transects, $H_i = \{W_{i-1}, W_i, W_{i+1}\}$. This procedure requires a condition to be fulfilled on W_i before it extends S to $S \cup H_i$. For example, y_i could belong to a set C , e.g. an interval $C = \{y: y \geq c\}$ for a constant c . Since transect lengths are highly variable, we let c equal δ , the number of animals observed per unit area but projected on each subunit. This is equivalent to the number of animals observed per subunit and is calculated individually for each species based on data from previous surveys in the same region. Thus, if $y_i \geq \delta$, then W_{i-1} and W_{i+1} are included in S and observed.

Following Thompson (1990), we call the collection of transects observed as a result of initial selection of W_i a cluster. Within a cluster, a subset of transects forms a network if selection of any transect leads to inclusion of the whole subset. This implies that all transects in a network satisfy condition C . Edge units do not satisfy C but are in the neighbourhood of those that do. They are only included in the sample estimator if they are part of the initial sample. By classifying transects for which $y_i < \delta$ as networks of size one, given y -values and condition C , the population is uniquely partitioned into K networks. K is unknown and is estimated from the data.

For surveys of single species, sampling proceeds as follows: We select an initial sample of size n using SRS (without replacement) from N transects. The total area A of the study region and that of each transect a_i ($i = 1, 2, \dots, N$) must be known. We define C based on δ for each species, yielding k_0 observed networks, where $k_0 \leq n$ since the sample consists of distinct transects. The i th transect in the k th network ($i = 1, 2, \dots, t_k$) contains y_{ik} animals

and has an area equal to a_{ik} . We estimate the density of animals per unit area in network k (D_k) by

$$\hat{D}_k = \frac{\sum_{i=1}^{t_k} y_{ik}}{\sum_{i=1}^{t_k} a_{ik}}. \quad (2.7)$$

Using (7) would overestimate D since the sampling design deliberately selects transects with high abundance. We, therefore, correct \hat{D}_k by multiply it by a weight w_k equal to the proportion contributed by the k th network to the total sample area, i.e.

$$w_k = \frac{a_{.k}}{\sum_{k=1}^{k_0} a_{.k}}, \quad (2.8)$$

where $a_{.k}$ is the area of all transects covering network k and $\sum_{k=1}^{k_0} w_k = 1$.

An improved estimate of D is obtained by averaging over all k_0 networks as follows

$$\hat{D} = \frac{\sum_{k=1}^{k_0} \hat{D}_k \cdot w_k}{k_0}, \quad (2.9)$$

which we use to estimate Y through $\hat{Y} = \hat{D} \cdot A$.

An approximate variance of the estimate is given by

$$\text{Var}(\hat{Y}) = \frac{\hat{K}(\hat{K} - k_0)}{k_0(k_0 - 1)} \sum_{k=1}^{k_0} (y_k - \hat{D} \cdot a_{.k})^2 \quad (2.10)$$

where y_k equals the number of animals observed in network k ($k = 1, 2, \dots, k_0$) and \hat{K} is an estimate of the total number of networks in the population, obtained for example as $\frac{1}{\Pi} \cdot k_0$,

where Π is the probability of hitting a network, estimated as the minimum of 1 and the ratio of the average projection of observed networks on the flight baseline (pn) with the distance between initial transects (dt), i.e. $\Pi = \min\left(\frac{pn}{dt}, 1\right)$.

2.2.4.2. The multivariate case

Since most Africa airborne surveys are multi-species, we extend the adaptive sampling procedure in a multivariate way by associating the i th window W_i with y_{ij} animals

of the j th species ($i = 1, 2, \dots, N$ and $j = 1, 2, \dots, p$). By letting rows and columns represent windows and species, respectively, we denote the data in matrix form as

$$\mathbf{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdot & \cdot & \cdot & y_{1p} \\ y_{21} & y_{22} & \cdot & \cdot & \cdot & y_{2p} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ y_{N1} & y_{N2} & \cdot & \cdot & \cdot & y_{Np} \end{pmatrix}. \quad (2.11)$$

For a sample S of size n , we consider the design as a function $\psi(S|\mathbf{Y})$ that assigns conditional probabilities of selecting every possible sample S given \mathbf{Y} , with an objective of estimating a vector of population totals $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)'$, where $Y_j = \sum_{i=1}^N y_{ij}$, $j = 1, 2, \dots, p$. Each transect has to satisfy a condition C , which is a p -dimensional region corresponding to p animal species i.e. W_i satisfies the condition if $y_i \in C$, where $y_i = (y_{i1}, y_{i2}, \dots, y_{ip})$ is a row vector of values from matrix (2.11).

We base C on individual species or on a function of some or all species. Following Thompson and Seber (1996, p. 202-203), we define the following conditions:

- $C_1 = \{y_{ij}: y_{i1} \geq c\}$ – based on observations of one representative species only;
- $C_2 = \{y_{ij}: y_{ij} \geq c_j \forall j\}$ – based on observations of all species simultaneously (AND);
- $C_3 = \{y_{ij}: y_{ij} \geq c_j \text{ for some } j\}$ – based on observations of any one species (OR);
- $C_4 = \{y_{ij} : \sum_{j=1}^p y_{ij} \geq c\}$ – based on the sum of observations of all species.

In actual surveys, C_1 is most often used.

After choosing C , the population is partitioned into unique networks given \mathbf{Y} , thereby allowing for the estimation of population parameters for each individual species as in section 2.3.1. We let y_{ijk} represent the number of animals in transect i of network k for species j and estimate the population total Y_j for the j th species of density \hat{D}_j by $\hat{Y}_j = \hat{D}_j \cdot A$. Its variance is estimated by

$$\text{Var}(\hat{Y}_j) = \frac{\hat{K}(\hat{K} - k_0)}{k_0(k_0 - 1)} \sum_{k=1}^{k_0} (y_{jk} - \hat{D}_j \cdot a_k)^2. \quad (2.12)$$

The covariance between estimates of population totals \hat{Y}_j and $\hat{Y}_{j'}$ for species j and j' , respectively, is given by:

$$\text{Cov}(\hat{Y}_j, \hat{Y}_{j'}) = \frac{\hat{K}(\hat{K} - k_0)}{k_0(k_0 - 1)} \sum_{k=1}^{k_0} (y_{j,k} - \hat{D}_j \cdot a_k)(y_{j',k} - \hat{D}_{j'} \cdot a_k). \quad (2.13)$$

This represents the off-diagonal elements of a p by p variance-covariance matrix for the p – species, formed by combining (2.12) and (2.13).

2.3. Case Study: The Masai Mara Ecosystem of Kenya

2.3.1. Site description

Mara, which is part of the Serengeti-Mara Ecosystem, is situated in Narok district in southern Kenya and lies between $0^\circ 45'$ South, $34^\circ 45'$ East and $2^\circ 00'$, $36^\circ 00'$ East (Figure 2.1). It consists of the Masai Mara National reserve (MMNR), the Loita, Siana and Mara plains and covers an area of approximately 6,600 km². It is home to high wildlife diversity and is also an important dry season refuge for migratory wildlife species like the wildebeest (Brotten and Said, 1995). This species, together with the elephant and kongoni, are among the keystone species in this ecosystem (Sinclair, 1995a).

Data for this study come from a survey conducted in May 1997 (survey Id 9703, i.e. the third survey of 1997), in which eighteen transects were flown with a strip width of 0.282 km, giving a sampling fraction of 5.14%. The spatial distribution of species shows the elephant to be found mostly in MMNR and on the Mara plains, while the kongoni occupy the middle section of the ecosystem, along boundaries of MMNR and the Siana and Mara plains (Figure 2.2). Wildebeest, the most abundant species, mainly occur in Loita plains but are also found in Mara and Siana plains. The kongoni is observed only in eight of the 18 transects surveyed (Table 2.1) indicating a sparse distribution for this species. The other two species appear evenly spread, although they too have several transects with zero counts. Jolly II estimates of the population totals for the three animal species confirm the high abundance of the wildebeest (Table 2.1).

2.3.2. Simulation

We base a simulation study on the observed data to evaluate the quality of results from the adaptive sampling design. We let animal counts resemble Poisson clusters (Diggle, 1983) and reproduce spatial patterns similar to the original data. We start by creating a rectangular frame that fully encompasses Mara and generate finite population values to represent y_i animals in transect i as point objects in a two-step procedure. We uniformly distribute m_1 parent locations in the frame and generate m_2 points at each location, following a Poisson distribution with species specific mean values obtained from the 1997 survey. Within the frame, we generate a maximum number ($N = 350$) of transects using the actual strip width and overlay them on a map of Mara to yield transects of unequal lengths.

We then generate a systematic sample of $n = 18$ transects by first selecting one transect at random among the first 20 transects and then selecting every 20th row.

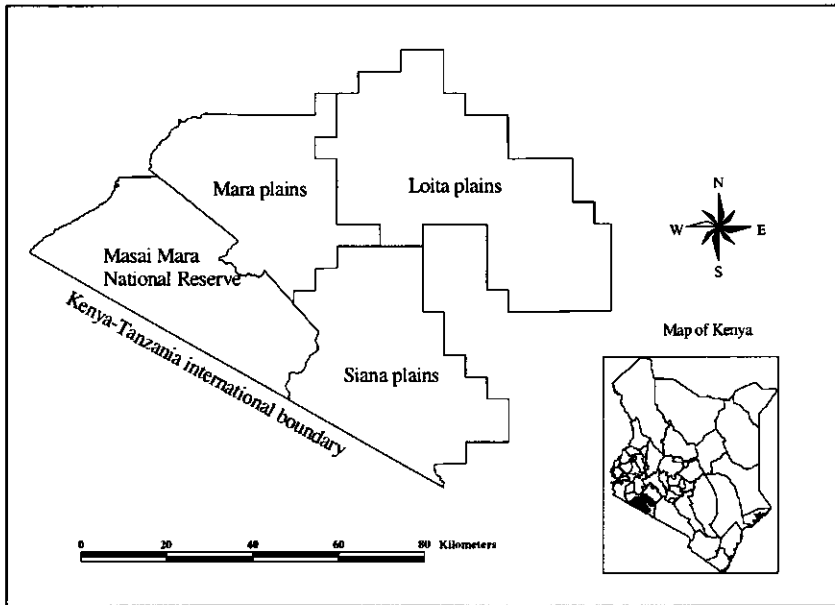


Figure 2.1. Location map showing the composition of the Mara ecosystem, which lies in Narok District of Kenya.

We use the Jolly II procedure to estimate population totals and standard errors. We estimate bias as the difference between the true and estimated population total. Similarly, we calculate population totals, standard errors and biases for adaptive samples generated with a random initial sample of size 10. For both designs, we calculate relative efficiencies as the ratio of the variance of the estimated population total obtained using a SRS design, to the variance of the same estimate obtained by either design at equivalent sample sizes. Sample sizes are standardised to the effective sample size of the adaptive design, which is defined as the number of transects in all networks. Mimicking the original observed patterns preserves correlation between species thereby allowing calculation of covariance in the multivariate case. The simulations are repeated 100 times, yielding 100 simulated populations for each species.

2.3.3. Simulated distributions

In the simulation of animal populations, each transect is checked for condition $C = \{y: y > \delta\}$ based on subunit density. Values of δ , expressed as animals per subunit, are equal to 1 for the elephant, 9 for wildebeest and 0.5 for kongoni, yielding mean effective

sample sizes of 20, 36 and 15, respectively. The estimated bias for the adaptive estimate of the elephant population total is much less in absolute terms than that obtained with the SS design, whereas the standard error reduces by 37% (Table 2.2). Less marked reductions are also observed for the other two species. Further, the adaptive design tends to give positive bias for solitary animals and small to medium herds like kongoni and elephant respectively, whereas abundance is underestimated by both methods for animals in large herds like the wildebeest. The adaptive design appears to be significantly more efficient in cases where effective sample size exceeds the 18 transects used by the SS design in this survey.

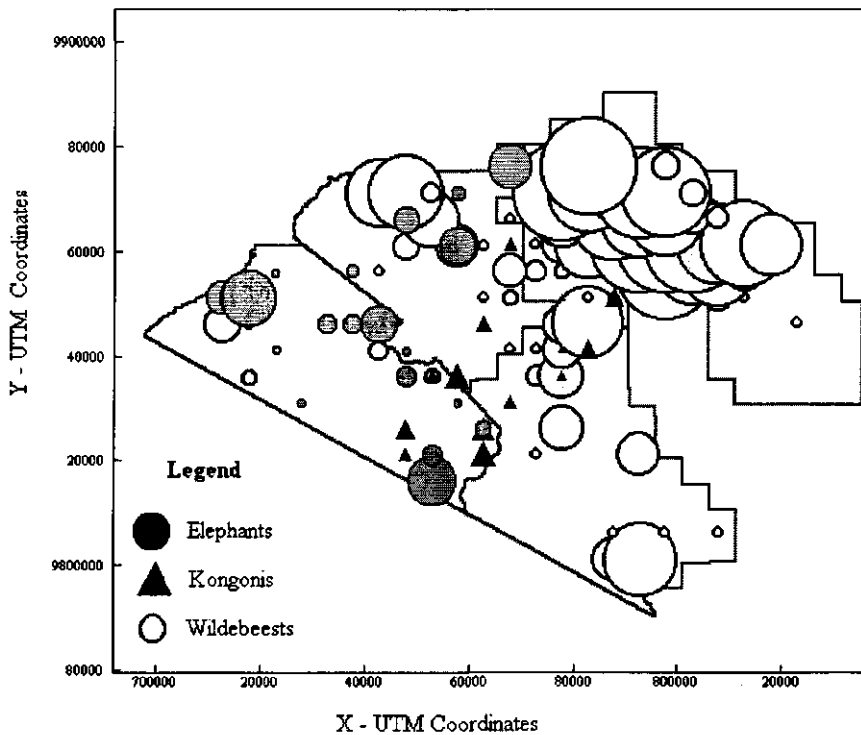


Figure 2.2. Point map, showing occurrence of elephant, kongoni and wildebeest in the study region.

Table 2.2. Comparison of adaptive design to the conventional systematic design for the three animal species. Table shows population totals Y , their estimates \hat{Y}_1 and \hat{Y}_2 , biases $B\hat{Y}_1$ and $B\hat{Y}_2$, standard errors $se\hat{Y}_1$ and $se\hat{Y}_2$ and relative efficiencies eff_{AD} and eff_{SS} obtained from adaptive and SS sampling designs, respectively, from simulations of three animal species' populations. Efficiency is defined as the ratio of the variance of \hat{Y} , estimated by a stratified random sampling design, to that obtained with an adaptive or systematic design for an equivalent sample size.

Species	Y	Adaptive method				SS design			
		\hat{Y}_1	$B\hat{Y}_1$	$se\hat{Y}_1$	eff_{AD}	\hat{Y}_2	$B\hat{Y}_2$	$se\hat{Y}_2$	eff_{SS}
Elephant	5,195	5,236	41	742	1.22	4,631	-564	1,178	1.19
Kongoni	2,398	2,482	84	600	1.03	2,301	-97	629	1.17
Wildebeest	44,329	43,187	-1,142	11,225	1.68	40,296	-4,033	13,908	1.06

Table 2.3. Results from a multivariate adaptive sampling. Table shows estimated totals (τ_e , τ_k and τ_w), biases (B_{τ_e} , B_{τ_k} and B_{τ_w}), standard errors (se_{τ_e} , se_{τ_k} and se_{τ_w}) and relative efficiencies (eff_{τ_e} , eff_{τ_k} and eff_{τ_w}), obtained from simulated populations of elephant, kongoni and wildebeest, respectively, for four multivariate conditions C_1 , C_2 , C_3 , and C_4 . Efficiency is defined as in Table 2.2.

Condition	Wildlife species											
	Elephant				Kongoni				Wildebeest			
	τ_e	B_{τ_e}	se_{τ_e}	eff_{τ_e}	τ_k	B_{τ_k}	se_{τ_k}	eff_{τ_k}	τ_w	B_{τ_w}	se_{τ_w}	eff_{τ_w}
C_1	5,231	36	728	1.27	2,385	-13	748	1.21	60,186	15,857	25,345	1.20
C_2	6,908	1,713	1,627	1.08	2,311	-87	601	1.05	52,561	8,232	18,842	1.29
C_3	7,683	2,488	2,320	1.35	2,590	192	823	1.42	38,436	-5,893	13,485	1.77
C_4	7,676	2,481	2,256	1.24	2,637	239	866	1.12	40,060	-4,269	14,703	1.64

Table 2.1. Observed numbers per transect for three animal species based on census 9703 conducted in Mara in 1997. Table shows transect area, distribution and Jolly II estimates of population totals for the elephant, kongoni and wildebeest.

Transect	Area (km ²)	Elephant	Kongoni	Wildebeest
1	2.784	0	0	0
2	8.212	0	0	72
3	11.101	0	0	3
4	12.758	0	0	0
5	14.097	37	0	0
6	14.485	7	21	18
7	14.661	4	20	20
8	21.923	2	3	0
9	25.092	7	23	29
10	26.857	2	9	20
11	29.712	33	6	79
12	33.975	73	11	17
13	31.882	3	0	293
14	31.996	22	4	684
15	22.766	12	0	296
16	20.106	3	0	495
17	13.87	31	0	108
18	6.156	0	0	0
Total observed numbers		236	97	2,134
Estimated population total		4,587	1,886	41,478
Standard error (s.e.)		1,391	632	14,332

Multivariate sampling results in mean effective sample sizes of 20, 16, 43 and 28 transects for conditions C_1 , C_2 , C_3 , and C_4 , respectively. Mean results indicate elephant abundance to be well estimated under C_1 , which is based on the elephant's density (Table 2.3). Estimated bias for kongoni remains small under all conditions, while that of the wildebeest is highly overestimated under conditions C_1 , and C_2 . Covariances between species are large and generally negative.

2.4. Discussion

This study shows how precision of estimates can be improved by using more information from observed populations. Though the method appears to underestimate population totals for animals in large herds, its results are better than those obtained from the conventional systematic sampling design. It could be a point for further research whether this underestimation is due to herd size or to species abundance. If the regression-through-the-origin model assumption is correct, more precision could be realised by deliberately choosing longer transects (Thompson, 1992). To provide robustness against such departures, initial samples should be randomly selected.

The conventional SS design may give better estimates than the adaptive design for uniformly distributed animal populations because such populations do not violate basic design assumptions. Such distributions, however, are unlikely to occur for wildlife populations in large areas, due to substantial spatial variation in factors affecting their distribution. Further, stochastic processes influencing this distribution are as yet poorly understood and are likely to differ between species (Maddock, 1979). The adaptive sampling design, therefore, becomes more efficient in such situations by capturing more information. It also responds better to underlying processes by depending on observed wildlife distributional patterns. For example, it is better suited to deal with a gradient in the data by sampling more intensively in areas with large densities.

The definition of relative efficiency in this study makes the adaptive design appear more efficient for abundant but clustered species like the wildebeest. Although an increase in sample size appears to favour the SS design, this is invalidated by the corresponding increase in transect variation. Defining condition C by using the number of observed animals for each subunit is superior to the mean number of animals per transect, as it takes variation in transect length into account.

Data quality often influences results obtained from airborne sampling of animal populations. For example, wildlife species like the wildebeest migrate each year from Serengeti in Tanzania to Mara during dry seasons (July – October) and return in wet seasons (December – June) (Maddock, 1979). Therefore, surveys carried out in Mara during these wet seasons would result in many zero counts for migratory species. If such prior information exists, it is reasonable to use an adaptive sampling design with a low initial sample size, most probably saving cost, particularly for clustered populations that are either sparse or abundant.

A multivariate extension of adaptive sampling strategies requires further research. The four conditions needed to add neighbouring transects to the sample indicate subjectivity in the adaptive procedure. The choice of a suitable condition must be based on prior information like individual species dynamics. In the absence of such knowledge, a pilot survey at a limited scale may be useful. If interest focuses on a single species, this study suggests that basing the formulation of C on values of the species may help in obtaining a

suitable condition. Even if an improved multivariate treatment was possible, implementation difficulty will increase with an increasing number of species.

2.5. Conclusions

We conclude from this study that an adaptive sampling design, using information from observed transects, can improve estimates of population parameters from airborne wildlife surveys. The method appears to perform better for clustered and highly abundant species as well as sparse populations. Extension to a multivariate setting does not substantially improve estimates but leads to more efficient data use. Modern ways of sampling, therefore, clearly improve airborne wildlife estimates, and more specific contributions may be anticipated in the near future.

Acknowledgements

The authors are grateful to DRSRS for permission to use part of their data.

CHAPTER 3*

Modelling wildlife population using GIS

* To be published as Khaemba, W. M. and Stein, A. (2000). Use of GIS for a spatial and temporal analysis of Kenyan wildlife with generalised linear modelling. *Int. J. Geographical Information Science*, Vol. 14, No. 8, 000-000.

Abstract

This chapter applies generalised linear statistical techniques in a GIS to analyse wildlife data from a Kenyan wildlife reserve and its surrounding areas. Attention focuses on the spatial distribution of elephant during nine successive surveys, analysing their temporal and spatial relationship and relating them to 12 explanatory variables. A principal component analysis identifies five major determining factors, thereby reducing dimensionality in the data, while a simple spatial analysis procedure, suitable for wildlife data obtained from airborne surveys, quantifies clustering for different animal species. The number of explanatory variables appearing in abundance models is found to be subject to large variations during successive surveys with a minimum and maximum of four and eight variables, respectively. Species from highly clustered populations are found to have over 20 times more observations within short distances compared to the rest. The study concludes that a combination of generalised linear modelling and GIS gives deeper insight into the dynamics of wildlife species in and around well-defined nature reserves.

3.1. Introduction

A lot of data on wildlife populations are routinely collected through airborne surveys. In most cases, these data are geographically referenced and therefore spatial in nature. Quite often too, the surveys are replicated over time, resulting in large amounts of spatial and temporal data that need to be translated into useful information. The advent of computer technology, particularly Geographical Information Systems (GIS), has led to better methods of data storage, retrieval and manipulation (Burrough and McDonnell, 1998). For example, wildlife surveys can now be carried out using sampling procedures that are optimised through consideration of landscape features and environmental factors stored in a GIS. Also, relationships between animal population dynamics and environmental factors can now be studied simultaneously. Suitable analysis techniques, which are capable of handling spatial and temporal variability, are however, indispensable in yielding useful and reliable information.

In most modelling situations, high correlation between and within variables directly affects the precision of estimated model parameters. Buckland and Elston (1993) use Principal Component Analysis (PCA) to reduce dimensionality and eliminate correlation between explanatory variables, while Augustin *et al.* (1996) model spatial autocorrelation in the distribution of wildlife. Both studies, however, do not simultaneously model spatial autocorrelation and correlation in the explanatory variables.

On the other hand, several studies have confirmed strong association between animal species abundance and environmental factors (Osborne and Tigar, 1992; Buckland and Elston, 1993). Among the statistical techniques used to establish these relations, logistic and multivariate regression analyses have been predominant (Nichols, 1989; Li *et al.*, 1997). A shortcoming of these linear models is the assumption of a normal distribution for the response variable. In situations where this assumption is not tenable, generalised linear models (GLMs) provide a useful alternative (McCullagh and Nelder, 1989; Dobson, 1990; Pereira and Itami, 1991). Both linear models and GLMs rely on an assumption that data are independently and identically distributed.

It is known, however, that factors influencing animal distribution i.e. vegetation cover, soil fertility, distance to water points etc., are spatial in nature. Therefore, there is need to combine GLM and GIS in modelling animal distribution so as to account for spatial dependence.

In the past, Walker and Moore (1988) interfaced the software package SIMPLE with a GIS to model wildlife distribution. In a subsequent study to relate kangaroo distribution to climatic conditions, Walker (1990) combined the package with GLMs to obtain results that compared favourably with those from the rule-based CART algorithm (CART, 1984). This indicated the usefulness of GIS modelling as a strategy to analyse wildlife data. In both studies, however, abundance was re-expressed into presence and absence data to allow the use of logistic regression in the modelling. This led to some loss of information.

The objective of the current study is to analyse observed spatial and temporal patterns of four wildlife species, i.e. elephant (*Loxodonta africana*), kongoni (*Alcelaphus buselaphus*), wildebeest (*Connochaetes taurinus*) and zebra (*Equus burchelli*), found in a Kenyan wildlife ecosystem. We use GIS techniques to identify and relate environmental factors to observed counts. We carry out a spatial correlation analyses as well as generalised linear modelling and reduce dimensionality in the data by means of PCA. Finally, we outline the ecological implications of these analyses to management of the species. We often use data on the elephant to illustrate certain procedures.

3.2. Study area

The study area is the Masai Mara ecosystem (Mara), which is fully described in section 2.3.1 and illustrated in Figure 2.1. High rainfall, tall grassland and permanent water make Mara an important dry season refuge for the last great migrating herds of wildebeest and zebra (Delany, 1982; Murray, 1995; Fryxell, 1995; Broten and Said, 1995). These two species, together with the elephant and kongoni are among the most important large herbivores found in Mara. Mara is also important from an economical perspective because of attracting many tourists who bring in foreign earnings (GOK, 1984). All these have led to an increasing interest in analysing factors that influence abundance and distribution of wildlife over time (Sinclair, 1995b).

3.3. Materials and Methods

3.3.1. Data

Data for this study are a subset of a comprehensive dataset of surveys conducted in Mara since 1977, by the Department of Resource Surveys and Remote Sensing (DRSRS), of the Ministry of Planning and National Development. We concentrated on post-1990 surveys because those carried out previously may not have been sufficiently uniform and were, therefore, subject to unrealistic variation. Typical surveys simultaneously observe several animal species. A number of pre-determined strip transects (18-55) are systematically flown by a light aircraft at fixed height (typically 122 m) and fixed strip widths (282 m or 304 m). These are designed so as to cover a sampling fraction of between 3.5% and 11.8% for the Mara (Grunblatt *et al.*, 1995). Distances between transects may vary for successive surveys, but 2,500 m, 5,000 m and 10,000 m are typical values (Grunblatt *et al.*, 1995).

Transects are usually divided into sub-units of 5,000 m length for which *x* and *y* co-ordinates, as well as the number of observed animals of different species, are recorded. Based on these co-ordinates, remotely sensed environmental data such as vegetation cover, composition and greenness are associated with each sub-unit. Similarly, the shortest distance to permanent rivers, major roads and the boundary of MMNR are calculated. Legislative changes have placed MMNR under favourable protection and conservation status as compared to the surrounding areas. Effects of this protection are studied by

defining an explanatory variable that assigns 1 to areas within MMNR and 0 otherwise. Average soil fertility as well as the extent of agricultural and fallow land were also recorded for each subunit. Finally, the data were verified and formatted before being stored into a GIS.

The three explanatory variables for distances to rivers, roads and MMNR were derived by rasterising and resampling digitised topographic maps of a 1:250,000 scale to pixels of size equal to 30 m. We obtained vegetation cover variables by a supervised classification of georeferenced Landsat TM images, which were complemented by ground-truthing. Further, Normalised Difference Vegetation Index (NDVI) values were calculated from NOAA-AVHRR images that had a resolution of 1.1 km. These were georeferenced using the Landsat TM image and resampled to a 30 m pixel size. Finally, the soil fertility map of Mara (Jeatzold and Schmidt, 1983) was digitised, rasterised and resampled to a 30 m pixel size. In all analyses, observed animal numbers were treated as response variables.

3.3.2. GIS Modelling

In most wildlife studies, modelling with GIS focuses on defining habitat suitability for various animal species (Tomlin *et al.*, 1983; Walker, 1990; Pereira and Itami, 1991). However, for modelling wildlife populations, both permanent and dynamic data recorded during surveys are required. Functional GIS layers, mainly containing topographical data like vegetation cover, drainage and road networks, can store these data as observed or interpolated counts of individual animal species (Burrough and McDonnell, 1998). For this study, major roads, the MMNR boundary, and the drainage pattern of permanent rivers form three such data layers (Figure 3.2).

Modelling was preceded by a descriptive and correlation analysis of all variables in this study. We restricted the analyses to survey 9703 of 1997 and only used data on the elephant to illustrate the procedure of modelling spatial and temporal distribution. The elephant are an important indicator species, being protected and subject to changing nutritional conditions (Delany, 1982; Dublin, 1995). They also possess solitary and clustering characteristics and are easily observable due to their large size. We hypothesised the elephant abundance to be related to several environmental factors. The exact form of this relation being unknown, but able to be modelled using a GLM. We used the 12 explanatory variables to represent different GIS layers, whose combined overlay would have been too complex for interpretation (Table 3.1). We, therefore, started modelling by making bivariate comparisons.

We followed a deductive modelling approach, although inductive modelling is also common in GIS studies (Walker, 1990; Pereira and Itami, 1991). Deductive models start from a general idea that is applied to specific observations, while inductive modelling derives general principles from observing many specific examples (Stoms *et al.*, 1992). As a starting point, we utilised the knowledge that elephant generally prefer tall grassland and shrubby vegetation (Rosero, 1997). Similarly, past studies of Mara suggest that elephant are

mainly found in MMNR (Broten and Said, 1995). We tested these ideas by successively overlaying observed elephant counts with data layers for tall grassland, shrubby grassland, boundary of MMNR, permanent rivers, major roads, average soil fertility and conservation status. A subset of layers that indicated high correlation with elephant abundance was identified and a combined overlay formed. We studied interaction between elephant and the other animal species by analysing spatial patterns observed during the same survey. We also modelled temporal changes in elephant abundance by analysing spatial patterns in the nine successive post-1990 surveys.

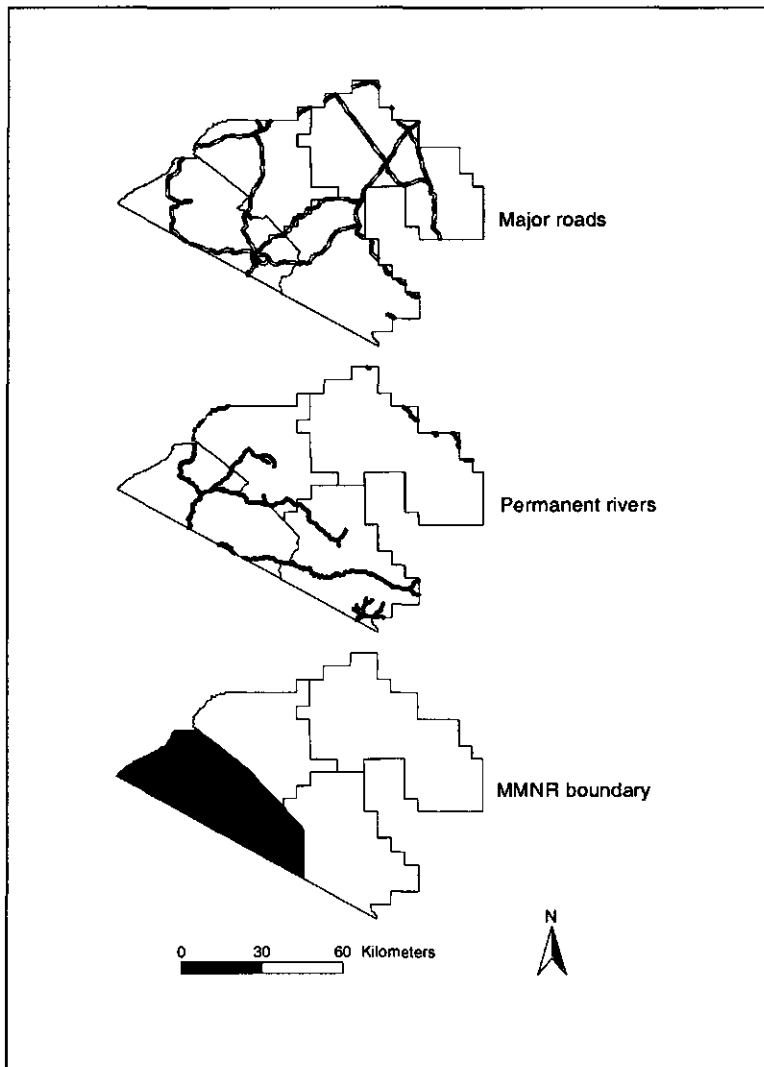


Figure 3.2. Three GIS layers formed from three explanatory variables, major roads in Mara, permanent rivers and the boundary of the Masai Mara National Reserve.

Table 3.1. Explanatory variables used in the study, together with their definitions. Shortened names are in brackets.

Explanatory variables used in the study	Definitions
Conservation Status (CONSTAT)	Binary variable, 1 for areas in MMNR and 0 otherwise
Distance to Reserve (DISTRESERVE)	Continuos variable, measures distance to Reserve boundary (km)
Distance to River (DISTRIVE)	Continuos variable, measures distance to permanent rivers (km)
Distance to Road (DISTROAD)	Continuos variable, measures distance to major roads (km)
Average Soil fertility (AVFERTI)	Continuos variable, measures average fertility for the unit
Agricultural or Fallow land (AGROFALL)	Cover of land under agriculture or left fallow (%)
Dwarf Shrubby Grassland (DWSHGRA)	Cover under dwarf shrubby grassland (%)
Forest (FOREST)	Cover under forest (%)
Short Grassland (SHORTGR)	Cover under short grassland (%)
Shrubby Grassland (SHRUBBYG)	Cover under shrubby grassland (%)
Tall Grassland (TALLGR)	Cover under tall grassland (%)
NDVI (NDVI)	Normalised Difference Vegetation Index; Measure for vegetation greenness

3.3.3. Distance measure of spatial correlation

Landscape features and local conditions may govern the spatial variability observed in wildlife data and hence affect distribution of different species. Some species may cluster close to available nutrients but become increasingly sparse in drier sub-areas. Such variability can be quantified by spatial correlation measures dependent upon distances between neighbouring sub-units. A high occurrence of short distances implies clustered

populations. In this study we used Euclidean distances between sub-units, together with absolute differences in observed animal counts to calculate a measure of spatial correlation.

The neighbourhood configuration for each sub-unit includes neighbours along the transect as well as on adjacent transects. Therefore, distances from the centre of a sub-unit to its adjacent neighbours equals 5,000 m along the transect, 2,500 m and 5,590 m perpendicularly and diagonally to adjacent transects, respectively (Figure 3.3). We took the diagonal distance between two adjacent sub-units (longest separation) to represent a unit d and classified all distances into multiples of d . We considered clustering by counting the distances within intervals of d , $2d$, $3d$ or $4d$. We standardised the number of distances in each interval for comparison by determining proportions with respect to all pairs of sub-units with non-zero counts per species. We also calculated absolute differences in abundance between pairs of sub-units having non-zero counts. We expected most observations of sub-units with non-zero counts from clustered populations to fall into intervals with the smallest distances (d or $2d$). This is the same as expecting most observed animals to be found in adjacent sub-units leading to some similarity between this procedure and the h -scatter plot (Isaaks and Shrivastava, 1989: p. 52)

3.3.4. Generalised Linear Models

Most wildlife modelling involves transformation of the response variable (animal counts) into a presence/absence outcome so as to allow use of the logit link (Buckland and Elston, 1993; Augustin *et al.*, 1996; Li *et al.*, 1997), which models the probability of an animal's presence without giving information on abundance. Further, for animals in herds, equal weights are assigned to sampling units having only one animal as to units with hundreds of animals. This is clearly undesirable when modelling animal abundance for a wildlife reserve. Therefore, an alternative link function is necessary to model wildlife abundance.

GLMs allow fitting of models to data without the requirement of normality by letting the random component assume any distribution from the exponential family. The random component is then related to the systematic component through a monotonic differentiable function known as a link function (McCullagh and Nelder, 1989). Several link functions are available depending on the type of response variable. For normally distributed data the expected value of each datum y is given by:

$$E(y) = \mu = \sum_{j=1}^p x_j \beta_j . \quad (3.1)$$

Where $\beta_1, \beta_2, \dots, \beta_p$ is a set of unknown parameters corresponding to a set of p explanatory variables to be estimated, producing a linear predictor given by:

$$\eta = \sum_{j=1}^p x_j \beta_j . \quad (3.2)$$

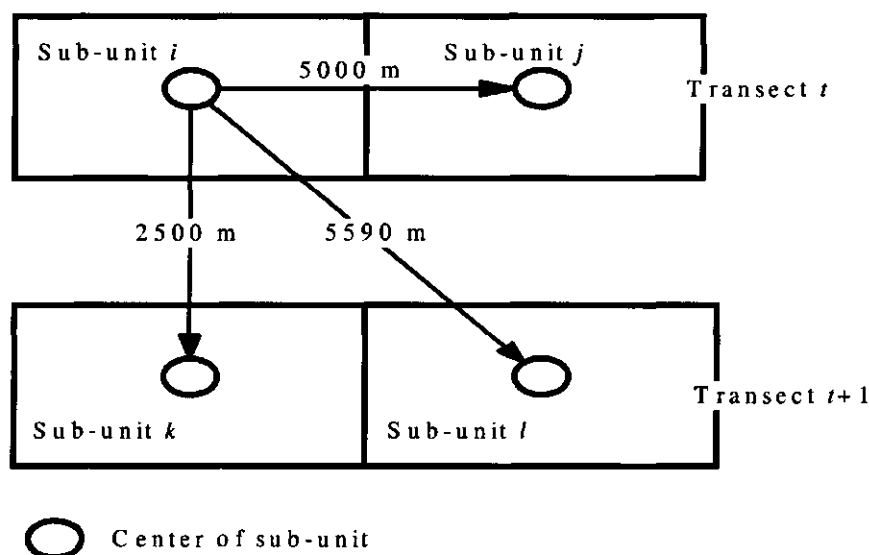


Figure 3.3. Distance calculation between neighbouring sub-units i, j, k and l , found on two adjacent transects t and $t+1$. The longest nearest neighbour distance is that between diagonal sub-units i and l , which is calculated using the Pythagorean theorem as shown above.

In case of normality, η equals μ and both can take any value on the real line leading to the identity link of classical linear models. For count data, however, μ must always be greater than zero, which invalidates the assumption of normality and the use of the identity link. We use a log link $\eta = \log \mu$ instead, leading to the inverse relationship $\mu = e^\eta$ and ensuring that μ is always positive. This formulation assumes a Poisson distribution for the observed counts and modelling proceeds via log-linear models. Again we only use data on elephant abundance from the nine surveys for this modelling.

3.3.4.1. Log - Linear Modelling

Poisson-like processes are best modelled using log-linear models, which avoid the restriction imposed by equality of the Poisson mean to the variance. The log-linear models used in this study assumed the following relationship between the variance and mean:

$$\text{Var}(Y_i) = \delta E(Y_i). \quad (3.3)$$

The dispersion parameter term δ is constant over the data, with $\delta > 1$ implying over-dispersion and $\delta < 1$ under-dispersion. Since most wildlife species exhibit clustering, we expect the variance to be higher than the mean (over-dispersion). We, therefore, assume a log-linear model of the form:

$$\eta_i = \log E(Y_i) = \mathbf{X}_i^T \boldsymbol{\beta}, \quad i = 1, 2, \dots, n \quad (3.4)$$

implying that

$$\mu_i = E(Y_i) = \exp(\eta_i) = \exp(X_i^T \beta) = \prod_{j=1}^p \exp(X_{ij} \beta_j). \quad (3.5)$$

This leads to a multiplicative relationship between μ_i and the covariate vector X_i . From (3.4), we regress $\log(\mu_i)$ on the set of explanatory variables to yield the regression equation:

$$\log(\mu_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi}. \quad (3.6)$$

We refer to this as log-linear regression, which is equivalent to generalised linear modelling with a log link. In common statistical software like S-plus, GLMs with the Poisson family imposes an over-dispersion parameter of one. This limitation is overcome by using the quasi family with a log link, which allows over-dispersion to vary from one model to another. We then estimate parameters via the quasi-likelihood estimating equations (qee) (McCullagh and Nelder, 1989: p. 327). This yields similar coefficients but with improved standard errors than those obtained via the usual maximum likelihood.

3.3.4.2. Principal component analysis

We use PCA to compose the original variables into linearly independent orthogonal principal components (PCs) and thereby reduce dimensionality in the data. Usually, there are as many PCs as there are variables, but with only the first few, which account for most of the variation in the data being considered in practice. For example, a PCA involving p variables decomposes into the following linear combination of original variables:

$$PC_i = \xi_{i1} X_{1i} + \xi_{i2} X_{2i} + \dots + \xi_{ij} X_{ji} + \dots + \xi_{ip} X_{pi}, \quad (3.7)$$

where PC_i is the i th PC and ξ_{ij} is the coefficient corresponding to variable X_{ji} . The coefficients ξ_{ij} form a matrix ξ composed of eigenvectors from the covariance matrix of explanatory variables. Eigenvalues derived from the covariance matrix represent the PCs' ability to reflect the original variables, with high eigenvalues implying high ability. Using data from survey 9703, new variables are derived from the PCs and used in a stepwise regression to determine which PCs best modelled the distribution of elephant in Mara for 1997.

In this study, we use the geographical information system software ARC/INFO 7.2 and ArcView 3.0a (ESRI, 1996) for map-production and spatial data analysis, while generalised linear modelling and PCA are carried out using S-Plus 4.5 release 2 (Mathsoft, 1997).

3.4. Results

3.4.1. GIS modelling

From descriptive statistics (Table 3.2), short and shrubby grasslands collectively account for nearly 70% of the vegetation cover type, with shrubby grassland being nearly

10% more abundant than short grassland (39.05% and 30.12%, respectively). We observe the highest correlation (0.654) between wildebeest counts and the number of sub-units under agricultural or fallow land, implying a possible conflict between man and beast. The negative correlation between the number of sub-units under the most abundant vegetation cover (short and shrubby grassland) and sub-units under agricultural or fallow land indicate possible clearance of land for human encroachment. The zebra are positively correlated with wildebeest as expected since both are migratory species preferring similar habitats. The amount of tall grass appear to decrease as one moves away from MMNR, which has a high conservation status as indicated by a high negative correlation coefficient (-0.495) between Tallgr and Constat. The positive correlation between Distreserve and distance to the river (Distrive) may be explained by the fact that permanent rivers in Mara are mainly found within MMNR. The high correlation between distance to MMNR and land under agricultural or fallow suggests that more agricultural activities are taking place close to or even within MMNR.

We observe a high occurrence of sub-units under tall grass cover in MMNR while modelling in GIS. This coincides with high elephant abundance in the same area (Figure 3.4a) thus supporting results from the correlation analysis. The many sub-units with a high shrubby grassland cover and no elephant support a weak correlation between shrubby grassland and elephant abundance. Other bivariate overlays suggest occurrence of more elephant within or close to MMNR as well as within short distances of permanent rivers compared to similar distances from major roads. The combined overlay of observed elephant counts, major roads and permanent rivers in Mara, as well as the boundary of MMNR reinforce this observation (Figure 3.4b). Overlaying these counts on other factors like NDVI, sub-units under agriculture or fallow land and human population density does not reveal a strong relation (Figure 3.4c). A high observation of elephant in MMNR, however, is in apparent avoidance of land under agriculture mostly found outside the reserve, which suggests a negative relationship between elephant abundance and increased agricultural activity in this ecosystem (Figure 3.4c).

Occurrence of elephant and kongoni interacts negatively with each other as well as with other species, as indicated by their occupation of spatially separate regions of the study area (Figure 3.5). We observe wildebeest and zebra in similar regions, again suggesting high interaction between their habitats. These two species were mainly concentrated in Loita plains, but were also observed in other areas of Mara.

From the nine successive surveys, elephant were mainly observed within or near MMNR (Figure 3.6). This supports the significance of Distreserve in explaining abundance and distribution of the elephant. Their spatial patterns differ significantly over time, with surveys 9102 and 9103 both being carried out in the same year but giving very different patterns. Survey 9102 was carried out at the end of the dry season (April), while 9103 was carried out after the rainy season (August), a period when migratory species come to Mara from Serengeti (Murray, 1995). This might explain the higher elephant abundance for 9103

(181 animals) compared to 19 animals for 9102, and suggests that season may be an influencing factor on elephant abundance. This is, however, contradicted by observations in the following year (1992) where two surveys 9202 (March) and 9204 (August) show nearly four times more elephant during the dry season (498) compared to the wet season (127). Even surveys carried out during the same month but in different years show little relation in the spatial distribution of observed elephant. For example, surveys 9003, 9103, 9204 and 9604, which were all carried out in August, show much year to year variation in their spatial distribution. These indicate the need to consider other factors in explaining elephant distribution.

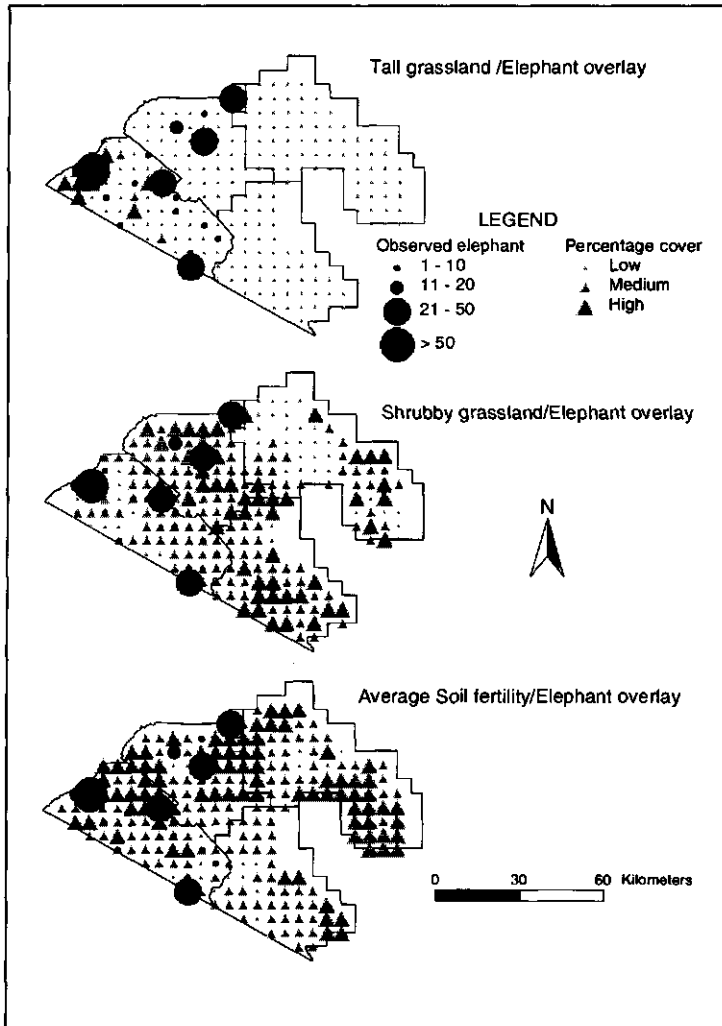


Figure 3.4a. Relating elephant abundance to several environmental factors using GIS modelling. Here elephant counts are overlaid singly on occurrence of tall grassland, shrubby grassland and average soil fertility at points of observation.

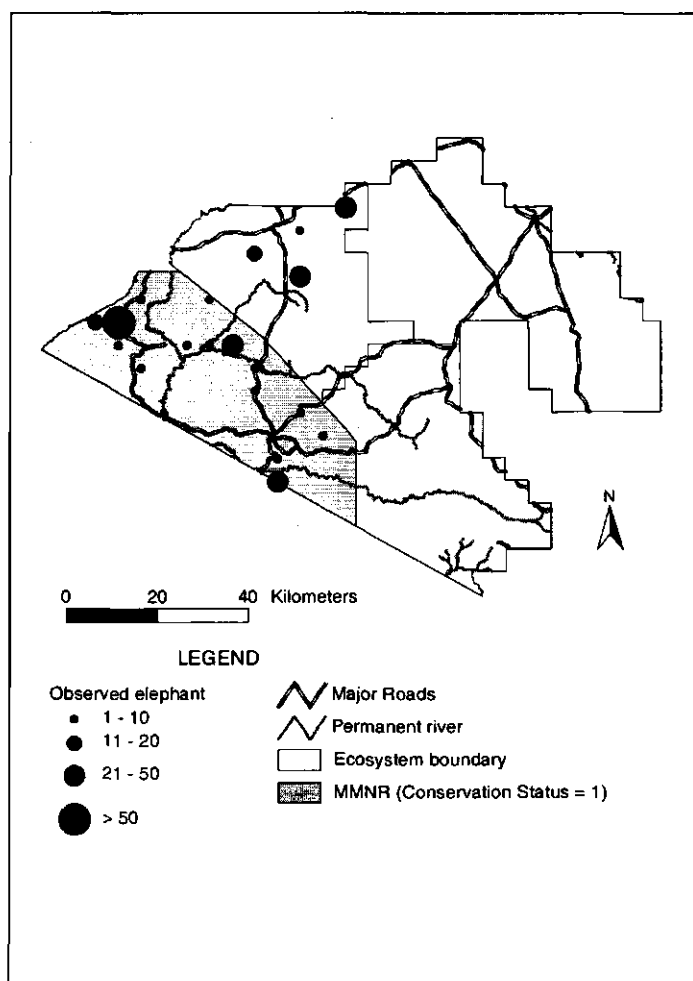


Figure 3.4b. Combined overlay of elephant abundance with major roads, permanent rivers in Mara and an area where some conservation is taking place (MMNR).

3.4.2. Distance measure for spatial correlation

We obtain measures of spatial correlation for the four animal species within distance intervals by computing proportions of distances between sub-units having non-zero counts (Table 3.3). We also calculate sums of absolute differences between all contributing pairs of sub-units. For example, survey 9703 yields 20 sub-units with non-zero elephant counts leading to 190 possible pairs. Three of these pairs are within distance d , representing a 1.6% proportion and a sum of absolute differences equal to 113. A similar approach is taken for all other distance intervals per species.

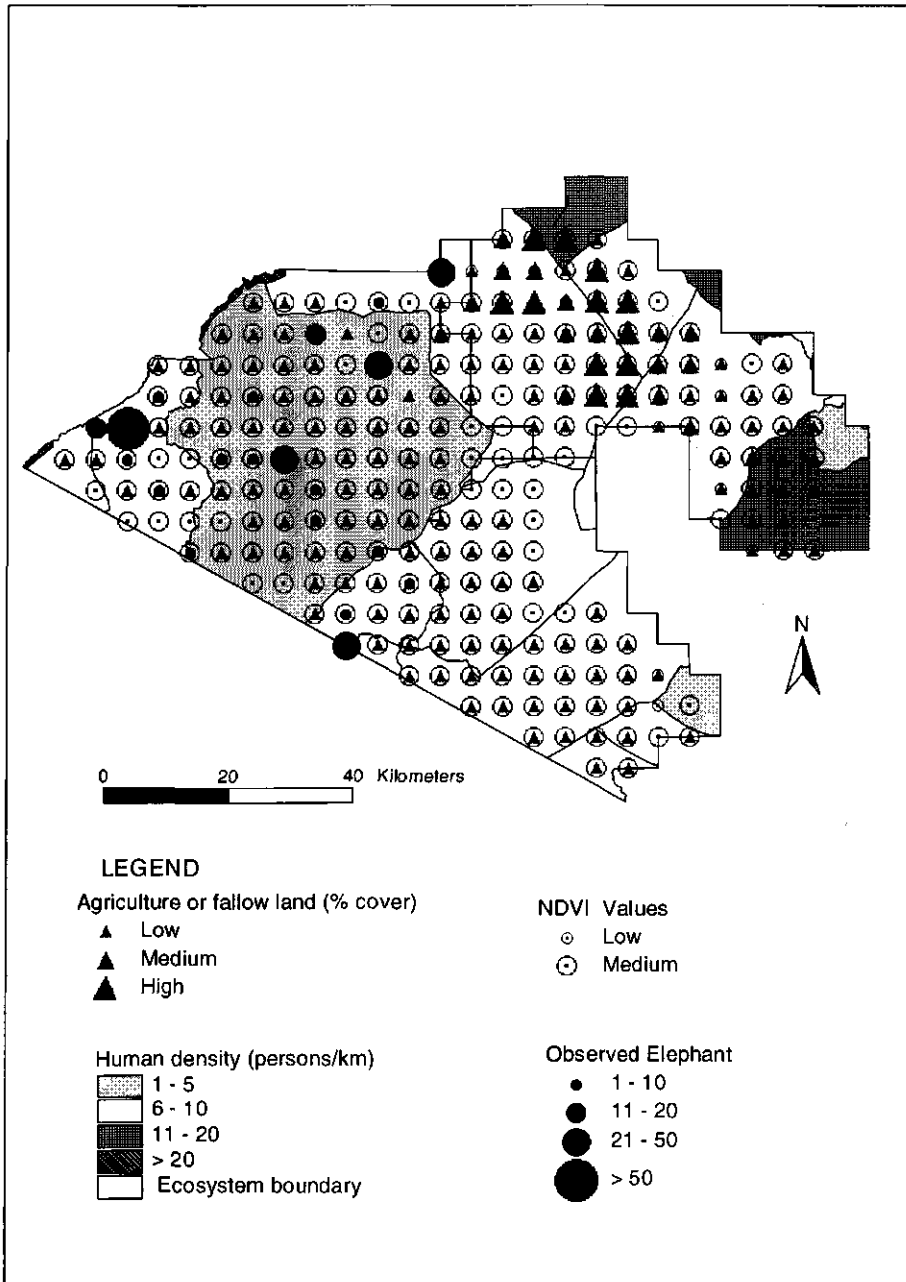


Figure 3.4c. Another combined overlay of elephant abundance with NDVI, places under agriculture or lying fallow and human population density. This is meant to show a simultaneous effect of these factors on the spatial distribution of the elephant.

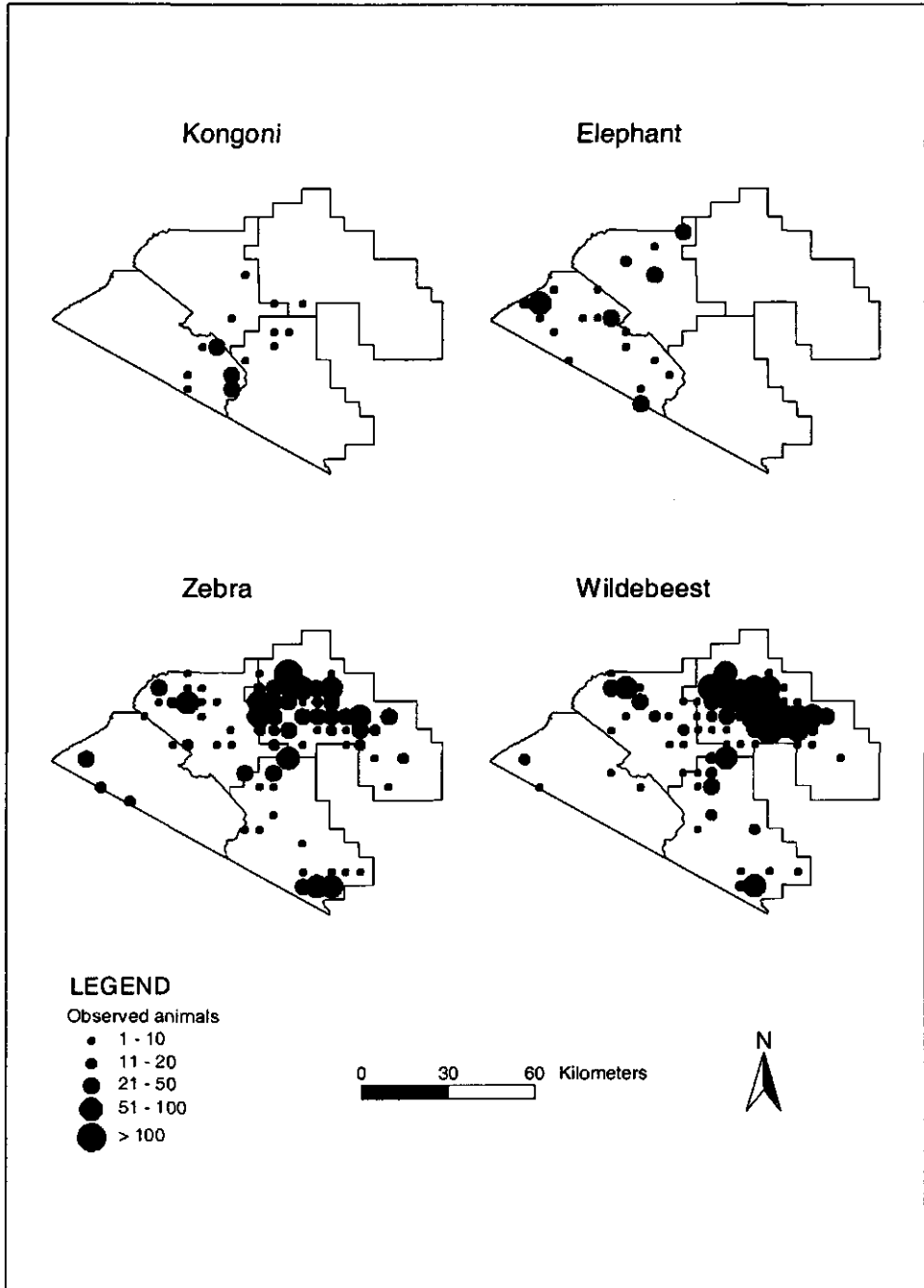


Figure 3.5. Observed counts for four common animal species observed during survey 9703. This shows the spatial relation between species and that surveys are multi-species in nature.

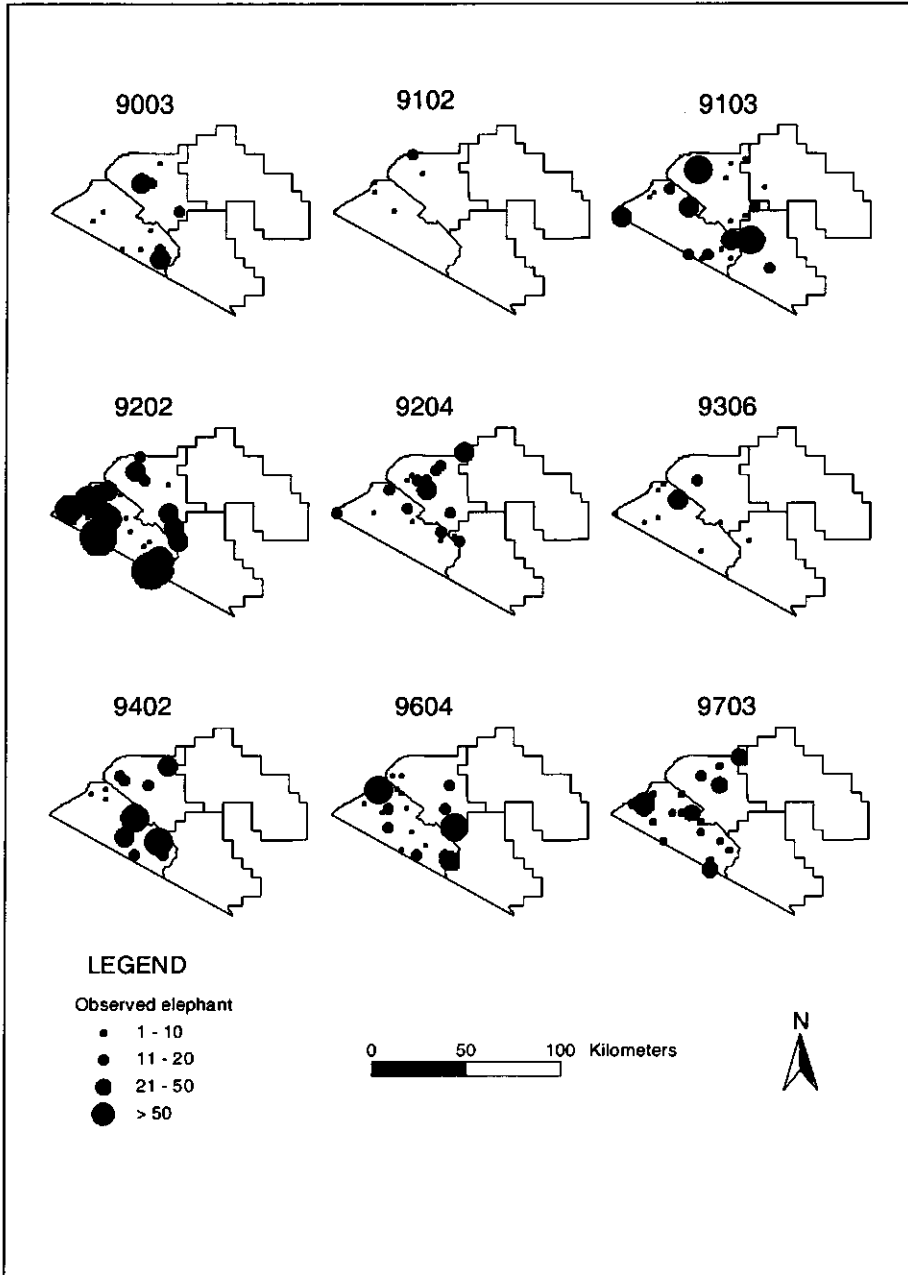


Figure 3.6. Elephant counts for nine surveys in Mara (1990-1997) to show temporal changes in abundance.

Table 3.3. Number of distances ($N(d)$ to $N(4d)$) and sum of absolute differences Σ , for pairs of sub-units with observed animals, at intervals of d , $2d$, $3d$ or $4d$ for the four animal species. Values in brackets are proportions based on all non-zero observations.

Animal Species	Separation Distance					
	$N(d)$	Σ	$N(2d)$	Σ	$N(3d)$	Σ
Elephant	3 (1.6)	113	8 (4.2)	120	21 (11.1)	327
Kongoni	2 (2.2)	19	3 (3.3)	10	9 (9.9)	67
Wildebeest	54 (2.2)	2 047	88 (3.5)	3 386	178 (7.2)	7 196
Zebra	49 (1.9)	1 183	87 (3.4)	2 472	204 (8)	5 452
						4 624

Table 3.4. Matrix of log-linear models for distribution of elephant counts observed in the Masai Mara Ecosystem over the period 1990-1997. Values in the table are regression coefficients for corresponding independent variables as given in table 3.1.

Variable	Census Year and Id									
	9003	9102	9103	9202	9204	9306	9402	9604	9703	
Constat	-1.766		0.919	-1.317	2.849			0.737	-1.150	
Distreserve			-0.064	-0.121	-0.071	-0.124	-0.130	-0.139		
Distrive	0.069	-1.161			-0.381			-0.110		
Distroad	-0.262		-0.126	0.083	-0.168	0.240	-0.179			
Avferti	-0.067	0.076		-0.018			0.015	-0.013	-0.035	
NDVI							-0.124			
Forest						-0.188		-0.078	0.054	
Tallgr	-0.042	0.049		-0.010					0.012	
Shortgr				-0.029			-0.031	0.014		
Shrubbyg							0.037			
Dwshgra					0.051					
Agrofall	-0.052				-0.239			0.046	-0.035	
Intercept	4.545	-6.927	-0.230	3.379	-0.509	-0.981	4.017	1.032	2.677	

The kongoni appear to be the least clustered of the four species by having a low number of pairs in the lower intervals compared to higher intervals (2, 3, 9 and 9 for d , $2d$, $3d$ and $4d$, respectively). Elephant exhibit both solitary and clustered behaviour by having comparable sums of absolute differences in three of the first four distance intervals (113, 120 and 156 for d , $2d$ and $4d$ units, respectively). The sum of absolute differences for both zebra and wildebeest are much higher than for kongoni and elephant in the first distance class (1,183 and 2,047 versus 19 and 113, respectively). We observe a similar situation in the other distance intervals indicating that zebra and wildebeest are highly clustered and spatially correlated. For the four animal species, this spatial correlation decreases with increasing distance as suggested by an increasing number of observations within higher distance intervals as well as large sums of absolute differences.

3.4.3. Generalised linear modelling

3.4.3.1. Log-linear modelling

For each of the nine successive surveys, elephant abundance is modelled with respect to the 12 explanatory variables (Table 3.1). We consider significance at the 5 % level and present the results as a matrix showing only significant variables in each model representing observations in one survey. The variables Constat, Distreserve, distance to road (Distroad) and average soil fertility (Avferti), are significant in six out of nine models and hence, may be useful in explaining elephant abundance (Table 3.4). Other significant variables include Distrive, Tallgr and Agrofali, all appearing in four models each. Also, forest cover (Forest) and short grassland (Shotgr) are significant in three models.

Coefficients for Distreserve have a negative sign in all models indicating a reduction in elephant counts with an increase in distance from MMNR. Similarly, the negative coefficients of Agrofali suggest that an increase in agriculture or fallow land lead to a decrease in elephant abundance. This is consistent with earlier results from the correlation analysis. The year to year variation in model composition may be attributed to differences in the numbers of observed elephant caused by other factors that were not studied here like periods of drought, seasonal changes, poaching etc.

3.4.3.2. Principal component analysis

From the 12 original explanatory variables, PCA yields five PCs, which collectively explain 94% of the total variance (Table 3.5). Based on loadings, contributions to each of the PCs are as follows:

- PC_1 is related to both Agrofali and Distreserve with a negative sign.
- PC_2 is related, also with a negative sign, to Shrubbyg and Distreserve, suggesting a positive relation between these two variables.
- PC_3 is similar to Shortgr, which yields the only large loading.

- PC₄ is related inversely to both Avferti and Tallgr.
- Finally, PC₅ is related to Distreserve and Avferti with a negative and positive sign, respectively, indicating an inverse relation between average fertility and distance to MMNR.

Other variables do not significantly contribute to the first five PCs.

Table 3.5. Five principal components derived from the 12 original explanatory variables and their corresponding scores. The largest scores for each principal component are highlighted in bold.

Principal Components and the corresponding Scores					
Variable	PC ₁	PC ₂	PC ₃	PC ₄	PC ₅
Eigen Values (%)	35.515	56.806	75.734	86.619	93.897
Constat	-0.0044	-0.0102	-0.0043	0.0039	-0.0003
Distreserve	-0.5065	-0.5735	0.1511	0.0544	-0.5583
Dstrive	-0.0369	-0.1251	0.1066	-0.0361	-0.2214
Distroad	-0.0054	-0.0443	0.0263	-0.0074	-0.0415
Avferti	-0.1538	-0.3354	0.4455	-0.6155	0.5133
Agrofall	-0.6736	0.1885	-0.1332	0.2868	0.2781
Dwshgra	-0.0839	-0.0135	-0.0832	-0.0546	0.2275
Forest	0.0190	0.0515	-0.0058	-0.0454	-0.0472
Shortgr	0.3396	-0.0345	0.7032	0.4292	-0.0387
Shrubbyg	0.3757	-0.5886	-0.4971	-0.0229	0.0575
Tallgr	0.0274	0.3936	0.0166	-0.5873	-0.4878
NDVI	0.0006	0.0006	-0.0001	0.0000	0.0003

A stepwise regression of the variables derived from the five PCs for survey 9703 results in the following model that relates elephant abundance (Y) to the PCs:

$$Y = 4.755 + 0.016PC_1 + 0.053PC_2 - 0.032PC_4 - 0.064PC_5. \quad (3.8)$$

This model is made up of PCs that are dominated by the variable Distreserve (PC₁, PC₂, and PC₅). The last two PCs (PC₄, and PC₃) are both influenced by Avferti, implying a possible effect of soil fertility on elephant abundance. The variables tall grassland, shrubby

grassland and agricultural or fallow land influence elephant abundance through PC₄, PC₂, and PC₁ in the model, respectively.

3.5. Discussion

From results of the log-linear modelling, the distance to MMNR turns out to be a strong variable in explaining elephant abundance. Its large, but negative contribution to PC₁ in the PCA reinforces the negative relation of elephant abundance to distance from MMNR. This relation is confirmed by PC₁ being significant in a stepwise modelling of abundance to derived variables (Table 3.5). A possible explanation of this relation lies in the fact that the three permanent rivers in the Mara ecosystem i.e. Mara, Talek and Sand rivers, all drain MMNR. Since elephant are water dependent, their distribution will be limited to areas close to permanent water sources like MMNR.

Said (1993) and Rosero (1997) observed a similar inverse relation between elephant abundance and the distance to MMNR. Rosero (1997) further reported the African elephant to be intermediate bulk feeders that are not very selective. They prefer grass to browse and are generally found in areas with high precipitation. This may explain why in all models, the elephant were predominantly observed on the western part of Mara. This is a region of flat plains with open and woody grasslands, which provide suitable habitat for the elephant.

The fluctuation of elephant abundance in Mara over the nine successive surveys may be related to poaching. In the 1980's severe poaching in Serengeti caused the elephant to migrate to Mara and seek refuge in the protected MMNR, thereby limiting their ranging patterns (Dublin and Douglas-Hamilton, 1987). This is supported by the consistent significance of Distreserve in the elephant abundance models.

Although conservation status is a significant variable in several models of GLM, its effect is suspect as its coefficient keeps changing signs from model to model. The fact that it fails to contribute significantly to any of the PCs in the PCA also supports this lack of significance. We, therefore, can not establish a significant effect of conservation status on elephant abundance in this study. This is surprising because this variable is highly correlated with the distance to MMNR, a protected area where many elephant are found. On the other hand, Mara is composed of a gazetted wildlife protected area (MMNR) plus adjacent group ranches that act as wildlife dispersal areas. The size of MMNR has, however, changed several times over the last 20 years with more land being de-gazetted to pave way for agricultural expansion and human settlements (GOK, 1984). This has reduced the elephant habitat and may explain why agricultural or fallow land is significant in several models, as well as contributing the largest negative loading to PC₁ in the PCA.

We do not observe a significant effect of the distance to road on elephant abundance in the PCA mainly because of the difficulty in detecting an effect at the coarse resolution of data collection (2.5×5 km). Similarly, although soil fertility in Mara ranges

from high to very low (Broten and Said, 1995), it has an insignificant effect on elephant abundance in this study. Its alternating signs in the log-linear models and its large loadings in the PCA only contribute to the less important fourth and fifth PCs.

The four animal species in this study have diverse ecological characteristics and social behaviour. The wildebeest and zebra are largely found in herds, elephant are solitary but also aggregate into medium herds, while the kongoni are mainly solitary. Further, the wildebeest and zebra are migratory, coming to Mara at the end of the rainy season. This study shows the two migratory species to have overlapping niches. Their dietary requirements, however, indicate differing trophic preferences, which results in their sequential arrival into Mara (Delany, 1982). The zebra, preferring tall and less nutritious grass, move in first, and their trampling and removal make grass suitable for consumption by the wildebeest, which prefer short grasses with a large proportion of leaf and a small amount of stem (Murray, 1995). This intensive grazing stimulates growth of new plants, which are preferred by the last of the migrating ungulates, i.e. the Thomson gazelle (*Gazella thomsoni*). The dietary requirements of wildebeest and zebra also lead them to congregate into large herds as observed in this study.

The vegetation in MMNR mainly consists of tall grasses like *Themeda triandra*, *setaria sphacelata* and *pennisetum mezianum* (Sinclair, 1975), which do not contain the high protein and soluble carbohydrates preferred by wildebeest. This may explain why this species migrates to the Siana and Loita plains, where forage has richer nutrient composition compared to MMNR. The apparent repulsion in habitat occupation between the resident kongoni and migratory species may be due to interspecies competition (Sinclair, 1995). However, further research is needed to confirm this hypothesis.

3.6. Concluding remarks

In this study, application of GIS allows the storage and analysis of wildlife data for both single species as well as for interaction between several species. The usefulness of GIS is highlighted by the ease with which we relate environmental data to both spatial and temporal changes in abundance. Therefore, GIS offers a useful framework for making precise statements about factors influencing wildlife variability.

Currently, a deficiency of standard GIS packages is the lack of elegant and systematic methods to simultaneously analyse spatial and temporal changes for wildlife data, the reason being that most GIS software can only describe and display geographic data, while lacking the ability to develop new hypotheses. Combining GIS and GLM provides an opportunity to model the effects of environmental factors on a quantitative response variable like animal counts, although not as a single step procedure. This lead to a better understanding of spatial and temporal relationships in wildlife data as illustrated in this paper. This study also demonstrates how taking advantage of spatial statistical routines, especially in a GIS environment, can extend standard statistical procedures.

An issue of continuing debate is the quality of data. In general, a lot of data exists on different aspects of wildlife dynamics, and if the scale at which it has been collected is fine enough, then conversion of different layers to a suitable resolution is not a major problem. For purposes of monitoring abundance and distribution of wildlife, these data can be considered to be sufficiently reliable as long as the method of collection is standardised over successive surveys. Any bias and/or imprecision is assumed to remain constant, thereby allowing inferences to be made on trends in abundance.

The data from DRSRS used in this study are fairly reliable, having been obtained through a consistent survey procedure. However, carrying out a more rigorous spatial analysis turns out to be far from easy. In particular, calculation of the variogram was unsatisfactory due to the structuring of data within transects. That is why we propose a measure that allows modelling of spatial dependence in a simple and straightforward way for wildlife data collected via transects. This is essential because once spatial distribution has been modelled, abundance and distribution can efficiently be monitored over time. Moreover, prediction of future changes becomes relevant when aspects of management are considered. For instance, the procedures we propose in this study may indicate at an early stage, where and when specific interventions have to be taken in terms of preserving available water and nutrient resources, or when faced with other threats like human encroachment.

Sampling is one of the main causes of poor abundance estimates (Caughley, 1974; Thompson, 1992). Recent developments in sampling techniques such as distance sampling (Buckland *et al.*, 1990) and adaptive sampling (Thomson and Seber, 1996) have been used to yield substantial improvements in estimation. With these techniques, using correct models of distribution may further help to improve sampling. For example, this study suggests that sampling efforts may be concentrated in sub-areas close to MMNR, where elephant are more likely to be found. More research is still being carried out to improve the sampling of dynamic populations.

This paper demonstrates how deductive-analytical spatial modelling in GIS can be used to identify environmental preferences for different animal species. The procedures outlined here are applicable in any situation where wildlife surveys are carried out and indeed for any large herbivore. Although a method-sensitivity was not carried out, the results obtained appear consistent with what is known about the species studied. The gist of this paper is in highlighting how combining statistical techniques with GIS provides a unified way of modelling animal distribution and abundance while quantifying the associated uncertainty in data from wildlife resorts.

Acknowledgements

The authors are grateful to Mr. Hesbon Mwendwa and Mr. Mohammed Said of the Department of Resource Surveys and Remote Sensing (DRSRS) for permission to use part of their dataset. Comments from two anonymous referees greatly improved the manuscript.

CHAPTER 4*

Modelling wildlife abundance using spatial statistics

* This chapter is based on Khaemba, W. M. and Stein, A. (*In press*). Spatial statistics for Modelling of Abundance and Distribution of Wildlife Species in the Masai Mara Ecosystem, Kenya. *Environmental and Ecological Statistics*.

Abstract

This study illustrates the use of modern statistical procedures for better wildlife management by addressing three key issues: determination of abundance, modelling of animal distributions and variability of diversity in space and time. Prior information is used in Markov Chain Monte Carlo (MCMC) methods to improve estimates of abundance. Measures of autocorrelation are included when modelling distributions of animal counts, and a diversity index to indicate species abundance and richness for large herbivores is developed. Data from the Masai Mara ecosystem are used to develop and demonstrate these procedures. The new abundance estimates are up to 35% more accurate than those obtained by existing methods. Significant temporal changes in spatial patterns are found from a space-time analysis of elephant counts over a 20-year period, with strong interactions over 5 km and 6 months space and time separations, respectively. The new diversity index is sensitive to both high abundance and species richness and is also able to capture year to year variation. It indicates an overall marginal decrease in diversity for large herbivores in Mara. The space-time analyses and diversity index can easily be computed thereby providing tools for rapid decision making.

KEYWORDS: Abundance estimation, distribution modelling, diversity measurement, Masai Mara ecosystem, MCMC methods, Space-time analysis.

4.1. Introduction

Correct and up-to-date information on wildlife provide necessary input in any decision support system for effective wildlife management. Such information may include species abundance, functional relationships between species distribution and environmental factors, or changes in biodiversity. It should be derived from data obtained by wildlife surveys conducted with optimal sampling techniques (Khaemba and Stein, *in press*). Moreover, influence of environmental factors on spatial and temporal distribution as well as abundance of wildlife population dynamics is as yet only partly understood for the Masai Mara Ecosystem (Mara) in Kenya.

Currently, wildlife abundance in East Africa is determined by using the Jolly II method (Jolly, 1969a), which interpolates sample characteristics like species density from data collected through systematic sampling along strip transects to larger management regions (Norton-Griffiths, 1978; Grunblatt, *et al.*, 1995). Spatial dependence within wildlife populations is ignored and abundance estimates often have wide confidence limits. Moreover, these methods do not take prior information into account although this has been known to improve sampling and interpolation (Stein, 1994), as well as abundance estimation in capture-recapture studies (Brooks, 1998). On the other hand, Markov Chain Monte Carlo (MCMC) methods have proved to be useful in medical studies (Spiegelhalter, *et al.*, 1996; Berzuini, 1996) and image analyses (Green, 1996), but have found little application in the analysis of strip transect data, particularly from wildlife surveys. For well-defined sampling units, data on wildlife and factors influencing their distribution can be collected easily and stored in geographic information systems (GIS). Such data can be subjected to spatial statistics to allow the interpolation and quantification of associated uncertainty (Cressie, 1993). In the recent past, several studies have investigated wildlife population dynamics using GIS (Walker, 1990; Pereira and Itami, 1991; Augustin, *et al.*, 1996; Li, *et al.*, 1997). Most of these studies, however, have only related environmental factors to presence and absence of wildlife without modelling the actual observed counts.

Conservation of wildlife diversity within an ecosystem requires an accurate determination of animal abundance as well as modelling changes in biodiversity over time. Diversity indices have been developed to capture a multidimensional concept in a single number, which is then used to give insight in both species abundance and richness for the wildlife in an ecosystem (Patil and Taillie, 1979; Magurran, 1988). Available indices, however, have been developed to be sensitive only to dominant or abundant species, and therefore, can not be used for large herbivores like the ones found in Mara, particularly under the framework of aerial surveys.

Therefore, objectives of this study are threefold:

- To investigate how prior information can improve estimates of wildlife population parameters.
- To use spatial statistics in a GIS context to model abundance and distribution and monitor distributional changes over time.

- On the basis of the two previous objectives, to develop a diversity index for large herbivore species.

We focus on four herbivore species: the elephant (*Loxodonta africana*), kongoni (*Alcelaphus buselaphus*), wildebeest (*Connochaetes taurinus*) and zebra (*Equus burchelli*). These have been chosen because of differences in abundance and observability as well as their social and migratory characteristics. Moreover, the wildebeest is considered a keystone species because its migratory characteristics influence many other components of the ecosystem, and its loss would cause a more than average change in other species population and dynamics (Sinclair, 1995b).

4.2. Materials and methods

4.2.1. Data

Data for this study come from multi-species surveys conducted in Mara over a 20-year period (1977-1997). During the surveys, high-winged Partenavia aircraft equipped with global navigation and positioning systems (GNS and GPS, respectively) are flown along systematic transects determined beforehand. Flying height varies between 70 m and 152 m, with 122 m being standard, and a speed of 190 km/h is maintained. Two rear seat observers count the number of animals falling within strips defined on either side of the aircraft. Common strip widths lie between 224 m and 304 m, and each strip is divided into equal subunits of 5 km in length to give distributional data. Ratio estimators are used to estimate animal abundance.

Let us partition a study region of size A and containing S animal species into N exhaustive and non-overlapping strips. Consider x_{ij} animals of species j occurring in the i th strip having an area of size a_i , $i = 1, 2, \dots, N$, $j = 1, 2, \dots, S$. We observe a sample of n strips in which y_{ij} animals of the x_{ij} are counted. The problem is to estimate the total number of animals of species j (τ_j) in the population based on sample observations.

Define a ratio R_j as:

$$R_j = \frac{\tau_j}{A} = \frac{\sum_{i=1}^N x_{ij}}{\sum_{i=1}^N a_i} = \frac{\sum_{i=1}^N x_{ij} / N}{\sum_{i=1}^N a_i / N} = \frac{\mu_j}{\mu_A}, \quad (4.1)$$

where μ_j and μ_A are population means for true animal numbers of species j and areas of all strips, respectively. This leads to the relationship:

$$\mu_j = R_j \mu_A, \quad (4.2)$$

where μ_j , μ_A and R_j are population parameters. The term R_j is estimated from sample values by:

$$\hat{R}_j = \frac{\sum_{i=1}^n y_{ij}}{\sum_{i=1}^n a_i} = \frac{\bar{y}_j}{\bar{a}}, \quad (4.3)$$

where \bar{y}_j and \bar{a} are sample means for counts of animals from species j and areas of n observed strips, respectively. We use MCMC modelling to determine \hat{R}_j .

4.2.2. MCMC modelling

MCMC methods are derived from Monte Carlo integration using Markov chains. They allow integration over high-dimensional probability distributions like the *posterior* distributions found in Bayesian analyses (Gelman, *et al.*, 1996; Gilks, *et al.*, 1996). Through algorithms like the Metropolis-Hastings or Gibbs Sampler, MCMC methods allow direct determination of population quantities like means or totals and corresponding standard errors from sample values (Geman and Geman, 1984; Gilks, *et al.*, 1996; Brooks, 1998). Moreover, 100(1-2 q) credible intervals [c_q , c_{1-q}] are estimated for these parameters by setting c_q and c_{1-q} equal to the q^{th} and $(1-q)^{\text{th}}$ quantiles of the ordered statistic, respectively (Gilks, *et al.* 1996).

In this study, we let Y_j denote counts of animal of the j th species with y_{ij} observed in the i th strip. We make the following structural assumptions:

- y_{ij} are independent conditional on the expectation μ_j and a sampling error parameter σ^2_j . In particular, y_{ij} is distributed as negative binomial with parameters p_j and k_j , which represent the probability of observing an animal and the number of successes, respectively. Both p_j and k_j are species dependent.
- μ_j is a deterministic function depending only on the expectation μ_A of strip areas of size a_i and a parameter R_j representing the population density for the j th species.
- Parameters R_j are drawn independently from a distribution parameterised by parameters μ_{R_j} and $\sigma^2_{R_j}$, which are referred to as hyperparameters. In particular $R_j \sim N(\mu_{R_j}, \sigma^2_{R_j})$.

We require prior distributions for p_j , k_j , μ_{R_j} and $\sigma^2_{R_j}$ to complete specification of the full probability model for each animal species. For these, we adopt priors that lead to standard forms of the conditional distributions, thereby allowing the use of Gibbs sampling. It is common to write sampling error parameters like $\sigma^2_{R_j}$ in terms of precision, defined as the reciprocal of the variance ($\sigma^{-2}_{R_j}$). With \hat{R}_j thus obtained, an estimate of τ_j follows for various species through the relation $\hat{\tau}_j = \hat{R}_j A$. We use data obtained from survey 9604 for modelling abundance and its precision for the four animal species. We also obtain 95%

credible intervals by setting the lower and upper limits to the 2.5% and 97.5% quantiles, respectively.

4.2.3. Space - Time modelling

4.2.3.1. Analysis

In this section, animal occurrence at different locations are considered as events in a space-time setting, where co-ordinates of subunits and a simple date provide spatial and temporal attributes. Proximity between events is judged by inter-event differences. For this analysis, we focus on elephant, which is both solitary and gregarious. We start by testing the null hypothesis H_0 that an equal proportion of events occurs in space at each survey. Data from all the surveys in Mara are arranged in a two-way table where columns represent repeated observations over time while rows represent locations. We only consider subunits having one elephant or more for the analysis.

Let E denote an event. Define $SS_c = \sum_c (\sum_r E)^2$ as the sum of squared column totals, $SS_r = \sum_r (\sum_c E)^2$ as the sum of squared row totals and $GT = \sum_r \sum_c E$ as the grand total of the table frequencies). The subscript on the summation indicates a sum over that particular term, and c and r correspond to the number of surveys and subunits where at least one elephant was observed, respectively. We define the Q -statistic as:

$$Q = \frac{(c-1)[c \sum_c (\sum_r E)^2 - (\sum_r \sum_c E)^2]}{c \sum_r \sum_c E - \sum_r (\sum_c E)^2} = \frac{(c-1)[cSS_c - GT^2]}{cGT - SS_r}, \quad (4.4)$$

which follows a χ^2 distribution with $c-1$ degrees of freedom when H_0 is true (Sokal and Rohlf, 1995, p.783). We compute pairwise differences and use Q to test for space-time independence between events. We use the Euclidean distance d_{uv} between events at locations l_v and l_u in space and define a (decreasing) time function T_{st} between events at times t_s and t_r as:

$$T_{st} = \frac{1}{e + |t_r - t_s|}, \quad (4.5)$$

where e (>0) is a small term added to avoid division by zero.

To classify events as close or distant in both space and time, we compare several threshold values D and T and use them to crosstabulate d_{uv} and T_{st} values into 2×2 tables. This allows use of standard contingency table analysis techniques. We test the null hypothesis H_0 of no significant patterns in space over time in the tables. This is equivalent to testing for independence between space and time.

From examination of the data, it is observed that events are likely to be correlated because some observations are made repeatedly at the same location over time. This is

taken into account by using the McNemar's test (Agresti, 1990) to test for H_0 . If we assume a b_{11} , b_{12} , b_{21} , and b_{22} configuration for the tables, with the main diagonal (b_{11} and b_{22}) representing close and distant events in time and space, respectively, a test statistic can be defined by:

$$z_0^2 = \frac{(b_{21} - b_{12})^2}{b_{21} + b_{12}}. \quad (4.6)$$

This follows a χ^2 distribution with one degree of freedom under H_0 .

4.2.3.2. Modelling with GIS

For GIS modelling, data are initially summarised to assess distribution variability and detect extreme values in an exploratory data analysis (EDA) (Tukey, 1977). A spatial EDA uses point maps of observed animal counts per subunit to investigate their spatial distribution. Further spatial exploration concerns the study of spatial variability using autocorrelation to model influences of neighbouring units. These methods serve to identify patterns in species distribution and to establish the validity of assumptions prior to modelling.

Spatial general linear modelling is applied to relate animal counts as a response variable to a set of environmental conditions. The Poisson distribution is used in log-linear modelling, after correcting for overdispersion. Consider the mixed regressive-spatial-autoregressive specification defined by Anselin (1993) as:

$$y = X\beta + \rho Wy + \varepsilon. \quad (4.7)$$

Here, y is a vector of counts, Wy transforms y through a square matrix W of spatial lags and ρ is a spatial autoregressive coefficient measuring spatial autocorrelation. The matrix X contains observations on k explanatory variables, β is a vector of regression coefficients and ε is a vector of error terms.

Neighbourhood memberships for each observation pair (y_i, y_j ; $i, j = 1, 2, \dots, n$) is expressed by the matrix W such that for a binary form of W , the element w_{ij} takes the value 1 when observations i and j are neighbours and zero otherwise. For other non-binary forms of W more complex forms apply. Including Wy in the model allows assessment of the degree of spatial dependence while controlling for effects of other explanatory variables. A model containing only y and Wy results in a purely auto-regressive model, useful for estimating ρ .

Ordinary least squares (OLS) is the most common parameter estimation method in regression analyses, yielding optimal estimates if appropriate assumptions are satisfied. In this model, however, Wy is correlated with ε , leading to inferior estimation by OLS. We therefore use generalised least squares (GLS) under assumption of the normal distribution after estimating the correlation between Wy and ε through residual analysis.

4.2.4. Diversity measurement

Most diversity indices characterise biodiversity within an ecosystem by a single digit. Diversity, however, consists of two components: species richness and evenness, which measure the variety and relative abundance of species, respectively. Therefore, differences between diversity indices lie in the relative weighting given to these two components.

In this study, we consider Shanon's index H' defined as $H' = -\sum p_j \ln p_j$ and Simpson D_{Sim} defined as $D_{Sim} = \sum_{j=1}^S \frac{y_{.j}(y_{.j} - 1)}{y_{..}(y_{..} - 1)}$. The quantity $\sum_{i=1}^n y_{ij} = y_{.j}$ is the number of individuals of the j th species, $\sum_{j=1}^S \sum_{i=1}^n y_{ij} = y_{..}$ is the total number of individuals for all species, $p_j = \frac{y_{.j}}{y_{..}}$, the proportion of the j th species in the ecosystem and \ln denotes the natural logarithm. Both indices are based on proportional species abundance. For further comparison, we also consider the Berger-Parker index ($B-P$), which is defined as $\frac{\max(y_{.j})}{y_{..}}$ and is based on absolute abundance. It has a simple formulation but is rarely used (Magurran, 1988).

Indices H' and D_{Sim} form a basis for formulating a new diversity index K , useful when the only available wildlife data come from aerial surveys. By assuming that all individuals in a community are sampled randomly from an infinitely large population, from which all species are sampled, we can apply the intraspecific encounter theory and define a dichotomous type diversity index (Patil and Taillie, 1979) as:

$$K = \sum_{j=1}^S \frac{y_{.j}(y_{.j} + 1)}{y_{..}(y_{..} + 1)}. \quad (4.8)$$

A single species community ($S = 1$) leads to $K = 1$ and as S increases, K decreases implying an increase in diversity with decreasing values of K . This suggests using its reciprocal as a more suitable diversity measure. The evenness K_E of K is calculated as:

$$K_E = \ln \left(\frac{S}{K} \right). \quad (4.9)$$

Performance of the four indices on the present data is compared through descriptive statistics and a trend analysis.

4.3. Results and discussion

4.3.1. Survey data

Data for various analyses in this study come from 21 surveys conducted in Mara between 1977 and 1997. Such surveys are typically multi-species in nature. An example is survey 9604, 96 being the year of survey and the identification (ID) number 04 indicating this to be the fourth survey of that year. This survey was carried out in the month of August and 55 transects were flown at a strip width of 282 m leading to a sampling fraction of 11.38%. Wildebeest and zebra were observed in 74.5% and 80% of the subunits, respectively, whereas elephant and kongoni were observed only in 21.8% and 45.5% of the subunits, respectively. Distribution maps of these species (Figure 4.1) confirm the high abundance of wildebeest and zebra. The elephant appears to prefer the Masai Mara National Reserve (MMNR), while kongoni were mainly observed in the middle section of the study region. High inter-strip variability is observed for wildebeest and zebra, with a range in counts of 2,105 and 852 animals for the two species, respectively.

4.3.2. MCMC modelling

MCMC modelling is carried out using the BUGS software, (Spielgelhalter, *et al.*, 1996). For animals of species j , the following non-informative priors for hyperparameters p_j , k_j , μ_j and $\sigma^{-2}_{R_j}$ are chosen from proper probability distributions with small precision in order to have minimal effect on the analysis (Gelman, *et al.*, 1995). The symbol \sim means 'distributed as' leading to the following.

- $p_j \sim \text{beta}(1, 1)$, denoting a beta distribution with shape 1 and shape 2 parameters both equal to 1,
- $k_j \sim \text{unif}(1, 1,000)$, denoting a uniform distribution with 1 and 1,000 as minimum and maximum values,
- $\mu_j \sim N(0, 10^6)$, denoting a normal distribution with mean 0 and variance 10^6 ,
- $\sigma^{-2}_{R_j} \sim \text{Ga}(0.001, 1,000)$, denoting a gamma distribution with shape parameter 0.001 and scale parameter 1,000.

We use the Gibbs sampler to sample iteratively from conditional distributions starting with 1,000 iterations and followed by another 10,000 iterations.

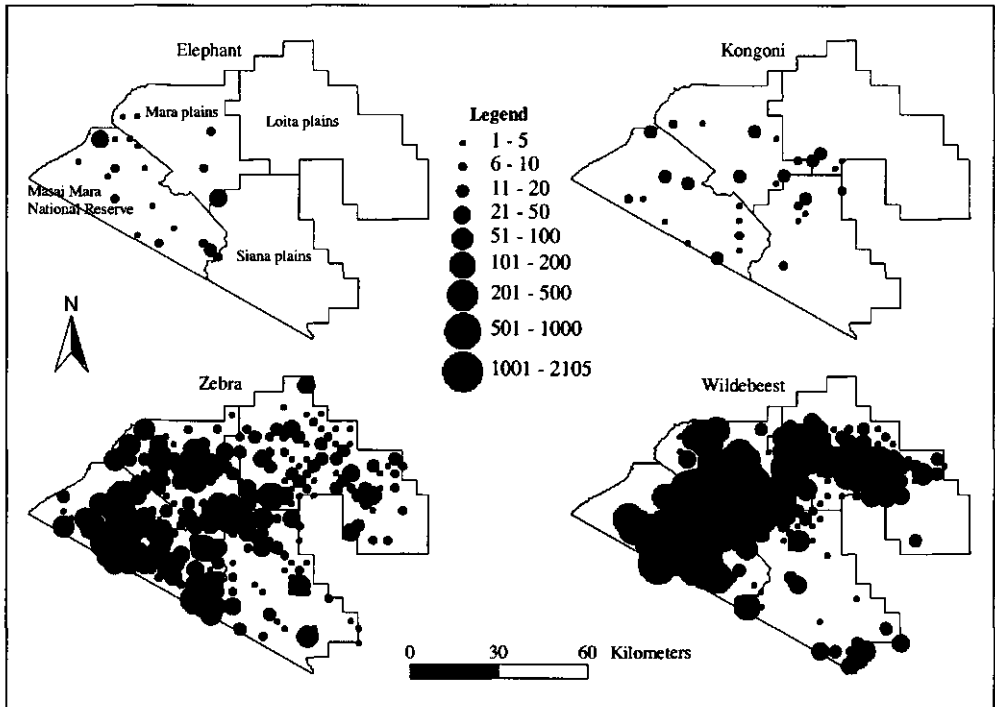


Figure 4.1. Point maps showing the distribution of elephant, kongoni, zebra and wildebeest as observed in Mara during survey 9604.

Population parameters for the four animal species ($S = 4$) are calculated using average values from these iterations. Population total estimates by MCMC modelling do not appear to differ much from those obtained by the conventional Jolly II method (Table 4.1). The standard errors (se) estimated by MCMC are, however, much lower. For example, the MCMC elephant total estimate of 1,633 with se equal to 360, is 1.5 times more accurate than the estimate of 1,626 with se equal to 545 obtained by the Jolly II method. Further, the accompanying 95% credible intervals [1062, 2253] obtained by MCMC for the elephant are much narrower than the corresponding 95% confidence intervals [558, 2695] given by the Jolly II method (Table 4.1).

This increase in accuracy is attributable to the use of prior information in estimation, something that the Jolly II method does not do. Indeed the main criticism of the Jolly II method lies in its large standard errors, which are a direct consequence of differences in the size of sampling units and observed counts, especially for gregarious species like wildebeest and zebra. These differences invalidate the assumption of a through-the-origin regression and proportional variance, which are necessary assumptions for ratio estimators (Thompson, 1992).

Table 4.1. Descriptive statistics based on transect counts of the four animal species, results of MCMC modelling of the same data and corresponding results from the Jolly II method for comparison.

STATISTIC	VARIABLE				
	Area size	Elephant	Kongoni	Wildebeest	Zebra
Minimum	0.71	0	0	0	0
Median	1.41	0	0	0	0
Mean	1.43	0.35	0.45	65.3	14.69
Maximum	1.83	45	20	2,105	852
Standard deviation	0.1	2.64	2.21	221.82	50.59
Autocorrelation (ρ)		0.461	0.194	0.013	0.019
MCMC MODELLING RESULTS					
Density estimate (\hat{R}_j)		0.25	0.32	46.02	13.07
Standard error (\hat{R}_j)		0.05	0.07	6.89	2.62
Estimate of population total ($\hat{\tau}_j$)		1633	2128	306033	86915
Standard error of $\hat{\tau}_j$		360	466	45818	17396
95% credible interval		[1062, 2253]	[1384, 2936]	[259842, 353042]	[73513, 100265]
JOLLY II METHOD RESULTS					
Estimate of population total ($\hat{\tau}_j$)		1626	2065	299357	87067
Standard error of $\hat{\tau}_j$		545	487	52541	18746
95% Confidence interval		[558, 2695]	[1110, 3019]	[196357, 402338]	[50325, 123809]

4.3.3. Space-time modelling

4.3.3.1. Analysis

Large and irregular variations are observed among the 21 surveys in Mara. The null hypothesis of an equal proportion of events per location over surveys is tested first. In total, at least one elephant was observed in 237 subunits from 21 surveys (time) and 121 locations (space). When arranged in a two-way table, the number of events per location ranges from 1 to 7, while the number of locations with an event ranges from 3 to 35. Column totals indicate a tendency for the number of events to increase with time (4 in 1977

and 20 in 1997). Sums of squares for rows SS_R and columns SS_C equal 651 and 3,927, respectively, while the grand total GT equals 237. These yield a Q value equal to 121.58, which at 20 degrees of freedoms, is highly significant ($p < 0.001$). This leads to rejection of the null hypothesis and we conclude that the proportion of events increases over time.

To compute W_{ij} for each time difference, we set the term e equal to $2/30 \approx 0.1$, based on duration of the shortest survey (converted to months). Several threshold values of T are used: from 120 months, i.e. half the time between 1977 and 1997 ($T = 1 / (0.1 + 120) = 0.008$), through 60, 36, 12 months to 6 months ($T = 1 / (0.1 + 6) = 0.2$). By considering half the range of the study region, D was similarly determined at values of 50 km, 25 km, 15 km, 10 km and 5 km. The 27,966 possible pairs resulting from the 237 events were classified using these different D and T values.

For $D = 50$ km and $T = 120$ months (Table 4.2), the sample odds ratio (OR) equals 1.11, indicating little association between space and time ($p = 0.08$). Sample proportions p_{1+} and p_{+1} , which indicate closeness in space and time, equal 0.836 and 0.849 for contemporaneous and close events in space, respectively. Similarity of these two values support the weak association indicated by the sample OR. McNemar's z_0 statistic equals 0.83, and with one df gives a p -value of 0.362, suggesting independence between space and time. However, for $T = 6$ months and $D = 5$ km, OR reduces to 0.69 while z_0 increases to 1,410 ($p < 0.001$), leading to rejection of the null hypothesis, a result in agreement with that found by the Q -statistic. We, therefore, conclude that space-time interaction occur for short time (< 6 months) and distance separations (< 5 Km), but reduces with increasing separation in space and time.

4.3.3.2. Modelling with GIS

Descriptive statistics show heavily skewed distributions for each species (Table 4.1). In all cases, the standard deviation is much larger than the mean, pointing to overdispersion and invalidating the use of a Poisson distribution to model these counts. Low pairwise correlation for the observed animal counts (from -0.14 to 0.19) indicate that the four species analysed do not encourage or constrain the presence of one another. This lack of a strong linear relationship is illustrated in the scatter plot matrix (Figure 4.2), which shows animal distributions influenced by a high occurrence of zero counts. A referee pointed out however, that no interaction is expected at the given scale of observation except for highly abundant species like the zebra. Corresponding box-plots and stem-and-leaf plots (not shown) reveal extreme values for wildebeest and zebra but not for elephant and kongoni.

Spatial autocorrelation is estimated using the pure auto-regressive model. This yields positive values of ρ indicating positive spatial autocorrelation for all four species (Table 4.1). Low values of ρ for both wildebeest and zebra emerge from their wide distribution (Figure 4.1). Units with high counts often surrounded units having low counts, thereby decreasing spatial dependence for these species.

Next, twelve explanatory variables (Table 3.1) are used together with the spatially-lagged W_y variable, to determine the best model for animal counts of the four species in a stepwise regression procedure (Table 4.4). In all models, a positive coefficient ρ is obtained, supporting the hypothesis of spatial dependence. The coefficient does not change much, even in the presence of information from explanatory variables, indicating a stable fit. Further, most other explanatory variables become insignificant after including the spatial variable in the analysis, confirming the dominance of spatial autocorrelation in explaining the distribution of these four animal species.

Table 4.2. Classification of the 237 events observed in Mara through all possible pairs (27,966) into a 2x2 table using two threshold values, $D = 50$ km and $T = 120$ months. The term d_{uv} denotes a Euclidean measure between events at locations l_u and l_v , while T_{st} denotes a time function between events taking place at time t_s and t_t .

		SPACE		Row totals
		$d_{uv} < D$ (Close together)	$d_{uv} > D$ (Far apart)	
TIME	$T_{st} > T$ (Contemporaneous)	19,687	3,704	23,391
	$T_{st} < T$ (Time gap)	3,783	792	4,575
Column totals		23,470	4,496	27,966

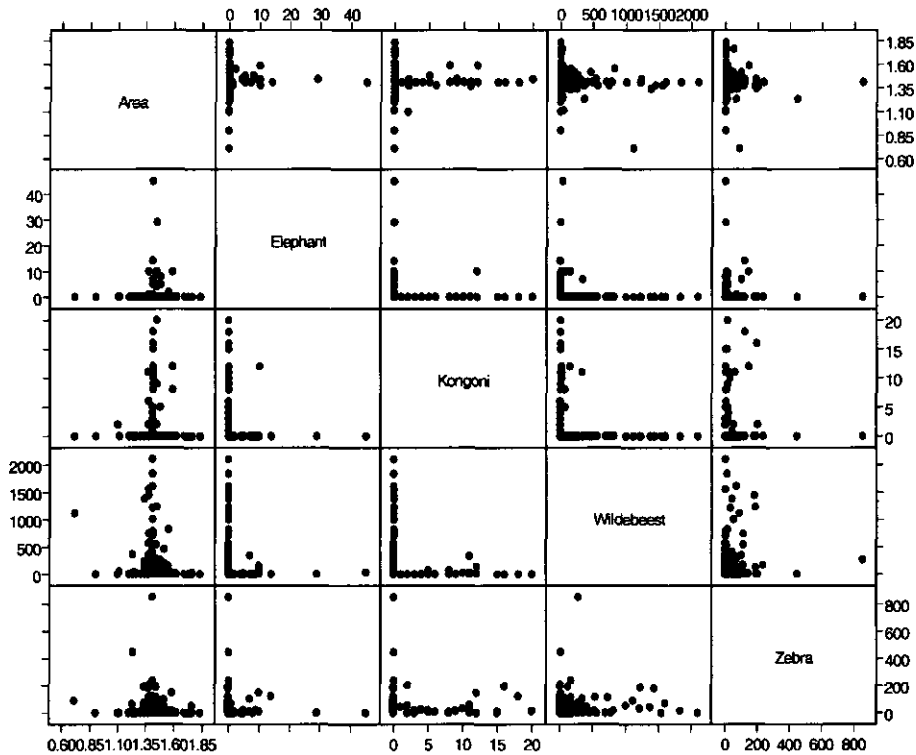


Figure 4.2. Matrix of scatter plots showing lack of linear relationships between counts of the four animal species and sizes of observed areas.

4.3.4. Diversity measurement

The four diversity indices described in section 4.2.4 are computed by considering all herbivore species surveyed in Mara since 1977. During these surveys, the number of observed species ranged from least abundant (1) to highly abundant (33,058). For illustration, we only report indices of six surveys (Table 4.4). We start by comparing the indices in an exploratory analysis. Summary statistics show little variation, suggesting a stable level of diversity of large herbivores in Mara during the last 20 twenty years (Table 4.5). Broten and Said (1995) obtained similar results when studying the population of ungulates in the same region.

Table 4.3. Significant coefficients, at 5% level, for explanatory variables of auto-regressive models for elephant, kongoni, wildebeest and zebra.

Variable	SPECIES			
	Elephant	Kongoni	Wildebeest	Zebra
Mean	0.982	-1.923	3.062	2.881
Spatial variable (W_y)	0.771	0.940	0.275	0.039
Average Soil fertility	-0.089			
Distance to river		-0.078		
Distance to road		0.317		
Conservation status			-0.349	-0.518
NDVI			-2.412	

Table 4.4. Four diversity indices calculated for six surveys conducted in Mara between 1977 and 1997. H' denotes Shanon's index, D_{Sim} represents Simpson's index, $B-P$ stands for the Berger-Parker index while K is the new index developed in this study.

	SURVEY YEAR AND ID					
	7701	7706	8015	8706	9003	9703
Number of species (S)	21	22	20	20	20	20
Number of individuals (y_i)	14,617	29,052	53,545	28,298	45,210	31,368
DIVERSITY INDICES						
H'	0.69	0.66	0.47	0.71	0.55	0.62
D_{Sim}	5.81	5.71	2.44	6.56	3.58	4.86
$B-P$	3.11	3.33	1.62	3.82	2.15	3.21
K	4.80	4.83	3.89	4.88	4.27	4.58

Pairwise correlation between indices show a high linear relation between the new K index and the two most common indices H' and D_{sim} ($r = 0.96$ and $r = 0.98$, respectively), indicating that K contains joint qualities of the other two. Correlation with $B-P$ is also high ($r = 0.92$). Index K shows no relationship with the number of species S , but similar to other indices, has a negative linear relation with the total number of individuals. A possible ecological interpretation of this observation is that larger communities are likely to be more diverse than small ones.

The number of observed species appears to remain steady, although abundance fluctuates over time. When plotted against time, values of all indices fluctuate around a slight downward linear trend (4.3). These results are consistent because species count represents one way of measuring diversity.

4.4. Concluding remarks

This study shows how use of prior information in MCMC modelling can improve accuracy of estimates of population totals obtained from aerial wildlife surveys. By relating animal counts to strip areas in a generalised linear model, density parameters are well estimated, leading to a gain in precision of up to 35 %.

Spatial statistics and GIS are useful to detect relations between the many variables measured during wildlife surveys. In this study we observe how spatial autocorrelation of four animal species in Mara increases for data close together in both time and space. The diversity index developed here shows year to year variation in diversity for large herbivore species, while capturing a marginal long-term decline. This result is consistent with the findings of De Leeuw *et al.* (1998), who observed a downward trend in both livestock and wildlife from 19 Kenyan districts over a 20-year period (1977-1997).

Data used in this study are not corrected for bias because of the tendency of bias correction methods to be highly site and survey specific, thereby lacking generalisation. To minimise bias, however, many survey agencies incorporate correction procedures in the sampling strategy like the use of aerial photography for animals found in herds of more than 10 (Grunblatt *et al.*, 1995; Wint, 1998). Moreover, most aerial surveys are used for long term monitoring of wildlife, and thus emphasise consistency in the survey methods. All the same, De Leeuw *et al.* (1998) found the DRSRS data to be of good quality in their study on the reliability and consistency of the data.

For a manager of a wildlife reserve, this study provides techniques that can be used to answer three relevant basic questions:

- How many animals are in my reserve?
- Where are they?
- Are the numbers changing with time?

Any efforts at management and conservation must address these questions. The space-time analysis and diversity measure developed in this study can easily be computed without need for specialised hardware and software, thereby providing tools for rapid decision making for such managers.

In future, estimation can be improved further by including other auxiliary information e.g. from data on other explanatory variables. Similarly, use of other algorithms apart from the Gibbs sampler should be investigated in future research of MCMC methods for wildlife surveys data.

Table 4.5. Summary statistics and pairwise correlation calculated from diversity indices defined in table 4.4. The statistics are based on computation of these indices for 21 surveys conducted in Mara from 1977 to 1997. *S* and *N* denote the number of species and individuals, respectively.

STATISTIC	DIVERSITY INDEX					
	<i>S</i>	<i>N</i>	<i>H'</i>	<i>D_{Sim}</i>	<i>B-P</i>	<i>K</i>
Minimum	19	14,617	0.41	2.13	1.49	3.89
Mean	20.77	33,080.73	0.64	5.33	3.25	4.66
Maximum	24	64,053	0.74	7.29	5.14	5.06
Std. Dev.	1.01	11,015.02	0.08	1.37	0.87	0.31
PAIRWISE CORRELATION						
<i>S</i>	1	0.17	-0.13	-0.01	0.02	0.09
<i>N</i>		1	-0.81	-0.74	-0.65	-0.77
<i>H'</i>			1	0.97	0.87	0.96
<i>D_{Sim}</i>				1	0.94	0.98
<i>B-P</i>					1	0.92
<i>K</i>						1

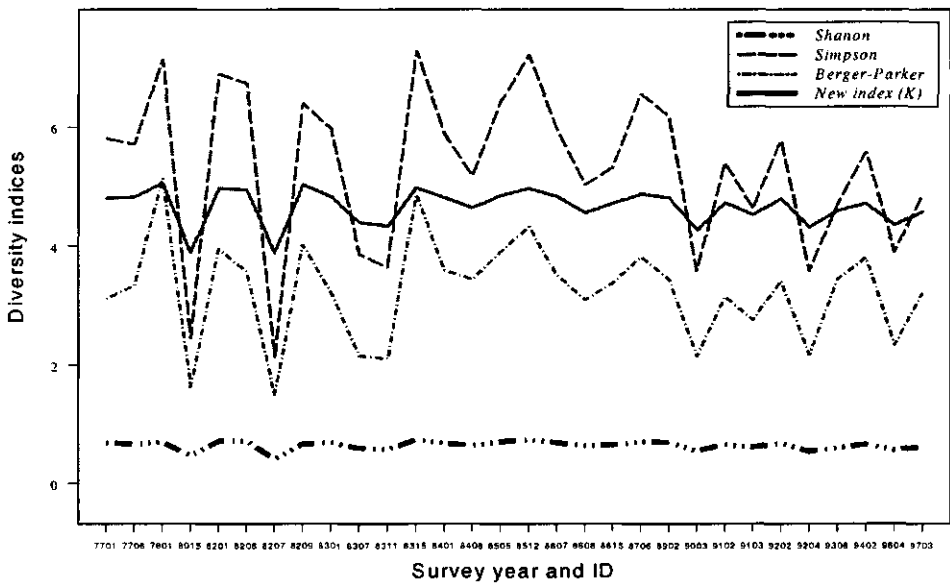


Figure 4.3. Chart showing calculated values of the four diversity indices given in table 4.4 for data obtained from the 21 surveys conducted in Mara between 1977 and 1997. ID denotes the position among the surveys carried out that year. The chart is used to analyse trend in diversity for large herbivore species in this ecosystem.

Acknowledgements

The authors are grateful to the department of resource survey and remote sensing (DRSRS) of the ministry of planning and national development in Kenya, for permission to use their data. Comments from three anonymous referees greatly improved the clarity and content of this paper.

CHAPTER 5*

Spatial point patterns for modelling wildlife distribution

* This chapter is based on Khaemba, W. M. Said, M. Y., Ottichilo, W. K. and Stein, A. (*In review*). Spatial point pattern analysis of total count data from aerial surveys of wildlife. *Journal of Agricultural, Biological and Ecological Statistics (JABES)*.

Abstract

Most wildlife populations exhibit clustering in their distributions. Assessment of clustering is important in identifying individual wildlife species by analysing their trophic requirements, which are related to the species spatial distribution. This assessment is, however, made difficult by the type of data obtained from aerial surveys. The shape and size of strip transects results in poor data support. GPS technology offers opportunities to record exact locations of observed animal groups and thereby obtain data that is amenable to spatial analysis. We use spatial point pattern to analyse such detailed data and explain observed patterns using environmental factors like vegetation type. This provides both an explanation of the distribution of animal species and differentiation between various species. Nearest neighbour distance measures like the *G*-statistic and *K*-function classify observed spatial patterns as clustered, regular or completely random. Independence between species is tested using a multivariate extension of the *K*-function. The techniques are illustrated with field data on six ungulates observed in an ecosystem in Kenya. Results indicate that spatial distribution is related to species dietary requirements. We conclude that spatial point pattern analysis is useful in determining and confirming species spatial distribution patterns.

KEYWORDS: Complete spatial randomness, *K*-function, Nearest neighbour distances, Spatial points pattern.

5.1. Introduction

Aerial surveys of wildlife in African nature reserves distinguish between total counts (TCs) and sample counts (SCs) depending on whether the whole or part of a study region is surveyed, respectively. These surveys are routinely used to collect data on wildlife dynamics (Caughley, 1977; Norton-Griffiths, 1978; Krebs, 1989). Analysis of data mainly concentrates on changes in animal populations and the results are used to monitor trends in abundance and map species distribution, sometimes in relation to environmental factors like vegetation types (Grunblatt, *et al.*, 1996).

A complete spatial analysis of the data from aerial surveys of wildlife, however, is hampered by poor data support. The problem with the data lies in the size and shape of strip transects, which are common sampling units in aerial surveys (Khaemba and Stein, *In press*). Equally sized subunits, created by regular subdivisions along transects, are often large compared to the size of individual animals or groups, thereby creating weak relation between area and position. This makes it difficult to classify data from SCs of wildlife into any of the three common spatial data categories i.e. geostatistical, lattice or spatial point patterns.

A possible solution is to collect data that are more suited to spatial analysis. Recent widespread use of global positioning systems (GPS) allows recording of observations of wildlife at individual geographic locations during aerial surveys (Wint, 1998). These yield point data that can be used to model stochastic processes that generate observed spatial patterns. Such patterns represent locations where animals are observed and can be considered as spatial point patterns. When the numbers of animals are attached to each location, they become marked point patterns (Ripley, 1981; Diggle, 1983; Cressie, 1993). Such data allow use of spatial point pattern analysis to classify observed point patterns, model spatial distribution and find explanations for it.

Various hypotheses have been put forward to explain variations in animal abundance, grouping and distribution for different species. Some authors have related these differences to physiological process i.e. metabolic mass and feeding style, competition and facilitation (Prins and Olff, 1998). Jaman (1974) and Estes (1974) have hypothesised that African antelopes with large body size form larger groups than smaller ungulates, and that grazers live in larger gatherings than mixed feeders. Voeten (1999) indicates that abundance and distribution of animals may be related to its food requirements and the distribution of this food.

Walker (1990), Buckland and Elston (1993), Augustin, Mugglestone and Buckland (1996) and Li, *et al.* (1997) have all reported carrying out some spatial analysis on wildlife data. A general overview on geostatistical tools useful in ecological modelling is found in Rossi (1992), while Steffens (1993) reports further applications of geostatistics on aerial wildlife data. None of these analyses, however, uses spatial point pattern analysis.

In this study, we explore the use of spatial point pattern techniques in analysing wildlife data collected from a detailed aerial survey of the Laikipia district of Kenya. The

data were collected with the help of GPS technology and represent point measurements. We focus on six ungulates that are ecologically important for this ecosystem namely: Burchell's zebra (*Equus burchelli*), Grant's gazelle (*Gazelle granti*), Thomson's gazelle (*Gazelle thomsoni*), impala (*Aepyceros melampus*), kongoni (*Alcelaphus buselaphus buselaphus cokii*) and Grevy's zebra (*Equus grevyi*). These species can be grouped into grazers (Burchell's zebra, Thomson's gazelle, kongoni and Grevy's zebra) and mixed feeders (Grant's gazelle and impala) by their dietary requirements. Further, Burchell's zebra is the most abundant species, representing more than half the total population of large herbivores in this ecosystem (Georgiadis, 1997). We, therefore, compare its spatial point pattern to those of the other five species in an effort to find comparative differences or similarities that can be taken into considerations during multi-species surveys. Understanding the spatial distribution of these species will help develop better sampling designs for their survey.

Further, we hypothesise that observed spatial patterns are related to both animal behaviour and environmental conditions. The main objective of this study is to provide methods and procedures to analyse point data from wildlife surveys. This is relevant because many wildlife species are often classified as having clustered distributions without any formal testing of their spatial point patterns.

Although this study involves only a few wildlife species from a TC of a single ecosystem, the methods and procedures are applicable to wildlife species in other ecosystems where GPS has been used to record exact locations of observed species. Trophic differences allow generalisation of results albeit for grazers and mixed feeders.

5.2. Materials

Data in this study were collected from a three-day TC of Laikipia district carried out in September 1996. The study region was divided into 3 sections, each sub-divided into blocks of 200 to 500 km² (Georgiadis, 1997). An aircraft systematically searched a block by flying at transects separated by a 1 km interval (Figure 5.1). Ten aircraft were used simultaneously to minimise survey time and reduce double counting caused by animal movements. Each aircraft carried a crew of pilot, front seat observer (FSO) and two rear seat observers (RSOs), and was assigned a fixed portion of the study region to survey per day.

Whenever an animal (or group) was spotted, the aircraft deviated from its flight-line and circled the observed animals until their exact number and geographical co-ordinates were counted and recorded by the RSOs. Co-ordinates were obtained using a Trimble GPS receiver. Flight paths, with all observation points for the entire study region, were mapped and corrected for count overlaps. This resulted in 2,381 data points over the whole study region, where at least one animal was observed (Figure 5.2). The observation

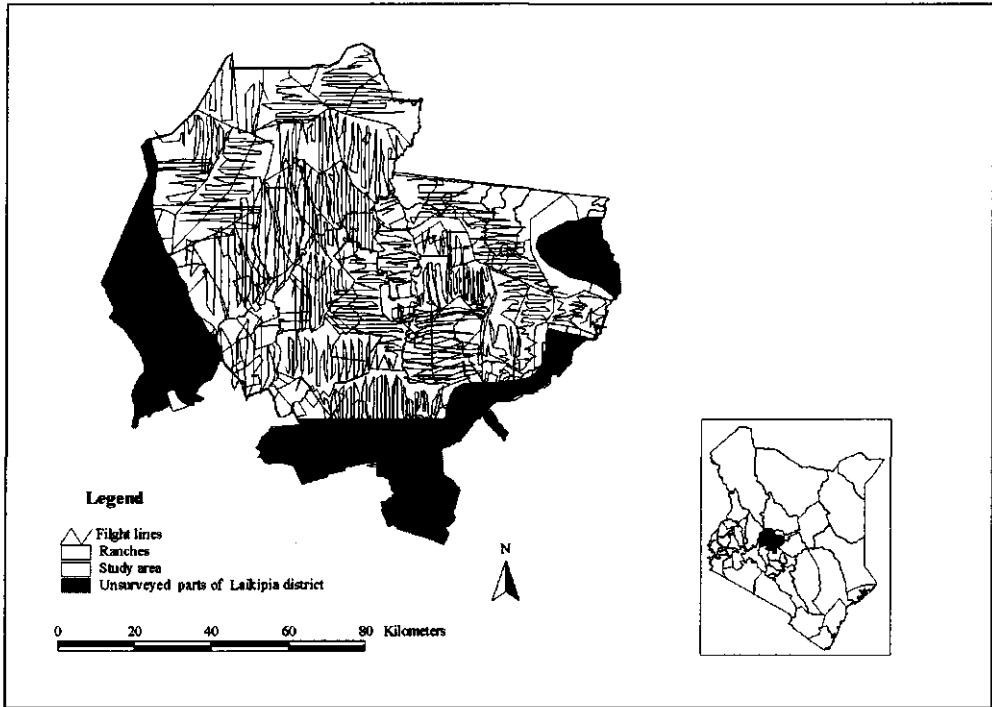


Figure 5.1. Map of the study region showing flight-lines and group ranches. Some sections of Laikipia district were not surveyed because of crop cultivation and human settlement.

of one or more animals at a given location was considered as an event. To facilitate calculation of nearest neighbour distances during spatial point pattern analysis (Kaluzny, *et al.*, 1998), we carry out all spatial point pattern analyses on a 50 by 50 km square block, which is the largest regular block that entirely falls within the study region.

Although aerial surveys are known to have Sightability problems and some animals are likely to be missed, there is considerable movement of animals that results in double counting, thereby balancing some of the missed animals. Also, since the observers used here are very experienced, the aircraft fly at fairly low heights and the species studied are easily observable, we assume that nearly all groups were sighted and do not, therefore, correct for assess sightability. Moreover, this dataset still represents the most detailed data ever collected in this ecosystem using aerial surveys.

5.3. Methods

Spatial processes yielding observed animal counts are characterised by a simple stochastic model (Cressie, 1993). We represent the location of an event in a region D of the 2-dimensional Euclidean space by $s \in D \subset \mathcal{R}^2$. A possible observation $z(s)$ at location s is a

random variable whose multivariate random field, defined by $\{Z(s): s \in D\}$, is generated by varying s over D . In this case, $Z(s)$ represents an underlying spatial process, of which a given data set denoted by $\{z(s): s \in D\}$ is a single realisation. We call this realisation a spatial point pattern. Since both $Z(\cdot)$ and D are random, analysis of the spatial point pattern aims to infer parameters of the point process (model) from the observed point pattern (data) (Cressie, 1993).

We start the analysis by using frequency diagrams and descriptive statistics to explore the data. We then map the observed spatial point patterns of the six animal species in the study and analyse whether each species exhibits spatial randomness, clustering or regularity. We test for complete spatial randomness (CSR) using the G statistic, which is based on distance d_i from the i th event to the nearest neighbouring event in A . Its empirical distribution function (EDF) is defined as:

$$\hat{G}(y) = n^{-1} \sum_{d_i \leq y} 1, \quad (5.1)$$

where n is the number of events in A .

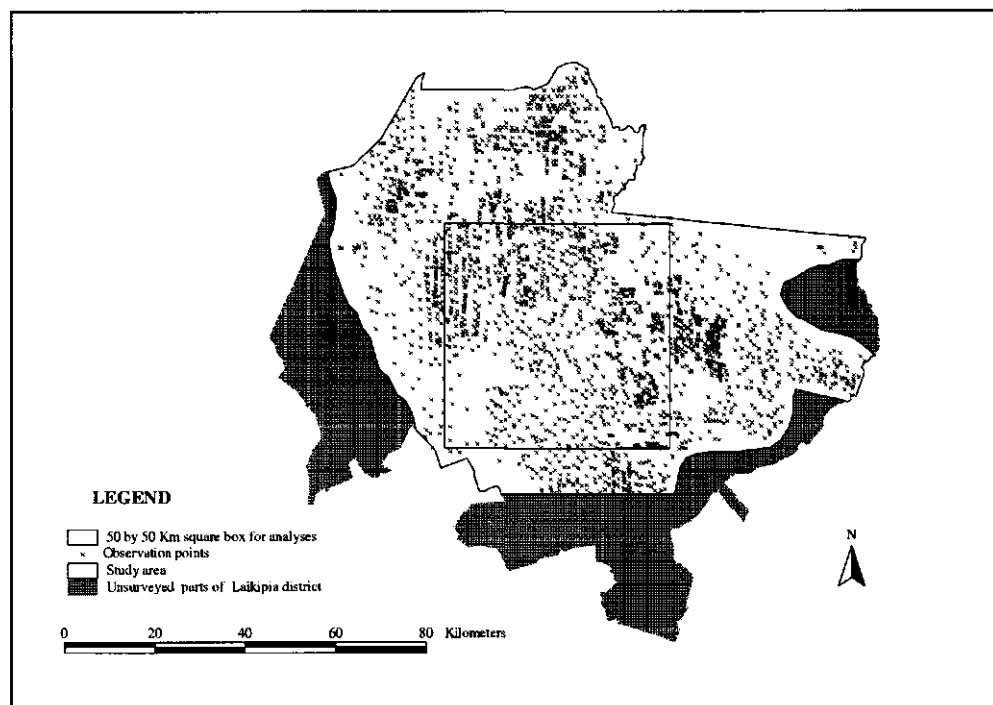


Figure 5.2. A map showing all point patterns in the study area and the 50 by 50 km square box selected for detailed spatial point analysis. Crosses represent a point where at least one animal was observed.

An excess of short distance neighbours indicates clustering, while regularity is indicated by an excess of long distance neighbours. For a formal test to assess CSR, s EDFs of the d_i s are derived from s realisations of a comparable Poisson CSR process on A . Their average provides a reference line, while the minimum and maximum provide an envelope that is used for testing in an approach that corrects for edge effects. If the \hat{G} calculated from the data falls outside this envelope, either at short or at long distances, there is evidence against CSR.

Finally, we model the underlying stochastic spatial process and estimate model parameters. To help fit models to observed point patterns, spatial data are reduced to informative descriptive statistics based on distances between events or between randomly sampled points in the study region and events.

Construction of a model to describe a spatial point pattern first examines the underlying spatial process for stationarity and isotropy by studying its first and second-order properties. First-order properties describe how the intensity λ , defined as the mean number of events per unit area, varies through space. For a stationary process, λ is constant over A . In this study, we estimate single point and local estimates of λ in A using the kernel method, which is based on a weighted function of the points in the surrounding region of influence. The radius of this region of influence determines the smoothing constant (or bandwidth) of the kernel (Green, 1994).

Second-order properties of spatial point process describe variation in spatial dependence. The K -function (Ripley, 1981; Diggle, 1983; Cressie, 1993) describes second order properties by capturing the spatial dependence between different regions of the point process. It is defined in the univariate case as:

$$K(d) \equiv \frac{E(M_d)}{\lambda}, \quad d \geq 0, \quad (5.2)$$

where E denotes expectation, M_d is the number of events within a distance d of an arbitrary event and λ represents intensity. An estimate of this function that adjusts for edge effects is defined for n locations (s_1, s_2, \dots, s_n) of all events in A as (Ripley, 1981):

$$\hat{K}(d) = n^{-2} |A| \sum_{i=1}^n \sum_{j=1}^n w_{ij}^{-1} I_d(d_{ij}) \quad \forall i \neq j \text{ and } d > 0. \quad (5.3)$$

The area of A is represented by $|A|$, d_{ij} is the distance between the i th and j th points, w_{ij} is the proportion of the circumference of a circle with its centre at i but passing through j and lies within A . The indicator function $I_d(d_{ij})$ takes the value 1 if $(d_{ij}) \leq d$ and 0 otherwise.

A multivariate extension of spatial point patterns is defined for n spatial locations of p animal species events as: $\{s_j^{(i)}: i = 1, 2, \dots, p; j = 1, 2, \dots, n\}$. For a bivariate spatial point pattern, we consider two animal species ($p = 2$) and use distance measures to test the hypothesis of independence in order to classify the two spatial points patterns as positively or negatively dependent. For second order analysis, the K -function is extended to its multivariate equivalent by the following set of functions (Diggle, 1983, p.91):

$$K_{ij}(d) \equiv \frac{E(M_d^j)}{\lambda_j}, \quad d > 0, \quad (5.4)$$

where M_d^j is the number of events of species j within a distance d of an arbitrary species i event.

We restrict our analyses to a bivariate cross K -function and define its estimator as:

$$\hat{K}_{12}(d) = (n_1 n_2)^{-1} \left| A \right| \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_{ij}^{-1} I_d(d_{ij}), \quad d > 0, \quad (5.5)$$

where n_1 and n_2 is the number of species 1 and species 2 events, respectively (Diggle, 1983, p. 107). We consider $K_{ij}(d) \neq K_{ji}(d)$ since species i may attract species j without the converse being true.

Theoretical values for $K(d)$ and $K_{ij}(d)$ are known for several types of spatial point processes. For example, a homogeneous process with no spatial dependence between events has $K(d)$ equal to πd^2 . Clustering suggest an excess of points at short distances leading to $K(d) > \pi d^2$ for small d 's while $K(d) < \pi d^2$ for regularly spaced patterns. We compute and plot the K -function for each of the six species in this study and compare results to expected values under assumptions of CSR. We also compute the multivariate K -function to test for interdependence between Burchell's zebra and other species.

5.4. Results

Abundance of the six animal species in this study ranges from 82 animals for Grevy zebra to 13,453 for Burchell's zebra, which also has the highest mean group size of 30.4 animals per point of observation (Table 5.1). Although the impala is the second most abundant species (2,732 animals), its group size (14.8) is almost 10 units lower than that of the Thompson's gazelle (23.4), implying larger groups for the Thompson's gazelle. The kongoni and Grant's gazelle appear to be solitary species due to their small groups (8.9 and 7.5, respectively). The Grevy's zebra is found in the smallest groups of the six species (4.6).

Frequency curves indicate long right-tailed skewed distributions for all species except the Grevy's zebra (Figure 5.3). This indicates clustering in the species distributions and supports use of spatial statistical analyses to account for possible spatial dependence. Accompanying descriptive statistics show varying group sizes for all species as discussed for Table 5.1 above. The Grevy's zebra appears to be observed in groups of not more than nine animals, while Grant's gazelle were not observed in groups of more than 32 animals. The kongoni also occur in small groups of not more than 25 animals. However, the two points of 40 and 91 animals shown on the frequency curve (Figure 5.3) are atypical. These results should not be viewed as models of the group size distributions, which are difficult to model.

Table 5.1. The six animal species under study, their mean group size as well as the total number of animals per species.

SPECIES NAMES			MEAN GROUP SIZE	TOTAL COUNT
Common name	Abbreviation	Scientific Name		
Burchell's zebra	ZB	<i>Equus burchelli</i> (Gray)	30.4	13,453
Grant's gazelle	GG	<i>Gazelli granti</i> (Brooke)	7.5	1,182
Thomson's gazelle	TG	<i>Gazelli thomsoni</i> (Gunther)	23.4	2,639
Impala	IM	<i>Aepyceros melampus</i> (Lichtenstein)	14.8	2,732
Kongoni	KG	<i>Alcelaphus buselaphus cokii</i> (Gunther)	8.6	755
Grevy's zebra	ZG	<i>Equus grevyi</i> (Oustalet)	4.6	82

A summary of the combined spatial point pattern of all the six animal species yields a bounding box with vertices of 225,3335, 274,880 and 10,032.9, 59,938.1 for the X and Y UTM co-ordinates, respectively and covers an area of approximately 2,470 km² (Table 5.2). Spatial point patterns for individual species cover different areal extents. The Burchell's zebra's covers the largest area of 2,464.2 km² compared to that of the Grevy's zebra, which is only 1,032.9 km², suggesting a restricted use of the habitat by the Grevy's zebra.

A plot of the combined spatial pattern is fairly dense with no apparent spatial pattern (Figure 5.2). There is an unoccupied section in the lower left corner of the box corresponding to the area closest to human settlements. Evidence of clustering emerges from individual species point patterns (Figure 5.4). The Burchell's zebra appear to have a highly clustered spatial point pattern but this may be due to its high density. The other species are less dense but appear to have clustered patterns. Such visual assessments have to be tested using distance measures for confirmation.

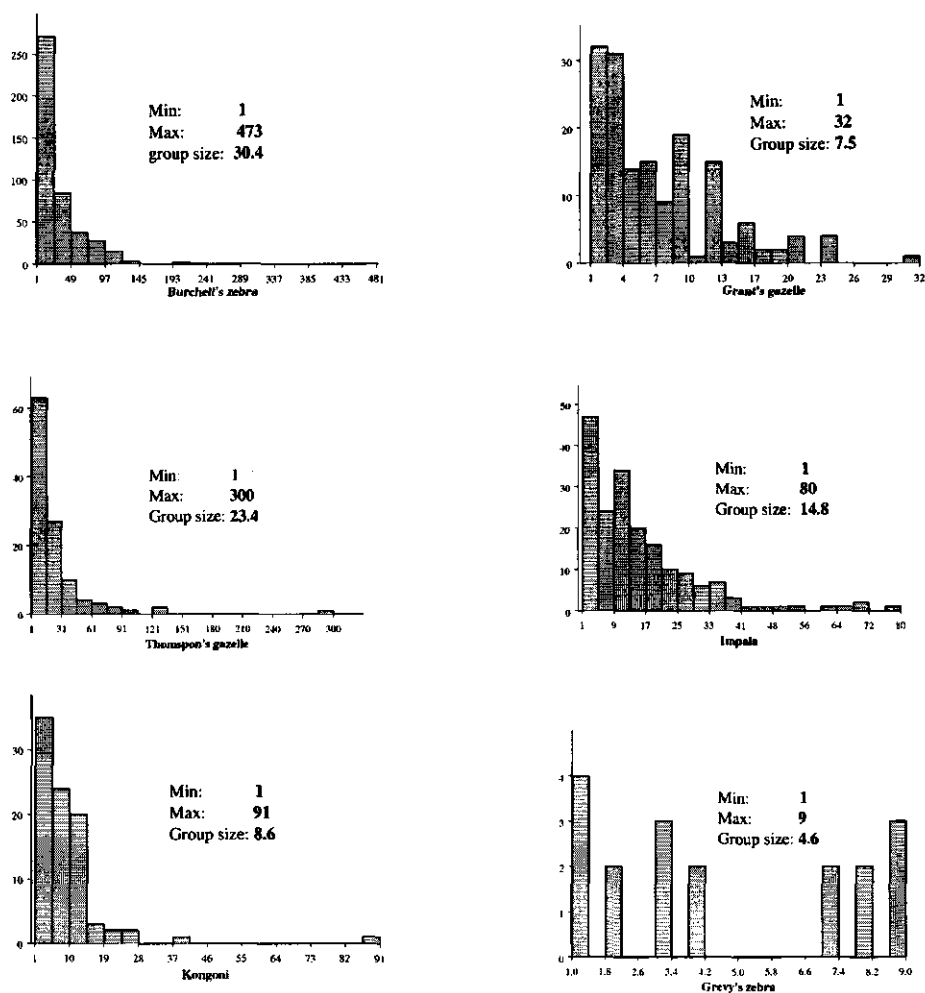


Figure 5.3. Frequency curves for all the six species with minimum, maximum and group size inserts.

Table 5.2. Summary statistics for different spatial point patterns.

SPECIES	NO. OF POINTS	CO-ORDINATE EXTENT		AREA (km ²)	POINT INTENSITY (λ)
		X	Y		
All six	670	(225335 , 274880)	(10032.9 , 59938.1)	2472.6	0.271
ZB	442	(225390 , 274809)	(10074.6 , 59938.1)	2464.2	0.179
GG	158	(226209 , 274880)	(10263.5 , 59938.1)	2417.7	0.065
TG	113	(227779 , 274259)	(10032.9 , 58592.4)	2257.1	0.050
IM	185	(225335 , 274806)	(10220.0 , 59938.1)	2459.6	0.075
KG	88	(226189 , 274691)	(10306.8 , 59515.8)	2386.7	0.037
ZG	18	(245847 , 273230)	(21228.2 , 58947.5)	1032.9	0.017

\hat{G} plots for the individual point patterns of the six species indicate strong clustering for Thompson's gazelle and Grevy's zebra, although its points are too sparse to lead to firm conclusions (Figure 5.6). Plots for Burchell's zebra and Grant's gazelle are consistent with those expected from a CSR spatial point pattern. The other species show high abundance of points at short distances suggesting clustering in their patterns.

Confirmatory assessment of CSR through simulation confirms lack of clustering for both Burchell's zebra and Grant's gazelle by having their \hat{G} plot lie within the CSR envelopes most of the time (Figure 5.6). The kongoni and Grevy's zebra also appear to have point patterns that are randomly distributed, although their envelopes are wide, possibly due to sparse distribution. Thompson's gazelle and impala appear to have highly clustered spatial point patterns since their \hat{G} plots lies outside the upper envelope for most of the distances.

Single point estimates of intensities confirm the abundance of Burchell's zebra. Its intensity estimate ($\lambda = 0.179$ points per unit area) is 66% of that of all the six species combined ($\lambda = 0.271$) (Table 5.2). Grevy's zebra has the lowest intensity estimate ($\lambda = 0.017$), perhaps as a consequence of its rare occurrence. All other species lie in between.

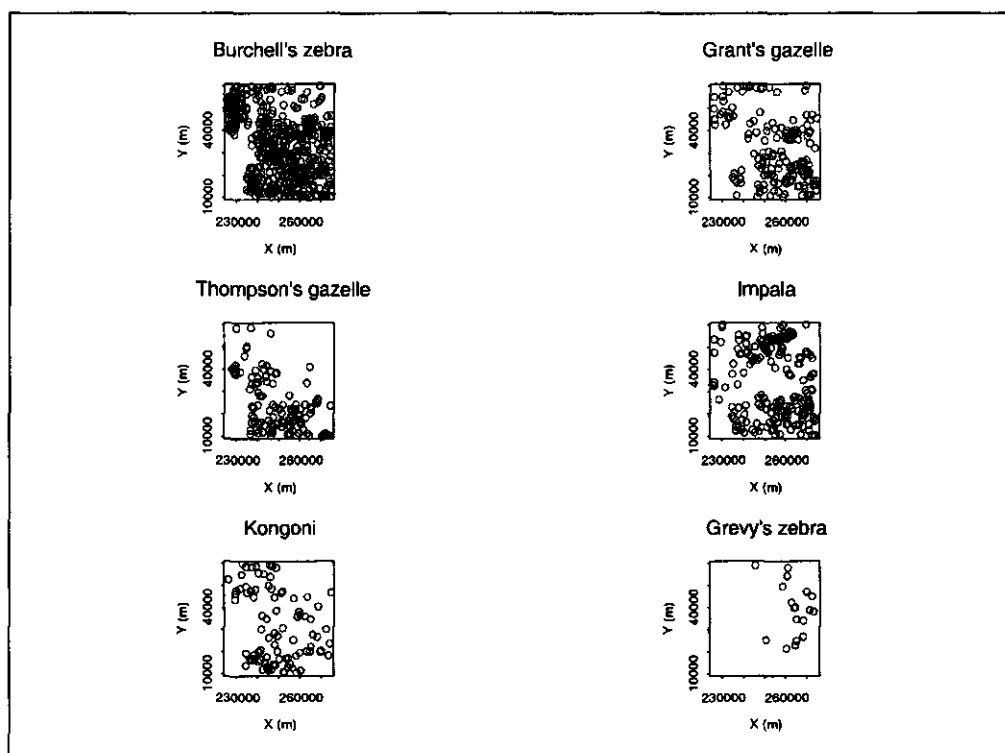


Figure 5.4. Spatial point patterns for the six individual species under study.

When intensities were estimated locally using the kernel method and plotted as 3-D figures, they were all observed to vary more than is expected from random fluctuations (Figure 5.7). This variation results from lack of observations in the lower left section of the study region for all species. It may also be due to a selective use of habitats by animals. Further, it indicates that other factors like human settlement or different vegetation types may be inhibiting species occurring in some places.

Interaction of the six animal species with vegetation shows the Burchell's zebra to occur in all the nine vegetation types, but being most abundant in *acacia drepanolobium* bushland (Table 5.3). It forms the largest groups in dwarf bush and grassland (41.8 animals) and the smallest in rocky habitats, where only 12 zebras are observed in three groups of four animals each. It is the only species whose group size differs significantly between vegetation types ($p = 0.035$). All species appear to prefer the *acacia drepanolobium* bushland, where all record the highest abundance with the exception of Grevy's zebra, which prefers open thicket. Apart from the rare Grevy's zebra, Grant's gazelle and kongoni form the smallest groups in all vegetation types with most having less than 10 animals. The Grevy's zebra evidently prefers special habitats, being absent in 3 out of the 9 vegetation types.

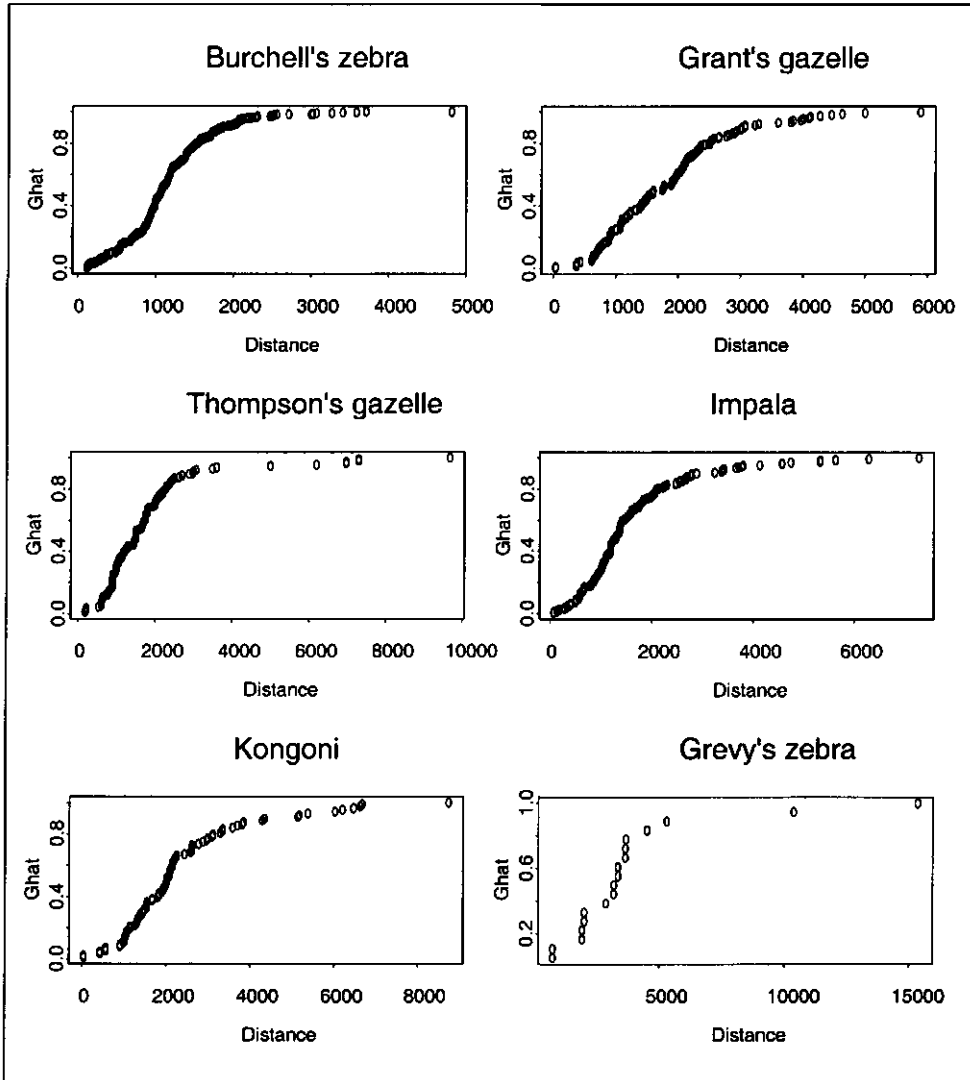


Figure 5.5. Plot of \hat{G} to check clustering in the point patterns for the six species.

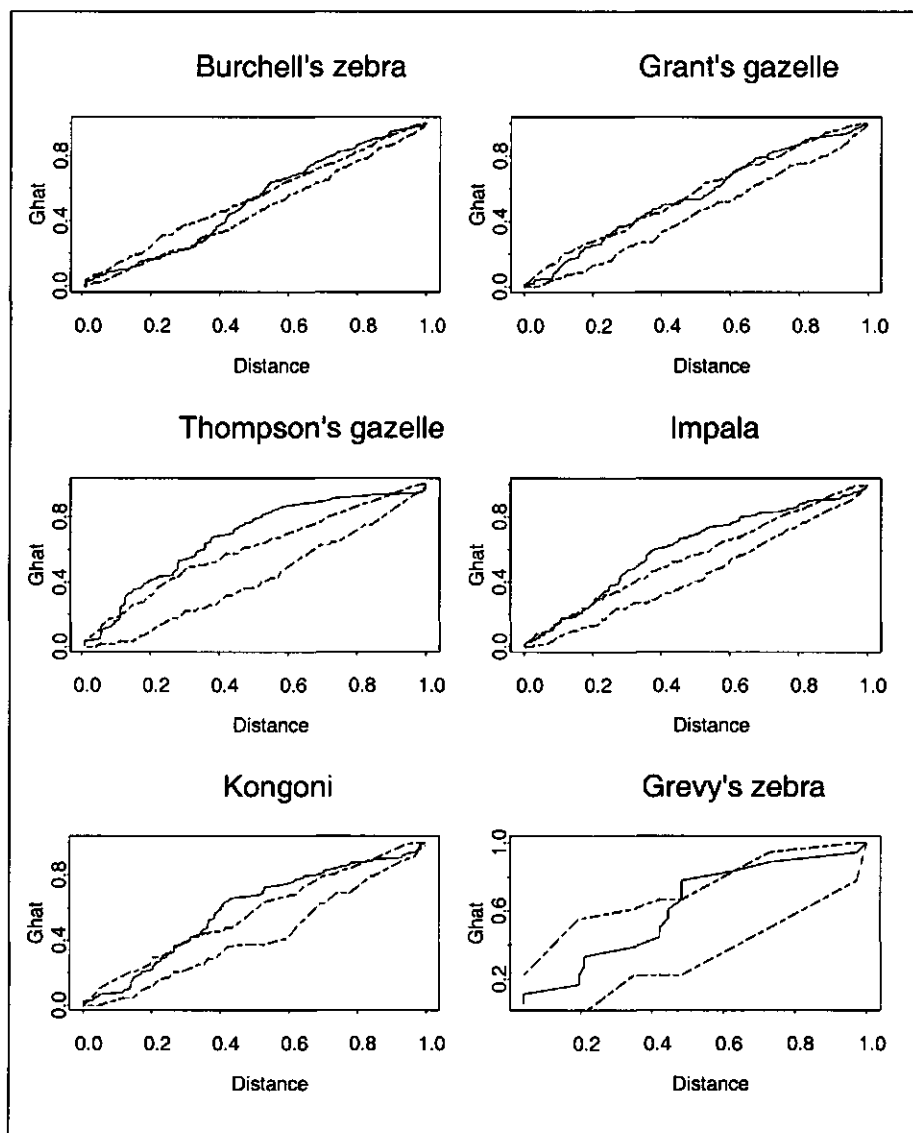


Figure 5.6. Plot of \hat{G} with envelopes to test for CSR for the above spatial point patterns.

When examining second order properties using the K -function, we plot \hat{K} values against corresponding distances and add a reference line to correspond to spatial randomness. For the combined spatial point pattern of all six species, the resulting function appears close to what is expected from a CSR process, implying that the combined observations of all six species in the study area does not follow a given pattern.

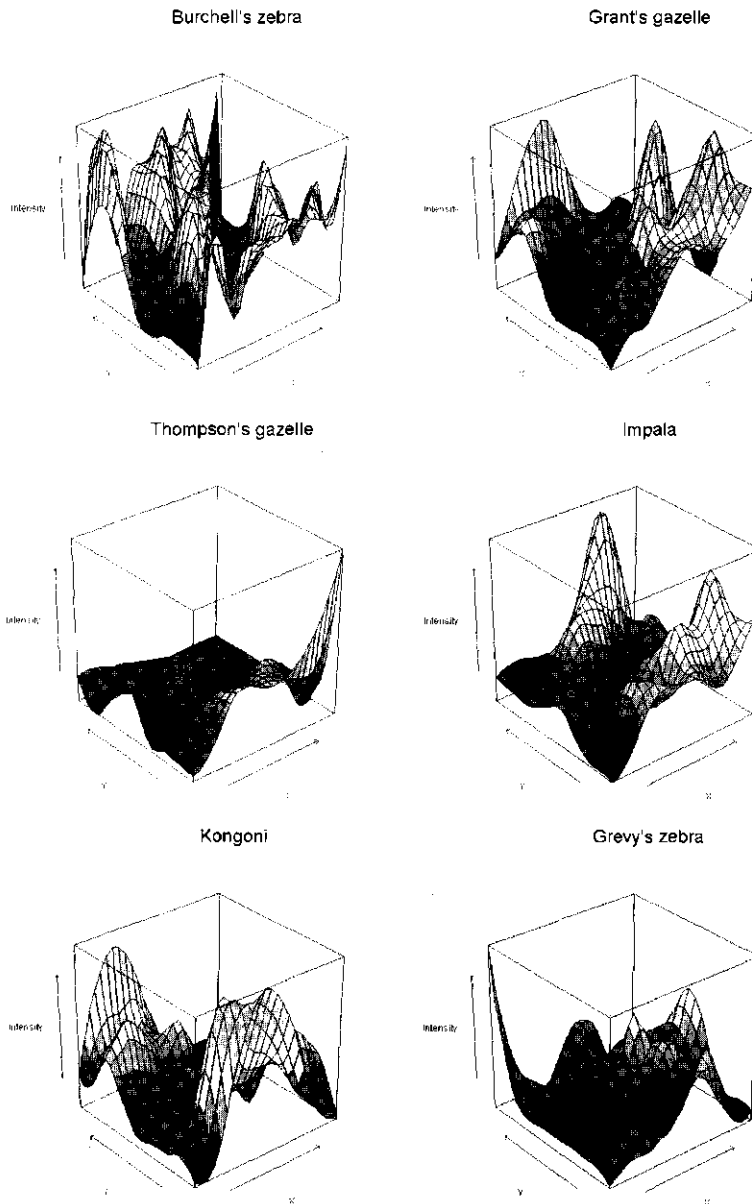


Figure 5.7. 3-D images showing how intensity varies in the study area for each of the six species.

Table 5.3. Distribution of the six animal species in nine different vegetation types (habitats) observed during the survey carried out in Laikipia. For each vegetation type, the minimum and maximum number of observed animals is given together with the number of locations, mean group size as well as the total number per vegetation. Also given is a *p*-value testing for equality of mean group sizes across vegetation types.

SPECIES	VEGETATION TYPE									Open thicket	<i>p</i> -value
	<i>Acacia drepanolobium</i> bushland	<i>Acacia seyal</i> bushland	Arid zone	Bare rock	Dwarf bush and grassland	Grassland	Leafy bushland thicket	Mpala scarpline vegetation			
ZB	Min	1	2	1	4	2	1	26	5	1	
	Max	311	473	20	4	111	139	26	22	93	
	N	184	75	13	3	40	88	1	6	30	0.035
	Mean	39.9	34.1	8.6	4	41.8	26.5	26	11.2	13.0	
	Sum	6245	2558	112	12	1670	2328	26	67	391	
GG	Min	1	1	1	4	1	1	-	2	4	
	Max	24	12	12	4	15	32	-	23	10	
	N	76	14	4	2	14	41	-	4	3	0.548
	Mean	7.9	5.2	5	4	9.1	7	-	10.6	6.3	
	Sum	602	76	20	8	127	287	-	43	19	
TG	Min	1	1	5	-	4	1	24	10	-	
	Max	90	300	5	-	23	129	24	10	-	
	N	48	18	1	-	5	39	1	1	-	0.976
	Mean	22.1	27.7	5	-	12.8	25.1	24	10	-	
	Sum	1059	499	5	-	64	978	24	10	-	

IM	Min	1	1	9	1	1	1	20	8	1	
	Max	70	20	14	1	50	80	20	20	61	
	N	65	16	3	1	9	35	1	3	51	0.294
	Mean	17.5	8.1	12.3	1	21.9	14.0	20	16	13.2	
	Sum	1134	130	37	1	197	491	20	48	673	
KG	Min	1	2	6	2	1	1	13	1	1	
	Max	25	91	6	3	15	40	13	12	13	
	N	46	8	1	2	7	17	1	3	3	0.634
	Mean	7.4	16.6	6	2.5	7.4	9.5	13	8	5.7	
	Sum	342	134	6	5	52	162	13	24	17	
ZG	Min	1	-	8	-	2	1	-	3	1	
	Max	7	-	9	-	2	3	-	9	9	
	N	5	-	2	-	1	3	-	2	5	0.236
	Mean	3.8	-	8.5	-	2	2	-	6	5.2	
	Sum	19	-	17	-	2	6	-	12	26	

To test for CSR, we consider points with a maximum nearest neighbour distance of five km and add envelopes to the K -function (Figure 5.8). Clustering is strongly suggested for both Thompson's gazelle and impala, whose $K(d)$ plots lie above and outside the CSR envelopes. All the other species have their $K(d)$ falling between the CSR envelopes suggesting lack of either clustering or regularity in their spatial point patterns. This may be explained for the Burchell's zebra by its high abundance and wide distribution, which tends to mask excesses in both short and long nearest neighbour distances. Envelopes for Grevy's zebra are too wide to lead to firm conclusions about its point pattern. This may result from its sparse spatial distribution.

Results from the K -function multivariate analysis do not reveal strong attraction between most species (Figure 5.9). However, Grevy's zebra appear to have an excess of large nearest neighbour distances with Thompson's gazelle and kongoni, suggesting inhibition between these two species and Grevy's zebra.

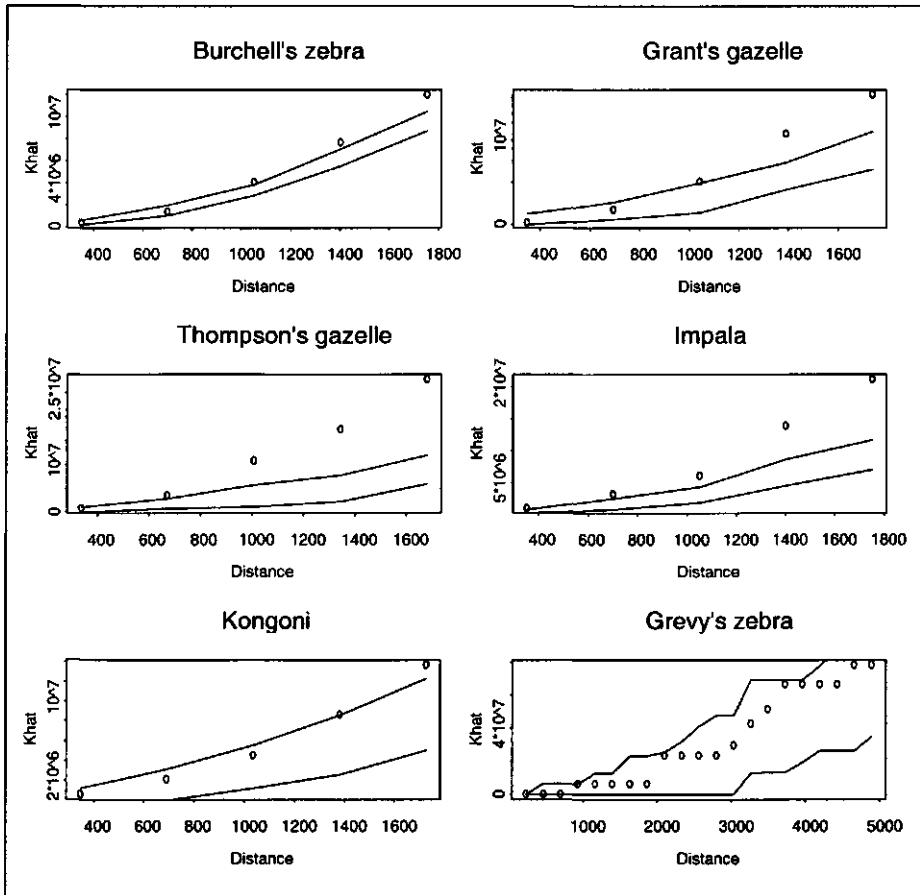


Figure 5.8. K -functions for spatial point patterns of individual species with envelopes to test for CSR.

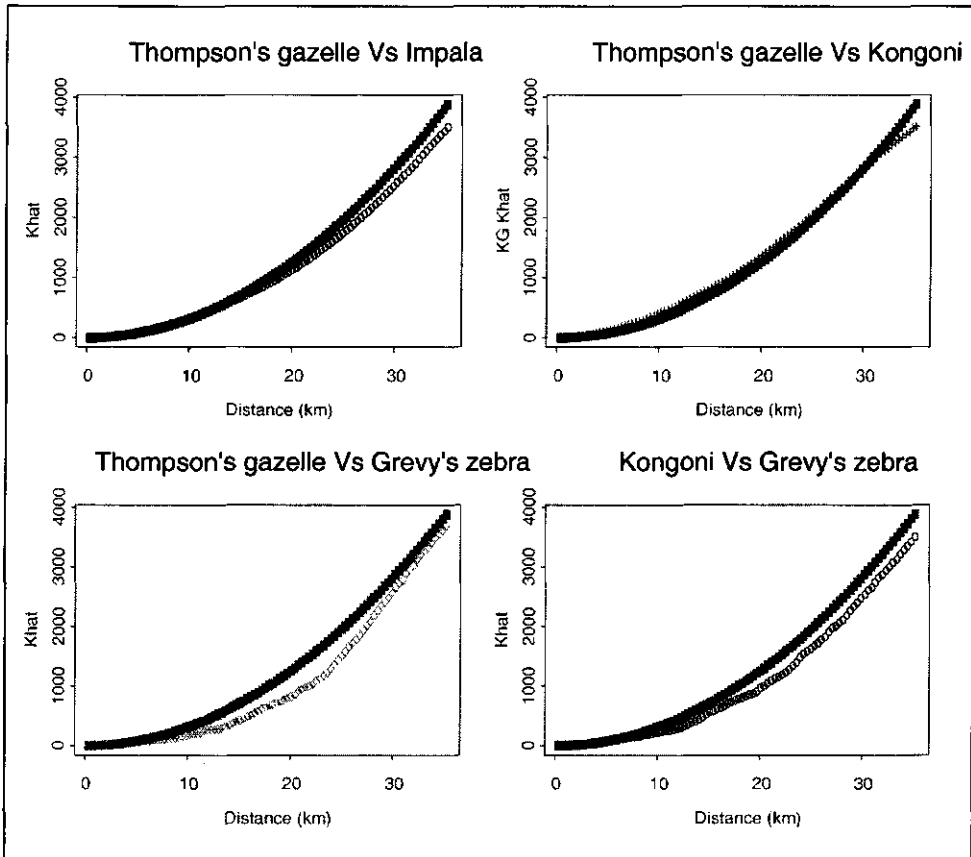


Figure 5.9. Multi-variate K -functions for spatial point patterns of some pairs of the six species. In all cases, the darker line corresponds to a spatial point pattern expected from a CSR process.

5.5. Discussion

This study illustrates that point data collected from aerial surveys of wildlife can effectively be used to determine species spatial distribution patterns. The patterns may be regular, clustered or random depending on effects of different factors like feeding habits, habitat preference, predator avoidance or human interference. Since it is not sufficient to determine clustering in species point patterns by mere visual assessment, there is need to use other objective procedures and techniques.

The study further shows that G plots and CRS simulation techniques are effective in determining and confirming species distribution patterns. Of the six species studied here, results show that all have clustered distributions with the exception of Burchell's zebra and Grant's gazelle, which indicates possible spatial dependence on factors other than what was analysed.

Further, results also confirm the hypothesis of species spatial distribution being related to trophic requirements. For example Burchell's zebra, which has a large body and is a grazer, occurs in large groups and appears to be distributed in relation to its food distribution (Table 5.3). However, one would expect the Grevy's zebra to also occur in large groups but this was not the case in this study. Perhaps this is because the species is rare and under threat. Moreover, the Grevy's zebra inhabits arid and semi-arid (xeric) environments and is exclusively a grazer, subsisting on low quality forage and generally surviving without water for up to 4 days (Ginberg, 1985, 1988; Rowen, 1992). It prefers grassy dwarf shrubland and dwarf shrubby grasslands (Stelfox and Ngatia, 1979, 1980; Churcher, 1982), and has a fluid intraspecific association (Klingel, 1974; Ginsberg, 1988), grouping into mares and foals, bachelor herds and mixed groups, which do not stay together for long periods of time (Wargute and Said, 1996).

On the other hand, mixed feeders like the impala occurred in moderate groups while Grant's gazelle, which is a small antelope, occurred in small groups. These patterns are explained by the fact that small-bodied animal species need higher quality diet than their larger counterparts. These high quality food items (e.g. leaves and fruits) preferred by mixed feeders are generally more dispersed than low quality food items like grass, which are selected by grazers (Voeten, 1999).

In their study on dietary components of ungulates in the Mara, Hansen *et al.* (1985) reported that the Grant's gazelle consumed more than 50% shrub and/or forbs, while impala's overall diet consisted of about 50% grasses and 'grass likes'. Other ungulates i.e. Thomson's gazelle, kongoni and Burchell's zebra were clearly grazers, consuming more than 50% grasses. Within grazers, some species have certain dietary similarities and preferences. In a study conducted in the Mara, high diet similarities were observed between kongoni and Thomson's gazelle (Hansen, *et al.*, 1985).

Intensities of distribution of different species appear to vary in the study area (Figure 5.7). This may be related to the distribution of food or vegetation types among other factors. For example because grasslands are distributed throughout the study block, Burchell's zebra, which are grazers, are also widely distributed. However, impala and Thomson's gazelle, which are mixed feeders, appear to be restricted or clustered in shrubby grassland vegetation.

Apart from providing a method to analyse data collected through the use of GPS, this study also investigates the relationship between observed spatial distribution and dietary preferences in relation to different vegetation (habitat) types. By relating abundance and distribution to different vegetation types this study provides an opportunity to classify animal species into different categories according to their feeding behaviour and habitat preference. The results could also help in explaining differences in distribution intensities and group sizes among different species. In areas where land allocation, wildlife cropping or hunting is undertaken on periodic basis, information on wildlife abundance and distribution patterns is essential for designating cropping and hunting areas.

It should be noted, however, that it was not possible to rigorously model group size because of the long right tails in species distributions. Further, the social structure of most animal species is very complex with non-dominant males often being in separate herds, causing changes in the composition of herds during mating seasons (Steffens, 1993). The herds may also spread out during grazing, leading to aerial surveys reporting several smaller groups that actually belong to the same herd. Therefore, spatial point pattern analyses techniques should be used together with ground observations.

5.6. Conclusions

The following conclusions can be drawn from this study:

1. Differences in abundance and spatial distribution of animal species are related to their food requirements as well as how food availability is distributed in the area.
2. Spatial statistical techniques like G plots and K -functions have great potential in studying factors influencing abundance from wildlife data collected with GPS. They allow one to estimate spatial distribution of animals and to generate information that is necessary for management of wildlife populations.
3. This study provides a rapid and cost effective method of assessing species patterns, thereby providing timely and necessary information for management of ecosystems where animal harvesting is practised.

CHAPTER 6*

Comparison of wildlife sampling procedures

* This chapter is based on Khaemba, W. M., Stein, A. and Rasch, D. (*Submitted*). An empirically simulated investigation and validation of sampling methods for aerial surveys of wildlife. *African Journal of Ecology*.

Abstract

This chapter reports a comparison of the distribution, sampling and estimation of abundance for two animal species in a Kenyan ecosystem through an intensive simulation in a geographic information system (GIS). It investigates two sampling designs commonly used in wildlife surveys: systematic and random sampling designs. It compares their performance to an adaptive sampling design through their root mean square errors (RMSE) at three increasing sampling intensities. Simulation is based on data collected on a prior survey conducted in the study area in which geographical locations of all observed animals were recorded. This provides more detailed data than what is usually collected from transect surveys. The study further assesses the impact of sampling designs and intensities on estimates of population parameters. Estimates of precision increase with increasing sampling intensity, while no significant differences are observed between random and systematic estimates. An increase in precision is demonstrated for the adaptive design, thereby validating the use of this design for sampling clustered populations. The study illustrates that statistical can be complemented with GIS techniques to offer greater insight in the dynamics of wildlife populations.

6.1. Introduction

Accurate and up-to date information on the size and distribution of animal populations is crucial for conservation and proper management of tropical savannah ecosystems. Such information can be obtained rapidly and at low costs through airborne surveys (Norton-Griffiths, 1978; Wint, 1998). However, because of the diversity and high mobility of animals, and also due to various forms of observer bias, results from these surveys tend to be inaccurate and sometimes unreliable (Caughley, 1974; Smith, 1981). A rigorous analysis of such inaccuracies using field tests is not possible because the sampling methods used are also subject to similar biases. This leaves simulation as a viable option to investigate and validate different sampling techniques used in aerial surveys of wildlife.

Simulation allows the study of systems that are expensive, difficult or impossible to observe in the field (Robinson, 1994). It also allows replication where only one realisation would have been possible, thereby giving the investigator more control when considering different scenarios. Furthermore, the rapid development in computer technology has made it possible to run several thousands of replications within a fairly short time.

Simulation has been used to study various phenomena. For example, Haile and Weidhaas (1977) used simulation to study mosquito population dynamics. Onyeanus (1986) used simulation to study the impact of tourists driving off designated roads in wildlife reserves, while Norton and Williams (1992) used simulation to model species habitat for nature conservation. Recently, a simulation study has been used to propose an improved sampling strategy for surveys of wildlife populations in a similar ecosystem (Khaemba and Stein, *in review*).

This study aims to investigate two sampling strategies commonly used in aerial surveys i.e. random and systematic sampling designs, and compare their performance to that of an adaptive sampling design at equivalent sampling intensities. The investigation is conducted through a simulation of the distribution of animals under a GIS environment. The simulation is based on data collected in Laikipia district of Kenya. The study focuses on two key wildlife species, namely the elephant (*Loxodonta africana*) and the Burchell's zebra (*Equus burchelli*) because of their influence on the ecosystem. The elephant migrates seasonally to the neighbouring Samburu district (Thouless, 1995), thereby posing great danger to agricultural crop in the two districts. On the other hand, the zebra constitutes half of the wild herbivores found in this ecosystem (Georgiadis, 1997). Moreover, the two species are easy to observe and a vast amount of literature and data exist on their behaviour to allow easy simulation of their dynamics. Data on which the simulation is based was obtained from a detailed airborne survey of the study area.

By simulating observed patterns of animals, it is possible to carry out empirical research on the sampling of wildlife populations for this ecosystem. It is also possible to investigate the effects of different sampling designs and intensities on estimates of

abundance and precision. Thus, the study proposes a faster and cost effective procedure for validating methods used in aerial surveys of wildlife.

6.2. Materials and methods

6.2.1. Study area

Laikipia district in Kenya covers an area of approximately 9,700 km², 33% of which contains little or no animals at all due to intense human settlement (Figure 6.1). Most of the district lies to the north of the Equator between latitudes 0° 17'S and 0° 45'N and longitudes 36° 15'E and 37° 20'E. It lies on a plateau with an altitude that varies gently from 1,600 m to 2,300 m above sea level, yielding a gentle to slightly undulating topography in most places. It is bounded in the Southwest by Mount Kenya, the highest mountain in the country, as well as the Aberdare ranges, leading to the highest elevation being in the Southwest while the lowest is found in the Northwest (Ojwang', 2000).

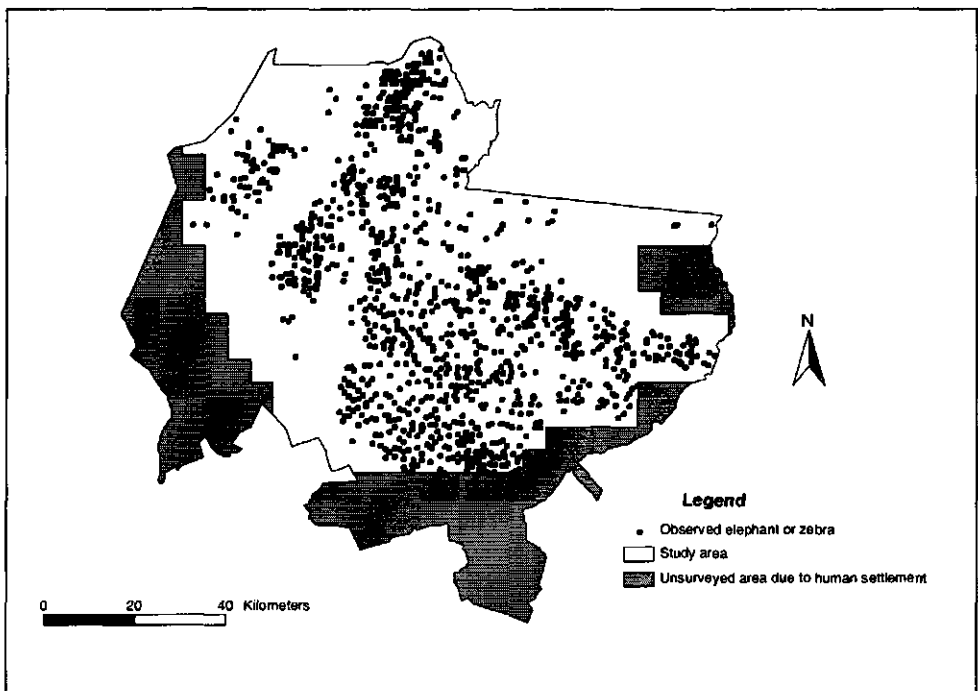


Figure 6.1. Map showing the study area in Laikipia district as well as the area not surveyed due to human settlement. Dots indicate points where an elephant or zebra was observed.

Mount Kenya and the Aberdare ranges are major catchments for perennial tributaries of Ewaso Ng'iro river - the main river whose basin drains most of the Laikipia plateau. These tributaries are crucial for the arid and semi-arid lowlands, where most streams dry during the larger part of the year. Other sources of water include shallow waterholes and boreholes sunk in the ground (Berger, 1989).

Three annual rainy seasons can be distinguished in this area: The long rains (March - May), middle rains (June - August) and the short rains (October - December). The average annual rainfall varies between 400 mm and 750 mm, with the highest being recorded on the foots of Mount Kenya and the Aberdare ranges (Brown and Cocheme, 1973). Annual potential evapotranspiration is high, rising to between 1,800 mm and 2,000 mm. Annual variation in air temperature is very low with an average of approximately 17.5°C and a daily amplitude of 1.4°C, which is affected by the high altitude (Berger, 1989). The district is composed of areas of unpredictable rainfall most of the year making it suitable only for ranching and wildlife farming (Georgiadis, 1997).

6.2.2. Data

Since this is an area of wildlife farming, a lot of data has been collected to monitor abundance and distribution of the various wildlife species found here. Data is commonly obtained through regular surveys conducted using the SRF method (see section 2.3). This leads to observations being concentrated along strip transects systematically covering the study area and may not allow modelling of the spatial distribution of different wildlife species. For this study, however, more detailed data that allows the modelling of observed spatial patterns is available. These data were collected from a detailed survey conducted in Laikipia district in September 1996, which resulted in 2,381 data points over the whole study area where at least one animal was observed.

Laikipia ecosystem is home to more than 13 species of wild herbivores with a total number of about 70,000 animals. The most abundant of these is the zebra, which accounts for about half of the total population. It is also the most widely distributed species being found in almost all vegetation types, although it is known to favour open grassland and avoids drier areas with less 500 mm mean annual rainfall (Haltenorth and Diller, 1996). Another ecologically interesting species in this area is the elephant, of which there are over 3,000 animals. It occurs in several habitats but prefers woodlands and thickets and is mainly found in the central and southern part of Laikipia district. However, it seasonally migrates for long distances (sometimes up to 500 km) in search of newly blooming growth and fruiting plants (Thouless, 1995; Haltenorth and Diller, 1996).

Other available data include vegetation cover types and structure. More than 50% of the vegetation cover types found in Laikipia belong to plant communities dominated by acacia species whose undergrowth mainly consists of several useful grass species and herbaceous layers (Ojwang', 2000). Pratt and Gwynne (1977) classified these communities as bushland and shrubland. Other vegetation types include grasslands and shrubby

grassland (approximately 25%) and forests and woodlands (approximately 13%). Data on precipitation, evapotranspiration, drainage, human settlement, agricultural activities, water availability, slope, height and aspect were also available.

6.2.3. Sampling designs in aerial surveys

Both stratified random sampling (SRS) and systematic sampling (SS) are applicable designs for aerial surveys of wildlife populations using strip transects (Jolly, 1969a; Caughley, 1977; Cochran, 1977; Smith, 1981; Thompson, 1992). With SRS, a sample of size n out of a total of N transects in the population is selected without replacement. This yields an inclusion probability of $\pi = n/N$ for any transect, making each transect equally likely to be selected into the sample. The total survey area A and that of each transect a_i ($i = 1, 2, \dots, N$) are known.

For SS, the N transects are numbered from 1 to N and an integer r is chosen such that $N = r \cdot n$. An initial transect is randomly chosen from the first r transects in the population and thereafter, every r th transect is selected. Since N is generally not an integral multiple of r , systematic samples obtained from the same finite population may vary in size. An improvement treats the N transects as being arranged round a circle and takes r to be the nearest integer to N/n (see Cochran, 1977, p. 206). Here, the first transect is randomly selected from integers between 1 and N , after which every r th transect is included in the sample until the required sample size n is obtained. This results in an equal inclusion probability for all transects in the population.

A modification of SS known as systematic reconnaissance flights (SRF) has been adopted as a sampling design for most aerial surveys in the tropics because of its low cost per sampled unit compared to other sampling methods (Norton-Griffiths, 1978; Steffens, 1993; Grunblatt, *et al.*, 1996; Ottichilo, 2000). Systematic flight lines, placed in an East-West direction in the study area, define transects for this design. The transects are generally separated by distances of 2.5 km, 5 km or 10 km, depending on the sampling intensity desired. Intensity is defined as the sampled area versus area of the whole study area, which is equivalent to ratio of the number n of transects in the sample to the total number N of transects in the whole study area. Common sampling intensities lie between 3% for a low-resolution survey and 15% for a high-resolution survey (Grunblatt, *et al.*, 1996). The method also allows collection of environmental data to be used for explaining observed animal distributions and for long-term monitoring purposes (Norton-Griffiths, 1978).

Estimation of population parameters for all designs is through the Jolly II procedure (Jolly, 1969a). If we let y_i represent observed animals in the i th transect of area a_i , an estimate of the population density D is given by

$$\hat{D} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n a_i}. \quad (6.1)$$

This leads to an estimate of the population total Y given by $\hat{Y}_j = \hat{D} \cdot A$ and a variance estimate equal to:

$$\text{Var}(\hat{Y}_j) = \frac{N(N-n)}{n(n-1)} \left(\sum_{i=1}^n y_i^2 + \hat{D}^2 \sum_{i=1}^n a_i^2 - 2\hat{D} \sum_{i=1}^n y_i \cdot a_i \right). \quad (6.2)$$

For the adaptive sampling design, an initial sample of size n is selected using SRS. A condition $C = \{y: y \geq \delta\}$ is defined for some constant δ and the number of animals y observed in the transect to determine whether neighbouring transects to those in the initial sample will be added to the sample or not. The constant δ is calculated individually for each species based on data from previous surveys in the same or similar region. For sequentially numbered transects T_{i-1}, T_i, T_{i+1} , C determines if T_{i-1} and T_{i+1} are observed. This yields k_0 observed networks ($k_0 \leq n$, since the sample consists of distinct transects). The i th transect in the k th network contains y_{ik} animals and has an area equal to a_{ik} ($i = 1, 2, \dots, t_k, k = 1, 2, \dots, k_0$). Thus, the density D_k of animals in the k th network is estimated by

$$\hat{D}_k = \frac{\sum_{i=1}^{t_k} y_{ik}}{\sum_{i=1}^{t_k} a_{ik}}. \quad (6.3)$$

Since the design deliberately selects transects with high abundance, D is overestimated by (3). This is corrected by multiplying \hat{D}_k by a weight w_k , which equals the proportion contributed by the k th network to the total sample area i.e. $w_k = \frac{a_{.k}}{\sum_{k=1}^{k_0} a_{.k}}$, where

the term $a_{.k}$ denotes the area of all transects found in network k and $\sum_{k=1}^{k_0} w_k = 1$. An improved estimate of D is obtained by averaging over all k_0 networks by

$$\hat{D} = \frac{\sum_{k=1}^{k_0} \hat{D}_k \cdot w_k}{k_0}. \quad (6.4)$$

An estimate of Y is given by $\hat{Y} = \hat{D} \cdot A$ with an approximate variance equal to

$$\text{Var}(\hat{Y}) = \frac{\hat{K}(\hat{K} - k_0)}{k_0(k_0 - 1)} \sum_{k=1}^{k_0} (y_k - \hat{D} \cdot a_k)^2, \quad (6.5)$$

where y_k equals the number of animals observed in network k ($k = 1, 2, \dots, k_0$) and \hat{K} is an estimate of the total number of networks in the population, obtained through a probabilistic derivation. Derivation and use of the above designs in wildlife surveys are treated in greater detail by Khaemba and Stein (*In review*).

6.2.4. Modelling and simulation

Three distinct stages of the simulation are envisaged. First, the study area, its boundary and topographical features are modelled through GIS based on available geographical data. The resulting terrain model remains fixed throughout the simulation. Second, distributions of the two selected wildlife species are generated according to a spatial model of observed patterns in the study area. For each generated point, a probability of an animal occurring at that point is calculated from a logistic model derived from explanatory variables observed in the study area resulting in some points being discarded. The remaining ones are overlaid on the terrain model obtained above to represent a distribution of animals in the terrain. The third stage involves simulating strip transects over the terrain to represent an aircraft flying over the study area. Observations falling within the strip transects are recorded and the data analysed to obtain estimates of population parameters. The procedure is replicated 100 times per species for the different sampling designs and intensities.

GIS modelling of the terrain starts by delimiting the boundary of Laikipia to serve as a background upon which all simulation takes place. All existing data on terrain, vegetation type, drainage patterns and other landscape features are stored in the same geographical database making it easy to apply GIS techniques. Buffers are created at varying distances of 0.5, 1, 2, 5, 10 and 15 km from permanent rivers and other water sources like boreholes, as well as points of cultivation and human settlement. These buffers are overlaid on other landscape features and topographical characteristics like aspect, elevation, height, slope, climatic and vegetation data to define a complex GIS terrain model.

Based on data from explanatory variables observed in the study area, a functional model is derived using stepwise logistic regression to indicate probabilities of an animal occurring at a given location. For p explanatory variables, the probability π_o of observing an animal is given by

$$\pi_o = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}, \quad (6.6)$$

where x_1, x_2, \dots, x_p and $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ represent p explanatory variables and corresponding regression parameters and \exp denotes an exponent to the power of the natural logarithm. The 16 classes of vegetation are converted into 15 dummy variables to yield a variable for each class. The model is derived for the two animal species and is kept constant during the survey period.

Distributions of the two animal species in this study are simulated as spatial patterns following a clustered distribution. One way of simulating such a distribution is through a Strauss process conditioned on the number n of observed points in a given area (Cressie, 1993, p. 676). The simulation is implemented in S-Plus© (Mathsoft, 1999) and is based on data from the survey of September 1996. The dynamics of each species are simulated so as to reflect real life situations. For example, differences in distance between groups of animals are incorporated in the simulation of the two wildlife species. Assuming the co-ordinate of a point where an animal has been observed falls at the centre of a shadow cast by a light shining directly above that animal, then the distance between two zebras will be smaller than between two elephants.

Resulting co-ordinates are used as input to create a point coverage in Arc Info© (ESRI, 1997), where each point represents occurrence of an animal. The coverage is overlaid with the terrain model from which the functional model is used to determine the probability of an animal occurring at each point. Points with probability less than 0.3 are considered unsuitable habitats and are discarded from the database. This yields a terrain with a known number of animals distributed on it. Sampling and abundance estimation can now be carried out.

In simulating an aircraft flying above locations of animals in the above terrain, a rectangular grid is defined to cover the whole study area. Horizontal lines separated by 2.5 km are systematically drawn on the grid to represent flight lines. Strips observed by the left and right observers during actual surveys (width $d_L = 141\text{m}$ and $d_R = 141\text{m}$, respectively), are obtained by buffering either side of the flight lines. These buffers are separated by a distance $d_B = 250\text{ m}$, corresponding to the blind spot under the aircraft (Figure 6.2). Combining the left and right strips yields a strip transect of width W equal to $d_L + d_R = 282\text{ m}$ and commonly employed for aerial surveys. The strip transects are clipped using the boundary of the study area to yield n transects of unequal lengths. Sampling intensity is varied for the systematic design by changing the interval between transects from 2.5 km to 5 km and finally to 10 km. Thus corresponding to common values used for surveys in this study area (Grunblatt *et al.*, 1996).

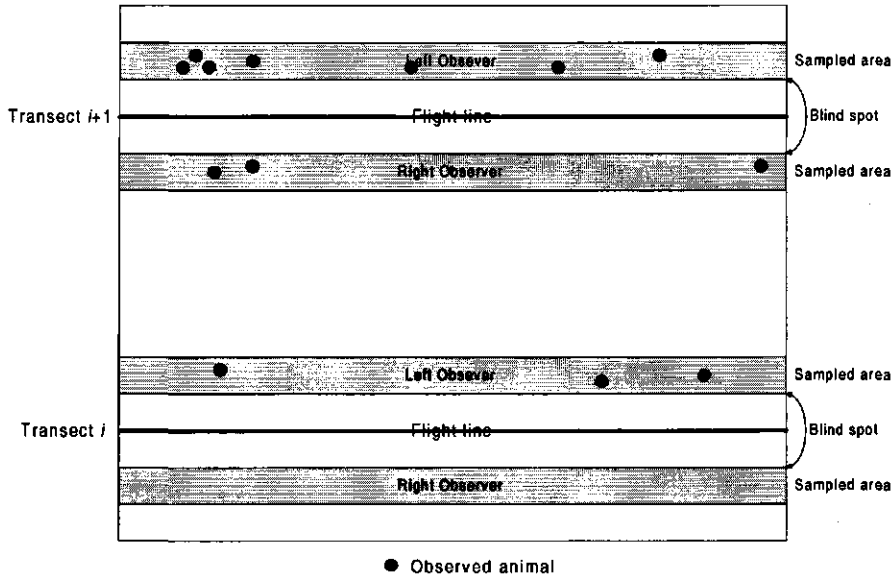


Figure 6.2. The figure shows buffering of grid lines to produce two transects (i and $i+1$) for a small section of the study area. The blind spot below the aircraft is indicated as well as the area searched by both left and right observers during surveys. Points at which animals are observed are also shown.

To locate sampling units for the random design, a vertical baseline is drawn in the middle of a grid covering the study area. This baseline is then divided into sections of the same length as the strip width $W = 282$ m and labelled 1, 2, ..., N to represent the total number of transects covering the whole study area. A random number generator is used to select n random numbers between 1 and N . Transects are placed at the location of these points along the baseline at right angles to yield samples of size n . These are overlaid with the data layer containing animal co-ordinates on the terrain model and points falling within transects are recorded as observed animals.

The implementation of random and systematic sampling designs is easy because sample transects are defined in advance, allowing them to be overlaid with simulated animal distributions. For the adaptive design, however, the final sample is not known at the start of the survey since it changes through observed patterns in the field. A modification to the selection procedure is to define an initial random sample of a small sample size. This size depends on several factors e.g. the precision required, available resources etc., but it could be 40% of the n used in the above designs. Define all possible transects in the study area and overlay this design with distributed animals in the terrain noting transects from the initial sample. Check condition C for this small number of transects, adding their neighbours to the sample if they fulfil C . For those neighbours that fulfil C , observe their neighbours too. Repeat the procedure until no more neighbours meet condition C .

Estimates of population parameters are obtained using the Jolly II method as described in section 2.3 for all designs. Since the correct number of animals in the study area is known during simulation, the root mean square error (RMSE) is calculated to give a measure of precision and allow comparisons between designs.

6.3. Results

Thirteen variables were used in this study to explain the distribution and presence of elephant and zebra. The variables ranged from those related to human habitation like presence of settlement, cultivation and water to vegetation cover, terrain variables like slope, height, aspect and variables related to rainfall like precipitation, evapotranspiration and temperature (Table 6.1). Sixteen different vegetation classes are distinguishable in the study area (Table 6.2). The high occurrence of observation points in acacia *drepanobila* bushland for both elephant and zebra suggests a preference of this vegetation class as a habitat for these two species. The zebra appears to prefer grassland as well, while the elephant favours open acacia *brevispica* thickets. From the results, these two species appear to be selective in their habitat preference.

Table 6.1. Variable used in the study, their names, descriptions and abbreviations.

VARIABLE	ABBREVIATION
Presence or absence of water	Water
Presence or absence of human settlement	Settlement
Presence or absence of cultivation	Cultivation
Vegetation class (see table 6.2)	Veg
Slope (levels 1-7)	Slope
Height (m)	Height
Aspect (levels 1-8)	Aspect
Annual precipitation (mm)	ANNPRE
Annual potential evapotranspiration (mm)	ANNPET
Ratio of ANNPRES and ANNPRE (ratio)	ANNPPE
Annual extreme temperature (°C)	ANNXT
Annual instantaneous temperature (°C)	ANNIT
Annual maximum extreme temperature (°C)	ANNMAXXT
Annual minimum instantaneous temperature (°C)	ANNMINIT

Correlation between explanatory variables and presence of both elephant and zebra is minimal with the highest observed value of -0.2 being between presence of zebra and human settlement. This suggests a negative relation between occurrence of zebra and that of human beings. The elephant shows even weaker correlation, perhaps because of its

Comparison of wildlife sampling procedures

sparse distribution in the study area. Climatic variables show strong correlation among themselves suggesting the use of only a few for modelling. In particular, the variable measuring annual potential evapotranspiration (ANNPET) appears to adequately explain the effects of other climatic variables. Slight changes in height appear to cause major changes in temperature and precipitation as indicated by high correlation between these variables. This may be an influence of the two mountain ranges in the study area.

Table 6.2. The distribution of elephant and zebra in different vegetation classes found in Laikipia. Ratio gives the proportion of points where each species was observed in relation to the points where the same species was observed in the whole study area.

Code	Vegetation class	Elephant		Zebra	
		Observations	Ratio	Observations	Ratio
1	Urban	0	0	0	0
2	Small scale cropland complex	0	0	1	0.001
3	Upland / riverine forest	5	0.03	6	0.01
4	Plantation forest	1	0.01	11	0.01
5	Degenerate dry upland forest	0	0	0	0
6	Leafy upland bush	1	0.01	3	0.003
7	Leafy bushland and thicket	27	0.17	48	0.05
8	Acacia <i>drepanobila</i> bushland	44	0.28	403	0.39
9	Acacia <i>seyal</i> bushland	4	0.03	89	0.09
10	Dwarf bush and grassland	14	0.09	55	0.05
11	Grassland	27	0.17	336	0.33
12	Open acacia <i>brevispica</i> thicket	33	0.21	30	0.03
13	Acacia <i>mellifera</i> bushland	4	0.03	37	0.04
14	Mpala scarpline vegetation	0	0	6	0.01
15	Bare rock, kopjes and escarpment	0	0	4	0.004
16	Water bodies and wetlands	0	0	3	0.003

(Adapted from Ojwang', 2000)

An attempt to reduce dimensionality in the data through a principal component analysis (PCA) was not very successful since 15 principle components (PCs) were required to explain 80% of sample variability or more instead of the acceptable, two or three PCs. Modelling the probability of observing an animal was carried out using actual variable values. Variables related to human presence and vegetation cover were significant in

modelling the probability of observing elephant and zebra (Table 6.3). Height, slope and evapotranspiration were also significant in determining the probability of observing zebra.

Table 6.3. Significant variables and their coefficients from the logistic regression analysis for modelling the probability of observing elephant and zebra, based on data from past surveys.

VARIABLE	COEFFICIENTS	
	Elephant ($R^2 = 0.63$)	Zebra ($R^2 = 0.71$)
Constant	-4.58	13.08
Presence of water within 2 km (WT2K)	-1.16	-1.62
Presence of water within 5 km (WT5K)	2.09	1.24
Presence of water within 10 km (WT10K)	4.13	-1.06
Presence of human settlement within 500 m (ST500)	-7.18	-2.03
Presence of human settlement within 5 km (ST5K)	3.97	4.21
Presence of cultivation within 500 m (CT500)	-6.24	-1.70
Within urban areas	3.36	0.51
Small scale cropland complex	-0.90	0.84
Upland / riverine forest	1.36	0.56
Plantation forest	0.60	0.62
Degenerate dry upland forest	0.66	0.50
Leafy upland bush	0.44	0.15
Leafy bushland and thicket	0.19	0.23
Acacia <i>drepanobila</i> bushland	0.07	0.19
Acacia <i>seyal</i> bushland	0.26	0.21
Dwarf bush and grassland	0.07	0.14
Grassland	0.19	-0.02
Open acacia <i>brevispica</i> thicket	0.01	0.01
Acacia <i>mellifera</i> bushland (Arid zone)	-0.45	0.02
Mpala scarpline vegetation	-0.39	0.07
Bare rock, kopjes and escarpment	-0.29	0.06
Slope		-0.01
Height		-0.0007
Annual potential evapotranspiration (ANNPET)		-0.008

Based on the sign of coefficients from the stepwise logistic regression, the presence of water, human settlement and cultivation appear to negatively influence the presence of both elephant and zebra at short distances. However, as distance increases, we notice the probability of observing both species increasing too. In particular, the three variables positively influence the presence of both animals at 5 km. The situation is

reversed at 10 km, where a positive coefficient (4.13) suggests an increase in the probability of observing the elephant as opposed to that of observing the zebra, which has a negative coefficient ($\beta_3 = -1.06$).

To study the effect of vegetation, its 16 levels were modelled as dummy variables (Hosmer and Lemeshow, 1989). The variable appears to have a strong influence on the presence of both elephant and zebra by having all its levels being significant in models for both species. The negative coefficient of grassland for zebra is a surprising result, especially if one considers the fact that 33% of observed zebra were found in this vegetation type, making it the second most favourite vegetation type for the zebra.

Results from the simulation are summarised by intensity (%), estimate of the population total (\hat{Y}_E) and the root mean square error (RMSE) of \hat{Y}_E for each sampling design. The results indicate consistent and similar estimates of the true population totals for all three sampling designs (Table 6.4). Although the population estimates for the adaptive design appear similar to those of the other two designs, they are more precise for both species. In general there is no difference in precision for both systematic and random designs. A gain in precision is observed with increasing intensity in almost all cases.

6.4. Discussion

The logistic models obtained yield results that are consistent with the ecology of the two animal species. For instance it is known that both elephant and zebra are water dependent species, with the elephant requiring daily drinking and the zebra not being able to go for more than three days without water (Haltenorth and Diller, 1996). This explains why both species have higher probabilities of being observed within 5 km of water points (Table 6.3). The negative influence at 2 km for both species can be explained by the fact that these same water points are also available for use by livestock, which creates competition with human beings. At further distances (i.e. 10 km), only the elephant predominates since it can walk for long distances compared to the zebra, which are generally found within 4 km of water points (Ottichilo, 2000).

There is further evidence of conflict for the use of habitat between man and beast in the variables measuring presence of settlement and cultivation. At short distances (0.5 km), both variables impact negatively on the presence of both animal species. However, their effect changes to a positive value as distance increases, showing a higher probability of animals occurring further from human settlements. This is bound to be a major topic of discussion in the near future for Laikipia district. Evidence already exists that the area occupied by wildlife in this area is reducing, mainly due to fragmentation caused by human resettlement (Georgiadis, 1997). Another impact of human encroachment is isolation of the ecosystem by surrounding wildlife areas with cultivated farms and excluding wildlife. This is particularly harmful for seasonally migrating species like the elephant and zebra.

Table 6.4. Summary of simulations giving the population estimates \hat{Y}_E , \hat{Y}_Z and the root mean square errors $RMSE(\hat{Y}_E)$ and $RMSE(\hat{Y}_Z)$ for the elephant and zebra, respectively. Results are given for three mean sampling intensities 3.27%, 6.54% and 12.75% for systematic and random sampling designs, while for the adaptive design, intensity averaged 3.19%, 6.58% and 13.21%.

SPECIES/STATISTICS					
Design	Intensity (%)	Elephant ($Y = 3495$)		Zebra ($Y = 33073$)	
		\hat{Y}_E	$RMSE(\hat{Y}_E)$	\hat{Y}_Z	$RMSE(\hat{Y}_Z)$
Systematic	3.27	3408	594	32981	9183
	6.34	3499	391	33750	6702
	12.75	3410	272	33859	3849
Random	3.27	3296	593	32782	10829
	6.34	3329	617	32913	6973
	12.75	3385	280	33092	3566
Adaptive	3.19	3476	483	33081	6219
	6.58	3493	278	33205	4229
	13.21	3468	169	34016	2531

Slope, height and annual potential evapotranspiration appear to affect only the presence of the zebra and not the elephant. Since the probability of observing an animal varies with its location and topographical feature of the study area, visibility is corrected for each species by values derived from the literature (Pollock and Kendall, 1987; Wint, 1998). The cut-off value used to accept a point as a suitable habitat is set to 0.3 because of weak correlation between the presence of the two species and variables influencing the distribution, as well as to account for other important variables not considered in model building.

Summaries of the 100 simulations per species validate known behaviour of the sampling designs used. For instance, an increase in sampling intensity leads to improvement in precision. However, this study shows that for clustered populations, the gain in precision may be marginal and should be weighed against the increase in cost accompanying increased sampling intensity.

Although there is a general gain in precision when using the adaptive sampling design, its implementation is difficult even under a simulation situation. Two possible sources of difficulty include:

- A clear definition of neighbouring units in transect sampling, and
- The choice of criteria for including neighbouring units in the sample, mainly due to great variation in transect counts for clustered populations.

Simulation studies like this are useful to determine this criteria. For this study area, the median was the most consistent statistic to use as a criterion for the widely distributed zebra but not the elephant, which needed observations of two or more animals before sampling the neighbouring transects. Using the median transect count as criterion helps to avoid the influence of clustering. However, this may not be possible in situations where no prior information exists. Furthermore, defining a different condition *C* for different animal species poses difficulties for multi-species surveys.

The increased difficulty associated with implementing the adaptive sampling design has to be carefully considered, with respect to the gain in precision, when adopting this sampling design for regular surveys of the tropical savannahs. This is pertinent, especially if one notes that many surveys are for monitoring purposes, where emphasis is on the study of trends in abundance as opposed to accurate one-off abundance estimates. In such cases, standardisation of current sampling strategies to ensure consistent results may be a cheaper option. There are situations, however, where a survey with high precision is important. Laikipia is one area where wildlife farming is undertaken, mainly on private ranches. Income is generated on these farms through wildlife utilisation such as cropping or fee-paying hunting. For these farmers, precise estimates are crucial and this is where the adaptive design is useful.

The aspect of visibility bias and correction factors is not treated in detail in this study. However, failure to see some of the animals during the survey is corrected by adjusting observed counts using proportions from the literature (Pollock and Kendall, 1987; Krebs, 1989; Wint, 1998). This is standard procedure in the absence of correction factors for animal species in the study area.

Populations are simulated as being independent of each other and interaction between the two animal species is not considered in this study. This is partly due to the knowledge that several animal species may generally occur in the same area but still be separated in terms of habitat requirements (Voeten and Prins, 1999). Furthermore, the distribution of herbivores in African ecosystems is modulated by resource partitioning and water requirements (Lamprey, 1963; Bell, 1970; Western, 1975; Voeten and Prins, 1999). Considering other none-competition variables that influence animal distribution like climatic conditions, relief and human settlement may eliminate this interaction. In any case, elephant and zebra are not generally known to compete for the same habitat since the elephant is both a grazer and a browser while the zebra is strictly a grazer (MacNoughton and Gorgiadis, 1986; Prins and Douglas-Hamilton, 1990).

The simulation in this study is based on a more detailed dataset collected from a previous survey conducted in ecosystem. This gives more information and allows a more complete modelling of the dynamics of the two animal species under study. Although the observations are obtained within a narrow temporal window, they still provide a representative picture of the spatial pattern. Consequently, simulations can only be as good as the data on which they are based. In this study, data on variables that are known to influence the distribution of elephant and zebra are used to build a suitability model. In particular, the logistic models only used data on variables collected during the survey, ignoring others that may have a greater influence on the distribution of these two species but were not observed. This is a common problem in modelling and is related to the fact that no single correct model exists for a given data set (Hosmer and Lemeshow, 1989; Burnham and Anderson, 1998). In this study, however, reasonable R^2 values (0.63 and 0.71 for elephant and zebra, respectively; Table 6.3) are obtained allowing us to treat both models as parsimonious.

6.5. Conclusions

The following conclusions can be drawn from this study:

- Data collected by a detailed survey provides a better basis for simulating animal dynamics allowing an investigation of different sampling designs.
- Population estimates obtained by the three sampling designs; random, systematic and adaptive do not differ significantly from each other.
- Estimates of precision from the adaptive design are higher than from the other two designs, which don't differ from each other.
- Presence of man has a negative impact on the presence of the two animal species.
- Common statistical techniques can be combined with GIS to gain more insight into the dynamics of wildlife populations.

Acknowledgements

The authors are grateful to DRSRS and Impala research station for permission to use their data. Discussions with Mohammed Said and Laurindo Jose Santos of ITC greatly improved the simulations.

CHAPTER 7*

An introductory treatment of decision making on statistical procedures used in wildlife management

* This chapter differs markedly from the previous five chapters by not being a fully-fledged treatment of a topic but an introduction to decision making.

Abstract

This study outlines some preliminary techniques useful in evaluating several sampling designs applicable in wildlife surveys. It introduces the multi-criteria evaluation method and applies it in assessing the performance of simple random sampling, systematic sampling and adaptive sampling designs. The study outlines how the choice of one of these sampling designs is affected by the importance we place on each criterion in terms of weights. The study also develops criterion scores that show to what degree sampling designs meet the given criteria. Illustrations are carried out using results of an analysis performed in an earlier chapter (chapter 6). The adaptive design is indicated to be the preferred design in eight out of twelve scenarios. However, the systematic sampling design also performs reasonably well and is actually found to be better in surveys that do not put much emphasis on precision as well as modelling of clustering in the population. Simple random sampling design is found to be the least preferable in all situations. In general, the final decision depends on the degree of importance placed on different criteria.

7.1. Introduction

The study of biological populations often requires an estimate of population density (D) or its size (Y). These parameters are commonly obtained through sampling procedures using aerial surveys. Due to mobility of wildlife populations, several sampling designs yield estimates of varying precision. Further, more data on wildlife populations are collected in space and time and their proper analysis requires appropriate statistical treatment. With an increasing number of possible statistical techniques that can be applied in a given situation, wildlife managers must weigh the advantages and disadvantages of adopting one statistical procedure over another, with considerations such as economics in mind. For example, selection of the right sampling and analysis design can result in significant savings in cost (Caughley, 1977; Norton-Griffiths, 1978; Thompson, 1992; Buckland, *et al.*, 1993).

This study distinguishes two levels of decision making. Decisions made on the selection of appropriate wildlife management options and those made on the selection of appropriate statistical procedures for wildlife data collection and analysis. An explanation of the relationship between the two levels is briefly outlined below.

Wildlife management involves looking after a wildlife population, where a population is defined as a group of coexisting individuals of the same species or type (Caughley and Sinclair, 1994). This management can either be manipulative or custodial. In manipulative management, something is done to the population or its habitat in order to change its size, while in custodial management, external influences on the population and its habitat are minimised with the prime aim of maintaining the population as it is. Custodial management is appropriate for protected areas where ecological processes determine the dynamics of the system (Caughley and Sinclair, 1994). However, its effectiveness in resolving conflicts between park managers and local communities is increasingly being questioned (Amuyunzu and Bijl, 1996; Mungatana, 1999).

Irrespective of the type of management, four options are distinguished in the management of a given wildlife population.

- To increase the wildlife population
- To decrease the wildlife population (culling)
- To harvest it in a sustainable manner (cropping)
- Do nothing to the population but monitor it from time to time (*Laissez-faire*).

These lead to three decisions that a wildlife manager needs to make from time to time in the course of managing a wildlife population.

- What is the desired management goal?
- What management option is appropriate?
- What method (action) is needed to best achieve the chosen option?

However, before a manager is in a position to make a decision on whether to increase, decrease, harvest or leave a wildlife population as it is, he needs to have information on the abundance and distribution on the population. Such information can only be obtained accurately, reliably and timely by application of proper statistical methods for data gathering and analysis. For wildlife management, this reduces to the choice of a suitable sampling strategy.

Further, in making wildlife management decisions, it is advantageous to outline all steps taken before arriving at the final decision. Wildlife managers must rely on their professional knowledge as well as advice sought from other professionals like system ecologists, statisticians, economists etc., for input on appropriate and acceptable procedures to adopt.

Statistical procedures have been used to support decision making in water resources (Walsh, 1993) or for smart farming (Stein and Van Groenigen, 1997). Amuyunzu and Bijl (1996) also integrated remote sensing and GIS in management decision support for managing the elephant, while Gorte (1996) used GIS to develop a statistical decision analysis for remotely sensed imagery.

The aim of this study is to introduce preliminary aspects on a framework for deciding on which design to apply in sampling wildlife populations using aerial surveys so as to obtain reliable estimates of abundance. The study focuses on situations where wildlife managers have to decide on the better sampling procedure to adopt given several alternatives. Quantitative and qualitative criteria are defined to differentiate the effect of adopting one of these designs. A number of hypothetical scenarios are outlined for different perspectives i.e. importance attributed to different criteria in order to illustrate the way each scenario leads to a different solution. The analysis is based on data collected in actual surveys of wildlife in Kenya.

7.2. Materials and methods

7.2.1. Decision making process

A decision is a specific commitment to action, usually in conjunction with a commitment of resources (Janssen, 1996). A decision making process is a set of actions and dynamic factors that begin with the identification of a stimuli for action (problem) and ends with a specific commitment to action (solution). Decision makers are individuals or groups of individuals who, directly or indirectly, provide value judgements or opinions on the decision process necessary to define and choose between alternative courses of action.

The decision making process can be thought of as a series of interrelated activities that lead to making a choice from a wide range of alternatives (Nyabenge, 1998). Several authors have broken this process into the intelligence, design and choice phases, which correspond to answers to the questions: What is the problem? What are the available

alternatives? Which is the best alternative? (Sharifi and Keulen, 1994; Janssen, 1996; Nyabenge, 1998).

In this study we focus on the third phase to help us select the best sampling design suitable for aerial surveys. We assume that any individual or group of individuals who have the mandate to carry out any of the four management options listed in section 7.1 for purposes of managing a designated wildlife population is a decision maker.

In general, the decision analysis in wildlife management must outline and assess the benefits resulting from making the right decision versus the penalties of getting it wrong (Voogd, 1983). This may take the form of examining social, political, biological and economic considerations and assigning them due weights. Different wildlife managers will assign different priorities to various possible criteria used to examine different statistical methods. We do not, however, carry out this pay-off table analysis in this study.

A suitable tool that inventories, classifies, analyses and conveniently arranges available information concerning different choice-possibilities is the multicriteria evaluation methods, which starts from a finite number of explicitly formulated criteria (Voogd, 1983). A basic feature of this method is the construction of a matrix that outlines a set of choice-possibilities that are evaluated against a set of criteria (Figure 7.1).

Development of an evaluation matrix in wildlife management involves a choice of alternative statistical methods with choice-possibilities representing different sampling designs while criteria represent standards by which these methods are judged.

		CHOICE POSSIBILITIES				
		A	B	C	P
CRITERIA	1	Criterion scores				
	2					
	3.					
	.					
	.					
	K					

Adapted from Voogd, (1983)

Figure 7.1. An evaluation matrix used in multicriteria evaluation methods for P choice possibilities that are evaluated using K evaluation criteria.

7.2.2. Sampling designs

Three common sampling designs suitable for aerial surveys of wildlife are simple random sampling (SRS), systematic reconnaissance flights (SRF) and adaptive sampling (AdS). With SRS, samples of size n out of a total of N units in the population, are selected with an equal inclusion probabilities of $\pi = n/N$, while SRF is a modification of systematic sampling, where transects are chosen at regular intervals. With AdS, an initial sample of size n_0 is selected by means of SRS and a condition $C = \{y: y \geq c\}$ is defined, for some constant c and a number y of observed animals in the transect, to determine whether transects neighbouring those in the initial sample will be added to the sample or not. These designs are defined and described in greater detail in chapter 2 of this thesis and their performance compared in chapter 6.

7.2.3. An evaluation matrix for wildlife management

In setting up an evaluation matrix, the three sampling designs represent three choice-possibilities for a wildlife manager. Criteria for this matrix are derived from information requirements of wildlife managers. Wildlife management requires reliable information on numbers, dynamics and movements of wildlife population within a clearly demarcated area. Such information is obtained with considerations of cost and precision, leading to three level 1 criteria that can be used to assess the three sampling designs listed above i.e. Precision, Cost and Information output.

Sample size, stratification and ability of the sampling design to adjust to observed population patterns influence precision, while two types of costs can be distinguished, i.e. fixed costs and variable costs. The type of information obtained from a survey should lead to proper modelling of the distribution and movements of wildlife under study. These can be measured in the ability of the different sampling designs to identify areas of high species diversity, to be understood by people using the results, to provide data to quantify species interaction and to identify trends over time. These represent level 2 criteria. Level 1 criteria provide an objective way evaluate the three sampling designs, all of which are repeatable.

The next step in defining the evaluation matrix is to determine a set of criterion scores, which determine to what degree a choice probability fulfils a certain objective (criteria). Criterion scores are often derived by means of a thorough investigation like an impact or effect analysis. They may also be obtained from intuitive estimation leading to qualitative rankings. This study assumes the latter approach for defining most of the criterion scores.

Based on results of chapter 6, sampling intensities for the three sampling designs (Table 6.4) are used as quantitative criterion scores for sample size. A higher sampling intensity implies larger sample and a subsequently higher precision. Fixed costs are estimated based on planning costs, personnel costs and the cost of processing information from the observed data. Variable costs include costs of going to and from the study area, costs for moving from one sampling unit to another and the cost of observing the sample.

Scores for both types of cost are ranked as low (3), medium (2) and high (1). Other scores in the evaluation matrix are derived as binary scores to indicate the capability of the sampling design to meet the desired objective.

The final definition in the evaluation matrix is to assign weights to each main criterion (level 1) and subcriteria (levels 2, 3, ..., k) so as to reflect the importance attached to each criterion by decision makers. These will vary from manager to manager and are often dictated by the main objective of the survey. The weights at each level must sum up to 1 (100%) as a way of standardisation.

In order to arrive at a complete evaluation of the three sampling designs, a weighted summation evaluation procedure is followed, whereby criterion scores are multiplied by the corresponding criterion weights and then summed for each design. Other arithmetic procedures exist but may not have the simplicity that weighted summation has (Voogd, 1983). To eliminate the effect of scale, criteria scores are often standardised by dividing each score by the maximum score for every criterion.

7.3. Results

To illustrate use of the evaluation matrix, consider an example where a survey is carried out to estimate the total number of animals in a given study area with some consideration on the cost of the survey. A manager in charge of such a survey may attach more importance on precision by assigning it a weight of 0.5 and less importance on the type of information obtained from the survey by assigning it a weight of 0.2. Since weights commonly have to add to 1, the remaining 0.3 goes towards the cost criteria. Within precision, the number of sampled units (sample size) may be considered to be of greater influence on precision than stratification. The manager also knows that there many clustered species in the population and would give a higher weight to a design with the ability to adjust sampling based on observed patterns. He therefore assigns the weights 0.6, 0.1 and 0.3 to correspond to sample size, stratification and accounting for clustering, respectively. Further, there is no reason to consider variable costs as more important than fixed costs and thus an equal weight of 0.5 is assigned to both. The same applies to the information output types, with each being assigned an equal weight of 0.25.

We consider the lowest sampling intensities given for each design in table 6.4 and obtain the evaluation matrix in table 7.1. The problem is to assess how each sampling design is likely to achieve the stated objectives by assigning it scores according to its importance. Most criterion scores are binary with 0 for no and 1 for yes. All scores are standardised by dividing them with the maximum value per criterion. For this configuration, the weighted summation evaluation procedure yields the following appraisal scores:

$$\text{SRS: } 0.5*0.5*1 + 0.5*0.1*1 + 0 + 0.3*0.5*1 + 0.3*0.5*0.667 + 0.2*0.25*1*1 = 0.60$$

Introduction to decision making in sampling procedures

$$\text{SRF: } 0.5 \times 0.5 \times 1 + 0.5 \times 0.1 \times 1 + 0 + 0.3 \times 0.5 \times 1 + 0.3 \times 0.5 \times 1 + 0.2 \times 0.25 \times 1 \times 3 = 0.75$$

$$\text{AdS: } 0.5 \times 0.5 \times 0.976 + 0 + 0.5 \times 0.4 \times 1 + 0.3 \times 0.5 \times 0.5 + 0.3 \times 0.5 \times 0.333 + 0.2 \times 0.25 \times 1 \times 3 = 0.72.$$

These indicate SRF to be a better design for the given weight configuration and sampling intensity. Next, the level 1 weights were fixed at 0.5, 0.3 and 0.2 for precision, cost and information output respectively, while level 2 weights of sample size, stratification and clustering were varied. At some stage, SRF was observed to be better than AdS (Table 7.2). Similar observation was made when the situation was reversed and level 1 weights were varied.

Table 7.1. An evaluation matrix to examine possible sampling designs i.e. Simple Random (SRS), Systematic Reconnaissance (SRF) and Adaptive sampling (AdS) against management objectives in the sampling wildlife populations. Raw scores are given in brackets where necessary.

		Sampling designs			
Level 1 weights	Criteria and respective level 2 weights		SRS	SRF	AdS
<hr/>					
	Precision				
0.5	0.5	- Sample size	1 (3.27)	1 (3.27)	0.976 (3.19)
	0.1	- Allows stratification	1	1	0
	0.4	- Adjusts to observed patterns	0	0	1
<hr/>					
	Cost				
0.3	0.5	- Fixed costs	1 (2)	1 (2)	0.5 (1)
	0.5	- Variable costs	0.667 (2)	1 (3)	0.333 (1)
<hr/>					
	Information output				
0.2	0.25	- Identify species diversity	0	1	1
	0.25	- Easy to understand	1	1	0
	0.25	- Identify species interaction	0	0	1
	0.25	- Capture trends	0	1	1

For fixed level 1 weights and low priority for stratification, AdS appears to decrease in importance with a decrease in the weight on the criterion on the ability to use information observed in the sample.

On the other hand, when level 2 weights were varied for fixed effects of level 2 criteria, AdS was preferable for surveys where cost was not of great importance, while SRF appeared important in surveys where precision and the type of information output were not

heavily weighted (Table 7.3). Both AdS and SRF performed well in a situation where all three were equally weighted with evaluation scores of 0.69 and 0.70, respectively. SRS had the lowest appraisal scores in all situations.

Table 7.2. Level 1 weights are fixed at 0.5, 0.3, 0.2 for precision, cost and information output, respectively, while only those corresponding to sample size, stratification and ability to detect clustering are varied.

Design	Level 1 Weights					
	.2, .1, .7	.3, .1, .6	.4, .1, .5	.5, .1, .4	.6, .1, .3	.7, .1, .2
SRS	0.45	0.49	0.54	0.59	0.64	0.69
SRF	0.55	0.59	0.64	0.69	0.73	0.79
AdS	0.67	0.70	0.72	0.73	0.74	0.73

Table 7.3. The weights corresponding to sample size, stratification and ability to detect clustering are fixed at 0.5, 0.1, 0.4, respectively, while weights for precision, cost and information output are varied.

Design	Level 2 weights					
	.5, .1, .4	.4, .1, .5	.5, .3, .2	.35, .35, .3	.5, .4, .1	.1, .5, .4
SRS	0.47	0.44	0.59	0.57	0.65	0.57
SRF	0.59	0.58	0.69	0.70	0.74	0.76
AdS	0.79	0.78	0.73	0.69	0.69	0.60

7.4. Concluding remarks

In this study, we develop a preliminary multi-criteria evaluation procedure that can be used to distinguish between three sampling designs. We identify appropriate criteria by which this can be done, derive criterion scores and choose weights to illustrate the method with results from a previous analysis. We use a weighted summation evaluation, but other procedures are also applicable.

There is need to develop methods by which wildlife managers can assess performance of different statistical procedures. The whole success of wildlife surveys rests on the correct application of sampling techniques. There is a running debate in the survey

literature on the advantages and disadvantages of both simple random and systematic sampling (Jolly, 1981; Norton-Griffiths, 1981). Addition of a third sampling design would make the debate complex and thus the need to an objective method to of choosing one design over the others.

Although the subjective method of choosing weights might appear a disadvantage at first, this is actually where the strength of the method lies. The fact that different decision makers can choose different values, allows them to explore several scenarios and see the results of their choices (Voogd, 1983).

However, there is still room to improve the determination of criterion scores. This study mainly adopted binary scores to outline the methods. In the presence of other information or data, quantitative scores may be assumed. It is also possible to extend the number of criteria from the current three to any number that a wildlife manager desires leading to a straightforward extension of the method. Indeed for other choice-possibilities, it may be relevant to take political, economic and cultural considerations into account.

The results of the analysis show that neither systematic nor the adaptive design is superior in all cases. This is consistent with our expectation because whereas the adaptive design is well suited for surveys of clustered species, it may not work very well for other species. A choice between the two must, therefore, be made with a clear purpose in mind.

In conclusion, decisions made in wildlife management would benefit a lot from a well formulated decision analysis process. This must outline the available alternatives to choose from and a list of several criteria by which the alternatives will be judged. Three such criteria, precision, cost and type of information obtained were able to select between simple random sampling, systematic sampling and an adaptive sampling design, with adaptive and systematic sampling designs consistently doing better than simple random sampling.

CHAPTER 8

Conclusions and Recommendations

8.1. Conclusions

This study deals with different aspects of obtaining efficient, reliable and timely information from surveys of wildlife populations. It is based on data obtained from surveys conducted in Masai Mara and Laikipia Ecosystems in Kenya. The major objective of estimating wildlife abundance and modelling wildlife distribution serves as a common thread. This concluding chapter gives the main conclusions of the study. The research questions posed and answered in different chapters of the thesis are related to the main question of improving information obtained from wildlife surveys for better management of wildlife populations.

This chapter outlines the main objectives of the study as formulated in section 1.3 and checks these against summaries of research findings. It concludes by giving recommendations for further research.

1st Objective: *To investigate current wildlife sampling procedures and propose an improvement.*

This study investigated and discussed advantages and disadvantages of two common sampling designs in wildlife surveys: simple random sampling and a modified systematic sampling design known as systematic reconnaissance flights. It proposed an adaptive sampling design as an alternative that takes clustering of wildlife populations into account and uses criteria on observed animal counts to maximise sampling information.

Common sampling designs are defined such that animals observed outside sampling units during the survey are not counted. Although this gives unbiased estimates, it is a waste of sampled information. A pragmatic approach is to use an adaptive design that responds to changes in distribution caused by stochastic processes leading to the observed wildlife patterns. Such an improvement is obtained from the developed adaptive design, which takes clustering in distribution of wildlife into account (section 2.2.4). Unbiased estimators for both the univariate and multivariate case are developed

The adaptive sampling method gives estimates with lower standard errors than those obtained by the commonly used systematic sampling design for both univariate and multivariate cases. It is also more efficient than the other design, showing a decrease in standard error of up to 37% for some species. Further, it uses information from observed patterns, thereby utilising more sample information. The adaptive sampling method performs better for clustered and highly abundant species as well as for sparse populations. Extension to a multivariate setting does not substantially improve estimates but leads to more efficient data use.

2nd Objective: *To model wildlife abundance and distribution using statistical techniques and geographic information system (GIS).*

Most data from wildlife surveys are stored in geographic information systems (GIS), allowing the use of GIS techniques to combine layers of information. Generalised linear modelling (GLM) is used in a GIS to model abundance and distribution of wildlife in space and time, focusing on the distribution of elephant during nine successive surveys. Their temporal and spatial distribution is related to 12 environmental variables. A principal component analysis identifies five principle components, thereby reducing dimensionality in the data. The number of variables explaining elephant abundance is subject to large variations during successive surveys with a minimum of four and maximum of eight variables. In general, variables related to the protected reserve have more influence on elephant abundance.

A simple distance measure that calculates spatial correlation and quantifies clustering for different animal species is developed for wildlife data obtained from aerial surveys. The procedure is illustrated by data on elephant, kongoni, wildebeest and zebra and it is able to capture clustering in wildebeest and zebra distributions, which have over 20 times more observations falling within short distances compared to the other two species.

Lack of an elegant and systematic method to simultaneously analyse spatial and temporal changes for wildlife data is a major deficiency in standard GIS packages. The cause is that most GIS software can only describe and display geographic data, while lacking the ability to develop new hypotheses. Use of GLM in a GIS allows to model effects of environmental factors on a quantitative response variable like animal counts. This chapter demonstrates how to take advantage of spatial statistical routines, especially in a GIS environment. Extension standard statistical procedures may therefore lead to a better understanding of spatial and temporal relationships in wildlife populations.

3rd Objective: *To use spatio-temporal procedures to model changes in wildlife populations in space and time.*

A manager of a wildlife population needs tools and techniques that can be used to answer the following three relevant basic questions:

- How many animals are in my reserve?
- Where are they?
- Are their numbers changing with time?

These questions concern estimation of abundance and modelling of the spatial distribution. The second and third questions require statistical techniques to model changes in wildlife

populations in space and time. The study shows how using prior information in Markov Chains Monte Carlo (MCMC) modelling improves precision for estimates of population totals obtained from aerial wildlife surveys. As an extension of the previous chapter, animal counts are related to the area of observed strip transects in a generalised linear model for better estimation of animal density. This leads to a gain in precision of up to 35 % in estimating elephant abundance (Table 4.1).

Such an increase in precision is not possible with the Jolly II method, which is now a standard tool for estimation. Indeed the main criticism of the Jolly II method are its large standard errors, which are a direct consequence of differences in sizes of sampling units and observed counts, especially for gregarious species like wildebeest and zebra. These differences invalidate the assumption of a through-the-origin regression and proportional variance, which are necessary assumptions for ratio estimators like the Jolly II.

Space-time interactions by means of modelling autocorrelation are assessed before estimating effects of environmental variables on observed counts. The space-time analysis reveals significant changes in the spatial patterns of elephant counts over a 20-year period. Strong interactions are observed within 5 km and 6 months space and time separations, respectively (Table 4.2), while spatial dependence accounts for significant variation in modelling observed counts of wildlife species. Therefore, it should always be taken into account when modelling animal abundance.

A diversity index is developed that is suitable for aerial surveys of large herbivores. This index is sensitive to both high abundance and species richness and is also able to capture year to year variation in diversity. It indicates an overall marginal decrease in diversity for large herbivores in the Masai Mara ecosystem. Both the space-time analyses and diversity index are easy to compute and therefore provide simple tools for rapid decision making as far as trends in wildlife populations are concerned.

4th Objective: *To model wildlife distribution using spatial point patterns.*

A more detailed data set, in which exact geographical locations of groups of animal have been recorded, is available from a second study area. This is used to model spatial point patterns of six ungulates in the area. Differences in abundance and spatial distribution of wildlife species are related to their food requirements and how this is distributed. Interaction between species is generally not strong as shown by the X^2 statistic and the multivariate K -function. Spatial point pattern analysis exists that can be very useful to study factors influencing spatial and temporal distribution of animals and in generating information necessary for management of wildlife populations. The use of these techniques, however, depend on the availability of better data than currently available from aerial surveys and calls for an increase in the use of GPS technology during collection of animal and environmental data. Spatial point pattern analysis provides a tool that can be used to

study wildlife species grouping and relate this to food availability. Such information is useful for designating cropping and hunting areas in areas practising consumptive utilisation.

5th Objective: *To compare performance of proposed sampling procedures to established sampling strategies.*

A simulation study compares performance of the conventional random and systematic designs to the proposed adaptive design. Based on observed spatial patterns, the distribution of animals is simulated under a GIS, where it is related to other environmental variables like vegetation type to build a statistical model that predicts the presence of a wildlife species. The resultant spatial patterns are then sampled using the different sampling designs and the data analysed. This is repeated 100 times for two key species, the elephant and zebra. Mean of population total estimates are compared and performance of each design is compared, at different sampling intensities, by means of the root mean squared error (RMSE).

Human presence has a negative impact on the presence of the elephant and zebra in this ecosystem. Population estimates obtained by the three sampling designs do not differ significantly from each other. Estimates of precision for the adaptive design is higher than that of the other two designs, thereby validating the use of adaptive sampling design for sampling clustered population. Common statistical techniques can be therefore be used in a GIS to gain more insight into the dynamics of wildlife populations. This works well with the availability of data collected by a detailed survey, which provides a better basis for simulating animal dynamics.

To help a wildlife manager choose between different sampling techniques and other statistical methods clear steps must be provided to choose a particular method. This study describes different scenarios encountered when making a decision related to the adoption of a statistical procedure useful in the management of wildlife. In particular, selection of a suitable sampling design and the analysis of subsequent data can result in significant savings in cost.

8.2. Recommendations

This thesis presents an adaptive sampling design as an improvement over the conventional systematic reconnaissance flights. The development of the sampling strategy is treated in somewhat greater detail for the univariate case as opposed to the multivariate case. This extension to the multivariate case is one area that would benefit from further research. Another related area of future research relates to the definition of criteria that

Conclusions and Recommendations

determine when to sample neighbouring units. In the validation analyses of chapter 6, it is observed that different criteria may be appropriate for different wildlife species. This presents a drawback, especially for multi-species surveys, which are common in the tropics. More research needs to be carried out to determine standard criteria for certain species grouping. Finally, there is urgent need to carry out detailed field tests to quantify various sources and forms of observer bias in wildlife surveys. It is hoped that this thesis presents a useful step towards the integration of statistical developments and ecology for the improvement of wildlife management.

References & Author bibliography

References

- Agresti, A. (1990). *Categorical data analysis*. Wiley: New York.
- Amuyunzu, L. C. (1984). *The use of remote sensing techniques in resource inventory: A case study of Narok District, Kenya*. Msc. Thesis, International Institute for Aerospace Survey and Earth Sciences (ITC): Enschede, The Netherlands.
- Amuyunzu, L. C. and Bijl, B. C. (1996). Integration of Remote Sensing and GIS for Management Decision support in protected areas: Evaluating and monitoring of wildlife habitats. *Conference on the application of remotely sensed data and GIS in environmental and natural resources assessment in Africa*. Harare, Zimbabwe, March 15-22, 1996. Pp. 297-306.
- Anselin, L. (1993). Discrete Space Autoregressive Models. In *Environmental modelling with GIS*, M. F. Goodchild, B. O. Parks and L. T. Steyaert, (eds). Oxford University Press: New York. Pp. 454-469.
- Augustin, N. H., Muggleston, M. A. and Buckland, S. T. (1996). An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology* 33:339-347.
- Bell, R. V. (1970). The use of the herb layer by grazing ungulates in the Serengeti. In *Animal population in relation to their food sources*, A. Watson (ed.). Blackwell: Oxford. Pp. 111-124.
- Berger, P. (1989). Rainfall and Agro-climatology of Laikipia plateau, Kenya. ASS Vol. A7. University of Bern: Bern, Switzerland.
- Berzuini, C. (1996). Medical monitoring. In *Markov Chain Monte Carlo in Practice*, W.R. Gilks, S. Richardson and D.J. Spiegelhalter (eds). Chapman & Hall: London. Pp. 321-337.
- Brooks, S. P. (1998). Markov chain Monte Carlo method and its application. *The Statistician* 47(1):69-100.
- Broten, M. D., and Said, M., (1995). Population Trends of Ungulates in and around Kenya's Masai Mara Reserve. In *Serengeti II: Dynamics, Management, and Conservation of an Ecosystem*, A. R. E. Sinclair and P. Arcese (eds). University of Chicago Press: Chicago. Pp. 169-193.
- Brown, L. H. and Cocheme, J. (1973). *A study of the Agro-Climatology of the highlands of Eastern Africa*. Tech. Note 125. WHO: Geneva.
- Buckland, S. T. and Elston, D. A. (1993). Empirical models for the spatial distribution of wildlife. *Journal of Applied Ecology* 30:478-495.
- Buckland, S. T., Anderson, D. R., Burnham, K. P. and Laake, J. L. (1993). *Distance sampling: Estimating abundance of biological populations*. Chapman & Hall: London; New York.
- Burnham, K. P. and Anderson, D. R. (1998). *Model selection and inference - A practical information-Theoretic approach*. Springer-Verlag: New York.
- Burrough, P. A. and McDonnel, R. A. (1998). *Principles of Geographic Information*

References

- Systems*. Oxford University Press: Oxford.
- CART, 1984, *CART*. California Statistical Software: Lafayette.
- Caughley, G. (1974). Bias in Aerial Survey. *Journal of Wildlife Management*. **38**:921-933.
- Caughley, G. (1977). Sampling in Aerial Survey. *Journal of Wildlife Management*. **41**(4):605-615.
- Caughley, G. and Sinclair, A. R. E. (1994). *Wildlife Ecology and Management*. Blackwell Science: Cambridge.
- Churcher, C. (1982). Grevy's: The Other Zebra. *Swara*, **5** (1): 12-18.
- Cochran, W. G. (1977). *Sampling Techniques*, 3rd edition. John Wiley and Sons: New York.
- Cressie, N. A. C. (1993). *Statistics for spatial data*. Wiley: New York.
- De Leeuw, J., Prins, H. H. T., Njuguna, E. C., Said, M. Y and de By, R. A. (1998). *Interpretation of DRSRS animal counts (1977-1997) in the rangeland districts of Kenya*. DRSRS: Nairobi, Kenya/ITC: Enschede, The Netherlands.
- Delany, M. J., (1982). *Mammals Ecology*. Blackie & Sons Limited: Glasgow.
- Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*. Academic Press: London.
- Dobson, A. J., (1990). *An introduction to generalised linear models*. Chapman and Hall: London; New York.
- Dublin, H. T. and Douglas-Hamilton, I., (1987). Status and trends of elephant in the Serengeti-Mara ecosystem, *African Journal of Ecology*, **25**:19-33.
- Dublin, H. T., (1995). Vegetation dynamics in the Serengeti-Mara Ecosystem: The role of the Elephant, Fire, and Other Factors, In *Serengeti II: Dynamics, Management and Conservation of an Ecosystem*. A. R. E. Sinclair and P. Arcese (eds). University of Chicago press: Chicago. Pp. 71-90.
- ESRI, (1996). ArcView GIS version 3.0a, Environmental Systems Research Institute, Inc.
- Estes, R.D. (1974). Social organisation of the African Bovidae. In: *The behaviour of ungulates and its relation to management*, Vol. I. V. Geist and F. Walther (eds). International Union for the Conservation of Nature and Natural Resources: Morges. Pp. 166-205.
- Fryxell, J. M., (1995). Aggregation and Migration of Grazing Ungulates in relation to Resources and predators, In *Serengeti II: Dynamics, Management, and Conservation of an Ecosystem*. A. R. E. Sinclair and P. Arcese (eds). University of Chicago Press: Chicago. Pp. 257-273.
- Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995). *Bayesian data analysis*, Chapman & Hall: London.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the

References

- Bayesian restoration of images. *IEEE Trans. Pattn. Anal. Mach. Intel.*, 6:721-741.
- Georgiadis, N. (1997). *Numbers and Distribution of Wild Herbivores in Laikipia district*. Laikipia Wildlife Forum: Laikipia.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). Introducing Markov Chain Monte Carlo. In *Markov Chain Monte Carlo in Practice*. W. R. Gilks S. Richardson and D. J. Spiegelhalter (eds). Chapman & Hall: London. Pp. 1-19.
- Ginsberg, J. R. (1985). The Grevy's zebra, a New Look at its Conservation Status. *Wildlife News*. 20(1):5-9.
- Ginsberg, J. R. (1988). *Social Organisation and Mating Strategies of an Arid Adapted Equid: The Grevy's zebra*. PhD Thesis. Princeton University.
- GOK, (Government of Kenya). (1984). *Narok district Development Plan (1984/87)*, Ministry of Finance and Planning. Government Printer: Nairobi.
- GOK, (Government of Kenya). (1998). *Economic Survey*. Ministry of Planning and National Development. Government Printer: Nairobi.
- Gorte, B. (1996). Efficient use of GIS in Statistical Decision analysis of remotely sensed imagery. *Proceedings of the workshop on spatial statistics from remote sensing*, 16th October 1996. International Institute for Aerospace Survey and Earth Sciences (ITC): Enschede, The Netherlands. Pp. 1-4.
- Green, P. J. (1996). MCMC in image analysis. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds). Chapman & Hall: London. Pp. 381-399.
- Green, P. J. and Silverman, B. W. (1994). *Non-parametric regression and generalised linear models: a roughness penalty approach*. Chapman & Hall: London.
- Grunblatt, J., Said, M. Y., Wargute, P., and Kifugo, S. C., (1995). *DRSRS Aerial Surveys Database*. Department of Resource Surveys and Remote Sensing, Ministry of Planning and National Development: Nairobi.
- Grunblatt, J., Said, M. Y. and Wargute, P. (1996). *DRSRS National Rangelands Report: Summary of Population Estimates for Wildlife and Livestock. Kenyan Rangelands 1977-1994*. Department of Resource Surveys and Remote Sensing, Ministry of Planning and National Development: Nairobi.
- Haile, D. G. and Weidhaas, D. E. (1977). Computer simulation of Mosquito populations, *Anopheles Albimanus*, for comparing the effectiveness of control technologies. *Journal of Medical Entomology*. 13:553-567.
- Haltenorth, T. and Diller, H. (1996). *Mammals of Africa including Madagascar*. HarperCollins: London.
- Hansen, R. M., Mugambi, M. M. and Bauni, S. M. (1985). Diets and Trophic Ranking of Ungulates of the Northern Serengeti. *Journal of Wildlife Management*. 49(3): 823-829.

References

- Hosmer, D. W. and Lemeshow, S. (1989). *Applied logistic regression*. Wiley: New York.
- Isaaks, E. H. and Shrivastava, R. M. (1989). *Applied Geostatistics (An introduction)*. Oxford University Press: New York, Oxford.
- Jaetzold, R., and Schmidt, H., (1983). *Farm management Handbook of Kenya: Central Kenya, Volume IIB*. Rosdorf: Germany.
- Jaman, P. J. (1974). The social organisation of antelope in relation to their ecology. *Behaviour*, **48**:215-267.
- Janssen, R. (1996). *Multi Objective Decision Support for Environmental Management*. Dordrecht: Kluwer Academic Publishers.
- Jolly, G. M. (1969a). Sampling methods for aerial censuses of wildlife populations. *East African Agricultural and Forestry Journal*. **34**:46-49.
- Jolly, G. M. (1969b). The treatment of errors in aerial counts of wildlife populations. *East African Agricultural and Forestry Journal*. **34**:50-56.
- Jolly, G. M. (1981). A review of the sampling methods used in aerial survey. *Low-Level Aerial Survey Techniques. ILCA Monograph No. 4*. ILCA: Addis Ababa.
- Kaluzny, S. P., Vega, S. C., Cardoso, T. P. and Shelly, A. A. (1998). S+ SpatialStats: User's manual for Windows® and Unix®. Mathsoft Inc.
- Klingel, H. (1974). Social organization and behaviour of the Grevy's zebra. *Z. Tierpsychol.* **36**:36-70.
- Krebs, C. J. (1989). *Ecological Methodology*. HarperCollins: New York.
- Lamprey, R. H. (1984). *Maasai Impact on Kenya Savannah vegetation: A remote sensing approach*. PhD Thesis. University of Aston in Birmingham: Birmingham, England.
- Legendre, P. and Legendre, L. (1998). *Numerical ecology*. Elsevier Science: Oxford.
- Li, W., Wang, Z., Ma, Z., and Tang, H., (1997). A regression model for the spatial distribution of red-crown crane in Yancheng Biosphere Reserve, China, *Ecological Modelling*, **103**:115-121.
- Lusigi, W. (1986). *Swara Magazine*: East African Wildlife Society: Nairobi
- Maddock, L. (1979) the "Migration" and grazing succession. In *Serengeti: Dynamics of an Ecosystem*. A. R. E. Sinclair and M. Norton-Griffiths (eds). University of Chicago Press: Chicago. Pp. 104-129.
- Magurran, A.E. (1988). *Ecological Diversity and its Measurement*. Croom Helm: London.
- Mathsoft, (1997). S-Plus 4.5 Professional, Mathsoft, Inc.
- MathSoft (1999). S-PLUS 2000 Professional, Release 1. MathSoft Inc.
- McCullagh, P., and Nelder, J. A., (1989). *Generalised linear models*. Chapman and Hall: London, New York.
- McNaughton, S. J. and Georgiadis, N. (1986). Ecology of African grazing and Browsing mammals. *Annual Review of Ecological Systems* **17**:39-65.

References

- Mungatana, E. D. (1999). *The welfare economics of protected areas: The case of Kakamega Forest National Reserve, Kenya*. PhD Thesis. Technical university Dresden, Germany.
- Murray, G. M., (1995). Specific nutrient requirements of the Wildebeest, In *Serengeti II: Dynamics, Management, and Conservation of an Ecosystem*. A. R. E. Sinclair and P. Arcese (eds). University of Chicago Press: Chicago. Pp. 231-256.
- Nichols, A. O., (1989). How to make biological; surveys go further with Generalised Linear Models, *Biological Conservation*, **50**:51-75.
- Norton, T. W. and Williams, J. E. (1992). Habitat modelling and simulation for nature conservation: A need to deal systematically with uncertainty. *Mathematics and Computers in Simulation*. 33:379-384.
- Norton-Griffiths, M. (1978). *Counting animals*. Handbook 1, 2nd edition. African Wildlife Foundation: Nairobi.
- Norton-Griffiths, M. (1981). Unstratified Systematic Sampling: Rationale and method. *Low-Level Aerial Survey Techniques. ILCA Monograph No. 4*. ILCA: Addis Ababa.
- Nyabenge, M. O. (1998). *Developing a prototype Multilevel Decision Support System for Land use planning in the Rangelands: A case study in Kenya*. MSc. Thesis. International Institute for Aerospace Survey and Earth Sciences (ITC): Enschede, The Netherlands.
- Ojwang', G. O. (2000). *On sampling variability of wildlife cropping-quota allocated on the basis of airborne sample surveys*. MSc. Thesis. International Institute for Aerospace Survey and Earth Sciences (ITC): Enschede, The Netherlands.
- Onyeausi, A. E. (1986). Measurements of impact of tourist off-road driving on grasslands in Masai Mara national reserve, Kenya: a simulation approach. *Environmental conservation*. 13:325-329.
- Osborne, P. E., and Tigar, B. J., (1992). Interpreting bird atlas data using logistic models: an example from Lesotho, Southern Africa, *Journal of Applied Ecology*, **29**:55-62.
- Ottichilo, W. K., (2000). *Wildlife Dynamics: An analysis of change in the Masai Mara Ecosystem of Kenya*. PhD Thesis. Wageningen university and Institute for Aerospace Survey and Earth Sciences (ITC): Enschede, The Netherlands.
- Patil, G. P. and Taillie, C. (1979). An overview of diversity. In *Ecological diversity in theory and practice*, J. F. Grassle, G. N. Patil, W. Smith and C. Taillie (eds), International Co-operative Publishing House: Fairland, Maryland. Pp. 3-27.
- Pereira, M. C. J., and Itami, R. M., (1991). GIS-based habitat modelling using logistic multiple regression: a study of the Graham red squirrel, *Journal of Photogrammetric engineering and Remote Sensing*, **51**(11):1475-1486.

References

- Pollock, K. H. and Kendall, W. L. (1987). Visibility Bias in Aerial Surveys: A review of Estimation Procedures. *Journal of Wildlife Management*. **51**:502-510.
- Pratt, D. J. and Gwynne, M. D. (1977). *Rangeland Management and Ecology in East Africa*. Hodder & Stoughton: London.
- Prins, H. H. T. and Douglas-Hamilton, I. (1990). Stability and multi-species assemblage of large herbivores in East Africa. *Oecologia*. **83**:392-400.
- Prins, H. H. T. and Olff, H. (1998). Species richness of African grazer assemblages: Towards a functional explanation. In *Dynamics of Tropical communities*, D. M. Newbery, H. H. T. Prins and N. Brown (eds). Blackwell Science: Oxford. Pp. 449-490.
- Ripley, B. D. (1981). *Spatial Statistics*. Wiley: New York.
- Rasch, D., Verdooren, L. R. and Gowers, J. I. (1999). *Fundamentals in the design and analysis of experiments and surveys*. Oldenbourg: München
- Robinson, S. (1994). *Successful simulation: A practical approach to simulation projects*. McGraw-Hill Book Company, London. 246 pp.
- Rosero, F. R., (1997). *Environmental factors influencing the spatial distribution of Livestock and Wildlife in Mara area of Narok district – Kenya*. M.Sc. Thesis. Institute for Aerospace Survey and Earth Sciences (ITC): Enschede, The Netherlands.
- Rossi, R. E. (1992). Geostatistical tools for modelling and interpreting ecological spatial dependence. *Ecological monographs*. **62**(2): 277-314.
- Rowen, M. (1992). Mother-infant behaviour and the ecology of the Grevy's zebra (*Equus grevi*). PhD Thesis, Yale University.
- Said, M. Y., (1993). *Modelling relationships between large herbivores, vegetation. Soils and climate in the Mara Ecosystem: A GIS approach*. M.Sc. Thesis. Institute for Aerospace Survey and Earth Sciences (ITC): Enschede, The Netherlands.
- Seber, G. A. F. (1982). *The estimation of animal abundance and related parameters*. 2nd edition. Edward Arnold: London.
- Sharifi, M. A. and Keulen, H. V. (1994). A decision Support System for Land Use Planning at Farm Enterprise Level. *Agricultural Systems* **45**, 239-257.
- Sinclair, A. R. E., (1975). The resource limitations of trophic levels in tropical grassland ecosystem, *Journal of animal ecology*. **44**:497-520.
- Sinclair, A.R.E. (1995a). Serengeti Past and Present. In *Serengeti II: Dynamics, Management and Conservation of an ecosystem*. A. R. E. Sinclair and P. Arcese (eds). University of Chicago Press: Chicago. Pp. 3-30.
- Sinclair, A. R. E., (1995b). Population limitation of resident herbivores, In *Serengeti II: Dynamics, Management, and Conservation of an Ecosystem*. A. R. E. Sinclair and P. Arcese (eds). University of Chicago Press: Chicago. Pp. 194-219.

References

- Smith, G. E. J. (1981). Some thoughts on Sampling design. In *Low level survey techniques. ILCA Monograph No. 4*. International Livestock Centre for Africa-ILCA: Addis Ababa. Pp. 159-166.
- Sokal, R. R. and Rohlf, J. F. (1995). *Biometry: The principles and practice of statistics in biological research*. W. H. Freeman and company: New York.
- Spiegelhalter, D. J., Best, N. G., Gilks, W. R. and Inskip, H. (1996). Hepatitis B: A case study in MCMC methods. In *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds). Chapman & Hall: London. Pp. 21-43.
- Spiegelhalter, D. J., Thomas, A., Gilks, W. R. and Best, N. G. (1996). BUGS 0.5, Version ii, User manual, Medical Research Council, Biostatistics unit: Cambridge.
- Steffens, F. E. (1993). Geostatistical estimation of animal abundance in Kruger National Park, South Africa. In *Geostatistics Troia 92*, A. Soares (ed). Kluwer Academic Press: Dordrecht. Pp. 887-897.
- Stein, A. (1994). The use of prior information in spatial statistics. *Geoderma* 62:199-216.
- Stein, A. and van Groenigen, J. W. (1997). Spatial and temporal statistical procedures to support decision making for smart farming. *Proceedings of the INRA-KCW workshop on decision support systems*. Laon, France. Pp 125-129.
- Stelfox, J. G. and M. Ngatia. (1979). *Grevy's zebra in North Central Kenya*. Technical Report. Series. No. 13. KREMU: Nairobi, Kenya.
- Stoms, D. M., Davis, F. W., and Cogan, C. B., (1992). Sensitivity of Wildlife habitat models to uncertainties in GIS data, *Photogrammetry Engineering and Remote Sensing*. 58(6):843-850.
- Thompson, S. K. (1990) Adaptive cluster sampling: Designs with primary and secondary units. *Biometrics*. 47:1103-1115.
- Thompson, S. K. (1992). *Sampling*. John Wiley and Sons: New York.
- Thompson, S. K. and Seber, G. A. F. (1996). *Adaptive sampling*. John Wiley and Sons: New York.
- Thouless C. R. (1995). Long distance movement of elephants in northern Kenya. *African Journal Ecology*. 33:321-334.
- Tomlin, C. D., Berwick, S. H., and Tomlin, S. M., (1983). *Cartographic analysis of deer habitat utilisation*. Prentice-Hall: Englewood Cliffs, New Jersey.
- Tukey, J.W. (1977). *Exploratory Data Analysis*. Addison-Wesley: Reading, MA.
- Van Groenigen, J. W. and Stein, A. (1998) Constrained Optimisation of Spatial Sampling using continuous Simulated Annealing, *Journal of Environmental Quality* 27:1078-1086.
- Vesey-Fitzgerald, D. (1960). Grazing succession among East African game animals. *Journal of mammals*. 41:161-172.
- Voeten, M. M. (1999). *Living with wildlife: Coexistence of wildlife and livestock in an*

References

- East African Savannah system*. PhD Thesis. Wageningen University: Wageningen, The Netherlands.
- Voeten, M. M. and Prins, H. H. T. (1999). Resource Partitioning between sympatric wild and domestic herbivores in the Tarangire region of Tanzania. *Oecologia*. **120**:287-294.
- Voogd, H. (1983). *Multicriteria Evaluation for Urban and Rural Planning*. London: Pion.
- Walker, P. A., (1990). Modelling Wildlife distributions using a geographic information system: Kangaroos in relation to climate, *Journal of Biogeography*. **17**:279-289.
- Walker, P. A., and Moore, D. M., (1988). SIMPLE: an inductive modelling and mapping tool for spatially oriented data, *International Journal of Geographic Information Systems*. **2**:374-363.
- Wargute, P.W. and Said, M.Y. (1996). *The Status of Wild Equids in Kenya: The Grevy's and Burchell's zebra*. Technical Report. Department of Resource Surveys and Remote Sensing: Nairobi, Kenya.
- Western, D. (1975). Water availability and its influence on the structure and dynamics of large community in savannah. *East African Wildlife Journal*. **13**:265-289.
- Wint, W. 1998. Rapid resource assessment and environmental monitoring using low level aerial surveys. Pp 277-301 in Squires, V. R. and Sidahmed, A. E. (ed.). *Drylands: Sustainable use of rangelands into the twenty-first century*. IFAD publication, Rome. 470 pp.

Author Bibliography

- Khaemba, W. M. and Stein, A., (2000). Use of GIS for a spatial and temporal analysis of Kenyan wildlife with generalised linear modelling. *Int. J. Geographical Information Science*, Vol. 14, No. 8, 000-000.
- Khaemba, W. M. and Stein, A., (*In press*). Spatial statistics for modelling of abundance and distribution of wildlife species in the Masai Mara ecosystem, Kenya. *Environmental and Ecological Statistics*.
- Khaemba, W. M. and Stein, A., (*In review*). Improved sampling for airborne surveys to estimate wildlife population parameters in the African savannah. *Wildlife Research*.
- Khaemba, W. M., Said, M. Y., Ottichilo, W. K. and Stein, A., (*In review*). Spatial Point Pattern Analysis of Total Count Data from Aerial Wildlife Surveys. *Journal of Biological, Agricultural and Ecological Statistics*.

References

- Khaemba, W. M., Stein, A. and Rasch, D. (*Submitted*). An empirically simulated investigation and validation of aerial wildlife sampling methods. *African Journal of Ecology*.
- Ottichilo, W. K. and Khaemba, W. M., (*In press*). Validation of observer and aircraft calibration for aerial surveys. *African Journal of Ecology*.
- Ottichilo, W. K., Khaemba, W. M. and Said, M. Y., (*In review*). An investigation into sampling strategies and bias correction for aerial surveys. *Journal of Wildlife Management*.

English Summary

Current methods of obtaining information on wildlife populations are based on monitoring programmes using periodic surveys. In most cases aerial techniques are applied. Reported numbers are, however, often biased and imprecise, making it difficult to use this information for management purposes. This thesis develops suitable statistical procedures to improve sampling of wildlife populations. It investigates survey and analysis procedures and proposes improvements and modifications to existing methods. Data analysed in the study originate two study areas in Kenya: Masai Mara National Reserve and Laikipia ecosystem.

Chapter 1 gives a general introduction to the thesis. It formulates the motivation, objectives and scope of the research.

Chapter 2 investigates different current sampling designs in aerial surveys, with particular focus on systematic and stratified random sampling. Sampling error is a major cause of biased and imprecise estimates of population parameters. Occurrence of several wildlife species in herds violates common assumptions in current sampling methods. This chapter investigates and discusses advantages and disadvantages of two common sampling designs in wildlife surveys: simple random sampling and a modified systematic sampling design known as systematic reconnaissance flights (SRF). It proposes an adaptive sampling design as an alternative that takes clustering of wildlife populations into account and uses criteria on observed animal counts to maximise sampling information. For such populations, the adaptive design is found to be more efficient than the common designs, showing a decrease in the standard error of up to 37%. The comparison focuses on three animal species of varying social behaviour: the elephant (*Loxodonta africana*), kongoni (*Alcelaphus buselaphus*) and wildebeest (*Connochaetes taurinus*).

Chapter 3 integrates generalised linear modelling with geographic information systems to model abundance and distribution of wildlife in space and time. The chapter focuses on the distribution of elephant during nine successive surveys. It analyses their temporal and spatial distribution and relates these to 12 environmental variables using generalised linear modelling. A principal component analysis identifies five principle components, thereby reducing dimensionality in the data. The number of variables explaining elephant abundance is subject to large variations during successive surveys with a minimum of four and maximum of eight variables. In general, variables related to the protected reserve have more influence on elephant abundance. This chapter also develops a simple distance measure to calculate spatial correlation for wildlife data obtained through aerial surveys by quantifying clustering for different animal species. The procedure is illustrated by data on elephant, kongoni, wildebeest and zebra (*Equus burchelli*). The measure captures clustering in the wildebeest and zebra, which have over 20 times more observations within short distances compared to the other two species.

Chapter 4 introduces more modern statistical procedures and applies them for better management of wildlife by addressing three key issues: determination of abundance, modelling of animal distributions and variability of diversity in both space and time. Prior

information is incorporated in Markov Chain Monte Carlo (MCMC) methods and used to improve estimates of abundance. The new abundance estimates are up to 35% more accurate when compared to those obtained by the common Jolly II method. Modelling distribution is improved by developing a simple space-time procedure within a geographical information system, which includes modelling of autocorrelation in wildlife counts. Significant temporal changes in spatial patterns are found from a space-time analysis of elephant counts over a 20-year period, with strong interactions over 5 km and 6 months space and time separations, respectively. Spatial dependence is found to account for most variation when modelling species distribution. The chapter further proposes a diversity index suitable for monitoring changes in trend of large herbivores and based on transect data. The index is sensitive to both high abundance and species richness and is able to capture year to year variation. It indicates an overall marginal decrease in large herbivore diversity for in the Masai Mara ecosystem. The diversity index is easy to compute, thereby providing a handy tool for rapid decision making.

Wildlife populations exhibit clustering in their distributions that is difficult to assess quantitatively by analysing transect data obtained from aerial surveys. Chapter 5 looks at this issue and analyses different clustering behaviour and characterises them using spatial point patterns analysis. This is made possible by the availability of a detailed data set, which gives geographic positions of each observed group of animals, leading to data that is amenable to spatial point pattern analysis. Nearest neighbour distance measures like the G -statistic and K -function are used to classify observed patterns as clustered, regular or completely random to correspond to three types of social behaviour, i.e. animals found in large herds, animals found in small to medium herds and solitary animals. Independence between species is tested using a multivariate extension of the K -function. Results show that spatial point patterns from Thomson's gazelle (*Gazelle thomsoni*) and impala (*Aepyceros melampus*) come from strongly clustered populations. Clustering is explained for different wildlife species by relating observed patterns to environmental factors like vegetation type. This chapter demonstrates spatial point pattern analysis to be useful in determining and confirming species distribution patterns.

In chapter 6 once more takes advantage of the detailed data set to develop a procedure that combines statistical simulation techniques and GIS to compare performance of the two common sampling designs, random and systematic, to the adaptive design. The intensive simulation in a GIS compares distribution, sampling and estimation of abundance. The chapter further assesses impact of sampling designs and intensities on estimates of population parameters from the three designs. Performance is compared by means of the root mean square errors at three increasing sampling intensities. Results show an increase in precision of estimates with increasing sampling intensity, while no significant differences are observed between estimates obtained with the two common sampling designs. The study demonstrates an increase in precision for the adaptive design, thereby stressing the importance of using such designs when sampling clustered populations.

A brief outline is given in chapter 7 to aid a wildlife manager choose between different spatio-temporal techniques and other statistical methods introduced in earlier chapters of the thesis. This short chapter describes different scenarios encountered when making decisions related to the statistical aspects of wildlife management. This is made relevant by the fact that more and more data are collected in space and time and their proper analysis requires appropriate statistical attention. Selection of the right design and analysis method can result in significant savings in cost.

In summary, this research shows that ecology can largely benefit from application of appropriate statistical techniques. In particular, estimation of population parameters like population size needs sound sampling strategies, while assumptions for each sampling design need to be carefully studied. Use of detailed data proved to be an important improvement in understanding spatial distribution of wildlife. This study suggests that it is better to first model spatial and temporal dependence, which is known to exist for many biological populations, before carrying out more detailed analysis. In general, this thesis shows that several existing techniques useful for studying dynamic populations can be extended and improved to provide tools that improve the information obtained from wildlife surveys.

In conclusion, the following are main findings of this research:

- An adaptive sampling strategy as presented in this thesis is an extension to current sampling strategies that allows to sample clustered wildlife populations.
- Modelling of spatial dependence for individual wildlife species improves estimation of wildlife abundance.
- Modelling of spatial distributions of wildlife benefits from a further integration of statistical techniques in geographical information systems.
- An extension of current statistical methods with procedures to analyse spatio-temporal data allows assessing changes in wildlife populations in space and time.
- A simple diversity index as developed in this study shows a marginal decrease in diversity of large herbivores in the Masai Mara ecosystem.
- Data that are currently being collected by aerial transects are usually not sufficient for a rigorous statistical analysis. A higher resolution, in particular recording of individual animal locations, is necessary to model spatial distributions of wildlife using spatial point pattern analysis. If a spatial point pattern analysis is carried out, detailed information about the ecology of different species becomes available.

Dutch Summary

Dit proefschrift richt zich het ontwikkelen van statistische procedures om het waarnemen van wild te verbeteren. Gangbare methodes om informatie in te winnen maken gebruik van monitoring programma's gebaseerd op periodieke inventarisaties. Meestal worden hierbij luchtopnames gebruikt. Waarnemingen en tellingen zijn echter vaak onzuiver. Dit maakt het moeilijk om ze te gebruiken voor beleid en beheer. In dit proefschrift onderzoek ik de procedures voor waarneming en analyse en stel ik verbeteringen en veranderingen voor op bestaande methodes. De gegevens komen uit het Masai Mara natuurspark in Kenia.

Hoofdstuk 1 geeft een algemene inleiding op het proefschrift. Het formuleert de motivatie, het doel en reikwijdte van de studie.

Hoofdstuk 2 onderzoekt verschillende gangbare bemonsteringsstrategieën die gebruikt worden bij luchtopnames, in het bijzonder systematisch bemonsteren en gestratificeerd willekeurig bemonsteren. De monsterfout is een belangrijke bron voor onzuivere en onnauwkeurige schattingen van parameters die de populatie beschrijven. Bovendien komen verschillende soorten wilde dieren in kuddes voor. Dit komt niet overeen met noodzakelijke aannames die in gangbare bemonsteringsstrategieën gemaakt moeten worden. Het hoofdstuk onderzoekt en bediscussieert voor- en nadelen van twee gangbare bemonsteringsstrategieën in wild observaties: enkelvoudige willekeurige bemonstering en een modificatie op de systematische strategie, die we kennen als systematische herkenningsvluchten. Het presenteert een alternatieve strategie die kuddevorming in beschouwing neemt en die criteria gebruikt om aantallen waargenomen dieren te maximaliseren bij het verzamelen van informatie. Voor dergelijke populaties is dit schema efficiënter dan gangbare schema's, in die zin dat de standaardfout afneemt met waardes tot 37%. De vergelijking is gebaseerd op drie soorten: olifanten (*Loxodonta africana*), kongoni's (*Alcelaphus buselaphus*) en gnoes (*Connochaetes taurinus*). Deze hebben alle een verschillende sociaal gedrag.

Hoofdstuk 3 integreert het gegeneraliseerde lineaire model in een geografisch informatie systeem ten einde het voorkomen en de verspreiding van olifanten in ruimte en tijd te modelleren. Het concentreert zich op hun verspreiding gedurende negen opeenvolgende inventarisaties. Hun ruimtelijke en temporele verspreiding wordt gekoppeld aan 12 omgevingsvariabelen met behulp van statistische methoden. Een hoofdcomponentenanalyse identificeerde 5 hoofdcomponenten, waarbij de dimensie van de ruimte waarin de gegevens voorkomen duidelijk is gereduceerd. Het aantal variabelen is overigens sterk wisselend gedurende opeenvolgende inventarisatie, met een minimum van 4 en een maximum van 8 variabelen. In het algemeen zijn variabelen die het beschermde gebied beschrijven het sterkst gekoppeld aan het beschrijven van voorkomen van de olifanten. Het hoofdstuk gaat dan verder met het ontwikkelen van een eenvoudige afstandsmaat om ruimtelijke correlatie te meten. Voor wilde dieren die met luchtopnames worden waargenomen gebeurt dit door kuddevorming van

verschillende soorten te kwantificeren. De procedures zijn toegepast op olifanten, kongoni's, gnoes en zebra's (*Equus burchelli*). Kuddevorming van zowel gnoes als zebra's is met succes gemodelleerd. Deze dieren kennen tot 20 keer zoveel waarnemingen op korte afstand als de twee andere onderzochte diersoorten.

Hoofdstuk 4 introduceert moderne statistische procedures en past die toe voor beheer en beleid, met daarbij aandacht voor het kwantificeren van het voorkomen van soorten, het modelleren van verdelingen en voor het kwantificeren van variabiliteit in ruimte en tijd. Voorinformatie kan worden meegenomen als gebruik gemaakt wordt van Markov Chain Monte Carlo methoden. Schattingen voor het voorkomen van soorten worden hiermee verbeterd. Deze zijn 35% nauwkeuriger dan die worden verkregen met de gangbare Jolly-II procedure. Het modelleren van de verdeling in ruimte en tijd wordt verbeterd via een eenvoudige procedure waarbij variatie in ruimte en tijd in een geografisch informatie systeem worden gemodelleerd en autocorrelaties worden meegenomen. Significante temporele veranderingen in ruimtelijke patronen zijn aangetroffen in een ruimte-tijds analyse van tellingen aan olifanten gedurende een 20-jarige periode met sterke interacties over afstanden tot 5 km in de ruimte en intervallen tot 6 maanden in de tijd. Ruimtelijke afhankelijkheid neemt de meeste variatie voor haar rekening bij het modelleren van verdeling der soorten. Het hoofdstuk stelt voorts een diversiteitindex voor die geschikt is voor het monitoren van veranderingen in langjarige trends van grote herbivoren. Deze is gebaseerd op 1-dimensionale transect gegevens. De index is gevoelig voor hoge waarden voor het voorkomen van soorten en voor soortenrijkdom en is in staat de jaarlijkse variatie te kwantificeren. Het laat zien dat er een geringe afname is in de diversiteit van grote herbivoren in het Masai Mara ecosysteem. De diversiteitindex kan eenvoudig worden berekend en is daarmee een geschikt gereedschap voor het ondersteunen van het maken van snelle beslissingen in ruimte en tijd.

Populaties van wilde dieren kunnen clustering vertonen in hun verdeling. Deze clustering is vaak moeilijk kwantitatief vast te stellen als gegevens worden gebruikt die via luchtopnames worden verkregen. Hoofdstuk 5 analyseert en kwantificeert verschillend gedrag in clustering en karakteriseert dit via het gebruik van procedures uit de punt-patroon analyse. Dit is mogelijk geworden doordat een gegevens bestand beschikbaar kwam met zeer gedetailleerde informatie over de geografische posities van iedere waargenomen groep dieren. Deze gegevens lijken daarom sterk op een puntpatroon. Maten gebaseerd op de kortste afstand tussen (clusters van) dieren zoals de G - en de K -functie worden gebruikt om waargenomen patronen te classificeren als geclusterd, regelmatig of totaal willekeurig. Dit correspondeert dan met drie vormen van sociaal gedrag: dieren die voorkomen in grote kuddes, dieren die voorkomen in kleine tot middelgrote kuddes en solitaire dieren. Onafhankelijkheid tussen soorten is getoetst met een multivariate uitbreiding van de K -functie en met de X^2 toets. Resultaten laten zien dat ruimtelijke patronen van de Thomson gazelle (*Gazelle thomsoni*) en de impala (*Aepyceros*

melampus) afkomstig zijn van sterk geclusterde populaties. Clustering kan verklaard worden voor de verschillende soorten wild door de waargenomen patronen te koppelen aan omgevingsfactoren, zoals het type vegetatie. Dit hoofdstuk laat zien dat een analyse van ruimtelijke punt patronen nuttig is bij het bepalen en bevestigen van verdelingspatronen van verschillende wildsoorten.

In hoofdstuk 6 maak ik nog eens gebruik van het gedetailleerde gegevens bestand en gebruik dat om een procedure te ontwikkelen die statistische methoden koppelt aan een geografisch informatie systeem om de kwaliteit te vergelijken tussen twee gangbare bemonsteringsstrategieën en de adaptieve strategie die in hoofdstuk 2 werd gepresenteerd. Ik rapporteer een vergelijking van de verdeling bemonstering en schatting van het voorkomen van wild door middel van een intensieve simulatie in een geografisch informatie systeem. De kwaliteit van het functioneren van twee gangbare bemonsteringsstrategieën t.o.v. de adaptieve strategie wordt gemeten via de wortel uit de gekwadrateerde verschillen tussen gemeten en gemodelleerde waarden bij drie bemonsteringsdichtende. Het hoofdstuk meet voorts de invloed van een bemonsteringsstrategie en dichtheden van voorkomen op schattingen van populatie parameters die met de drie strategieën verkregen worden. De simulatie laat een toename zien van de nauwkeurigheid van de schattingen met toenemende bemonsteringsintensiteit, terwijl geen significante verschillen bestaan tussen de schattingen die met gangbare schattingsprocedures worden verkregen. De simulatie laat verder een toename zien in de precisie van de schattingen als de adaptieve strategie gebruikt wordt, waarbij de kwaliteit voor het gebruiken van deze strategie nog eens wordt onderschreven voor het bemonsteren van geclusterde populaties van wilde dieren.

In hoofdstuk 7 geef ik een kort overzicht dat een beheerder van een wildreservaat ten dienste kan zijn om te kiezen tussen verschillende ruimtelijk en temporele technieken en andere statistische methoden die eerder in het proefschrift behandeld zijn. Het hoofdstuk beschrijft verschillende scenario's die gebruikt kunnen worden als een beslissing moet worden genomen die betrekking heeft op de statistische aspecten van wildbeheer. De motivatie is dat steeds meer gegevens beschikbaar komen in ruimte en tijd en dat een goede analyse hiervoor de juiste statistische aandacht moet krijgen. Een keuze voor de juiste bemonsteringsstrategie en de juiste analyse methode resulteert in een aanzienlijke besparing van kosten.

Samengevat laat dit onderzoek zien dat ecologie in hoge mate baat kan vinden bij het toepassen van de juiste statistische procedures. In het bijzonder wordt het nut getoond van het gebruik van geschikte bemonsteringsstrategieën voor het schatten van populatie parameters zoals de omvang van een populatie. De aannames die bij iedere strategie gemaakt worden moeten zorgvuldig getoetst worden. Een nuttige aanwijzing is verder om eerst de ruimtelijke en temporele variatie te modelleren die in veel biologische populaties aanwezig is. Tenslotte is het van groot belang om zo nauwkeurig mogelijke gegevens te verzamelen die door een zo klein

mogelijke aggregatie een zo groot mogelijke betrouwbaarheid hebben. Dit proefschrift met de studie naar dynamische populaties in een natuurreserveat in Afrika laat zien dat verschillende technieken bestaan die kunnen worden uitgebreid en verbeterd om de gereedschappen te geven die de informatie uit luchtopnames van wildpopulaties kunnen verbeteren.

De belangrijkste resultaten van dit onderzoek zijn:

- Een adaptieve bemonsteringsstrategie zoals gepresenteerd in dit proefschrift vormt een uitbreiding op bestaande bemonsteringsstrategieën voor wild om geclusterde populaties te kunnen bemonsteren.
- Het modelleren van ruimtelijke afhankelijkheid binnen individuele soorten wild komt ten goede aan het schatten van de totale populatie.
- Het modelleren van ruimtelijke verdelingen van wild profiteert van een verdere integratie van statistische technieken in geografische informatie systemen.
- Een uitbreiding van gangbare statistische methoden met procedures voor het analyseren van spatio-temporele gegevens maakt het mogelijk om veranderingen in wildpopulaties te modelleren in ruimte en tijd.
- Een eenvoudige diversiteitindex, die in dit proefschrift wordt gepresenteerd laat zien dat er een geringe afname is in de diversiteit van grote herbivoren in het Masai Mara ecosysteem.
- Gegevens zoals die standaard via vluchttransecten worden verzameld zijn als regel onvoldoende om een gedegen statistische analyse mee uit te voeren. Een betere resolutie, met name registratie van individuele dieren, is nodig om verdelingen van wild te modelleren met behulp van ruimtelijke punt-patronen.

Curriculum Vitae

Wilson Mwale Khaemba was born on 16th of March 1967 in Uasin Gishu district, Kenya. After successfully completing his secondary education at Starehe Boys Centre in Nairobi, he joined Moi University to study Mathematics and Statistics in 1987. He graduated in 1990 with a Bachelor of Science degree (with honours) and immediately went to work as an actuarial officer in the Ministry of Finance. He left this position to take up a graduate fellowship at Moi University, where he completed a Master of Philosophy course in Mathematical Statistics in 1993. He was subsequently employed as an assistant lecturer in the department of mathematics and statistics at the same university. In mid 1995, he left Kenya to join Limburgs University in Belgium where he pursued a Master of Science degree course in Biostatistics. He graduated in 1996 and returned to his teaching position at Moi University. He later won a scholarship from the Institute of Aerospace Survey and Earth Sciences (ITC) in the Netherlands and was admitted as a PhD researcher in Spatial Statistics for natural resource management (NRM) in March 1997. He is currently employed as a lecturer in the department of mathematics and statistics at Moi University, Kenya.

Appendix 1

Completed Ph.D. studies at ITC

This list is also available at [Http://www.itc.nl/Research](http://www.itc.nl/Research), where it updated regularly.

Completed Ph.D. studies at ITC

Author	Year	Dissertation No.	ISBN	Dissertation title
Mr. Akinyede (Nigeria)	1990	1	NA	Highway cost modelling and route selection using a geotechnical information system
Mr. Pan He Ping (China)	1990	2	90-9003-757-8	Spatial structure theory in machine vision and applications to structural and textural analysis of remotely sensed images.
Mr. G. Bocco Verdinelli (Mexico)	1990	3	NA	Gully erosion analysis using remote sensing and geographic information systems: a case study in Central Mexico
Mr. M. Sharif (ITC staff)	1991	4	NA	Composite sampling optimization for DTM in the context of GIS
Ms. J. Drummond (ITC staff)	1991	5	NA	Determining and processing quality parameters in geographic information systems
Ms. S. Groten (ITC staff)	1991	6	NA	Satellite monitoring of agro-ecosystems in the Sahel.
Mr. A. Sharifi (ITC staff)	1991	7	90-6164-074-1	Development of an appropriate resource information system to support agricultural management at farm enterprise level
Mr. D. van der Zee (ITC staff)	1991	8	90-6164-075-X	Recreation studied from above: Air photo interpretation as input into land evaluation for recreation
Mr. C. Mannaerts (ITC staff)	1991	9	90-6164-085-7	Assessment of the transferability of laboratory rainfall-runoff and rainfall -soil loss relationships to field and catchment scales: a study in the Cape Verde Islands
Mr. Ze Shen Wang (China)	1991	10	90-393-0333-9	An expert system for cartographic symbol design
Mr. Zhou Yunxian (China)	1991	11	90-6164-081-4	Application of Radon transforms to the processing of airborne geophysical data.
Mr. M. de Zuñivia (Argentina)	1992	12	90-6164-077-6	Mapping agro-topoclimates by integrating topographic, meteorological and land ecological data in a geographic information system: a case study of the Lom Sak area, North Central Thailand
Mr. C. van Westen (ITC staff)	1993	13	90-6164-078-4	Application of Geographic Information Systems to landslide hazard zonation
Mr. Shi Wenzhong (China)	1994	14	90-6164-099-7	Modelling positional and thematic uncertainties in integration of remote sensing and geographic information systems
Mr. R. Javelosa (Philippines)	1994	15	90-6164-086-5	Active Quaternary environments in the Philippine mobile belt.
Mr. Lo King-Chang (China)	1994	16	90-9006526-1	High Quality Automatic DEM, Digital Elevation Model Generation from Multiple Imagery
Mr. S. Wokabi (Kenya)	1994	17	90-6164-102-0	Quantified land evaluation for maize yield gap analysis at three sites on the eastern slope of Mt. Kenya.
Mr. O. Rodriguez (Venezuela)	1995	18	NA	Land Use conflicts and planning strategies in urban fringes: a case study of Western Caracas, Venezuela
Mr. F. van der Meer (ITC staff)	1995	19	90-5485-385-9	Imaging spectrometry & the Ronda peridotites
Mr. O. Kufoniya (Nigeria)	1995	20	90-6164-105-5	Spatial coincidence: automated database updating and data consistency in vector GIS
Mr. P. Zambezi (Zambia)	1995	21	NA	Geochemistry of the Nkombwa Hill carbonatite complex of Isoka District, north-east Zambia, with special emphasis on economic minerals
Mr. T. Woldai (ITC staff)	1995	22	NA	The application of remote sensing to the study of the geology and structure of the Carboniferous in the Calañas area, pyrite belt, SW Spain
Ms. P. Verweij (ITC staff)	1995	23	90-6164-109-8	Spatial and temporal modelling of vegetation patterns: burning and grazing in the Paramo of Los Nevados National Park, Colombia
Ms. C. Pohl (Germany)	1996	24	90-6164-121-7	Geometric Aspects of Multisensor Image Fusion for Topographic Map Updating in the Humid Tropics
Mr. Jiang Bin (China)	1996	25	90-6266-128-9	Fuzzy overlay analysis and visualization in GIS
Ms. G. Metternicht (Argentina)	1996	26	90-6164-118-7	Detecting and monitoring land degradation features and processes in the Cochabamba Valleys, Bolivia. A synergistic approach.

Appendix 1

Mr. Hoanh Chu Thai (Vietnam)	1996	27	90-6164-120-9	Development of a Computerized Aid to Integrated Land Use Planning (CAILUP) at regional level in irrigated areas: a case study for the Quan Lo Phung Hiep region in the Mekong Delta, Vietnam
Mr. A. Roshannejad (Iran)	1996	28	90-9009-284-6	The management of spatio-temporal data in a national geographic information system
Mr. M. Terlien (AiO)	1996	29	90-6164-115-2	Modelling Spatial and Temporal Variations in Rainfall-Triggered Landslides: the integration of hydrologic models, slope stability models and GIS for the hazard zonation of rainfall-triggered landslides with examples from Manizales, Colombia
Mr. J. Mahavir (India)	1996	30	90-6164-117-9	Modelling settlement patterns for metropolitan regions: inputs from remote sensing
Ms. S. Al-Amir (Syria)	1996	31	90-6164-116-0	Modern spatial planning practice as supported by the multi-applicable tools of remote sensing and GIS: the Syrian case
Mr. M. Pilouk (Thailand)	1996	32	90-6164-122-5	Integrated modelling for 3D GIS
Mr. Duan Zengshan (China)	1996	33	90-6164-123-3	Optimization modelling of a river-aquifer system with technical interventions: a case study for the Huangshui river and the coastal aquifer, Shandong, China
Mr. W.H. de Man (ITC staff)	1996	34	90-9009-775-9	Surveys: informatie als norm: een verkenning van de institutionalisering van dorp - surveys in Thailand en op de Filipijnen
Mr. Z. Vekerdy (Hungary)	1996	35	90-6164-119-5	GIS-based hydrological modelling of alluvial regions: using the example of the Kisaföld, Hungary
Ms. Luisa Pereira (Portugal)	1996	36	90-407-1385-5	A Robust and Adaptive Matching Procedure for Automatic Modelling of Terrain Relief
Ms. M. Fandino Lozano (Colombia)	1996	37	90-6164-129-2	A Framework of Ecological Evaluation oriented at the Establishment and Management of Protected Areas: a case study of the Santuario de Iguaque, Colombia
Mr. B. Toxopeus (ITC staff)	1996	38	90-6164-126-8	ISM : an Interactive Spatial and temporal Modelling system as a tool in ecosystem management : with two case studies : Cibodas biosphere reserve, West Java Indonesia : Amboseli biosphere reserve, Kajiado district, Central Southern Kenya
Ms. Wang Yiman (China)	1997	39	90-6164-131-4	Satellite SAR imagery for topographic mapping of tidal flat areas in the Dutch Wadden Sea
Ms. Asun Saldana-Lopez (Spain)	1997	40	90-6164-133-0	Complexity of soils and Soilscape patterns on the southern slopes of the Ayllon Range, central Spain: a GIS assisted modelling approach
Mr. T. Ceccarelli (Italy)	1997	41	90-6164-135-7	Towards a planning support system for communal areas in the Zambezi valley, Zimbabwe: a multi-criteria evaluation linking farm household analysis, land evaluation and geographic information systems
Mr. Peng Wanning (China)	1997	42	90-6164-134-9	Automated generalization in GIS
Ms. C. Lawas (Philippines)	1997	43	90-6164-137-3	The Resource Users' Knowledge, the neglected input in Land resource management: the case of the Kankanaey farmers in Benguet, Philippines
Ms. W. Bijker (ITC staff)	1997	44	90-6164-139-X	Radar for rain forest: A monitoring system for land cover Change in the Colombian Amazon
Mr. A. Farshad (ITC staff)	1997	45	90-6164-142-X	Analysis of integrated land and water management practices within different agricultural systems under semi-arid conditions of Iran and evaluation of their sustainability
Mr. B. Orlic (Serbia)	1997	46	90-6164-140-3	Predicting subsurface conditions for geotechnical modelling
Mr. Y. Bishr (Egypt)	1997	47	90-6164-141-1	Semantic Aspects of Interoperable GIS
Mr. Zhang Xiangmin (China)	1998	48	90-6164-144-6	Coal fires in Northwest China: detection, monitoring and prediction using remote sensing data
Mr. R. Gens (Germany)	1998	49	90-6164-155-1	Quality assessment of SAR interferometric data
Mr. J. Turkstra (ITC staff)	1998	50	90-6164-147-0	Urban development and geographical information: spatial and temporal patterns of urban development and land values using integrated geo-data, Villaviciencia, Colombia
Mr. C. Cassells (UK)	1998	51	NA	Thermal modelling of underground coal fires in northern

Completed Ph.D. studies at ITC

				China
Mr. M. Naseri (Iran)	1998	52	NA	Monitoring soil salinization, Iran
Mr. B.G.H. Gorte (ITC staff)	1998	53	90-6164-157-8	Probabilistic Segmentation of Remotely Sensed Images
Mr. Tenalem Ayenew (Ethiopia)	1998	54	90-6164-158-6	The hydrological system of the lake district basin, central main Ethiopian rift
Mr. Wang Donggen (China)	1998	55	90-6864-551-7	Conjoint approaches to developing activity-based models
Ms. M. Bastidas de Calderon (Venezuela)	1998	56	NA	Environmental fragility and vulnerability of Amazonian landscapes and ecosystems in the middle Orinoco river basin, Venezuela
Mr. A. Moameni (Iran)	1999	57	NA	Soil quality changes under long-term wheat cultivation in the Marvdasht plain, South-Central Iran
Mr. J.W. van Groenigen (The Netherlands)	1999	58	90-6164-156-X	Constrained optimisation of spatial sampling: a geostatistical approach
Ms. Cheng Tao (China)	1999	59	90-6164-164-0	A process-oriented data model for fuzzy spatial objects
Mr. Piotr Wolski (Poland)	1999	60	90-6164-165-9	Application of reservoir modelling to hydrotopes identified by remote sensing
Mr. B. Acharya (Nepal)	1999	61	90-6164-168-3	Forest biodiversity assessment: A spatial analysis of tree species diversity in Nepal.
Mr. Ali Akbar Abkar (Iran)	1999	62	90-6164-169-1	Likelihood-based segmentation and classification of remotely sensed images
Mr. T. Yanuariadi (Indonesia)	1999	63	90-5808-082-X	Sustainable Land Allocation: GIS-based decision support for industrial forest plantation development in Indonesia
Mr. Mohamed Abu Bakr (Sudan)	1999	64	90-6164-170-5	An Integrated Agro-Economic and Agro-Ecological Framework for Land Use Planning and Policy Analysis
Ms. M. Eleveld (The Netherlands)	1999	65	90-6461-166-7	Exploring coastal morphodynamics of Ameland (The Netherlands) with remote sensing monitoring techniques and dynamic modelling in GIS
Ms. Yang Hong (China)	1999	66	90-6164-172-1	Imaging Spectrometry for Hydrocarbon Microseepage
Mr. Félix Mainam (Cameroon)	1999	67	90-6164-179-9	Modelling soil erodibility in the semiarid zone of Cameroon
Mr. Mahmoud Bakr (Egypt)	2000	68	90-6164-176-4	A Stochastic Inverse-Management Approach to Groundwater Quality
Ms. S. Zlatanova (Bulgaria)	2000	69	90-6164-178-0	3D GIS for Urban Development
Mr. Wilber K. Ottichilo (Kenya)	2000	70	90-5808-197-4	Wildlife Dynamics: An Analysis of Change in the Masai Mara Ecosystem
Mr. Nuri Kaymakci (Turkey)	2000	71	90-6164-181-0	Tectono-stratigraphical Evolution of the Cankiri Basin (Central Anatolia, Turkey)
Ms. Rhodora Gonzalez (Philippines)	2000	72	90-5808-246-6	Platforms and Terraces: Bridging participation and GIS in joint-learning for watershed management with the Ifugaos of the Philippines
Mr. Ernst Schetselaar (ITC staff)	2000	73	90-6164-180-2	Integrated analyses of granite-gneiss terrain from field and multisource remotely sensed data. A case study from the Canadian Shield