STATISTICAL ANALYSES OF ARTIFICIAL WATERPOINTS: THEIR EFFECT ON THE HERBACEOUS AND WOODY STRUCTURE COMPOSITION WITHIN THE KRUGER NATIONAL PARK

A thesis submitted in partial fulfilment of the requirements for the degree
of
MASTER OF SCIENCE
at
RHODES UNIVERSITY

by

VICTORIA LUCY GOODALL

December 2006

ABSTRACT

The objective of this project is to link the statistical theory used in the ecological sciences with an actual project that was developed for the South African National Parks Scientific Services. It investigates the changes that have occurred in the herbaceous and woody structure due to the closure of artificial waterpoints; including the impacts that elephants and other herbivores have on the vegetation of the Kruger National Park.

This project was designed in conjunction with South African National Parks (SANP) Scientific Services and it is a registered project with this department. The results of this project will be submitted to Scientific Services in accordance with the terms and conditions of a SANP research project. A major concern within the KNP is the declining numbers of rare antelope and numerous projects have been developed to investigate possible ways of halting this decline and thus protecting the heterogeneity of the Kruger National Park.

Three different datasets were investigated, covering three aspects of vegetation structure and composition within the KNP. The first investigated the changes that have occurred since the N'washitsumbe enclosure in the Far Northern KNP was fenced off from the rest of the park. The results show that over the 40 years since the enclosure was built, changes have occurred which have resulted in a significant difference in the abundance of Increaser 2 and Decreaser grass species between the inside and the outside of the enclosure. Increaser 2 and Decreaser categories are the result of a grass species classification depending on whether the species thrives or is depressed by heavy grazing. The difference in grass species composition and structure between the inside and the outside of the enclosure indicates that the grazing animals within the KNP have influenced the grass composition in a way that favours the dominant animals. This has resulted in a declining roan antelope population – one of the species that is considered as a 'rare antelope'.

Many artificial waterpoints (boreholes and dams) have also been closed throughout the KNP in the hope of resulting in a change in vegetation structure and composition in favour of the roan. Veld condition assessment data for 87 boreholes throughout the Park was analyzed to determine whether the veld in the vicinity is beginning to change towards a more Decreaser dominated sward which would favour the roan. The results were analyzed for the different regions of the Park; and they indicate that changes are becoming evident; however, the results are not particularly conclusive, yet. The majority of the boreholes were closed between 1994 and 1998 which means that not a lot of data were available to be analyzed. A similar study conducted in another 10 years time might reveal more meaningful results. However the results are moving in the direction hoped for by the management of the KNP. The results show that the grass composition has a higher proportion of Decreaser grasses since the closure of the waterpoints, and the grass biomass around these areas has also improved. The results were analyzed on an individual basis; and then on a regional basis as the minimal data meant that the individual analyses did not provide any significant results.

A third study was then done on the impact that the rapidly increasing elephant population on the vegetation within the Riparian zone along three rivers in the Far Northern region of the KNP. The riparian zone is an important part of the landscape, in terms of providing food for many animals as well as shade. The elephant population has increased substantially since the termination of the culling program and this means that the feeding requirements of the population has increased which could result in severe damage upon the vegetation, as elephants can be very destructive feeders.

The results show surprising differences between the three years of data that were analyzed; however the results indicate that the elephants are targeting specific height ranges of trees when feeding; however they do not seem to consistently target specific tree species. This is positive for the diversity of the Riparian zone as this region is very important both ecologically and aesthetically for the tourists who visit the Park.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

There are a number of people who I owe my gratitude to; for their assistance, guidance and support throughout this project. Firstly I would like to thank Prof. Sarah Radloff who has been my Head of Department, supervisor and mentor throughout my time in the Rhodes University Statistics department. She has encouraged me, and had faith in me; and has been available at any time that I needed her advice and knowledge. I could not have hoped for a more supportive supervisor. I would also like to extend my thanks to all the members of the Rhodes University Statistics department who have all helped and encouraged me during my time at Rhodes University in the department.

Then I would like to thank Dr. Rina Grant from the South African National Parks Scientific Services. She has provided me with amazing support, encouragement and words of wisdom about the ecological issues that have been dealt with in this project. I would also like to thank Abri de Buys, formerly of the South African National Parks Scientific Services for his friendship, advice, and numerous discussions about the issues discussed in this paper.

I would also like to thank Angela Gaylard, of SANParks; and Dr. Susi Vetter, of Rhodes University; for their advice and comments on the ecological issues of this paper.

Then, a very special thank you to my parents who have supported me throughout my academic studies from day one at school – I cannot thank them enough for the support, love and encouragement. They also introduced me to the Kruger National Park; and have instilled a deep love and respect for the bush in me. I also need to thank my Dad for the advice and help in the preparation of this document.

Thanks to Volkswagen South Africa Pty (Ltd) for their financial assistance for this project and flexible working hours to allow me to study further. Thanks also go Dr. Jan Hofmeyr and Synovate Pty (Ltd) who provided me with much support and flexibility in the final months of this thesis.

DECLARATION

The following thesis has not been submitted to a university other than Rhodes University, Grahamstown, South Africa. The work presented here is that of the author.

CHAPTER 1: INTRODUCTION

The mission of the Kruger National Park (KNP) is "To develop and manage a system of national parks that represents the biodiversity, landscapes and associated heritage assets of South Africa for the sustainable use and benefit of all" (Pienaar, 2004).

In order to fulfil the goals set out by this mission, the South African National Parks has a dedicated team of scientists conducting research and monitoring projects throughout the KNP, as well as all the other parks throughout South Africa. A key to the understanding and analysis of the results obtained through these projects is the underlying statistical analysis. Numerous statistical methods are used in the analysis of ecological data depending on the type of data, and the desired outcome depending on the ecological meaning of the data.

This study focuses on three projects that have been conducted in the KNP and the theory behind the statistical methods that were used for the data analysis. It is hoped that the project will provide some insight into some of the key ecological issues facing the rare roan antelope, the artificial waterpoint closures and the effect that the increasing elephant population is having in the KNP and at the same time, provide insight to the robust analytical methods of logistic regression and log-linear analysis that are common analytical methods used in the analysis of ecological data.

The first project investigates the changes that have occurred in the proportion of different grass categories in the N'washitsumbe enclosure in the Far Northern KNP. This enclosure was built to provide protection to the rare antelope species. This has meant that no other animals are present in the enclosure; and in the 40 years since the erection of the enclosure, differences in the vegetation are apparent.

In another effort to conserve the rare antelope species in the north of the Park, artificial waterpoints have been closed. The aim of the closures is to restrict bulk grazing, water-dependent animals from utilizing areas that are the preferred territory of the roan antelope. The chapter investigates whether changes in the vegetation have occurred in a similar way to the changes that have occurred in the N'washitsumbe enclosure. The majority of the borehole closures happened in the 1990's which means that far less data is available, unlike the enclosure which is nearly 40 years old.

The elephant population of the Kruger National Park has increased substantially in the last 10 years since the culling of elephants was stopped. Elephants are destructive feeders; and the trees that they damage can die or have their growth severely stunted. The riparian zone is the section of vegetation on the banks of a river; in the Kruger Park, the riparian zones contain some large trees and are both ecologically and aesthetically very important areas. The third part of this thesis investigates whether elephants are targeting specific tree species, height of tree or certain wood types when they utilize trees in these areas.

It is hoped that this research will assist with providing a more comprehensive understanding of vegetation dynamics and the factors that play a major role in the change of vegetation composition and structure. This thesis will also be presented to the South African National Parks Scientific Services as the result of a registered project.

CHAPTER 2: PROPOSAL TO SOUTH AFRICAN NATIONAL PARKS

The following proposal was submitted to the Scientific Services division of South African National Parks in February 2006. It is noted that due to the limitations of the data obtained, not all of the goals for this project have been achieved.

KRUGER NATIONAL PARK RESEARCH PROPOSAL
Victoria Goodall (8108260244085)
Department of Statistics, Rhodes University

Title

Artificial waterpoints: how does their distribution affect the herbaceous and woody structure and composition?

Project Statement

Who is the project being developed for?

It is hypothesized that the wide distribution of artificial waterholes in the KNP led to an increase in bulk grazers. This would have changed the grass composition towards herbaceous species that are more resilient to herbivory. Plant surveys comparing the vegetation inside enclosures with only selective grazers with vegetation outside will provide information to test this hypothesis. Furthermore herbaceous surveys done before and after the closure of about 200 waterpoints across the KNP in 1998 would provide further information towards testing this hypothesis.

Although grass forms a substantial part of elephant diets in summer, they utilize mostly woody species in winter. As elephant are water and shade dependent, it is hypothesized that a wider distribution of artificial water especially near non-perennial rivers will increase utilization in these areas. As elephant tend to select specific species, it could lead to a change in species composition, especially if some of the selected species do not recover readily from the intense herbivory by elephant.

What is the project's main purpose?

The major aims of the project are to:

1. Determine change in herbaceous species composition and biomass as a result of utilization by bulk and other grazers associated with a higher density of waterpoints.
2. Determine the change in herbaceous species composition and biomass as a result of closure of artificial waterpoints.
3. Determine the woody species and structure favoured by elephant in the riparian zone where artificial water is provided.
4. Determine the effect of increasing distance from artificial water on the utilization of woody species by elephant. Determine if certain riparian tree species or height classes are more susceptible to certain types of elephant damage in relation to distance from water.

What change will the project achieve?

The results from the analysis of the enclosures will contribute to our understanding of herbivore-vegetation interactions in the context of the effects on the herbaceous layer of bulk vs. selective grazers. This has implications for water provisioning issues, which change the relative abundance of bulk to selective grazers.

The analysis of the effects of the closure of the waterpoints will be able to examine what impact the closure has had and what the possible effects of further closure of waterpoints could be. It will also enable an investigation of the success of closing the artificial waterpoints and whether this closure has resulted in the availability of grazing more suitable for the selective grazers.

The results of the elephant data analysis will focus on the effect of elephant utilization on a species- and height-specific level. This will contribute towards our understanding of whether the current concern about the effects on vegetation of increasing elephant numbers is indeed justified, and if so, which species or size classes are at highest risk from this damage. The riparian vegetation is ecologically important to the biodiversity of the Park but it also has a huge aesthetic appeal for the tourists who visit the Park. If species- and height-specific elephant impacts are demonstrated, then monitoring programmes and management actions can be fine-tuned to prioritize these species and height classes for protection.

What is the overall time scale?
The project will run from 1 February 2006 until 30 November 2006.

What are the phase time scales?
Literature review and planning – January 2006
Preparation of data – February 2006
Analysis of data – March to July 2006
Thesis and paper Write-up – August to November 2006

Who will benefit from the project being achieved?
The analysis of the enclosure dataset will contribute towards a better understanding of the effects of selective grazing on vegetation. In addition, it will contribute towards our understanding of the effects of enclosures, which can assist with planning for the erection of new enclosures should they become necessary in the future. It will also provide a better understanding of the long-term effects of a population of roan on the vegetation by comparing the N'washitsumbe enclosure that has been utilized by selective grazers for 30 years, to the Capricorn enclosure that has only been exclusively utilized by selective grazers for 2 years.

The nature of enclosed game reserves, with no migration and hardly any natural predators for the elephant, has meant that the population of African Elephant has increased rapidly. The resultant damage to the habitat and impact this has on the other species will continually be researched; and no one solution will solve all the problems. However, each project will result in a bit more information being added to the puzzle of Elephant Management. The riparian zone is a particularly important component of the landscape, both aesthetically and from a biodiversity point of view.

This project will also serve a 50% thesis for a Master of Science Degree through Rhodes University.

Who is the funding provider and what is the estimated cost of the project?

All funding for the MSc will be provided by Victoria Goodall and Volkswagen South Africa (Pty) Ltd. There is no expected cost for this research.

Where will the study area for the project be?

The Northern Plains aspect of the project will focus on the effects of herbivore exclusion on the composition of the grass layer, using data collected from the N'washitsumbe enclosure; while waterpoint closure effects will be examined at all VCA sites across the KNP.

The impact by elephants on the riparian vegetation was examined in the granitic areas of the Phugwane, Mphongolo and Shingwedzi rivers in the Kruger National Park

Assistance from the KNP

GIS data layers: waterpoint distribution, geology, VCA survey results, rainfall etc.

Data collected inside and outside the enclosures

Work Procedure

Part of this project will investigate the changes in vegetation composition around waterpoints and in particular the changes that have occurred in the herbaceous composition and biomass since the closure of some of these artificial waterpoints. The other part of the project will focus on data collected by Angela Gaylard as part of her PhD study under the River and Savanna Boundaries Program. The latter analysis will investigate various types of impact by elephants on the riparian vegetation in the granitic areas of the Phugwane, Mphongolo and Shingwedzi rivers in the Far Northern Kruger National Park. The project will be a joint venture with the Statistics Department of Rhodes University, Grahamstown.

Enclosure Comparison

1. This analysis aims to investigate the changes to the composition of the grass layer within the enclosures as a result of selective grazing, by comparing the grass species composition inside and outside the enclosures.
2. To analyze the differences in grass biomass between the inside and the outside of the enclosures.

3. To analyze the differences and similarities between the 3 enclosures in terms of vegetation composition

Effects of the Closure of Artificial Waterpoints

1. This analysis aims to investigate the extent of changes in vegetation since the closure of waterpoints in the Kruger National Park between 1994 and 1998. This will involve a ‘before’ and ‘after’ comparative study that will investigate the influence of increasing distance to water on the vegetation change; as well as the influence of the density of waterpoints in the area. It could further be investigated whether there are significant changes in the vegetation between the granites and basalt plains or alternatively these changes could be investigated for waterpoints throughout the Kruger National Park; not just the Northern Plains.
2. Investigate the change in the ratio between increaser and Decreaser grasses; as well as annuals and perennials since the closure of the waterpoints.

Elephant damage

1. Analyze the tendencies of the African Elephant to target specific tree species and/or height classes within the riparian zones of seasonal rivers in the northern KNP.
2. Further analyze these trends to determine if certain tree species or height classes are more susceptible to certain types of elephant damage.

The field work for the enclosure study was collected by Abri de Buys as part of his work for the Northern Plains Program. The N’washitsumbe enclosure is in the Vlakteplaas ranger section in the Far Northern Kruger National Park in the basalt areas. The Capricorn enclosure is in the Mooiplaas ranger section in the north of the Kruger National Park; while the Hlangwine enclosure is in the Pretoriuskop sourveld in the south east of the Park.

The elephant damage data was collected along the riparian vegetation in the granite areas of the Phugwane, Mphongolo and Shingwedzi Rivers in the Far Northern Kruger National Park by Angela Gaylard as part of her PhD study.

The N'washitsumbe enclosure study involved categorizing each grass species as either an Increaser 2 (grasses that increase in abundance in veld that is overgrazed) or a Decreaser (grasses that decrease in abundance in veld that is either over or undergrazed). A relative abundance was then calculated; while a disc pasture meter reading was also taken in order to calculate a grass biomass reading ($kg.ha^{-1}$). The desired outcome of this research is to determine if there are differences in the herbaceous vegetation composition and biomass inside and outside the enclosure; and if so, explain these through the selective grazing of the roan antelope as they have been the dominant species within the enclosure for the last 37 years.

The statistical analysis will include a log-linear analysis to determine the significance, if any, of the differences in increaser2 and Decreaser grasses inside and outside the enclosure; as well as the location of increaser2's and Decreasers (as vegetation patches) within the enclosure. An analysis of variance will also be run to determine the significance, if any, of the grass biomass within and outside the enclosure. It can be assumed that the grass measurements inside and outside the enclosure will be independent of each other since the rainfall, climatic conditions and wind will be the same in both locations. Hence any differences can be attributed to the selective grazing of the roan antelope. The erection of man-made enclosures within the KNP creates an environment where the normal predator-prey relationship no longer exists. This ensures that any differences in the vegetation can be attributed to the dominant species and that selective grazing can influence the quality and diversity of the grass layer.

Log-linear analysis and ANOVA tests on the waterpoint data will be carried out to determine if significant vegetation changes have occurred in the vicinity of the closed waterpoints. Five years of VCA data from before 1998, will be used that has the same rainfall regime as 2000 to 2004. Changes in vegetation can then be analyzed to determine whether in fact the vegetation is now becoming more suitable for the type II grazers (roan and sable) rather than the bulk grazers, type 1 (waterbuck, elephant, zebra and wildebeest). The ANOVA will consider data prior and subsequent to the closure of the waterpoints; and possibly could investigate the time period over which the vegetation changes occurred. This could provide further insight into the planning for closure of further waterpoints.

A nested Analysis of Variance (ANOVA) will be run on the elephant data to determine if there are significant trends in the type and intensity of damage that the tree species suffers. By refining the model, we can determine if there are species trends; or if groups of trees, for example the *Acacias*, suffer a particular type (e.g. main stem breakage) of damage more intensively than other groups of trees. By including the age of the damage (old or new damage) into the model, this will demonstrate whether certain species can survive being damaged. This will give an indication of the long-term impact that the destructive feeding habits of elephant can have on the species within the riparian zone and whether this damage could pose a serious threat to particular species. This analysis will indicate whether elephants do target particular species while feeding or whether their feeding in the riparian zone is random. A random feeding pattern would indicate that no one species is more at risk than any other species and the management policy could focus on the maintenance of the riparian zone, rather than individual species. This would contribute towards a better understanding of elephant-vegetation interactions for the management of the riparian zone by highlighting the areas of immediate concern.

Major milestones

The major milestones for the project are:

Literature review on

1. Northern Plains Program Reports and current projects
2. The N'washitsumbe, Capricorn and Hlangwine enclosures in the Kruger National Park
3. The dietary preferences of the roan Antelope, sable Antelope, tsessebe and African Elephant
4. The Elephant Management Policy of the Kruger National Park
5. Elephant Effects on Biodiversity
6. Statistical analysis within the ecological sciences

Statistical analysis of the data making use of Statistica, Microsoft Excel and SPSS, and other statistical software that may become necessary during the project.

Thesis on the results of the analysis, and the theory behind the methods used.

Project Risk

The project risk is very low as the data has already been collected, and the project will only involve analysis.

Researchers

Communication regarding this proposal should be directed to Victoria Goodall.

Research Officer
Victoria Goodall
Dept. of Statistics
Rhodes University
Grahamstown
6140
Phone: (046) 6038111
Cell: 083 349 4636
g00g0101@campus.ru.ac.za

Principal Researcher
Prof. Sarah Radloff
Dept of Statistics
Rhodes University
Grahamstown
6140
Phone: (046) 6038346
Cell: 082 666 5492
s.radloff@ru.ac.za

SANP Liaison
Dr. Rina Grant
Northern Plains Program
Skukuza
Kruger National Park
3150
Phone: (013) 7355524
Cell: 082 925 6299
rinag@sanparks.org

SANP Liaison
Mrs. Angela Gaylard
RSBP
Shingwedzi Camp
Kruger National Park
3150
Phone: (013) 7356685
Cell: 082 960 6176
gaylard@sanparks.org

Rhodes University Botany Liaison
Dr. Susi Vetter
Dept. of Botany
Rhodes University
Grahamstown
6140
Phone: (046) 6038595
s.vetter@ru.ac.za

CHAPTER 3: LOGISTIC REGRESSION

A generalized linear model is defined as having three components: a random component, which is the response variable (Y) with a probability distribution which can be normal, binomial, Poisson, gamma and negative binomial; a systematic component which represents the predictor variables (X) which can be continuous and/or categorical; and finally a link function which links the random and the systematic components (Quinn & Keough, 2002). Logistic regression can also be applied using categorical predictor variables which are ordinal. The versatility of logistic regression makes it a very valuable analytical technique for the analysis of categorical data.

Logistic regression is often applied in the analysis of ecological data when the response variable is binary. Examples of this would be the when the response variable is dead/alive or present/absent. A binary response variable is analogous to a categorical with only two categories.

The logit function is defined to be the link function for the generalized linear model such that this type of data can be analyzed. If $g(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots$ is the link function where the β's are the parameters to be estimated and μ is the expected value of Y; then the logit function is given by:

$$g(\mu) = \log[\mu(1 - \mu)].$$

This link function will be used for binary data and logistic regression. Logistic regression models are also referred to as *logit models* (Agresti, 2002).

3.1 SIMPLE LOGISTIC REGRESSION

Logistic regression can be split into simple logistic regression and multiple logistic regression. Simple logistic regression involves a generalized linear model with one predictor variable and a binary response variable. A binary response variable is in fact a categorical variable with two levels, in the ecological sciences this could be described as one level being equal to zero (e.g. dead) and the other being equal to one (e.g. alive). Since the response variable is binary; the model therefore is actually predicting the probability that the response Y is one for a given value of the predictor variable X. A logistic regression model is different from an ordinary regression for two main reasons. The binary nature of the response variable means that it can only have a value of zero or one; hence the regression model is essentially predicting the probability of the response variable being equal to one. The logistic regression model restricts the prediction of the response variable to between zero and one; whereas an ordinary regression model may lead to a prediction of greater than one as the predictor variables may have any unlimited range of values and this violates the definition of a probability. This means that ordinary regression is not applicable in the case of a binary response variable. Secondly, there is a problem of additivity since if the probability was thought to be a linear function, estimates could once again exceed 1 and thus the model is not applicable. Another important distinguishing feature between logistic regression and linear regression is that in the case of linear regression the mean response is linear in the parameters while for the logistic regression, a function of the mean response is linear. A usual technique for data analysis will involve a plot of the response variables against the predictor variables; however in the case of logistic regression, this will be not be helpful since the response variable only takes on values of 0 and 1 (Agresti, 2002).

The logistic regression model is defined as being a nonlinear model with a sigmoidal shape. Due to the response variable being binary, the residual errors are not normally distributed, but in fact have a binomial distribution. The method of maximum likelihood is used to estimate the parameters. The logistic model is given by:

$$\pi(x) = [\exp(\beta_0 + \beta_1 x)] / [1+\exp(\beta_0 + \beta_1 x)]$$

where β_0 and β_1 are the parameters that need to be estimated and $\pi(x)$ is the probability that the response variable is equal to one. As with linear models β_0 is the coefficient of the intercept and β_1 is the regression coefficient or the gradient which measures the rate of change of $\pi(x)$ for a specific change in X. If β_1 is positive, this indicates that $\pi(x)$ is increasing as x increases. This model can be fitted using nonlinear modelling techniques but the simpler option is to transform the model so that it resembles a linear model and then the modelling techniques are far simpler. In order to perform this transformation, the odds that an event occurs are calculated. This is in fact, calculating the probability that the response variable is equal to one relative to the variable being equal to zero. This odds ratio is given by:

$$\pi(x) / [1 - \pi(x)].$$

If the odds ratio is greater than 1 this means that the probability of the event occurring is greater than the probability of it not occurring; and then the converse is true if the odds ratio is less than 1. The logit transformation or the link function is then given by:

$$\ln\{ \pi(x) / [1 - \pi(x)]\}.$$

If $\ln\{ \pi(x) / [1 - \pi(x)]\} = g(x)$, then $g(x)$ can be modelled far more easily as:

$$g(x) = \beta_0 + \beta_1 x_i.$$

This transformation has two important results: firstly although $\pi(x)$ ranges from 0 to 1, $g(x)$ ranges from $-\infty$ to ∞ and secondly the binomial distribution of the errors is now modelled. The method of maximum likelihood is used to estimate the parameters β_0 and β_1. This is done by maximizing the likelihood function which is given by:

$$L = \Pi^{n}_{i=1}\, \pi(x_i)^{y}{}_{i}\, [1 - \pi(x_i)]^{1-y}{}_{i}\, .$$

As usual, the natural log of L is maximized for simplicity and the maximum likelihood estimators of β_0 and β_1 determined using an iterative process. The standard errors used for β_0 and β_1 are the asymptotic standard errors for large sample sizes.

3.1.1 Statistical Inference for logistic regression

The null hypothesis for a logistic regression is whether the binary response variable is independent of the predictor variable. If the response variable is dependent on the predictor variable, this means that there is a relationship between the two variables. The null hypothesis for a logistic regression is that $\beta_1 = 0$. There are two ways to test this hypothesis; firstly a Wald test statistic can be calculated, which is a Maximum Likelihood (ML) version of the *t*-test. Thus it is the parameter estimate divided by the standard error of that parameter estimate (b_1 / SE(b_1)) (Quinn & Keough, 2002). Wald's asymptotic results for ML estimators and parameter estimators in logistic regressions have large-sample normal distribution (Agresti, 2002). The standard errors are asymptotic which means that the distribution of b_1 approaches normality for large sample sizes. If the sample size is small, then the standard error should be considered as an approximation. Some sources and statistical software may call the Wald statistic the Wald *t*-statistic due to its similarity to the traditional *t*-statistic. The Wald statistic is traditionally compared to the standard normal z distribution (Agresti 1996, cited by Quinn & Keough, 2002).

The hypothesis $H_0: \beta_1 = 0$ is tested by comparing the fit of the full model:

$g(x) = \beta_0 + \beta_1 x_i$ to the fit of the reduced model $g(x) = \beta_0$.

The likelihoods are compared using a likelihood ratio statistic (Λ), which is defined as the ratio of the log-likelihood for the reduced model to the log-likelihood of the full model. By definition the larger the value of Λ, the better the fit of the model; thus for a logistic regression model, if the value of Λ is near to one then the conclusion is that β_1 contributes little to the fit of the full model. However, if Λ is much less than one then the β_1 does contribute to the fit of the model. The sampling distribution of Λ is not user-friendly, so instead a G^2 statistic is calculated. This statistic is defined as:

$$\begin{aligned} G^2 &= -2\ln(\Lambda) \\ &= -2(\text{log-likelihood reduced} - \text{log-likelihood full}). \end{aligned}$$

The G^2 is also called the likelihood ratio χ^2 statistic (Quinn & Keough, 2002).

If $H_0: \beta_1 = 0$ is true and the following assumptions hold, the sampling distribution of G^2 is very close to a χ^2-distribution with one degree of freedom. The assumptions that need to be met are that the binomial distribution adequately describes the probability distribution of the response variable – since the response variable is binary, this assumption is easily met. The second assumption is that the logistic model is appropriate and hence the model estimation will be reliable. This needs to be checked and will be discussed further in section 3.2.

The sample size is critical and the Wald statistic is most reliable when the sample size is large. If the sample size is small then an alternative strategy for hypothesis testing is preferable. In this case a method similar to that used by Ordinary Least Squares regression is used where the full and reduced models are compared and log-likelihood is used as a measure of goodness of fit, rather than least squares.

The Wald test and the G^2 can both be used to test H_0: $\beta_1 = 0$ and the two tests are different. The G^2 is recommended as the better of the two tests as the Wald statistic tends to be less reliable and lacks power for smaller sample sizes (Agresti 1996, cited by Quinn & Keough, 2002). The G^2 is sometimes termed the *deviance* and some statistical software programs will produce output for logistic regressions using the deviance term for the log-likelihood statistic, for example Statistica. The deviance is usually defined as:

[-2(log-likelihood specific model – log-likelihood saturated model)].

The saturated model is one which fits the data perfectly and has as many parameters as observations. This means that the saturated model will account for all of the variation of the data. The *full* model mentioned above is not a *saturated* model since it does not fit the data perfectly. Thus for simple logistic regression, the deviance of the full and reduced models can be compared to the saturated model can be examined and if the deviances for the two models are different this indicates that there is a difference in the two models and hence that β_1 does contribute to the model.

In practice the log-likelihood for the saturated model is not normally calculated as it will always be zero, since the model is a perfect fit. Thus the comparison mentioned in the paragraph above of comparing the differences of the deviances from the saturated model is in fact [-2(log-likelihood reduced – log-likelihood full)] which is the same as the G^2 statistic defined above. Hence the G^2, or likelihood ratio χ^2 statistic equals the difference in deviance if the two models. The deviance of a Generalized Linear Model (GLM) can be thought of as a similar measure to the $SS_{Residual}$ (Residual Sum of Squares) in a linear model; and both are a measure of the unexplained variation for the given model.

There are a number of statistical methods that can be used when fitting a logistic regression model. The most commonly used is the *Newton-Raphson* method which is an iterative process for solving non-linear equations. It begins by using an initial guess for the solution to the equations and then obtaining the second guess by approximating the function to be maximized in a neighbourhood of the original guess by a second degree polynomial. This is then repeated until the series of guesses converge to the location of the maximum in the vicinity of the original guess.

3.2 MULTIPLE LOGISTIC REGRESSION

The concepts for multiple logistic regression are the same as for the simple regression discussed above. The difference is that the multiple logistic regression involves more than one predictor variable explaining the binary response variable. Despite having numerous predictor variables, the same log-likelihood approach will be used in order to compare models. As with the simple case, the model can involve continuous are categorical predictor variables, or a combination of the two. The interaction between the predictor variables also needs to be taken into account.

The general logistic model for j parameters is given by:

$$g(x) = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_j x_{ij}$$

where $g(x)$ is the natural log of the odds ratios of the response variable occurring vs. the outcome of the response variable not occurring, as in the simple logistic regression model. β_0 is the intercept or constant and alone this would be the log of the odds of the response variable occurring if all other predictors were equal to zero. β_1 is the partial regression coefficient for X_1, holding the remaining predictors constant. In general β_k is the partial regression coefficient for the k-th predictor X_k, holding the other predictors constant.

The method for testing logistic regression models is similar to the method used for multiple linear regression models and involves first testing the overall significance of the overall regression model by comparing the log-likelihoods for the full model with that of the reduced model (β_0 only). A G^2 statistic can be calculated to test the hypothesis that at least one of the regression coefficients equals zero. Then the individual coefficients can be tested by calculating the Wald statistics for each coefficient. However the concerns about the reliability of the Wald tests still apply for the multiple regression models. In order to avoid the problems caused by the Wald statistic, a preferred method is to fit a series of reduced models to the data and compare their fit to that of the full model. If the G^2 test is significant, it indicates that the inclusion of that predictor variable results in the full model having a better fit to the data than the reduced model and hence the hypothesis H_0: $\beta_k = 0$ is rejected for the predictor variable X_k.

All the assumptions mentioned for the simple logistic regression model apply to the multiple logistic regression model. These are that the response variable is binary and the need for the logistic model to be appropriate and hence the adequacy of the model still needs to be checked. However, there are further assumptions that need to hold for the multiple logistic regression model. Since there are multiple predictor variables, it needs to be checked whether strong relationships (or collinearity) exist between these variables. This is an important feature of multiple logistic regression, and although collinearity of the predictors will not necessarily reduce the predictive power of the model; however it will result in inflated standard errors of the estimates of the model coefficients (Quinn & Keough, 2002) and hence can provide unreliable results. For categorical predictor variables, a contingency table will be needed in order to identify the collinearity; and a correlation matrix will indicate between which continuous variables collinearity exists. The existence of multi-collinearity will mean that the logistic regression model shows that no one variable is important in the model, when all the others are present. Deletion of the redundant variable is usually the best option and it will result in lowered standard error estimates (Agresti, 2002).

If a categorical predictor is used in a logistic regression, it will be included in the model as a dummy variable; and this is done automatically by most statistical software when the predictor is defined as a categorical variable. A logit model is defined as a model with a binary response variable and one or more categorical predictor variables. If all the predictors are categorical, then log-linear modelling is a more suitable process. This method is discussed further in Chapter 4.

3.3 MODEL DIAGNOSTICS

There are two type of residuals created from a logistic model. Quinn and Keough define the Pearson residual for an observation as “the contribution of the difference between the observed and the predicted value for an observation to the Pearson χ^2 statistic and it is usually expressed as a standardized residual". The second type of residual is the contribution of the difference between the observed and the predicted values of an observation to the total deviance. For large sample sizes, these two statistics will follow a normal distribution when the model is correct and residuals that are greater than 2 indicate a lack of fit. Once again, for continuous predictors, there is only a single value of the response variable for each combination of values of predictor variables and hence the sample size condition will not be met and hence the residuals will be very difficult to interpret.

3.4 GOODNESS-OF-FIT AND RESIDUALS

The goodness-of-fit test is used to determine the adequacy of the logistic model. This is done by comparing the observed and the predicted values for the response variable. The fit of a particular model is calculated as the difference between the observed and the predicted values. These differences are called the residuals. As mentioned in sections 3.1.1 and 3.2 dealing with the assumptions for both the simple and the multiple regression models, the adequacy of the regression model needs to be checked.

Goodness-of-fit statistics test that the observed data came from a population in which the fitted logistic regression model is true (Quinn & Keough, 2002). The Pearson χ^2 is based on the observed and expected (predicted) observations and Quinn and Keough state that "the χ^2 statistic for logistic regression is best visualized by treating the data as two (binary response, Y) by *n* (different values of X) contingency table. The χ^2 statistic for the goodness-of-fit is the usual χ^2 for contingency tables". The χ^2 statistic will be discussed in further detail in Chapter 4. The G^2 statistic can also be used to test the goodness-of-fit of the model; and the lower the value of both statistics, the better the fit of the model to the data; since this means that there is a small difference between the observed and the predicted values or the binary response variable. Quinn and Keough state that the G^2 and the Pearson χ^2 approximately follow a χ^2-distribution provided that the minimum frequency of either of the predicted binary outcomes is not too small. However, if the predictor variables are continuous, there will usually be one or few observations for each combination of values for the predictor variables. In this case the

Pearson χ^2 and the G^2 statistics do not follow a χ^2 distribution but they can still be used to measure the goodness-of-fit, however their *p*-values will be unreliable.

It is also possible to group the observed and fitted values according to the estimated probabilities of obtaining a "success" in the test. One approach is to form groups so that they have approximately the same size. Each group has an observed count of subjects with each outcome and a fitted value for each outcome. The fitted value of the outcome is the sum of the estimated probabilities for that outcome for all observations in that group. This construction forms the basis of the Hosmer-Lemeshow test which proposes a Pearson statistic for the comparison of the observed and fitted counts for each of the partitions (Agresti, 2002). And, the statistic is not limited by the χ^2-distribution since the observations in the group are not identical trials since they do not share a common probability of success. Unfortunately the Hosmer-Lemeshow test does not have a good power for detecting different types of lack-of-fit. When the test indicates that the lack of the fit of the model is poor, further diagnostics need to be done in order to describe the influence of individual observations on the model fit.

3.5 MODEL SELECTION

The model which fits the data the best is the model that explains the response variable most accurately, using the fewest number of predictor variables. One of the problems associated with model selection is that it becomes far more complicated as the number of parameters increases. This is due to the rapid increase in the effects and interactions as predictors are added to the model. The criteria for determining which the best model is can be done using either the Pearson χ^2, deviance (G^2) or an information criterion such as the Akaike Information Criterion (AIC). The AIC adjusts the deviance for a given model for the number of predictor variables. The AIC is defined as:

$$\text{AIC} = G^2 - n + 2p$$

where n is the number of observations and p is the number of predictors. For categorical predictors, the AIC is given by:

$$\text{AIC} = G^2 - D + 2p$$

where *D* is the number of different combinations of the categorical predictors (Larntz 1993, cited by Quinn & Keough, 2002). The best model is chosen as the one with the lowest AIC, or if numerous models have similarly low AIC's then the one involving the fewest number of parameters is chosen. Only the saturated model will fit the data perfectly, all the other models are simplified models, and hence they will not provide a completely accurate model. The best model will provide a sufficiently accurate view of reality, using as few predictors as possible in order to do so.

Other regression diagnostic tools can also be used in the analysis of goodness-of-fit of the model. These can include graphical displays of ordered residuals against normal percentiles and analyses that indicate an individual observation's effect on the fit of the parameters and the overall fit of the model (Agresti, 2002). This will indicate where one particular observation has a large effect on the fit of the model and will force the inclusion of the parameter; or it will show whether the observation is in fact an outlier and the influence of the parameter and the fit of the model will actually improve if the observation is deleted. This type of analysis is termed *Influence Measures* and Agresti (2002) discusses the following influence measures for an individual observation. *Dfbeta* is the change in the parameter estimate for each model parameter when the observation is deleted, divided by its standard error. Another influence measure *c* is defined as the measure of the change in a joint confidence interval for the parameters produced by deleting the observation; and finally a measure is calculated for the change in G^2 or χ^2 when the observation is deleted. In general, the larger the value of either of the measures defined above indicates a greater influence of the parameter.

Agresti (2002) explains that an algorithmic method for searching among models for the "best" model can be useful; in particular a forward or backward selection process similar to that used in ordinary regression. Forward selection starts with the simplest model (no interactions) and adds terms sequentially until a further addition does not improve the fit of the model. At each stage, the term which improves the fit of the model the most is added. Backward elimination starts with a complex saturated model and sequentially removes the terms which do not significantly contribute to the overall goodness-of-fit of the model. In this case the process is terminated when the reduction of another term results in a significantly worse fit of the model. If dummy variables are used in the model, then the process should consider the entire variable at any stage and

not the dummy variable as it is impractical to delete a category of a variable in the case of a dummy variable. Backward elimination is the preferred method of model selection as forward selection might result in a premature termination of the process; and it is considered more robust to start with the most complex model and work downwards.
However, statistical significance should not be the only criterion in the decision to include or delete a variable from the model. The analysis of ecological data often requires the inclusion of a particular predictor variable even through it may not necessarily be significant, in order to arrive at ecologically meaningful conclusions. Inclusion of such a variable in the model may help to reduce the bias of the other predictors (Agresti, 2002).

3.6 PREDICTIVE POWER

In ordinary regression, "R^2 describes the proportional reduction in variation in comparing the conditional variation of the response to the marginal variation" (Agresti, 2002). Together with R which describes the multiple correlation; it describes the power of the explanatory variables to predict the response. However, due to the categorical nature of the data when discussing logistic regression, no measures have been found that are as useful as the R and R^2. For any GLM, the correlation between the observed responses $\{y_i\}$ and the model's fitted values measures the predictive power of the model. It is noted that the larger the number of parameters used in the model, the greater the possibility of bias in higher values of r, the correlation. Other measures use the log-likelihood – in this case let L_M denote the maximum log-likelihood; L_S the log-likelihood of the saturated model and L_0 the log-likelihood of the null model containing only an intercept term.

As the model complexity increases, it is logic that $L_0 \leq L_M \leq L_S \leq 0$ and that this measure will fall between 0 and 1 since the probabilities will always be less than one and hence the log-likelihoods are non-positive. Agresti (2002) then shows that the measure:

$$[L_M - L_0] / [L_S - L_0]$$

falls between 0 and 1. If this measure equals 1 then the model fits the data as well as the saturated model and if it is 0 then there is no improvement in fit over the null model.

Agresti (2002) concludes by saying that a problem with this measure is that, it is not an easily interpreted scale.

3.7 SUMMARY

Logistic regression is a commonly used analytical technique for analyzing categorical data where the response variable is binary. This is a common situation in the ecological sciences. It was hoped that this technique could be used for the Kruger National Park project, however due to the nature of the data obtained from the KNP Scientific Services, this was not possible. When using a logistic regression, the prediction of the binary response variable needs to be ecologically meaningful in relation to the predictor variables.

CHAPTER 4: LOG-LINEAR ANALYSIS

Log-linear modelling is the analysis of association and interaction of categorical variables using cell counts in contingency tables; as well as hypothesis testing. These models are primarily used when at least two variables are response variables; however for a single categorical response it is simpler to use logit models. Cell count frequencies are treated as random variables with a Poisson distribution.

Log-linear models are examples of generalized linear models (GLM) where the expected cell frequencies are modelled against the variables using the log link and a Poisson error term (Quinn & Keough, 2002). A characteristic of log-linear models is that they do not distinguish response and predictor variables; all variables are considered equally as response variables. Linear models use the theory that the expected values of a series of observations can be calculated using a combination of a number of parameters. The maximum likelihood and the least squares methods are then used to estimate these parameters. These estimates are used to determine which parameters are most influential in determining the values of the observations.

The analysis of contingency tables using log-linear analysis is very similar to those methods used in the analysis of variance and linear regression. A consequence of this similarity is that there is an *interaction* term included in the model that describes the association between two or more variables within the contingency table. This association is either described as a “first-order interaction” between *pairs* of variables; or “second-order interaction” between *triplets* of variables etc. Everitt (1977) described the advantages of log-linear analysis as providing a systematic approach to the analysis complex multidimensional tables and that the analysis techniques provide estimates of the magnitude of effects of interest, consequently they allow the relative importance of different effects to be judged.

4.1 TWO-WAY TABLES

Two-way tables are constructed where sampling or experimental units are cross-classified by two variables. This means having an IxJ contingency table that crossclassifies a multinomial sample of *n* subjects on two categorical responses. The cell probabilities are π_{ij} and the expected frequencies are $\mu_{ij} = n\pi_{ij}$. Log-linear model formulas use the expected frequencies rather than the cell probabilities, so they also apply for Poisson sampling for $N = IJ$ independent cell counts Y_{ij} having $\mu_{ij} = E(Y_{ij})$. The observed cell count is denoted by n_{ij}.

For the saturated (full) model, with multinomial sampling, $\mu_{ij} = n\pi_{i.}\pi_{.j}$; let the row variable be denoted by *X* and the column variable by *Y*. The model is given by

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$$

- where λ is the mean of the logs of all the expected frequencies (a constant)
- λ_i^X is the effect of the category *i* of variable *X*
- λ_j^Y is the effect of the category *j* of variable Y
- and λ_{ij}^{XY} is the effect of any interaction between *X* and *Y*. This interaction measures deviations from independence of the two variables.

The linear model for the logarithms of the frequencies is what is known as a *log-linear* model. The saturated model is hierarchical, this means that the model includes all lower-order terms composed from variables contained in higher-order model terms; i.e. when the model contains λ_{ij}^{XY}, it also contains λ_i^X and λ_j^Y. This restriction is due to the constraints that are imposed by the maximum likelihood estimation procedures (Everitt 1977). The values taken by the “main effect” parameters simply reflect differences between the row and the column marginal totals; and so, in the context of contingency table analysis, are of little importance. This is different to the analysis of variance situation where the main effects are of greatest importance (Everitt, 1977).

For the independence (reduced) model, the formula expressing independence is multiplicative; thus $\log\mu_{ij}$ has an additive form thus:

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y.$$

Agresti (2002) explains that this is the log-linear model of independence and identifiablilty requires constraints such as $\lambda_I^X = \lambda_J^Y = 0$. The maximum likelihood fitted values for the estimate of μ_{ij} are $(n_{i.}n_{.j})/n$ – the estimated expected frequencies for Chi-squared tests of independence.

The log-linear model parameters of log-linear models are the effects of a particular category of each variable on the expected frequencies. Hence the larger the value of λ, the larger the values of the expected frequencies for that row or column. Estimation of the interaction effects is used in identifying which categories are responsible for a lack of independence between the parameters (Quinn & Keough, 2002).

4.1.1 *Interpretation of Parameters*

From Agresti (2002), in the log-linear model of independence:

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y$$

the model treats both variables jointly as responses, modelling μ_{ij} for combinations of their levels. To interpret the parameters, it is easier to treat the variables asymmetrically. In an example for Ix2 tables the logit for the i-th row is given by:

$$\begin{aligned}\text{logit}[P(Y=1|X=i)] &= \log [P(Y=1|X=i)/ P(Y=2|X=i)]\\ &= \log (\mu_{i1} / \mu_{i2})\\ &= \log \mu_{i1} - \log \mu_{i2}\\ &= (\lambda + \lambda_i^X + \lambda_1^Y) - (\lambda + \lambda_i^X + \lambda_2^Y)\\ &= \lambda_1^Y - \lambda_2^Y.\end{aligned}$$

The above equation does not depend on i; this means logit[P(Y=1|X=i)] is identical at each level of X. Thus independence implies a model of form logit[P(Y=1|X=i)] = α. In each row, the odds of response in column 1 = exp(α) = exp($\lambda_1^Y - \lambda_2^Y$).

{Note: a logit link is given by the transformation $g(\mu) = \log(\mu/(1-\mu)$}.

4.1.2 *Null hypothesis of independence*

The null hypothesis H_0 of independence in a two way table is also a test of the H_0 that $\lambda_{ij}^{XY} = 0$ (there is no interaction between the two variables). This null hypothesis can be tested by comparing the fit of the saturated model compared to the independence model. The saturated model will fit the data perfectly since the expected values are the observed values. The fit of each model is determined by calculating the expected frequencies under each model, comparing the observed and expected frequencies and calculating the log-likelihood of each model. The comparison of fit of the two models is done with the likelihood ratio statistic (Λ), that is the ratio of the two log-likelihood functions. However the sampling distribution of Λ is not well known, so instead the G^2-statistic (or likelihood ratio χ^2 statistic) is used.

G^2 = -2log (Λ) and G^2 follows approximately a χ^2 distribution for reasonable sample sizes. This can be generalized to:

$$G^2 = -2\ (\text{log-likelihood reduced model} - \text{log-likelihood saturated model}).$$

G^2 is also termed the deviance and measures the difference in fit of the two models. If the H_0 of independence is true, then the reduced model (no interaction) should fit as well as the full model and the deviance G^2 will be close to zero. If the H_0 is false, the G^2 will be greater than zero, meaning a difference in the fit of the two models. The calculated G^2 is compared to a χ^2 distribution with $(I-1)(J-1)$ degrees of freedom (Quinn & Keough, 2002).

4.1.3 *The Saturated Model*

In a saturated model, $\{\lambda_i^X\}$ and $\{\lambda_j^Y\}$ can be thought of as coefficients of dummy variables. $\{\lambda_i^X\}$ will be the coefficient of the first $(I-1)$ categories of X. Similarly, $\{\lambda_j^Y\}$ will be the coefficient of the first $(J-1)$ categories of Y. This is as a result of the constraints in both the saturated and the reduced models that $\lambda_i^X = \lambda_j^Y = 0$. The product of the dummy variables for λ_i^X and λ_j^Y has λ_{ij}^{XY} as its coefficient.

Since there are $(I-1)(J-1)$ such cross-products, $\lambda_{Ij}^{XY} = \lambda_{iJ}^{XY} = 0$; this means that $(I-1)(J-1)$ of these parameters are not redundant. Testing whether the $(I-1)(J-1)$ are equal to zero is a test of the independence of the model. This means that the residual degrees of freedom will be equal to $(I-1)(J-1)$.

The saturated model:

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$$

has $1+(I-1)+(J-1)+(I-1)(J-1) = 1+I-1+J-1+IJ-I-J+1 = IJ$ parameters, which is equal to the number of cells. Hence this model perfectly describes any $\{\mu_{ij} > 0\}$.

For the saturated model, direct relationships exist between log odds ratios and $\{\lambda_{ij}^{XY}\}$. For 2x2 tables:

$$\log\theta = \log(\mu_{11}\mu_{22} / \mu_{12}\mu_{21})$$
$$= \log\mu_{11} + \log\mu_{22} - \log\mu_{12} - \log\mu_{21}$$
$$= (\lambda+\lambda_{1}^{X}+\lambda_{1}^{Y}+\lambda_{11}^{XY})+(\lambda+\lambda_{2}^{X}+\lambda_{2}^{Y}+\lambda_{22}^{XY})-(\lambda+\lambda_{1}^{X}+\lambda_{2}^{Y}+\lambda_{12}^{XY})-(\lambda+\lambda_{2}^{X}+\lambda_{1}^{Y}+\lambda_{21}^{XY})$$
$$= \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY}.$$

Thus the $\{\lambda_{ij}^{XY}\}$ determine the association.

Unsaturated models are usually used in data analysis since they smooth the sample data and interpretation of the parameters is easier. An unsaturated model does not show a completely accurate model of the data; however, the model chosen will provide a sufficiently accurate view in order to draw accurate conclusions about the data.

4.2 THREE-WAY TABLES

For a three-way table, a saturated model is:

$$\log \mu_{ijk} = \lambda + \lambda_{i}^{X} + \lambda_{j}^{Y} + \lambda_{k}^{Z} + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}.$$

This model includes three main effects (*X, Y, Z*); three two-variable interaction (*XY, YZ, XZ*) and one three-variable interaction (*XYZ*). And, μ_{ijk} is the expected frequency in cell *ijk*; while the constant is the mean of the logs of all the expected frequencies.

Log-linear models are usually fitted in a hierarchical manner. This means that the inclusion of a higher order term means that all other lower order terms of these variables must be included. This is illustrated in Table 4.1, where if the term *XYZ* is included, then all other terms involving *X*, *Y* and *Z* are included in the model. This hierarchical sequence for adding terms is as a result of the constraints of the maximum likelihood estimation procedure.

The saturated model allows for complete dependence of the three variables by including the three-way interaction term.

Table 4.1 Range of log-linear models for a three-way table

Log-linear model	**Degrees of freedom**
$X+Y+Z$	$IJK-I-J-K+2$
$X+Y+Z+XY$	$(K-1)(IJ-1)$
$X+Y+Z+XZ$	$(J-1)(IK-1)$
$X+Y+Z+YZ$	$(I-1)(JK-1)$
$X+Y+Z+XZ+YZ$	$K(I-1)(J-1)$
$X+Y+Z+XY+YZ$	$J(I-1)(K-1)$
$X+Y+Z+XY+XZ$	$I(J-1)(K-1)$
$X+Y+Z+XY+XZ+YZ$	$(I-1)(J-1)(K-1)$
$X+Y+Z+XY+XZ+YZ+XYZ$	0

4.2.1 *Independence and Odds Ratios*

A three-way table can best be interpreted by considering it as a set of partial tables, each of which is a two-way table for each level of the third variable.

Conditional independence is where two variables are independent of each other given the level of the third variable. When the two variables are not conditionally independent, they are said to have a partial association (they are not independent for all levels of the third variable). “Bishop, Feinberg and Holland proved that a three-dimensional contingency may be collapsed over any variable that is independent of at least one of the remaining pair, and the reduced table can be examined *without* the danger of misleading conclusions” (Everitt, 1977).

Odds ratios are important in the interpretation of conditional independence in three-way tables but are more difficult to calculate because there are three variables and odds ratios can only be calculated for two by two tables. Odds ratios for larger tables can be derived by partitioning the table into two by two subsets, so conditional odds ratios can be calculated for each set of partial tables. Conditional independence between Y and Z means that all the odds ratios between Y and Z equal one.

If conditional independence between two variables does not hold, then two possible patterns may occur. Firstly, the odds ratios for two variables may all be different from one but still may be equal for all levels of the other variable (conditional dependence exists between two variables but is the same for all levels of the third variable). This pattern is called a homogenous association between two variables. A homogenous association implies no three variable interactions; and conditional independence is a special case of a homogenous association. Secondly, the pattern of dependence between two variables may differ between levels of the third variable and therefore the odds ratios for the two variables vary between the levels of the other variable. This pattern indicates an interaction between all three variables and that the two variable associations will not have a simple interpretation.

The odds ratio for an I equals two by J equals two by K table, for a given level k of K, can be estimated as:

$$\theta_{ijk} = (\pi_{11k}\, \pi_{22k}) \,/\, (\pi_{12k}\, \pi_{21k}).$$

Marginal tables are two way tables completely ignoring the third variable (the frequencies for X by Y by pooling Z). Marginal odds ratios can also be calculated using the marginal table. The effects of the individual variables represent complete independence; with no two- or three-way interaction.

4.3 INFERENCE FOR LOG-LINEAR MODELS

A log-linear model which closely fits the data provides a basis for describing and making inferences about associations among categorical responses. Standard methods apply for checking the fit and making inferences about model parameters. Fitting a log-linear model to the frequencies in a contingency table is equivalent to testing particular hypotheses about the table. Estimates of the parameters in a log-linear model may be obtained and this is a major advantage of log-linear modelling. This allows for the effects of the various variables; and the interactions between the variables to be quantified. Estimates of the parameters in the fitted model are obtained as functions of the logarithms of the expected values; similar to the estimates obtained in the analysis of variance.

A problem that may arise in analyzing contingency tables using log-linear analysis is the occurrence of zero cell counts. Since $\log(0) = -\infty$ this causes problems. A zero cell entry may occur in two ways. The first is when it is impossible to observe values for certain combinations of variables, in which case they are known as *a priori* zeros. Secondly they may arise due to sampling variation when a relatively small sample is collected for a table having a large number of cells; in this case zero cell entries are referred to as *sampling* zeros. In this case, the best method is to increase the sample size; if this is not possible, a common procedure is to increase all cell counts by 0.5 (a small constant) before proceeding with the analysis.

4.3.1 *Chi-Squared Goodness-of-Fit Tests*

The χ^2 and G^2 statistics test whether a model holds by comparing cell fitted values with the observed cell counts. The degrees of freedom are the number of cell counts minus the number of model parameters. Thus the G^2 statistic for any model represents the difference in fit of that model to the fit of the saturated model. Another method is to minimize the Akaike Information Criterion; this criterion "penalizes" the model for the number of parameters used. The Akaike Information Criterion is defined as:

$$\begin{aligned} \text{AIC} &= G^2 - (\text{df}_{\text{saturated model}} - 2\text{df}_{\text{particular model}}) \\ &= G^2 - 2\text{df}_{\text{Test of Model}}. \end{aligned}$$

The "best" model is that which minimizes either G^2 or the AIC (Quinn & Keough, 2002).

The purpose of fitting a log-linear model is to determine the model with the fewest number of parameters that adequately fits the data.

4.4 LOG-LINEAR MODELS FOR MULTI-WAY TABLES

In the same way that log-linear models for two-way tables can be extended to three-way tables; so these can be extended to higher dimension tables. However, as the number of dimensions increases, so does the complexity of the model. This is due to the increase in the number of possible association and interaction terms, making the selection of the "best" model more difficult. There is also an increase in the number of cells; which can cause difficulties with existence of estimates and appropriateness of asymptotic theory. However, all the results for the three-way classification remain essentially true in the higher dimensions (Upton, 1980). The analysis of multi-dimensional tables by fitting log-linear models requires the use of a computer program to handle the complexity of the model.

4.4.1 *Sample sizes vs. Practicality*

As in any test, a statistically significant effect need not be practically important. With very large sample sizes, it is more important to focus on estimation rather than hypothesis testing. And, in many cases, a simpler model is adequate for most purposes. As the number of dimensions of a multi-dimensional table increases, so does the number of possible models. In this case, procedures are required to determine which models may reasonably fit the data and are least likely to provide an adequate fit of the data.

4.4.2 *Model selection*

A procedure to determine which models are least likely to provide an adequate fit of the data involves examining the standardized values for the saturated model. These values may indicate which parameters can be excluded from the model and hence which unsaturated models should be investigated further. Mostly this process will indicate that many models adequately describe the data as indicated by the non-significance of the likelihood ratio criterion. In general, the model with the least parameters is preferable, however in some cases; a test between rival models may be needed in order to determine the “best” model. In general, a model including an extra parameter would only be favoured if it significantly improves the fit of the data.

4.4.3 *Dissimilarity Index*

For a table of arbitrary dimension with cell counts $\{n_i = np_i\}$ and fitted values $\{\mu_{ie} = n\pi_{ie}\}$, the closeness of a model fit to the data can be summarized by the *dissimilarity index*:

$$\begin{aligned}\Delta_e &= \Sigma_i|n_i - \mu_{ie}|/2n \\ &= \Sigma_i|p_i - \pi_{ie}|/2\end{aligned}$$

where the subscript e indicates an estimate of the parameter.

The dissimilarity index lies between zero and one; and it represents the proportion of cases that are not correctly predicted by the model. Thus, the smaller the value of the dissimilarity index, the better the fit of the model. The dissimilarity index Δ_e estimates a corresponding population index Δ describing model lack of fit. The value $\Delta=0$ occurs when the model fits perfectly; which is unrealistic in the case of unsaturated models

{This index lies between 0 and 1, with smaller values representing a better fit. It represents the proportion of sample cases that must move to different cells for the model to fit perfectly}.

4.4.4 *Fixed Marginal Totals*

Everitt (1977) stated that in certain cases, corresponding to particular hypotheses, particular sets of expected value marginal totals are constrained to be equal to the corresponding marginal totals for the observed values. In terms of the parameters in log-linear models, this means that the λ_i^X, λ_j^Y and λ_k^Z included in the model, determine the marginal constraints imposed on the expected values. For example, in some datasets, certain marginal totals are fixed by the sampling design and so the corresponding λ term must be included in the model so that the corresponding marginals of expected values are similarly fixed.

4.4.5 *Determining expected values using iterative methods*

Expected values corresponding to some models cannot be obtained directly from marginal totals of observed values since in these cases the maximum likelihood equations have no explicit solution. Various iterative methods have been discussed, including the Newton-Raphson method and the method of iterative proportional fitting. The method of iterative proportional fitting begins by assuming a starting value for each expected value of unity and then adjusting them proportionally to satisfy the first marginal constraint. The revised expected values are then adjusted to satisfy the second marginal constraint and this is continued until the differences between succeeding expected values differ by less than a fixed small amount, generally 0.01. As explained by Everitt (1977), this algorithm operates by proportionally fitting the marginal totals fixed by the model.

4.4.6 *Log-linear models for tables with ordered categories*

Log-linear analysis for contingency tables can easily be extended to deal with contingency tables in which one or more of the variables have categories that fall into a natural order; for example, a variable could involve age classes 0-5 years; 5-10years and +10 years. The λ parameters used above measure effects as deviations from an overall mean and hence sum to zero. In the case of ordered categories, these are now replaced by parameters representing linear, quadratic and, if appropriate, higher order effects. Everitt (1977) explains that this process is greatly simplified if the levels of the ordered variable can be assumed to be equally spaced, in which case orthogonal polynomials may be used. These effects represent trends in the single variable marginal totals of the ordered variable. Linear and quadratic effects may be found for the interaction between the ordered and the unordered variables; the magnitude of these effects indicates the similarity or otherwise of the trend of the ordered variable in different categories of the unordered variable.

4.5 LOG-LINEAR – LOGIT MODEL CONNECTION

Log-linear models treat categorical response variables symmetrically, focusing on associations and interactions in their joint distribution. On the contrary, logit models describe how a single categorical response depends on explanatory variables. Log-linear models are Generalized Linear Models (GLM) that treat the cell counts as independent Poisson variables, whereas logit models are GLM's that treat the cell counts as binomial. Agresti (2002) explains that the log-linear model has an identical fit as the logit model; however, it contains a general interaction term for the relationships amongst the explanatory models. The logit model does not make assumptions about the relationships between the explanatory variables and therefore it allows an arbitrary interaction for them. Logit models apply best when only one variable is a response variable while log-linear models are best when at least two variables are response variables.

4.5.1 *The Generalized Log-linear Model*

As discussed by Agresti (2002), let $\boldsymbol{n} = (n_1,\ldots, n_N)'$ and $\boldsymbol{\mu} = (\mu_1,.., \mu_N)'$ denote column vectors of observed and expected counts for the N cells of a contingency table, with $n = \Sigma_i n_i$. In this case a single index is used however the table may be multidimensional. Log-linear models for positive Poisson means have the form

$$\log \boldsymbol{\mu} = \boldsymbol{X\beta}$$

for model matrix $\boldsymbol{X}$ and column vector $\boldsymbol{\beta}$ of model parameters. The *generalized log-linear model* is

$$\boldsymbol{C}\log(\boldsymbol{A\mu}) = \boldsymbol{X\beta}$$

for matrices $\boldsymbol{C}$ and $\boldsymbol{A}$. The ordinary log-linear model results when $\boldsymbol{C}$ and $\boldsymbol{A}$ are identity matrices. Other special cases include logit models for binary or multi-category cases.

4.6 LOG-LINEAR MODEL FITTING

The sufficient statistics and the likelihood equations are derived in order to fit log-linear models. For simplicity, derivations use the Poisson sampling method which does not require a constraint on the parameters such as the multinomial does.

4.6.1 *Likelihood Equations for Log-linear Models*

Agresti (2002) explains that the fitted values for a model are solutions to the likelihood equations. The likelihood equations are derived using the general expression for a log-linear model

$$\log \boldsymbol{\mu} = \mathbf{X\beta}$$

with **n** a vector of counts, $\boldsymbol{\mu} = E(\mathbf{n})$ and $\log(\mu_i) = \Sigma_j x_{ij}\beta_j$ for all i.

Extending

$$L(\mu) = \Sigma_i\Sigma_j\Sigma_k\, n_{ijk}\log\mu_{ijk} - \Sigma_i\Sigma_j\Sigma_k\, \mu_{ijk},$$

for Poisson sampling the log likelihood function is

$$\begin{aligned} L(\boldsymbol{\mu}) &= \Sigma_i n_i \log\mu_i - \Sigma_i\mu_i \\ &= \Sigma_i n_i(\Sigma_j x_{ij}\beta_j) - \Sigma_i \exp(\Sigma_j x_{ij}\beta_j). \end{aligned}$$

The sufficient statistic for β_j is its coefficient $\Sigma_i n_i x_{ij}$. Since

$$\frac{\partial}{\partial \beta_j}[\exp(\Sigma_j x_{ij}\beta_j)] = x_{ij}\exp(\Sigma_j x_{ij}\beta_j) = x_{ij}\mu_i$$

$$\frac{\partial L(\boldsymbol{\mu})}{\partial \beta_j} = \Sigma_i n_i x_{ij} - \Sigma_i \mu_i x_{ij}\ ;\quad j = 1,2,\ldots,p.$$

The likelihood equations equate these derivatives to zero. They have the form

$$\mathbf{X}'\mathbf{n} = \mathbf{X}'\boldsymbol{\mu}$$

From a result obtained in the General Linear Model theory, these equations equate the sufficient statistics to their expected values.

4.7 SUMMARY

Log-linear models involve techniques that systematically examine the relationships that exist between variables in complex multidimensional tables and provide a powerful analysis of contingency table data. The estimates of first, second and higher order interaction effects allow for the assessment of how accurately an unsaturated model will fit the data and they also allow for the analysis of the relative importance of variable in the analysis. This allows for the reduction in the number of categories if some of them do not have a significant effect. Log-linear analysis will be utilized in the analysis of data from the Kruger National Park in Chapter 5 looking at the changes that have occurred in vegetation types in an enclosure for rare antelope.

CHAPTER 5: VEGETATION CHANGE IN THE N'WASHITSUMBE ENCLOSURE

5.1 INTRODUCTION

Rare species within the Kruger National Park (KNP) have continuously received the attention of the scientists and management of South African National Parks (SANParks) throughout the management history of the KNP. In response to the concern over the low numbers of roan antelope (*Hippotragus equinus)* (Figure 5.4), Lichtenstein's hartebeest (*Sigmoceros lichtensteinii)* and sable antelope (*Hippotragus niger)* within the KNP, three rare game enclosures have been erected over a period of time.

These enclosures have meant that other herbivores no longer have access to the vegetation within the enclosure and that predators can no longer prey upon the animals on the inside. It was noted that the sable population was disturbingly low; and the research indicated that predation was an important factor in limiting population growth (Grant & de Buys, 2004). Hence enclosures would mean that the rare antelope within the enclosure could exist without the threat of predation. The fencing of these areas leads to a reduction in the herbivory within the enclosure, in particular the absence of elephants has resulted in visible differences now existing between the vegetation of the inside and the outside of the enclosure (personal observation).

In 1967 the N'washitsumbe enclosure in the Far Northern KNP was erected, in the Vlakteplaas ranger section, near Babalala (a map of the KNP is shown in Figure 5.1). This enclosure was built to provide a predator free environment for the roan antelope so that they could breed in safety. However Lichtenstein's hartebeest and a number of sable antelope were also kept in the enclosure at various times. The N'washitsumbe enclosure was named after the creek which is a large tributary of the Shisha River. The name was connected to a hill somewhere between the Phugwane and Mphongolo rivers; and it literally means 'a species of edible bean' (Kloppers & Bornman, 2005). During a workshop in 2000; it was decided that "both the roan and the tsessebe (*Damaliscus lunatus*) populations have reached a threshold where recovery of the population without boosting was highly unlikely" (Grant & Freitag-Ronaldson, 2004). Due to this, resources and scientific study will be concentrated on these regions and populations, to facilitate the expansion of the enclosures and the optimal expansion of the rare antelope populations within the KNP.

The results from this analysis hope to provide an insight into the impact on vegetation of the current populations of browsers and grazers. It will determine whether the vegetation differences are in fact significant, and perhaps this knowledge could be used to enhance the management of the current populations of animals within the Park. This analysis will provide SANParks with some more evidence on what effect the current high populations, particularly of elephant, are having on the habitat and landscape of the KNP; as well as the vegetation changes caused by selective grazing. This will be investigated further in Chapter 6 in terms of the effect that the closure of waterpoints has had on the herbaceous composition of the veld. The closure of the waterpoints has meant that water dependent animals are no longer able to access areas during the dry season, with the hope that this will result in changes to the vegetation similar to what has happened inside the enclosure.

5.2 BACKGROUND

The roan antelope is able to go for a longer period of time without water, in comparison to most other ungulates. The erection of artificial waterpoints throughout the KNP, was putting grazing pressure on the roan, in their historical habitat since the availability of water in areas previously inaccessible to water-dependent ungulates meant that there was an increased number of species in these areas. The added number of other animals that were present in the area meant an increase in the number of predators, which also was having an impact on the roan population. From 1939 until 1980 there was a steady increase in artificial waterpoints in the KNP; with the majority of boreholes drilled between 1960 and 1980. Initially the numbers of Burchell's zebra (*Equus burchelli*) and roan antelope both increased in response to the greater availability of water. However, by 1986, the numbers of roan had started to decline alarmingly and this decline was attributed to the longer-term consequences of the increased water availability. The hypothesized reason for this decline was that the perennial availability of water had allowed the zebra to stay permanently in areas that had previously only been seasonally accessible. These grazers dominate a landscape and keep the grass layer very low, which is not optimal for the roan (Grant & de Buys, 2004). The combined effect of a long, dry climatic cycle and unusually high numbers of zebra and their associated predators had a dramatic impact on the roan population. It was also noted that the increased numbers of grazers that would gather around an artificial borehole would result in an increase in the available nutrients in the soil due to urination and defaecation (Grant & de Buys, 2004). This compounded the problem by providing a suitable environment for the development of more Increaser grass species and this would attract more animals and hence created a circle of events that were not in favour of the roan antelope.

In response to this, in 1994, 12 artificial waterpoints were closed in the prime roan antelope area in order to encourage a departure of zebra from this habitat. The closure of these waterpoints resulted in the movement of the zebra population to areas with water; and there was also a drop in the lion numbers; however the lion population has subsequently stabilized. There was also an improvement in the grass species composition and an increase in the grass biomass around the closed waterpoints and the significance of these improvements will be analyzed in Chapter 6; however the roan population has not increased as hoped. In fact the free-ranging population has stabilized, but not increased.

Since the closure of the waterpoints, it was expected that the numbers of rare antelope would steadily increase. However, since the closure, the roan numbers have stabilized, while the sable and tsessebe numbers have declined further. In fact, the tsessebe numbers have declined alarmingly; dropping from 70 animals in 1993 to 20 animals in 2000. A detailed analysis of the vegetation surrounding these closed waterpoints will be done in Chapter 6.

The N’washitsumbe enclosure has been identified as a valuable area to monitor vegetation changes as elephants have been excluded from the enclosure since its construction in 1967 (Hofmeyr & Eckhardt, 2004).

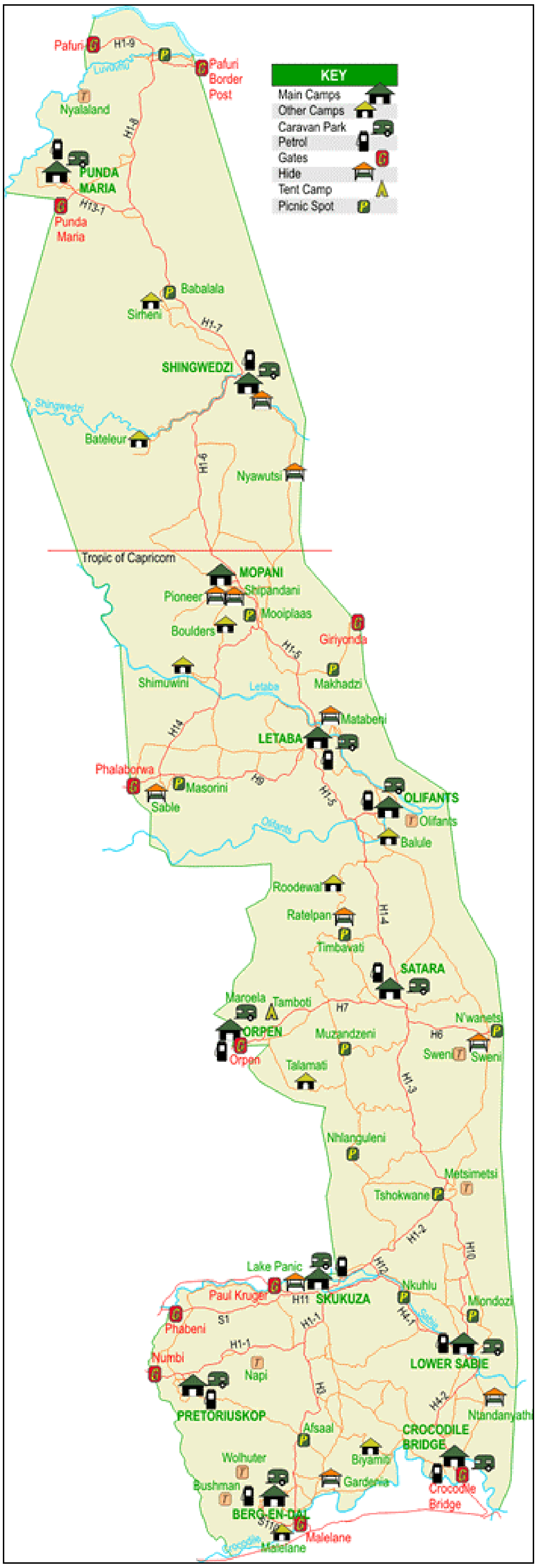

Figure 5.1 Map of the Kruger National Park showing tourist roads and rest camps; from the SANParks website (http://www.sanparks.org/parks/kruger/tourism/map.php)

There are currently 3 enclosures within the KNP for the rare antelope. The N'washitsumbe enclosure was originally 254ha when it was erected in 1967. In 1984 a further 48ha was included; which contained an adjacent vlei. There are still a number of roan antelope that are present in the areas adjacent to the enclosures and they are constantly being monitored. The surplus antelope from within the enclosures are re-located into these areas.

The Capricorn enclosure in the Mooiplaas section, near Mopani Rest Camp of the KNP was completed in 2002; and is 500ha in area. Initially the enclosure contained 8 tsessebe that had originally lived in the area and 19 roan antelope were translocated from the N'washitsumbe enclosure. In September 2004, the Capricorn enclosure contained 25 roan antelope and 27 tsessebe. Unlike the N'washitsumbe enclosure which has obvious differences in appearance of the veld between the inside and outside of the enclosure, there are currently very few differences in the vegetation composition between the inside and the outside of the Capricorn enclosure. Veld condition assessments (VCA's) have been conducted in the Capricorn enclosure and these will provide a valuable base dataset in the future as differences become more apparent due to the absence of browsers and bulk feeders inside the enclosure.

The population of the roan antelope within the N'washitsumbe enclosure is shown in Figure 5.2. The data for this graph were obtained from "Northern Plains Programme" by Rina Grant and Abri de Buys (2004).

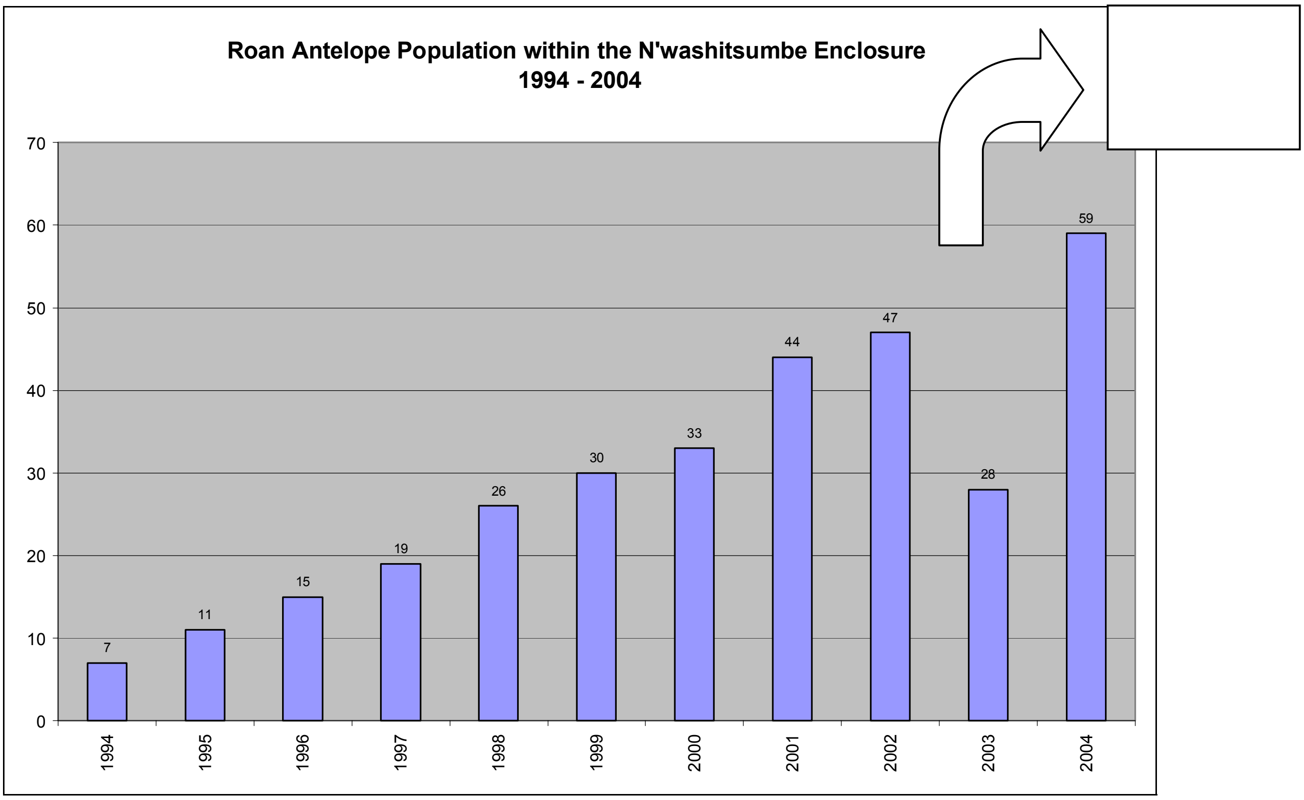

Figure 5.2 Graph showing the numbers of roan in the N'washitsumbe enclosure

Animal densities peaked inside the N'washitsumbe enclosure at 3.4ha per animal in 1974 but the camp was cleared of animals and rested in 1974. Roan antelope densities have varied over the years; from 43ha per animal in August 1994 to 6.8ha per animal at the time of the study. When the N'washitsumbe enclosure was first erected, there would have been no difference between the vegetation structure and composition between the inside and the outside of the enclosure as this area was merely part of the entire KNP. However presently, 39 years later, there are visible differences (personal observation). The main factors that have resulted in this difference are the absence of elephants and other bulk grazers and browsers within the enclosure. These mammals can cause noticeable differences in the appearance of the veld; while the presence of selective grazers only within the enclosure will also have contributed to the visible differences.

There is another enclosure in the south, near Pretoriuskop, called Hlangwine, which focuses mainly on the breeding of tsessebe. This enclosure is 220ha and was erected in 1972; when it was originally built to house eland. However, since 1994 it has been used primarily for tsessebe. The purpose of these 3 enclosures is currently being redefined and a longer-term approach is being investigated. The new approach will take into account the KNP objectives hierarchy.

Kruger National Park Rare Biota Objective – revised draft, 2004

To prevent the extinction within the Kruger Park of any species on the IUCN's global critically endangered or endangered lists[1]*, and to work with other conservation initiatives to secure and strengthen the future of such species over their historic distribution ranges. To put in place appropriate monitoring an conservation efforts of other threatened*[2] *species or lower taxonomic division*[3]*, including considering recommendations of experts of invertebrate taxa for which no formal redlisting has been done, according to a realistic framework. Except in crucial instances for the survival of globally critically endangered species, management for system integrity and biodiversity must take precedence over species management.*

[1]Until such time as the plants have been assigned to the latest IUCN criteria they will be evaluated on the old system; [2] Threatened includes critically endangered, endangered and vulnerable according to the IUCN classification; [3]Includes subspecies, variant or population.

Figure 5.3 Kruger National Park Rare Biota Objective – revised draft, 2004 (Grant & Freitag-Ronaldson, 2004)

In 2000, at the conclusion of a workshop on rare antelope conservation (Grant and van der Walt, 2000 cited by Grant & de Buys, 2004) it was stated that "the general consensus was that the current free ranging roan are the most valuable animals and these herds should be strengthened as soon as possible. Both the roan and tsessebe populations have reached a threshold where recovery of the population without boosting is highly unlikely. At this stage resources will be concentrated on the use and expansion of existing enclosures allowing the roan and tsessebe in the enclosures to increase sufficiently to supply a large group to be released simultaneously. Before animals are released the habitat and available forage will be evaluated to determine whether the environment into which these rare antelope will be released is ideal". The goal of this workshop can be obtained using effective monitoring of the current situation within the enclosures; and it is hoped the closure of waterpoints throughout the KNP will result in a vegetation composition that will favour the roan; and thus result in a strengthening of the population numbers.

The aim of this research was to quantify the differences between the inside and outside of the enclosure, focusing only on the grass layer and not the trees.

Figure 5.4 *Hippotragus equinus* - roan antelope (http://www.outdoorphoto.co.za/forum/photopost/showphoto.php)

5.3 RARE ANTELOPE

The roan antelope is the second largest antelope species in Africa after the eland (*Taurotragus oryx*). It has a general grey-brown colouring, often with a reddish tinge. The under parts are lighter with the face having distinct black and white markings. The long, narrow ears have prominent tassels of hair at the tip. The tail is long and tufted; and it has a light-coloured but darker-tipped mane that runs from between the ears to just beyond the shoulders. Both sexes carry back-curved horns but those of the female are lighter and shorter than the bull's (Stuart and Stuart, 1992).

This species is restricted to the northern and north-eastern areas of southern Africa where it is considered rare. Beyond the Southern African Subregion it occurs widely in central Africa, western East Africa and through the savanna zone to West Africa. However despite its wide distribution, it is considered to be rare and endangered throughout much of its range. The distribution of the roan antelope in Africa is shown in Figure 5.5.

The roan antelope requires open or lightly wooded grassland with medium to tall grass and limited access to water. They avoid areas with short grass. They usually live in small herds of between 5 and 12 animals usually led by an adult bull. However larger herds of up to 80 animals have been recorded. Nursery herds of cows and young calves occupy a fixed territory which is defended by dominant bulls. The herd is led by a matriarch cow. Roan antelope are grazers, selecting medium or long grasses. They will seldom browse; they have even been recorded to completely submerge their heads to feed on underwater plants (Apps, 1992). They chew bones for their calcium and phosphorous content. Calves are born at any time of the year; the mother will stay with the calf for the first few days and then she returns to the herd, visiting the calf morning and evening to suckle and clean it. After feeding the calf goes alone to a new hiding place so that predators cannot find it by following the mother's scent; as the calf itself has no scent.

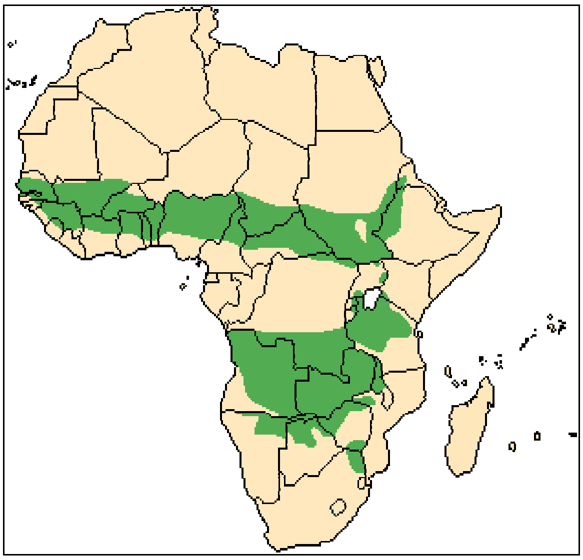

Figure 5.5 Map showing the distribution of the roan antelope (http://www.wild-about-you.com/GameRoan.htm)

The sable (Figure 5.6) is a large antelope with an adult bull being black with white underparts and inner thighs. Younger bulls and females are usually reddish-brown instead of black. They have distinct facial markings; the face is mainly white, with a broad black stripe from the forehead to the nose and a thinner black stripe from below the eye almost to the muzzle. They have an erect mane that runs from the top of the neck to just beyond the shoulders. The ears are long and narrow but lack the tufted tips which the roan have. Both sexes carry horns; however the male's are longer and thicker. The horns rise up from the scull and then sweep backwards in a pronounced curve.

Sable are usually associated with dry, open woodland with medium to tall grass. They avoid dense woodland and open grassveld. Water is essential; they are predominantly grazers however they will turn to browsing and fruit in times when grazing is unavailable (Fourie, 1984). The sable antelope usually lives in herds from between 10 and 30 animals; but occasionally larger groups come together. Territorial bulls establish themselves in territories overlapping those of nursery herds (cows and calves) that move in a home range. Sable are most active during the early morning, late afternoon and early evening.

Figure 5.6 *Hippotragus niger* - Sable Antelope (http://www.bushveld.co.za/animal-pictures-sable.htm)

A sable cow leaves the herd to give birth in dense cover where the calf remains for the first two weeks. The mother visits the calf once or twice a day to suckle and clean it. The calf moves to a new hideaway every day to avoid building up an odour that a predator could detect. The sable’s main enemies are lion, leopard, cheetah and wild dog. However sable can defend themselves viciously and even lions have to be careful when attacking a fully-grown bull. Calves and females are the main target of predators.

Sable are restricted to the north-eastern parts of southern Africa however their distribution is patchy and not continuous. They occur as far north as southern Kenya and marginally in south-western Angola. An isolated population occurs in northern Angola and is considered to be a separate sub-species, the so-called Giant Sable.

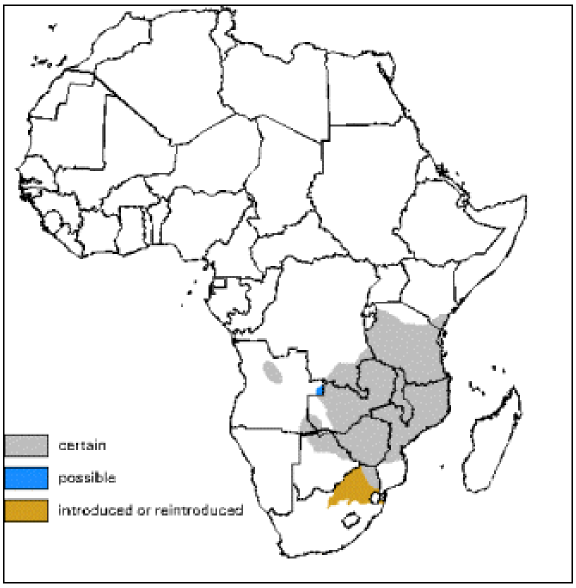

Figure 5.7 Map showing the distribution of the sable antelope (http://www.wild-about-you.com/GameSable.htm)

5.4 STUDY AREA

Research was conducted on the Vlakteplaas section in the Far Northern Region of the KNP in the Limpopo Province of South Africa; in the area within and surrounding the N’washitsumbe enclosure (Figure 5.8). The research was done as part of the Northern Plains program, and was carried out be Abri de Buys.

This region on average receives 515mm of rain annually, mainly during October to March. The summer months often experience temperatures in excess of 40 degrees Celsius during the day and during winter the temperature never drops below freezing (de Buys, pers comm.). Within the enclosure there is a high local density of the roan antelope and a near absence of other grazers (Grant & de Buys, 2004).

Fig. 1. Layout of the N'washitshumbe roan enclosure. The Marula tree section is now very obviously different from the area outside the enclosure, where most of the large Marula trees have disappeared. The roan antelope tend to use the ecotone between the drainage line and the upland very intensively as can be seen by the amount of faeces deposited there.

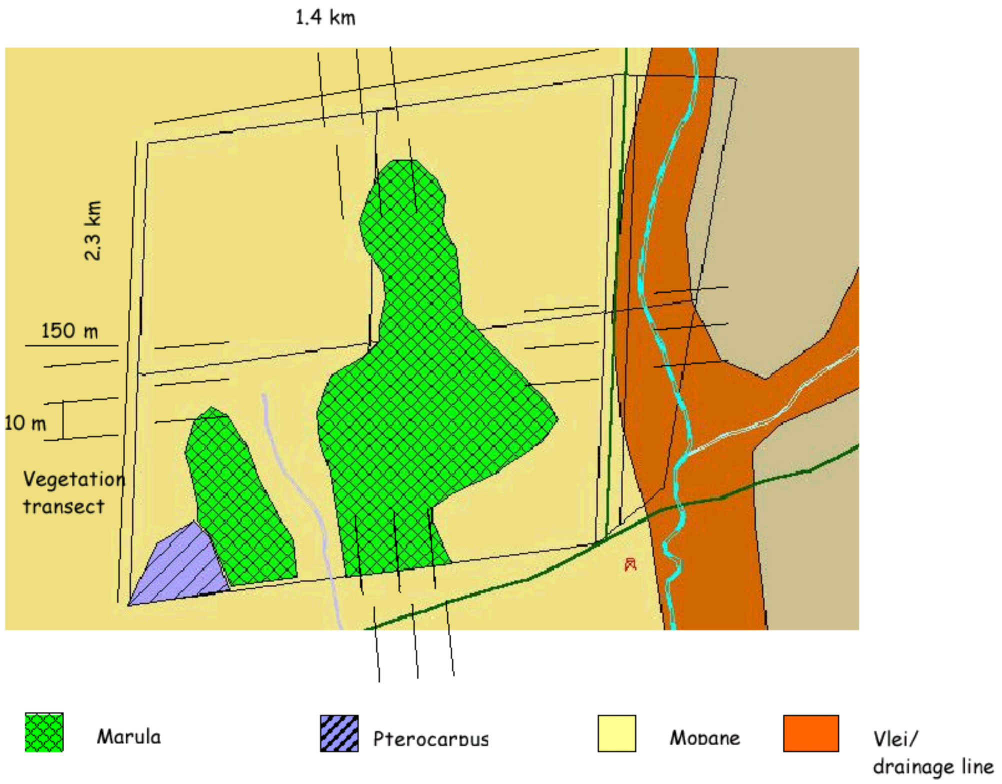

Figure 5.8 Map showing the layout of the N'washitsumbe enclosure (Grant & Freitag-Ronaldson, 2004)

5.5 DATA COLLECTION

The research was conducted along the boundary fences of the enclosure, using a 150m long transect. A set of 3 transects were done along each of the 4 fence lines, resulting in a total of 12 transects. This was done both inside and outside the enclosure. Within each transect, at 2m intervals, a pointer was stuck into the soil and the closest plant was then classified into one of 3 grass categories. This resulted in 75 points being recorded in each transect.

In the grass layer, different species have different requirements for survival and the presence of these species therefore indicate the presence of the factors that they require. The method used, classifies the grass into one of 3 main categories namely Increaser 1, Increaser 2 and Decreaser. However the N'washitsumbe enclosure is located in one the drier areas within the KNP and Increaser 1 species do not occur in this part of the veld due to the nutrient rich soils that occur in this area. Increaser 1 species are more prevalent in the wetter areas of the KNP in the regions around Punda Maria and Pretoriuskop. Hence the grass species within the N'washitsumbe enclosure are classified as either an Increaser 2, or a Decreaser. There are also categories for forbs (weed), and if there are no plants within 25cm of the pointer, it is classified as bare ground. This grass classification was first developed by Dyksterhuis (1949). However in the overall classification process, forbs and bare ground are classified as Increaser 2's. The same survey procedure was used both inside and outside the enclosure. It was recorded which fence line, the data were recorded against (e.g. north / south / east or west). Due to most of the environmental variables such as rainfall, soil and wind being the same both inside and outside the enclosure, a difference between the fence line data would indicate a finer scale difference within the camp. This would indicate that herbivores are responsible for small scale patches of different vegetation.

At each survey point, a reading was taken with a disc pasture meter. This measures the compressed grass height under a disc of a certain weight. This is used to calculate the grass biomass from a regression analysis to convert the settling height of the disk to dry mass per unit area (in kg.ha^{-1}). In total 900 readings were taken inside the enclosure and 900 readings outside.

Table 5.1 shows the list of species that were recorded during the assessment both inside and outside the enclosure. The most abundant species within the enclosure were *Panicum maximum* and *Panicum coloratum* while the most abundant species outside the enclosure were *Schmidtia pappophoroides, Themeda triandra* (Figure 5.9) and *Urochloa mosambicensis.*

Figure 5.9 *Themeda triandra* (http://www.sabisandslodges.co.za/Plants.asp)

Table 5.1 Grass species sampled in the N’washitsumbe enclosure

Scientific Name	**Common Name**
Aristida congesta subsp. barbicollis	Spreading Three-awn
Aristida canescens	Pale Three-awn
Bothriochloa radicans	Stinking Grass
Cenchrus cilliarus	Blue Buffalo Grass
Digitaria eriantha	Common Finger Grass
Enneapogon cenchroides	Nine-awned Grass
Eragrostis rigidior	(Broad) Curly Leaf
Eragrostis superba	Saw-tooth Love Grass
Heteropogon contortus	Spear Grass
Ischaemum afrum	Turf Grass
Panicum coloratum	Small Buffalo Grass
Panicum maximum	Guinea Grass
Schmidtia pappophoroides	Kalahari Sand Quick
Setaria incrassata	Vlei Bristle Grass
Setaria sphacelata	Common Bristle Grass
Sporobolus fimbriatus	Dropseed
Sporobolus ioclados	Pan Dropseed
Sporobolus pyramidalis	Catstail Dropseed
Themeda triandra	Red Grass (Rooi gras)
Urochloa mosambicensis	Bushveld Signal Grass
Urochloa panicoides	Herringbone Grass

5.6 METHODS

Initially it was hypothesized that the frequency of Increaser 2 species was different between the inside and the outside of the enclosure; and a similar hypothesis was investigated for the Decreaser frequencies. This would be done by determining if there is a significant difference in the frequency of Increaser 2 species inside and outside the enclosure; and repeating the analysis for the Decreasers. This would indicate that selective grazing has a significant impact on the type of vegetation that is most abundant and in this way alters the natural biodiversity of the ecosystem. In order to test these hypotheses, a log-linear analysis was run; and then χ^2-tests were then run using the same data, except fixing it for one of the factors. The grass biomass was analyzed by using a *t*-test to determine whether there is a significant difference between the inside and the outside of the enclosure.

5.7 RESULTS

5.7.1 *Grass composition*

The results from the transects were placed into a contingency table (Table 5.2) and graphed in Figure 5.10.

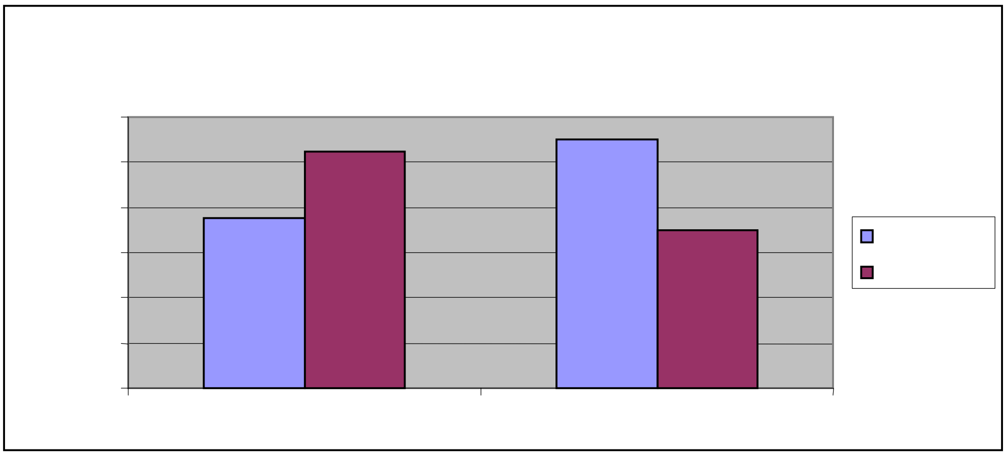

Figure 5.10 Graph showing comparative grass composition frequencies

Table 5.2 Grass species classification frequencies

Contingency Table for Position vs. Location vs. Classification

Position	Location	Classification	Frequency
East	Inside	Decreaser	119
East	Inside	Increaser	106
East	Outside	Increaser	120
East	Outside	Decreaser	105
North	Inside	Decreaser	118
North	Inside	Increaser	107
North	Outside	Increaser	177
North	Outside	Decreaser	48
West	Inside	Decreaser	120
West	Inside	Increaser	105
West	Outside	Increaser	124
West	Outside	Decreaser	101
South	Inside	Decreaser	164
South	Inside	Increaser	61
South	Outside	Increaser	130
South	Outside	Decreaser	95

From Figure 5.10 we can see that there appears to be a noticeable difference between the frequency of both the Increaser 2 and the Decreaser species between the inside and the outside of the enclosure.

A log-linear analysis was then run; and this could be used to determine if the differences observed in Figure 5.10 are, in fact, statistically significant. It is expected that the Decreaser species will be more abundant within the enclosure since by their definition they decrease under heavy utilization. Then Increaser 2 species is expected to be more abundant outside the enclosure as here there will be more homogenous grazing due to the variety of grazers that exist outside the enclosure in comparison to the limited number of species within the enclosure and these are all selective grazers.

The first analysis run was testing the complete independence of the 3 variables, namely the position (**P**) (north / south / east / west); its location (**L**) (inside or outside the enclosure) and the classification (**C**) (Increaser 2 or Decreaser).

Table 5.3 Fitted values from the log-linear analysis

FITTED VALUES Increaser 2 vs. Decreaser Classification							
			Log-linear Model				
Position	**Location**	**Classification**	**(P,L,C)**	**(PL,C)**	**(PC,LC)**	**(PL,PC,LC)**	**(PLC)**
North	In	Increaser 2	116.25	116.25	115.74	121.55	107.00
		Decreaser	108.75	108.75	99.41	103.45	118.00
	Out	Increaser 2	116.25	116.25	168.26	162.45	177.00
		Decreaser	108.75	108.75	66.59	62.55	48.00
South	In	Increaser 2	116.25	116.25	77.84	74.00	61.00
		Decreaser	108.75	108.75	155.10	132.52	164.00
	Out	Increaser 2	116.25	116.25	113.16	117.00	130.00
		Decreaser	108.75	108.75	103.90	88.48	95.00
East	In	Increaser 2	116.25	116.25	92.10	90.98	106.00
		Decreaser	108.75	108.75	134.14	134.02	119.00
	Out	Increaser 2	116.25	116.25	133.90	135.03	120.00
		Decreaser	108.75	108.75	89.86	89.98	105.00
West	In	Increaser 2	116.25	116.25	93.32	92.48	105.00
		Decreaser	108.75	108.75	132.35	151.00	120.00
	Out	Increaser 2	116.25	116.25	135.68	136.52	124.00
		Decreaser	108.75	108.75	88.65	108.00	101.00

The fit for the model (PC, LC) is the closest to the observed data; where the observed data are given by the fitted values for the saturated model (PLC). The model (PL, PC, LC) is the next closest fit while the other models fit poorly. A good fitting model provides the basis for describing and making decisions about associations among the categorical responses. To do this, Chi-squared Goodness-of-Fit tests were run; where the χ^2 and G^2 statistics test whether a model holds by comparing the fitted values to the observed counts. This is done to confirm the assumptions made by visually analyzing the data in Table 5.3. A more detailed explanation of the theory involved in log-linear analysis is given in Chapter 4.

For the grass composition data captured along the various transects, Table 5.4 shows results of testing; fitting poorly for several log-linear models. Models that lack any association term fit poorly; while the model (PC, LC) is the best fitting model, although in this case, it would be more accurately described as the 'least inaccurate' model.

Ecologically this can be explained as there being very little or no correlation between the position, location and classification of the grasses.

Table 5.4 Goodness-of-Fit results

Goodness-of-Fit Tests for Log-linear Models in Table 5.3

Model	G^2	**df**	**p-value**	χ^2	**df**	**p-value**
(P,L,C)	136.73	10	<0.000001	130.34	10	<0.000001
(P,LC)	70.51	9	<0.000001	67.13	9	<0.000001
(L,PC)	96.96	7	<0.000001	95.03	7	<0.000001
(PL,C)	136.73	7	<0.000001	130.34	7	<0.000001
(PC,LC)	**30.73**	**6**	**0.000029**	**30.34**	**6**	**0.000034**
(PL,LC)	70.51	6	<0.000001	67.13	6	<0.000001
(PL,PC)	96.96	4	<0.000001	95.03	4	<0.000001
(PL,PC,LC)	29.20	3	0.000002	29.02	3	0.000002
(PLM)	0.00	0	1.00	0.00	0	1.00

The (PC, LC) model can be explained as the conditional independence between the position and the location, which is ecologically meaningless as these factors are determined by the experiment and are not variables which we wish to be explained via the log-linear analysis. From this we can conclude that the apparent differences between the frequencies of Increaser 2 and Decreaser vegetation cannot be explained by the position and location of the transect and hence no model can accurately predict the classification, given the position and location.

As no significant result could be determined from running the log-linear analysis looking at the interaction between the position, location and classification – it was decided to fix the position and test whether the classification can be determined by the location, at a specific position. To do this, the data were placed into a contingency table for each position and analyzed using a χ^2-test.

A χ^2-test was run for the north position and gave the following results (Table 5.5). The null hypothesis is that the species classification is independent of the location either inside or outside the enclosure. The χ^2 statistic for testing this hypothesis is 46.8 with one degree of freedom and $p<0.001$. Therefore we have sufficient evidence to conclude that there is a location effect on the distribution of Decreaser and Increaser 2.

Table 5.5 Grass composition for the North showing row percentages

Location vs. Grass Composition	Increaser 2	Decreaser
Inside	47.56%	52.44%
Outside	78.67%	21.33%

The analysis was repeated with the data fixed for the transects done along the south of the enclosure. The data are tabulated in Table 5.6.

Table 5.6 Grass composition for the South showing row percentages

Location vs. Grass Composition	Increaser 2	Decreaser
Inside	27.11%	72.89%
Outside	57.78%	42.22%

The χ^2 statistic for testing hypothesis for the south is 43.3 with one degree of freedom; and $p<0.001$. Once again, we have significant evidence that the grass species distribution for the south of the enclosure is dependent on the location inside or outside of the enclosure.

Table 5.7 Grass composition for the East showing row percentages

Location vs. Grass Composition	Increaser 2	Decreaser
Inside	47.11%	52.89%
Outside	53.33%	46.67%

The χ^2 statistic for testing the same hypothesis for the eastern side of the enclosure is 1.7 with one degree of freedom and $p=0.187$. Hence we do not have significant evidence to prove dependency of the grass species distribution on the location. This means, that for the eastern fenceline of the enclosure, the species distribution is independent of its position either inside or outside the enclosure.

Table 5.8 Grass composition for the West showing row percentages

Location vs. Grass Composition	Increaser 2	Decreaser
Inside	46.67%	53.33%
Outside	55.11%	44.89%

The data for the western side of the enclosure are recorded in Table 5.8. For this analysis, the χ^2 statistic for testing the hypothesis for the western fence line is 3.2 with one degree of freedom and p=0.073. Once again we do not have sufficient evidence to reject the hypothesis of independence of the grass species distribution in the west. The data for all the above χ^2 tests are summarized in Table 5.9.

Table 5.9 Summary of results for fixed position analysis

Summary of Nonparametric χ^2 Tests

Fixed Position	χ^2 statistic	df	p
North	46.8	1	<0.001
South	43.3	1	<0.001
East	1.7	1	0.187
West	3.2	1	0.073

From the data in Table 5.9, we can see that there is a statistically significant difference in the grass species distribution between the inside and the outside of the enclosure in the north and the south of the enclosure; however, there is no statistically significant difference in the eastern and western fenceline. In 1984 an additional 48ha was added to the enclosure along the eastern boundary to include the adjacent vlei and vlei-upland boundary/ecotone vegetation. This combined effect of recent inclusion and different vegetation type could explain why there is no significant difference between the grass composition inside and outside of the enclosure. The animals outside the enclosure would be expected to utilize the vlei area; and the roan could also be expected to use this area within the enclosure as, particularly in winter, this might be the only green grass available and this would of course be more nutritious than the dry grass. This consistent utilization of the veld both inside and outside the enclosure; and this area's recent inclusion into the enclosure could explain why there is no significant difference in the grass composition along this fence line. The result for the western boundary of the enclosure is unexpected as the north and the west have similar soil colour and vegetation (de Buys, pers comm.). This lack of difference in the west would need to be further analyzed in order to find a more ecologically robust reason for the similarity in grass composition.

It was then investigated whether the grass species distribution can be explained by the location inside or outside the enclosure only. To determine this, all the data for the enclosure were lumped together for the north, south, east and west and a χ^2 test was run. For this analysis, the hypothesis being investigated is that the grass species classification is dependent on the location; irrespective of the position.

Table 5.10 Result of the χ^2 test for the N'washitsumbe enclosure

$\chi 2$ **statistic**	**df**	***p***
65.8	1	<0.001

The results (Table 5.10) obtained provide a χ^2 statistic of 65.8 with one degree of freedom and $p<0.001$. Thus we have significant evidence that the grass species distribution for the enclosure as a whole is dependent on the location. This means that at the higher-level scale of the entire enclosure, there are significant differences in the vegetation; and this has been caused by the selective grazing within the enclosure and the impact of elephants and bulk grazers outside the enclosure.

5.7.2 Grass Biomass

At each survey point a reading was taken with a disc pasture meter, which was used to calculate the grass biomass measured in kg.ha^{-1}. In total 900 readings were taken inside the enclosure and 900 readings outside. It can be assumed that these readings inside and outside the enclosure are independent as the weather conditions, fire patterns and rainfall will be identical in both areas.

For this analysis, we hypothesize that there is a significant difference in the grass biomass inside and the grass biomass outside the enclosure. This would indicate that there has been a significant impact due to the selective grazing of the roan antelope within the enclosure. It is expected that the grass biomass would be lower within the enclosure due to the higher density of selective grazers within the enclosure.

The data were collated and a *t*-test for independent samples was required. Firstly Levene's test for homogeneity of the variances was run. The result of Levene's test gave: F=7.7 with (1,1798)df and p=0.006. We therefore, reject the hypothesis of homogenous variances and conclude that the variances are significantly different from one another.

Due to the heterogeneity of the variances, it was investigated whether a log transformation of the data would result in a distribution that was closer to a normal distribution with homogeneity of variances. The natural logarithm of all values were calculated; and then Levene's test for homogeneity of the variances was run with the following results: F=0.6 with (1,1798)df and p=0.427. We therefore fail to reject the hypothesis and conclude that the variances are homogenous; and hence we could run a *t*-test using a pooled sample variance. Next the Shapiro-Wilks test for normality was run. The hypotheses tested whether the transformed data, for both the inside of the enclosure and the outside of the enclosure, followed normal distributions.

The results for the inside data gave W=0.976 and p<0.001 and hence we have significant evidence to show that the log transformed data from within the enclosure is not normally distributed. The results for transformed data from outside the enclosure gave W=0.994 and p<0.001 and once again we have significant evidence to show that the log transformed data from outside the enclosure are not normally distributed. However, the histograms revealed that the data approximates a normal distribution quite closely except for some extremely low values which are causing the test for normality to fail. It was decided to run the *t*-test for a pooled sample variance, since this test is reasonably robust to the normality assumption; and at least the variances had been stabilized. The results of the t-test gave t=5.504 with 1798df and p<0.001. The means and standard deviations of the biomass were 14.15±1.9kg.ha^{-1} for the inside of the enclosure and 16.61±1.8kg.ha^{-1} for the grass biomass outside the enclosure. This indicates a difference of approximately 17% between the grass biomass of the veld inside and outside the enclosure. Although this is not a dramatic difference, it is a statistically significant difference.

From these results there is sufficient evidence to conclude that there is a significant difference in the grass biomass inside and outside the enclosure and this may have resulted from the constraint of a high density of selective grazers within the enclosure. We can therefore conclude that any increased density of a particular species may have a significant impact on the vegetation and intensive grazing of a particular grass type may reduce the biomass of the grazing. This result shows that too many of a particular grazer in a small area may reduce the biomass of the vegetation and may also have a dramatic impact on the proportions of the grasses available. This confirms previous research that has found that large concentration of zebra and wildebeest around waterpoints will change the type of vegetation available and this is not optimal for the roan antelope. In the same way, the roan when protected in an enclosed area may result in a transformation of the vegetation in order to favour their own feeding requirements.

5.8 SUMMARY

In the late 1990’s a number of artificial waterpoints throughout the Park were closed in order to allow selective grazers to reclaim the areas that had become available to the bulk grazers due to the increased availability of water. The above study has shown that there has been a change in the vegetation proportions when the grazing is only available to selective grazers. It is expected that similar results will begin to show in the areas outside the enclosures around the artificial waterpoints, where although it is open to all the animals within the KNP; only the selective grazing, water independent animals will be able to utilize these regions. This will be investigated in Chapter 5 to see whether changes in vegetation proportions are already becoming apparent.

The N’washitsumbe enclosure is a very valuable source of ‘control’ data for the KNP since it has been carefully monitored and the changes that have occurred to the grass and tree proportions have been investigated in numerous studies. The Capricorn enclosure will also be very useful in order to monitor the time frame over which the changes begin to be significant as this enclosure has had VCA’s conducted as it was erected and regularly thereafter.

CHAPTER 6: ARTIFICIAL WATERPOINT CLOSURES

6.1 INTRODUCTION

From 1939 until 1980, artificial waterpoints throughout the Kruger National Park (KNP) were erected. This eventually led to water being available in many areas that had previously not had a consistent supply of water throughout the year. The theory behind the artificial water sources was that it would provide a constant supply of water throughout the KNP during the low rainfall periods. Since the erection of the western boundary fence in 1970, natural migration routes have been cut-off for the animals. Usually animals, particularly Zebra (*Equus burchelli*) and Blue Wildebeest (*Connochaetes taurinus*) would migrate to the private game reserves to the west of the KNP during the dry periods which they were now unable to do and many died against the fence. Hence it was decided to provide many more permanent water sources for the animals (Travers, 2006). By the end of the 1980's more than 300 boreholes had been drilled, 50 earth dams were constructed and various seasonal and perennial rivers dammed (Travers, 2006).

However, the artificial water points meant that bulk grazers such as the zebra were dominating areas that had previously been inaccessible, particularly during the dry seasons. A decision was made to close many of the boreholes and it is hoped that the population of water independent grazers, such as the roan, will be able to strengthen in these areas. Figure 6.1 shows the location of the boreholes that have been closed throughout the Park. The majority of the closures have been made in the northern regions of the KNP which is the preferred territory of the roan antelope.

It was noted that the "most rational and reliable way to detect overgrazing is to recognize the replacement of one type of plant cover by another" (Dyksterhuis, 1949); and this theory will be used to determine whether the effects of overgrazing are being reversed due to the closure of the boreholes. This project will examine whether changes in the vegetation structure and composition have occurred in the areas surrounding the boreholes since the closures; this is a similar study to that done in the N'washitsumbe enclosure, discussed in Chapter 5. If the results show that there is no significant difference in the vegetation composition since the closure of the boreholes, this does not mean the borehole closures have not achieved the desired goal; but rather that it is a long term project and that a similar analysis will need to be done in a few years time, and the changes might then be evident. It is expected that by the end of the borehole closures, only 50 open artificial waterpoints will remain; with the majority of the closures happening in the regions far away from tourist roads and other human activity (Travers, 2006).

6.2 BACKGROUND

Studies done in the northern areas of the KNP have investigated the hypothesis that the even distribution of permanent water has allowed water dependent bulk feeders, such as zebra and wildebeest, to increase in number (Grant & de Buys, 2005). The availability of water makes vegetation available that would previously have been too far away from water for many animals during the dry seasons. It has also been hypothesized that this intense utilization of the forage has resulted in changes in the herbaceous and woody structure composition (Grant & de Buys, 2004). This is due to the fact that intense grazing promotes the dominance of lawn type grasses that increase under heavy grazing (Grant & de Buys, 2004). This type of forage is preferred by species such as impala (*Aepyceros melampus*); but decreases the forage available to selective grazers like the roan and the sable (*Hippotragus niger)* which prefer taller grass species. Chapter 5 contains a more detailed description of the roan and sable antelopes and the vegetation that they favour. The increased availability of water also attracts larger numbers of elephants into these areas. They are destructive feeders that can alter the appearance of a landscape due to the decline in tall trees; this is discussed in more detail in Chapter 7. Elephant can also result in more open watercourses which further hamper the roan, since they hide their offspring in the dense riverine vegetation to avoid predators. If this cover is not available, the predation rate on the calves increases. The increased utilization of the areas around the artificial waterpoints results in increased nutrient deposits through urination and defecation; and the increase in the nutrient levels stimulates the growth of more Increaser grasses. This results in a vicious cycle, to the detriment of the roan antelope (Grant & de Buys, 2005).

A study by Whyte (Whyte, 2004 cited by Grant & de Buys, 2004) showed a concurrent increase in the buffalo and elephant population with the increase in artificial waterpoints; and this supports the hypothesis that an increase in waterpoints will lead to an increase in numbers of the water-dependent bulk grazers.

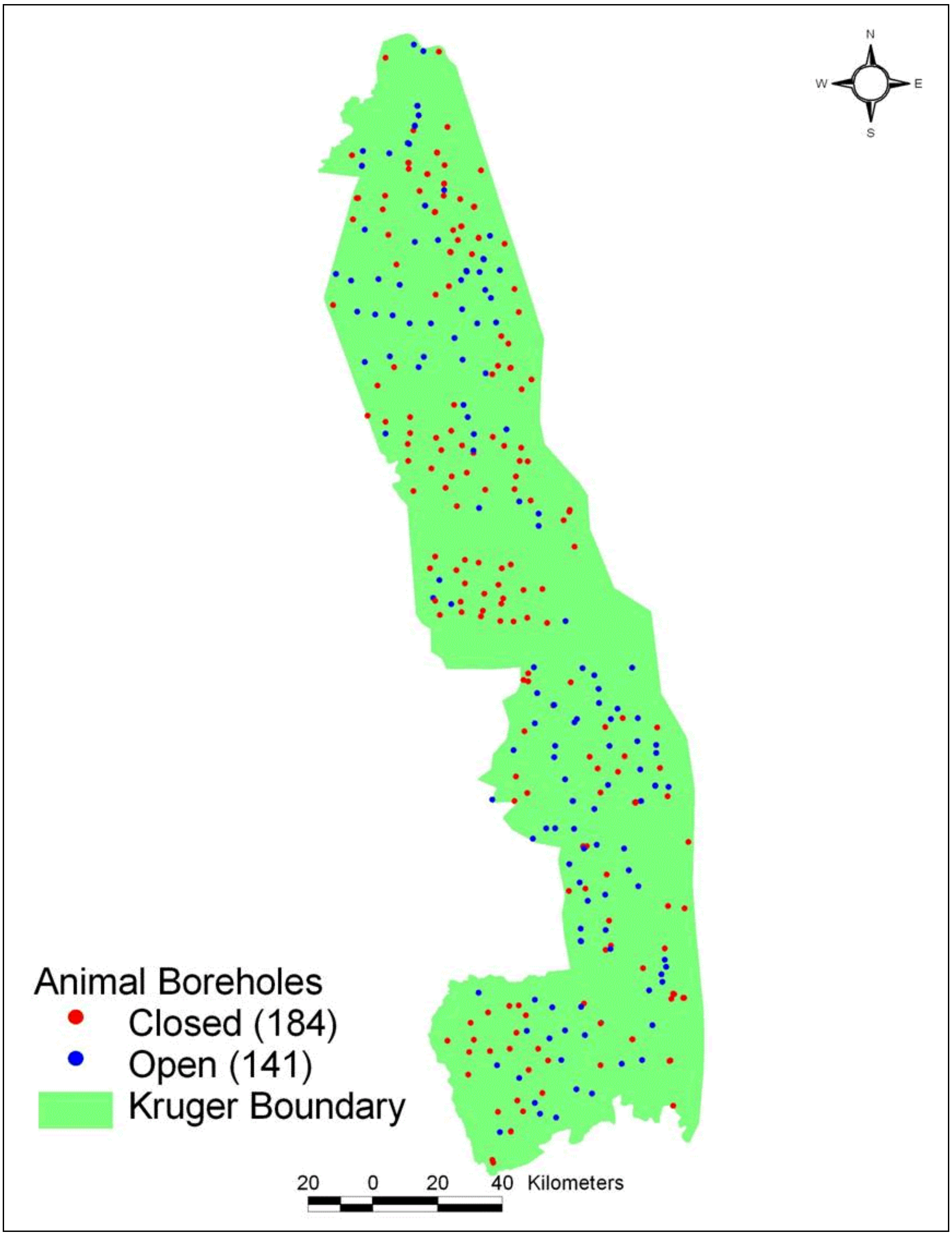

Figure 6.1 Map showing the location of closed boreholes (Travers, 2006)

Grant and de Buys (2004) showed that since 1986, the increase in number of bulk grazers on the northern plains coincided with a decline in the numbers of selective grazers. It was postulated that this was as a direct result of the change in grass composition. The findings in Chapter 5 showed that vegetation, available to selective grazers only within the N'washitsumbe enclosure, was Decreaser-dominated; whereas outside the enclosure, in areas accessible to the bulk grazers, the vegetation was Increaser 2 dominated. A more detailed description of the grass classification process is given in Chapter 5. The proportion of Increaser 2 species increases under conditions of intense and heavy grazing; whereas Decreaser species thrive under light grazing.

The aim of this research is to investigate whether the closure of many artificial waterpoints has resulted in a change in the habitat that is now more suitable to selective grazers. It is expected that the vegetation would have become more Increaser dominated during the period that the waterpoints were open since these areas would have been heavily utilized. It is expected that, over time, these areas will become more Decreaser dominated. It is acknowledged that, as most boreholes were closed in the period 1994 - 1998, the time period and data availability since the closure of the boreholes may not be sufficient to pick up significant changes in the vegetation composition. In Chapter 5 it was seen that significant differences in the vegetation composition exist between the N'washitsumbe enclosure and its surrounds – however this area has been fenced off for almost 40 years and in the case of the borehole closures, the vegetation has only had approximately ten years to recover from the heavy bulk grazing.

Figure 6.2 shows Boyela, an open borehole in the Far Northern Region of the KNP, north of the Shingwedzi rest camp. The photograph illustrates the fact that regular use of this borehole has resulted in the grass cover being very sparse and low. This waterpoint is not located near to a watercourse and it is regularly used by most animal species including buffalo and elephant (personal observation).

Figure 6.2 Boyela borehole – an open waterpoint showing signs of intense vegetation utilization

In contrast, Nshawu No 1 borehole shown in Figure 6.3 was closed in 1999. This picture shows very few signs of vegetation utilization; in fact the grass is tall and abundant. Despite the different status of these two boreholes, it is noted that Nshawu No 1 is very close to a small watercourse and hence the animals in the region would not have been as dependent on the borehole for a regular water supply. The Nshawu No 1 borehole is situated near to Mopani rest camp in the northern region of the KNP.

Since the presence or absence of a particular grass species will also be dependent on the rainfall, the annual rainfall figures for the areas in close proximity to the boreholes were obtained and will be included in the analysis as it needs to be determined whether changes in vegetation composition structure have been caused by the rainfall or the grazing intensity changes since the borehole closures. In other studies it has been found that changing rainfall did not affect tree growth significantly; however grass growth increased and decreased according to the rainfall (Kruger Park Times, 1–14 September, 2005). It is also noted that higher rainfall will cause more water to be available for a longer period in temporary pools and pans that will allow the bulk, water dependent grazers to remain in the areas that the closures were intended to make inaccessible due to lack of water. It was noted by Grant and de Buys (2004) that "an increase in rainfall from 2003 is reflected in the higher percentage Decreaser species in 2004" and hence it was decided that the rainfall figures for the areas around the boreholes needed to be factored into the analysis as changes in the vegetation composition could have been as a result of the rainfall rather than the closure of the borehole. This will be further examined later on.

This study focuses on three main hypotheses; namely whether the removal of the permanent artificial water sources has led to an increase in herbaceous species that decrease with intense herbivory (Decreasers). Secondly, whether the removal of the artificial waterpoints has led to an increase in the grass biomass and finally whether the changes that have occurred are dependent on the rainfall, locality and geology of the landscape.

6.3 STUDY AREA

This waterpoint investigation will not be isolated to the northern plains, but will look at waterpoints throughout the KNP. Data were obtained from Sandra McFaden and Dr. Nick Zambatis of Scientific Services in the KNP. The boreholes that have been closed since 1989 were identified and then veld condition assessment (VCA) data for the VCA-points closest to these boreholes were obtained. The VCA's were performed according to the Kruger Park VCA method and data were collected on the herbaceous and woody strata as developed by Nick Zambatis. There are 500 VCA sites throughout the KNP covering all of the different landscape types (Hofmeyr & Eckhardt, 2004).

Figure 6.3 Nshawu No1 – a waterpoint closed in 1999, showing very little vegetation utilization

6.4 DATA COLLECTION

The VCA data from 1989-2004 (inclusive) were obtained; and the relative percentages of Increaser and Decreaser grass composition was examined. The grass biomass was also examined to see if there have been changes in the biomass since the closure of the boreholes. The VCA's were performed in April-May of each year. The rainfall for each of these sites was also obtained as it was hypothesized that significantly high or significantly low rainfall might also have resulted in changes in the grass composition at the time of the veld condition assessment.

A high grass biomass can be indicative of high grass production or low consumption or both. The closure of the boreholes should have resulted in change in the consumption levels, rather than the grass production. Rainfall will play an important role in the grass production, this will be investigated by the correlation between rainfall and grass biomass.

Data were obtained for 87 boreholes that have been closed. It is noted that the VCA sites are varying distances from the closed boreholes, in some cases up to 6kms away from the site of the borehole. Unfortunately this was the only data available for this study but it is expected that differences will not be as noticeable at this distance from the borehole. For future closures of boreholes, it would be statistically interesting to conduct a VCA survey at the site of the borehole and then monitor the changes over a number of years to see whether this localized investigation would provide more significant results.

6.5 METHODS

To begin with, the VCA data for all the closed boreholes were analyzed on an individual basis. The data were in a time series format, since the VCA data were captured on an annual basis as were the rainfall figures. The data considered for the analysis were the status of the borehole, i.e. whether it was open or closed; the relative percentages of Decreaser, Increaser 1, Increaser 2, Forbs and Other plant classification as well as the rainfall for the rainfall station closest to the borehole.

Firstly means were calculated for the relative percentages of main grass classes within the KNP namely the Increaser 2 and Decreaser. Two means were calculated per borehole, one for the years before the closure and one for the years after the closure. An inspection of the differences between the means would give an indication of which boreholes had experienced changes in the relative grass composition in the years after the closure of the boreholes.

Initially time series methods were considered as a way to examine the data for the 87 boreholes. However, due to the limited number of years of data, none of the time series techniques resulted in stationary data for the individual boreholes.

By calculating a difference percentage score for both the Decreaser percentage and the grass biomass, it was possible to use Analysis of Variance methods to determine the effect of the region, the rainfall and the geological base on the changes that have occurred since the closure of the boreholes. The difference was done as the raw data for these variables are highly dependent.

The geological base is either granitic or basaltic. The granitic layer is very sandy and water will absorb and dissipate very quickly in this area. However, on the basaltic layer, the water will be held on the surface for a lot longer and this will mean that the availability of water will be a far more influential factor in the distribution of water dependent mammals. The granite derived soils are nutrient poor, where as the basalt derived soils are nutrient rich.

The KNP is divided in 22 ranger sections (see Figure 6.4) that are then grouped into the following regions: south, central, north and far north. There are either five or six ranger sections per region (see Table 6.1).

Table 6.1 Ranger sections in each of the four regions and the number of analyzed boreholes within each region

Far North	**No.**	**North**	**No.**	**Central**	**No.**	**South**	**No.**
Pafuri	2	Mooiplaas	7	Houtboschrand	4	Lower Sabie	0
Punda Maria	1	Mahlangeni	8	Kingfisherspruit	1	Skukuza	4
Vlakteplaas	7	Letaba	3	Satara	2	Pretoriuskop	6
Shangoni	3	Phalaborwa	9	Nwanetsi	2	Stolsnek	7
Woodlands	2	Olifants	5	Tshokwane	5	Malelane	2
Shingwedzi	8					Crocodile Bridge	1

These regions are not an ecological division of the KNP however, the climate, vegetation, animal abundance, temperature, rainfall, geology and burning patterns will vary across each of the sections.

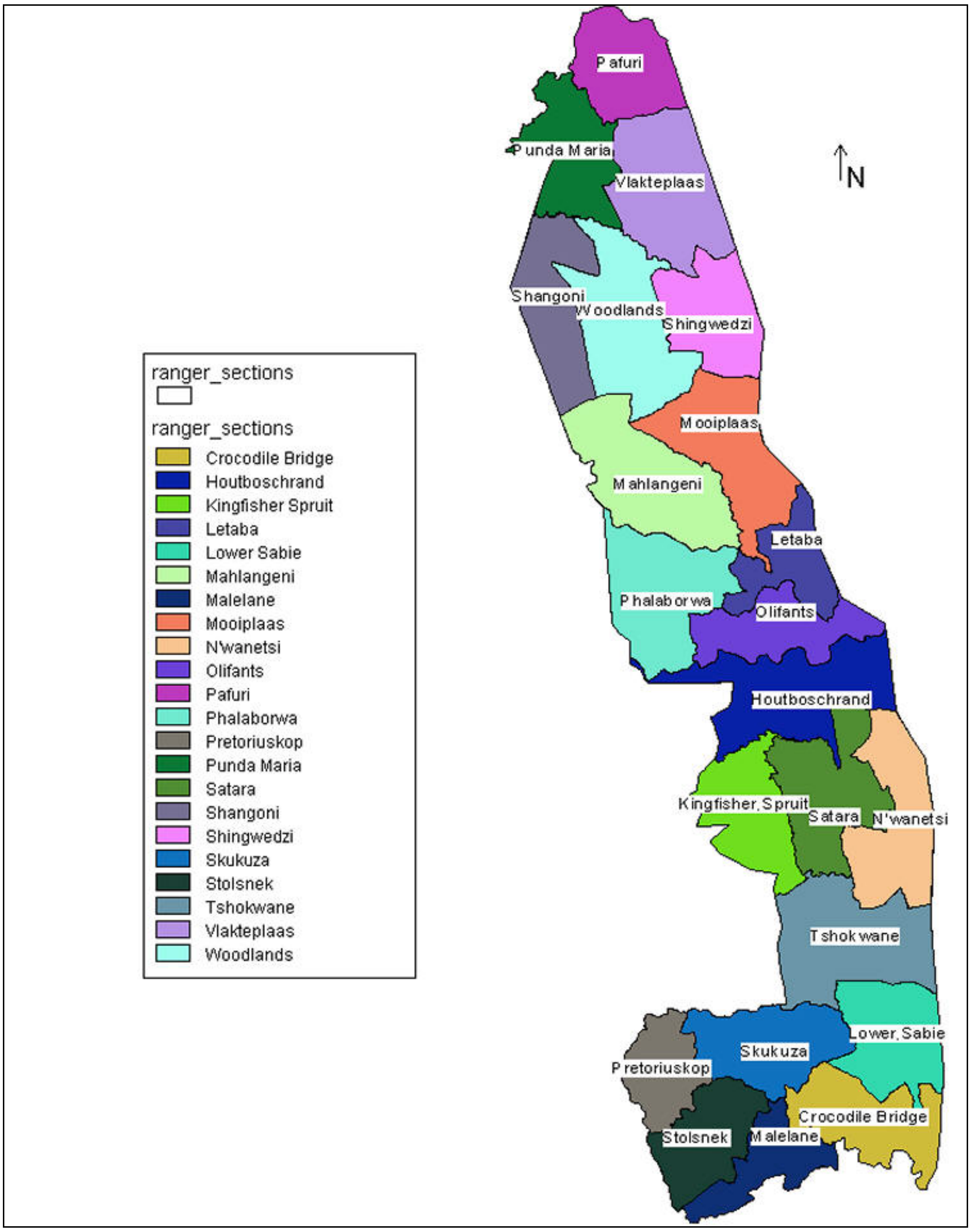

Figure 6.4 Map showing the 22 ranger sections (http://thekruger.com/category/lists-and-maps/)

6.6 RESULTS

The difference in mean percentage of Decreaser proportion, before and after the closure of the boreholes is shown in Figure 6.5. It can be seen that at 24 of the waterpoints, the relative percentage of Decreaser species has actually decreased since the closure. However, the majority of the areas around the closed waterpoints have experienced an increase in the relative percentage of this category. The increase in Decreaser proportion is the desired outcome of the borehole closures. The results for the Increaser 2 species were nearly a mirror image of Figure 6.5, since the two categories are heavily dependent on one another. Hence this graph is not shown.

The rainfall for the period from 1989 until the closure of the boreholes and then the period subsequent to the closures was analyzed to investigate the possibility that improved rainfall might have resulted in the changes in vegetation composition seen in Figure 6.5. The results show that the mean rainfall before closure was 452.9±167.2mm and the mean after the closure was 604.9±285.5mm. This is a significant difference ($t_{1309} = -12.1$, $p<0.01$) and this needs to be considered when making conclusions about the impact that the closures have had on the vegetation composition around the boreholes.

It is noted that since the percentages of Increaser 1, 2 and Decreaser species will be relative to one another, they are clearly dependent on one another and a large degree of collinearity is expected between the variables. Due to this, only the Decreaser proportions will be analyzed in the Analysis of Variance procedures. A correlation matrix was computed in order to determine whether there was a linear relationship between the grass categories and the annual rainfall. Correlations between the Decreaser, Increaser 1 and Increaser 2 proportions as well as the rainfall, lagged rainfall by one year, and the distance from the borehole to the VCA site were computed. As expected, the strongest correlation is between the Increaser 2 and Decreaser proportions ($r = -0.83$, $p<0.001$). The annual rainfall has a very weak positive correlation with the Decreaser abundance ($r = 0.09$, $p=0.002$) and a very weak negative correlation with the Increaser 2 abundance ($r = -0.12$, $p<0.001$). The positive nature of the correlation with the Decreaser abundance is expected and this is in line with the observations by Grant and de Buys (2005) that an increase in rainfall led to a subsequent increase in the relative abundance of Decreaser grasses. This would in turn result in a decrease in the relative abundance of the Increaser 2 species.

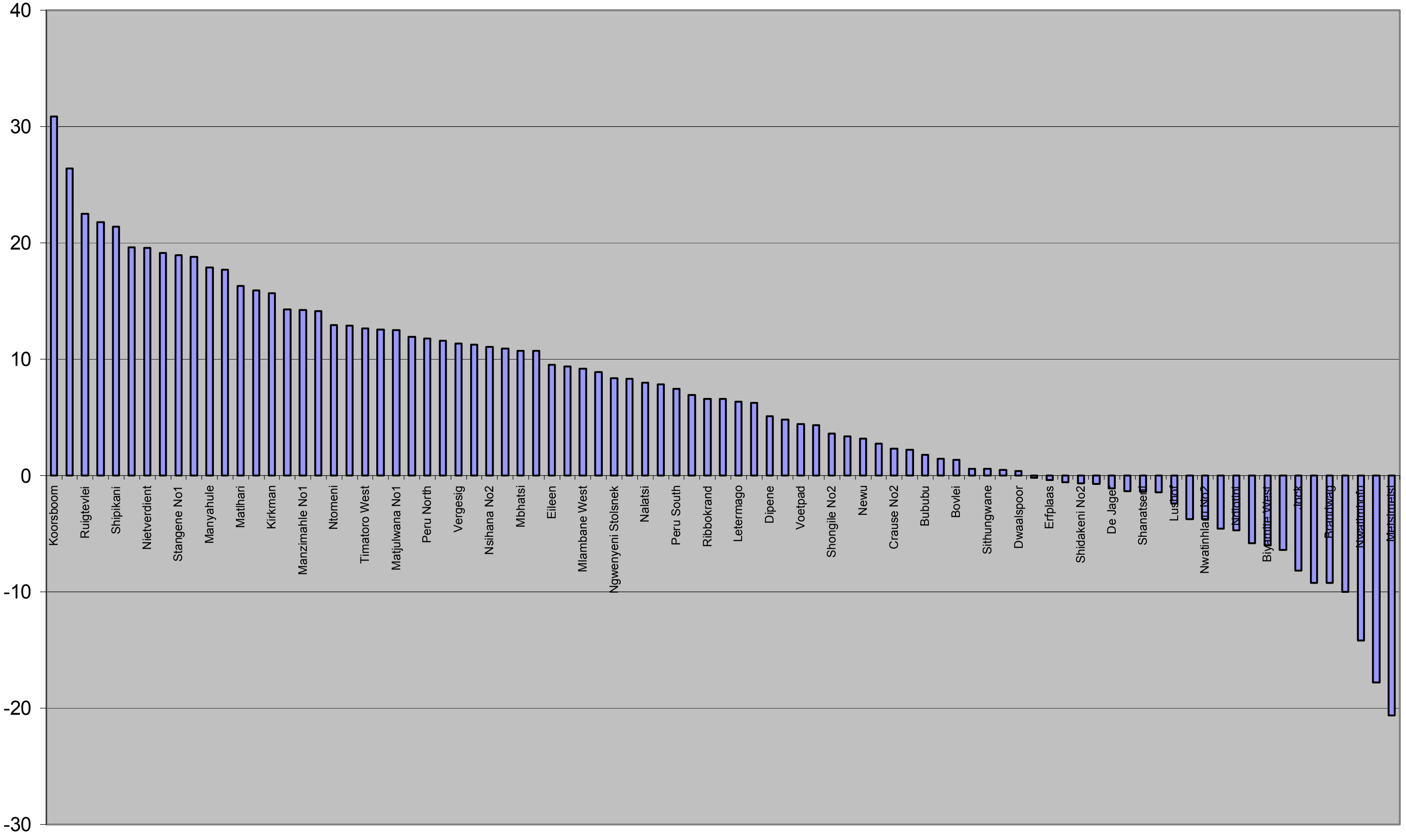

Figure 6.5 Change in mean Decreaser proportion since by borehole since closure

The rainfall is significantly correlated with the grass biomass ($r = 0.47$, $p<0.001$); the grass biomass, measured in kg.ha^{-1}, is estimated by measuring the compressed height of the grass and then uses regression analysis to convert the settling height of the disk to dry mass per unit area. The correlations were also calculated for a lagged rainfall, calculated by lagging the rainfall by one time period. The correlations slightly improved to $r = 0.17$, $p<0.001$ with Decreaser and $r = -0.22$, $p<0.001$ for the Increaser 2 proportions.

Correlations were also calculated to determine whether there is a relationship between the grass composition proportions and the distance from the borehole to the VCA site. None of these correlations were higher than 0.03 which indicates that the distance from the VCA site to the borehole is not related to the grass composition or grass biomass. This is a slightly unexpected result, as one would expect the intensity of the effects of grazing to be less evident further away from the borehole.

Since the grass biomass is positively correlated with rainfall, and the rainfall was significantly higher in the years after the closure of the boreholes, any improvement in the grass biomass cannot be attributed only to the closure of the boreholes.

6.6.1 *Analysis of Variance - Decreaser*

Figure 6.6 shows the raw Decreaser percentage for all the boreholes from 1989 until 2004. The boreholes were closed during different years between 1994 and 1999, so we cannot identify a specific closure year; however the graph shows a declining trend for the Decreaser proportion before the closures occurred. The reason for the decreasing Decreaser proportion from 1989 until 1993 is unknown. It is possible that it could have been caused by the added grazing pressure due to the artificial waterpoints which allowed an increasing number of bulk grazers, such as zebra, into these areas; or it may have been a result of changes in other factors that have not been identified in this project.

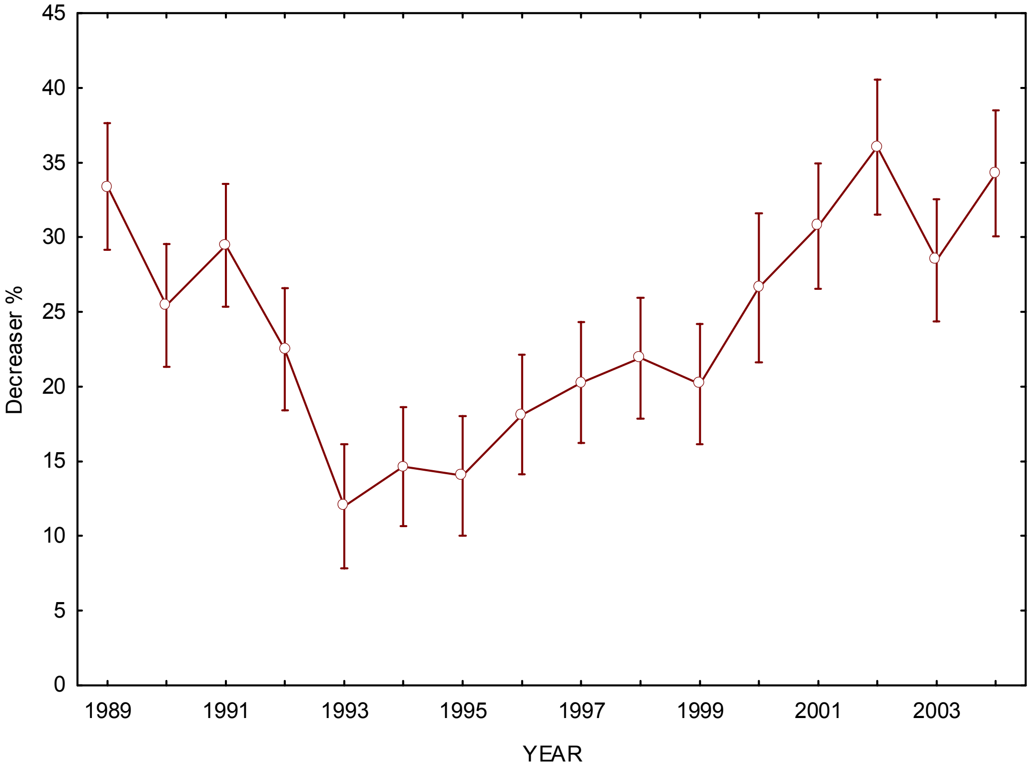

Figure 6.6 Decreaser percentage for the period 1998 to 2004 (mean ± SE)

It is noted that throughout these analyses, the "general trend" for the grass composition for the years before and after the closures is not taken into account. This means that natural fluctuations in the grass composition as a result of temperature, burning and utilization have not been investigated; and that any changes after the closure of the boreholes will be considered as a result of the closure. In order to limit the inaccuracy this creates, the rainfall pattern before and after the closure will be investigated.

Following on from the means analysis done above, a new measure of the change in composition was computed. A difference percentage score for the Decreaser percentage was calculated by subtracting the Decreaser percentage for the first year of VCA data from the Decreaser percentage for the year before the closure. This difference percentage score is a measure of the change in Decreaser percentage for the years before the closure occurred. A second difference percentage score was calculated by subtracting the Decreaser percentage for the year of closure from the Decreaser percentage for the final year of data in the study. This is a measure of the change in Decreaser percentage subsequent to the closures. These two difference percentage scores (before and after closure) are considered as repeated measures of the Decreaser change for each borehole. The need to use this difference variable is due to the annual scores being highly dependent on the previous year's percentage. The differencing also removes the high variation of Decreaser percentage among all of the boreholes due to rainfall and other factors.

A repeated measures Analysis of Variance procedure was used to determine whether the effect of the closure, the regional locality or the geological base (basalt or granite) were significant factors in the Decreaser percentage change. The dependent variables were the Decreaser percentage change before and the Decreaser percentage change after the closure of the borehole. These repeated measures were used to define the borehole status. The categorical, independent factors were the region and the geological base.

Table 6.2 Repeated measures ANOVA results for the Decreaser proportion change

	SS	df	MS	F	p
Intercept	40.29	1	40.29	0.18	0.68
Region	215.54	2	107.77	0.47	0.63
Geological base	2.78	1	2.78	0.01	0.91
Region*Geological base	1270.40	2	635.20	2.78	0.07
Error	13698.47	60	228.31		
Borehole Status	27451.34	1	27451.34	82.73	<0.001
Status*Region	194.86	2	97.43	0.29	0.75
Status*Geological base	109.13	1	109.13	0.33	0.57
Status*Region*Geological base	1658.79	2	829.40	2.50	0.09
Error	19910.06	60	331.83		

The results are shown in Table 6.2. The regional effect, the geological base and the interaction between both of these factors are insignificant. The only significant effect is the borehole status ($p<0.001$) which shows that vegetation changes are occurring as a result of the closure of the boreholes. Figure 6.7 is a graphical representation of these data. It is noted that none of the closed boreholes used in this study were located on a basalt base in the Southern region.

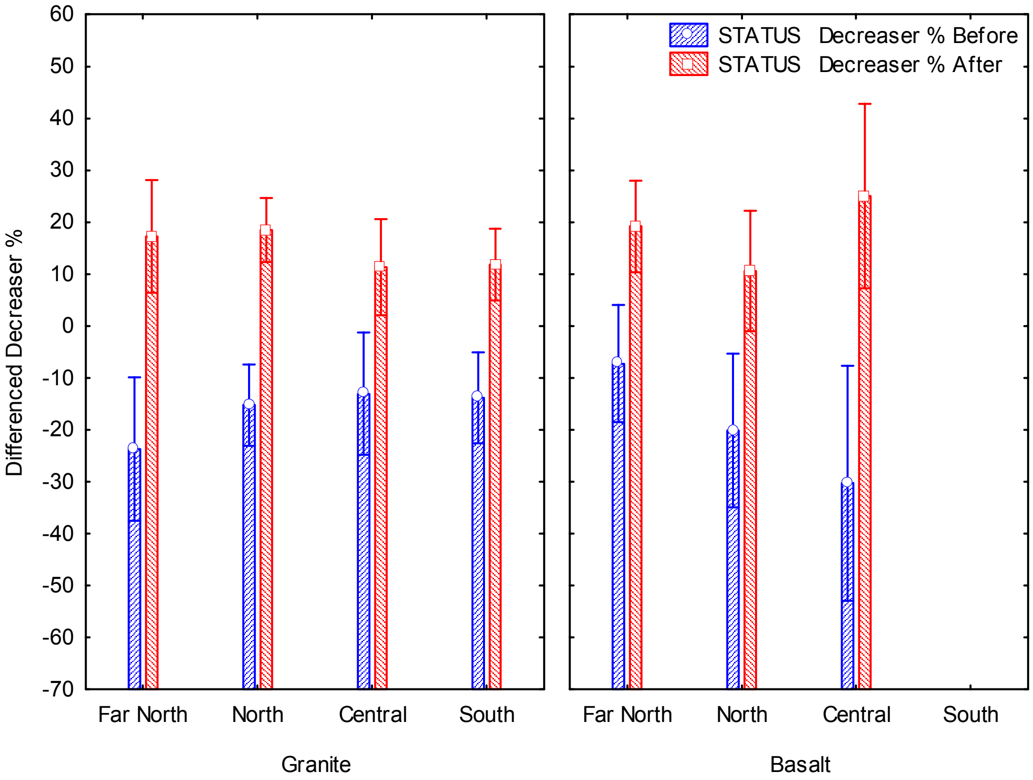

Figure 6.7 Graph showing Decreaser difference before and after closure by region and geological base (mean ± SE)

It is clear that the difference percentage score is significantly higher subsequent to the closure of the boreholes. This means that the Decreaser percentage in 2004 was significantly higher than when the boreholes were closed. Thus, it can be concluded that it is possible that the closure of the boreholes has contributed, possibly along with other factors, to a higher percentage of Decreaser species in the areas around the boreholes.

The rainfall effect has not been included as this would have been alleviated by the calculation of the difference variable. To confirm this, the analyses were re-run, using a weighted difference score. This was calculated by dividing the difference score before the closure by the mean rainfall for the years prior to closure. Similarly, the difference score for after the closure was divided by the mean rainfall for the years subsequent to the closure. This was done to remove the effect of rainfall, if it did exist, despite the use of the difference score. The analyses were re-run and the trends and patterns were found to be exactly the same as with the unweighted variable. This confirms that the differencing removes any effect due to annual rainfall.

6.6.2 *Analysis of Variance – Grass Biomass*

An identical set of analyses were run, replacing the Decreaser proportion with the grass biomass. Figure 6.8 shows the mean grass biomass (kg.ha^{-1}) by the years.

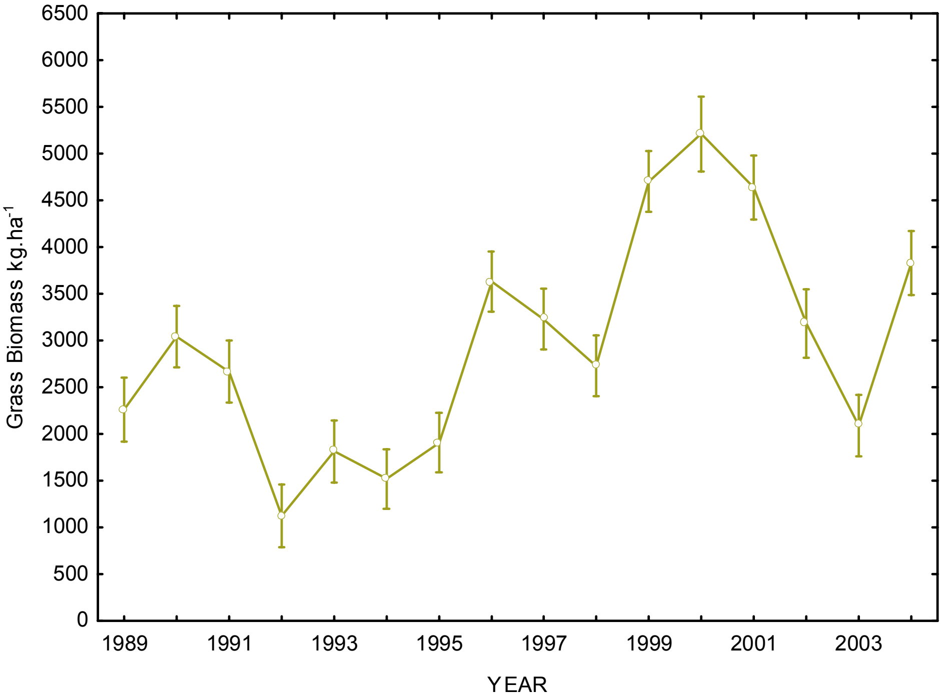

Figure 6.8 Grass Biomass (kg.ha^{-1}) for the period 1998 to 2004 (mean ± SE)

Once again, there is a slight decreasing trend up to 1993 and then different fluctuations occur between 1993 and 2004, from those seen for the Decreaser proportion. The differenced biomass scores were calculated by subtracting the biomass for the year prior to closure from the biomass for the first year of data to create the ‘Biomass difference before’ variable; and the ‘Biomass difference after’ variable was the year of closure’s biomass score subtracted from the final year’s biomass score. The repeated measures ANOVA was run with the dependent variables being the difference scores, and the status effect being the within effect of these two repeated measures. Once again, the independent variables were the region and the geological base. The results are shown in Table 6.3.

Table 6.3 Repeated measures ANOVA results for the Grass Biomass change

	SS	df	MS	F	p
Intercept	142901050.35	1	142901050.35	66.09	<0.001
Region	11667749.11	2	5833874.56	2.70	0.08
Geological base	6686850.96	1	6686850.96	3.09	0.08
Region*Geological base	30890499.83	2	15445249.92	7.14	<0.001
Error	129732268.56	60	2162204.48		
Borehole Status	1630826.84	1	1630826.84	0.43	0.51
Status*Region	308656.21	2	154328.11	0.04	0.96
Status*Geological base	5276208.07	1	5276208.07	1.40	0.24
Status*Region*Geological base	12276554.41	2	6138277.20	1.63	0.20
Error	226312649.97	60	3771877.50		

The results show that the region and the geological effects are insignificant, however the interaction between these two effects is significant. More importantly, the borehole status effect is not significant which indicates that the biomass has not changed as a result of the closures. The graphical representation of these data is shown in Figure 6.9.

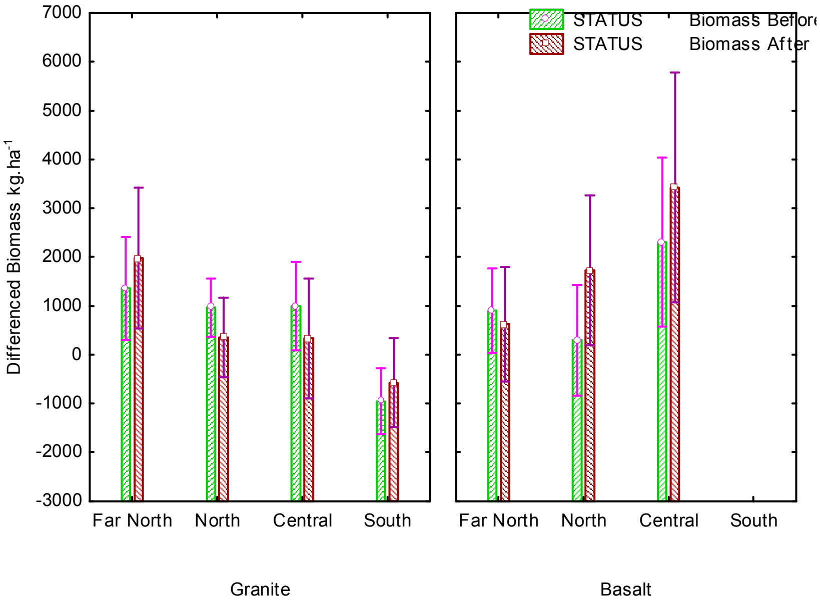

Figure 6.9 Graph showing Grass Biomass (kg.ha^{-1}) difference before and after closure by region and geological base (mean ± SE)

The biomass is not consistently higher or lower for each of the regions, which confirms the insignificance of the borehole status effect. This means that the closures of the boreholes have not resulted in a changed grass biomass.

Once again, the rainfall effect has not been included due to the use of the differenced variable.

6.7 SUMMARY

The results of these analyses have shown that the Decreaser proportion has increased significantly since the closure of the boreholes. This is the desired result of the closures as the vegetation is now more favourable for the selective grazers such as the roan. It is hoped that these changes will result in an increase in the roan population and other rare antelope. The changes are universal, and are not restricted to certain regions or landscape types within the KNP. The grass biomass however has not changed significantly since the closures of the boreholes.

This analysis of the borehole data is very simple, with no sophisticated statistical methods being used. This is due to the lack of data; and as time goes by and more time series observations are obtained, it is possible that more sophisticated modelling techniques could be used in the future, to test similar hypotheses.

Thus to answer the three hypotheses discussed in Section 6.2, the removal of the permanent artificial water sources has led to an increase in herbaceous Decreaser species, which indicates a decline in intense herbivory. Secondly, the closure of the boreholes has not resulted in a significant increase in the grass biomass; and finally the changes that have occurred have not been caused by the rainfall, regional locality and the geology of the landscape.

CHAPTER 7: ELEPHANT DAMAGE IN THE RIPARIAN ZONE

7.1 INTRODUCTION

“Elephants are major agents of change and are often indicated as those large herbivores possessing the ability of changing entire ecosystems in terms of the vegetation structure and composition, thereby affecting a whole series of other ecosystem components as well. The exclusive role of elephants as agents of change could thus far not be completely isolated from the multitude of factors involved in ecosystem dynamics” (Hofmeyr & Eckhardt, 2004).

A concern for the management of the Kruger National Park (KNP) is the impact which the African Elephant (*Loxodonta africana*), (see Figure 7.1) is having on the vegetation within the KNP. The population of elephant is increasing at an estimated rate of 6.6% per annum (Whyte, 2004) which means that the number of elephants is increasing rapidly and the concern is that the amount of damage being done to the vegetation, that these animals are feeding upon, is probably also increasing. Elephant once occurred throughout Southern Africa, but now their distribution is restricted to the extreme northern and eastern areas. There are also isolated populations in the Addo Elephant National Park and surrounding Eastern Cape game reserves (Stuart & Stuart, 1992).

Figure 7.1 *Loxodonta africana* – African Elephant near Shingwedzi rest camp

Research and extensive fieldwork has been done into the effect that elephants are having on the habitat within the KNP. The Elephant Management Policy is constantly being debated and will always be changing as new data are analyzed and the effects of the rising number of animals within the population is being seen. Numerous studies are being done throughout the whole of southern Africa into various aspects of the elephant population and the sustainability of the current numbers of this “super-herbivore”. Elephants are mixed bulk feeders and they are able to utilize both the woody and herbaceous components. They compensate for ingesting low-quality food characterized by low-protein and high fibre contents, by consuming vast amounts of vegetation (Hofmeyr & Eckhardt, 2004). This coupled with destructive feeding behaviour, has large scale implications for the woody vegetation; and extended intensive utilization of areas of vegetation can alter the structure and sometimes even the composition of the vegetation (Hofmeyr & Eckhardt, 2004). This damage is believed to have a significant effect on the habitat in which the elephants live, due to the large amount of vegetation that they can consume (Codron *et al.*, 2004).

Up until 1993 the KNP was a closed ecosystem due to it being fully fenced on all sides and since there are no natural predators of elephant, SANParks had a policy of culling through which the elephant population was held at approximately 7000 animals. This was considered to be the carrying capacity for the KNP that would not result in the elephant population having significant impacts on the biodiversity within the KNP. However, in 1995 a decision was taken to stop the culling process and let the elephant population expand naturally and monitor the effects that it was having on the ecosystems within the KNP (Pienaar, 2004). Numerous options for controlling the elephant population have been discussed, including translocation, contraception, culling and others. As yet, SANParks has not adopted any of these in a determined effort to stop the increase of the population. In 1993 part of the KNP's western boundary fence was dropped which has meant that animals are free to move into the Sabie-Sand Reserve and the Klaserie/Timbavati complex of reserves, known as the Associated Private Nature Reserves (Whyte, 2004). Some of the fence-line has also been removed between South Africa and Mozambique, creating the new Greater Limpopo Transfrontier Park and it is hoped that the new areas that are now accessible to the game will help to alleviate the problem of growth in the elephant population. Research is needed to determine how many elephants have moved into these areas and whether this is reducing the pressure on the KNP biodiversity sufficiently.

7.2 BACKROUND

The riparian zone (vegetation along streams and rivers) is an ecologically very important part of the ecosystem. It will usually contain some of the biggest trees in the region and is used by numerous other species of animal, bird, plant, reptile and insect. It also holds a lot of aesthetic appeal for tourists; and many of the tourist roads throughout the Park follow the rivers and wind their way through the riparian vegetation. The riparian zone contains some of the most beautiful and iconic tree species of the Lowveld bush; for example, an Umbrella Thorn (*Acacia tortilis*) (see Figure 7.3); a species which provides important shade and forage for many herbivores, particularly in the dry season.

Elephants can be destructive feeders and can damage trees very seriously that they have been feeding on and sometimes result in the death of the tree. It has been noticed in the last few years, since the termination of elephant culling, that there has been an apparent increase in the amount of visible elephant damage to the vegetation, especially in certain landscapes (personal observation and see Figure 7.2). This is a concern for the long term sustainability of the riparian zones. Social behaviour also has an implication for the woody vegetation since aggressive behaviour often results in uprooting or breakage of trees (Hofmeyr & Eckhardt, 2004). Some researchers believe that tree damage, especially by male elephants, is a way of building strength and improving tusk dexterity (Wray, 2005). For this reason Norman Owen-Smith suggests that disturbance culling could be used to restrict the use of sensitive areas by the mature male segment in addition to the removal of all artificial water sources in these areas (Owen-Smith, 2004a). Jaco Badenhorst, a wilderness trails ranger in the KNP noted that he has seen increasing elephant damage in last 12 years since the population started increasing and he believes that the KNP is slowly changing from woodland into scrubland (Badenhorst, 2006).

Figure 7.2 An example of elephant damage near Letaba rest camp

Although, it is not so much the severity of the damage inflicted by elephants on the tree species that is the concern, it is the extent of the damage (Owen-Smith, 2004b). Elephants are water dependent and hence the severest damage is expected to be in the areas near to water. This means that woodland elimination may occur locally but it is unlikely to occur in areas remote from water (Owen-Smith, 2004b). This however, means that the riparian zones which are adjacent to water will be under the most immediate threat of elephant damage. Elephant distribution over the past 20 years shows a strong association with rivers (Grant *et. al.*, 2005) However, many of the tree species in the riparian zone have deep roots in order to access underground water (Owen-Smith, 2004b); and this will make them less susceptible to being pushed over and killed by the damage inflicted by elephants.

It is important that younger trees species can grow to one day replace these larger trees, in order to maintain the heterogeneity of the riparian zone. However, if the damage inflicted upon the younger trees is restricting the development of them, the large trees in the riparian zone may begin to disappear.

In order to understand the effects that the elephant population is currently having on the riparian zone, this project will be investigating whether elephants damage certain tree types more seriously than others; and whether their tree utilization is focused on certain height classes of tree. The results of the elephant data analysis will focus on the effect of elephant utilization on a species- and height-specific level. This will contribute to the understanding of whether the current concern about the effects on vegetation of increasing elephant numbers is indeed justified, and if so, which species or size classes are at highest risk from this damage. If species- and height-specific elephant impacts are demonstrated, then monitoring programs and management actions can be fine-tuned to prioritize these species and height classes for protection.

Gaylard (2004) states that “species- and size specific susceptibility to elephant impacts determines whether the individual will survive the elephant impact or simply be knocked back into a smaller size class by the removal of biomass”. It is also noted that elephants are not the only species to alter the vegetation status; fire, other herbivores and climatic conditions also play a role in changes of savanna woodland to scrubveld. The impacts of all of these factors will become evident over different time periods.

7.3 STUDY AREA

The data for this project was captured along granitic areas of the Shingwedzi, Phugwane and Mphongolo rivers in the northern areas of the Kruger National Park by Angela Gaylard as part of her PhD study under the River and Savanna Boundaries program. The Phugwane River is a tributary of the Mphongolo River that has its source to the west of the KNP. It was named after a chief who used to live along the banks of the river (Kloppers & Bornman, 2005). The Mphongolo River is one of the main tributaries of the Shingwedzi River and it was also named after a chief who lived near the river in times gone by. It literally means ‘a barrel or cask’ (Kloppers & Bornman, 2005). The Shingwedzi River is a tributary of the Rio dos Elefantes in Mozambique (Kloppers & Bornman, 2005). These rivers are seasonal and the research was conducted in the granite areas of these rivers.

In the course of her research, Gaylard found that the elephant impact intensities in riparian zones were spatially variable in relation to the locality’s surface water context (Gaylard, 2004). She tested whether/how this varying elephant impact intensity in relation to surface water distribution had produced distinct differences in the compositional and structural diversity of riparian trees, particularly considering the abundant supply of artificial water sources and the current density of elephants (Gaylard, 2004). The results also suggest that current elephant densities have not yet reached a level where they are significantly altering either the compositional or structural diversity of riparian woody vegetation (Gaylard, 2004). In her research, Gaylard performed a Principal Components Analysis to determine whether there were particular suites of species and/or size classes of woody species that characterized areas with different surface water distribution attributes (Gaylard, 2004).

In a different study, it was found that elephants in the northern regions of the KNP tend to eat a higher percentage of grass than those in the southern regions – this is unexpected since the north is dominated by *Colophospermum mopane*, (Codron *et al.*, 2004). However, this does not mean *Colophospermum mopane* is not an important part of an elephant's diet. In fact, in the dry season it is one of the few trees that still carries leaves and hence it is a very important food source (Lagendijk *et al.*, 2004).

The distance to the nearest water was also included in the dataset, however this was not analyzed at the request of SANParks since it has already been covered in other research done using this data.

Figure 7.3 *Acacia tortilis* – Umbrella Thorn on the banks of the Letaba River

7.4 DATA COLLECTION

The data captured includes a scoring system which quantifies the damage that has been done to a tree. This damage is classified as either old or new damage. New damage is that done within the last season, while old damage is an accumulation of damage done over a long term period. This old damage actually gives a long term view of what the elephants have done to the riparian woody zone, and this could be used to develop a forecast of the long term sustainability of this zone. This scoring system was initially developed as a probability of the tree dying due to the damage inflicted to the tree.

However an overall scoring system was then developed which is no longer a probability, but rather a lumped sum, and this is used as a scoring system to quantify the damage done.

A score of zero would indicate that no damage was visible on the tree and higher damage scores indicate more severe damage had been inflicted upon the tree, The more severe the damage, the greater the chance that the tree will die as a result of the damage. The total damage scores range between 0 and 3.22.

For three successive years (1997 – 1999), numerous trees within the riparian zone where analyzed and their damage quantified. The data that will be used for the purposes of this project includes the tree species, the wood type (soft, medium, hard or very hard), the height of the tree when sampled and the new and old damage scores as well as a total damage score which is the sum of the two damage scores.

The dataset has been split into the three separate years of data as it is possible that some trees that were given a new damage score in one year might have been re-sampled in the following year and this would now have become part of the old damage score. In order to avoid the complication that this repeated measure might have on the data, the data will be analyzed for the three years separately. It is expected that the results will be similar for the three years due to the large sample sizes; however if the conclusions drawn are different for the three years, then reasons for the differences will be investigated.

The height of the tree was recorded and then categorized into one of the following levels: less than 20cm, 20cm – 50cm, 51cm – 100cm, 101cm – 200cm, 201cm – 400cm, 400cm – 800cm and greater than 800cm. During the course of this study, 26 different tree species (see Table 7.1) were sampled with some being more common than others. Details of some of the more common species are given below.

Table 7.1 List of species sampled during the research

Species	**Common Name**
Salvadora australis	Narrow-leaved Mustard-tree
Terminalia sericea	Silver Cluster-leaf
Sclerocarya birrea	Marula
Shakama	Shakama-plum
Lannea schweinfurthii	False Marula
Ficus sycamorus	Sycamore Fig
Combretum mossambicense	Knobbly Climbing Bushwillow
Nuxia oppositifolia	Water Nuxia
Xanthocercis zambesiaca	Nyala-tree
Peltophorum africanum	African-wattle
Combretum apiculatum	Red Bushwillow
Ziziphus mucronata	Buffalo-thorn
Combretum imberbe	Leadwood
Acacia nigrescens	Knob Thorn
Dichrostachys cinerea	Small-leaved Sickle-bush
Flueggea virosa	White-berry Bush
Combretum hereroense	Russet Bushwillow
Euclea divinorum	Magic Guarri
Acacia robusta	River Thorn
Spirostachys africana	Tamboti
Acacia tortilis	Umbrella Thorn
Diospyros mespiliformis	Jackal-berry
Lonchocarpus capassa	Apple-leaf
Colophospermum mopane	Mopane
Maytenus heterophylla	Common Spikethorn
Croton megalobotrys	Feverberry

The most commonly sampled tree was *Croton megalobotrys* or the Feverberry which is a medium sized tree that grows to heights of 15m. The wood is light and soft and the bark and seeds can be used as a purgative or fish poison (Schmidt *et al.*, 2002). This tree is heavily utilized by elephant and other game species (Schmidt *et al.*, 2002). The seeds have the reputation of being effective in the prevention and cure of malaria. Another common tree in the northern regions of the KNP is the *Colophospermum mopane* or the Mopane (see Figure 7.4). This species can range in height from 5-12m and on some occasions up to 22m. The Mopane worm is the larvae of the moth *Gonimbrasia belina* which feed exclusively on the leaves of this tree; and these worms are an important source of protein to many local people (Schmidt *et al.*, 2002). Mopane wood is very hard; and the leaves and twigs are consumed by many game species (Schmidt *et al.*, 2002).

Figure 7.4 *Colophospermum mopane* – Mopane on Mphongolo River

7.5 METHOD

Firstly mean damage scores were calculated for each of the height class of the different wood types and then an Analysis of Variance was run for each of the three year's data separately to investigate whether there is a significant difference in the damage done to the different height class of tree, wood type or species. The analysis was run using the damage score as the dependent variable and then either the height class and the wood type or the height class and the species as the independent categorical variables. It is noted that some of the wood types only have one or two species that have sufficient data in order to run the analysis. In the case of only one species, a one-way ANOVA was run using the height class as the independent categorical variable.

It is noted that there are many inconsistencies and omissions in the dataset that were beyond our control for this project. For example, the sampling has meant that there are no values for some of the height classes and not all the species sampled in one year have a similar sample size in the other years. This means that only basic analyses can be run on the data and that conclusions cannot be drawn for some of the species over the different years.

7.6 RESULTS

The mean damage scores for the height classes of each wood type were examined in Table 7.2. The results show that there appears to be differences between the mean scores for the different wood types and height classes. The significance of this impression was then analyzed using an Analysis of Variance (ANOVA). The ANOVA's which will be run for the individual years so that there is no possibility of repeated measures as it is feasible that a tree that was sampled in 1997 could then have been re-sampled again in 1998 or 1999. This would mean that the damage score would be made up of an old damage score for the second measure; and this damage could have been recorded as a new damage score a year or two previously.

Table 7.2 Mean and standard deviation for damage scores for 1997, 1998 and 1999 by wood type (n = sample size)

		1997			1998			1999		
Wood type	**Height Class (cm)**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
Soft	<20			0	0.55	0.55	4			2
	20-50	0.14	0.35	8	0.39	0.50	31			4
	50-100	0.54	0.75	46	0.46	0.55	58	0.16	0.26	12
	100-200	0.28	0.33	290	0.36	0.45	216	0.37	0.44	123
	200-400	0.33	0.24	230	0.43	0.56	187	0.62	0.44	105
	400-800	0.51	0.33	66	0.75	0.59	29	0.81	0.41	18
	>800	0.27	0.35	11	0.47	0.43	15	0.25	0.13	2
Medium	<20	0.96	.	1	0.57	1.02	12			4
	20-50	0.17	0.44	32	0.31	0.58	34	0.03	0.09	11
	50-100	0.30	0.51	42	0.26	0.47	57	0.31	0.53	26
	100-200	0.20	0.30	80	0.29	0.40	185	0.45	0.43	101
	200-400	0.30	0.32	43	0.40	0.46	46	0.63	0.45	20
	400-800	0.15	0.23	62	0.31	0.44	27	0.27	0.31	28
	>800	0.07	0.09	35	0.42	0.53	79	0.17	0.28	24
Hard	<20	0.00	0.02	54	0.22	0.35	50	0.30	0.66	10
	20-50	0.28	0.56	69	0.35	0.45	122	0.16	0.36	49
	50-100	0.36	0.55	92	0.31	0.42	215	0.33	0.45	54
	100-200	0.41	0.43	174	0.48	0.49	494	0.47	0.49	252
	200-400	0.53	0.52	105	0.46	0.55	143	0.57	0.40	92
	400-800	0.34	0.34	110	0.40	0.50	56	0.45	0.39	31
	>800	0.20	0.19	74	0.44	0.57	67	0.18	0.17	20
Very hard	<20	0.00	0.00	57	0.22	0.32	42	0.30	0.80	14
	20-50	0.12	0.30	80	0.17	0.39	48	0.40	0.53	13
	50-100	0.18	0.34	89	0.53	0.61	39	0.46	0.65	5
	100-200	0.24	0.36	434	0.42	0.61	103	0.72	0.61	52
	200-400	0.38	0.40	228	0.38	0.41	61	0.52	0.58	40
	400-800	0.23	0.28	193	0.41	0.52	54	0.20	0.24	56
	>800	0.14	0.16	91	0.33	0.44	75	0.16	0.19	18

7.6.1 *Results for 1997*

Firstly an ANOVA was run to investigate whether the wood type and the height class were significant factors in terms of the damage that was inflicted upon the tree.

Table 7.3 ANOVA results for Height class vs. Wood type - 1997

Effect	**SS**	**df**	**MS**	**F**	***p***
Intercept	88.66	1	88.66	647.75	<0.001
Wood type	8.44	3	2.81	20.54	<0.001
Height class	6.6	5	1.32	9.64	<0.001
Wood type*Height class	9.01	15	0.6	4.39	<0.001
Error	364.09	2660	0.14		

These results show that the damage scores are significantly different for the different wood types and height classes. This means that the elephant are targeting specific wood types more than others; as well as focusing their feeding attention on tree species at specific heights.

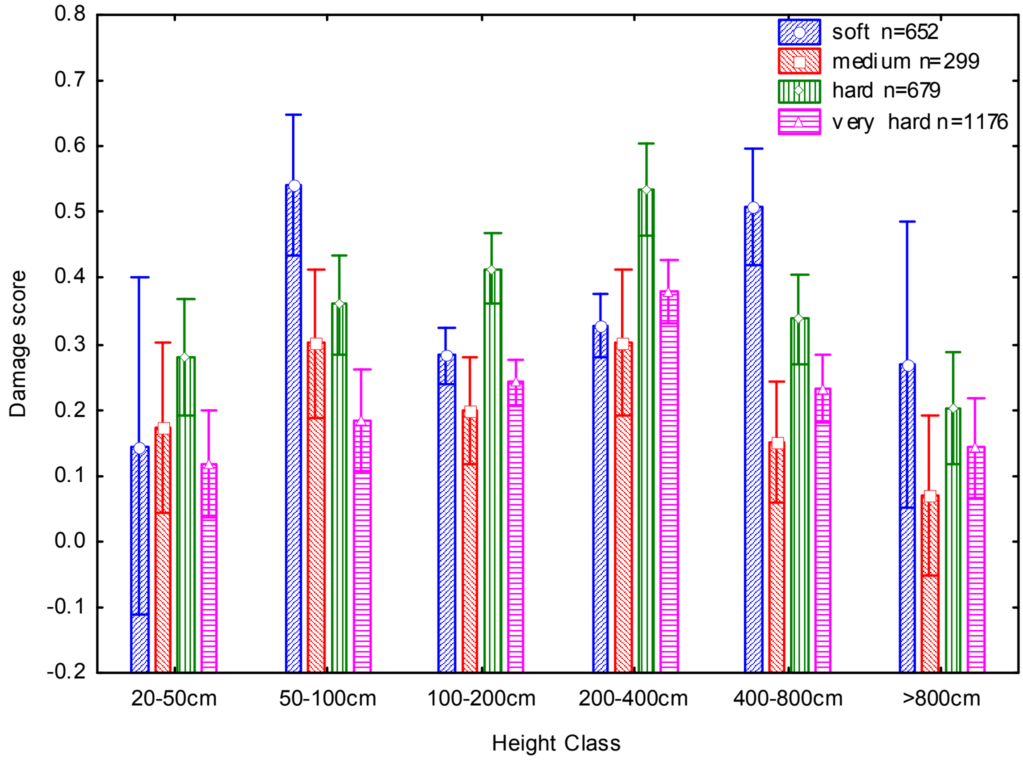

Figure 7.5 Graph showing differences in damage score for the wood types of different height classes – 1997 (mean ± SE)

This result is further illustrated in Figure 7.5 where the hard wood and the soft wood types have a consistently higher damage score than the medium and very hard woods. It also shows that the height class of 200-400cm shows more damage than the others, particularly for the soft and hard woods. Tukey's test confirms that there are significant differences between the wood types; and shows that there is no significant height class effect for the medium woods. The very hard woods have a significant damage difference between each height class; as do the hard woods. The soft wood type has significant differences in damage scores for the categories between 50cm and 800cm.

Since the wood type has a significant effect on the damage score; further ANOVA's will be run examining whether there is a species specific effect on the results of the total damage scores. The wood type narrows down the species list, since tree species will belong to only one wood type group. These results are shown in Table 7.4. These analyses were run by fixing the wood type and running a factorial ANOVA for the species vs. the height class. Due to the lack of data for some species, only those, for which there was sufficient data to perform the ANOVA, were included in the analysis. For the soft wood and medium wood types, there was only one tree species with sufficient data; and hence a one-way ANOVA was computed. In all cases, the height class of lower than 20cm had insufficient data and hence was excluded from the analysis.

For the soft wood, *Croton megalobotrys* was the only species with sufficient data (n=629); while for the medium wood type, *Lonchocarpus capassa* had sufficient data in order to compute the one-way ANOVA (n=299). The hard wood analysis included *Diospyros mespiliformis*, *Acacia robusta* and *Combretum hereroense*; and the very hard wood species were *Colophospermum mopane, Spirostachys africana*, and *Acacia nigrescens*. The *Croton megalobotrys* was the only soft wood species analyzed; and it was found above that the soft wood type had a generally higher damage score than the medium and very hard woods. This is expected since this species is heavily utilized by elephant and other game species (Schmidt *et al.*, 2002).

Table 7.4 Results of ANOVA's for fixed wood type - 1997

Wood Type	Effect	SS	df	MS	F	*p*
Soft	**Intercept**	18.15	1	18.15	153.03	<0.001
	Height class	4.22	4	1.05	8.89	<0.001
	Error	74.02	624	0.12		
Medium	**Intercept**	10.57	1	10.57	96.34	<0.001
	Height class	1.64	5	0.33	3.00	0.012
	Error	31.59	288	0.11		
Hard	**Intercept**	16.29	1	16.29	88.40	<0.001
	Species	0.39	2	0.20	1.06	0.346
	Height class	2.93	5	0.59	3.18	0.008
	Species*Height class	1.19	10	0.12	0.65	0.772
	Error	89.00	483	0.18		
Very Hard	**Intercept**	6.10	1	6.10	54.71	<0.001
	Species	0.69	2	0.34	3.07	0.047
	Height class	0.37	5	0.07	0.66	0.658
	Species*Height class	1.72	10	0.17	1.55	0.118
	Error	86.62	777	0.11		

The height class effect was significant for all wood types except the very hard woods; and the species effect was significant for the very hard woods, but not the hard woods. From Figure 7.6, it can be seen that the damage caused to hard wood tree species, is highest when the trees are between 1m and 4m in height however the differences were not show to be significant in the post-hoc tests. However, for the very hard wood tree species, the height class is not significant (p=0.658) whereas the species effect is significant (p=0.047). This seems to indicate that elephants have preferred species such as *C. mopane* and *A. nigrescens* and despite the wood being very hard, they will utilize these trees quite heavily; no matter what the size of the tree. The Tukey post-hoc tests confirm that the differences between *C. mopane* and *S.africana* are significant for three of the height classes.

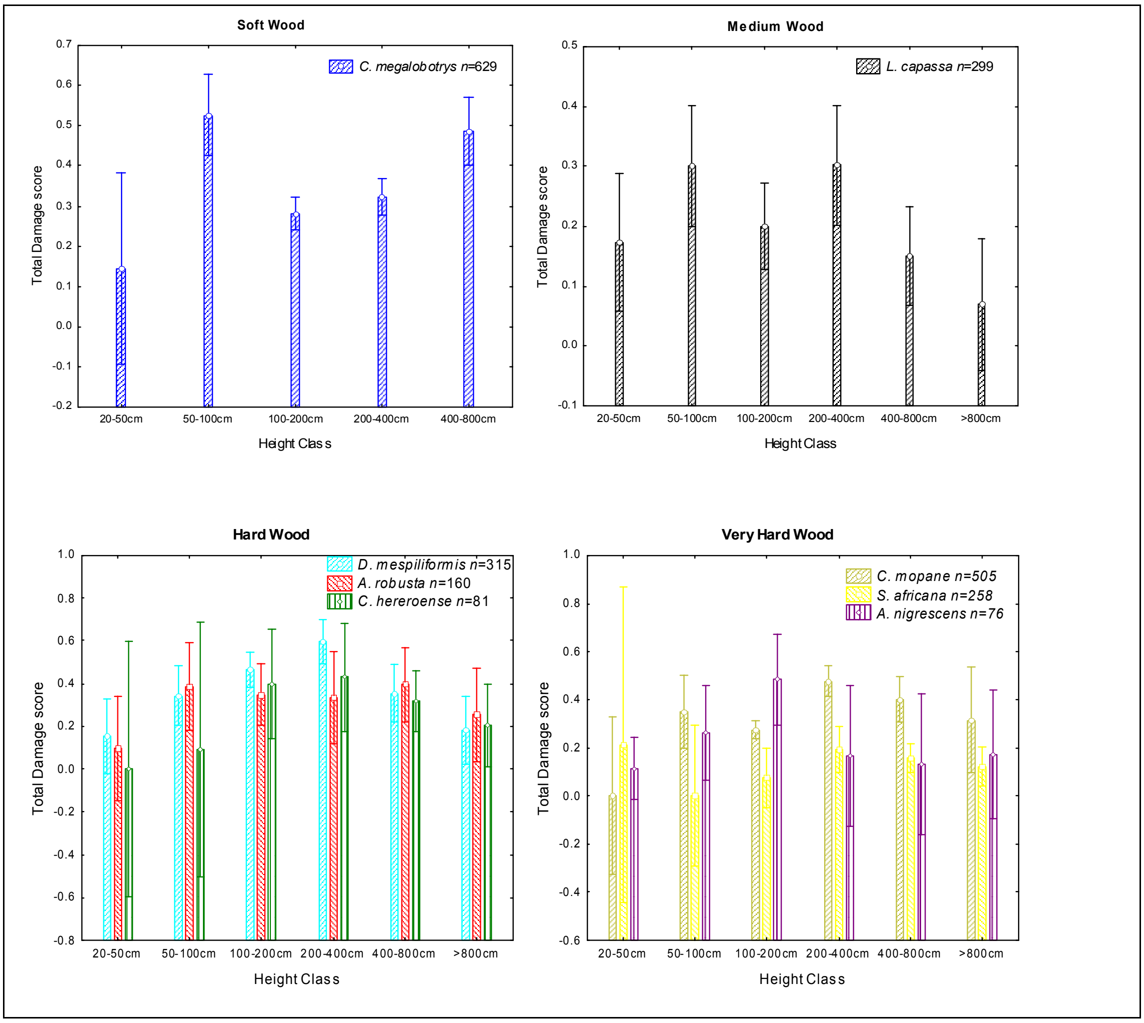

Figure 7.6 Graphs showing the effect on the damage score of the height class and species for fixed wood type – 1997 (mean ± SE)

7.6.2 *Results for 1998*

The analyses were then repeated for 1998. The initial wood type vs. height class analysis was run and the results are shown in Table 7.5.

Table 7.5 ANOVA results for Height class vs. Wood type - 1998

Effect	**SS**	**df**	**MS**	**F**	***p***
Intercept	149.97	1	149.96	624.83	<0.01
Wood type	1.74	3	0.58	2.41	0.065
Height class	2.54	6	0.42	1.76	0.102
Wood type*Height class	12.13	18	0.67	2.81	<0.01
Error	605.06	2521	0.24		

For 1998, the results differ from 1997 in that the wood type and the height class are not significant factors. Ecologically this means that during 1998, the elephants were not necessarily targeting specific wood types or height classes of trees; but rather that certain wood types within specific height classes were being impacted more heavily than others. This difference between the results for 1997 and 1998 is further illustrated in Figure 7.7; where there is no one wood type that is suffering consistently more damage than another. This is in contrast to 1997 where the hard and soft woods were more heavily impacted than the other two wood types.

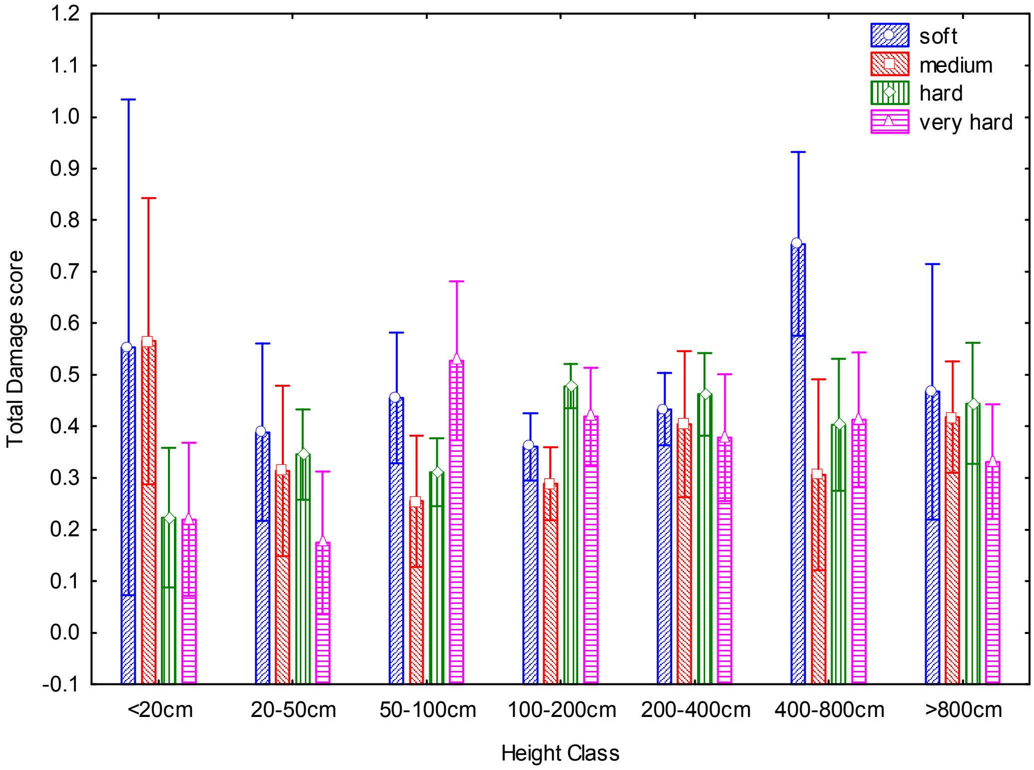

Figure 7.7 Graph showing differences in damage score for the wood types for different height classes – 1998 (mean ± SE)

Despite the lack of significance of the wood type factor, a closer analysis was run to see whether certain species within the wood types were more heavily damaged than others. A similar analysis as that done for the 1997 data was done; and the results are summarized in Table 7.6. However, it is noted that the number of species sampled in each year is different; and in some case different species and hence the species that are analyzed in the four wood types for the three years does vary. Once again, a one-way

ANOVA was run for the soft wood type due to lack of data; and factorial ANOVA's were run for the other three wood types.

Table 7.6 Results of ANOVA's for fixed wood type - 1998

Wood Type	**Effect**	**SS**	**df**	**MS**	**F**	***p***
Soft	**Intercept**	21.84	1	21.84	84.06	<0.001
	Height class	3.57	5	0.71	2.75	0.019
	Error	104.95	404	0.26		
Medium	**Intercept**	12.28	1	12.28	55.28	<0.001
	Species	0.00	1	0.00	0.00	0.969
	Height class	2.08	5	0.42	1.87	0.098
	Species*Height class	1.21	5	0.24	1.09	0.365
	Error	77.52	349	0.22		
Hard	**Intercept**	17.53	1	17.53	64.50	<0.001
	Species	0.33	2	0.16	0.60	0.547
	Height class	1.38	5	0.28	1.01	0.410
	Species*Height class	1.81	10	0.18	0.67	0.754
	Error	81.55	300	0.27		
Very Hard	**Intercept**	14.51	1	14.51	60.06	<0.001
	Species	1.03	1	1.03	4.25	0.040
	Height class	5.47	6	0.91	3.77	0.001
	Species*Height class	1.65	6	0.28	1.14	0.340
	Error	68.15	282	0.24		

The results confirm that there is no significant difference in the damage effect between the species, except for the very hard woods. For the soft wood, the height class is significant (p=0.019); however this factor is not significant for the medium wood (p=0.098); and hard wood (p=0.410). For the very hard woods, there is a significant effect on the damage score as a result of the species and the height class.

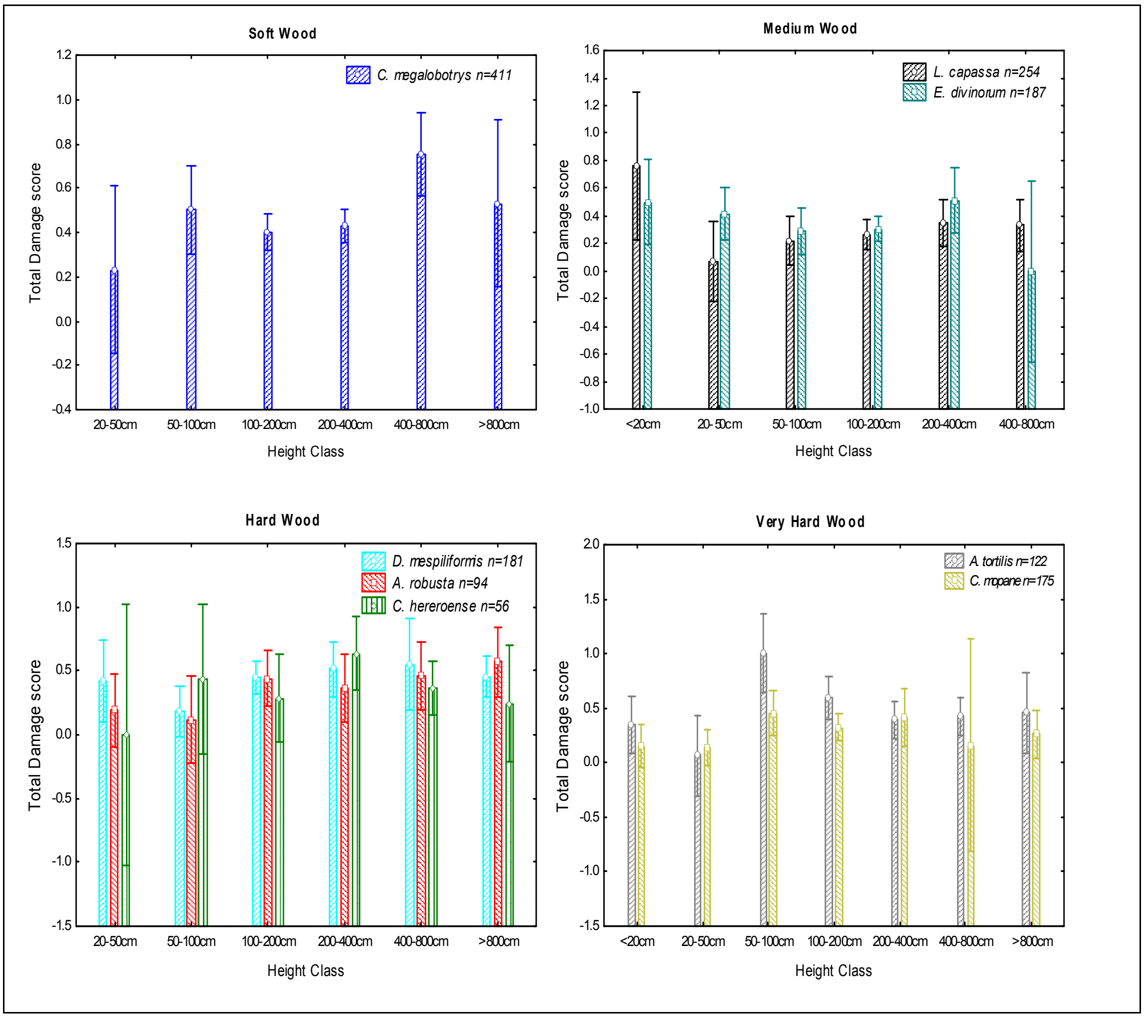

Figure 7.8 Graphs showing the effect on the damage score of the height class and species for fixed wood type – 1998 (mean ± SE)

The graphs in Figure 7.8 show that for the very hard wood types, *C. mopane* has a significantly lower damage score than *A. tortilis*; and the damage does not appear to be concentrated on one particular height class for *C. mopane* whereas the damage is severest on *A. tortilis* trees between 50cm and 1m in height. These differences were confirmed with the post-hoc tests which showed significant differences between the two species at the 50-100cm height class. For the soft wood, *C. megalobotrys* has a significant difference in damage score between the 1-2m and the 4-8m height classes, with the severest damage being inflicted upon the larger trees. This means that there are noticeable differences in the elephant utilization between 1997 and 1998. The possible reasons for these differences are further examined in section 7.7.

7.6.3 *Results for 1999*

The results of the ANOVA for height class and wood type are summarized in Table 7.7. Once again, the results show a different outcome than those that were found for the two preceding years. For 1999, the height class had a significant effect on the damage score; however the wood type alone was not a significant factor.

Table 7.7 ANOVA results for Height class vs. Wood type - 1999

Effect	**SS**	**df**	**MS**	**F**	***p***
Intercept	32.90	1.00	32.90	164.97	<0.01
Wood type	0.84	3.00	0.28	1.40	0.240
Height class	11.28	6.00	1.88	9.43	<0.01
Wood type*Height class	11.77	18.00	0.65	3.28	<0.01
Error	230.97	1158.00	0.20		

These results are illustrated in Figure 7.9 where a trend can be seen for the damage score for the different height categories. The severest damage is inflicted to those trees in the 2m to 4m height range; but once again there is no consistent pattern of damage in terms of the wood types. The post-hoc tests reveal some differences are significant, particularly between the 20-50cm height and the 2m-8m range.

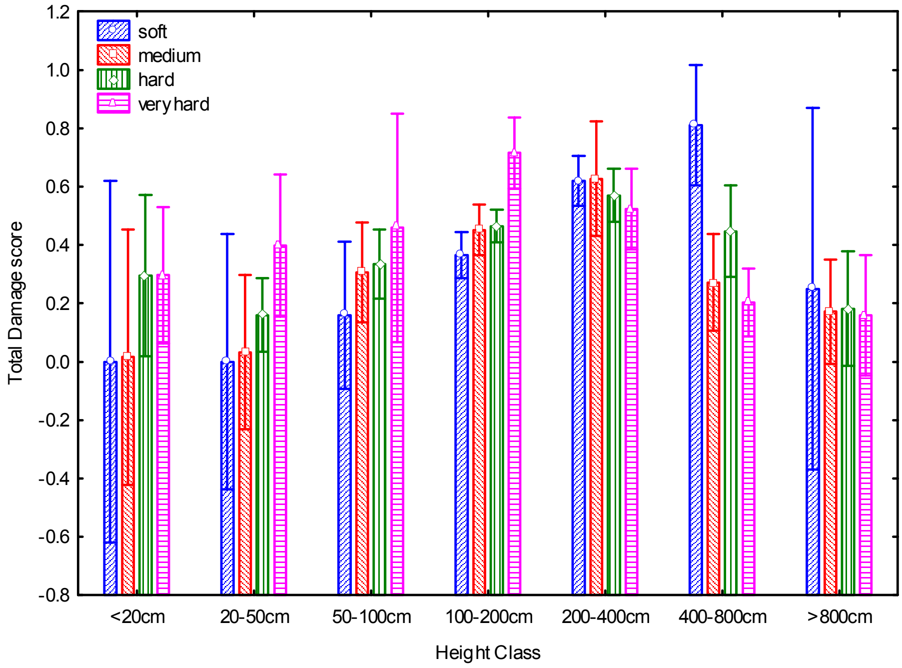

Figure 7.9 Graph showing differences in damage score for the wood types in different height classes – 1999 (mean ± SE)

Once again, despite their being no significant effect on the damage score due to the wood type; it was decided to determine whether any significant differences in damage severity do exist between the different tree species. The results are summarized in Table 7.8.

Table 7.8 Results of ANOVA's for fixed wood type - 1999

Wood Type	**Effect**	**SS**	**df**	**MS**	**F**	***p***
Soft	**Intercept**	5.27	1	5.27	25.92	<0.001
	Height class	3.99	4	1.00	4.90	0.001
	Error	40.67	200	0.20		
Medium	**Intercept**	3.88	1	3.88	21.40	<0.001
	Species	0.01	1	0.01	0.08	0.783
	Height class	3.78	4	0.95	5.22	0.001
	Species*Height class	1.88	4	0.47	2.59	0.039
	Error	27.55	152	0.18		
Hard	**Intercept**	15.11	1	15.11	83.16	<0.001
	Species	2.08	2	1.04	5.73	0.004
	Height class	2.42	4	0.61	3.33	0.011
	Species*Height class	3.53	8	0.44	2.43	0.014
	Error	67.98	374	0.18		
Very Hard	**Intercept**	9.47	1	9.47	57.88	<0.001
	Species	0.54	1	0.54	3.32	0.072
	Height class	3.19	3	1.06	6.49	<0.001
	Species*Height class	1.03	3	0.34	2.11	0.105
	Error	15.54	95	0.16		

Once again, there was insufficient data to run a factorial ANOVA on the soft wood species; and only *C. megalobotrys* was analyzed. The height class for all species did have a significant effect on the damage score. In most cases the damage was most severe for trees between 1m and 4m tall. The species effect was significant for the hard woods only; and from Figure 7.10 it is shown that *A. robusta* has a significantly higher damage score than the other two species for the tree species lower than 50cm. For the very hard woods, the post-hoc tests confirm that there is a significant difference in the damage inflicted upon C. *mopane* and *S. africana* for the 2m-4m and the 4m-8m height classes, with the Mopane being more damaged.

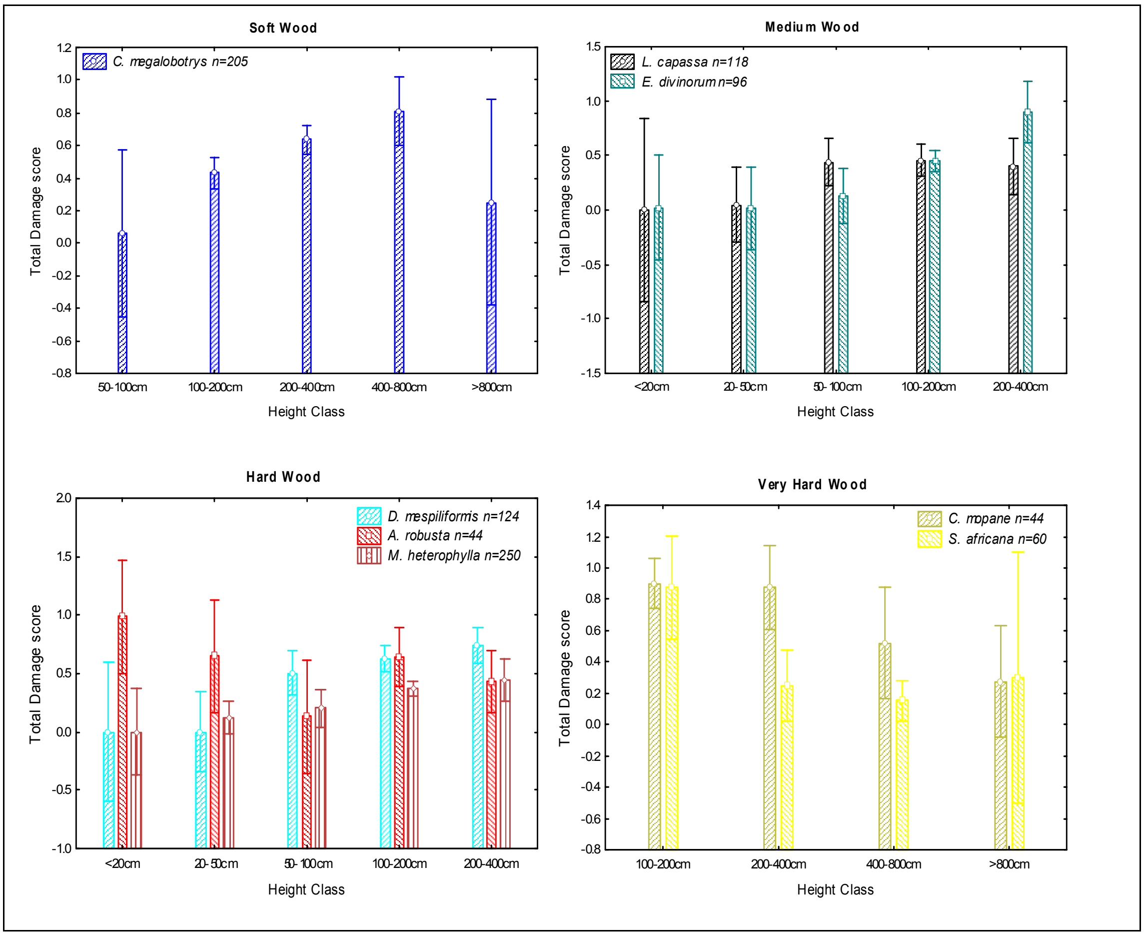

Figure 7.10 Graphs showing the effect on the damage score of the height class and species for fixed wood type – 1999 (mean ± SE)

7.7 SUMMARY

The results of the ANOVA for wood type vs. height class were surprisingly different for the 3 years of data; one possible reason for this difference is the different rainfall that was experienced over these 3 years. 1997 was a low rainfall year, 1998 was average and 1999 experienced high rainfall. It has been shown that elephants can change their diets depending on the season and the amount of vegetation available; Scholes (2004) noted that elephants are preferentially grazers that can switch to browsing when the quantity and quality of appropriate grazing is depleted. In the dry season, they tend to eat a higher proportion of woody vegetation; whereas in the wetter seasons they will concentrate their feeding on the grass species. This theory is supported by the analyses done for this project since the elephant utilization of woody species is most apparent in 1997 which was a dry year. The low rainfall would have resulted in less abundant grazing and hence the elephant would have consumed more woody vegetation.

The data for the three years does show that the heaviest utilization of the tree species occurs between the heights of 1m and 4m. Before the tree has reached one metre, it is probably ignored as not being worth eating; and once it has exceed 4m; it is utilized but the tree is strong enough that the damage does not pose such a severe threat to the tree. This is a similar result as that obtained by Matthews and Page (2004) who did a study of woody species use by elephants in the Tembe Elephant Park in northern Kwazulu-Natal. Their research showed that trees in the 3m – 5m height range are used more frequently than other height classes (Matthews & Page, 2004). This adds confirmation to the theory that elephant damage is similar to fire damage; and that beyond 4m in height, the tree is able to withstand both burning and elephant usage far better than when it is smaller.

Although the tree species within the riparian zone are being heavily utilized by elephants, there does not appear to be one species that is suffering noticeably more than any other. In fact, the changing climatic conditions from year to year seem to influence the elephants to concentrate their feeding on different species; thus giving the others a chance to recover and grow. This growth will bring them closer to the 4m point where they should then be able to recover from the utilization more successfully. This adds weight to the theory that none of the riparian zones in the KNP appear to be threatened (Owen-Smith, 2004c).

One of the conclusions of Gaylard’s study was that the natural spatial and temporal heterogeneity of elephant impacts needs to be restored and this can be done by either reducing the elephant population and/or closing of artificial waterpoints; in order to restore the maximum diversity of the riparian zones (Gaylard, 2004). These sentiments are echoed by Ron Thompson who has spent five decades working in and around Africa’s national park systems. He believes that elephant numbers should be reduced to below the current habitat carrying capacity level until the habitats recover from the elephant damage; and then the numbers are only allowed to increase to, and be maintained at their then optimal carrying capacity (Thomson, 2005). However, the feeding of elephants is not all negative. They coppice the trees which results in feeding opportunities being created for the smaller herbivores (Lagendijk *et al*., 2004).

There are always two sides to any story and in terms of the elephant debate, there are far more than two sides to the story. Elephants are a natural asset to the KNP as are the trees and vegetation that make up the KNP; and all the other species that live within the KNP. It should be the aim of the management to safeguard the threatened components of the ecosystem while letting the natural components run their course as much as possible without human interference, bearing in mind that it is a fenced park (Owen-Smith, 2004c).

CHAPTER 8: CONCLUSION

As stated in the Introduction, the mission of the Kruger National Park is “to develop and manage a system of national parks that represents the biodiversity, landscapes and associated heritage assets of South Africa for the sustainable use and benefit of all” (Pienaar, 2004). Statistical analysis has been used in this project to analyse the changes that have occurred in the natural system and investigate the possible results of current management policies and their goal of maintaining the biodiversity of the KNP.

The KNP is a constantly changing and evolving ecosystem; and the animals within the reserve are some of the agents of change that are driving the evolution. The N’washitsumbe enclosure is an important example of what the vegetation would look like without the variety of species that are present throughout the Park. The significant differences in the vegetation structure and composition between the inside and the outside of the enclosure provide an indication of the direction and types of changes that occur when the species within an area are limited. Although the artificial waterpoint closures are not expected to result in such extreme differences, it is expected that the vegetation changes will follow a similar trend. This is encouraging for the prospects of the roan antelope and it is hoped that their population can begin to strengthen in their prime territories. The elephant population, on the other hand, is increasing to levels not seen in the history of the KNP. Although it seems that their current feeding habits are not threatening particular riparian trees, it is noted that the years of that study were between 7 and 9 years ago and in that time the elephant population has increased quite substantially. Thus, the management of the KNP need to monitor the effects that the population increase is having on the entire ecosystem and changes need to be made to the current elephant management policy and implemented rapidly when it appears that the elephants are negatively impacting on their surroundings and the other species of animal, bird, plant, and reptile.

Statistical analysis will always play a crucial role in ecological research projects and the predictions obtained will be able to provide valuable insight into the changing dynamics of the wildlife ecosystems.

REFERENCES

Agresti, A. (1996) Categorical Data Analysis, Wiley, New Jersey.

Agresti, A. (2002) Categorical Data Analysis, 2nd edition, Wiley, New Jersey.

Apps, P. (1992) Field Guide to the Behaviour of Southern African Mammals Wild Ways, Southern Book Publishers, Halfway House.

Badenhorst, J. (2006) "Don't blame the elephants… an opinion", Kruger Park Times, July 12 to July 28, p. 13.

Christensen, R. (1990) Log-linear Models, Springer-Verlag, New York.

Codron, J., Lee-Thorp, J.A., Sponheimer, M., Grant, C.C., Codron, D. (2004) "Carbon Isotopic Assessment of the Diets of Elephants in Kruger National Park - Seasonal and Spatial Variability". Available online: http://www.sanparks.org/parks/kruger/conservation/scientific/key_issues/elephants.php (Downloaded: November 2005).

Dyksterhuis, E. J. (1949) Condition and management of range land based on quantitative ecology, Journal of Range Management, 2:104-115.

Everitt, B.S. (1977) The Analysis of Contingency Tables, Chapman & Hall, London.

Fourie, P.F. (1984) Kruger National Park – Questions and Answers, SA Country Life, Nelspruit.

Gaylard, A. (2004) “Introduction to the use of frameworks for knowledge management and conflict resolution” and "Elephant-induced changes to the compositional and structural diversity of riparian woody vegetation in relation to surface water". Available online: http://www.sanparks.org/parks/kruger/conservation/scientific/key_issues/elephants.php (Downloaded: November 2005).

Grant, R., de Buys, A. (2004) "Northern Plains Programme". Available online: http://www.sanparks.org/parks/kruger/conservation/scientific/programmes/North_Plains_Prog.pdf (Downloaded November 2005).

Grant, R., de Buys, A. (2005) "Northern plains programme Annual Report 2004". Available online: http://www.sanparks.org/parks/kruger/conservation/scientific/programmes (Downloaded November 2005).

Grant, R., Freitag-Ronaldson, S. (2004) Rare Game Breeding Enclosures in the Kruger National Park, Internal Report, Skukuza, Kruger National Park.

Hofmeyr, M., Eckhardt, H. (2004) “Elephant Effects of Vegetation”. Available online: http://www.sanparks.org/parks/kruger/conservation/scientific/key_issues/elephants.php (Downloaded: November 2005).

http://www.bushveld.co.za/animal-pictures-sable.htm (Downloaded October 2006).

http://www.outdoorphoto.co.za/forum/photopost/showphoto.php (Downloaded March 2006).

http://www.thekruger.com/category/lists-and-maps/ (Downloaded October 2006).

http://www.sabisandslodges.co.za/Plants.asp (Downloaded July 2006).

http://www.wild-about-you.com/GameRoan.htm (Downloaded November 2005).

http://www.wild-about-you.com/GameSable.htm (Downloaded March 2006).

Kloppers, J.J., Bornman H. (2005) A Dictionary of Kruger National Park Place Names, SA Country Life, Barberton.

Kruger Park Times (2005) "Changing rainfall has surprising results for the trees", September 1 to September 14, p. 8.

Lagendijk, D.D.G., de Boer, W.F., van Wieren, S.E. (2004) "Can African Elephants Survive and Thrive in Monostands of Colophospermum mopane Woodlands?" Available online: http://www.sanparks.org/parks/kruger/conservation/scientific/key_issues/elephants.php (Downloaded: November 2005).

Larntz, K. (1993) Analysis of categorical response variables. In: Applied Analysis of Variance in Behavioural Science, Edwards, I.K., Ed., Marcel Dekker, New York.

Matthews, W.S., Page, B.R. (2004) "The comparitive use of woody species in different habitats by elephants in Tembe Elephant Park, Maputuland, Northern Kwazulu-Natal". Available online: http://www.sanparks.org/parks/kruger/conservation/scientific/key_issues/elephants.php (Downloaded: November 2005).

Owen-Smith, N. (2004a) "Ecosystem Resources Influencing Elephant Populations". Available online: http://www.sanparks.org/parks/kruger/conservation/scientific/key_issues/elephants.php (Downloaded: November 2005).

Owen-Smith, N. (2004b) "Summary and Conclusions". Available online: http://www.sanparks.org/parks/kruger/conservation/scientific/key_issues/elephants.php (Downloaded: November 2005).

Owen-Smith, N. (2004c) "Vegetation - Elephant Interactions". Available online: http://www.sanparks.org/parks/kruger/conservation/scientific/key_issues/elephants.php (Downloaded: November 2005).

Pienaar, D. (2004) “Elephant Effects on Biodiversity; an assessment of current knowledge and understanding as a basis for elephant management in SANParks”. Available online: http://www.sanparks.org/parks/kruger/conservation/scientific/key_issues/elephants.php (Downloaded: November 2005).

Quinn G.P., Keough M.J. (2002) Experimental Design and Data Analysis for Biologists, Cambridge University Press, Cambridge.

Schmidt, E., Lotter, M., McCleland, W. (2002) Trees and Shrubs of Mpumalanga and Kruger National Park, Jacana, Johannesburg.

Scholes, R.J. (2004) "Some recent findings relevant to the elephant population question". Available online: http://www.sanparks.org/parks/kruger/conservation/scientific/key_issues/elephants.php (Downloaded: November 2005).

Stuart, C., Stuart, T. (1992) Field Guide to the Mammals of Southern Africa, Struik, Cape Town.

Thomson, R. (2005) "Your wildlife is being hung out to dry", Kruger Park Times, August 4 to August 17 2005, p. 7.

Travers, R. (2006) Borehole Water Association. “Borehole Closures in the Kruger National Park”. Available online: http://www.bwa.co.za/articles.htm (Downloaded: July 2006).

Upton, G.J.G. (1980) The Analysis of Cross-tabulated Data, Wiley, New York.

Whyte, I. (2004) “Background to Elephant Growth”. Available online: http://www.sanparks.org/parks/kruger/conservation/scientific/key_issues/elephants.php (Downloaded: November 2005).

Wray, M. (2005) “Do elephants test their tusks on trees?”, Kruger Park Times, December 22 to January 16, p. 5.