



Statistical distributions and modelling of GPS-Telemetry elephant movement data including the effect of covariates

by

ROBERT MATHENGE MUTWIRI

B.Sc Applied Statistics,

M.Sc Social Statistics

A thesis submitted in fulfillment of the requirements for the Degree
of Doctor of Philosophy in Statistics

The School of Mathematics, Statistics and Computer Science

University of KwaZulu Natal

South Africa

Declaration of Authorship

I, Robert Mathenge MUTWIRI, declare that this thesis titled, ‘Statistical distributions and modelling with applications to GPS-telemetry data of elephant movement with covariates.’ and the work presented in it are my own work. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

Signature (Robert Mathenge Mutwiri)

Date

As the candidates supervisor, I certify the above statement and have approved this thesis for submission.

Signature(Prof Henry Mwambi supervisor)

Date

List of manuscripts

The material presented in this thesis have resulted in the following manuscripts submitted or in their final stage for submission.

1. Mutwiri, R.M, Mwambi H and Slotow R.(Under review). Stable distributions: Theory and Applications.
2. Mutwiri, R.M, Mwambi H and Slotow R.(Under review). Approaches for analysing heavy tailed and skewed animal movement: application to elephant movement time series data.
3. Mutwiri, R.M, Mwambi H and Slotow R.(under review). A nonlinear regression model with stable law errors with application to animal movement data.
4. Mutwiri, R.M, Mwambi H, Slotow R and Vanak A. T. (Under review). A set of circular Statistical models of GPS tracking data with application to elephant movement. (*Presented at South Africa Statistical Association(SASA) conference 2013 at university of Limpompo, South Africa*).
5. Mutwiri, R.M, Mwambi H and Slotow R.(Under review). Approaches for testing the circular uniformity hypothesis in movement of large mega-herbivores in Kruger National park South Africa.

Abstract

In this thesis, I investigate the application of various statistical methods towards analysing GPS tracking data collected using GPS collars placed on large mammals in Kruger National Park, South Africa. Animal movement tracking is a rapidly advancing area of ecological research and large amount of data is being collected, with short sampling intervals between successive locations. A statistical challenge is to determine appropriate methods that capture most properties of the data is lacking despite the obvious importance of such information to understanding animal movement. The aim of this study was to investigate appropriate alternative models and compare them with the existing approaches in the literature for analysing GPS tracking data and establish appropriate statistical approaches for interpreting large scale mega-herbivore movements patterns. The focus was on which methods are the most appropriate for the linear metrics (step length and movement speed) and circular metrics (turn angles) for these animals and the comparison of the movement patterns across herds with covariate. A four parameter family of stable distributions was found to better describe the animal movement linear metrics as it captured both skewness and heavy tail properties of the data. The stable model performed favourably better than normal, Student's t and skewed Student's t models in an ARMA-GARCH modelling set-up. The flexibility of the stable distribution was further demonstrated in a regression model and compared with the heavy tailed t regression model. We also explore the application circular linear regression model in analysing animal turn angle data with covariate. A regression model assuming Von Mises distributed turn angles was shown to fit the data well and further areas of model development highlighted.

A couple of methods for testing the uniformity hypothesis of turn angles are presented. Finally, we model the linear metrics assuming the error terms are stable distributed and the turn angles assuming the error terms are von Mises distributed are recommended for analysing animal movement data with covariate.

Contents

Declaration of Authorship	i
List of manuscripts and Conferences attended	i
Abstract	ii
Table of Contents	iv
List of Figures	vii
List of Tables	x
Abbreviations	xi
Dedication	xiii
1 Introduction	1
1.1 General introduction	1
1.2 GPS Tracking data and metrics of animal movement	4
1.2.1 Derivation of movement metrics	5
1.2.2 Drivers of animal movement patterns	6
1.2.3 Probability distributions and their importance in animal movement	7
1.2.4 Modelling autocorrelated heavy tailed data of animal move- ment	9
1.2.5 Regression Models for heavy tailed linear metrics of animal movement with covariates	10
1.2.6 Modelling Circular responses with linear covariates	12
1.2.7 Testing circular hypothesis of uniformity	14
1.3 Research Objectives	15
1.4 Thesis Outline	15
1.5 Contribution of the present study	17
2 Stable Distributions: Theory and Application to Animal move- ment data	20

2.1	Introduction	21
2.2	Stable distribution	24
2.3	Biological motivations for the stable distribution	27
2.4	Parameter estimation	29
2.4.1	Asymptotic properties of the power law distribution	29
2.4.2	Quantile estimation methods	31
2.4.3	Koutrouvelis algorithm and Regression-type technique	32
2.4.4	Maximum likelihood estimation	33
2.4.5	Assessing the Goodness of fit	34
2.5	Real data application	35
2.6	Results of analysis and interpretation	36
2.7	Final remarks	42
3	Analysing the effects of landscape factors on animal movement step lengths with stable law regression models	44
3.1	Introduction	45
3.2	Stable distribution	47
3.2.1	Stable Paretian regression model	49
3.2.2	Regression model with t errors	51
3.3	Application to elephant movement data	53
3.3.1	Data description	53
3.3.2	Vegetation cover types	54
3.3.3	Model formulation	54
3.4	Results of the analysis	55
3.4.1	Biological implications and applications	58
3.5	Conclusion	59
4	Approaches for analysing heavy tailed and skewed animal movement: application to elephant movement time series data	61
4.1	Introduction	62
4.2	Methods and Materials	64
4.2.1	Stable autoregressive models	64
4.3	Statistical properties of step lengths time series	67
4.4	Parameter estimates and interpretation	72
4.5	Conclusion	75
5	Application of circular regression models to animal movement data	76
5.1	Introduction	77
5.2	The Von Mises Distribution	78
5.3	Circular orientation model	80
5.3.1	Computing the mean parameter estimates	81
5.3.2	Tests and confidence intervals	83
5.3.3	Model checking and diagnostics testing	83
5.4	Applications	84

5.4.1	An illustration using simulated data	84
5.4.2	Elephant movement data	86
5.5	Results	87
5.6	Discussion	90
6	Approaches for testing the circular uniformity hypothesis in movement of large mega-herbivores in Kruger National park South Africa	92
6.1	Introduction	93
6.2	Theory of hypothesis testing and notation	94
6.2.1	Rayleigh's test	96
6.2.2	Kuiper's test statistic	96
6.2.3	Watson's test	97
6.2.4	Rao's spacing test	97
6.3	Application to elephant movement data	98
6.3.1	Ethics statement	98
6.3.2	Study area and GPS Data on a elephants	98
6.4	Discussion	100
7	Conclusion	105
7.1	Summary conclusions	105
7.2	Further Work	107
	Bibliography	108

List of Figures

2.1	Effects of α, β, γ and μ on the stable density function (matlab image)	26
2.2	Histograms of the step lengths of the six elephant herds collected from Kruger National Park (2006-2009)	37
2.3	Stable distribution density function plots for elephant movement step lengths	40
2.4	Probability plots (pp) of stable distribution of six elephant herds movement step data collected from KNP between 2006-2009 lengths	41
3.1	Diagnostic analysis of elephant movement data	57
4.1	Time series plots of the step lengths of Herd AM108 of GPS-derived telemetry data collected from Kruger National Park South Africa (May 2006 -April 2009) (30minutes time intervals)	70
4.2	Histogram plot of the step lengths of Herd AM108 GPS-derived telemetry data collected from Kruger National Park South Africa (May 2006 -April 2009)	71
4.3	QQ plots of standardized residuals from an ARMA-GARCH(1,1,1,1) to step lengths of elephant movement.	71
5.1	Simulation summaries for the von Mises distribution	85
5.2	Histogram of elephant turn angle data.	87
5.3	Rose diagram of elephant turn angle data.	88
5.4	Displays of wind rose diagram of elephant turn angle data.	88
5.5	Diagnostic plot of the fitted von mises regression model of elephant movement.	90
5.6	Probability plot of elephant movement turn angle model.	90
6.1	Flowchart representing the sequence of hypothesis tests based on circular distributions.	95
6.2	Histogram of herd AM108, AM307 and AM308 GPS-telemetry derived turn angles data collected from Kruger National Park South Africa (May 2006 -April 2009)	102
6.3	Histogram of herd AM108, AM307 and AM308 GPS-telemetry derived turn angles data collected from Kruger National Park South Africa (May 2006 -April 2009)	102

6.4	Histogram of herd AM108, AM307 and AM308 GPS-telemetry derived turn angles data collected from Kruger National Park South Africa (May 2006 -April 2009)	103
6.5	Circular rose diagrams of the elephant herd AM108, AM307 and AM308 GPS-telemetry derived turn angles data collected from Kruger National Park South Africa (May 2006 -April 2009). Circular bars indicate the number of observations within each class range.	104

List of Tables

2.1	Descriptive analysis of six herds of Elephant movement step length in Kruger National Park collected every 30minutes	38
2.2	Comparison of parameter estimates and goodness of fit for the stable and power law distribution of elephant movement step lengths data sets collected from Kruger National Park South Africa recorded every 30 minutes from 2006 to 2009.	39
3.1	Dummy Variables of vegetation cover type under investigation . . .	55
3.2	Summary of Heavy tailed -t distribution and stable law regression model	57
4.1	Descriptive analysis of two female elephant movement step length in Kruger National Park collected every 30 minutes	69
4.2	KPSS tests and Augmented Dickey-Fuller	69
4.3	Step Length autocorrelation	70
4.4	Maximum likelihood estimates and goodness of fit statistics of Stable, normal, student's t and skewed student's t distribution ARMA-GARCH(1,1,1,1) of elephant movement step lengths	73
5.1	ML Estimates of directionality of turn angles of elephant movement	88
5.2	ML Estimates of landscape features on the mean direction of turn angles on Elephant movement($n = 4000$)	89
6.1	Descriptive analysis of the turn angle data of three elephant herds collected from Kruger National Park, South Africa.	101
6.2	Tests for uniformity for three elephant herds turn angles driven from GPS tracking data.	101
6.3	Testing if the elephant movement data is Von Mises distributed using Rao's spacing test.	102

Acknowledgement

Good science is rarely done by an individual working in a vacuum. I am indebted to many statisticians and scientist who have been my mentors, friends and collaborators over the past three years. My supervisor Professor Henry Mwambi, who has provided invaluable support throughout the course of my Ph.D. studies. He made himself available and offered assistance even at the busiest of the times. His contribution to both my thesis and my experience in the university of KwaZulu-Natal is greatly appreciated.

I also wish to thank Professor Rob Slotow (School of Life Sciences, University of KwaZulu-Natal) who gave me this topic of study and elephant data, his comments and discussions on movement ecology especially of elephants helped shape the scope of this thesis. As mentioned in the abstract, Statistical ecology is an interdisciplinary research area.

This thesis has benefited from the valuable contributions of Rob Slotow (The Amarula elephant research project), Vanak Abi(Ashoka Trust for Research in Ecology and the Environment, Bangalore, India) and Patricia Birkett (Amarula elephant research project) who has worked greatly with the elephants data.

Last, but not least, would like to thank my girlfriend Ann for providing continued encouragement, my fellow Ph.D. students for making my time at KwaZulu-Natal an enjoyable one, and my family for their support.

This Ph.D. was funded by University of KwaZulu-Natal college of Engineering and Agricultural sciences Ph.D. bursary, a bursary from my supervisor professor Henry Mwambi and Amarula elephant research project through professor Rob Slotow.

Abbreviations

ADF	Augmented Dickey-Fuller
AR(P)	Autoregressive models of order p
MA(q)	Moving average model of order q
ARMA(p,q)	Autoregressive moving average model of order p and q
FMEs	Fundamental movement elements
GARCH(r,s)	Generalized autoregressive conditional heteroscedasticity model of order r and s
GIS	Geographical information system
GLM	Generalized Linear Model
GOF	Goodness-of-Fit
GPS	Global positioning system
IID	Identically and independently distributed
IRLS	Iteratively re-weighted least squares
KNP	Kruger National park
KPSS	KwiatkowskiPhillipsSchmidtShin tests
MLE	Maximum likelihood estimation
SANPARKS	South Africa National parks

Dedication

I dedicate this work to God almighty for his mercy and care that endures forever. To my late Grandfather Zachariah whose dreams I cherish; may the Almighty God grant his soul eternal rest. To my sweetheart Ann for her moral support and patience during this long academic adventure. To my parents Henry and Margaret who have laboured to see me through school. To my two beloved brothers and a sister for being there for me.

CHAPTER 1

Introduction

1.1 General introduction

The large scale movement of mega-herbivore is difficult to monitor in situ. Therefore, little is known about the biological motives for the movement and behaviour of these animals. The location of migration corridors and residency areas in relation to habitat characteristics inform about the environmental preference of a species. Thus, access to this type of knowledge would increase the biological understanding of life in the wild and provide valuable insights into possible climate change responses of elephant and other mega-herbivores.

The population of numerous commercially targeted mega-herbivores are declining. Differences in abundance and size composition of predators indicate that a likely reason for the decline is hunting for Ivory (cite). It is therefore important for scientists and conservation commissions to understand the movement and behaviour of these species to determine the actions required to obtain sustainable levels for tourist attraction and ecosystem balance. The prime example is the African elephant (*Loxodonta africana*), which is heavily targeted by poachers for ivory.

Statistical methods for analyzing animal tracking data is one of the major challenges facing movement ecology today. Understanding the behaviour and movement patterns of foraging animals is a challenging endeavour due to the complex environment in which the search for food items takes place and lack of statistical framework for analysing the data[91]. Correspondingly, broad are the topics in movement and foraging ecology that require to be analysed. Advances in GPS technology are providing opportunities to document sequential movement patterns and habitat use in response to changing environmental conditions [131]. The GPS satellite link has enabled large amounts of data to be collected automatically without the effort required from and associated with conventional GPS radio-tracking devices. Geographic locations of the animal shows the regions in the landscape which are favoured, the habitat features and food or other resources sought by the animals bearing the collars [27].

Several metrics can be derived from the GPS- telemetry data to characterise animal movement [18, 53, 62]. These metrics includes the linear metric (step length, change in y and x axis, mean square displacement, net square displacement, etc.) and circular metric (turn angles, path tortuosity, sinuosity index etc). Given time, effort and monetary expenses devoted to obtaining such data from individual animals in the wild, it is important to analyze the data with valid statistical methods. This is particularly so because conclusions concerning animal movement may have management implications [99]. For example the analysis of GPS tracking data can reveal the distribution and space use pattern of species [23],foraging search strategies of several animal species [132, 133], and navigation strategies [35], that can facilitate applications into conservation ecology [27, 99].

Appropriate statistical assumptions that adequately capture the properties of the data are therefore necessary. In the following chapters, we will analyze real elephant movement data to answer more specific questions about the heavy tails and skewness properties in animal tracking data, the interaction between elephants and their environment and how can we model the foraging behaviour. Elephants are known to be destructive foragers [45]. Several factors are known to drive there movement. Among them distance to water points, vegetation cover types,

rain fall, temperature and seasons [27, 28]. Statistical models that incorporate such factors have important biological implications in understanding the animal movement pattern and behaviour. Getz and Saltz [45] classifies the drivers of animal movements as either internal or external factors. These factors are also known as fundamental movement elements(FMEs). To identify appropriate statistical models for analyzing GPS-telemetry data, a biological question that needs to be addressed by the data is required. Specifically, “what can we learn from the data about elephant movement and where can we apply the it in wildlife management?” have been used as the guiding questions through out this thesis.

Differences in abundance and the decline in elephant population has been attributed to increased poaching activities occasioned by high demand for ivory in the world markets [22]. One of the mechanisms of protecting these endangered species is monitoring the movement trajectory using the GPS tracking devices. Data generated by such devices can offer valuable conservation insights on the movement patterns of elephants. To make decisions based on GPS tracking data, relevant statistical tools for analyzing such data are required.

Another major problem in tracking data that is recorded over time is autocorrelation. Temporal autocorrelation, is an important understudied biological signal that can be critical to predictive modelling of population distribution [26]. Models in movement ecology that capture this property are lacking, as the linear metrics are heavy tailed and skewed. Most studies have recommended sub-sampling the data to remove autocorrelation. However, sub-sampling of the data leads to loss of information that may contain important biological signals [22]. We note some of the models used in financial time series to model heavy tailed and skewed data can be applied in movement ecology. In this study we have investigated the ARMA-GARCH models with Student’s t, skewed Student’s t or stable Paretian distribution error terms. These models are critical modelling the elephant movement data as the skewness and heavy tail properties are captured.

The analysis of circular metrics is lagging behind in ecology due to computational challenges and lack of tracking data rich with covariate. Circular regression models

can provide an important link between animal movement through the turn angle as a response variable and the landscape features such as rivers, water points, roads and fences [42, 124]. These models are important in assessing the effects of landscape features on animal movement. This thesis seeks to move statistical analysis of animal movement forward with the aim of evaluating the current statistical methods and proposing practical ones to obtain biologically and ecologically relevant information from the tracking data.

1.2 GPS Tracking data and metrics of animal movement

Recorded trajectories are often analyzed by considering the collection of fixes as a series of random events (e.g., displacements, turns) whose spatial and temporal distributions are assumed to possess certain statistical regularities. A useful conceptual model for animal locations is the home range model [32, 120]. Worton [140] and Solow [120] reviewed some issues arising in home range analysis and emphasized the significance of home ranges in animal movement. A common model for the home range of an individual animal is that the sequence of the locations through time follows a bivariate normal distribution in the plane [32, 120]. Under this model, the locations of the animal to be restricted to an elliptical home range, with the locations concentrated in the center. While the limitations of this model are well recognized, in many cases it is useful approximation under which inferences can be made about the parameters of the home range. Bivariate normal distribution is an important tool through which several metrics of animal movement can be derived and investigated.

In analyzing the sequence of locations, it is usually assumed that the observations are serially independent [123]. Essentially, this requires that the time between successive observations is large relative to the ranging characteristics of the animal. However, it is not easy to record all the information of the animal between two time points using the GPS telemetry as the animal may move around and return to the

same place it was before. The choice of temporal resolution is therefore important in understanding animal behavioral processes. Theoretical modeling of random searches and other encounter processes relies on the fact that the rate of biological encounters is determined by the statistical properties of the movement when all other things remain constant, such as target density, learning, and information availability.

1.2.1 Derivation of movement metrics

The most basic information that can be collected on the movement path of an individual is a sequence of positions (x_i, y_i) at time t , $i = 0, 1, 2, \dots, n$. Such information can be used in deriving biologically meaningful information about animal movement such as estimating the error of the GPS device and predicting the animal position at time t . The orientation of the vector joining (x_{i-1}, y_{i-1}) and (x_i, y_i) is given by

$$\theta_i = \begin{cases} \tan^{-1}((y_{i+1} - y_i)/(x_{i+1} - x_i)) & ; x_{i+1} - x_i > 0, \\ \tan^{-1}((y_{i+1} - y_i)/(x_{i+1} - x_i)) + \pi & ; x_{i+1} - x_i < 0, \\ \pi/2 & ; x_{i+1} - x_i = 0 \text{ and } y_{i+1} - y_i > 0, \\ -\pi/2 & ; x_{i+1} - x_i = 0 \text{ and } y_{i+1} - y_i < 0, \\ \text{undefined} & ; x_{i+1} - x_i = 0 \text{ and } y_{i+1} - y_i = 0. \end{cases} \quad (1.1)$$

and its resultant length is

$$\ell = r * (a \cos(\sin(y_1) * \sin(y_2) + \cos(y_1) * \cos(y_2) * \cos(x_1 - x_2))) \quad (1.2)$$

where

ℓ step length measured in meters,

x_1, y_1 = latitude and longitude at time 1,

y_2, x_2 = latitude and longitude at time 2

and r is the radius of the earth.

Then the movement of an individual can be represented by a sequence of turn angles and step lengths given its current position and its subsequent positions denoted by coordinates (x_i, y_i) , $i = 1, \dots, n$ computed using ArcView or ArcGis tools.

1.2.2 Drivers of animal movement patterns

In this section, we discuss the factors that drive elephant movement. The movement of animals depends on a variety of factors, among the food, reproduction water and home range. A nice introduction to the a variety of ecological factors and matching theories relating to animal movement can be found in [27, 45, 98]. The ability of an animal to forage and the resulting movement patterns depend on a large number of biological drivers. While some of them are related to environment, others are given by the internal constraints, such as its navigational ability, its energy need and storage capacity [45]. Nearly all animals live in a highly complex and heterogeneous environment. One common cause of environmental heterogeneity is non-uniform food resource distribution, [6, 24, 45, 98, 99, 114]. This is of concern to ecologists when optimal foraging strategies are investigated. Even seemingly monotonous environments such as the wildlife game parks have spatially heterogeneous food resources, in this case structured shrubs [98]. External spatially varying parameters, e.g., food availability, temperature or surface water distribution, affect the movement of foraging animals [131]. To relate the linear and circular metrics with these factors, two statistical approaches are of interest.

1.2.3 Probability distributions and their importance in animal movement

Theoretical and empirical ecology have developed over the past two decades including the classical foraging theory, behavioural theory and movement ecology paradigms, rested upon the assumption that step lengths of animal movement followed the power law distribution. However, it has long been known that step lengths of animal movement are heavy tailed and skewed a characteristic that cannot be addressed by the conventional symmetrical distributions. The heavy tailed or leptokurtic character of the distribution of step lengths has been repeatedly observed and demonstrated in several species [5, 34–37] of animals. Leptokurtotic behaviour is quantitatively measured by the kurtosis in excess of 3, a value obtained for the normal distribution. Though the power law distribution has been shown to fit the data, it misses important properties of the data such as skewness and scale because the third and fourth moments do not exist under the restriction $1 < \alpha < 3$ [19]. However, from numerous empirical studies, the power law assumption cannot be justified for several species [34–37].

Step lengths, which is the distance between two animal movement locations, is assumed to be a proxy measure of animal movement behaviour and pattern. It measures the cumulative outcomes of a vast number of pieces of information and individual decisions arriving almost continuously in time regarding animal movement [63]. As such, Viswanathan et al. [132], fitted a power law distribution to describe animal movement. The power law is a simpler variant of the stable distributions which only models the tails. The strongest statistical argument for it is based on the generalized central limit theorem, which states that the sum of a large number of i.i.d variables from a finite variance covariance distribution will tend to be stable distributed. However, as mentioned, animal movement step lengths are heavy tailed and skewed.

In response to Viswanathan et al. [132, 133] and subsequent studies, Kawai [63] proposed the use of a flexible four parametric family of stable distribution to model

the animal movement linear metrics. In many empirical studies such as finance, economics and energy, the non-Gaussian stable Paretian distribution of Mandelbrot [76] of which the power law distribution is special case have been found to be the most appropriate models for heavy tailed and skewed data. The probability of extreme events in ecology are so great such that many statistical techniques, which depend on asymptotic theory of finite variance for their validity, are not applicable [63]. Stable distributions are supported by the generalized central limit theorem, which states that stable laws are the only possible limit distributions for property normalized and centered sums of independent identically distributed random variables [94]. Application of stable distribution theory in ecology has biological significance. We note that it is difficult to determine analytically when and where a step of a foraging animal ends [63]. By the additivity property of the stable distribution, it is possible to accomodate such a worry in a modelling framework. The tails of step lengths contain important biological signals which can help understand animal pattern and behaviour. Such biological signals if left out in statistical assumptions, may lead to wrong conclusions of animal movement patterns and may have serious ecological implications [98].

Stable distributions have been mentioned in movement ecology by Bartumeus [5], though they have never been applied fully in the analysis of real animal movement data. One advantages of stable distribution in movement ecology is that they allow for continuous transition from diffusive (brownian) random walks when $\alpha = 2$, through super diffusion when $0 < \alpha < 2$, to ballistic (line) motion which occurs in the limit $\alpha \rightarrow 0$ which is associated with destructive foraging common in large mammals such as elephants. Defining search efficiency as the distance travelled, Viswanathan et al. [133], showed that Lévy walks ($1 < \mu \leq 3$) are more efficient than non-Lévy walks ($\mu > 3$) and the optimal Lévy exponent is approximately 2 where μ is the power law exponent James et al. [58]. Diffusive (i.e., Brownian) movement which is attained when $\alpha = 2$ in stable Paretian distribution involves much backtracking, which can be advantageous in keeping the forager in a food patch, but can also entail repeatedly searching an empty space, when not in a patch [6, 132]. Ballistic movement ($\mu \rightarrow 1$), which is attained when the stable parameter

$\alpha = 0$, avoids repeatedly searching the same space but is less suited to exploiting the patchy nature of the food environment [106]. Another reason advanced by [63] for the application of stable distribution is that the additivity property allows the coarse temporal resolution of recording animal tracking data to treat two or more consecutive steps as a single step without changing the distribution of the step lengths. Such a property is important when modelling movement data as it is difficult to define and hard to determine analytically when a step ends [8, 26, 34]. We note that the tails of animal movement data may contain important biological signals of animal movement that may be lost if the data is subjected to treatment. one thing that we discussed is the importance of the biological information in the tail, and that it is not often used. This new approach will indicate that there are important biological signals (information) at the tails that should be looked at in more detail in future studies.

Studies exploring alternative probability distributions are found in [34–37, 80]. These studies have compared the power law distribution assumption with the exponential, lognormal, weibull and gamma distributions. However, potential alternative distributions such as the student’s t, skewed student’s t, Laplace and the stable distributions remains an explored. Such models have potential for moving the movement ecology modelling forward. We note that the skewed student’s has the capability of handling the skewness and the heavy tail properties of animal movement linear metrics. In chapter 2, we give further details of the power law distribution and the stable distribution. In chapter 3 we compare the stable law regression with the heavy tailed t regression model.

1.2.4 Modelling autocorrelated heavy tailed data of animal movement

Despite several studies demonstrating that animal movement step length time series data are autocorrelated, heavy tailed and skewed, models for analyzing such data remains largely unexplored. Autocorrelation has been viewed as a major

problem in many telemetry studies because sequential observations are not independent in time and space and violate the assumptions for statistical inference [15, 22, 23, 26]. There is a lot to learn about the structure of ecological and behavioural data from the autocorrelation patterns of animal movement. These includes periodicity and patchiness in spatial data, characterized by autocorrelogram, semivariogram or spectrum [15].

To make valid biological interpretation and conclusions, the heavy tailed and skewed nature of the data must be taken into account. One such distribution that captures the properties of the tracking data is the four parameter stable distribution [65]. However, previous time series models of animal movement have relied on the Gaussian distribution assumption [52] and fitted the autoregressive models of order one (AR(1)). This model does not capture the heavy tails and skewness properties of the animal linear metrics (step length). Models based on stable distribution, student's t and skewed student's t distribution have been proposed in the financial literature as best alternative[21, 87]. The application of time series models has been hampered by lack of computational tools in the mainstream statistical softwares. However, with the ever improving technology and computational capacity, statistical methods that measure the level of autocorrelation in the data may be particularly useful for comparing patterns of animal behaviour and range use among individuals so long as the same sampling interval is used [15]. Dai et al. [23] investigated short-day time duration of African elephant movement in Pangola game reserve and reported autocorrelation at lags of 10 and 15 minutes and no autocorrelation at 20minutes interval.

1.2.5 Regression Models for heavy tailed linear metrics of animal movement with covariates

Exploring a techniques that investigates the relation between the linear metric step length with the environmental covariates of the data is a step towards diversifying the understanding of animal movement [91]. The most commonly used linear

regression models suffer from lack of robustness against departure from distributional assumptions as other statistical models based on the Gaussian distribution and may be too restrictive to provide an accurate representation of the structure of the data. To overcome this deficiency, a stable distribution assumption was proposed by [143]. The non-normal stable distributions all have heavier tails than the Normal distribution and allow skewness. Heavy tails and skewness implies that extreme observations are given a greater probability of occurring and are thus given less weighting in maximum likelihood estimation so that fitted lines are not biased towards these extreme observations [83, 138]. It is therefore a natural extension of normal regression models to assume stable distributions as the distribution of the error terms.

Stable distributions have found wide applications in other areas of research such as modelling financial, biological and geological problems [127]. The assumption that step lengths of animal movement follow a stable Paretian distribution has far reaching implications for both foraging and statistical theory [63]. For example, the problem of Lévy search patterns has been addressed in [34–37], robustness to the sampling frequency rate is studied in [65] and for a specific discussion of movement ecology issues see, [63]. However, in all these studies, the link between step lengths and the environmental covariates remains unexplored.

Kawai [63] and Kawai and Petrovskii [65] showed that lack of closed form density for stable distribution for all but a few special cases, stable distribution has received limited attention in movement ecology. Moreover, the direct numerical integration techniques are non-trivial and burdensome from a computational perspective. As a consequence, maximum likelihood estimation algorithms based on such approximations are difficult to implement especially for huge data encountered in movement ecology [63]. However, with increasing computational power and efficient algorithms, maximum likelihood estimation and other alternative techniques have been implemented by Nolan and Ojeda[95]. Due to the above mentioned drawbacks, stable distributions have not been exploited as a tool for modelling animal movement data. Other alternative models for heavy tailed data assume either a student's t, skewed student's t and skewed normal distribution errors of

a regression model. We note that despite the computational challenges and lack of rich data with covariates, the regression model with t distributed error terms is more computationally feasible than the stable law regression model. However, based on model selection tools, the stable law regression provides the best fit to the elephant movement data.

Fonseca et al. [44] demonstrates that the regression model with Student's t errors also suffers from monotone likelihood. This is because the likelihood function is ill-behaved for v close to zero and may be ill-behaved when $v \rightarrow \infty$. Lange et al. [72] generalized the traditional regression model with normal distributed errors to more robust regression models with t distributed errors. It is well known that the t distribution provides a convenient description for regression analysis when the residual term has a density with heavy tails. From Kotz and Nadarajah [67], the classical linear model can be modelled as follows.

We have identified several key areas to be pursued in future studies. Some of them are straightforward, such as increasing the number of covariates and allowing stable distribution parameters to be a function of the covariates. Diagnostic testing and model checking tools is another area which needs to be developed in future studies to check the adequacy of the fitted models. Further implementations of the stable regression and allied alternative models such as skewed normal and skewed t regression will provide researchers with a robust statistical framework for analysing environmental factors driving the movement patterns. An implementation of the student's t and skewed student's regression model is implemented in the *'heavy'* R statistical package. Building such a model that captures the skewness and heavy tails is a necessary step in being able to elucidate the links between the pattern and the environment of free-ranging Mega-herbivores.

1.2.6 Modelling Circular responses with linear covariates

Unlike the linear statistics, circular or angular data which is measured in degrees or radians. Angular data arise in many scientific disciplines, for instance, in

oceanography (wave directions), meteorology (wind direction) and biology (study of animal movement)[7]. Handling such data creates difficulties due to the restriction of support to the unit circle, $[0, 2\pi)$, and to the sensitivity of descriptive and inferential results to the starting point on the circle. Hence, the analysis of angular data is more challenging than the linear data. The circular nature of such data prevents the use of commonly used statistical techniques, as these would provide wrong or misleading results. A substantial literature on circular data exists with studies by [4, 7, 17, 43, 59, 78, 142], but broadly, it is confined to descriptive statistics and limited to inference for simple univariate models. Application of circular statistics is limited to descriptive analysis.

There are a number of probability distributions designed for use with angular data. These include circular normal distributions, wrapped normal distribution, wrapped Cauchy distribution, and various bimodal distribution [7, 50, 51]. Because of the difficulties inherent in the use of normal distributions, these (circular) distribution should be used when theoretical distributions are invoked to characterize angular data or to simulate movement paths. Circular normal distributions have been found to provide good statistical fits to insect turning angles and clonal plant branching angles [16]. Despite the fact that circular statistics is still in a very active stage of development, several monographs and textbook lay a standard foundation of the methods have been published [7, 43, 59, 142]. In two dimensions, this means that we allow the animal to move in any direction θ on the unit circle, where $-\pi < \theta \leq \pi$ correspond to the same direction. Such models generate data and statistics on the movement direction of the animal. So, the resulting summary statistics depend strongly on the point where the circle is cut.

Cain [17] reviewed the statistical methods for analyzing angular ecological data and found that standard statistical methods were not appropriate for analysis of circular data. Fisher et al. [43] provide a general introduction and methodology for dealing with statistics of circular and spherical data respectively, while [7] studies biological problems using circular statistics, and [78] provides a large amount of theoretical background and models for use with directional data. Many studies in movement ecology rely on circular distributions to develop complex models. For

instant [7, 119, 124] fitted von Mises random variables to turn angle data and come up with non-linear model of relating animal movement to object orientation. Previous studies consider turn angles to be a by-products of complex behavioural processes related to orientation strategies of animals [9, 14, 90, 126]. For instance, elephants use direct movement strategy when needing to get to a destination more quickly (e.g. towards mates), rather than significantly increasing their speed [27]. Statistical analysis of angular distributions can help determine whether or not an orientation component exists [8, 9], estimate quantitatively such components [20] and characterize the orientation component as local or global [9].

One of the approaches for parameter estimation is the Maximum likelihood estimation approach. The iterated re-weighted least square algorithm of Green [49] is used in a maximum likelihood method to estimate the parameters of the circular linear regression model. We note that though this approach is simple for linear regression models, it is computationally challenging for huge data set. Other approaches suggested in the literature are the Fisher scoring algorithm [124] and Newton-Raphson algorithm for circular regression models. We only look at the IRLS algorithm in this study. Further discussion of this topic is in chapter 6.

1.2.7 Testing circular hypothesis of uniformity

There are several methods for testing the uniformity hypothesis of circular data. Statistical methods emphasizing on accurate choice of assumptions play an important role in understanding animal movement data. A first step towards understanding animal movement orientation is the exploratory analysis such as display of data on a rose diagram or linear histogram [43].

Four commonly used methods in testing circular uniformity of angular data are explored in chapter 6 and applied to the elephant movement data. These methods are: Rao's spacing test, Kuiper's test, Rayleigh's test and Watson's test. Among the four, we note that Rao's method is more prone to rejecting the null hypothesis than the Kuiper's test and the Rayleigh's test in the event of a small sample size,

unless the empirical distribution of the data is appreciably uniform in atleast some of its sub- domains [109]. Put in otherwise, Rao's test carries a similar Type 1 error than the other test statistic; however, it follows from the nature of statistical data that the rejection of the null hypothesis cannot be absolutely certain.

1.3 Research Objectives

The general aim of this thesis was to investigate the various statistical methods of analyzing animal movement data sets derived from the GPS telemetry studies using applications example to elephant movement data and determine appropriate statistical tools from among the many. The study also aimed to deal with environmental drivers of animal movement and the effects of landscape features on animal movement. The specific objectives are:

- To study the statistical probability models for describing heavy tailed and skewed data sets with a particular focus on the stable distribution family.
- To investigate the potential of stable law assumption in a regression set-up to analyze animal movement data sets with covariates.
- To investigate time series models for analyzing heavy tailed and skewed data sets and compare them.
- To explore the circular linear regression models for analyzing animal movement data sets with covariates.
- To test the uniformity hypothesis of animal movement data sets.

1.4 Thesis Outline

This thesis is organized as a collection of five papers which have been presented for peer review in international journals. Each chapter has been written as a stand-alone article that can be read separately from the rest of the thesis but draws

separate conclusions that link to the overall research objectives. As a result a number of replications and overlaps occur in some sections “Stable distribution” and “application to real data” in different chapters. This problem is negligible when one considers the critical peer review process and the fact that different chapters can be read separately without losing the overall aim of the thesis. In chapter 2, 3 and 4, the stable distribution is central, while in chapter 5 and 6, the circular statistic is the main focus.

Chapter 1: This chapter overviews some past and recent developments in the theory of animal movement and outlines some points that have not been discussed in the literature. The missing gaps in the statistical analysis are explained in this chapter and constitute the main source of motivation in the writing of this thesis.

Chapters 2 This chapter focuses on the heavy tailed and skewed characteristic of animal movement data sets. The power law distribution which is commonly used in the literature is compared with the four parametric family of stable distributions. The method of moments, Kogon-Williams algorithm and maximum likelihood methods of parameter estimation are reviewed. An application study including the description of the data set used in the analysis is presented. The results of the different methods are compared. Finally the chapter concludes with a discussion of the results and future extensions in movement ecology.

Chapter 3: In this chapter, a statistical assumption based on the stable probability distribution described in chapter 2 is used in a regression model framework and presented. An alternative regression model for heavy tailed data based on student t distribution is also proposed and contrasted with the stable regression model. The method of maximum likelihood is used in the parameter estimation. An application study using data sets from five elephant herds is described and the results presented. Finally the chapter ends with a discussion of the findings.

Chapter 4: This chapter breaks from the conventional Gaussian assumption and presents a sets of statistical models for animal movement in a GARCH modelling framework. The Normal, students t, skewed student’s t and the stable Paretian ARMA-GARCH models are considered and evaluated. The chapter also presents

an application and description of the five elephant movement data sets used in the analysis. The statistical properties of the data and the empirical results of the are given. Lastly the chapter ends with a discussion of the findings and recommendations for future development.

Chapter 5: This chapter deals with modelling circular data. A circular linear regression model is presented. An application and description data set of animal movement used in the analysis is presented. A simulation study to determine the boundaries of the initial values of the model is also presented and discussed. The results of elephant movement data is presented and discussed in the of elephant ecology. Finally we highlight various limitations of modelling circular linear relationships using ordinary linear and nonlinear regression models.

Chapter 6: In this chapter, we consider the uniformity hypothesis in the animal movement turn angles data sets. A set of hypothesis tools is reviewed and used in a four stage process to evaluate the uniformity hypothesis. An illustration data set of elephant movement is described. The results of the analysis are presented and their biological implications in ecology discussed.

Finally, this chapter gives a synthesis of the study. The findings are summarized and conclusions are derived from the preceding chapters. For future research on the modelling of animal movement, relevant recommendations are made. Special focus is directed toward the operational use of the mixed models in explaining herd to herd effects. A single reference list is provided at the end of the thesis.

1.5 Contribution of the present study

Two statistical methods were selected for application to animal movement data obtained using the GPS-telemetry tracking collars. These methods were selected in other research with data of similar statistical properties. Throughout the work conducted in this study, the desired statistical outcome was to identify appropriate

statistical methods for analyzing heavy tailed and skewed animal movement metrics and relates the metric's to the environmental variables. The choice of model can have biological implications on the interpretation of the output and subsequent inferences of animal movement data. These may in turn affect decision made regarding conservation of the ecosystem. Each of the models investigated are based on slightly different assumptions and vary in complexity. From the biological perspective, it is important to find statistical distributions and models which do not only provide a theoretical fit to the data, but can also provide easily interpretable results which are realistic in terms of biology of the species under investigation. These output and inferences are then intended to be integrated with other ecological studies which make use of the movement data to answer much wider ecological questions.

The motivation for this study comes from the ability to collect movement data that has been outpaced by the ability to statistically analyze such data. GPS-telemetry data are expensive to collect and basic analysis based on conventional methods are inadequate. The link between animal movement linear and circular metrics with environmental covariates remains largely unexplored due to lack of appropriate statistical approach. The output from these analysis are needed to link more complex ecological studies of the animal tracking are opening up new research areas of study into behavioural ecology of the tracked animals. The need to better understand animal movement process and its relationship with the environmental drivers is acknowledged particularly in the light of climate changes and issues such as disease outbreak and biological invasions [45]. Statistical modelling approaches need to be advanced and adapted in order to cater for these developing areas of science. Kawai and Petrovskii [65] identified one challenge is that animal movement is never scale free as asserted in several species [23, 34, 34, 35, 37, 117]. Bartumeus [5] noted that animal movement is heavy tailed and skewed, properties which are not adequately covered using the power law distribution. This models used in the literature assume the heavy tailed and skewed time series data of animal movement linear metrics is Gaussian distributed and fit the stationary

time series models to investigate animal movement behaviour [52]. Such an assumption is inappropriate for animal movement step lengths time series as they have been shown to be heavy tailed and skewed. Another challenge highlighted by Nathan [91] is to integrate the animal movement data and analysis within statistical framework that includes environmental covariates, remote sensing and GIS to enhance the analysis. All these four challenges require statistical analysis and this study focuses on techniques and approaches of improving the analysis of animal movement GPS-telemetry derived metrics to achieve the outcomes of the four challenges. Stable law approach has been suggested as the likely method to gain popularity as the sampling frequency of movement tracking increases and the ability to collect landscape data improves with technological advances and computational capacity.

This study can be considered a first kind of its own, since a number of studies have been published using a wide variety of the techniques. Each of the methods used in this study have not been used in movement studies in its current form. However, a few studies have applied the circular linear models to movement data with only one covariate. This study adopts the stable distribution and customizes it in order to make it more suitable for this specific application.

CHAPTER 2

Stable Distributions: Theory and Application to Animal movement data

Abstract

Statistical distributions used in movement ecology applications attempt to capture the properties exhibited by the random variables describing the movement behaviour of the animal in heterogeneous environment. In a standard parametric approach, the underlying data are modelled using assumed probability distributions. The model parameters are fitted to the observed data using empirical or maximum likelihood methods. The model is then used to make decisions on sampled data for any potentially heavy tailed and skewed events. In practical applications, large data sets obtained from GIS telemetry studies of animal tracking exhibit heavier tails than power law distributions and have non-zero skewness. The tail behaviour of such a distribution offer valuable information for decision support and conservation management tools used by wildlife ecologists. This paper introduces the application of stable distributions to a data set of animal step lengths acquired from a set of individual elephants tracked in Kruger National Park in South Africa. Stable distributions are a family of probability distribution's that generalize the central limit theorem and can easily accommodate heavy tails and skewness. The results show that stable distributions describe tail steps of real life

animal movement data and can be used as an alerting tool in wildlife management.

Keywords: Heavy-tailed distribution; Stable distribution; Parameter estimation; GPS-telemetry; animal movement.

2.1 Introduction

Statistical analysis of movement data is key to understanding animal foraging patterns and movement behaviour. Understanding how animals move as an indicator of their responses to the environmental conditions they experience, and is important in deciding how we should manage our ecosystem, or how best to use the natural resources of our environment more effectively. Technological advances in GPS tracking devices are revealing new insights regarding animal movements on the landscape [107, 130]. GPS tracking devices tagged to animals are becoming smaller in size and larger in memory capacity yielding huge data [45, 130, 139]. Given the time, effort and monetary expenses devoted to obtaining data from individuals in the game park, it is reasonable to analyse the data with valid statistical methods. This is so because conclusions concerning animal movement may have management implications [27, 99]. Good conservation policies require accurate statistical analysis and meaningful biological interpretation of animal movement patterns and distribution [91].

The movement of an individual animal on a landscape can be represented by a sequence of successive positions (u_i, v_i) collected at time t_i where $i = 1, 2, \dots, n$ is the number of steps. These positions are characterized by short clustered steps and long rare steps as animals alternates between foraging or resting and searching for scarce food items or migration [45]. Such a path can be represented by a broken line with nodes indicating the position of the animal [63] and the distance between two points defined as the step length of the animal after an observation time interval [45, 53, 62]. Many popular models of animal movement have been developed under the assumption that step lengths are power law distributed. However, from numerous empirical studies , the power law assumption cannot be justified for

several species [34–37]. The step lengths of several species of animals are typically heavy tailed and skewed [5]. In other words, the probability that extreme events can happen is larger than the power law distribution can explain [5]. There may be important biologically relevant information in these extreme movement events which may influence managers’ conclusions.

In many empirical studies such as finance, economics and energy, the non-Gaussian stable Paretian distribution of Mandelbrot [76] of which the power law distribution is special case have been found to be the most appropriate models for heavy tailed and skewed data sets. The probability of extreme events in finance are so great such that many statistical techniques, which depend on asymptotic theory of finite variance for their validity, are not applicable [38, 39]. Movement of animals is the cumulative outcome of vast pieces of information and individual decisions arriving almost continuously in time [45] which typically lead to skewed, peaked and heavy tailed step lengths [63] hence it is natural distribution for modelling such a data is the four parameter family of stable distributions [76]. The power law distribution is by far the most well known and analytically tractable variant of the stable distribution and for these and other practical reasons it has been routinely postulated to govern the animal movement step lengths. However, animal step lengths are usually more leptokurtic i.e., have much heavier tails. This leads to considering the non-gaussian stable distribution $S(\alpha, \beta, \gamma, \mu)$ first described by Mandelbrot [76], where α is the index of stability which governs the movement steps of the animal, β is the skewness parameter which governs the infrequent of large rare steps, γ is a measure of scale as the movement step lengths are not scale free, and μ is the measure of location.

Stable distributions have been mentioned in movement ecology by Bartumeus [5], though they have never been applied fully in the analysis of real animal movement data. One advantages of stable distribution in movement ecology is that they allow for continuous transition from diffusive (brownian) random walks when $\alpha = 2$, through supper diffusion when $0 < \alpha < 2$, to ballistic (straight line) motion which occurs in the limit $\alpha \rightarrow 0$ which is associated with destructive foraging common in large mammals such as elephants. Defining search efficiency as the

distance travelled, Viswanathan et al. [133], showed that Lévy walks ($1 < \mu \leq 3$) are more efficient than non-Lévy walks ($\mu > 3$) and the optimal Lévy exponent is approximately 2 where μ is the power law exponent James et al. [58]. Diffusive (i.e., Brownian) movement which is attained when $\alpha = 2$ in stable Paretian distribution involves much backtracking, which can be advantageous in keeping the forager in a food patch, but can also entail repeatedly searching an empty space, when not in a patch [6, 132]. Ballistic movement ($\mu \rightarrow 1$), which is attained when the stable parameter $\alpha = 0$, avoids repeatedly searching the same space but is less suited to exploiting the patchy nature of the food environment [106]. Another reason advanced by [63] for the application of stable distribution is that the additivity property allows the coarse temporal resolution of recording animal tracking data to treat two or more consecutive steps as a single step without changing the distribution of the step lengths. Such a property is important when modelling movement data as it is difficult to define and hard to determine analytically when a step ends [8, 26, 34]. In this paper, we focus on the basic properties and the theory of stable distributions and show how they can be applied in movement ecology.

The rest of this paper is organized as follows. In the next section, we give a brief introduction to the α -stable distributions and the concepts related to stable distributions which will be necessary for the continuation of this paper. In section 2.4 we discuss methods of parameter estimation and statistical inference. In Section 2.5 we present the analysis of elephant (*Loxodonta Africana*) movement data. We fit and compare the power law distribution and α -stable distributions. The power law distribution can be regarded as a restricted version of the Lévy stable distribution and these restrictions can be tested by varying the values of α [63]. Finally, we discuss how our findings may help researchers working on individual animal to reveal the pattern and identify the movement type. We note that the tails of animal movement data may contain important biological signals of animal movement that may be lost if the data is subjected to treatment. one thing that we discussed is the importance of the biological information in the tail, and that it is not often used. This new approach will indicate that there are important

biological signals (information) at the tails that should be looked at in more detail in future studies.

2.2 Stable distribution

Stable Paretian distributions are a class of probability laws with both interesting theoretical and practical properties. They are appealing for ecological modelling of movement data since they generalize the Gaussian distribution and allow heavy tails and skewness, properties which are common in animal movement data. In this section, we present basic definitions of stable distributions.

Definition 2.1. A random variable X is said to have a stable distribution if for any $n \geq 2$, there is a positive number c_n and d_n real number such that

$$X_1 + X_2 + \dots + X_n \stackrel{d}{=} c_n X + d_n \quad (2.1)$$

where X_1, X_2, \dots, X_n are independent identically distributed copies of X . The symbol $\stackrel{d}{=}$ means equality in distribution. The stable law is strictly stable if $d_n = 0$ for all n .

For ecological modelling, this key characteristic implies that for coarse temporal resolution two or more steps can be combined without changing the distribution of step lengths. This property is of biological importance since it is difficult to determine analytically when the step begin and when the steps end [63].

The class of all laws that satisfy (2.1) is called stable and is described by four parameters: α index of stability that determines the tail weights or distributions of kurtosis with $0 < \alpha \leq 2$, skewness parameter β which determines the distribution's skewness and is in the range -1 to 1 , scale parameter γ and location parameter μ which measures variability and mean as in the case of Gaussian distributions respectively [93, 138].

Following [93], let $S(\alpha, \beta, \gamma, \mu)$ denote a stable distribution with tail exponent $\alpha \in (0, 2]$, skewness $\beta \in [-1, 1]$, scale $\gamma > 0$ and location parameter $\mu \in \mathbb{R}$, given by the characteristic function

$$E[\exp(itX)] = \begin{cases} \exp\left(-(\gamma|t|)^\alpha \left[1 + i\beta(\text{sign } t) \tan \frac{\pi\alpha}{2} (|t|^{1-\alpha} - 1)\right] + i\mu t\right) & \alpha \neq 1 \\ \exp(-\gamma|t| \left[1 + i\beta(\text{sign } t) \frac{2}{\pi} \ln |t|\right] + i\mu t) & \alpha = 1 \end{cases} \quad (2.2)$$

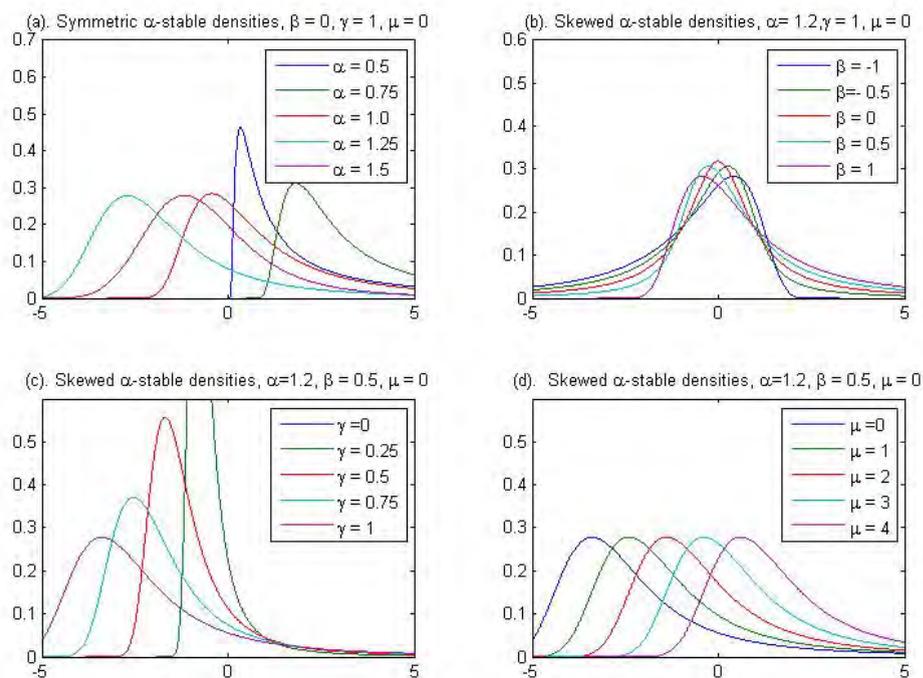
where $i = \sqrt{-1}$, $\text{sign}(t) = -1, 0, \text{ or } 1$ if $t <, =, > 1$, respectively. Parameters α and β determines the shape of the distribution. If the index of stability $\alpha = 2$, then the stable distribution reduces to the normal distribution. That implies that, relative to the normal distribution, a higher probability of large extreme rare steps exists in animal movement when $\alpha < 2$. The impact of α for values of α less than 2 on the density of the distribution is twofold. First, it has an effect on the tail thickness of the density. Second, it has an effect on the peakedness of the origin relative to the normal distribution [138]. Jointly, these two effects are known as ‘‘leptokurtosis’’, the index of stability α can be interpreted as a measure of leptokurtosis. As the value of α becomes smaller, the more leptokurtic the distribution. Thus, for $\alpha < 2$, stable distributions are more peaked around the center than the normal and have fatter tails (see Figure 2.1 a). In fact, for $\alpha < 2$ they are so heavy that the variance is infinite; and for $\alpha \leq 1$, even the first moment does not exist [110].

For $\beta = 0$, the distribution is symmetric around the location parameter μ . If $\beta > 0$, the distribution is skewed to the right and if $\beta < 0$, it is skewed to the left. Larger values of β indicate greater skewness. A symmetric stable distributions is a stable distribution with $\beta = 0$ and $\mu = 0$ and the stable distribution is symmetric around μ if $\mu = 0$. In movement ecology, the distribution of animal movement step lengths is never symmetric [5]. The scale parameter generalizes the definition of standard deviation and can interpret it as volatility clustering in animal movement. It allows any stable random variable X to be expressed as $X = \gamma X_0$, where the distribution of X_0 has a unit scale parameter and the same α and β as X . Figure (2.1 a) shows the effect of α on the kurtosis when $\beta = 0$,

$\mu = 0$ and $\gamma = 1$. Figure (2.1 b) illustrates the effect of β on the skewness of the density function for $\alpha = 1.2$, $\mu = 0$ and $\gamma = 1$.

There are three special cases of stable distributions: the distribution reduces to a Gaussian distribution when $\alpha = 2$ and $\beta = 0$, a Cauchy distribution when $\alpha = 1, \beta = 0$ and a Levy distribution when $\alpha = 0.5, \beta = 1$. Figure (2.1 (a)) illustrates the α -stable distributions given by various parameter values.

FIGURE 2.1: Effects of α, β, γ and μ on the stable density function (matlab image)



Capturing rare events of animal movement step lengths, such as skewness and heavy tails, requires the specification of appropriate distributions or models. Clearly, the stable distribution with $\beta > 0$ and $\alpha < 2$ is a natural candidate. Increasing $\beta \rightarrow 1$ results in skewness to the right and lower values of α lead to stronger leptokurtosis. When $\alpha > 1$, the location parameter measures the mean of the distribution. In empirical ecology α usually takes values in the interval $(1, 2)$. This implies the assumption that the step lengths modelled with stable Paretian distribution exhibits finite mean but infinite variances. This is because when $1 < \alpha < 2$ the tails of the distribution taper off too slowly for the variance to be finite and

only the mean exists. Empirical evidence suggests that animal movement step lengths data sets are heavy tailed meaning that it does not have finite variance [116].

2.3 Biological motivations for the stable distribution

We provide only relatively short overview of the essential properties of stable distribution and there motivating application in movement ecology. Models that characterize animal movement behaviour describe part or some of its characteristics. Several of the previous studies in movement ecology have advanced the theory that animal movement generates heavy tailed which is modelled using the power law distribution. One aspect of the data ignored by ecologists is the skewness. This parameter is difficult to obtain as the second moment of the stable distribution and the power law distribution does not exist. However, a measure of the skewness parameter of the stable distribution. The flexibility of the stable distribution, can be is demonstrated by the varying the parameters α and β . We note that when $\alpha = 2$, we recover the normal distribution from the stable distribution and all moments exists. Such a property is advantageous in ecology as Brownian motion is described by the Gaussian assumption [63]. The heavy tails of the stable laws follow the power law decay is characterized by

$$1 - F(x) = P(X > x) \sim \gamma^\alpha C_\alpha (1 + \beta)x^{-\alpha} \quad (2.3)$$

$$f(x|\alpha, \beta, \gamma, \delta; 0) \sim \alpha \gamma^\alpha C_\alpha (1 + \beta)x^{-(\alpha+1)}$$

where $C_\alpha = \sin \frac{\pi\alpha}{2} \Gamma(\alpha) / \pi$. When $\beta = 1$, the left tail decays faster than any power which can be taken to mean that the is making long rare steps or the animal is a mode of searching scarce food resources. The right tail behaviour when $\beta = -1$ is similar by the reflection property.

The difference between the Gaussian and the non-Gaussian stable distribution as $x \rightarrow \infty$ can be seen by expressing the tail behaviour as

$$P(X > x) \sim \frac{1}{2\sqrt{\pi\gamma x}} \exp\left(\frac{-x^2}{4\gamma^2}\right) \quad (2.4)$$

which is different from the asymptotic behaviour of the tails of stable distribution that follow a power law decay with exponent α stated as

$$P(X > x) \sim cx^{-\alpha}, x \rightarrow \infty \quad (2.5)$$

One important consequence of power law decay of the tails is that only certain moments exist. For a stable random variable X , the p^{th} absolute moment is given by

$$E|X|^p = \int_0^\infty P(|X|^p > x) dx \quad (2.6)$$

exist if and only if $p < \alpha$ or $\alpha = 2$, that is, when the integral given by (2.6) converges [110]. If the tails are too heavy, the integral will diverge. Thus the second moment of any non-Gaussian stable distribution is infinite [110]. One obvious consequence in ecological applications is that, since stable distributions have infinite variances, one cannot describe movement behaviour of animals in terms of variance and correlation of step length dependence on one another. Alternative measures namely covariation function, co-difference function and the dynamic function have been proposed in the literature [108, 110].

The third important property of stable laws is the stability or additivity property, which allows us to use stable laws in movement ecology to analyse animal movement data. Following (2.1), if X_1 and X_2 are independent and identically distributed stable random variables, then for any given positive numbers a and b , there exist a positive number c and real number d , such that

$$aX_1 + bX_2 \stackrel{d}{=} cX + d$$

For $d = 0$, we obtain the strictly stable distribution of X . This statistical property is suitable for describing telemetry data of animal movement as the temporal

resolution of collecting tracking data allows combining two or more steps without changing the distribution of step lengths [63].

2.4 Parameter estimation

The lack of known closed-form density functions also complicates statistical inference for stable distributions [30]. For instance, maximum likelihood (ML) estimates have to be based on numerical approximations or direct numerical integration formulas of [92]. Consequently, ML estimation is difficult to implement and time consuming for large samples encountered in movement ecology. There are two basic strategies to estimation of distribution parameters of heavy tailed and skewed data. The first solely focuses on the estimation of the tail exponent and the second on the estimation of the four parameters of the stable distribution. In the second case we will focus on a three-stage approach that can be used to estimate the four parameters of animal movement GPS-derived telemetry data sets in a three stage process: i) the first stage, we will use the quantile method of McCulloch to estimate the initial parameters of the stable distribution; ii) in the second stage we use the estimates from stage 1 as initial values in stage two and iii) the refined estimates of stage two as initial estimates of the final stage which we discuss further below.

2.4.1 Asymptotic properties of the power law distribution

The asymptotic tail properties of the stable distributions have been applied in to study the step lengths of several animal species in ecology [34–37, 58]. The limiting tail behavior of the stable distribution are referred to as the Pareto random walk [63, 65] due to the power law decay of the stability index α given by the probability density of the form

$$p(x) = \frac{\alpha x_{min}^\alpha}{x^{\alpha+1}} \quad \text{for all } x \in (x_{min}, \infty), \quad (2.7)$$

where $\alpha > 0$ and $x_{min} > 0$ and the tail probability given by

$$P(|X_1| > x) = (x_{min}/x)^\alpha, \quad x \in (x_{min}, \infty) \quad (2.8)$$

where x_{min} indicates the minimum step length possible, while α is the tail exponent (usually written with μ related to the stable distribution through $\alpha = \mu - 1$). In applications to animal movement, this pareto jump is usually referred to as a “power law distribution” (for instant of step size x), and the resultant random walk is called a “Lévy flight” and the movement pattern “Lévy movement pattern” [6, 132]. A serious drawback of pareto distribution is attributed to its infinite variance [19] for some specified values of α .

The ML estimator of the tail index is obtain by maximizing the log-likelihood equation

$$ln(\alpha, x_{min}) = n \ln \alpha + n\alpha \ln x_{min} - (\alpha + 1) \sum_{i=1}^n \ln x_i \quad (2.9)$$

To find the estimator for α , we compute the corresponding partial derivative and determine where it is zero:

$$\frac{\partial \ln}{\partial \alpha} = \frac{n}{\alpha} + n \ln x_{min} - \sum_{i=1}^n \ln x_i \quad (2.10)$$

Thus the maximum likelihood estimator for α is

$$\hat{\alpha} = \frac{1}{\sum_i (\ln x_i - \ln \hat{x}_{min})} \quad (2.11)$$

with standard error given by

$$\sigma = \sqrt{n+1} \left(\sum_i \ln(x_i/x_{min}) \right)^{-1} \quad (2.12)$$

The maximum likelihood estimator is strongly consistent and asymptotically normal [19]. Asymptotic normality of the ML estimator is given by

$$\sqrt{nI(\alpha_0)}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N(0, \alpha^2/n)$$

2.4.2 Quantile estimation methods

One of the earliest methods of estimating the stable distribution parameters is the Quantile method of Fama and Roll [40]. This technique, estimates the tail index α and γ when the $\beta = 0$ and $\mu = 0$ are fixed to zero. McCulloch [81] generalized Fama and Roll's method and provided consistent estimators of all the four stable parameters with the restriction that $\alpha \geq 0.06$. Let $X \sim S(\alpha, \beta, \gamma, \mu)$ and denote the p -th quantile of this distribution by X_p . McCulloch's estimator used five quantiles to estimate $\alpha \in [0.6, 2]$ and $\beta \in [-1, 1]$, and is given by

$$v_\alpha = \frac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}} \quad \text{and} \quad v_\beta = \frac{(x_{0.95} - x_{0.5}) - (x_{0.5} - x_{0.05})}{x_{0.95} - x_{0.5}} \quad (2.13)$$

so that $S(\alpha, \beta, \mu, \gamma)(x_p) = P[X < x_p] = p$, v_α is a measure of the relative size of the tails and the middle of the distribution and v_β is a measure of the spread between the right tail and the left tail of the distribution. A large value of v_α means that the tails will be fatter and a small value means the tails will be thinner or smaller. This relationship may be inverted and the parameters α and β may be viewed as functions of v_α and v_β

$$\alpha = \psi_1(v_\alpha, v_\beta) \quad \text{and} \quad \beta = \psi_2(v_\alpha, v_\beta) \quad (2.14)$$

Substituting v_α and v_β by their sample values and applying linear interpolation between values found in tables given in, McCulloch [81], yields estimators $\hat{\alpha}$ and $\hat{\beta}$. Scale and location parameters, γ and μ , can be estimated in a similar way. However, due to the discontinuity of the characteristic function, for $\alpha = 1$ and $\beta \neq 0$ in representation (2.2), this procedure is much more complicated than the Koutrouvelis algorithm [68] described below. Despite the complexities mentioned above, this method is the fastest way to estimate the parameters of the stable distribution, since it avoids optimization. However, theoretical properties remain unclear and the extension to the case of linear combinations of stable random variables is not possible.

2.4.3 Koutrouvelis algorithm and Regression-type technique

A second method of estimating parameters of stable distribution we consider in this study is the Koutrouvelis algorithm and regression type approach. Given a sample x_1, x_2, \dots, x_n of n i.i.d random variables, the characteristic function of the stable distribution is defined by $\hat{\Phi}(t) = \frac{1}{n} \sum_{j=1}^n \exp(itx_j)$ [102]. All the moments of $\hat{\Phi}(t)$ are finite since $|\hat{\Phi}(t)|$ is bounded to unity for any fixed t , it sample average of i.i.d random variables $\exp(itx)$. Hence, by the law of large numbers, $\hat{\Phi}(t)$ is a consistent estimator of the cf $\phi(t)$.

Press [102] was the first to use the sample characteristic function in the context of statistical inference for stable laws. He noted that one of the limitations of this method is that the convergence to the population values depends on the choice of the four estimation points, whose selection is rather problematic.

Koutrouvelis [68] extended the characteristic function method of Press [102] to a more accurate regression type approach. This method, starts with an initial estimate of the parameters and proceeds iteratively until convergence criterion is satisfied. Each iteration consists of two weighted regression runs. The number of points to be used in this regression method depends on the sample size and starting values α . Typically no more than two or three iterations are needed [71]. The speed of convergence however depends on the initial estimates and the criterion used [69]. The regression method is based on the following observations concerning the characteristic function $\phi(t)$. First from (2.2), we can easily derive

$$\log(-\log|\phi(t)^2|) = \log(2\gamma^\alpha) + \alpha \log(t) \quad (2.15)$$

The real and imaginary parts of $\phi(t)$ are for $\alpha \neq 1$ given by

$$\Re\{\phi(t)\} = \exp(-|\gamma t|^\alpha) \cos [\mu t + |\gamma t|^\alpha \beta \operatorname{sign} t \tan \frac{\pi\alpha}{2}] \quad (2.16)$$

$$\Im\{\phi(t)\} = \exp(-|\gamma t|^\alpha) \sin [\mu t + |\gamma t|^\alpha \beta \operatorname{sign} t \tan \frac{\pi\alpha}{2}] \quad (2.17)$$

Apart from considerations of principal values, equations (2.16) -(2.17) lead to

$$\arctan \left(\frac{\Im\{\phi(t)\}}{\Re\{\phi(t)\}} \right) = \mu t + \beta \gamma^\alpha \tan \frac{\pi\alpha}{2} \text{sign}(t)|t|^\alpha. \quad (2.18)$$

Equation (2.15) depends only on α and γ and suggests that we can estimate the two parameters by regressing $y = \log(-\log|\phi(t)^2|)$ on $w = \log|t|$ in the model $y_k = m + \alpha w_k + \epsilon_k$, where t_k is an appropriate set of real numbers, $m = \log(2\gamma^\alpha)$, and ϵ_k denotes an error term. Koutrouvelis [68] proposed to use $t_k = \frac{\pi k}{25}$, $k = 1, 2, \dots, K$; where K lies between 9 and 134 for different values of α and sample sizes.

Once $\hat{\alpha}$ and $\hat{\gamma}$ have been obtained and α and γ have been fixed at these values, estimates of β and μ can be obtained using (2.18). Next the regressions are repeated with $\hat{\alpha}, \hat{\gamma}, \hat{\beta}$ and $\hat{\mu}$ as the initial parameters. The iterations continue until a pre-specified convergence criterion is satisfied.

2.4.4 Maximum likelihood estimation

Let $X = (x_1, x_2, \dots, x_n)$ be a vector of n independent and identically distributed (i.i.d) stable Paretian random variables, i.e., $X_i \sim S(\alpha, \beta, \gamma, \mu)$ and let $x = (x_1, x_2, \dots, x_n)$ denote the corresponding vector of observations where the likelihood function is defined by

$$L(x|\alpha, \beta, \gamma, \mu) = \prod_{i=1}^n \frac{1}{\gamma} S(\alpha, \gamma, \beta, \frac{x_i - \mu}{\gamma}, 0).$$

Defining $\theta = (\alpha, \beta, \gamma, \mu)'$ the ML estimate of θ is obtained by maximizing the log-likelihood function

$$L_\theta(x) = \sum \ln f(x_i; \theta) \quad (2.19)$$

with respect to the unknown parameter vector θ . DuMouchel [29] investigates the theoretical properties of the ML estimator for θ and shows its asymptotic normality under certain regularity conditions.

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0)), \quad (2.20)$$

where " \xrightarrow{d} " stands for convergence in distribution and I denotes the fisher information matrix which can be approximated using the Hessian matrix arising in the maximization [92] by direct numerical integration.

In our ML estimation algorithm, we maximize the log-likelihood function (2.19) numerically. Rather than employing constrained optimization, we estimate a transformed version of θ , say $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\mu})$ such that $\theta = h(\hat{\alpha})$. The transformation can take the form

$$\alpha = \frac{2}{1 + \hat{\alpha}^2}, \quad \beta = \frac{2 - \hat{\beta}^2}{2 + \hat{\beta}^2}, \quad \gamma = \hat{\gamma}^2 \quad (2.21)$$

In many applications, it is assumed that first moments of X_i exist. Then one may restrict $\alpha \in (1, 2]$ by adopting the transformation $\alpha = 1 + 1/(1 + \hat{\alpha}^2)$ [94]. With the parameter transformations in place and defining the gradient $\Delta_0 h(\theta) = \frac{\partial h(\hat{\theta})}{\partial \hat{\theta}}$, (2.20) becomes

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Delta_{\hat{\theta}} h(\theta) I^{-1}(\hat{\theta}_0) h'(\hat{\theta})). \quad (2.22)$$

The maximum likelihood estimation has been implemented in two packages of R statistical software: *StableEstim package* by Kharrat[66] and *stable package* by Nolan [95] respectively. In this study we use the [95] stable package due to computational speed and its variety of faster algorithms.

2.4.5 Assessing the Goodness of fit

Kolmogorov-Sminorv (KS) test is applied to test the goodness of the fitted of the stable distribution [122]. This method is based on the empirical cumulative distribution function (ECDF), it measures the supremum distance between the cumulative distribution function of the theoretical distribution and the empirical distribution function, over all the sample points. The general procedure consists of defining a test statistic which is some function of the data measuring the distance between the hypothesised distribution and the data, and then calculating the probability of obtaining data which have a still larger value of this test statistic than the value observed. The KS test is distribution free since its critical values do not depend on the specific distribution being tested. The KS test is relatively

insensitive to differences in the tails but more sensitive to points near the median of the distribution[79]. The KS test statistic allows us to quantify the extent to which a theoretical or estimated distribution of the step lengths distribution describes the observed data [80]. A p-value is calculated as the proportion of the artificial data showing a poorer fit than fitting the observed data [19]. When this value is less than 0.1, the null hypothesis that the distribution fits the data can be considered to be drawn from the fitted distribution, and if not, is rejected. Following Clauset et al. [19], we choose 0.1 as the threshold p-value for validating the statistical test taken on the data.

2.5 Real data application

Animals are often equipped with radio-tracking tags containing GPS receivers which record the location of the animal at set intervals, and either store the data for later retrieval, or transmit the data in real time via satellite or GSM networks. Traditional monitoring approaches use an ad-hoc alerting thresholds to monitor the variables for abnormal events such as poaching or human wildlife conflict. The GPS radio tracking data-driven approach is a useful alternative to this, in which we construct a probabilistic model to capture the structure of the data and aim to identify and monitor the movement patterns and behaviour of animals. In the latter, inappropriate assumptions about the distribution of the random variables may result in underestimation of the tail mass, as discussed earlier [34–37]. This will lead to inaccurate classification of the sample data, and so careful consideration should be given to such assumptions during model development. In this section, we illustrate the application of stable family of distributions to animal movement step lengths. Stable distribution are fitted to step lengths of six mature female elephants tagged with GPS radio-collars following the strict ethical standards of animal handling with specific approval from the University of KwaZulu-Natal Animal Ethics sub-committee (*Ref. 009/10/Animal*) and registered and approved by SANParks project, in association with Kruger National Park and Scientific Services (*Ref: BIRPJ743*). The study area is located at Kruger National Park

(KNP) and associated private reserves along the western boundary (Sabie Sand, Klaserie, Timbavati, Umbabat and Manyaleti), covering an area of approximately equal to 21,281 km², in the north-eastern Lowveld region of the South Africa. Our study area covers the southern, central and western regions of KNP and includes the associated private reserves, since elephant are able to move freely between these areas as described in [11]. Step lengths(linear metrics) of animal movement were derived using the tracking tool in ArcGis 10.0 and the resulting ArcView shapefile was exported as a tab delimited text file for input to R (R Development Core Team 2008) environment in which we perform the remaining analysis. Step lengths of elephant movement were filtered to weed out erroneous fixes caused by GPS error using an upper, biologically based threshold speed of 7 km/h [135].

2.6 Results of analysis and interpretation

Table 2.1 gives the summary statistics for the step lengths of the six elephant herds. The maximum step length show the longest distance an elephant can cover while migrating or foraging within a period of 30 minutes. The maximum step length varied from herd to herd with AM255 having the largest value of 3.68km and herd AM254 having the least value of 2.45km. The average step length for the six herds ranged from 169 to 240 meters with herd AM99 having the highest value (238.5 meters) and herd AM208 having the lowest average of 169.5 meters. The step lengths appears to be asymmetric as reflected by the positive skewness estimates: there are more observations in the right-hand tail than in the left-hand tail. The skewness statistic varied from herd to herd and ranged from 24.09 to 2.69. Further, the histogram in Figure 2.2 confirms that the data of all the six elephant herds are right skewed. The kurtosis statistic for the step lengths of the six elephants herds is higher than that of a standard Gaussian distribution which is 3, showing the fat tails patterns of these empirical distributions. Kurtosis varied from herd to herd, which implies the biological significance of individual herd.

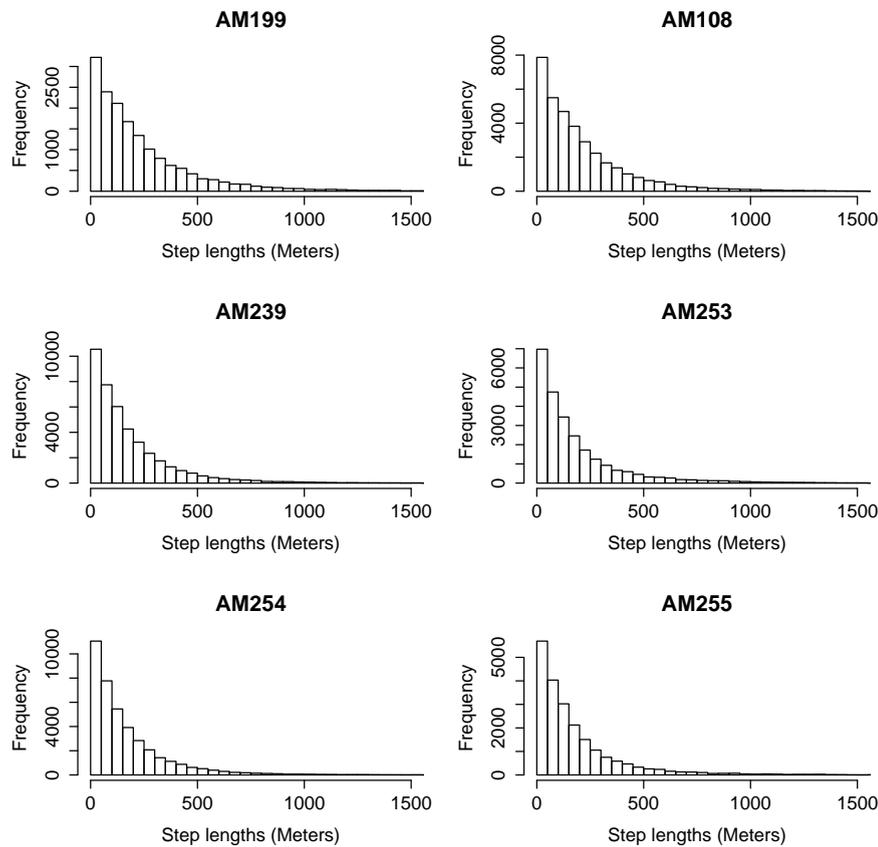


FIGURE 2.2: Histograms of the step lengths of the six elephant herds collected from Kruger National Park (2006-2009)

Figure 2.2 displays a histogram of the elephant movement step lengths. The plot illustrates that, the step lengths of the six herds heavy tailed and skewed to the right which suggests a leptokurtic distribution family. Table 2.2, gives the three stage parameter estimates of stable distribution for step lengths of the six elephant herds. In the analysis we assume that the time steps are equal and independent. The parameter α in the Stable Paretian and the power law distributions is the stability index which measures the heaviness of the tails. The parameter β in the Stable Paretian distribution is the skewness parameter, while γ and μ are the scale and location parameter respectively.

The estimates for the tail index α of the stable distribution nearly take almost the same values regardless of which herd they come from, suggesting an intrinsic characteristic of the data [23, 104]. The values of α ranged between 1 and 2 which is consistent with previous studies that fitted the power law distribution to study

the tail behaviour of animal movement step lengths. Note that $\alpha < 2$ departing from the Gaussian behaviour and capturing the heavy tail effects. We note that tails of animal movement step lengths may contain important biological signals which are necessary in understanding animal movement patterns and behaviour. Comparing the tails index α of the stable distribution and the power law exponent α we find that the power law tail estimate is less than the stable tail index which implies that the power law exponent overestimates the tail behaviour if skewness in the data is not accounted for during analysis. A reanalysis of data sets used in the studies of [34–37] needs to be carried out to establish this claim.

TABLE 2.1: Descriptive analysis of six herds of Elephant movement step length in Kruger National Park collected every 30minutes

stats	Am99	AM108	AM239	AM253	AM254	AM255
Sample size(n)	16058	35272	41876	25388	39567	21097
Minimum	5.07	5.04	5.02	5.04	5.05	5.05
Maximum	3211	2616	2830	3429	2450	3675
Mean	238.50	213.19	184.09	189.11	169.52	180.42
Median	157.99	145.16	120.28	112.17	107.27	112.07
SE Mean	2.04	1.20	1.00	1.43	0.97	1.48
Stdev	258.45	225.52	205.10	227.37	192.67	214.51
Skewness	2.44	2.25	2.71	2.65	2.69	2.91
Kurtosis	9.28	7.39	11.80	10.33	11.43	13.81

The skewness parameter β for all the six herds is greater than 0 ($\beta = 1$) which indicates that the step lengths are positively skewed. However, this may be a problem in future model and is known as a boundary problem. In movement ecology, a value of $\beta = 0$ would be interpreted to mean an animal is performing a Brownian motion search foraging, while a value of $\beta > 0$ indicates the animals make large infrequent rare steps especially when search for scarce food resources like water and forage.

The parameter γ measures the amount of variation in animal movement step lengths. The values of γ were positive and varied from herd to herd irrespective of the method of estimation. The parameter μ is a measure of central tendency was positive and differed from herd to herd.

TABLE 2.2: Comparison of parameter estimates and goodness of fit for the stable and power law distribution of elephant movement step lengths data sets collected from Kruger National Park South Africa recorded every 30 minutes from 2006 to 2009.

McCulloch quantiles method(stage 1)						
	AM99	AM108	AM239	AM253	AM254	AM255
α	1.62	1.69	1.61	1.42	1.60	1.48
β	1.00	1.00	1.00	1.00	1.00	1.00
γ	126.27	114.75	96.69	97.37	90.35	92.26
μ	150.64	138.32	114.23	106.38	101.32	104.82
KS.dist	0.14	0.15	0.14	0.13	0.15	0.13
p-value	0.10	0.13	0.10	0.102	0.13	0.102
Koutrovelis-Kogon-Williams algorithm(stage 2)						
parameter	AM99	AM108	AM239	AM253	AM254	AM255
α	1.52	1.56	1.49	1.37	1.48	1.40
β	1.00	1.00	1.00	1.00	1.00	1.00
γ	110.21	100.89	83.27	80.04	77.29	76.50
μ	112.67	105.04	85.85	75.04	75.04	75.54
KS.dist	0.16	0.16	0.16	0.16	0.17	0.15
p-value	0.15	0.15	0.15	0.155	0.14	0.14
Maximum likelihood estimation(final stage)						
	AM99	AM108	AM239	AM253	AM254	AM255
α	1.67	1.69	1.61	1.42	1.60	1.58
β	1.00	1.00	1.00	1.00	1.00	1.00
γ	99.98	91.78	73.32	64.39	67.22	68.05
μ	148.74	137.45	108.77	89.73	97.37	97.42
KS.dist	0.10	0.10	0.09	0.08	0.10	0.10
p-value	0.11	0.15	0.18	0.17	0.15	0.13
Paramter estimates and goodness of fit of the power law distribution to six herds of elephant movement						
	AM99	AM108	Am239	AM253	Am254	AM255
α	1.31	1.32	1.33	1.34	1.34	1.34
xmin	5	5	5	5	5	5
N	16060	35272	41876	25388	39567	21097
KS.dist	0.31	0.3	0.29	0.27	0.28	0.28
p-value	0.102	0.13	0.19	0.12	0.145	0.143

Though the values of the α vary from herd to herd, they still retain the biological meaning. From the interpretation of the four stable distribution parameters, we note that complex models can be built based on the stable distribution assumption to investigate the effects of environmental drivers on animal movement. The stable generalized linear model of [71] can be extended to stable generalized mixed models with herds as the random effects.

Despite the desirable statistical properties and empirical evidence of the tail index in ecological data [23, 34–37], the power law distribution has almost been used exclusively to model step lengths of animal movement ignoring the skewness in the data. However, in this paper, we have demonstrated the stable Paretian distribution fits the step length better than the power law by capturing both the skewness and the heavy tails properties of the data.

FIGURE 2.3: Stable distribution density function plots for elephant movement step lengths

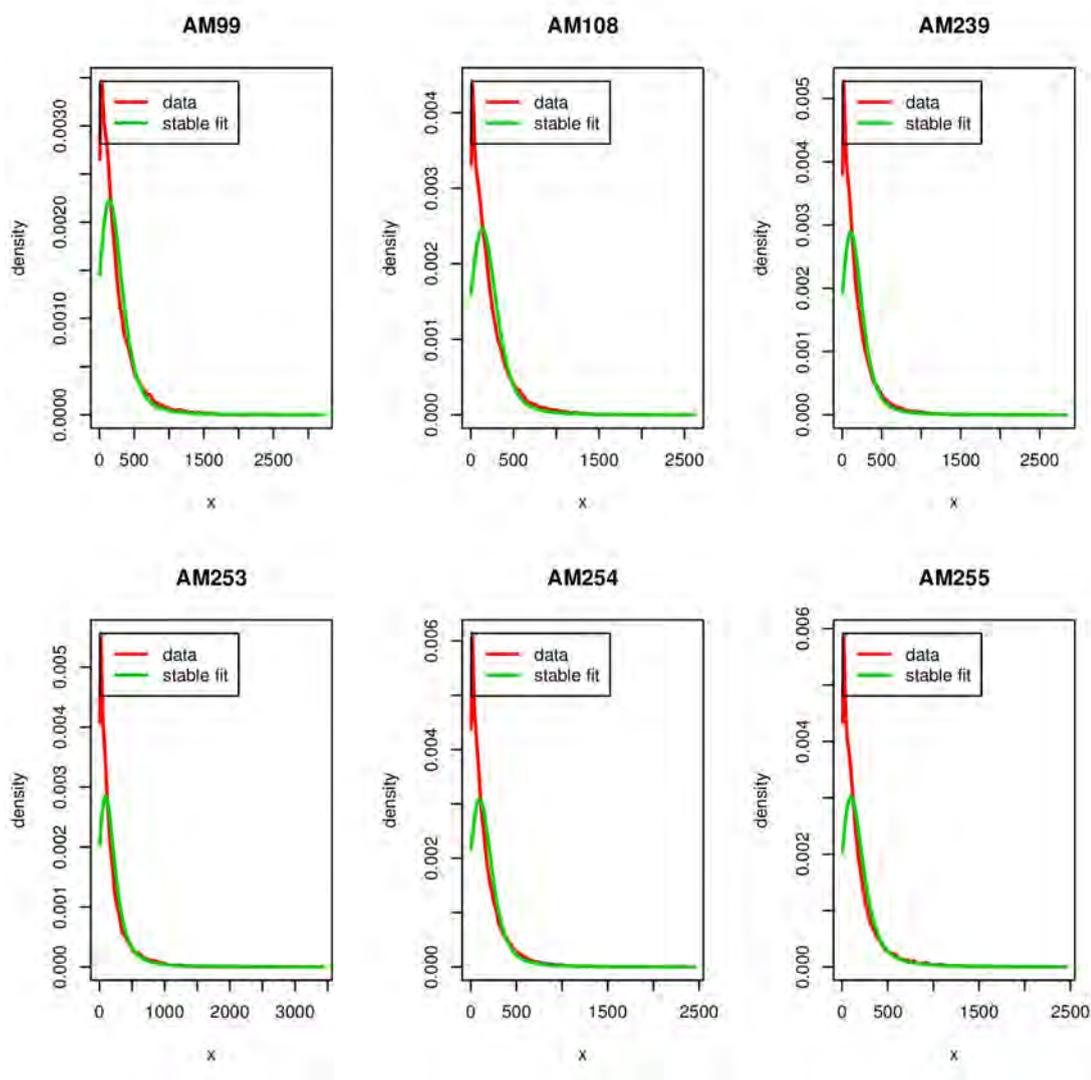


Figure 2.3 illustrates the stable density fit diagnostic tests for step length of six elephant herds. The density plot illustrates that the stable distribution fits the

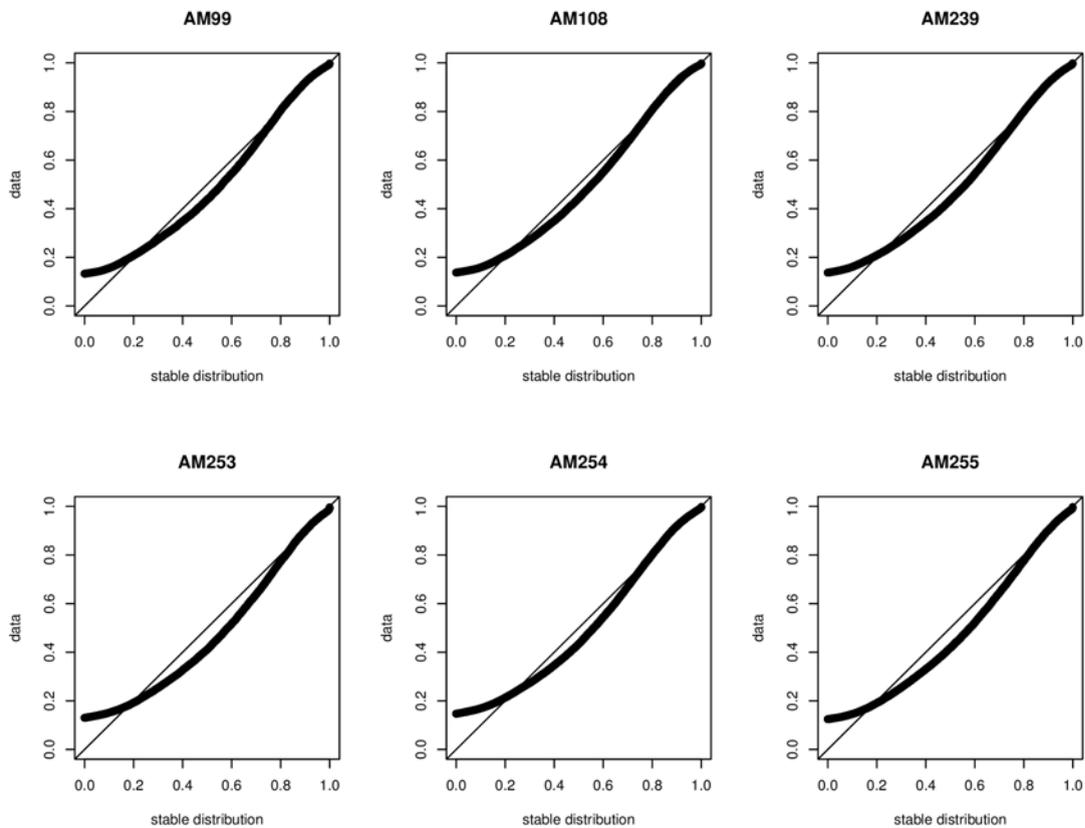


FIGURE 2.4: Probability plots (pp) of stable distribution of six elephant herds movement step data collected from KNP between 2006-2009 lengths

data well compared to the theoretical probabilities of the stable distribution. Figure 2.3 displays the P-P plot of the six elephant herd movement step lengths data. The stable distribution P-P plot displays a perfect fit of the probabilities of the data to the theoretical probabilities of the stable distribution. Based on the K-S test statistic as indicated by the results in Table 2.1 and 2.2 we found that the stable Paretian distribution fitted the step lengths data better than the conventional power law distribution. Therefore, elephants (*africana loxadonta*) are representative of the five large mammals and the stable Paretian distribution with $\alpha < 2$ clearly outperform the power law distribution.

2.7 Final remarks

Analysis and interpretation of animal movement data is one of the most controversial issues in movement ecology [65]. The biological importance of rare events such as long steps, and on the other hand, availability of large amounts of high-resolution animal tracking data owing to recent developments in the methodology and technology [11, 45, 56, 57], pose a considerable challenge for data analysis [91]. The commonly used statistical approaches do not always allow distinguishing unambiguously between qualitatively different processes such as Brownian motion, correlated random walk and the Lévy flight, and that has caused a heated debate [36, 101, 105]. Moreover, it is an open question as to what extent the animal movement step lengths can be adequately described by the Lévy walk models as it is intuitively clear the results may depend on the choice of appropriate probability distribution that adequately captures the statistical properties of the data [120].

In this paper, we have demonstrated that, with the recent technology and efficient computing power, stable distributions can adequately capture the properties of large data of animal movement. Having applied various methods of parameter estimation, we compared the goodness of fit of the stable distribution, and the power law distribution. We have shown that the family of stable distribution of which the power law is a variant, is the most biologically tractable and flexible model for animal movement data sets. The results of elephant herd movements have major implications for most of the traditional statistical modelling approaches. One of the main challenges in individual animal movement studies is to identify movement pattern. The tails of animal step lengths may contain important biological signals which characterise animal long term behaviour. Another issue in animal movement studies is the argument that the movement is scale free which is unrealistic as animal step lengths have scale [65]. It is this issue that we seek to address by applying the four parameter stable distribution which is more robust. The scale parameter of the stable distribution captures the scale in animal movement data.

For instance, unlike the power law distribution which is controlled by the x_{min} , the stable distribution is not. This leads to a counter-intuitive conclusion that, for an individual elephant movement step length described by power law distribution [23], we cannot separate the apparently different scales of small and large step lengths and the result of the analysis of search patterns in dispersed food patches essentially depends on animal movements at a small scale [98]. We refer to this situation as a genuine Lévy flight. It is not scale-free, as it was suggested by earlier studies [65, 80], but its specifics are that the small and large scales appear to be equally important and cannot be separated.

We note that animal movement is never uninterrupted, usually it alternates between periods of fast and slow movement or rest [65, 80]. The desire to capture as much information from the data as possible, can lead to inclusion of values of x_{min} even when the animal is resting i.e., $x_{min} = 0$. In this case the power law distribution in previous studies has been shown to be sensitive to the choice of x_{min} . However, in the case $x_{min} = 0$, a power law-type distribution is not valid. This problem has been solved by using the stable distribution which does not require the choice of the scale parameter x_{min} and hence the hypothesis of the Lévy flight movement pattern is restored.

It is notable that due to herd to herd variation, future studies should consider a mixed effects models with herds as the random effects. All values of the stability index α were in the range of 1 to 2 which shows that elephants have evolved to use the Lévy flight search strategy to optimize scarce food resources. These findings supports an earlier study by Dai et al. [23] who reported Lévy flight foraging patterns in elephants from Pongola Game Reserve in South Africa. The result of this application opens perspectives for new approaches of the analysis of step lengths of animal movement, particularly incorporating both skewness and heavy tails to identify animal movement pattern and extract important biological signals from the tails of the data. These and many more we leave them as a future research possibility.

CHAPTER 3

Analysing the effects of landscape factors on animal movement step lengths with stable law regression models

Abstract

The potential advantage of stable distribution assumption in modelling ecological disturbances of animal movement is the central theme of this paper. Studies relating animal movement paths to structured landscape data are particularly lacking despite the obvious importance of such information to understanding animal movement. Previous studies of elephant movement have shown that speed is heavy tailed and skewed. In this paper we model the heavy tails using the student t regression model, the skewness and the heavy tails with the stable law regression. The new models add substantial flexibility and capabilities, including the ability to incorporate multiple variables. We use a likelihood based approach that utilizes the Fourier Transform technique to evaluate the densities and demonstrate the approach with movement data from five elephant herds (*Africana Loxadonta*). The proposed methodology can be useful for GPS tracking data that is becoming more common in monitoring of animal movement behaviour. We discuss our results in

the context of the current knowledge of animal movement and in particular elephant ecology highlighting potential applications of our approach to the study of wide ranging animals.

Keywords: Stable law regression model; Elephant movement; habitat types; GPS data; Student-t regression model.

3.1 Introduction

Regression analysis is one of the most popular methods in ecology and statistics; where most variables of interest such animal movement step lengths are assumed to be normally distributed. However, the normality assumption is not appropriate for many ecological variables, especially animal movement linear metrics (speed, step lengths, etc) variables and also, in some cases circular metrics (turn angles) [5]. Animal movement linear metric are typically heavy tailed and excessively highly peaked around zero. A stable distribution, whose shape is governed by the stability index parameter α , represent one such alternative. Thus, such a distribution is better suited to describing such variables; the normal distribution is a special case of the stable distribution. To this end, the four parameter family of stable distribution is more of a generalization of the central limit theorem than an alternative.

The flexibility of stable distribution can be explored in a regression modelling framework to overcome some of the deficiencies of linear regression models when analysing heavy tailed and skewed data. The non-Gaussian stable distributions have heavier tails than the Normal distribution and allow skewness [136]. Heavy tails and skewness implies that extreme observations are given a greater probability of occurring and are thus given less weighting in maximum likelihood estimation so that fitted lines are not biased towards these extreme observations [71, 136]. Therefore, it is a reasonable extension to the regression models to assume a stable distribution as the distribution of the error terms. Alternative models to the stable law regression models are the Student's t, skewed Student's t and skewed normal

regression models [136]. Fonseca et al. [44] demonstrates that the regression model with Student's t errors also suffers from monotone likelihood as it is ill-behaved for v close to zero and may be ill-behaved when $v \rightarrow \infty$. The use of the normal distribution to model the errors of linear model is under increasing criticism for its inability to model fat or heavier tailed distributions as well as being non-robust. Lange et al. [72] generalized the traditional regression model with normal distributed errors to more robust regression models with t distributed errors. It is well known that the t distribution provides a convenient description for regression analysis when the residual term has a density with heavy tails.

The stable distribution has found wide applications in financial problems, biology, genetics, ecology and geology [127] with a few applications in movement ecology [63]. The assumption that animal step lengths or speed follow a stable distribution has far reaching consequences for both foraging ecology and statistical theory [63]. For example, the problem of Lévy flight search patterns is well studied [see 34–37], robustness to the sampling frequency is studied in [65] and for a specific discussion of movement ecology and statistical issues see, [63]. However, in all these studies, the link between linear metrics and the environmental heterogeneity variables remains unexplored. According to [27], different vegetation cover types have varying impacts on elephant movement. Surface water availability, patch quality, rainfall and distance to the water bodies is known to affect elephant movement. The effects of artificial water points and fences has been investigated [74]. In this study we examine the effects of vegetation cover type in a stable regression model setup in order to understand elephant movement.

Advances in statistical computation have made it possible to estimate the unconditional stable density as well as incorporate covariates [71]. However, estimates of the stable distribution conditional on a set of explanatory variables in the context of regression framework used by applied researchers poses an overwhelming computational problem [136]. One of the methods used for evaluating the stable density (the direct numerical integration techniques) is non-trivial and burdensome from a computational perspective [95]. As a consequence, maximum likelihood estimation algorithms based on such approximations are difficult to implement especially for

huge data sets encountered in movement ecology [63]. However, with increasing computational power and efficient algorithms, maximum likelihood estimation and other comparative techniques have been implemented by Nolan and Ojeda [95]. Due to the above mentioned drawbacks, stable distributions are not well explored in movement ecology.

The aim of this paper is to introduce the stable regression model to better understand the dependence of animal movement on vegetation cover types and relaxing the reliance on the normal assumption which is commonly used. In section 3.2 we introduce the stable distribution and its statistical properties. In section 3.2.1, the stable regression model is developed by assuming the stable distribution assumptions. The student's t regression model is provided and its properties discussed. Finally, section 3.3 gives an applications to a real data set of five elephant herds, including a comparison between the student t regression model and the stable distribution error terms regression model fits.

3.2 Stable distribution

Stable distributions are a four parameter family of probability models, which was first introduced by [73] in a study of normalized sums of independent and identically distributed (i.i.d) terms. A random variable X is said to be stable distributed if for any positive integer $n > 2$, there exist constant $a_n > 0$ and $b_n \in \mathbb{R}$ such that $X_1 + \dots + X_n \stackrel{d}{=} a_n X + b_n$ where X_1, \dots, X_n are independent identically distributed copies of X and “ $\stackrel{d}{=}$ ”-signifies equality in distribution. The coefficients a_n is necessarily of the form $a_n = n^{1/\alpha}$ for some $\alpha \in (0, 2]$ [41]. The parameter α is called the index of stability (tail index) of the distribution and a random variable X with index α is called α -stable. An α -stable distribution is described by four parameters and will be denoted by $s(\alpha, \beta, \gamma, \mu)$. Closed form expressions for the probability density function of the α -stable distribution is known to exist only for three special cases (cauchy, normal and Lévy distribution).

The work of Mandelbrot [76] and Fama [38] elicited a lot of interest in using stable distributions to model heavy tailed and skewed phenomena, but research has been restricted to theoretical context due to computational complexities involved in calculating the probability densities and the consequently what this has for the maximum likelihood procedures. Notable contributions in this field of study are found in DuMouchel [29, 30, 31], Zolotarev [143], [110], Janicki and Weron [60] and more recently Nolan [92, 94], Nolan and Ojeda-Revah [95].

Although the probability density function of the stable distribution cannot be written in closed form, the characteristic function, which can be specified in a closed form for all stable distributions, allows the only opening for practical use of the distributions in real life problems [83]. The characteristic function can be expressed in several different forms, each of which has advantages over others, for example formula simplicity over computational consistency. However, the Zolotarev's form has the advantage of being continuous in all the four parameters, and behaves more intuitively than in other forms [95]. Lambert and Lindsey [71] discuss complexities in fitting their regression model caused by the sensitivity of the location parameter to the skewness parameter. For numerical purposes, several authors have recommended the use of Zolotarev's parameterization as the most practical in application to real life data sets [95].

The characteristic function of a stable random variable X is given by

$$E \exp(iuZ) = \begin{cases} \exp(-\gamma|u|^\alpha [1 + i\beta(\text{sign } u) \tan \frac{\pi\alpha}{2} (|u|^{-\alpha-1} - 1)] + i\mu u) & \alpha \neq 1 \\ \exp(-\gamma|u| [1 + i\beta(\text{sign } u) \frac{2}{\pi} \log|u|] + i\mu u) & \alpha = 1 \end{cases} \quad (3.1)$$

The family of stable probability density can be calculated using the Fourier Transform of the characteristic function, given by

$$S(x, \alpha, \beta, \gamma, \mu) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \Phi(x; \alpha, \beta, \gamma, \mu) dt \quad (3.2)$$

Statistical software to fit stable distribution and density functions are available in Rmetrics for R [141], stable [95] or as standalone program *STABLE* [92]. These

resources allow one to evaluate the consequences of replacing the normal assumption with the more general stable distribution. Further advances in theory and computation will aid the development of new models in the coming years and the use of the stable distribution will become more common.

3.2.1 Stable Paretian regression model

In many practical applications in animal ecology, it is known that animal movement rate can be affected by a number of covariates (explanatory variables) such as the nearest distance to the water point, vegetation cover type, distance to tourist roads, soil topology, seasons, amount of rainfall, temperature and many others [27]. However animal movement data is characterized by skewed and heavy tailed distributions. Thus a model that provides a good fit to movement data will definitely yield more precise estimates of the quantities of interest. Based on the stable distribution assumption, we propose a linear regression type model linking the response y_i and the explanatory variables $X = (x_1, x_2, \dots, x_n)$ as

$$y_i = \beta_0 + \sum_{i=1}^n \beta_{ij} x_i + \epsilon_i, \quad i = 1, \dots, n \quad (3.3)$$

where β is a vector of the unknown parameters to be estimated and ϵ_i is the random error term. The notion of stable regression models (SRMs) was developed by [82] for symmetric stable distribution and discussed in detail by McHale and Laycock [83]. In SRMs, the error terms $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ are assumed to be independent identically distributed stable random variables denoted by $\epsilon_i \sim S(\alpha, \beta, \gamma, \mu)$.

Standard methods of approximating such integrals are of unknown accuracy in real settings. Instead, [29] suggested the use of numerical inversion of the First Fourier Transform (FFT) to obtain a closed density and hence the likelihood for stable distributions. In a similar manner, numerical inversion of the First Fourier Transform can be used to obtain the parameters of the stable Paretian regression model. DuMouchel [29–31] showed that subject to certain conditions, the maximum likelihood estimates of the parameters of an α -stable distribution

have the usual asymptotic properties of a maximum likelihood estimator. They are asymptotically normal, unbiased and have an asymptotic covariance matrix $n^{-1}I(\alpha, \beta, \gamma, \mu)^{-1}$ where $I(\alpha, \beta, \gamma, \mu)$ is the fisher information matrix. McCulloch [82] examines the linear regression model in the context of α -stable distribution paying particular attention to the symmetric case. Here the symmetry constraint is not imposed. If we denote the stable density function by $S(\epsilon_i; \alpha, \beta, \gamma, \mu)$ then we may rewrite the density of ϵ_i as

$$S(x; \alpha, \beta, \gamma, \mu) = \frac{1}{\gamma} S\left(\frac{y_i - \sum_{j=1}^k x_{ij}\beta_j}{\gamma}, \beta, 1, 0\right),$$

the likelihood as

$$L(\epsilon_i; \alpha, \beta, \gamma, \mu) = \frac{1}{\gamma} \prod_{i=1}^n S\left(\frac{y_i - \sum_{j=1}^k x_{ij}\beta_j}{\gamma}, \beta, 1, 0\right)$$

hence the log-likelihood function for the vector of parameters $\theta = (\alpha, \beta, \gamma, \mu, \beta_0, \beta_1, \dots, \beta_p)$ from model (3.3) has the form

$$l(\epsilon_i, \alpha, \beta, \gamma, \mu, \beta_0, \beta_1, \dots, \beta_p) = -n \sum_{i=1}^n \log(\gamma) + \sum_{i=1}^n \log\left(s\left(\frac{y_i - \sum_{i=1}^n x_{ij}\beta_j}{\gamma}, \beta, 1, 0\right)\right) \quad (3.4)$$

The ML estimator $\hat{\theta}$ of the vector θ of unknown parameters can be calculated by maximizing the log-likelihood (3.4) to obtain the solution to the equations

$$\begin{aligned} \frac{\partial \ell}{\partial \beta_m} &= \sum_{i=1}^n -\Phi(\epsilon_i) x_{im} = 0, m = 1, \dots, k \\ \sum_{i=1}^n -\frac{\Phi(\hat{\epsilon}_i)}{\hat{\epsilon}_i} \hat{\epsilon}_i x_{im} &= 0, m = 1, \dots, k \\ \sum_{i=1}^n -\frac{\Phi(\hat{\epsilon}_i)}{\hat{\epsilon}_i} (y_i - \sum_{i=1}^n x_{ij}\beta_j) x_{im} &= 0, m = 1, \dots, k \\ \sum_{i=1}^n -\frac{\Phi(\hat{\epsilon}_i)}{\hat{\epsilon}_i} y_i x_{im} &= -\sum_{i=1}^n -\frac{\Phi(\hat{\epsilon}_i)}{\hat{\epsilon}_i} \sum_{i=1}^n x_{ij}\beta_j, m = 1, \dots, k \end{aligned} \quad (3.5)$$

where $\epsilon_i = y_i x_{im} - \sum_{i=1}^n x'_i \hat{\beta}$. If we let W be the diagonal matrix given by

$$W = \begin{bmatrix} \frac{\Phi(\hat{\epsilon}_i)}{\hat{\epsilon}_i} & 0 & \dots & 0 \\ 0 & \frac{\Phi(\hat{\epsilon}_i)}{\hat{\epsilon}_i} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \frac{\Phi(\hat{\epsilon}_i)}{\hat{\epsilon}_i} \end{bmatrix}$$

then using the least squares notation, we may write the normal equations (3.5) as

$$X'W y = (X'W X) \hat{\beta} \quad (3.6)$$

and if $X'W X$ is not singular, the parameter estimates of β are given by

$$\beta = (X'W X)^{-1} X'W y \quad (3.7)$$

Nolan and Ojeda [95] showed that the evaluation of the likelihood function is made possible by using efficient non-linear optimizers. Maximum likelihood algorithm used in this work are provided by Nolan and Ojeda-Revah [95] within the R package *stable 5.1* which can be obtained commercially from www.RobustAnalysis.com. Initial values for $\alpha, \beta, \gamma, \mu, \beta_0, \beta_1, \dots, \beta_p$ can be taken from the fit of the stable distribution model.

3.2.2 Regression model with t errors

We consider the univariate nonlinear regression model where the observations $y = (y_1, \dots, y_n)'$ are independent, y_i having a student t distribution with location parameter μ_i , scale parameter σ and v degrees of freedom. The density of y_i , for each $i = 1, \dots, n$, is therefore given by

$$f(\beta, \delta, v; y, X) = \frac{\Gamma(\frac{v+1}{2})^n}{\Gamma(\frac{v}{2})^n (\pi v)^{n/2} \delta^n} \left(\left[1 + \frac{1}{v} \left(\frac{y_i - x'_i \beta}{\delta} \right)^2 \right] \right)^{-(v+1)/2} \quad (3.8)$$

where $\sigma > 0$ and $v > 1$ are both unknown. We define a linear regression for y by

$$y_i = \beta_0 + \sum_{i=1}^n \beta_{ij} x_i + \epsilon_i, \quad i = 1, \dots, n \quad (3.9)$$

where $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$ is the error vector where the components are independent and identically distributed according to the student t distribution with location zero and scale δ and degrees of freedom v [72]. $X = (x_1, \dots, x_n)'$ is the $n \times k$ matrix of explanatory variables. The parameter space is given by $\theta = (\beta, \delta, v)$. The likelihood is given by

$$L(\beta, \delta, v; y, X) = \frac{\Gamma(\frac{v+1}{2})^n}{\Gamma(\frac{v}{2})^n (\pi v)^{n/2} \delta^n} \prod_{i=1}^n \left[1 + \frac{1}{v} \left(\frac{y_i - x_i' \beta}{\delta} \right)^2 \right]^{-(v+1)/2} \quad (3.10)$$

The parameter estimates θ are obtained by maximizing the log-likelihood equation

$$\text{Log}L = \ln \Gamma\left(\frac{v+1}{2}\right)^n - \ln \Gamma\left(\frac{v}{2}\right)^n (\pi v)^{n/2} \delta^n - \frac{(v+1)}{2} \sum_{i=1}^n \log \left[1 + \frac{1}{v} \left(\frac{y_i - x_i' \beta}{\delta} \right)^2 \right] \quad (3.11)$$

The least squares estimator of β is

$$\beta = (X'X)^{-1} X'Y$$

The variance-covariance matrix for $\hat{\beta}$ is

$$\text{var}(\hat{\beta}) = E[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)] = \frac{v\sigma^2}{v-2} (X'X)^{-1}.$$

Lange et al. [72] noted that this is also the maximum likelihood estimate of β . Singh [118] provided the following estimate of the degrees of freedom parameter

$$\hat{v} = \frac{2(\hat{\alpha} - 3)}{\hat{\alpha} - 3}$$

where

$$\hat{\alpha} = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - x_i' \hat{\beta})^4}{\frac{1}{n} \sum_{i=1}^n (y_i - x_i' \hat{\beta})^2}$$

The maximum likelihood estimator of σ^2 is

$$\sigma^2 = \frac{1}{n}(y_i - x\hat{\beta})'(y_i - x\hat{\beta})$$

as in the normal case. For $v > 2$, $E(\hat{\sigma}^2) = \frac{(n-p)}{n}\sigma_u^2$ where $\sigma_u^2 = v\sigma^2/v - 2$ is the common variance of the elements of ϵ . Thus $\hat{\epsilon}'\hat{\epsilon}/n - p$ is an unbiased estimator for σ_u^2 while

$$\sigma^2 = \frac{v-2}{v(n-p)}\hat{\epsilon}'\hat{\epsilon} \quad (3.12)$$

is unbiased estimator for σ^2 . In the class of estimator $q\hat{\epsilon}'\hat{\epsilon}$, with q being a positive scalar, the minimal mean squared error estimator for σ^2 is

$$\sigma^2 = \frac{v-4}{v(n-p+2)}\hat{\epsilon}'\hat{\epsilon}, \quad (3.13)$$

while the minimal mean squared error estimator for σ^2 in this class is $(v-4)\hat{\epsilon}'\hat{\epsilon}/(v-2)(n-p+2)$. The variances of the unbiased and the minimal mean squared error estimators of σ^2 are

$$\text{var}(\hat{\sigma}^2) = \frac{2\sigma^4}{(n-p)} \frac{n-p+v-2}{v-4}, \quad (3.14)$$

Maximum likelihood algorithm used in this work are provided by [96] in R statistical package *'heavy'*.

3.3 Application to elephant movement data

3.3.1 Data description

The telemetry data employed in this study was collected by the South African National parks (SANPARKS). In May 2006, 18 African elephants were fitted with GPS -argos telemetry collars (Telenics). Capturing and handling was done according to the University of KwaZulu Natal animal care regulations. GPS locations were recorded every 30 minutes during the first three years after collaring and

transmitted to SANPARKS via an Argos satellite uplink every day when the elephant was within network range [11]. Telemetry points collected within the first 24 hours after capturing and those with obvious errors were excluded from the analysis. Overall, the telemetry data set was composed of more than 50,000 GPS points, taken over a period of three years, across a 19, 485km² area [11, 131].

3.3.2 Vegetation cover types

To determine the effects of various habitat types in the pattern of elephant movement, we extracted the vegetation cover types data of before and after the breakpoints. Land cover types and distances to different landscape features within a spatial resolution of 30 m pixels were obtained from the Kruger National Park Land cover database. This database is based on the Thematic mapper sensor on Landsat Earth-resource satellites using data frames recorded between 2006 and 2009 (spectral analysis Inc.2009). Dummy variables of vegetation cover types were created and fitted to a regression model assuming stable distributed error terms. The land cover of Kruger national park (KNP) consist of twelve vegetation cover types in Table 3.1.

3.3.3 Model formulation

The observations of the response variable y_1, y_2, \dots, y_n represent the movement rate of five elephant herds derived before and after breakpoint home ranges [see 11, for further details]. The covariate vector x_i is the dummy variables representing the vegetation cover type created from the habitat variable. Due to computational complexity of the stable regression model and lack of rich data set with covariates of elephant herds, we shall demonstrate the results of habitat cover types only in this study. The dummy variables created from vegetation cover type are presented in Table 3.1 Now we present the results by fitting the model

$$y_i = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{23} \quad (3.15)$$

TABLE 3.1: Dummy Variables of vegetation cover type under investigation

y_i	Speed of the animal
x_{12}	comb cover
x_{13}	thicket cover
x_{14}	Mixed Combretum/ <i>Terminalia sericea</i> woodland
x_{15}	Combretum/mopane woodland of Timbavati
x_{16}	<i>Acacia welwitschii</i> thickets on Karoo sediments
x_{17}	Kumana Sandveld
x_{18}	Punda Maria Sandveld on Cave Sandstone
x_{19}	<i>Sclerocarya birrea</i> subspecies <i>caffra</i> / <i>Acacia nigrescens</i> savanna
x_{20}	Dwarf <i>Acacia nigrescens</i> savanna
x_{21}	Bangu Rugged Veld
x_{22}	Combretum / <i>Acacia nigrescens</i> Rugged Veld
x_{23}	Lebombo South

where the dependent variable y_i speed of elephants follows the stable law distribution or the Student's t distribution for $i = 1 \dots, 200$. The dependent variable y_i is the speed of elephant before and after breakpoint home ranges obtained as described in [11] and [131]. The MLEs of the model parameters are calculated using the procedure `nlm` in R statistical software. Iterative maximization of the logarithm of the likelihood function of the stable law regression starts with some initial values for the θ taken from the linear regression model.

3.4 Results of the analysis

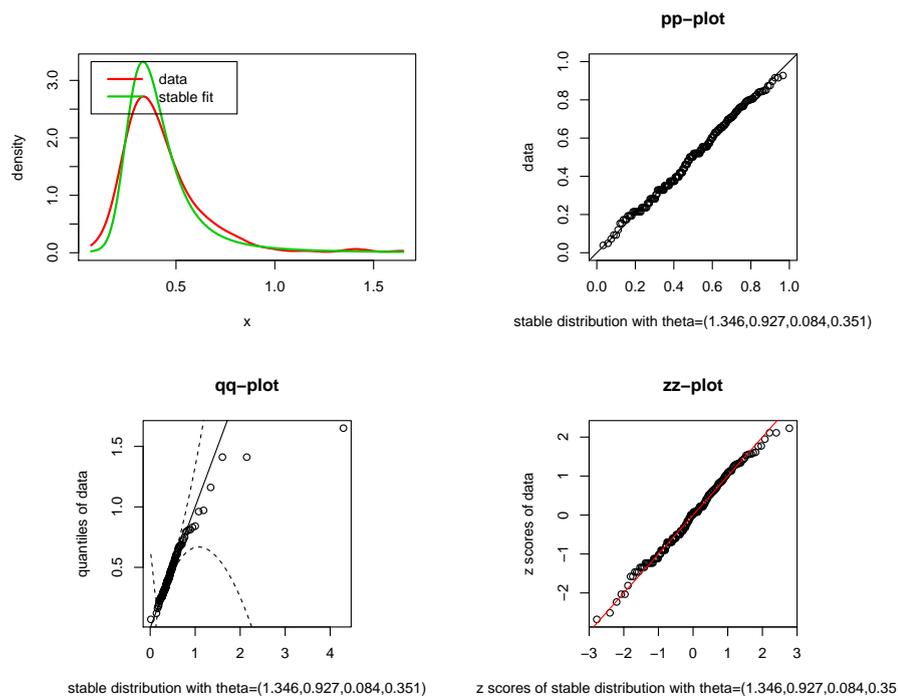
Table 3.2 lists the MLEs of the parameters for the SRMs and HTRMs models fitted to the current data. The SRMs model involves four extra parameters which gives it more flexibility to fit the elephant movement data. Due to lack of rich data set of animal movement with covariates we investigate only the effects of habitat types as dummy variables. Most of the environmental variables considered here were selected as drivers of movement rates before and after break point analysis of home ranges [see 131]. The fitted SRM indicates that the dummy variables X12, X13, X14 X18 and X23 are significant at 5% level of significance. The linear regression intercept was, however, significantly less than 1 indicating the ability of our models to predict the movement of the elephant at moderate speed.

Since we have demonstrated that the residuals are non-Gaussian, we will now compare the stable estimates with those obtained from the heavy tailed regression model with student's t distributed disturbances. The results of student's t regression model in Table 3.2 indicates that the vegetation covers combo, thicket, mixed combretum, punda maria sandveld and lebombo south significantly reduced the movement rates of elephants while mopane woodland, acacia welwisitchii, Kumana sandveld, Sclerocarya birrea subspecies, dwarf acacia savanna, Bangu rugged and Combretum acacia rugged increased the movement rates though not significant. The results of stable law regression model indicates that the vegetation covers combo, thicket, mixed combretum, punda maria sandveld and lebombo south significantly reduced the movement rates of elephants while mopane woodland, acacia welwisitchii, Kumana sandveld, Sclerocarya birrea subspecies, dwarf acacia savanna, Bangu rugged and Combretum acacia rugged increased the movement rates though not significant. We note that the movement of elephants in the resource poor patches are positive and significant indicating that elephants increased there movement speed when moving from search of food and water while in resource rich patches the move at a slower speed as they forage. The stability index parameter estimated is 1.31 which is less than 2 with a standard error of 0.0511 indicating that the data is heavy tailed. Clearly we can reject the null hypothesis that the random disturbance follows a Gaussian distribution (the hypothesis $\alpha = 2$) in favour of the alternative that the disturbance follows a non-Gaussian stable distribution with infinite variance. Figure 3.1, further supports the findings of the fitted model with residuals of the stable distribution plotted along the empirical density of the data. The density plot shows that the empirical distribution has heavier tails and a higher more concentrated peak compared to the Gaussian distribution. These attributes convey the ecological importance of the tails with appropriate statistical assumption. We used the Akaike information criterion to compare the student's t and the stable law regression models. The results of Table 3.2 indicates that the student t regression model better fits the data than the stable law regression model with an AIC of -170.46 and -162.78 respectively. However, we note that the student t regression model only captures the heavy tails

TABLE 3.2: Summary of Heavy tailed -t distribution and stable law regression model

parameter	Heavy tailed model			Stable law regression model		
	Estimate	Std.err	p-value	Estimate	Std.err	p-value
α				1.309	0.051	1.000
β				0.857	0.081	1.000
γ				0.091	0.003	1.000
intercept	0.376	0.025	1.000	0.34	0.011	0.000
X12	-0.089	0.054	0.000	-0.086	0.024	0.000
X13	-0.114	0.047	0.000	-0.144	0.02	0.000
X14	-0.068	0.034	0.000	-0.053	0.015	0.000
X15	-0.001	0.044	0.000	0.008	0.019	0.663
X16	0.067	0.037	0.9999	0.033	0.016	0.978
X17	0.228	0.058	1.000	0.165	0.026	1.000
X18	-0.047	0.045	0.0202	-0.08	0.02	0.000
X19	0.008	0.034	0.6772	0.008	0.015	0.5
X20	0.064	0.039	0.9999	0.043	0.017	0.994
X21	0.006	0.042	0.6103	-0.006	0.019	0.382
X22	0.021	0.041	0.8508	0.014	0.018	0.773
X23	-0.038	0.047	0.0559	-0.048	0.020	0.009
loglike	88.00481			84.16449		
AIC	-170.4644			-162.7838		

FIGURE 3.1: Diagnostic analysis of elephant movement data



and fails to account for the skewness in the data.

3.4.1 Biological implications and applications

The empirical analysis shows that the effect of vegetation cover types is to reduce the movement rates of elephant in abundant food patches and increased the movement rate in poor resource areas. This finding is consistent with the descriptive analysis of [27] who hypothesized that the quality and availability of forage suppresses movement rates of elephants. The findings also support the argument by [55] who found that resource rich vegetation cover reduced the movement rates of animals in Serengeti Game reserve, Kenya.

The stable regression model estimated in this paper sheds more light on the earlier results by identifying how the underlying ecological processes result to differential habitat use. African elephants have large effects on vegetation cover and high numbers can lead to extensive habitat modifications. Driven by the need to manage these impacts several models have been developed to better understand the interaction between elephants and trees. Therefore this understanding can be used for both management and habitat conservation [23]. Another implication of the stable regression analysis is that the distribution of movement rates-even when conditioning on the vegetation attributes has infinite variance. This means that the point predictions are useless because they lack precision especially when the stable parameters α and β are at the boundaries.

Confirming the foraging success and measuring the impact of environmental drivers is one of the challenges facing ecologists today. Thus the finding of this paper provides a direct link of inferring the effects of vegetation cover types on elephant movement speed. However, the stable Paretian model does not permit the conditional distribution of movement rates to be quantified and it can be used to make probability statements that may be useful in practice; for example optimal foraging theory. A potentially important practical application of the stable analysis is movement strategies analysis. A strategy that includes both the Lévy stable walks

and the Lévy flights are thought to optimize foraging. Kawai and Petrovskii [65] show in movement ecology applications that stable models- because they capture both skewness and heavy tails in movement rates- perform considerably well than models based on power law distribution or the empirical distribution. Further, due to the analytical tractability of the stable distributions, it is possible to use the stable models to construct optimal search strategies for animals within the framework of movement ecology. In animal movement studies, where rare steps in the upper tail of the distribution drive search optimality, it appears promising to use the stable regression models developed above as an input into constructing an optimal search strategies for animals that help understand the relationship between elephant herds and their habitats.

3.5 Conclusion

We have described the theoretical justification for the use of stable law regression models and Student's t regression models in analysing animal movement data. To be useful in practice, a statistical model of the speed of animal movement should capture asymmetry, the heavy tails implied by the importance of extreme events and allow the speed to be conditioned on a vector of explanatory variables. Recent advances in the statistical theory of non-symmetric density functions and their estimation make it feasible to estimate statistical models based on the stable law and the student's t distribution. It is also possible to estimate Student's t regression models using standard maximum likelihood techniques. The t regression model has one demerit in that it does not capture the skewness property of the data though based on the AIC values, it is superior to the stable law regression model

Despite several studies detailing analogous statistical approaches, application of such models to GPS tracking is limited due to computational difficulties [63] and lack of adequate data rich in covariates in ecology. One limitation of the Stable law regression model is the lack of convergence when the values of the parameters

α and β are at the boundary. The Student's t regression model is particularly appealing in ecology where the data are characterized by heavy tails and where we are interested in conditional distributions. Unlike Stable law regression model, the Student's t regression models does not account for skewness and is not in the domain of attraction of sum of independent and identically distributed random variables. However, a skewed Student's t regression model is suggested in future studies to assess the effects of covariates on heavy tailed and skewed data in movement ecology. Advancement in technology and availability of computational softwares have made the implementation of these models easier.

We have identified several key areas to be pursued. Some of these areas are straightforward, such as increasing the number of explanatory variables, allowing the parameters of stable distribution to vary with the explanatory variables in a generalized linear model (GLM) framework. Diagnostic testing and model checking tools need to be developed to check the adequacy of the fitted models. Similarly, the random effects can be included in the model to explain herd variability between herds. While each is an extension of the simple models demonstrated they entail estimation of many more parameters.

Our empirical application demonstrates the importance of modeling explicitly the asymmetries and heavy tails that characterize animal movement linear metrics (step length or speed) if one is to make the accurate probability statements required to manage the environmental fluctuations. Typically, elephant movement is not predictable as it is difficult to determine analytically when a step starts and ends. However quantifying the distribution of the movement rate conditional on specific-environmental variables is one way to describe the effects of the drivers on the elephant movement. The stable regression models appears to be a useful tool for quantifying this relationship and it may have an important and practical application in assessing the value of artificial incentives in wildlife management especially on private game ranches in South Africa.

CHAPTER 4

Approaches for analysing heavy tailed and skewed animal movement: application to elephant movement time series data

Abstract

Many time series models in movement ecology are often based on the assumption that the random variables follow a Gaussian distribution. It is well known that empirical data sets have infrequently occurring rare events and cannot be modelled with the Gaussian distribution. Such data sets can be described by ARMA-GARCH models with an appropriate underlying distribution that can cater for skewed and heavy tailed distributions. In this study we compare four ARMA-GARCH models as potential candidates for modelling heavy tailed and skewed time series data of animal movement step lengths. Estimation difficulties have however hindered the use of stable Paretian distributions among practitioners in fields such as movement ecology. Four distributional hypothesis are tested: the Gaussian, student's t , skewed student's t and the stable Paretian distributions and fitted to time series step lengths of single female elephant herd. We found that the stable Paretian distributed ARMA-GARCH model performs better than the Gaussian and student's t distribution ARMA-GARCH models.

Keywords: Stable Paretian ARMA-GARCH, Heavy tails, Volatility clustering, animal movement, Ecology, stable Paretian distribution.

4.1 Introduction

Many ecological models of animal movement rely on the assumption that step lengths follow a Normal distribution [52]. However, animal movement data often depart from the Normality assumption, in that their marginal distributions are heavy tailed and skewed. Animal location data are being collected at an increasing high resolution (0.1-30 minutes sampling intervals of large mammals are now quite common) over several seasons. For such data, first order autoregressive models that assume normal distribution may miss important properties of the data. The use of stable distributions in movement ecology has been mentioned in a few studies via the works of Viswanathan et al. [132] and emphasized by Bartumeus [5]. Since then, the power law a simpler variant of the stable distribution, has been used to model the step lengths of animal movements [23, 34–37]. Such an approach ignores the temporal correlation potentially leading to spuriously precise estimates of parameters [see also 36, for a discussion of the statistical issues with Lévy approaches]. Stable distributions in general have not gained much attention in movement ecology due to lack of closed form density and computational difficulties [63]. Biological signals modelled using these models can provide a useful approach for statistically detecting and characterizing the temporal dependency in animal movement data.

While different heavy tailed distributions such as the Weibull and Log-normal distribution can be used for modelling animal movement metrics, stable distributions are preferred due to the generalized Central Limit Theorem [84]. According to this theorem, the limiting distribution of a sum of i.i.d random variables is stable [13, 92]. A notable feature of the stable distribution is the slow decay in the tails so that large rare steps are captured naturally. All the stable distributed variables

have the useful property of additive invariance adding two or more independent realisations yields another variable of this type. The normal distribution is a limiting case of the generalised central limit theorem when $\alpha = 2$ [143]. These and many other properties makes stable distribution more tractable to the analysis of animal movement data. Thus they have gained popularity in modelling heavy tailed data [21, 143] in finance and economics and a few cases in animal movement [63, 64]. Animals typically move in a non-random manner with short clustered steps when foraging or resting and long rare steps when migrating or in search of scarce food resources like water resulting to heavy tailed and skewed step lengths. These data sets are autocorrelated a property which most ecologists tend to eliminate through de-trending or filtering [22]. However, [26] notes that such autocorrelation is an intrinsic property of biological data thus eliminating it could reduce the relevance of ecological studies. Traditional analysis often assume uncorrelated or weakly correlated temporal structure in animal movement time series constructed using sequential location data [26].

The conditional distribution of step length is assumed to be Normal in Autoregressive models. However, this model specification is not proper for many animal movement time series because of the leptokurtic nature of the data [5]. Therefore, distributions such as the student's t, Log-normal and the Laplace have been suggested [84]. Previous studies have shown that movement step length data set is heavy tailed and skewed, but little has been done to evaluate time series models that capture both the skewness and heavy tails properties. It is this gap that this study seeks to contribute by proposing an ARMA-GARCH model with stable Paretian distributed innovations to analyse animal movement data sets.

Recent methodological developments have improved the ability of statistical models to handle the biological complexity of animal movement data [91]. This paper examines the statistical properties of the step lengths, using a set of symmetrical and asymmetrical time series. These tools are applied to high-frequency telemetry data within the framework of movement ecology.

The remainder of this chapter is organized as follows. Section 4.2 describes the models used in this chapter. In section 4.3 we explain the statistical properties of animal movement step lengths data used in this paper and presents the initial findings. We also discuss parameter estimation and compare the goodness-of-fit of the fitted models. Section 4 summarizes the findings and gives the concluding remarks.

4.2 Methods and Materials

4.2.1 Stable autoregressive models

In this section, we describe time series models that can be adopted to analyse animal movement step length time series data in ecology. We also suggest an extension to a new class of time series models based on the stable distribution whose density has the capacity to capture both skewness and heavy tailed nature of animal movement data.

Let y_t be an ARMA(p,q) process of order p and q given mathematically by the equation:

$$y_t = a_0 + \sum_{i=1}^p a_i y_{t-i} + \sum_{j=1}^q b_j u_{t-j} + u_t \quad (4.1)$$

where a_i and b_j are the model parameters and $i = 0, \dots, p, j = 0, \dots, q$ are the orders of the autoregressive and moving average process respectively, u_t is the error term and y_t is the step length of an individual animal at time t.

The generalized autoregressive heteroscedasticity (GARCH) models can be used to model animal movement data by expressing the conditional variance as a linear function of past information, allowing the conditional heteroscedasticity of step lengths. Animals alternates between fast movement while migrating and slow movement while foraging or resting which can be visualized as small clustered steps with long rare steps whose outcome is heavy tailed and skewed data. The assumption that the residual terms ϵ_t in equation (4.1) are i.i.d is not valid in

ecological time series data. The width of the error terms is typically clustered and depends on the time itself. This property in finance is called volatility clustering. Bollerslev [12] introduced a class of GARCH(r,s) models which models this property. The biological clustering is denoted by γ_t and can be modelled by the expression:

$$u_t = \gamma_t \epsilon_t$$

with the recursive relation

$$\gamma_t^2 = \omega + \sum_{m=1}^r \psi_m u_{t-i}^2 + \sum_{n=1}^s \phi_n \gamma_{t-j}^2, \quad u_t | U_{t-i} \sim N(0, \gamma_t^2) \quad (4.2)$$

where $\omega > 0$ is the constant coefficient of the variance equation, ϵ_t is a sequence of i.i.d random variables with mean zero and variance 1, $\psi_m > 0$ and $\phi_n \geq 0$ are the ARCH and GARCH parameters respectively and $\sum_{m=1}^r \psi_m + \sum_{n=1}^s \phi_n < 1$, $m = 1, \dots, r$ and $n = 1, \dots, s$ are the lags of the ARCH and GARCH effects and $u_t | U_{t-1}$ is the history of the movement. Here it is understood that $\alpha_i = 0$ for $m > r$ and $b_j = 0$ for $n > s$. The latter constraint on $\alpha_i + \beta_j$ implies that the conditional variance of u_t is finite, whereas its conditional variance γ_t^2 evolves overtime [87]. Thus, volatility clustering depends on its previous values and on the squared residuals ϵ_t^2 [12, 21].

While a GARCH model can describe volatility clustering also known as Lévy flight in movement ecology, one needs a statistical model that can adequately estimate the conditional mean of the animal tracking data. If the prediction of the conditional mean is not reasonable, the construction of the conditional volatility clustering is not possible [89]. We combine the equations (4.1) and (4.2) to form an ARMA-GARCH(p,q,r,s) model defined by equation:

$$y_t = a_0 + \sum_{i=1}^p a_i y_{t-i} + \sum_{j=1}^q b_j u_{t-1} + \epsilon_t \gamma_t \quad (4.3)$$

where γ_t is given by equation (4.2) and ϵ_t is assumed to have a mean zero and variance 1. For ϵ_t , we assume various probability density functions to test different types of ARMA-GARCH models. For example, the normal, the Student's t, the

skewed Student's t and the asymmetric stable distributions can be used. In the Gaussian case, $\epsilon_t \sim N(0, 1)$, and in the student's t case, $\epsilon_t \sim t(v)$ where v is the degrees of freedom and the scale parameter is equal to one. We denote the Gaussian model as Normal-ARMA-GARCH, the student's t model as t -ARMA-GARCH and the skewed student's t model as st -ARMA-GARCH model. We refer to these models as the conventional ARMA-GARCH models. An important property that makes ARMA-GARCH process attractive to models of animal movement ecology data is that its tail is heavier than that of a normal distribution which allow a time varying volatility clustering of animal movement step lengths. The clustering of animal movement step lengths is known as Lévy flight pattern in ecology [63]. It is expected that the skewed student's t ARMA-GARCH model which has slightly heavier tails will be able to capture rare large steps better than the student's t and the normal ARMA-GARCH process. However, models that can better captures both heavy tails and skewness are still of interest.

It has been argued in the literature [5, 23, 63] that a model that captures both the heavy tails and skewness inherent in ecological data is the four parametric family of stable distributions. However, the stable density for a random variable X does not have a simple mathematical description, so it is instead represented using its characteristic function given by

$$E[\exp(itx)] = \begin{cases} \exp\left(-(\gamma|t|)^\alpha \left[1 + i\beta(\text{sign } t) \tan \frac{\pi\alpha}{2} (|t|^{1-\alpha} - 1)\right] + i\mu t\right) & \alpha \neq 1 \\ \exp(-\gamma|t| \left[1 + i\beta(\text{sign } t) \frac{2}{\pi} \ln|t|\right] + i\mu t) & \alpha = 1 \end{cases} \quad (4.4)$$

where $\alpha \in (1, 2]$ is the tail index of stability, $\beta \in [-1, 1]$ is the skewness parameter, $\gamma > 0$ is the scale parameter and $\mu > 0$ is the location. The stable distribution is symmetric about the location μ when $\beta = 0$ and the characteristic exponent α determines the thickness of the tails [94]. When $\alpha = 2$, the underlying stable Paretian distribution is Normal distribution: $N(0, 2)$, with finite moments of all orders [143]. As α tends to 0 from 2, the tail of the stable distribution becomes increasingly fatter than the Normal case and thus, the scale parameter γ is used as the analog of the variance [21]. In this paper we note that the k^{th} moments

of the stable distribution exist if $k < \alpha$ and we assume that $E|U| < \alpha$. In other words, we restrict α to be in the interval $(1, 2]$ through out.

Assuming that step lengths of animal movement follow a stable paretian distribution, we define the volatility model of the stable ARMA-GARCH process y_t denoted by *S-ARMA-GARCH* and given by the equation

$$\gamma_t = \psi_0 + \sum_{i=1}^r \psi_i |u_{t-i}| + \sum_{j=1}^s \phi_j \gamma_{t-j}. \quad (4.5)$$

where $u_t = \gamma_t^{1/\alpha} \epsilon_t$ and ϵ_t is specified to have a stable density [88]. The *S-ARMA-GARCH*(p, q, r, s) process defined by equations (4.1) and (4.5) with $\alpha \in (1, 2]$ has a unique strictly stationary solution if $\alpha_i > 0, i = 0, 1, \dots, r, \beta_j > 0, j = 1, \dots, s$ and the measure of volatility persistence, $V_s = \lambda_{\{\alpha, \beta, \gamma, \mu\}} \sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j$, satisfies $V_s \leq 1$ [21, 88], where

$$\lambda_{\alpha, \beta, \gamma, \mu} = E[\epsilon_t] = \frac{2}{\pi} \Gamma \left(1 - \frac{1}{\alpha} (1 + \tau_{\alpha, \beta, \gamma, \mu}^2)^{1/2\alpha} \cos \left(\frac{1}{\alpha} \arctan \tau_{\alpha, \beta, \gamma, \mu} \right) \right), \quad (4.6)$$

and $\tau_{\alpha, \beta, \gamma, \mu} = \beta \tan \frac{\pi\alpha}{2}$. If V_s is strictly less than one, this implies a conditional volatility equation where the impact of long rare steps dies out over time [88]. In practice, the estimated volatility persistence, V_s tends to be quite close to one for highly volatile series which would offer a reasonable description of the animal tracking data.

4.3 Statistical properties of step lengths time series

To illustrate the utility of the reviewed methods, we analyse a time series data set of an individual female elephant herd (*Africana loxadonta*) collected via radio-telemetry ([see, 11, 131, for details on data collection] for details on data collection) from May 2006 to April 2009. Animal capture was undertaken using chemical immobilisers by South African National parks staff, following approved ethical

procedures (*University of KwaZulu Natal, Ref.009/10/Animal*). These elephants inhabited Kruger national park isolated by paved roads and rivers in South Africa. *Africana loxadonta* is an elephant species mostly found in Southern of Africa.

One of the objectives of the radio-telemetry study was to assess the effects of autocorrelation in the presence of heavy tails and skewness in the step lengths of elephant movement. Spatial coordinates of the elephant location were obtained using global positioning system (GPS) receivers and differential correction techniques were applied to improve their accuracy [11]. These animals were typically located at 30 minutes intervals between May 2006 and April 2009 yielding 35272 observations for that herd. Step lengths (linear metrics) of animal movement were derived using the tracking tool in ArcGis 10.0 and the resulting ArcView shapefile was exported as a tab delimited text file for input to R (R Development Core Team 2008) environment in which we perform the remaining analysis.

To summarize the statistical properties of step lengths, (i) we carry out a visual inspection of the time series graphs. (ii) we compute the relevant descriptive statistics. (iii) we carry out two unit root tests: the augmented Dickey-Fuller (ADF) and the KPSS tests to decide whether the series is stationary. (iv) we discuss the autocorrelation of step length and test the presence of non-linearities in the data.

Figure 4.1 gives the time series plot of step lengths of herd AM108 and shows the long run movement pattern of elephant spanning a period of 3 years. The plots show relatively large number of high spikes in the step lengths which shows the presence of clustering of steps and the possibility that the step length innovations of the herd are non-Gaussian.

Table 4.1 summarizes the basic descriptive statistical properties of the data. The mean step lengths is 213 meters and the median is 145 meters signifying the presence of high skewness in the data. The step lengths of the elephant herd are asymmetric as reflected by the positive skewness which implies the presence of large rare steps in the tracking. The kurtosis is also positive and higher than that of a standard Gaussian distribution which is 3, which implies the presence of heavy

tails for empirical distributions. The Jarque-Bera tests of normality has a value beyond the critical value at 5%, indicating that the normal distribution hypothesis for the empirical step lengths x_t , is rejected. We computed the ADF and the KPSS tests as shown in Table 4.2 to determine the stationarity of elephant movement step lengths data. We reject the the ADF unit root hypothesis at 5% level of significance and the KPSS stationarity hypothesis at 10% level of significance, we conclude that the elephant movement is non-stationary. For the purpose of this study, and according to the descriptive statistic obtained for the herd AM108, there is evidence for higher order lags but we have confined our to the autoregressive models of order 1 Gurarie et al. [52]. The first order autocorrelation of the AR of order 1 is positive which implies that the elephants have short memory and a small of movement step lengths is predictable. Hence the Lévy flight foraging hypothesis does not hold strictly as stipulated in the literature [5, 23, 36]. This behaviour could be attributed to environmental heterogeneities such as uneven distribution of food patches and water sources. In order to determine our mean equation, we

TABLE 4.1: Descriptive analysis of two female elephant movement step length in Kruger National Park collected every 30 minutes

Stats	Observations	Minimum	Maximum	Mean	Median
	35272	5.039	2616	213	145
Stats	Stdev	Skewness	Kurtosis	Jarque-Bera	p-value
	226	2.25	7.39	164028.9	0.000

note that the lag value which has the minimum AIC value corresponding to the ARMA terms was chosen to model the step lengths, thus supporting the choice of our model.

Even though the series of step lengths seems to be correlated over time, the Ljung-Box statistic for up to 20th order serial correlation is highly significant at any level

TABLE 4.2: KPSS tests and Augmented Dickey-Fuller

Test	statistic	lag	p-value
KPSS	0.3567	43	0.096
ADF	-31.378	32	0.010

for the animal movement linear metrics, suggesting the presence of strong non-linear dependence in the data. From the time series plot in Figure 4.1, volatility clustering of steps are clearly visible suggesting the presence of conditional heteroscedasticity which suggests that non-linearities in the tracking data must enter through the variance of the processes. Thus the non-linear dependencies can be explained by the presence of conditional heteroscedasticity and one might represent this behaviour using a model in which successive conditional variances of step lengths are positively autocorrelated. We note that such specification is consistent with the optimal foraging hypothesis of animal movement.

The Lagrange multiplier (LM) is used to test formally the presence of conditional heteroscedasticity and the evidence for ARCH effects. The LM test for a first order linear effect is shown in Table 4.3 which implies that the herd AM108 exhibit ARCH effects and the null hypothesis is rejected. In the next section we present the results of the ARMA-GARCH models.

TABLE 4.3: Step Length autocorrelation

Test	Statistic	p-Value
Ljung-Box	98.02398	0.00
LM Arch	133.6647	0.00

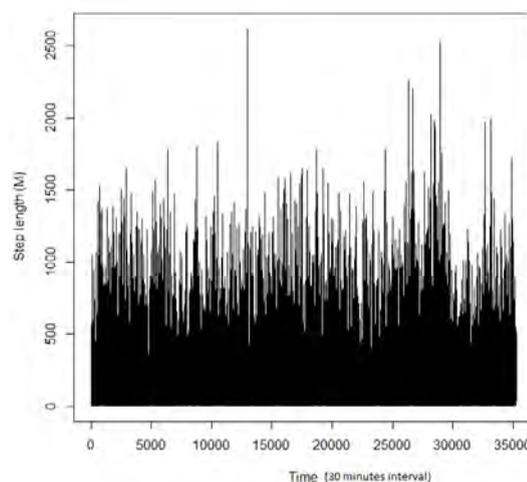


FIGURE 4.1: Time series plots of the step lengths of Herd AM108 of GPS-derived telemetry data collected from Kruger National Park South Africa (May 2006 -April 2009) (30minutes time intervals)

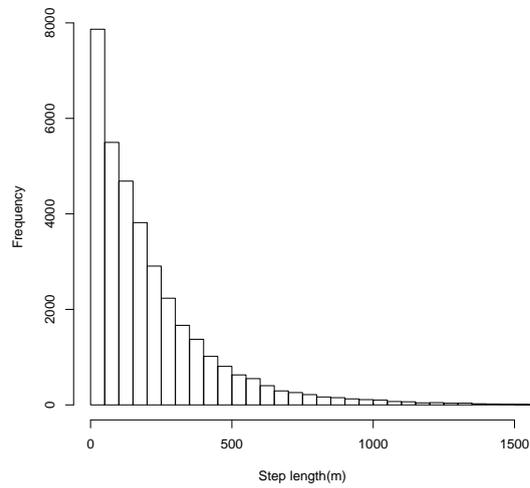


FIGURE 4.2: Histogram plot of the step lengths of Herd AM108 GPS-derived telemetry data collected from Kruger National Park South Africa (May 2006 -April 2009)

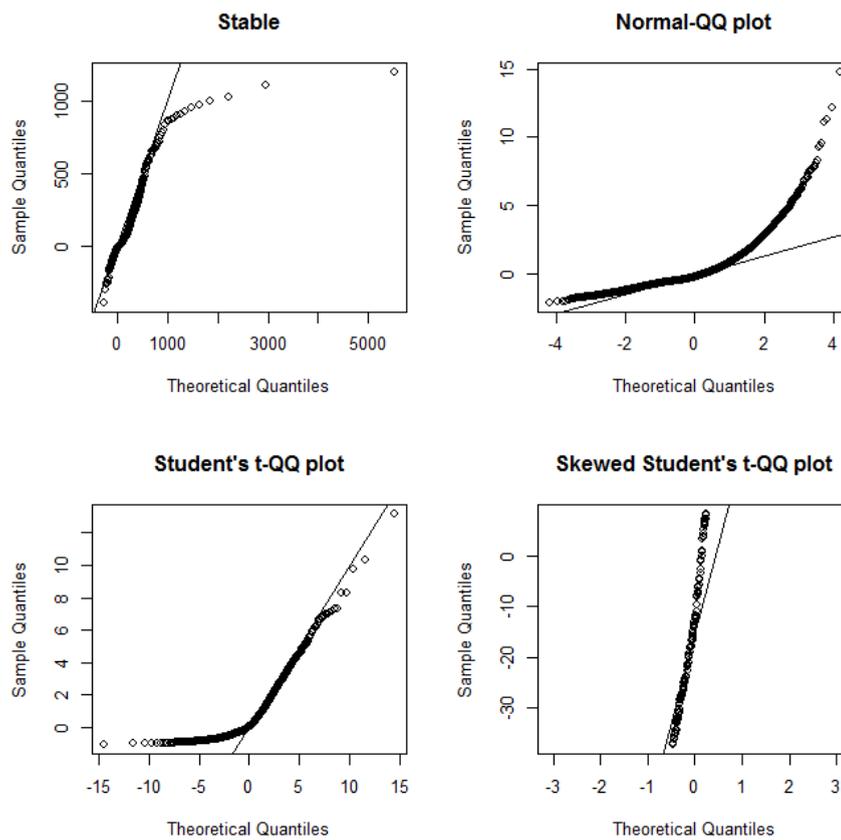


FIGURE 4.3: QQ plots of standardized residuals from an ARMA-GARCH(1,1,1,1) to step lengths of elephant movement.

4.4 Parameter estimates and interpretation

The maximum likelihood approach was used in estimating the parameters and we assumed the error term, ϵ_t to be either Normal, Student's t, skew Student's t or stable Paretian distributed and the γ_t satisfies ARMA-GARCH recursions 4.3 and 4.5, respectively. We follow the algorithm described by Mittnik et al. [86] who approximates the stable Paretian density function through the Fast Fourier Transform (FFT) of the characteristic function.

To compare the fitted models, we employ three likelihood based goodness of fit criterion described by Mittnik and Paoella [87]. The first is the maximum log-likelihood value obtained from ML estimation (MLE). This value allows us to judge which model is more likely to have generated the data. The second method is the bias-corrected Akaike [3] information criteria (AICC) and the third method is the Schwarz Bayesian criteria (SBC, [112]):

$$AICC = -2 \log L(\hat{\theta}) + \frac{2n(k+1)}{n-k-2}, \quad SBC = -2 \log L(\hat{\theta}) + \frac{2 \ln(n)}{n}, \quad (4.7)$$

where $\log L(\hat{\theta})$ is the maximum log-likelihood value, n is the number of observations and k is the number of the model parameters. The model with a lowest value for these information criteria is judged to be preferable [21]. Table 4.4 presents the estimation results for the stable Paretian, normal, Student's t and skewed Student's t distribution ARMA-GARCH(1,1,1,1) models for herd AM108. The estimates of the intercept which is denoted by a_0 is statistically significant for all the models except for the stable ARMA-GARCH model. The autoregressive parameter denoted by a_1 is positive and statistically significant for all the four models implying that there is a positive autocorrelation in the data. This indicates that the AR(1) model is able to model the conditional mean of animal movement linear metrics as suggested by [52].

TABLE 4.4: Maximum likelihood estimates and goodness of fit statistics of Stable, normal, student's t and skewed student's t distribution ARMA-GARCH(1,1,1,1) of elephant movement step lengths

	Stable Paretian			Normal			Students' t			Skewed student t		
	Estimate	std.error	p-value	Estimate	Std.Error	p-value	Estimate	Std.Error	p-value	Estimate	Std.Error	p-value
a_0	2.411	2.419	0.319	34.514	9.4351	0.000	17.3791	4.2481	0.000	45.6341	7.5732	0.000
a_1	0.636	0.038	0.000	0.765	0.0631	0.000	0.7589	0.0410	0.000	0.6651	0.0462	0.000
b_1	-0.350	0.055	0.000	-0.382	0.1000	0.000	-0.3252	0.0630	0.000	-0.3753	0.0591	0.000
ω	1.000	1.748	0.567	7852.4	1761.51	0.000	5322.16	2656.35	0.045	8782.56	3844.38	0.022
ψ_1	0.267	0.035	0.000	0.307	0.0824	0.000	1.0000	0.4291	0.020	0.3586	0.1382	0.009
ϕ_1	0.534	0.052	0.000	0.318	0.1073	0.003	0.4418	0.0968	0.000	0.3335	0.1219	0.006
β	0.980	0.006	0.000							1.9218	0.1344	0.000
α	1.068	0.056	0.000				2.3935	0.2016	0.000	2.8265	0.4018	0.000
loglik ^a	222701			233782			227900			223671		
AICC ^b	445420			467577			455816			447360		
SBC ^c	445402			467563			455800			447342		

^a "LLH" refers to the maximum log-likelihood value

^b "AICC" is the bias-corrected Akaike Information Criterion

^c "SBC" is the Schwartz Bayesian Information Criterion

Both ψ_1 and ϕ_1 are highly significant which implies that there is persistent volatility clustering in the step length time series data of the herd. These results indicate that as the animals increase their step lengths during the search for scarce food resources or migration, so does volatility. Since the step length is the displacement between two positions covered by the animal within a given period of time is assumed to be a proxy measure of animal movement pattern or unobserved information flow. We conjecture that as the food resources becomes more scarce or decreases, the volatility also increases. This clearly makes sense because the animal is in the search mode state.

From the stable distribution parameterisation, the estimates of the shape parameter α is statistically significant at 5% and less than 2 which means that the step length time series of the herd have heavy tailed patterns and thus rejecting a Gaussian hypothesis. The estimates of the skewness parameters β are positive and significant which means that the step length time series are skewed. These findings agree with the results of the descriptive analysis which indicates that the data set is highly skewed and leptokurtic as shown in Table 4.1.

One important characteristic of these estimates relates to the connection between the skewness and kurtosis of a normal distribution, the asymmetry and the tail index parameter of the stable distribution [87]. Based on the sample skewness and kurtosis obtained in Table 4.1, we expect a fit with the stable distribution will probably give stable skewness with the same sign and with tail index less than 2. Indeed this happens in Table 4.4. However, there no quantitative connection between the sample kurtosis and the population kurtosis as it is undefined under the stable distribution [87]. The real difficulty in analysing the tracking data in ecology is that the exact distribution of the underlying population is not known and sampling variations make this problem even more complex [63].

Based on the log-likelihood, AICC and the SBC results of Table 4.1, we find that the ARMA-GARCH(1,1,1,1) model with stable distributed innovations performed better than the normal followed by the skewed Student's t and Student's t ARMA-GARCH models respectively. We also note that the skewed Student's t performed better than the Student's t and the normal ARMA-GARCH models respectively. This implies that the asymmetric models are better suited for analysing heavy tailed and skewed data of animal movement like this one. This finding indicates that asymmetric models perform better with heavy tailed and skewed data. The results further confirm the advantage of using the stable distributions that can take into account the non-normal nature of the time series data examined. However, we point out that despite the computational challenges highlighted by [63], stable distribution through the index of stability and skewness parameter can capture more properties of animal movement data and provide more biological insights than the conventional normal models.

4.5 Conclusion

We investigated the empirical 30 minutes time series data of elephant movement step lengths collected from Kruger national park using both asymmetric and symmetric distribution in ARMA-GARCH models. We note that models that appropriately capture the heavy tails and skewness properties of animal movement and other ecological applications remain largely unexplored. Stable Paretian distribution has rich properties that can appropriately capture the skewness and heavy tailed nature of animal movement data better than the conventional normal distribution. Using asymmetric models, we are able to extract biological insights from a rich statistically challenging empirical GPS animal location data which could otherwise be lost if the data was subjected to treatment to remove the autocorrelation. These models provide an objective way of modelling animal movement data sets without loss of biological information due to data treatment as proposed by Cushman et al. [22]. By incorporating the Biological signals through stability index and skewness parameter of the stable distribution, the model provides an approach of statistically detecting and characterizing the temporal dependency in animal movement facilitating model development.

However, stable Paretian time series models are computationally challenging and no well developed softwares have been implemented. We note that the skewed student's t distribution can be a better alternative as it comes second in goodness of fit to the stable Paretian model. Models of higher order lags can be of interest as a future research. We also note that our models can be extended through the random coefficient autoregressive models to investigate the effects of several herds. We leave this and other considerations for future investigation.

CHAPTER 5

Application of circular regression models to animal movement data

Abstract

In a number of biological and physical sciences studies, a set of techniques have been developed to analyse the relationship between the circular and linear data derived from the geographical positioning system (GPS) telemetry to describe animal movement. Yet, many of the models used by ecologists do not provide a link between the circular and linear variables. This chapter demonstrates the application of the circular-linear regression in describing such a relationship. We describe numerical methods of obtaining maximum likelihood model parameter estimates. We discuss the technical limitations of the model through simulation and application to real elephant movement data with covariates collected from Kruger national park, South Africa. These results provide a new statistical paradigm for understanding the need to landscape features in elephant and similar animal models and evolutionary forces driving unpredictable.

Keywords: Animal movement, circular statistics, von Mises, regression model, GPS tracking, turn angles.

5.1 Introduction

Advances in GPS tracking technology is revealing new insights regarding animal movements [107, 130]. Tracking devices tagged on animals are becoming smaller in size yet larger in memory capacity yielding huge datasets [45, 130]. GPS metrics(circular or linear) of animal movement can be derived from the relocation data. A large body of literature exists on the analysis of linear metrics (step length, speed and mean square displacement) [see references in, 35, 36] but little is known about the circular metrics due to lack of computational software. Accurate and objective interpretation of biological data on movement patterns of animals is needed because conclusions may have management implications [98, 99]. For instance, the analysis of turn angles data can reveal animal space use and the ecosystem drivers of movement [27], and efficient foraging strategies [134] that can facilitate applications in conservation ecology [27, 98, 99].

Circular data can also be visualized as being distributed on the circumference of a unit circle in the range of 0 to 2π [78]. Handling such data creates difficulties due to the restriction of support to the unit circle, the sensitivity of descriptive and inferential statistics to the starting point on the circle. There exists a substantial literature on circular data but broadly, it is confined to descriptive statistics and limited to inference for simple univariate models [7]. For instance, [17] reviewed the statistical methods for analysing angular ecological data and found that standard statistical methods were not appropriate for circular data.

Many ecological models have relied on circular statistics to develop complex models. For instant the authors in [119] and [124] fitted von Mises distribution to turn(move) angle data and come up with a set of non-linear regression models relating animal movement to landscape features orientation. Cain [16] found that von Mises distribution provides good statistical fit to insect turn angles and clonal plant branching angles. According to behavioural and foraging theory, magnitude of the turn angle and the step lengths depends on fundamental movement elements [45] such as habitat type, distance to the river (water point), landscape structure, quality of food patches etc. Elephants locate their home ranges in areas of high food density, water source and low human disturbance [24] while [114] argued that the location of artificial water points can have negative impact on the bio-diversity by putting pressure on habitats closure to them. Landscape features are physical barriers such as rivers and roads which constrain animal movement by preventing access to adjacent patches and impeding dispersal [131]. Redfern et al. [103] found correlation between distance to water and elephant distribution across the poor and rich forage habitat patches in KNP. To model the effects of landscape feature on animal turning behaviour a tool that relates the circular and linear variables is needed.

We apply circular statistics, a specific area of methods for analysing data arising from animal movement studies. The primary motivation of our approach arises from the rich data set of

elephant turning angle and landscape features covariates data of an individual elephant collected from Kruger National park using GIS telemetry tags. The turn angles of the animal are derived from the location and the covariates extracted. Using a circular-linear regression model [42], we model the turn angles of the elephant movement as a function of landscape and climatic covariates. Using this model, we can determine the mean orientation of the elephant and identify important predictors of animal orientation.

This chapter is organized as follows. First, in Section 5.2, we describe the von Mises distribution and its properties. Section deals with the von Mises regression model, parameter estimation, confidence intervals and model checking methods. In Section we discuss the limitations of the model based on simulated data and real elephant movement data with covariates. In Section 5.6 we present the results of the fitted model and there description. Finally we end the chapter with discuss and conclusion of the findings.

5.2 The Von Mises Distribution

One of the most important distributions for circular data is the von Mises distribution, represented by $vm(\mu, \kappa)$, whose probability density function (pdf) is given by

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\} \quad (5.1)$$

where $[2\pi I_0(\kappa)]$ is the normalizing constant, $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero. The parameter μ represents the mean direction while κ is the concentration parameter of the distribution [85]. These parameters must satisfy $\mu \in [-\pi, \pi)$ and $\kappa \in [-\infty, \infty)$.

The general modified Bessel function of the first kind and order p is defined by

$$I_p(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp[\kappa \cos(\phi - \mu)] d\phi$$

Note that $vm(\mu, \kappa)$ and $vm(\mu + \pi, -\kappa)$ have the same distribution. For this model, the values of κ are set to be non-negative, and the ranges of θ and μ are $[-\pi, \pi]$. The von Mises distribution is unimodal with two parameters μ and κ , and is symmetrical about mean direction $\theta = \mu$. The larger the value of concentration parameter κ , the denser the clustering around the mean direction μ . For $\kappa = 0$, the von Mises distribution tends to the uniform distribution. As $\kappa \rightarrow \infty$ it becomes concentrated at the point $\theta = \mu$.

From a random sample $\theta_1, \theta_2, \dots, \theta_n$, we can calculate C and S according to equation 5.1 so that the maximum likelihood estimate of κ is given [25] by the solution from

$$A(\kappa) = (C^2 + S^2)^{1/2}$$

where the function $A(\kappa) = I_1(\kappa)/I_0(\kappa)$ is defined with $I_1(\kappa)$, the modified Bessel function of the first kind (order 1; [85]). Meaning that we need

$$\kappa = A^{-1}(C^2 + S^2)^{1/2}$$

This is sufficiently unwieldy that tabular lookup was used historically and we are now advantaged by software [46].

Unfortunately, the MLE here is also biased for finite sample [129]. Defining

$$R^2 = [nI_1(\kappa)/I_0(\kappa)]^2$$

Schou [111] and Batschelet [7] tabulated unbiased values of $\hat{\kappa}$, which is the solution for $A(\kappa) = RA(R\kappa)/n$ when $R^2 \geq n$. These authors recommends that $\hat{\kappa} = 0$ when $\hat{R} < n$. A very useful set of approximations is given by [43] given by

$$\hat{\kappa} = \begin{cases} 2(R/n) + (R/n)^3 + 5(R/n)/6 & \text{for } R < 0.53n, \\ -0.4 + 1.39(R/n) + 0.43(1 - R/n) & \text{for } 0.53n \leq R < 0.85n, \\ 1/((R/n)^3 - 4(R/n) + 3(R/n)) & \text{for } R \geq 0.85n \end{cases}$$

which needs to be adjusted for small sample size and small R/n [43, 129]:

$$\hat{\kappa} = \begin{cases} \max(\hat{\kappa} - \frac{2}{n\hat{\kappa}}) & \text{for } \hat{\kappa} < 2, \\ \frac{(n-1)^3 \hat{\kappa}}{n^3 + n} & \text{for } \hat{\kappa} \geq 2 \end{cases} \quad (5.2)$$

The MLE for the mean direction is the value $\hat{\mu}$ that satisfies the equation $\hat{\mu} = C/R$, $\hat{\nu} = S/R$, with C, S and R defined as above. Upton [128] gives a likelihood ratio test statistic for $H_0 : \mu = \mu_0$ verses $H_0 : \mu \neq \mu_0$, provided that κ is not large, that rejects H_0 if

$$R^2 > X^2 + \frac{(2n^2 - X^2)}{4n} Z_\alpha, \quad X^2 = [R \cos(\mu_0 - \hat{\mu})]^2$$

for an a significance level α . Upton [129] extends his hypothesis test to derive two confidence intervals for μ :

$$\begin{aligned} \hat{\mu} \pm \cos^{-1} \left[\left(\frac{4nR^2 - 4nR^2 Z_\alpha}{4nR^2 - R^2 Z_\alpha} \right)^{1/2} \right] & \text{ for } R \leq 0.9n, \\ \hat{\mu} \pm \cos^{-1} \left[\left(\frac{n^2 - (n^2 - R^2) \exp(Z_\alpha/n)}{R^2} \right)^{1/2} \right] & \text{ for } R > 0.9n. \end{aligned}$$

Under the normal approximation, this reduces to the simpler form

$$\mu \pm (Z_\alpha / (R\hat{\kappa}))^{1/2}$$

Finally, note that the R package `circular` [1] provides many of these calculations using the von Mises distribution.

5.3 Circular orientation model

In order to incorporate the effects of landscape features and environmental covariates in animal movement, we assume that the angular displacement of the animal follows a von Mises distribution with mean μ_i and concentration parameter κ given by the equation

$$f(y_i; \mu_i, \kappa) = [2\pi I_0(\kappa)]^{-1} \exp\{\kappa \cos(y_i - \mu_i)\}, \quad 0 \leq y, 2\pi \leq 2\pi, \kappa_i > 0 \quad (5.3)$$

where $I_p(\cdot)$ is the modified Bessel function of the first kind and order $p, p = 0, 1, \dots$. The positive parameter κ measures the concentration of the distribution: as $\kappa \rightarrow 0$, the Von Mises distribution converges to uniform distribution around the circumference, whereas for $\kappa \rightarrow \infty$, the distribution tends to the point distribution concentrated in mean direction.

For our purpose we assume the response y_i has a von Mises distribution with mean μ_i and the concentration parameter κ . We denote the vector of turn angles $Y_i = (y_1, \dots, y_n)'$ which corresponds to the mean orientation μ_i . We note that each y_i is a circular random variable while the corresponding vector of covariates X_i are not circular i.e., are continuous or categorical as with standard regression approaches. Hence in order to study the relationship between Y' s and X' s, we use a so-called circular-linear regression model [42, 48, 124]. This model assumes a monotonic link function that maps the explanatory variables to a circle. Though a variety of choices of link function can be used as discussed in Fisher and Lee [42] or Fisher et al. [43], a generalized linear model (GLM) for the mean turn angle $\mu_i = E[Y_i X_i]$ may be formulated as follows:

$$\mu_i = \mu_0 + 2 \arctan(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) \quad (5.4)$$

where μ_0 is an offset mean parameter representing the origin. If $Y_i^* = Y_i - \mu_0$ is taken as a surrogate response, then the corresponding mean turn angle is $\mu_i = \mu_i - \mu_0 = 2\arctan(\eta_i)$ with origin of 0° . This implies that $\tan(\mu_i^*/2) = \eta_i = X_i\beta$.

We note that equation 5.4 can incorporate covariates measured at both individual and herd levels. However, we focus solely on the individual level covariates as our model results in a single mean orientation model. Individual level predictors in our model would output a vector of mean turn angle for an individual, which though statistically correct, would not be easily interpretable in our setting. This challenge is analogous to covariates in a cross sectional data. Our goal is to find individual level characteristics that may be associated with animal turning pattern. We note that some of the individual level covariates can be a summary statistics, i.e., average distance to landscape features, environmental drivers throughout the mean orientations when the animal is moving, seasons, rainfall [27] and distance to roads [124].

One of the crucial assumptions of the von Mises distribution is that the turn angle data of the animal movement is unimodal [47]. We note that no conclusive proof in ecology regarding whether unimodality is a valid statistical assumption, as any evidence of multi-modality is based mostly on the evidence from very small samples of the data with no formal attempt to determine analytically whether more than one mode can be detected relative to the amount of variability in the data. SenGupta and Ugwuowo [113] derived recent advances in the analyses of directional data in ecological and environmental sciences. Otieno and Anderson-Cook [97] found Measures of preferred direction for environmental and ecological circular data. Tracey et al. [124] studied the effects of landscape features on rattle snake move angles and found that distance to the roads reduced the move angle of the rattle-snakes. One limitation of there model is that it cannot accommodate more than one covariate. Tracey et al. [125] used neural network approach and extended their single covariate model to a population model with several covariates in a semi-parametric approach. However, their model is computationally challenging. Fisher and Lee [42] studied the orientation of birds and discusses the challenges of the model.

5.3.1 Computing the mean parameter estimates

Our goal here is to estimate the regression coefficient $\beta = (\beta_0, \dots, \beta_k)'$ in equation 5.3 and derive consistent variance estimates of the parameter estimates. Our data is composed of turn angles of an individual elephants herd. The parameter estimates of μ , β and κ of equation 5.5 are obtained by maximizing the log-likelihood:

$$\log L = -N \log I_\circ(\kappa) + \kappa \sum_{i=1}^N \cos(\theta - \mu - 2 \arctan(x_i \beta)). \quad (5.5)$$

We have

$$\begin{aligned}\frac{\partial L}{\partial \mu} &= \kappa \sum_{i=1}^N \sin \{\theta_i - \mu - g(\beta x_i)\} \\ &= \kappa [\cos \mu \sum_{i=1}^N \sin \{\theta_i - g(\beta x_i)\} - \sin \mu \sum_{i=1}^N \cos \{\theta_i - g(\beta x_i)\}];\end{aligned}\tag{5.6}$$

$$\begin{aligned}\frac{\partial L}{\partial \beta} &= \kappa \sum_{i=1}^N \sin \{\theta_i - \mu - g(\beta x_i)\} \frac{2x_i}{1 + \beta^2 x_i^2} \\ &= 2\kappa H u,\end{aligned}\tag{5.7}$$

where $u_i = \sin \{\theta_i - \mu - g(\beta x_i)\}$, $H = \text{diag}(g'(x_1\beta), \dots, g'(x_n\beta))$, $g'(\beta x_i) = 2x_i/(1 + \beta^2 x_i^2)$, and

$$\begin{aligned}\frac{\partial L}{\partial \kappa} &= -N \frac{I'_0(\kappa)}{I_0(\kappa)} + \sum_{i=1}^N \cos \{\theta_i - \mu - g(\beta x_i)\} \\ &= -NR + C \cos \mu + S \sin \mu.\end{aligned}\tag{5.8}$$

From which the estimates are the solutions of the following equations:

$$X' H u = 0,\tag{5.9}$$

where $S = \sum_{i=1}^n \sin(\theta_i - g(x_i\beta))/n$, $C = \sum_{i=1}^n \cos(\theta_i - g(x_i\beta))/n$, $R = (S^2 + C^2)^{1/2}$ and $A(\kappa) = \frac{d}{d\kappa} \log I_0(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$.

We solve equation (5.9) using iteratively re-weighted least squares (IRLS) algorithm of Green and Williams [49] to obtain the parameter estimates μ , β and κ . In order to obtain the starting values of the IRLS algorithm, we assume the data are uncorrelated. We fit the von Mises distribution to the data and obtain the maximum likelihood estimates of μ and κ as described by Fisher and Lee [42]. We choose the MLE of κ as its starting value for our algorithm and the MLE of μ is not used further. From equations (5.7), the values of β that maximizes the log-likelihood, assuming independent observations, is equal to the value that maximizes the log-likelihood equation. So, we begin with an initial value for $\hat{\beta}$, then calculate S, C and R and hence $\hat{\mu}$ and $\hat{\kappa}$ using (5.9). These estimates are then used to solve (5.9) for an updated value $\hat{\beta}$. This procedure is repeated until convergence is achieved. The updating equations for $\hat{\beta}$ in the IRLS algorithm is

$$X' H^2 X (\hat{\beta}^+ - \hat{\beta}) = X' H^2 r,\tag{5.10}$$

From the theory of maximum likelihood, the asymptotic variance covariance matrix for the $\hat{\beta}$ is given by

$$\text{var}(\hat{\beta}) = \frac{1}{\kappa A(\kappa)} \left((X'H^2X)^{-1} + \frac{(X'H^2X)^{-1}X'h h'X(X'H^2X)^{-1}}{N - h'X(X'H^2X)^{-1}X'h} \right), \quad (5.11)$$

where h is a vector whose elements are the diagonal elements of H . The asymptotic variance of $\hat{\kappa}$ is equal to $(nA'(\kappa))^{-1}$ and the asymptotic circular variance of μ_0 is $(n(n-p)\kappa A(\kappa))^{-1}$ where $A'(\kappa) = \frac{\partial A(\kappa)}{\partial \kappa} = 1 - \frac{A(\kappa)}{\kappa} - A^2(\kappa)$.

5.3.2 Tests and confidence intervals

Statistical inference on random variables comprises estimation and testing procedures that allow to characterize the underlying distribution regardless of the variables nature and or/dimensions. Tests and confidence intervals can be derived utilizing the standard normal distribution. The confidence interval of β is derived as

$$\hat{\beta} \pm z_{N-1, \alpha/2} \sqrt{\widehat{\text{var}} \hat{\beta}};$$

which is

$$\hat{\beta} \pm z_{N-1, \alpha/2} \sqrt{\frac{1}{\kappa A(\kappa)} \left((X'H^2X)^{-1} + \frac{(X'H^2X)^{-1}X'h h'X(X'H^2X)^{-1}}{N - h'X(X'H^2X)^{-1}X'h} \right)}$$

Similarly, a test of $H_0 : \beta \leq 0$ versus $H_0 : \beta > 0$ would be to reject the H_0 if

$$\hat{\beta} / \sqrt{\frac{1}{\kappa A(\kappa)} \left((X'H^2X)^{-1} + \frac{(X'H^2X)^{-1}X'h h'X(X'H^2X)^{-1}}{N - h'X(X'H^2X)^{-1}X'h} \right)} > t_{N-1, \alpha/2}.$$

Tests and confidence intervals for α can be derived in the same manner. A confidence interval for σ^2 can be calculated as

$$\left(\frac{N\hat{\sigma}^2}{\chi_{N-2, 1-\alpha/2}^2}, \frac{N\hat{\sigma}^2}{\chi_{N-2, \alpha/2}^2} \right).$$

5.3.3 Model checking and diagnostics testing

Residuals are used to identify discrepancies between models and data. It is useful to establish residuals as contributions made by individual observations on the goodness of fit measures. One of the most useful in GLMs is the deviance $D(y; \hat{\mu}) = \sum_{i=1}^n d_i^2$ where

$$d_i = \pm \sqrt{2} (\ell_i(y_i; \tilde{\mu}_i, \kappa) - \ell_i(y_i; \hat{\mu}_i, \kappa))^{1/2}$$

and $\ell_i(y_i; \cdot)$ is the contribution of y_i for the total log-likelihood, $\tilde{\mu}_i$ is the maximum likelihood estimate of μ_i based only on y_i , $\hat{\mu}_i$ is the maximum likelihood estimate of μ_i based on the whole sample and the sign of d_i is the same as that of $y_i - \hat{\mu}_i$. In the case of the von Mises distribution, $VM(\mu_i, \kappa)$, the deviance residual is defined by $d_i = d(y_i; \hat{\mu}_i, \kappa) = \pm\sqrt{\kappa}(1 - \cos(y_i - \hat{\mu}_i))^{1/2}$. By using relations between the trigonometric functions we can rewrite d_i in a more convenient form,

$$d(y_i; \hat{\mu}_i, \kappa) = \pm\sqrt{\kappa} \cos \frac{1}{2}(y_i - \hat{\mu}_i), \quad (5.12)$$

that is useful in the development of the approximations.

5.4 Applications

In this section we first consider simulated examples to illustrate the proposed von Mises regression model and highlight the limitations of the model. Application to data collected in elephant tracking study is then presented.

5.4.1 An illustration using simulated data

To investigate the behaviour of the parameters of the fitted circular linear model, turn angles of animal movement are simulated from a Von Mises distribution $vm(\mu_i, \kappa)$ with $\mu_i = \mu + \arctan(x_i\beta)$ and x values from a uniform distribution on $(0,50)$. Such model checking is required since wrong choice of initial parameter values in model specification can lead to non-identifiability and non-convergence of the model [48, 61]. Figure 5.1 shows the scatter plots of the raw data (first row), the densities (second row) and the corresponding log-likelihood as a function of the estimated β coefficients where $\hat{\beta}$ coefficients are estimated using the IRLS algorithm (third row). The scatter plots in Figure 5.1 displays the possible patterns in the plot of (θ, x) and the additional points $(\theta + \pi, x)$ in the cartesian coordinates. For the two strongly non-uniform cases, we see the effect of rolling past zero point as the mode determines a well defined point cloud. The density plots in second row of Figure 5.1 shows that these are not strongly modal forms despite the patterns in the first row. The loglikelihood plot in the third row of Figure 5.1 demonstrates the difficulty in naively applying a mode finding algorithm, and this is why we recommend always using multiple starting points to find the global maxima (dashed line).

The first column of Figure (5.1) shows that if the true β is close to or near zero, the log-likelihood function not only has peaks near $\hat{\beta} = 0$, but also asymptotes out to $\pm\infty$ as β gets big in absolute value. In such a case, the only practical estimate of the mode is the peak near zero, which is

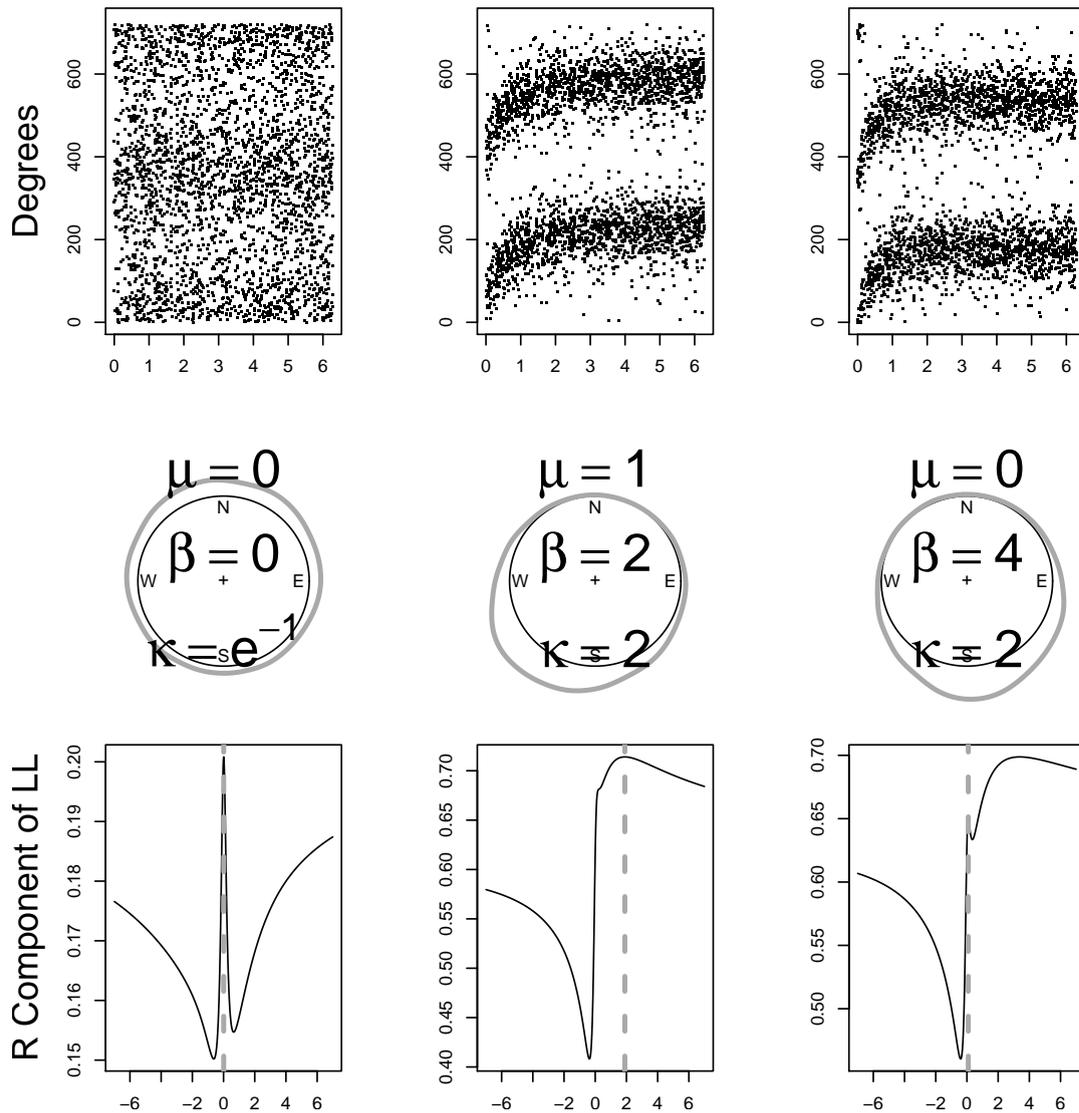


FIGURE 5.1: Simulation summaries for the von Mises distribution

the solution produced by the probability density function of a von Mises distribution. Despite this seemingly arbitrary choice, the parameter $\hat{\beta}(\text{mod}2\pi)$ is fully identified in the mathematical sense, as opposed to the ecological sense. This difference in identification definitions is a direct result of wrapping around the circle.

If the true value of $|\beta| > 0$, as illustrated in the second and third column of figure (5.1), the mean resultant length (R) of the log-likelihood function is not globally concave and local maximum exist quite close to the global maximum. This makes the maximization of the likelihood function difficult. The inspection of R as a function of β usually produces good starting values for the maximization of the likelihood. If computationally feasible, a grid of starting values can be used in subsequent runs of the IRLS algorithm. However, due to lack of global log-concavity, a proper exploration of the likelihood plot is always highly advisable when using the maximum likelihood approach in the von Mises regression model [42, 46].

5.4.2 Elephant movement data

In this section, we apply the discussed regression model to a real elephant movement data of orientation with covariates. We analyse the data that depicts the reorientation of the Elephants collected using GPS telemetry radio tracking in Kruger National Park (KNP). The data set consists of 63,265 observations recorded at an interval of 30 minutes from April 17 2006 to May 2009. However, due to computational challenges we only use 4,000 observations in our analysis. The turning angle (θ) is defined as the change in the direction of movement made by an individual from one location to another. The turning angle is a right-hand turn that ranged from $-\pi$ to π . To illustrate the utility of the reviewed methods, we analysed movement data from foraging African elephant in KNP. Spatial locations of the movement path of an individual foraging herd of female African elephant (*Loxodonta*) were recorded using GPS radio-telemetry device during 2006-2009. The GPS locations provided data every 30 minutes during an entire day with an accuracy of the locations within 50 meters. This information was sent via cellular phone (GSM) network to a website from where the information was downloaded.

GPS tags provided by African wildlife tracking (<http://www.awt.co.za>) transmitted location data through the GSM (cellular phone) network to a website, from where these records were received. Collars bearing GPS tags were placed on female elephants representing the movements of the breeding herds with which they were associated. Animal capture was undertaken using chemical immobilisers by South African National parks staff, following standard ethical procedures. An individual elephant representing a heard of eight elephants in the south-western (Pretorious Kop) region of Kruger National Park received collars in may 2006. If the animal was outside of the cell phone reception, the location data were stored and then downloaded once the animal came into a reception area [11, 131]. Movement turn angles were calculated from GPS locations recorded every 30 minutes apart using the methods given by [124].

In the South Africa, elephant live in protected habitats fenced by the government. An objective of the radio-telemetry study was to assess the effects of landscape features on the movement of the elephant. Location data and other associated information were collected by radio-tracking the elephant [11]. Elephant herd locations were obtained by tagging a small transmitter on the elephant's body and then locating the transmitter by a receiver attached to a directional antenna. Spatial coordinates for the elephant location were obtained using global positioning system(GPS) receivers and differential correction techniques were applied to improve their accuracy. Data for relocation intervals during which the animal did not move were excluded from this analysis [131]. The set of closed line segments, where the from-vertex of the first segment was identical to the to-vertex of the last segment, form a polygon representing the habitat patch. The resulting ArcView shapefile was exported as a tab delimited text file for input to R (R Development Core Team 2008) environment in which we perform the remaining analysis. A R function *clm*

developed by extending the source code of the circular R package was used to generate the results [2, 75]. Due to computational challenges and the desire to obtain biologically meaningful results from the data, we used only the first 4000 observations in our analysis.

5.5 Results

In Figure 5.4, we show a wind rose diagram representing the elephant movement turn angles data. Unlike linear data, failure to account for the cyclic nature of the circular data in graphics deceives viewers because it appears there are explicit endpoints. For example, consider the animal movement data. The turn angles derived from the GPS tracking data recorded half hourly certainly affect how animals move. Figure 5.2 is a histogram of turn angles are unimodal and bell-shaped distributed. Conversely, Figure 5.3 is a linear histogram of the elephant turn angles, it accurately depicts the orientation pattern of elephant movement during foraging or moving to a target.

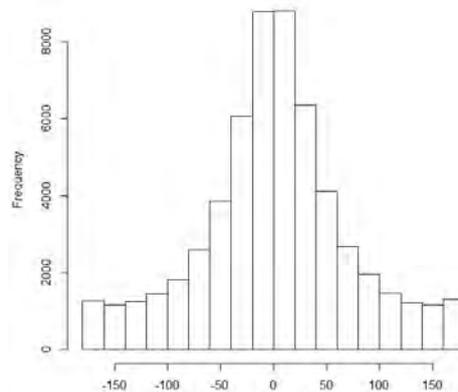


FIGURE 5.2: Histogram of elephant turn angle data.

We model the relationship between the turn angles and the step lengths. To make the IRLS algorithm converge to the global maxima, we carefully select the starting values based on visual inspection of the likelihood plot in Figure 5.5 of the simulated data. Table 5.1 gives the output from the circular regression model for these data, which is fit using our version of the [42]

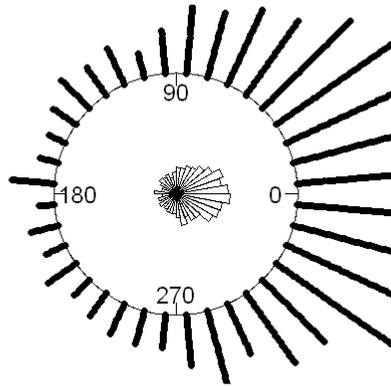


FIGURE 5.3: Rose diagram of elephant turn angle data.

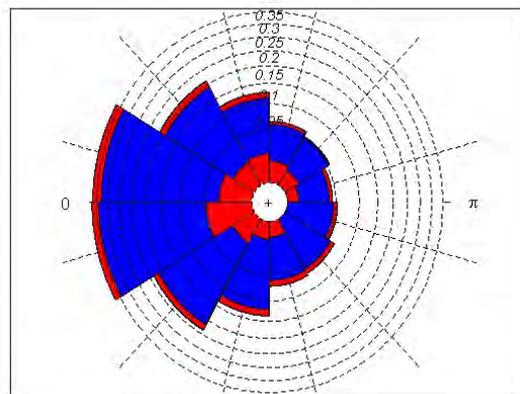


FIGURE 5.4: Displays of wind rose diagram of elephant turn angle data.

algorithm described in section 5.3.1. To ensure that the IRLS algorithm converged to the global maxima, the algorithm in section 3 was run from several different starting values since the likelihood function is not guaranteed to be unimodal, making the model somewhat fragile as seen in Figure 5.5. The fitted model provides a substantial improvement over the null (μ only) model since the difference in deviance is (3669.977) is far into the tail of the chi-squared distribution with 15 degrees of freedom. Applying the iterative process 5.10 to solve 5.9, we found the

TABLE 5.1: ML Estimates of directionality of turn angles of elephant movement

Parameters	coef	std.err	$ z $	p value
Mean direction (μ)	2.182	0.3978	5.4839	0.0000
β	0.0101	0.0068	1.4835	0.01379
κ	0.1126	0.0448	2.5109	0.012042

maximum likelihood estimates $\mu_0 = 2.182(se(\mu_0) = 0.3978)$, The step length parameter $\hat{\beta} =$

0.0101(0.00068) is significantly positive which implies that the length of a step an animal makes increases the magnitude of the turn angle. The concentration parameter of the elephant turn angles is statistically significant $\hat{\kappa} = 0.1126(0.0448)$ which implies that the turn angles are not uniformly distributed. Table 5.2 presents the parameter estimates of the circular regression of

TABLE 5.2: ML Estimates of landscape features on the mean direction of turn angles on Elephant movement($n = 4000$)

Parameters	coefficients	Std.err	$z - value$	p-value
μ	-2.6640	0.3010	8.8494	0.0000
Step Length	0.0101	0.0043	2.3456	0.0190
Distance to roads	0.0012	0.0005	2.5869	0.0097
Distance to water point	-0.0004	0.0003	1.4263	0.1538
Distance to river	-0.0004	0.0002	1.9097	0.0562
Temperature	0.0700	0.0827	0.8462	0.3974
Rainfall	-0.0008	0.0046	0.1774	0.8592
κ	0.0744	0.3010	0.2470	0.8049

the animal movement and so as expected there is a significant and positive relationship between the turning behaviour of elephants and the landscape features. The coefficient of step length is positively and significantly associated with the turning angle of the animal ($\beta = 0.0101, p < 0.05$). This implies that the more elephants move, the more they turn clockwise within a given interval of time in search of food resources [27]. The distance to the river is negatively and significantly associated with the turn angle the animal makes ($\beta = -0.0012, p < 0.05$). This finding supports the earlier argument by Duffy et al. [27] who noted that elephants forage in habitats closer to the river than habitats far away from the river. The distance to the water points is negatively associated with the turn angle but not significant. The distance to the roads was positively and significantly related with the elephant turning pattern ($\beta = 0.0012, p < 0.05$). This implies that elephants adjust their turning angles when avoid approaching the roads, or crossing the roads, thereby avoiding risks involved in moving through the human-dominated landscape elements. The distance to the water points and distance to the road were also negatively associated with the turn angles ($\beta = -0.004, p < 0.05$). This result indicates that elephants reduce their turn angles as they approach the water points and move in a more directed manner. The mean annual rainfall has negative effects on elephant movement orientation ($\beta = -0.0008, p > 0.05$) which indicates that elephants turned less as rainfall increased. The temperature is positively associated with turn angle but not significant ($\beta = 0.070, p > 0.05$). The result indicates that the elephant did not alter the turn angle as the temperature increased. Figure 5.6 presents von Mises Q-Q plot of the residuals. While some deviations from the theoretical distribution become apparent, there is little evidence of a serious departure from a von Mises model. Figure 5.5 displays a scatter plot of the fitted versus the residual of the von Mises regression model. The plot indicates that the fitted model provides a good fit to the data of elephant movement. The

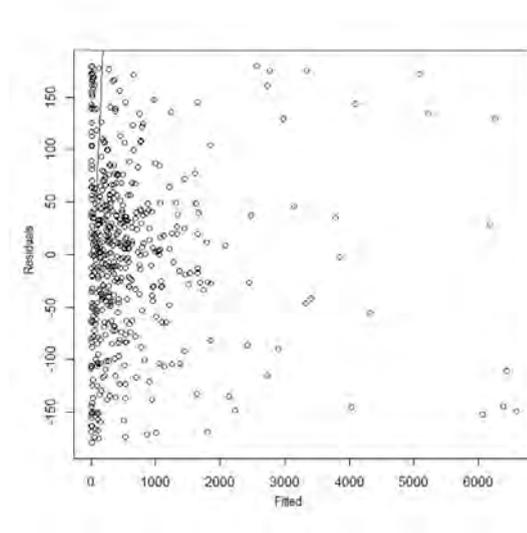


FIGURE 5.5: Diagnostic plot of the fitted von Mises regression model of elephant movement.

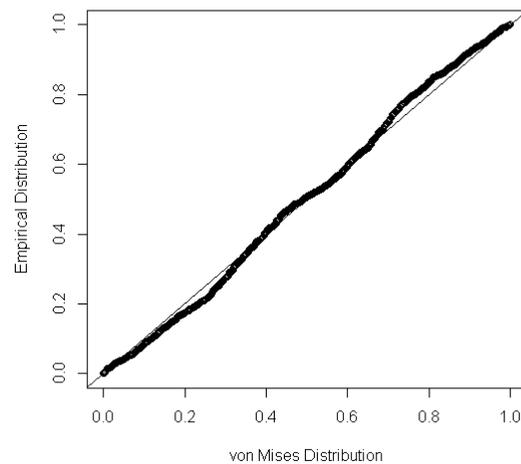


FIGURE 5.6: Probability plot of elephant movement turn angle model.

red line indicates that most of the data values are concentrated around 0° which is consistent the the hypothesis that animals have target oriented.

5.6 Discussion

Our interest in this chapter was to apply the circular linear regression methodology to animal movement data. In particular, we show that the models presented here can be an alternative to modelling animal movement data with circular responses and linear or categorical covariate.

Despite the substantial literature involving information from several species, there is no clear application of circular statistical methods in ecology and especially on animal movement GPS tracking data.

The application of circular statistical methods to optimal foraging theory across in situ systems has been hindered by the inability to account for all critical variables. The regression models provide a link between the circular metrics (turn angles) and the environmental drivers of animal movement. This allows broader inspection of the influence of ecological fluctuations on movement behaviour and the behavioural mechanisms adopted by species to cope with ecological constraints they face. An important characteristic of animal movement pattern is the clustering of the visited locations which can be modelled through the concentration parameter of the von Mises model [42, 124, 125]. Further, complex models that capture the effects of landscape covariates on the concentration parameter and the mean orientation are possible. However, a major challenge in implementing such a model occasioned by lack statistical software and computational capacity. In particular, the turning angle models enabled us to evaluate the effects of internal states of the animal on movement orientation such as moving to water and foraging separately. The findings of this study support the descriptive analysis of [27] who argued that animals turned less further away from large permanent water bodies and more from semi-permanent water bodies like seasonal rivers. Elephants used a direct movement strategy when needing to get a destination more quickly (e.g. toward water or mates), rather than significantly increasing their speed. Further, it supports the hypothesis that the elephant optimizes energy efficiency while still varying their foraging approach and search intensity. Movement ecology will advance in parallel to developments in circular statistics, and the development of circular statistic will be promoted by the practical demands made from movement ecology.

Although estimating the parameters of the von Mises distribution was straightforward, in that estimates of μ and κ were always obtained, the angular regression models did not always tend to a sensible solution immediately. Thus, careful note of the contour of the likelihood surfaces is required. However, these problems will in part be data-specific particularly (we postulate) if the value of β is small. More recent methodological work on circular regression, and which also refers to some early animal movement examples concerned with the study of landscape features on animal movement orientation [124].

CHAPTER 6

Approaches for testing the circular uniformity hypothesis in movement of large mega-herbivores in Kruger National park South Africa

Abstract

Circular statistics is an area not used very much by ecologists to describe animal movement patterns. Nevertheless, the connection between the evaluation of temporal recurring events and the analysis of directional data have converged in several papers, and show circular statistics to be an outstanding tool for understanding animal movement better. The aim of this chapter is to evaluate the applications of circular statistical tests to check uniformity hypothesis in animal movement and its potential interpretation within the general framework of movement ecology. Four methods of circular statistics: Rayleigh's, Watson's, Rao's spacing and Kuiper's test based on the mean resultant length are applied to examine the uniformity hypothesis of GPS derived telemetry data of elephant movement collected from Kruger National Park(KNP) South Africa. Overall, circular statistical uniformity tests methods represent a useful tool for evaluation of directionality elephant movement with applications including (i) assessment of animal foraging strategies; (ii) determination of orientation in response to landscape features and (iii)determination of the relative strengths of landscape features present bin a complex environment.

Keywords: Circular statistics, animal movement, turn angles, uniformity hypothesis.

6.1 Introduction

The elaboration of appropriate conservation management and protection of endangered species of vegetation cover should be based on accurate interpretation of data and knowledge of animal movement on its habitat. One metric in movement ecology critical to this understanding is animal turn angles. It is urged in the literature that animals turn more during the wet season and less during the dry season. This variation is attributed to heterogeneous food resources and landscape features [27]. The movement of animals in protected areas (PA) and distribution of artificial water points has great impact on the vegetation and ecological dynamics of the ecosystem in general.

Circular statistical test of uniformity provides an opportunity to ecologists for understanding the turning patterns of the animals. In most studies, unimodal orientation may be the expected outcome. In several situations where for instant species or population preferences have been studied or when compass cues were set in experimental conflict, bimodal or multi-modal orientation may be expected [42, 124]. Simulation studies based on random walk and Lévy flight theory of animal movement assume that turn angles are uniformly distributed [132]. As in the case linear statistics, the main objective here is to draw objective, reliable and biologically meaningful statistical inference about the population parameters on the basis of samples. Observations are either geometric or temporal in nature, where time related distributions can be fitted into a circular or spherical pattern [7]. Parametric and non-parametric statistical methods can be used to test hypotheses concerning angular data [7].

These methods are based on the assumption that observed angles are independent, a condition that may not be satisfied when multiple angles are recorded from a single individual. Among the distributions used to describe data on a circle are the circular uniform and Von Mises distributions. Goodness-of-fit tests exist for these distributions. Four widely known methods for carrying out such a task are the subject of the current paper. Biological applications related to the technique described in section 6.2.1 and 6.2.4 illustrate the earlier predominant role of one particular life science in predicting the potential utility of circular statistics in other domains of scientific and technical endeavor.

In recent years new data sources and GIS tools have been increasingly used in ecological studies. A peculiar characteristic of these data sources is that, often, only information about the locations of the animal trajectory/path have been recorded over some sampling resolution. Methodologies targeted especially for these data need to be developed. Motivation of this work arises from the a study about the movement patterns of Elephant in Kruger National park, South Africa and the apparent lack of tangible awareness of circular statistics in the movement ecology literature.

This study was performed in order to support decisions for the management, in particular the conservation strategies for Elephants.

A limitation of the most common approaches is that they treat turn angles qualitatively, that is, converting angular data as either north, south, east or west which may lead to loss of valuable information [23]. The specific aim here is to demonstrate the potential usefulness of circular statistics to animal movement analysis, in deciding whether GPS derived telemetry data justify the inference of uniformity in animal movement. Following a short presentation of theoretical aspects, numerical examples illustrate calculations required for drawing proper conclusions about the animal movement patterns.

6.2 Theory of hypothesis testing and notation

Circular hypothesis testing for uniformity is a valuable tool in movement ecology [42]. Statistical test should depend on previous assumptions supported by the descriptive or exploratory analysis [43]. In applied research, critical decisions based on data depends on objective and reliable assumptions. Due to technological advances and huge data collected on animal movement, testing such assumptions on animal movement data requires the knowledge of hypothesis testing.

A common question in circular statistics is whether a sample of data is uniformly distributed around the circle or has a common mean direction [10]. A multiple of test statistics have been designed for testing this hypothesis. These methods includes :(i) Rao spacing test; (ii) Kuiper's test, (iii) Rayleigh test and (iv) Watsons test. These four tests can be used to assess the evidence for a uniform, unimodality and the goodness-of-fit for the von mises distributions respectively [78]. Previous investigations shows that the Raos spacing method is more susceptible to rejecting H_0 than the kuipers test and the Rayleigh test in the face of a small data sample, unless data distribution is appreciably uniform at least in some of its sub-domains [43, 109]. This implies that the Rao's spacing test carries a similar Type 1 error compared to other test; however, we point out that from the nature of the data, rejection of H_0 cannot be absolutely certain [78].

In all the testing methods, the null hypothesis H_0 states that the population samples are uniformly distributed around the circle and the alternative hypothesis H_A the population samples do not show a uniform (or random) circular distribution. Each method rejects H_0 if its test statistic exceeds a critical value depending usually on sample size n , and level of significance α [59]. In the theory of statistics, $\alpha = 0.05$ regarded as significant and $\alpha = 0.01$ is highly significant, meaning that a Type 1 error, is made by rejecting H_0 is 5% or 1%, respectively. In the theory of statistics, $\alpha = 0.05$ is regarded as significant and $\alpha = 0.01$ is regarded as highly significant, meaning that a Type 1 error, is made by rejecting H_0 is 5% level of significance,

respectively. The current approach is more flexible by allowing the test statistic to determine the rejection of H_0 on the basis of the p-value. According to Hogg and Craig [54], the p-value is the magnitude of the error committed in rejecting H_0 in face of the computed test statistic.

Rejecting the null hypothesis thus implies that deviations from uniformity are too large to assign them to chance factors, hence they are of deterministic origin [100]. Since rejection of a null hypothesis is statistically stronger than its opposite, Rao's method is more inviting than any other test when at least medium-size deviations from non-uniformity are expected from prior inspection descriptive statistics or exploratory analysis [78]. Pewsey et al. [100] emphasize the importance of testing uniformity hypothesis in circular. We note that if the data fits neither a von Mises distribution or a uniform distribution and contains a single mode, then this data is said to follow a unimodal distribution [10, 43, 100]. In this case, although it is not possible to identify the actual distribution, the presence of a single mode not only indicates a preferred orientation in the sample, but also enables the use of non-parametric methods to estimate a mean direction with a confidence interval [43, 78, 100]. We then tested the uniformity hypothesis following the procedure outlined in figure 6.1 . Figure 6.1 represents the sequence of hypothesis tests

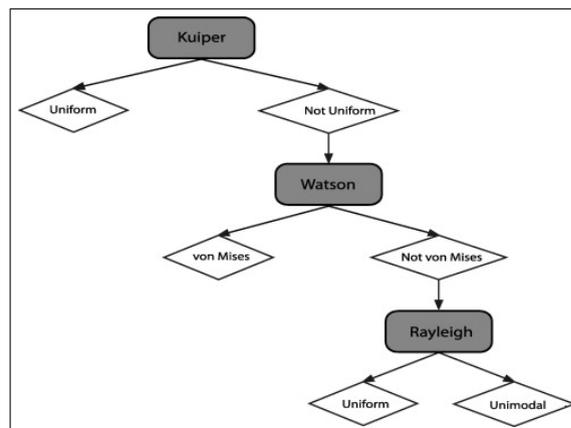


FIGURE 6.1: Flowchart representing the sequence of hypothesis tests based on circular distributions.

performed in this analysis. The Kuiper test is an omnibus test, meaning it tests the hypothesis of the sample following a uniform distribution, against any alternative distribution. If the null hypothesis is rejected, then there is evidence against uniformity and the possibility that the data fit the von Mises distribution is tested. The goodness of fit of the von Mises model can be formally assessed using Watson's test [78]. The null hypothesis in this test is that the data are drawn from a von Mises distribution, against the alternative that they are not drawn from a von Mises distribution. Rejection of the null hypothesis in the Watson's test, leads to rejecting the hypothesis that the data fits a von Mises model. The next step is to determine whether the data presents a single modal direction, using Rayleigh's test, where the alternative hypothesis is unimodality [78].

6.2.1 Rayleigh's test

The Rayleigh's test is based on the intuitive idea of rejecting uniformity when the vector sample mean $\bar{\theta}$, is far from 0, when \bar{R} is large [43, 78, 100]. The Rayleigh's test is the score test of uniformity within the von Mises model [78]. Put $w = (\kappa \cos(\mu), \kappa \sin(\mu))'$, the log likelihood Von Mises based on circular observations $\theta_1, \dots, \theta_n$ is

$$l(\kappa : \theta_1, \dots, \theta_n) = n\bar{\kappa}\bar{\theta} - n \log I_0(\kappa),$$

where $I_o(\kappa)$ is the modified Bessel function and $\bar{\theta} = \frac{1}{n} \sum_{i=1}^n (\cos \theta_i, \sin \theta_i)^T$ is the sample mean vector. The score is:

$$U = \frac{\partial l}{\partial \kappa^T} = n\bar{\theta} - nA(\kappa)(\cos \mu, \sin \mu)^T \quad (6.1)$$

From the moments properties of \bar{R} , [78] it is possible to note that the score statistic is

$$U' \mathbf{var}(U)^{-1} U = 2n\bar{R}^2. \quad (6.2)$$

From the general theory of score test, [78] the large sample asymptotic distribution of $2n\bar{R}^2$ under uniformity is a χ^2 with two degrees of freedom:

$$2n\bar{R}^2 \sim \chi^2 \quad (6.3)$$

where n is the sample size. It has been demonstrated, also that the Rayleigh's test coincides with the likelihood test of uniformity within the von Mises family.

6.2.2 Kuiper's test statistic

Kuiper's test is used to determine if a given set of data can be a sample from a specific distribution. It is similar to the Kolmogorov-Smirnov (KS) test, as both compare cumulative distributions [10]. For the one-sample test, the empirical cumulative distribution is compared to a theoretical cumulative distribution. As for circular case, this test measures the deviation between empirical distribution, $S_n(x)$, and the Uniform cumulative distribution functions (cdf), $F(x) = \frac{\theta}{2\pi}$. In the case of circular data, the definition of cumulative distribution is not obvious and is quite different from the in line cdf [100]. In the circular data case, in fact, we first have to choose the circle zero point and orientation, then we need to augment the ordered observations, $\theta_1, \dots, \theta_n$ of $x_0 = 0$ and $\theta_{n+1} = 2\pi$. The S_n is then defined by:

$$S_n(\theta) = \frac{i}{n} \quad \text{if } \theta(i) \leq \theta \leq \theta(i+1) \quad i = 0, 1, \dots, n. \quad (6.4)$$

Just as in Kolmogorov-Smirnov's test for in line distribution [33], the following quantities are defined:

$$D_n^+ = \sup_{\theta} \{S_n(\theta) - F(\theta)\}, \quad D_n^- = \sup_{\theta} \{F(\theta) - S_n(\theta)\}.$$

To overcome the dependence of D_n^+ and D_n^- on the choice of the initial direction, [70] defined

$$V_n = D_n^+ + D_n^- \quad (6.5)$$

The statistic (6.5) has been demonstrated [59, 78] to be invariant under the change of initial direction. The null hypothesis of uniformity is rejected for large values of V_n . Moreover, the Kuiper's test is consistent against all alternative to uniformity [100]. For practical purposes, the following modification of V_n is used:

$$V_n^* = n^{1/2} V_n \left(1 + \frac{0.155}{\sqrt{n}} + \frac{0.24}{n} \right). \quad (6.6)$$

6.2.3 Watson's test

Another common test of uniformity in circular statistics is the Watson U^2 statistic Watson [137] which is a modification of the Cramér-von Mises test [33]. This test is used as a goodness-of-fit statistics for the von Mises distribution [77]. As a test of goodness of fit for circular data, it is invariant to the choice of the origin. The watson test statistic is defined as

$$U_n^2 = \int_0^{2\pi} \left[(S_n(\theta) - F(\theta)) - \int_0^{2\pi} (S_n(\theta) - F(\theta)) dF \right] dF \quad (6.7)$$

It follows from this definition that the Watson's statistic is invariant under rotation and reflections. As for the Kuiper's test, it is useful to consider the following modified statistic:

$$U_n^{*2} = \left(U_n^2 - \frac{0.1}{n} + \frac{0.1}{n^2} \right) \left(1 + \frac{0.8}{n} \right). \quad (6.8)$$

Stephens [121] provides the in a tabular form the quantiles of the Watson test statistic.

6.2.4 Rao's spacing test

Rao's Spacing test is a useful and powerful statistic for testing uniformity of circular data. As with other circular statistics, Rao's Spacing test is applicable for analysis of angular data, in studies of movement and spatial trends in geographical research [100]. In many cases, particularly

with an underlying distribution that is multimodal, it is more powerful than the popular Kuiper's Test and Rayleigh Test.

Rao's Spacing Test is based on the idea that if the underlying distribution is uniform, successive observations should be approximately evenly spaced, about $360/N$ apart. Large deviations from this distribution, resulting from unusually large or unusually short spaces between observations, are evidence for directionality. It is related to the general class of linear statistical tests that are based on successive order statistics and spacing. The spacing tests sample arc lengths T_1, \dots, T_n defined as:

$$T_i = \theta(i) - \theta(i-1), \quad i = 1, \dots, n-1, \quad T_n = 2\pi - (\theta(n) - \theta(1)). \quad (6.9)$$

Under uniformity $E[T_i] = \frac{2\pi}{n}$. Hence, it is reasonable to reject uniformity for large values of

$$L = \frac{1}{2} \sum_{i=1}^n \left| T_i - \frac{2\pi}{n} \right| \quad (6.10)$$

Large values of L indicate clustering of observations [10, 109]. An extensive table of quantiles of L is given [109], while Sherman [115] shows that a suitable transformation of L is asymptotically standard Normally distributed.

6.3 Application to elephant movement data

6.3.1 Ethics statement

Elephant capture and handling was conducted in strict accordance with ethical standards. Specific approval for this particular research project was obtained through the University of KwaZulu-Natal Animal Ethics sub-committee (*Ref. 009/10/Animal*). This research also forms part of a registered and approved SANParks project, in association with Kruger National Park and Scientific Services (*Ref: BIRPJ743*) [11, 131].

6.3.2 Study area and GPS Data on a elephants

The methods described were applied to GPS location data for three female elephant herds in Kruger national park of South Africa. Elephant movement have been previously studied by [11, 131]. The elephant population in KNP was estimated to be 14,000 individuals during 2010 (SANParks, unpublished data). From 2006 to 2010, we collected geographical location data,

downloaded from GPS/GSM Collars (Africa Wildlife Tracking cc., South Africa), fitted to three elephant cows from different herds. To ensure the independence of sampling, a single female in each herd was selected and collared. Because GPS coordinates were measured continuously at frequent time intervals, the trajectory we obtained was almost smoothly connected [131]. The GPS data are freely available upon request.

The turning patterns of elephants were monitored by computing the angle θ between consecutive relocations over 30 minutes interval of time between May, 2006 and June 2009. The data set consists of 36395, 29221 and 29908 observations for herd AM108, AM307 and AM308 respectively recorded at an interval of 30 minutes. The turning angle (θ) were computed for the three herds as the change in the direction of movement made by an each individual elephant tagged from one location to the next. The turning angle is a right-hand turn that ranged from $-\pi$ to π . The GPS locations provided data every 30 minutes during an entire day with an accuracy of the locations within 50 meters. This information was sent via cellular phone(GSM) network to a website from where the information was downloaded. The descriptive analysis of elephant turn angles shown in Table 6.1 the elephant turn angles are oriented towards a preferred direction as evidenced by the mean orientation 359 and 1 degrees. The circular variances of the three herds ranged between 0.518 and 0.566 for the three year period. However the circular variance is larger in the wet season than in the dry season. Angular data of three elephant herds was used to construct circular histograms that depict the mean and frequency of movement orientations of animals Figure 6.2 to 6.4. The rose diagram in Figure 6.5 indicates that elephants move in a non-random manner during the dry and the wet seasons. This variation is attributed to uneven distribution of resources during dry season than in the wet season. Similarly, the linear histograms in Figure 6.4 shows that elephants turn angles are concentrated around zero degrees which indicate that the movement is target oriented. The tendency is known as the Unimodal movement pattern better described by Unimodal distributions. However, the histograms in Figures 6.2 and 6.3. indicate there is no season deference in the distribution of elephant turn angles. All the three herds display an oriented movement across the wet and dry seasons. The histograms further show that the turn angles follow a Unimodal pattern.

Based on the results in Table 6.2, the circular uniformity hypothesis is rejected in all the three elephant herds. The Kuiper's test result indicates that the elephant movement turn angles are not uniform (p-value; 0:05). This implies that the turn angles are non-randomly distributed. Kuiper's and Rayleigh's test exhibit small p-values although the former is less realistic with respect to the powerful rejection of uniformity hypothesis by Rao's test statistic. The result of Table 6.3 indicates that the turn angles of the three elephants follow a von Mises distribution across the wet and dry seasons. This finding implies that it's valid assume that elephant turn angles data follows a von Mises distribution. Similar results are obtained for the three year data. Based on the Rayleigh's test statistic in Table 6.2, the claim that the data is unimodal

for the three elephant herds is not rejected. A p-value is not a sufficient justification for the rejection of the unimodality distribution as evidenced by the histograms in Figure 6.5, 6.2 and 6.3 indicating a tendency of elephants to move in one direction while approaching a target [27]. There distribution was not different between the wet and dry season 6.3.

6.4 Discussion

This chapter focuses on the test of uniformity hypothesis for the circular data of animal movement. The primary objective was to determine whether elephant orientation patterns were uniform. We also investigated the hypothesis that elephant orientation patterns would vary between seasons. Testing the null hypothesis can be accomplished by comparing any test statistic for uniformity to a reference distribution obtained by simulation. This is appealing in that precise distributions consistent with turn angle data need not to be assumed for the procedure to have the proper type I error rate. These applications, in summary entail calculation of animal orientation variables mean turn angle, concentration parameter, distribution, and determination of modality and testing hypothesis about uniformity of animal turn angles. Our results suggest that elephant turn angles data is not uniformly distributed and showed no seasonal variation. The four statistical tests reject the uniformity hypothesis. We also found that the elephant turn angles follow a von Mises distribution. This means that making an assumption that animal turn angles are von Mises distributed is valid. It is important to note that the circular statistics applies to any level and scale of analysis, from individuals to several species of animals, and that the descriptive measures calculated are easily compared statistically by means of a variety of two or more test statistics.

TABLE 6.1: Descriptive analysis of the turn angle data of three elephant herds collected from Kruger National Park, South Africa.

Statistic	AM108			AM307			AM308		
	Angle	Seasons		angle	Seasons		angle	Seasons	
Variable	Angle	Wet	Dry	angle	Wet	Dry	angle	Wet	Dry
Number of Observations	36395	17736	18659	29221	12652	16569	29908	13550	16358
Mean Vector (μ)	359.425°	359.434°	359.418°	0.706°	0.84°	0.609°	359.957°	359.7°	0.16°
Length of Mean Vector (r)	0.481	0.464	0.498	0.434	0.418	0.446	0.482	0.471	0.492
Concentration	1.096	1.046	1.144	0.962	0.92	0.995	1.099	1.065	1.127
Circular Variance	0.519	0.536	0.502	0.566	0.582	0.554	0.518	0.529	0.508
Circular Standard Deviation	69.292°	71°	67.686°	74.044°	75.65°	72.833°	69.19°	70.336°	68.248°
Standard Error of Mean	0.414°	0.618°	0.556°	0.519°	0.821°	0.668°	0.455°	0.695°	0.602°

TABLE 6.2: Tests for uniformity for three elephant herds turn angles derived from GPS tracking data.

	Kuiper's		Rayleigh		Rao spacing		Watson's	
	statistic	p-value	statistic	p-value	statistic	p-value	statistic	p-value
AM108	61.6574	< 0.01	0.4813	0.000	166.2602	< 0.001	452.2261	< 0.01
AM307	50.7271	< 0.01	0.4339	0.000	162.4811	< 0.001	301.1729	< 0.01
AM308	56.3221	< 0.01	0.4823	0.000	166.2796	< 0.001	376.4143	< 0.01

TABLE 6.3: Testing if the elephant movement data is Von Mises distributed using Rao's spacing test.

	Wet Season		Dry season		Both seasons	
	statistic	p-value	statistic	p-value	statistic	p-value
AM108	0.0196	> 0.05	0.0177	> 0.05	0.0169	> 0.05
AM307	0.0302	> 0.05	0.0278	> 0.05	0.0292	> 0.05
AM308	0.0484	> 0.05	0.0882	> 0.05	0.0154	> 0.05

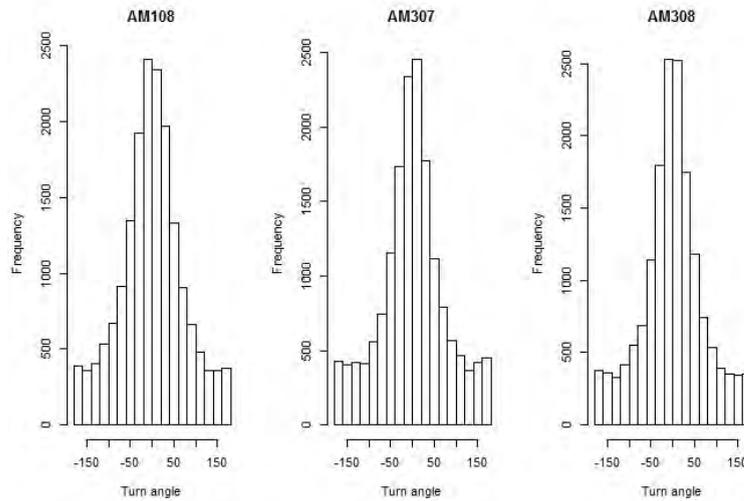


FIGURE 6.2: Histogram of herd AM108, AM307 and AM308 GPS-telemetry derived turn angles data collected from Kruger National Park South Africa (May 2006 -April 2009)

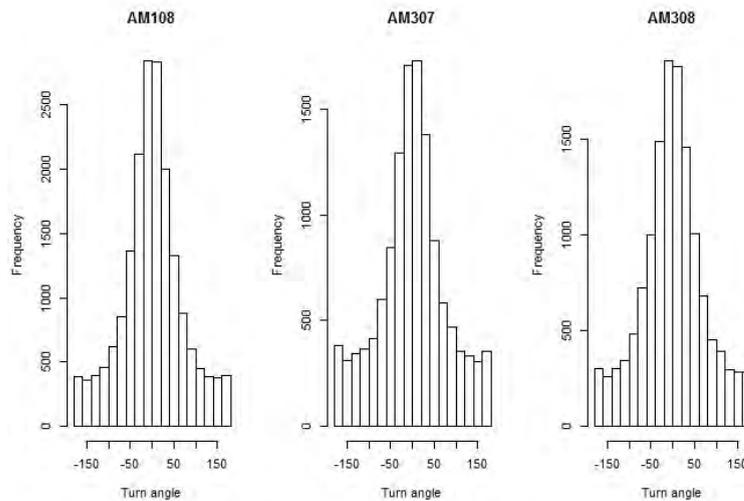


FIGURE 6.3: Histogram of herd AM108, AM307 and AM308 GPS-telemetry derived turn angles data collected from Kruger National Park South Africa (May 2006 -April 2009)

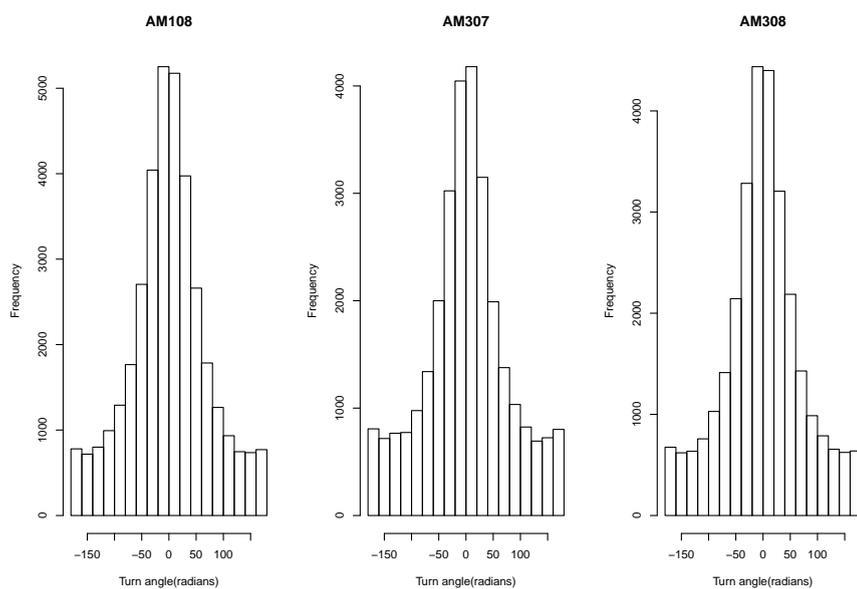


FIGURE 6.4: Histogram of herd AM108, AM307 and AM308 GPS-telemetry derived turn angles data collected from Kruger National Park South Africa (May 2006 -April 2009)

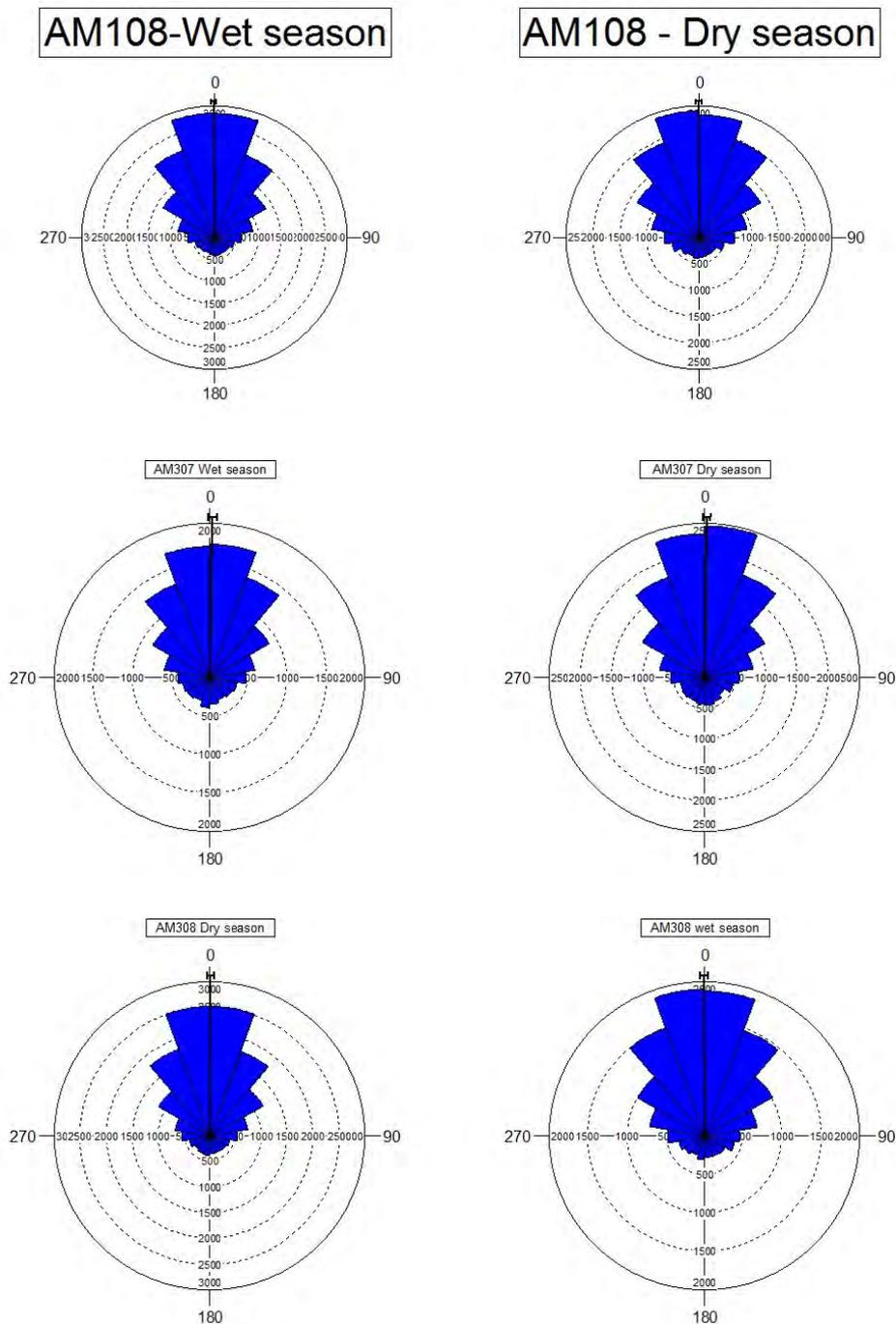


FIGURE 6.5: Circular rose diagrams of the elephant herd AM108, AM307 and AM308 GPS-telemetry derived turn angles data collected from Kruger National Park South Africa (May 2006 -April 2009). Circular bars indicate the number of observations within each class range.

CHAPTER 7

Conclusion

7.1 Summary conclusions

In this thesis, we have investigated how the movement of animals can be studied in terms of statistical analysis and the biological meaning drawn from it. Specifically, we examined the appropriateness of the various probability distributions for the non-linear variables in describing the elephant movement linear and circular metrics. In our work, we find that the stable distribution assumption is inadequate to capture the heavy tail and skewness properties of the step length data. We further demonstrate the flexibility of the stable distribution can be exploited in a linear regression models to investigate the effects of habitat type covariate on heavy tailed and skewed data. A model for describing the effect of the habitat type covariate on the tails is presented. However, one limitation of the stable distribution is that the confidence intervals estimated when the tail index and the skewness parameters are at the boundary are unreliable. Our models are a milestone to the modelling of animal movement. With the recent advances in technology and increasing ability to collect more data on animal movement, our findings provide alternative statistical tools for analysing tracking data and investigating the effects of environmental factors on movement metrics.

In chapter 2, I introduced and critiqued a range of analysis methods for analysing tracking data. In particular, I discuss methods for fitting stable distribution including the method of moments, the regression type characteristic methods and the maximum likelihood method. I also critique alternative to stable distribution and focus on the stable Paretian and the power law distribution only in this chapter. For most applications, power law distributions are the most rigorous analytical approaches but they require a high level of programming knowledge and

statistical expertise to implement effectively. We also compare the stable distribution against the power law across several herds. We find that the stable distribution fitted the data better than the power law and was robust across the six elephant herds. Our findings support an assertion by [101] who pointed out that wrong statistical assumption of the distribution generating the step lengths can lead to misidentification of the movement pattern and hence wrong conclusions which can impact management and conservation policies if the goal is to protect endangered species. Interestingly, we find herd to herd variation which indicates that future model developments should endeavor to explain herd effects through a random effect term in a regression model.

In chapter 3, we assumed the error terms of the linear regression model linking the animal movement step lengths to environmental variables to be stable distributed and fitted a stable Paretian regression model. The potential of using a simple heavy tailed model known as t regression is demonstrated and found to perform more better than stable law regression models. The analysis demonstrated a clear relationship between vegetation cover type and animal movement speed, with elephants moving more slowly and tending to forage in rich vegetation cover types, while tending to moving quickly through poor vegetation cover types. Our findings indicates that elephant movement is influenced by the vegetation cover types. Further analysis can be carried out with more covariates.

In chapter 4, we demonstrated the application of ARMA-GARCH modelling structure to ecological data of animal movement. GPS telemetry data of animal locations contains important biological signals which are appropriately captured by the tail index and skewness parameters of the stable distribution and the Skewed t distribution. However, the student's t distribution only captures the heavy tails. In this chapter, we found that the residuals of the stable Paretian ARMA-GARCH model fitted the elephant movement data better than the alternative normal, Student's t and skewed student t distributed GARCH models. However, we note that the diagnostic testing and model checking for the stable GARCH models remains to be developed for heavy tailed and skewed data sets. We note that future rime series models of animal movement should be based on robust choice of appropriate distributions assumption that can appropriately capture both the biological and the statistical properties of the data.

Chapter 5 and 6 focuses on statistical analysis of circular data from animal movement. Chapter 5 employed the circular statistical methodology and more on the biology of elephant movement. In that chapter, i used the results of the circular regression model to describe the relationship between animal turn angles and landscape features. The results indicates that there is a strong relationship between distance to water point, distance to the river amount of rainfall and temperature. In chapter 6, we test the circular uniformity hypothesis. A number of methods for testing the uniformity of turning angles are explored and its implications in ecology discussed.

We demonstrate that during model building, appropriate statistical assumptions are necessary to attain robust results.

7.2 Further Work

This study has used both sound statistical methodology and ecological theory to answer questions on the movement patterns and behaviours of elephant. Specifically a number of statistical methods have been explored and sound biological interpretation provided. The elephant data has been used as the main focus in this thesis though the methods discussed here can be applied to data from other animals. The study indicates further the need to apply statistical and ecological theory rather than just answering specific descriptive questions about individual species that are not in bigger ecological picture. For a huge data set which is highly skewed and heavy tailed, the confidence intervals seems to be unrealistic as the upper limit lies outside the confidence boundary. Such a problem warrants further investigations as the confidence intervals of the stable distribution parameter when α and β are in the boundary.

The results of descriptive analysis have also shown to understand the biological variation in the elephant population, future models need to capture the herd to herd variation. This can be captured through a linear mixed model framework with the herds as the random term. To assess the effects of covariates on each parameter of the stable distribution, a regression model in the framework of glm should be adopted as described by Lambert and Lindsey [71]. Computational softwares also need to be implemented to enable fitting such models.

The analysis of heavy tailed and skewed autocorrected time series data of elephant movement has revealed important biological signals which could be lost if the data through sub-sampling. Time series models incorporating environmental drivers remains to be tested in ecology. We further note to make more valid understanding of animal movement data, future studies should focus on developing diagnostic tools for model checking, selection based on the stable Paretian, student's t and skewed student's t distribution assumption in a ARMA-GARCH modelling framework. Higher order lags time series models can be of interest in future studies of animal movement. Second we mention that the time series models can be generalized to study biological variation between herds through the random coefficient autoregressive models. Asymmetric models such as skewed Student's t and skewed normal are the best next alternative to the stable Paretian ARMA-GARCH models. More research needs to be carried out to establish the worth of these models.

Circular linear regression models assuming a von Mises distribution of turn angles offers an alternative framework of evaluating the effects of environmental variables on animal movement.

Such models are computationally challenging as they are lacking in main stream statistical softwares. Though a few statistical packages have been implemented in R statistical software, tools for diagnostic checking, model selection and validation needs to be implemented. To necessitate further development of such models, ecological data of animal movement rich with covariates is needed. We leave this for future research. It would also be interesting to investigate the problem in a time series framework.

Further, the models highlights the biological importance of using appropriate statistical assumption that capture the properties of the data adequately. Previous studies of animal movement failed to account for both heavy tail and skewness in the data. For a huge data set which is highly skewed and heavy tailed, the confidence intervals seems to be unrealistic as the upper limit lies outside the confidence boundary. Such a problem warrants further investigations as the confidence intervals of the stable distribution parameter when α and β are in the boundary.

Bibliography

- [1] Agostinelli, C. and Lund, U. (2011a). R package 'circular': Circular statistics (version 0.4-3). URL <https://r-forge.r-project.org/projects/circular>.
- [2] Agostinelli, C. and Lund, U. (2011b). R package circular: Circular statistics (version 0.4-3). CA: Department of Environmental Sciences, Informatics and Statistics, CaFoscari University, Venice, Italy. UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA, page 1.
- [3] Akaike, H. (1978). On the likelihood of a time series model. *The Statistician*, pages 217–235.
- [4] Barragán, S., Fernández, M. A., Rueda, C., and Peddada, S. D. (2011). isocir: An r package for isotonic inference for circular data. an application in cell biology.
- [5] Bartumeus, F. (2007). Lévy processes in animal movement: an evolutionary hypothesis. *Fractals*, 15(02):151–162.
- [6] Bartumeus, F., Francesc, P., Salvador, P., Celia, M., and Jordi, C. (2003). Helical lévy walks: adjusting searching statistics to resource availability in microzooplankton. *Proceedings of the National Academy of Sciences*, 100(22):12771–12775.
- [7] Batschelet, E. (1981). *Circular Statistics in Biology*. Mathematics in Biology. Academic Press.
- [8] Benhamou, S. (2004). How to reliably estimate the tortuosity of an animal's path:: straightness, sinuosity, or fractal dimension? *Journal of Theoretical Biology*, 229(2):209 – 220.
- [9] Benhamou, S. (2006). Detecting an orientation component in animal paths when the preferred direction is individual-dependent. *Ecology*, 87(2):518–528.
- [10] Berens, P. (2009). Circstat: A matlab toolbox for circular statistics. *Journal of Statistical Software*, 31(10):1–21.

-
- [11] Birkett, P. J., Vanak, A. T., Muggeo, V. M. R., Ferreira, S. M., and Slotow, R. (2012). Animal perception of seasonal thresholds: Changes in elephant movement in relation to rainfall patterns. *PLoS ONE*, 7(6):e38363.
- [12] Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Econom.*, 31:307–327.
- [13] Borak, S., Härdle, W., and Weron, R. (2005). Stable distributions. In *Statistical tools for finance and insurance*, pages 21–44. Springer.
- [14] Bovet, P. and Benhamou, S. (1988). Spatial analysis of animals’ movements using a correlated random walk model. *Journal of Theoretical Biology*, 131(4):419–433.
- [15] Boyce, M. S., Pitt, J., Northrup, J. M., Morehouse, A. T., Knopff, K. H., Cristescu, B., and Stenhouse, G. B. (2010). Temporal autocorrelation functions for movement rates from global positioning system radiotelemetry data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1550):2213–2219.
- [16] Cain, M. L. (1985). Random search by herbivorous insects: A simulation model. *Ecology*, 66(3):pp. 876–888.
- [17] Cain, M. L. (1989). The analysis of angular data in ecological field studies. *Ecology*, 70(5):pp. 1540–1543.
- [18] Calenge, C., Dray, S., and Royer-Carenzi, M. (2009). The concept of animals’ trajectories from a data analysis perspective. *Ecological informatics*, 4(1):34–41.
- [19] Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *Society for Industrial and Applied Mathematics*, 51(4):661–703.
- [20] Codling, E. and Hill, N. (2005). Sampling rate effects on measurements of correlated and biased random walks. *Journal of Theoretical Biology*, 233(4):573 – 588.
- [21] Curto, J. D., Pinto, J. C., and Tavares, G. N. (2009). Modeling stock markets’ volatility using garch models with normal, studentst and stable paretian distributions. *Statistical Papers*, 50(2):311–321.
- [22] Cushman, S., M, C., and C, G. (2005). Elephants in space and time. *Oikos*, 109:331–341.
- [23] Dai, X., Shannon, G., Slotow, R., Page, B., and Duffy, K. J. (2007). Short-duration daytime movements of a cow herd of african elephants. *Journal of Mammalogy*, 88(1):151–157.
- [24] De Beer, Y. and Van Aarde, R. (2008). Do landscape heterogeneity and water distribution explain aspects of elephant home range in southern africa’s arid savannas? *Journal of Arid Environments*, 72(11):2017–2025.

- [25] Downs, T. D. and Mardia, K. V. (2002). Circular regression. *Biometrika*, 89(3):683–698.
- [26] Dray, S., Royer-Carenzi, M., and Calenge, C. (2010). The exploratory analysis of autocorrelation in animal-movement studies. *Ecological Research*, 25(3):673–681.
- [27] Duffy, K. J., Dai, X., Shannon, G., Slotow, R., and Page, B. (2011). Movement patterns of african elephants (*Loxodonta africana*) in different habitat types. *South African Journal of Wildlife Research*, 41(1):21–28.
- [28] Duffy, K. J., van Os, R., Vos, S., van Aarde, J., Elish, G., and Stretch, A. M. B. (2002). Estimating impact of reintroduced elephant on trees in a small reserve. *South African Journal of Wildlife Research*, 32(1):23–29.
- [29] DuMouchel, W. H. (1971). *Stable distributions in statistical inference*. University Microfilms.
- [30] DuMouchel, W. H. (1973). On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *The Annals of Statistics*, 1(5):948–957.
- [31] DuMouchel, W. H. (1975). Stable distributions in statistical inference: 2. information from stably distributed samples. *Journal of the American Statistical Association*, 70(350):386–393.
- [32] Dunn, J. E. and Gipson, P. S. (1977). Analysis of radio telemetry data in studies of home range. *Biometrics*, 33(1):pp. 85–101.
- [33] Durbin, J. (1973). *Distribution theory for tests based on sample distribution function*, volume 9. Society for Industrial Mathematics.
- [34] Edwards, A. M. (2008). Using likelihood to test for lévy flight search patterns and for general power-law distributions in nature. *Journal of Animal Ecology*, 77(6):1212–1222.
- [35] Edwards, A. M., Freeman, M. P., Breed, G. A., and Jonsen, I. D. (2012). Incorrect likelihood methods were used to infer scaling laws of marine predator search behaviour. *PloS one*, 7(10):e45174.
- [36] Edwards, A. M., Phillips, R. A., Watkins, N. W., Freeman, M. P., Murphy, E. J., Afanasyev, V., Buldyrev, S. V., da Luz, M. G. E., Raposo, E. P., Stanley, H. E., and Viswanathan, G. M. (2007). Revisiting lévy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature*, 449(7165):1044–1048.
- [37] Edwards, Andrew, M. (2011). Overturning conclusions of lévy flight movement patterns by fishing boats and foraging animals. *Ecological Society of America*, 92(6):1247–1257.

-
- [38] Fama, E. F. (1963). Mandelbrot and the stable paretian hypothesis. *Journal of Business*, pages 420–429.
- [39] Fama, E. F. and Roll, R. (1968). Some properties of symmetric stable distributions. *Journal of the American Statistical Association*, pages 817–836.
- [40] Fama, E. F. and Roll, R. (1971). Parameter estimates for symmetric stable distributions. *Journal of the American Statistical Association*, 66(334):331–338.
- [41] Feller, W. (2008). *An introduction to probability theory and its applications*, volume 2. John Wiley & Sons.
- [42] Fisher, N. and Lee, A. (1992). Regression models for an angular response. *Biometrics*, pages 665–677.
- [43] Fisher, N., Lewis, T., and Embleton, B. (1993). *Statistical analysis of spherical data*. Cambridge University Press.
- [44] Fonseca, T. C., Ferreira, M. A., and Migon, H. S. (2008). Objective bayesian analysis for the student-t regression model. *Biometrika*, 95(2):325–333.
- [45] Getz, W. M. and Saltz, D. (2008). A framework for generating and analyzing movement paths on ecological landscapes. *Proceedings of the National Academy of Sciences*, 105(49):19066–19071.
- [46] Gill, J. and Hangartner, D. (2010). Circular data in political science and how to handle it. *Political Analysis*, 18(3):316–336.
- [47] Goodwin, B. J. and Fahrig, L. (2002). Effect of landscape structure on the movement behaviour of a specialized goldenrod beetle, *trirhabda borealis*. *Canadian Journal of Zoology*, 80(1):24–35.
- [48] Gould, A. L. (1969). A regression technique for angular variates. *Biometrics*, pages 683–700.
- [49] Green, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 149–192.
- [50] Gumbel, E. (1954). Applications of the circular normal distribution. *Journal of the American Statistical Association*, 49(266):267–297.
- [51] Gumbel, E., Greenwood, J. A., and Durand, D. (1953). The circular normal distribution: Theory and tables. *Journal of the American Statistical Association*, 48(261):131–152.

- [52] Gurarie, E., Andrews, R. D., and Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters*, 12(5):395–408.
- [53] Hall, R. L. (1977). Amoeboid movement as a correlated walk. *Journal of Mathematical Biology*, 4:327–335. 10.1007/BF00275081.
- [54] Hogg, R. V. and Craig, A. T. (1978). *Introduction to mathematical statistics*. New York: Macmillan.
- [55] Hopcraft, J. G. C., Morales, J. M., Beyer, H. L., Haydon, D. T., Borner, M., Sinclair, A. R., and Olf, H. (2007). Serengeti wildebeest and zebra migrations are affected differently by food resources and predation risks. zebra.
- [56] Jachowski, D. S., Montgomery, R. A., Slotow, R., and Millspaugh, J. J. (2013a). Unravelling complex associations between physiological state and movement of african elephants. *Functional Ecology*, 27(5):1166–1175.
- [57] Jachowski, D. S., Slotow, R., and Millspaugh, J. J. (2013b). Corridor use and streaking behavior by african elephants in relation to physiological state. *Biological Conservation*, 167:276–282.
- [58] James, A., Plank, M. J., , and Edwards, A. M. (2011). Assessing lévy walks as models of animal foraging. *Journal of the Royal Society Interface*, 8:1233–1247.
- [59] Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*,. World Scientific, New York.
- [60] Janicki, A. and Weron, A. (1994). *Simulation and Chaotic Behavior of Stable Stochastic Processes*. Marcel Dekker, New York.
- [61] Johnson, R. A. and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, 73(363):pp. 602–606.
- [62] Kareiva, P. and Shigesada, N. (1983). Analyzing insect movement as a correlated random walk. *Oecologia*, 56(2/3):234–238.
- [63] Kawai, R. (2012a). Continuous-time modeling of random searches: statistical properties and inference. *Journal of Physics A: Mathematical and Theoretical*, 45(23):235004.
- [64] Kawai, R. (2012b). Sampling rate of spatial stochastic processes with independent components in modeling random search paths. *Physical Review E*, 85(2):021907.
- [65] Kawai, R. and Petrovskii, S. (2012). Multi-scale properties of random walk models of animal movement: lessons from statistical inference. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 468(2141):1428–1451.

-
- [66] Kharrat, T. and Boshnakov, G. N. (2014). *StableEstim: Estimate the 4 parameters of stable law using different methods*. R package version 2.0.
- [67] Kotz, S. and Nadarajah, S. (2004). *Multivariate t Distributions and their Applications*. Cambridge University Press, Cambridge.
- [68] Koutrouvelis, I. A. (1980). Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, 75(372):918–928.
- [69] Koutrouvelis, I. A. (1981). An iterative procedure for the estimation of the parameters of stable laws: An iterative procedure for the estimation. *Communications in Statistics-Simulation and Computation*, 10(1):17–28.
- [70] Kuiper, N. H. (1962). Tests concerning random points on a circle. *Nederl. Akad. Wetensch. Proc. Ser. A*, 63:38–47.
- [71] Lambert, P. and Lindsey, J. K. (1999). Analysing financial returns by using regression models based on non-symmetric stable distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48(3):409–424.
- [72] Lange, K. L., Little, R. J. A., and Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84(408):pp. 881–896.
- [73] Lévy, P. (1925). *Calcul des probabilités*, volume 9. Gauthier-Villars Paris.
- [74] Loarie, S. R., Aarde, R. J. V., and Pimm, S. L. (2009). Fences and artificial water affect african savannah elephant movement patterns. *Biological Conservation*, 142(12):3086 – 3098.
- [75] Lund, U. and Agostinelli, C. (2007). Circstats: Circular statistics, from topics in circular statistics(2001). *S-plus original by Lund, U. R port by Agostinelli, C. R package version 0.2-3*.
- [76] Mandelbrot, B. (1963). The stable paretian income distribution when the apparent exponent is near two. *International Economic Review*, 4(1):111–115.
- [77] Mardia, K. (1972). *Statistics for directional data*. Academic press.
- [78] Mardia, K. and Jupp, P. (2009). *Directional statistics*, volume 494. Wiley.
- [79] Marsaglia, G., Tsang, W. W., and Wang, J. (2003). Evaluating kolmogorov’s distribution. *Journal of Statistical Software*, 8(18):1–4.
- [80] Mashanova, A., Oliver, T. H., and Jansen, V. A. (2010). Evidence for intermittency and a truncated power law from highly resolved aphid movement data. *Journal of The Royal Society Interface*, 7(42):199–208.

- [81] McCulloch, J. H. (1986). Simple consistent estimators of stable distribution parameters. *Communications in Statistics-Simulation and Computation*, 15(4):1109–1136.
- [82] McCulloch, J. H. (1998). Linear regression with stable disturbances. *Adler et al.(1998)*, pages 359–376.
- [83] McHale, I. G. and Laycock, P. J. (2006). Applications of a general stable law regression model. *Journal of Applied Statistics*, 33(10):1075–1084.
- [84] Meerschaert, M. M., Roy, P., and Shao, Q. (2009). Parameter estimation for tempered power law distributions. *Communications in Statistics–Theory and Methods*.
- [85] Milne-Thomson, L., Abramowitz, M., and Stegun, I. (1972). Handbook of mathematical functions. *Handbook of Mathematical Functions*.
- [86] Mittnik, S., Doganoglu, T., Chenyao, D., et al. (1999). Maximum likelihood estimation of stable paretian models. *Mathematical and Computer Modelling*, 29(10):275–293.
- [87] Mittnik, S. and Paoella, M. S. (2003). Prediction of financial downside-risk with heavy-tailed conditional distributions. *Handbook of Heavy Tailed Distributions in Finance, Elsevier, New York*, pages 385–404.
- [88] Mittnik, S., Paoella, M. S., and Rachev, S. T. (2002). Stationarity of stable power-garch processes. *Journal of Econometrics*, 106(1):97–107.
- [89] Mittnik, S., Rachev, S. T., and Paoella, M. S. (1998). Stable paretian modeling in finance: Some empirical and theoretical aspects. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications. Boston: Birkhaüser*, pages 79–110.
- [90] Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., and Fryxell, J. M. (2004). Extracting more out of relocation data: Building movement models as mixtures of random walks. *Ecology*, 85(9):2436–2445.
- [91] Nathan, R. (2008). An emerging movement ecology paradigm. *Proceedings of the National Academy of Sciences*, 105(49):19050–19051.
- [92] Nolan (1998). Parameterizations and modes of stable distributions. *statistics and probability letters*, 38:187–195.
- [93] Nolan, J. (2003). *Stable distributions: models for heavy-tailed data*. Birkhauser.
- [94] Nolan, J. P. (2001). Maximum likelihood estimation of stable parameters. *Levy processes: Theory and applications*, pages 379–400.

-
- [95] Nolan, J. P. and Ojeda-Revah, D. (2013). Linear and nonlinear regression with stable errors. *Journal of Econometrics*, 172(2):186–194.
- [96] Osorio, F. and Galea, M. (2006). Detection of a change-point in student-t linear regression models. *Statistical Papers*, 47(1):31–48.
- [97] Otieno, B. S. and Anderson-Cook, C. M. (2006). Measures of preferred direction for environmental and ecological circular data. *Environmental and Ecological Statistics*, 13(3):311–324.
- [98] Owen-Smith, N., Fryxell, J., and Merrill, E. (2010). Foraging theory upscaled: the behavioural ecology of herbivore movement. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1550):2267–2278.
- [99] Owen-Smith, N., Goodall, V., and Fatti, P. (2012). Applying mixture models to derive activity states of large herbivores from movement rates obtained using gps telemetry. *Wildlife Research*, 39(5):452–462.
- [100] Pewsey, A., Neuhäuser, M., and Ruxton, G. D. (2013). *Circular statistics in R*. Oxford University Press.
- [101] Plank, M. J. and Codling, E. A. (2009). Sampling rate and misidentification of lévy and non-lévy movement paths. *Ecology*, 90(12):3546–3553.
- [102] Press, S. J. (1972). Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association*, 67(340):842–846.
- [103] Redfern, J. V., Grant, R., Biggs, H., and Getz, W. M. (2003). Surface-water constraints on herbivore foraging in the kruger national park, south africa. *Ecology*, 84(8):pp. 2092–2107.
- [104] Reynolds, A. M. (2006). On the intermittent behaviour of foraging animals. *Europhysics Letters*, 5(4):517–520.
- [105] Reynolds, A. M. (2010). Bridging the gulf between correlated random walks and lévy walks: autocorrelation as a source of lévy walk movement patterns. *Journal of The Royal Society Interface*, 7(53):1753–1758.
- [106] Reynolds, A. M. and Rhodes, C. J. (2009). The lévy flight paradigm: Random search patterns and mechanisms. *Ecology*, 90(4):pp. 877–887.
- [107] Ropert-Coudert, Y. and Wilson, R. P. (2005). Trends and perspectives in animal-attached remote sensing. *Frontiers in Ecology and the Environment*, 3(8):437–444.
- [108] Rosadi, D. (2012). The covariation function for symmetric α -stable random variables with finite first moments. *Journal of the Indonesian Mathematical Society*, 15(1):1.

- [109] Russell, G. S. and Levitin, D. J. (1995). An expanded table of probability values for rao's spacing test. *Communications in Statistics-Simulation and Computation*, 24(4):879–888.
- [110] Samoradnitsky, G. and Taqqu, M. S. (1994). *Stable non-Gaussian random processes: stochastic models with infinite variance*, volume 1. CRC Press.
- [111] Schou, G. (1978). Estimation of the concentration parameter in von mises–fisher distributions. *Biometrika*, 65(2):369–377.
- [112] Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461–464.
- [113] SenGupta, A. and Ugwuowo, F. I. (2006). Asymmetric circular-linear multivariate regression models with applications to environmental data. *Environmental and Ecological Statistics*, 13(3):299–309.
- [114] Shannon, G., Matthews, W. S., Page, B. R., Parker, G. E., and Smith, R. J. (2009). The affects of artificial water availability on large herbivore ranging patterns in savanna habitats: a new approach based on modelling elephant path distributions. *Diversity and Distributions*, 15(5):776–783.
- [115] Sherman, B. (1950). A random variable related to the spacing of sample values. *The Annals of Mathematical Statistics*, 21(3):339–361.
- [116] Shlesinger, M. F., West, B. J., and Klafter, J. (1987). Lévy dynamics of enhanced diffusion: Application to turbulence. *Physical Review Letters*, 58(11):1100–1103.
- [117] Sims, D. W., Righton, D., and Pitchford, J. W. (2007). Minimizing errors in identifying lévy flight behaviour of organisms. *Journal of Animal Ecology*, 76(2):222–229.
- [118] Singh, R. S. (1988). Estimation of error variance in linear regression models with errors having multivariate student- t distribution with unknown degrees of freedom. *Economics Letters*, 27(1):47–53.
- [119] Siniff, D. B. and Jessen, C. (1969). A simulation model of animal movement patterns. In Cragg, J., editor, *Advances in Ecological Research*, volume 6 of *Advances in Ecological Research*, pages 185–219. Academic Press.
- [120] Solow, A. R. (1990). A note on the statistical properties of animal locations. *Journal of Mathematical Biology*, 29:189–193.
- [121] Stephens, M. A. (1970). Use of the kolmogorov-smirnov, cramér-von mises and related statistics without extensive tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 115–122.

- [122] Stephens, M. A. (1974). Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737.
- [123] Swihart, R. K. and Slade, N. A. (1985). Influence of sampling interval on estimates of home range size. *Journal of Wildlife Management*, 49(4):1019–1025.
- [124] Tracey, J. A., Zhu, J., and Crooks, K. (2005). A set of nonlinear regression models for animal movement in response to a single landscape feature. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(1):1–18.
- [125] Tracey, J. A., Zhu, J., and Crooks, K. R. (2011). Modeling and inference of animal movement using artificial neural networks. *Environmental and Ecological Statistics*, 18(3):393–410.
- [126] Turchin, P., Odendaal, F., and Rausher, M. (1991). Quantifying insect movement in the field. *Environmental Entomology*, 20(4):955–963.
- [127] Uchaikin, V. V. and Zolotarev, V. M. (1999). *Chance and Stability: Stable Distributions and their applications*. Walter de Gruyter.
- [128] Upton, G. J. (1973). Single-sample tests for the von mises distribution. *Biometrika*, 60(1):87–99.
- [129] Upton, G. J. (1986). Distance and directional analyses of settlement patterns. *Economic Geography*, pages 167–179.
- [130] Urbano, F., Cagnacci, F., Calenge, C., Dettki, H., Cameron, A., and Neteler, M. (2010). Wildlife tracking data management: a new vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1550):2177–2185.
- [131] Vanak, A., Thaker, M., and Slotow, R. (2010). Do fences create an edge-effect on the movement patterns of a highly mobile mega-herbivore? *Biological Conservation*, pages 2631–2637.
- [132] Viswanathan, G. M., Afanasyev, V., Buldyrev, S. V., Murphy, E. J., Prince, P. A., and Stanley, H. E. (1996). Lévy flight search patterns of wandering albatrosses. *Nature*, 381(6581):413–415.
- [133] Viswanathan, G. M., Buldyrev, S. V., Havlin, S., da Luz, M. G. E., Raposo, E. P., and Stanley, H. E. (1999). Optimizing the success of random searches. *Nature*, 401(6756):911–914.
- [134] Viswanathan, G. M., Raposo, E. P., Bartumeus, F., Catalan, J., and da Luz, M. G. E. (2005). Necessary criterion for distinguishing true superdiffusion from correlated random walk processes. *Physical Review. E*, 72:011111.

-
- [135] Wall, J., Wittemyer, G., Klinkenberg, B., LeMay, V., and Douglas-Hamilton, I. (2013). Characterizing properties and drivers of long distance movements by elephants (*Loxodonta africana*) in the gourma, mali. *Biological Conservation*, 157:60–68.
- [136] Walls, W. D. (2005). Modelling heavy tails and skewness in film returns. *Applied Financial Economics*, 15(17):1181–1188.
- [137] Watson, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika*, pages 109–114.
- [138] Weron, R. (2001). Levy-stable distributions revisited: tail index > 2 does not exclude the levy-stable regime. *International Journal of Modern Physics C*, 12(02):209–223.
- [139] Wittemyer, G., Getz, W., Vollrath, F., and Douglas-Hamilton, I. (2007). Social dominance, seasonal movements, and spatial segregation in african elephants: a contribution to conservation behavior. *Behavioral Ecology and Sociobiology*, 61(12):1919–1931.
- [140] Worton, B. (1987). A review of models of home range for animal movement. *Ecological Modelling*, 38(3-4):277–298.
- [141] Wuertz, D. (2009). fbasics: Rmetrics-markets and basic statistics. *R package version*, 2100.
- [142] Zar, J. H. (1999). *Biostatistical analysis*, volume 564. Prentice hall Upper Saddle River, NJ.
- [143] Zolotarev, V. M. (1986). *One-dimensional stable distributions*, volume 65. American Mathematical Society.

Appendix

circular codes

```
library(circular)
circ.lin.reg <- function(x, theta, beta0, trace = FALSE, print = TRUE, tol = 1e-10, maxiter=1000)
{
  if(is.vector(x))
    x <- cbind(x)
  n <- length(theta)
  betaPrev <- coef(lm(theta~x))[2:(dim(x)[2]+1)]
  # CHANGE 1:1 TO 1:4 FOR DIFFERENT STARTING VALUES
  for(i in 1:1) {
    if(i==2) betaPrev <- betaPrev
    if(i==3) betaPrev <- betaPrev+1
    if(i==4) betaPrev <- betaPrev-1
    if(i==1) betaPrev <- rep(0,length(betaPrev))
    S <- sum(sin(theta - 2 * atan(x %*% betaPrev)))/n
    C <- sum(cos(theta - 2 * atan(x %*% betaPrev)))/n
    R <- sqrt(S^2 + C^2)
    mu <- atan2(S, C)
    k <- A1inv(R)
    diff <- tol + 1
    iter <- 0
    S.function <- function(betaPrev,x) {
      2/(1 + (t(betaPrev) %*% x)^2) }
    while(diff > tol & iter < maxiter) {
      iter <- iter + 1
      u <- k * sin(theta - mu - 2 * atan(x %*% betaPrev))
```

```

A <- diag(k * A1(k), nrow = n)

g.p <- diag(apply(x, 1, S.function, betaPrev = betaPrev), nrow = n)
D <- g.p %**% x
betaNew <- lm(t(D) %**% (u + A %**% D %**% betaPrev) ~ t(D) %**% A %**% D - 1)$coefficients
diff <- abs(max(betaNew - betaPrev))
brokened = 0
if(iter==1000) {
  brokened = 1
}
if(is.na(diff)==TRUE) {
  brokened = 1
  break
}
if(max(betaNew) > 100) {
  brokened = 1
  break
}
betaPrev <- betaNew
S <- sum(sin(theta - 2 * atan(x %**% betaPrev)))/n
C <- sum(cos(theta - 2 * atan(x %**% betaPrev)))/n
R <- sqrt(S^2 + C^2)
mu <- atan2(S, C)
# mu <- asin(S/R)
k <- A1inv(R)
if(trace == T) {
  log.lik <- - n * log(2*pi*I.0(k)) + k * sum(cos(theta -
    mu - 2 * atan(x %**% betaNew)))
  cat("Iteration ", iter, ":    Log-Likelihood = ",
    log.lik,"mu ", mu, "k ", k, "b ", betaNew, "\n")
  cat("Starting values ", i, "\n")
}
}
log.lik <- - n * log(2*pi*I.0(k)) + k * sum(cos(theta - mu - 2 * atan(
  x %**% betaNew)))
log.lik.old <- - n * log(2*pi*I.0(k)) + k * sum(cos(theta - mu - 2 * atan(
  x %**% betaPrev)))
cov.beta <- solve(t(D) %**% A %**% D)

```

```

se.beta <- sqrt(diag(cov.beta))
se.kappa <- sqrt(1/(n * (1 - A1(k)^2 - A1(k)/k)))
circ.se.mu <- 1/sqrt((n - ncol(x)) * k * A1(k))
z.values <- abs(betaNew/se.beta)
p.values <- (1 - pnorm(z.values))*2
result.matrix <- cbind(Coef = betaNew, SE = se.beta, Z = z.values,
                      p = p.values)
dimnames(result.matrix) <- list(dimnames(x)[[2]], c("Coef", "SE", "|z|",
                                                  "p"))
cat("\n", "Circular-Linear Regression", "\n", "\n")
print(result.matrix)
cat("\n", "\n")
betaNew <- as.matrix(betaNew)
dimnames(betaNew) <- list(dimnames(x)[[2]], c("Estimate"))
list(mu = mu, kappa = k, beta = betaNew, log.lik = log.lik, log.lik.old = log.lik.old,
     circ.se.mu = circ.se.mu, se.kappa = se.kappa, cov.beta = cov.beta,
     se.beta = se.beta, result.matrix = result.matrix, breaked=breaked)
}
}

kappa <- out$kappa
kappa.var <- (out$se.kappa)^2
out$mu <- 2*pi - abs(out$mu)
out.par <- as.vector(c(out$mu,out$beta))
varcov.mat <- matrix(0,length(out$beta)+1,length(out$beta)+1)
varcov.mat[1,1] <- (out$circ.se.mu^2)/10
varcov.mat[2:(length(out$beta)+1),2:(length(out$beta)+1)] <- out$cov.beta
se <- sqrt(diag(abs(varcov.mat)))
z.val <- abs(out.par/se)
par.list <- cbind(out.par,as.vector(se),as.vector(z.val))
rownames(par.list) <- c("mu",colnames(cbind(X)))
colnames(par.list) <- c("coef", "se", "|z|")
print(round(par.list,3))
var.mat <- diag(1,dim(varcov.mat)[1]) * diag(varcov.mat)

out$mu / out$circ.se.mu
out$kappa / out$se.kappa

```

```

conf.level <- 0.95

mlogl <- function(alpha, x) {
  if (length(alpha) < 1) stop("alpha must be scalar")
  if (alpha <= 0) stop("alpha must be positive")
  return(- sum(dgamma(x, shape = alpha, log = TRUE)))
}

n <- length(x)
out <- nlm(mlogl, mean(x), x = x, hessian = TRUE,
          fscale = n)
alpha.hat <- out$estimate
z <- qnorm((1 + conf.level) / 2)
# confidence interval using expected Fisher information
alpha.hat + c(-1, 1) * z / sqrt(n * trigamma(alpha.hat))
# confidence interval using observed Fisher information
alpha.hat + c(-1, 1) * z / sqrt(out$hessian)

mean(log.lik[index.new]) # MEAN LL FULL MODEL
out.null$log.lik # LL NULL MODEL
(dev1 = 2 * (mean(log.lik[index.new]) - out.null$log.lik)) # Fahrmeier/Tutz DEVIANCE
(dev2 = 2 * (-mean(log.lik[index.new]))) # RESIDUAL DEVIANCE
(dev3 = 2 * (0 - out.null$log.lik)) # NULL DEVIANCE
(AIC <- 2*dim(X)[2] - 2*mean(log.lik[index.new])) # AIC

```

Chapter 4 time series analysis

```

# This examples uses the dataset of the package
fGarch to estimate # an ARMA(1,1)-GARCH(1,1) with GEV conditional
distribution. library(fGarch) data(dem2gbp) x = dem2gbp[, 1] gF.new
= GSgarch.Fit(data = x , 1,1,1,1, cond.dist = "sstd", intercept =
TRUE, APARCH = FALSE, algorithm = "nlminb", printRes = TRUE, get.res
= TRUE)

library(fGarch) data(dem2gbp) x = dem2gbp[, 1] fitarmagarch0=

```

```
GSgarch.Fit(data = Lx1 , 0,0,1,2, cond.dist = "stable", intercept =  
TRUE, APARCH = FALSE, algorithm = "nlminb", printRes = TRUE, get.res  
= TRUE)
```

```
##### #Fitting  
stable paretian ARMA-GARCH(1,0,1,2) model  
##### fitgarch1=  
GSgarch.Fit(data = Lx1 , 1,0,1,2, cond.dist = "stable", intercept =  
TRUE, APARCH = FALSE, algorithm = "nlminb", printRes = TRUE, get.res  
= TRUE)
```

```
fitgarch2= GSgarch.Fit(data = Lx2 , 1,0,1,2, cond.dist = "stable",  
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =  
TRUE, get.res = TRUE)
```

```
fitgarch3= GSgarch.Fit(data = Lx3 , 1,1,1,2, cond.dist = "stable",  
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =  
TRUE, get.res = TRUE)
```

```
fitgarch4= GSgarch.Fit(data = Lx4 , 1,0,1,2, cond.dist = "stable",  
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =  
TRUE, get.res = TRUE)
```

```
fitgarch5= GSgarch.Fit(data = Lx5 , 1,0,1,2, cond.dist = "stable",  
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =  
TRUE, get.res = TRUE)
```

```
fitgarch6= GSgarch.Fit(data = Lx6 , 1,0,1,2, cond.dist = "stable",  
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =  
TRUE, get.res = TRUE)
```

```
##### #AIC for the  
stable paretian GARCH models
```

```
#####  
AIC1<-GSgarch.FitAIC(data = Lx1,1,0,1,2,cond.dist = "stable")  
AIC2<-GSgarch.FitAIC(data = Lx2,1,0,1,2,cond.dist = "stable")  
AIC3<-GSgarch.FitAIC(data = Lx3,1,0,1,2,cond.dist = "stable")  
AIC4<-GSgarch.FitAIC(data = Lx4,1,0,1,2,cond.dist = "stable")  
AIC5<-GSgarch.FitAIC(data = Lx5,1,0,1,2,cond.dist = "stable")  
AIC6<-GSgarch.FitAIC(data = Lx6,1,0,1,2,cond.dist = "stable")  
  
fitgarch1= GSgarch.Fit(data = Lx1 , 1,1,1,1, cond.dist = "sstd",  
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =  
TRUE, get.res = TRUE)  
  
fitgarch2= GSgarch.Fit(data = Lx2 , 1,1,1,1, cond.dist = "sstd",  
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =  
TRUE, get.res = TRUE)  
  
fitgarch3= GSgarch.Fit(data = Lx3 , 1,1,1,1, cond.dist = "sstd",  
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =  
TRUE, get.res = TRUE)  
  
fitgarch4= GSgarch.Fit(data = Lx4 , 1,1,1,1, cond.dist = "sstd",  
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =  
TRUE, get.res = TRUE)  
  
fitgarch5= GSgarch.Fit(data = Lx5 , 1,1,1,1, cond.dist = "stable",  
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =  
TRUE, get.res = TRUE)  
  
fitgarch6= GSgarch.Fit(data = Lx6 , 1,1,1,1, cond.dist = "std",  
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =  
TRUE, get.res = TRUE) n1<-length(Lx1) plot(sort(Lx1), (1:n1/n1),  
main = "Probability", col = "steelblue4") lines(Lx1, pstable(x =  
Lx1, alpha = 1.67))  
  
hist(Lx1, breaks=200,xlim=c(0,1500), probability = TRUE, border =  
"white", col = "steelblue4") x = seq(min(Lx1), max(Lx1), 0.02)
```

```
lines(x, dstable(x =x, alpha = 1.67, beta =1),col="red")

fit1<-garchFit(Lx1~arma(1,1)+ garch(1,1),data=Lx1 ,cond.dist="sstd"
,trace=FALSE)

fitarmagarch1= GSgarch.Fit(data = Lx1 , 0,0,1,2, cond.dist = "gev",
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =
TRUE, get.res = TRUE)

fitarmagarch2= GSgarch.Fit(data = Lx2 , 0,0,1,2, cond.dist = "gev",
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =
TRUE, get.res = TRUE)

fitarmagarch3= GSgarch.Fit(data = Lx3 , 0,0,1,2, cond.dist = "gev",
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =
TRUE, get.res = TRUE)

fitarmagarch4= GSgarch.Fit(data = Lx4 , 0,0,1,2, cond.dist = "gev",
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =
TRUE, get.res = TRUE)

fitarmagarch5= GSgarch.Fit(data = Lx5 , 0,0,1,2, cond.dist = "gev",
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =
TRUE, get.res = TRUE)

fitarmagarch6= GSgarch.Fit(data = Lx6 , 0,0,1,2, cond.dist = "gev",
intercept = TRUE, APARCH = FALSE, algorithm = "nlminb", printRes =
TRUE, get.res = TRUE)

fitsst1<-garchFit(x = "Lx1", cond.dist = "sstd") fitsst2<-garchFit(x
= "Lx2", cond.dist = "sstd") fitsst3<-garchFit(x = "Lx3", cond.dist
= "sstd") fitsst4<-garchFit(x = "Lx4", cond.dist = "sstd")
fitsst5<-garchFit(x = "Lx5", cond.dist = "sstd") fitsst6<-garchFit(x
= "Lx6", cond.dist = "sstd")

fitstd1<-garchFit(x = "Lx1", cond.dist = "std") fitstd2<-garchFit(x
= "Lx2", cond.dist = "std") fitstd3<-garchFit(x = "Lx3", cond.dist =
```

```
"std") fitstd4<-garchFit(x = "Lx4", cond.dist = "std")
fitstd5<-garchFit(x = "Lx5", cond.dist = "std") fitstd6<-garchFit(x
= "Lx6", cond.dist = "std")
```

```
fitsnormd1<-garchFit(x = "Lx1", cond.dist = "snorm")
fitsnormd2<-garchFit(x = "Lx2", cond.dist = "snorm")
fitsnormd3<-garchFit(x = "Lx3", cond.dist = "snorm")
fitsnormd4<-garchFit(x = "Lx4", cond.dist = "snorm")
fitsnormd5<-garchFit(x = "Lx5", cond.dist = "snorm")
fitsnormd6<-garchFit(x = "Lx6", cond.dist = "snorm")
```

```
fitsnigd1<-garchFit(x = "Lx1", cond.dist = "snig")
fitsnigd2<-garchFit(x = "Lx2", cond.dist = "snig")
fitsnigd3<-garchFit(x = "Lx3", cond.dist = "snig")
fitsnigd4<-garchFit(x = "Lx4", cond.dist = "snig")
fitsnigd5<-garchFit(x = "Lx5", cond.dist = "snig")
fitsnigd6<-garchFit(x = "Lx6", cond.dist = "snig")
```