TESI DI DOTTORATO

Dipartimento di Scienze Economiche, Aziendali e Statistiche

Modelling Spatio-Temporal Elephant Movement Data: a Generalized Additive Mixed Models Framework

Modellazione Spazio-Temporale del Movimento degli Elefanti in un approccio attraverso i Modelli Additivi Generalizzati Misti

Angela VITRANO

Tutor: *Dr. Vito M. R. Muggeo*

Coordinatore Dottorato: *Prof. Marcello Chiodi*

Dottorato di Ricerca in "Statistica, Statistica Applicata e Finanza Quantitativa", XXIV Ciclo - 2013
Settore Scientifico Disciplinare: SECS-S/01 - Statistica

Università degli Studi di Palermo

To my family

Acknowledgements

I would like to give my most sincere thanks to my supervisor, Dr. Vito Muggeo, who continuously guided and supported my PhD project with his proficiency, constant inspiration, interest and contributions. Special thanks also go to Professor Simon Wood, for his hospitality at the University of Bath (UK) during my PhD abroad and who helped to improve the work with his comments on some points of statistical methodology and on the use of his R package, mgcv. I would also like to thank Dr. Patricia Birkett from the University of Kwazulu-Natal, Durban, South Africa, for sharing spatio-temporal elephant movement data, collected by researchers at her University, and for her useful and fruitful discussion about data and results. I wish to extend my acknowledgement to all the members of the Department of "Statistical and Mathematical Science S. Vianelli" for their hospitality and tireless everyday help. A big thank you to my office mates for the friendly climate in which this thesis was prepared. In particular, a special thank goes to Dr. Clara Romano and Fabio Tuzzolino, for their continued support during my PhD study. I would like to extend my gratitude also to my friends for their encouragement, understanding and extraordinary support. Most important, I am extremely grateful to my mother and father. Their unreserved love, support and much more... during these years almost

always far from their home is what makes this thesis valuable, and I would like to dedicate this thesis to them and my siblings.

Abstract

This thesis focuses on understanding how environmental factors influence elephant movement and in investigating the spatio-temporal patterns.
The thesis analyses movement data of some African elephants (*Loxodonta Africana*) living in the Kruger National Park and its associated private reserves of South Africa. Due to heterogeneity among elephants, and nonlinear relationships between elephant movement and environmental variables, Generalized Additive Mixed Models (GAMMs) were employed. Results showed delayed effects of rainfall and temperature and particular trends in time and space.

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Literature review 2

1.3 Biological and Ecological Background 7

1.4 The thesis contribution 9

2 Data and some exploratory analyses **13**

2.1 Study Area and Elephant Collaring 13

2.2 Elephant Data 16

2.3 Environmental data 24

3 Penalized Spline Smoothing **33**

3.1 Idea of penalized smoothing 33

4 Generalized Additive (Mixed) Models **47**

4.1 Penalized smooth term as Mixed model representation . . 47

4.2 Generalized Additive Mixed Models 50

4.3 Parameter Estimation 51

4.3.1 GAMM estimation based on GLMMs settings . . . 51

4.3.2 GAMM estimation based on GAMs settings 52

5 Modelling spatio-temporal elephant movement data 61

5.1 Introduction 61

5.2 The proposed model framework 62

5.2.1 A three-dimensional spatio-temporal smoother . . 64

5.3 Results 66

5.3.1 Model diagnostics 73

6 A possible improvement 79

6.1 Including *cumulative rainfall* 79

7 Discussion 85

Bibliography 101

List of Figures

2.1 Kruger National Park situated in north-eastern South Africa. The Kruger is represented by the dark green area in the north-eastern boundaries of South Africa adjoining at East with the Monzambique. 14

2.2 Elephant distribution in the KNP from January 1, 2006 to January 17, 2010. Each small grey dot represents the daily average elephant location and each black dot represents the rainfall station distributed in the Kruger. 19

2.3 Trajectories of the 11 studied elephants in the KNP from January 1, 2006 to January 17, 2010. The initial and final locations are indicated with a blue and a red triangle, respectively. 20

2.4 Kernel density estimates of speed for each elephant. 22

2.5 Daily time series of individual elephant movement from January 1, 2006 to January 17, 2010. The dashed red lines represent the years from 2006 to 2010. 23

2.6 Individual elephant movement profiles versus month. Each elephant is represented by a different color. 24

2.7 Observed local rainfall time series for each elephant across years. The dashed red lines represent the years from 2006 to 2010. 26

2.8 Observed temperature time series for each elephant across years. The dashed red lines represent the years from 2006 to 2010. 27

2.9 Scatterplot of mean daily speed and local rainfall for each elephant, from January 1, 2006 to January 17, 2010. 29

2.10 Scatterplot of mean daily speed and temperature for each elephant, from January 1, 2006 to January 17, 2010. 30

2.11 Scatter plot of local rainfall and temperature. 31

3.1 Penalized regression spline fits to simulated data using three different values for the smoothing parameter. 45

5.1 Baysian Information Criteria values for the Tweedie distribution from a grid over 100 values for the p index. The red dashed line represents the lowest BIC value at the p index for the Tweedie distribution ($p = 1.83$). 63

5.2 Estimated seasonality effect. The dashed red lines indicate the calendar seasonal thresholds (October-March, wet season; April-September, dry season). 69

5.3 Estimated rainfall effect at lag 0, 1 and lag 2, with ± two standard errors (grey area). 70

5.4 Estimated rainfall effect at lag 0, 1 and lag 2, with ± two standard errors (grey area). 71

5.5 Estimated elephants movements over different years of study (2006-2010). Lighter colours represent areas with higher movements. 72

5.6 Model validation plots of the selected model. 74

5.7 Variograms of standardized residuals for each elephant. The envelops represent minimum and maximum values derived by permuting data points on the spatial locations for each elephant. The variograms indicate model mis-specification if the empirical points lie outside the envelops. 76

5.8 Partial autocorrelation functions of standardized residuals for each elephant. 77

5.9 Partial autocorrelation functions of standardized residuals for each elephant for the selected model with a specified autoregressive structure of order 1 for the within-group error. The estimated correlation parameter is $\hat{\rho} = 0.26$. . . . 78

6.1 Estimated lagged effects of rainfall (left column) and temperature (right column) with ± two standard errors (shade grey area). 82

6.2 Estimated seasonality and cumulative rainfall effects with ± two standard errors (shade grey area). The dashed red lines indicate the calendar seasonal thresholds (October-March, wet season; April-September, dry season). 83

List of Tables

2.1 *Collared* elephants by area in KNP. 16

2.2 Number of observations for each elephant by year. 18

2.3 Names of rainfall stations distributed in KNP. 25

2.4 List of explanatory variables and their used abbreviations. . 28

5.1 Comparing models with different lag structures for rainfall and temperature. 67

5.2 Type of basis and size per smooth term of the selected model. 67

5.3 Main results from the fitted model M_4. 68

6.1 Main results from the fitted model in (6.1). 81

Chapter 1

Introduction

1.1 Motivation

In movement ecology and other applied research fields, studying the environmental factors influencing animal movement is important to understand drivers of animal movement. Studies relating fine-scale movement paths to spatio-temporally structured landscape data, such as vegetation productivity or human activity, are particularly lacking despite the obvious importance of such information. In part, this may be due to the fact that few approaches have the ability to characterize the complexity of movement behavior and relate it to diverse, varying environmental stimuli. A large amount of statistical literature is devoted to animal movement data analyses. These techniques, which are mainly based on random walks and diffusion processes, such as Correlated Random walks, Hidden Markov Models, Ornstein-Uhlenbeck processes, and Levy Walks, provide knowledge on ecological systems and they are instrumental in understanding how populations might respond spatially to threatened or fragmented landscapes

(Kareiva and Wennergren, 1995), but they do not give information on how landscape features actually influence the movement process. Nowdays, understanding how organisms explore and exploit their environment, and assessing the rule of environmental factors is a central topic in Ecology. Recent advances in global positioning system (GPS) radio telemetry provide data on individual movements for many species. Such information is increasing the ability to better understand patterns in animal movement data.

1.2 Literature review

Many different approaches have used time-series methods to model animal movement data. We briefly review them discussing relevant advantages and disadvantages. These methods include models based on random walks and diffusion processes, such as Correlated Random walks, Hidden Markov Models, Ornstein-Uhlenbeck processes, Levy Walks; there are also recent progresses in Hierarchical Bayesian State Space models and Stochastic Differential Equations. Although these are not necessarily modern statistical tools, their application to animal tracking is very recent, and research is ongoing with respect to their suitability and applicability for animal movement data analyses. A common general model of animal movement is the *Correlated Random Walk* (CRWs) (Kareiva and Shigesada, 1983; Bovet and Benhamou, 1988), which hypothesizes some distribution of step-lengths and turning angles (angles between successive steps in the tracks). The location of an animal at any time is dependent upon its position in previous time periods. Bovet and Benhamou (1988) developed a first order correlated random walk where the directions of successive moves are correlated and this introduces a persistence to move forward. However, several important features of movement data complicate the straightforward application of CRWs. The first one is the error in the measurement process. A fruitful body of research, that addresses measurement error with the use of state-space models (Jonsen *et al.*, 2006; Patterson *et al.*, 2008), has emerged recently. A second important and commonly encountered feature of movement data that confounds the application of CRWs is the irregular timing of measurements. One of the advantages of this method is that it is possible to compare different animals within the same environment or the same species across different environments. A disadvantage is that the model does not include landascape features that might influence the animal movement.

Dunn and Gipson (1977) and Dunn and Brisbin (1985) employed *Ornstein-Uhlenbeck process* (O-U). This process is effective in modelling animal movements where animals are attracted towards a central point (Preisler *et al.*, 2004). The O-U process is Markovian in continuous time, with states given by the locations of the animal. While this approach allows to account for dependence between observations (Blackwell, 1997; Beichelt and Fatti, 2002), a disadvantage of O-U process is that the stationary distribution is always Gaussian, which limits its flexibility (Blackwell, 1997) and there is no way to incorporate any behavioural information. However, the Ornstein-Uhlenbeck process can be generalised to describe different movements for different behavioural states and can have a number of different "centres of attraction" (Blackwell, 1997). Blackwell (2003) used bivariate Ornstein-Uhlenbeck process to model movement using radio tracking data from a single mouse, while Preisler *et al.* (2004) used telemetry data for elks.

Another form of a random walk is *Lévy walk*, where the successive steps are distributed according to a power-law (or long-tailed) distribution of the turning angles (Bartumeus *et al.*, 2005). Lévy walks are Markov processes and they are similar to classical random walks as they are uncorrelated, but they have an infinite step-length variance (Benhamou, 2007). As for correlated random walks, the Lévy walk incorporates animal tendency to continue to move in a specific direction between successive steps. This directional persistence is introduced into the model through the power-law distribution of move lengths (Bartumeus *et al.*, 2005). Viswanathan *et al.* (1999) tested the theory of Lévy walk by analysing experimental foraging data from selected insects, mammals and bird species and found that the movement was consistent with the power-law distribution. A difficult and challenging issue is to identify how the pattern was actually generated (Benhamou, 2007): in fact not all seemingly Lévy walk patterns are necessarily produced by Lévy walk processes. Benhamou (2007) further discussed that a disadvantage of applying this method to animal tracking data captured at equal time intervals is that the step length corresponds to speed travelled in a pre-determined time rather than to distances between ecologically meaningful events.

Franke *et al.* (2004) used multiple-observation *Hidden Markov models* as an individual-based predictive modelling to explain the use of space, movement and behavior of caribou (*Rangifer tarandus*) in central Alberta, Canada. They defined the “hidden” states as the animal bedding, feeding and relocating and they assumed that distance-between-location and turn-angle were suitable “observations” for encapsulating movement behaviour and use of space. This model allowed them to estimate inferred behavioural states, their relative bout length and transitions as well as the most likely behavioural state. The authors described an advantage of hidden Markov models over other modelling techniques being that, assuming the states are known, hidden Markov models are able to provide the optimal state sequence from the observed data through optimal inference of a casual model (Franke *et al.*, 2004). A disadvantage is that the movement model has to be defined a *priori* as a correlated random walk with a hidden discrete variable (behavioural state).

Hierarchical Bayesian State-Space Models (HBSSMs) represent an advance in animal movement modelling (Jonsen *et al.*, 2006). These models combine a statistical model of the observation method with a model of the movement dynamics, which can include effects due to the environment and other variable factors (Patterson *et al.*, 2008). By employing a hierarchical structure, inference on movement processes is carried out straightforwardly. Without such a structure, it is difficult to make inference on the movement process underlying the inherently messy movement data. Jonsen *et al.* (2006) used robust hierarchical Bayes state-space models to test the hypothesis according to which leatherback turtles (*Dermochelys coriacea*) off the coast of Canada travel faster during the day, when the turtles are closer to the water surface, than at night, when they dive to greater dephts. A secondary analysis was to determine whether differences could be seen in travel rates between males and females, and between those who were breeding and those who were not. The authors used a fully Bayesian state-space approach which enabled them to combine individual results in order to analyse ‘among individual’ variation in movements rates.

Yet another approach in recent literature is based on the use of *Stochastic Differential Equations*. Preisler *et al.* (2001) used Stochastic Differential Equations (SDEs) to characterize the direction and speed of animal movements and to study the effects of explanatory variables (e.g., habitat characteristics) on movement patterns. Analyses of animal movements demand the use of complex models and computationally intense techniques. A univariate stochastic differential equation (SDE) is defined by

$$dY(t) = \mu(Y, t, \theta)dt + \sigma(Y, t, \theta)dB(t) \tag{1.1}$$

where $Y(t)$ is a random variable, $\{B(t), t = 0\}$ is a random process, and θ is a set of known and unknown parameters. The term

$$\mu(Y, t, \theta) = E\{dY(t)|Y(s), s < t\}/dt$$

is interpreted as the instantaneous velocity of the individual (drift coefficient), and $\sigma(Y, t, \theta) = se\{dY(t)|Y(s), s < t\}/dt$ is interpreted as the speed or the diffusion coefficient.

The simplest model for the SDE in (1.1) is a pure diffusion model where $\mu(Y, t, \theta) = 0$ and $B(t)$ is a Brownian process, i.e., each individual's movement is a random walk independent of others. Another special case of (1.1) is the mean-reverting Ornstein-Uhlenbeck (O-U) process where $\mu(Y, t, \theta) = \alpha[T(t) - a]$ and $\sigma(Y, t, \theta)dt = \sigma^2$. The O-U process was used to estimate home ranges of animals where a is the center of the home range (Turchin, 1998; Dunn and Brisbin, 1985). More complicated animal movement behavior may be studied by modelling the drift and diffusion coefficients as functions of explanatory variables. Bengtsson *et al.* (2002) modelled the drift term as a function of the distance between individuals in their attempt to characterize dispersal patterns of soil-living invertebrates. In the bark beetle example presented below, Preisler and Akers (1995) modelled the drift term as a function of the heading angle between the direction along the path of female beetles and a point source emitting male pheromones. Preisler *et al.* (2001) used bivariate SDEs to study trajectories of radio-collared elks and deers as they forage in Oregon. Preisler *et al.* (2004) used bivariate stochastic differential equations (SDEs) to model movements of

216 radiocollared female Rocky Mountain elk at the Starkey Experimental Forest and Range in Northeastern Oregon. Using the concept of a potential function, they succeeded in studying the influence of roads and grassland foraging areas on elk movements. They identified broad spatial patterns of elk movements and showed the time dependent effects of habitat features within the habitat mosaic at Starkey.

Each approach reviewed above allows to deal with different and important aspects of animal movement data. For example, the random walk/diffusion framework provides a great deal about the spatial dynamics of populations. However, ecologists have an increasing interest in spatial processes at the individual level and in animal environment interactions, and none of these approaches has the ability to test how landscape features actually influence the movement process.

1.3 Biological and Ecological Background

African elephants are regarded as a high-impact megaherbivore species of the savanna. They are believed to have a significant effect on local habitat conditions because they can consume large amounts of woody vegetation (Ben-Shahar, 1998; Bowland and Yeaton, 1997). African elephants range widely and can exhibit multiple movement strategies within the same ecosystem (Wittemyer *et al.*, 2007).

Elephant movement behaviour is mainly driven by changes in water availability and vegetation, but generally several factors influence their spatial movement, such as rainfall, temperature and primary productivity.

African savannas are characterised by dry and wet seasons (Sankaran *et al.*, 2005). Winter (April-September) is known as ‘dry season’, and during this period the quality of food resources deteriorates, and seasonal water sources dry up. Therefore, elephants appear to concentrate their movements closer to water sources (in particular riparian zones), where they feed on a wide variety of trees (Young *et al.*, 2009b). Summer (January-March/October-December) is known as 'wet season', because rain falls during this period and temperatures are high, allowing the grass to flush so elephants feed more on grass during the post-rainfall period (Young *et al.*, 2009b) and they switch to browse when the grass becomes unpalatable. During the wet season, food resources are more abundant and of higher quality (Owen-Smith, 1988). Water is distributed widely during the wet season (summer) and may not therefore limit elephants (Leuthold, 1977; Western and Lindsay, 1984; De Beer *et al.*, 2006). Elephants need to drink regularly and their water requirements are central to understand patterns of their spatial use. For instance, in Kruger National Park elephants drink on average every two days during the dry season (Young *et al.*, 2009b). In drier environments, bull elephants probably drink every 3-5 days and breeding herds drink every 2-4 days (Leggett, 2006). Elephants, especially breeding herds, therefore, seldom roam far away from drinking water. However, elephants move greater distances during the dry season to obtain food (Harris *et al.*, 2008).

Certainly, the distribution and availability of food are related to rainfall. Food availability may be greater in wet than in dry savannas, where the annual precipitation is relatively low. In wet savannas a longer duration and greater volume of rainfall may render seasonal differences in food availability less pronounced than in dry savannas. These differences may influence elephant returns to previously utilized areas within seasons, between seasons and between years.

Temperature also does affect elephant movement, particularly higher temperatures in summer (Kinahan *et al.*, 2007). Elephants usually respond to high midday summer temperatures by moving into shady areas, resting under trees or cooling themselves in pools. They resume their search for food once temperatures drop a little. Elephants have distinct home ranges which may shift from summer to winter.

Understanding the relationship between environmental dynamics and movement is particularly important to wide-ranging species whose mobility can be critical for persistence of high temporal variability of local food resources (Fryxell *et al.*, 2008). Environmental drivers known to affect timing of migration may, therefore, provide a useful basis to understand seasonality movement of this species.

1.4 The thesis contribution

The main aim of this thesis is to propose a statistical framework to analyse animal movement data taking into account the influence of environmental variables on movements and the effect of the spatial area where animals live. More specifically, the sample of study is represented by African elephants (*Loxodonta Africana*) which live in the Kruger National Park and its associated private reserves in South Africa. The goal is to assess and to quantify environmental factors affecting changes in movements, which is crucial to evaluate how movement patterns could change due to variation in climate, and to estimate the overall trend of elephant movements in the study area.

Furthermore, since seasonality and long term spatial trends affect both water availability and vegetation phenology (Scholes *et al.*, 2003), it is clear that changes in elephant movements would be affected by rainfall and temperature patterns (Loarie *et al.*, 2009; Birkett *et al.*, 2012).

In studying environmental factors affecting elephant movements, at least three important features have to be taken into account in the data modelling process:

- the typically nonlinear relationships between movement and environmental variables;
- the possible delayed effects of environmental factors;
- heterogeneity among elephants which leads to random effects.

We propose an approach that attempts to meet these requirements. More specifically, we assume a *Generalized Additive Mixed Model* (GAMM) (Wood, 2006a; Pinheiro and Bates, 2000) using penalized splines (Eilers and Marx, 1996). GAMMs provide a suitable framework to model animal movement data because explanatory variables, including seasonality, spatio-temporal effects and environmental factors, can be fitted via parametric or nonparametric terms and random effects can be incorporated straightforwardly. Moreover, any residual dependence among observations can be modelled using proper correlation structures (Pinheiro and Bates, 2000; Wood, 2006a).

The present thesis is structured in the following way. Chapter 2 deals with a detailed description of data under study and shows some exploratory analyses to better understand features inside data, while Chapter 3 provides some aspects of penalized spline smoothing. In Chapter 4 first we deal with smoothing in presence of random effects, where we show that the mixed model representation of penalized splines results in smoothing parameter estimation. Then, we discuss the generalized additive mixed models (GAMMs) (Lin and Zhang, 1999; Wood, 2006a). In Chapter 5, we show the results of spatio-temporal elephant movement data modelling. In Chapter 6 we show some possible improvements of the proposed model discussed in Chapter 5, and finally in Chapter 7 we discuss the results obtained from our analyses.

Chapter 2

Data and some exploratory analyses

2.1 Study Area and Elephant Collaring

The study area is the Kruger National Park (KNP), situated in north-eastern South Africa, and its associated private reserves along the western boundary (Sabie Sand, Klaserie, Timbavati, Umbabat and Manyaleti). The overall KNP covers an area of $18,992\ km^2$, forming part of the "lowveld" savanna (approximately 300 m above sea level) in the North-East (Codron *et al.*, 2006). A map of South Africa, where it is possible to identify the KNP, is shown in Figure 2.1. KNP is characterised by very heterogeneous systems which experience different seasons. Climate and geological substrate varies throughout KNP, and the resulting vegetational differences allow elephants to use a variety of plant foods and to adjust their diets according to season and food availability. KNP lies within two climatic zones. The Southern and Central portions, that is, the Southern area of the centrally located Olifants River (Southern KNP), lie in the lowveld bushveld zone with an average annual rainfall of 500 – 700 *mm* (Venter *et al.*, 2003). In the Northern area of the Olifants River (Northern KNP), which is in the arid bushveld, the mean annual rainfall is 300 – 500 *mm* (Venter *et al.*, 2003).

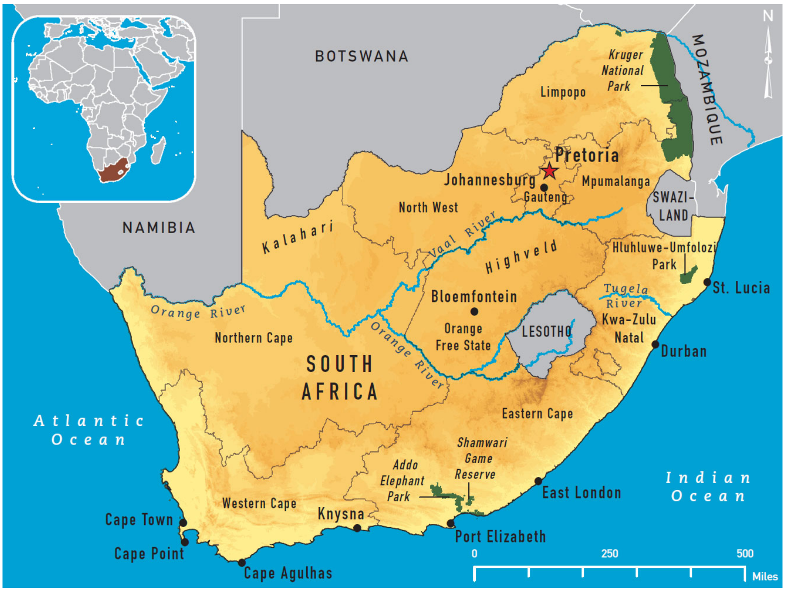

Figure 2.1: Kruger National Park situated in north-eastern South Africa. The Kruger is represented by the dark green area in the north-eastern boundaries of South Africa adjoining at East with the Monzambique.

Rainfall occurs in the austral summer between November and March (i.e., the wet season), with a peak in January and February (Venter *et al.*, 2003). Surface water is a complex phenomenon in KNP, because there are numerous water sources including rivers (seasonal, annual, perennial), boreholes, dams, wetlands, pans, distributed throughout Kruger. Kruger management has, over the years, shut down some of the man-made water-points because a year-round overabundance of water allows certain water-dependent species (such as elephant) to access many areas of Kruger that they perhaps would not use if the water was not present. The elephant population in KNP is estimated to be ∼14,000 individuals during 2010 (SANParks, unpublished data).

African elephant (*Loxodonta africana*) movements were tracked using global positioning system (GPS) collars and the capture occured 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).

The collaring operation was carried out by an experienced team which identified the elephant and shot a dart with a sedative at it. After about ten minutes, the elephant went down and the team fixed the radio collar and activated it, soon the veterinarians injected a drug to revive the elephant which, within three minutes, was fully up. The entire operation took about 50 minutes. During this operation, also a GPS-sensor, which measures the ambient temperature, was attached to the elephant. 14 female elephants, the *matriarchs*, from different herds, were collared in different areas of KNP (Orpen-Shukuza, Satara-Nhlanguleni-Muzanduzi, Lower Sabie, Satara and Shuzuka) from January 1, 2006 to January 17, 2010. The matriarch of a herd is the old female elephant which guides each member of its herd. The movements of these collared females are thus assumed to represent the movement behaviour of the breeding herd they belong to (Vanak *et al.*, 2010; Polansky and Wittemyer, 2011). A summary of the studied elephants by collarization areas is shown in Table 2.1. Hereafter the abbreviation

KNP will also include its private associated reserves, where elephants also move.

Table 2.1: *Collared* elephants by area in KNP.

Location	Elephant ID
Lower Sabie	AM105, AM108
Orpen-Skukuza	AM306, AM307,AM308
Satara	AM110,AM91,AM93,AM99
Satara-Nhlanguleni-Muzandzeni	AM239, AM253,AM254
Skukuza	AM106, AM107

2.2 Elephant Data

Data on elephant movements were collected by researchers of the University of Kwazulu-Natal, Durban, South Africa. They consist of daily movement observations of 14 female African elephants from different herds to ensure the indipendence of sampling. Data relative to three of 14 GPS-collars had erroneous points (AM106, AM108, and AM239), thus, in total 11 elephants were used for the analyses.

The object of interest is the daily movement of elephants, measured as mean daily speed in km/h (`speed`) obtained averaging the “hourly” speed (actually every 30 minutes) within a day for each elephant series, for a total of 11216 observations. The hourly speed was calculated by the following equation

$$speed_t = \frac{\sqrt{\Delta x_t^2 + \Delta y_t^2}}{\Delta t}/1000, \tag{2.1}$$

where $\Delta x_t = x_{t+1} - x_t$ is the distance between two consecutive hourly observations in x direction in meter, also known as *Northing*, and $\Delta y_t = y_{t+1} - y_t$ is the distance between two consecutive hourly observations in y direction in meter, also known as *Easting*. Δt is the time during $t + 1$ and t (30 minutes). The x, and y points were measured according to the *Universal Transverse Mercator* (UTM) projected coordinate system, which uses a 2-dimensional Cartesian coordinate system to give locations on the surface of the Earth in x-direction (XUTM) and y-direction (YUTM). Also the geographic hourly coordinates of elephant locations, longitude and latitude, were obtained from the XUTM and YUTM projections. Specifically, daily coordinates of elephant locations were obtained averaging the hourly longitude and latitude of elephant locations within a day for each elephant series. The distribution of elephants in Kruger is shown in Figure 2.2, where the small grey dots represent the mean daily elephant locations in the period under study, and the black points with associated labels represent the rainfall stations distributed in the Kruger as it will be discussed in the next Section. Figure 2.3 shows trajectories of the 11 considered elephants which move in KNP in the period under study, where the initial and final locations are indicated with a blue and red triangle, respectively. The initial and the final locations are very close to each other when they are observed in the same season, but if the initial and final locations are far from each other when they are not observed in the same season. It is interesting to notice that elephants remain in well-established home ranges which correspond with the collarization area. This is probably a result of a combination of factors. First of all it is due, most likely, to the fact that within their home ranges ele-

phants have adequate resources (forage and water), and they would waste energy to move further, especially in presence of young babies that cannot travel very far. Another motivation is that Kruger has a large number of elephants which live in herd guided from the dominant female matriarchs, and this can cause competition between more dominant female matriarchs and other subordinate ones (Wittemyer and Getz, 2007). However, this is not always the case, and herds may actually come together often to socialize and feed together forming close clans (Archie *et al.*, 2006).
Table 2.2 shows the number of observations for each elephant during each year under study.

Table 2.2: Number of observations for each elephant by year.

Elephant ID		Year				
		2006	2007	2008	2009	2010
1	AM105	231	363	364	300	0
2	AM107	281	365	360	249	0
3	AM110	276	365	358	318	0
4	AM253	0	173	344	0	0
5	AM254	0	224	363	318	0
6	AM306	0	0	297	349	12
7	AM307	0	0	300	340	0
8	AM308	0	0	300	360	16
9	AM91	357	359	359	299	0
10	AM93	283	356	359	313	0
11	AM99	280	365	358	306	0

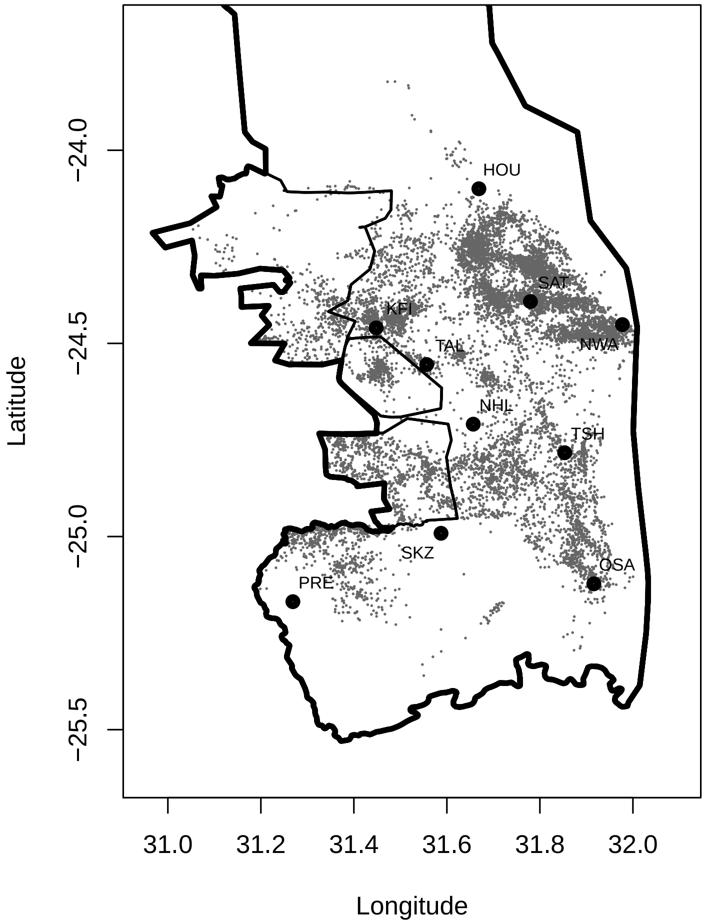

Figure 2.2: Elephant distribution in the KNP from January 1, 2006 to January 17, 2010. Each small grey dot represents the daily average elephant location and each black dot represents the rainfall station distributed in the Kruger.

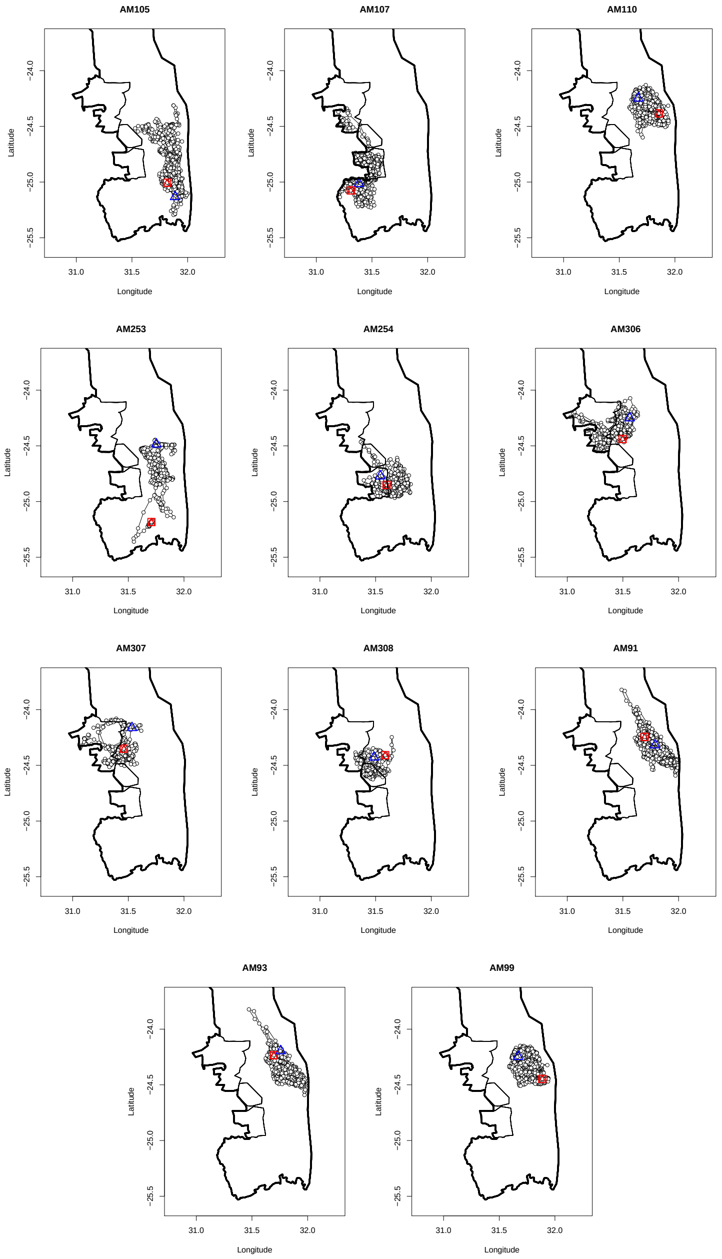

Figure 2.3: Trajectories of the 11 studied elephants in the KNP from January 1, 2006 to January 17, 2010. The initial and final locations are indicated with a blue and a red triangle, respectively.

It is important to notice that observations are not balanced and for 2010 (January, 17^{th}) there are a lot of missing values, and only for two elephants (AM306 and AM307) there are some observations in this year (2010). Missing data are ruled out, it means that no type of imputation is done to data.

The observations of the response, mean daily speed, are within (0.007,1.845), and the average of mean daily speed is about 0.416 km/h.

Figure 2.4 shows the density plots of the mean daily speed by each elephant: we observe quite similar values in the range $0.007 - 1.845$ km/h, with elephants AM107, AM307 and AM254 having lower variability.

Individual elephant movement profiles versus the days from 2006 to 2010, are shown in Figure 2.5. Here the patterns indicate that a strong seasonality effect is present in the data. To emphasise the seasonal patterns of individual elephant movements, Figure 2.6 represents the monthly averages of movements for each elephant. While the seasonal pattern is substantially the same, we observe some heterogeneity in the ‘intercepts‘, i.e. in the general mean level.

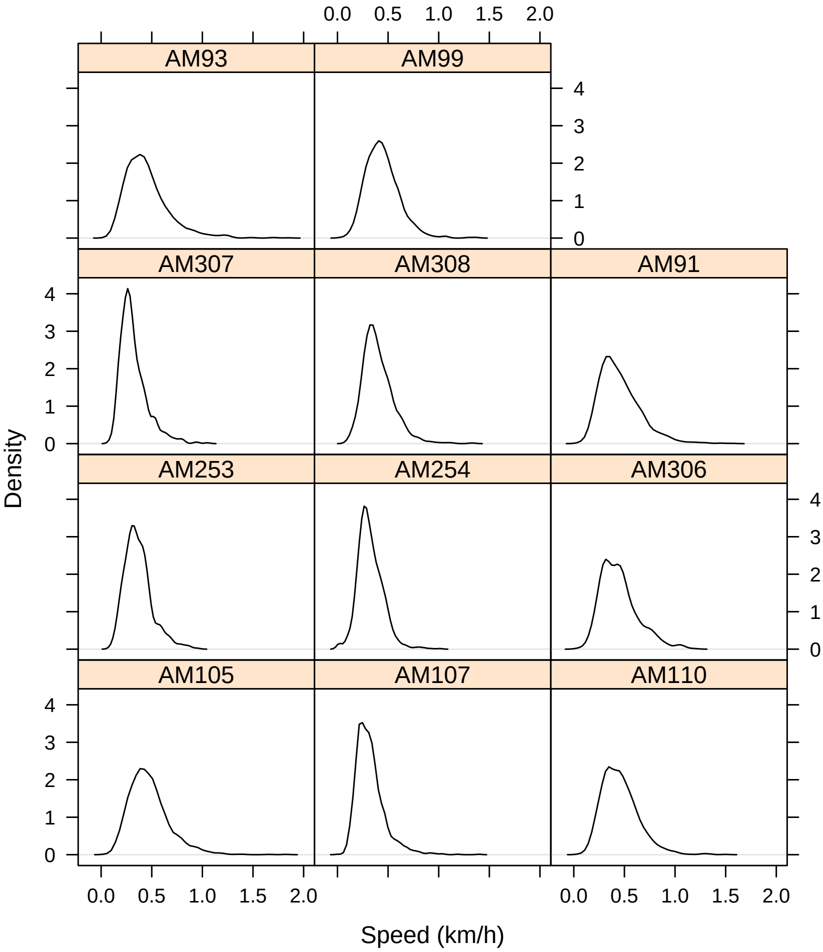

Figure 2.4: Kernel density estimates of speed for each elephant.

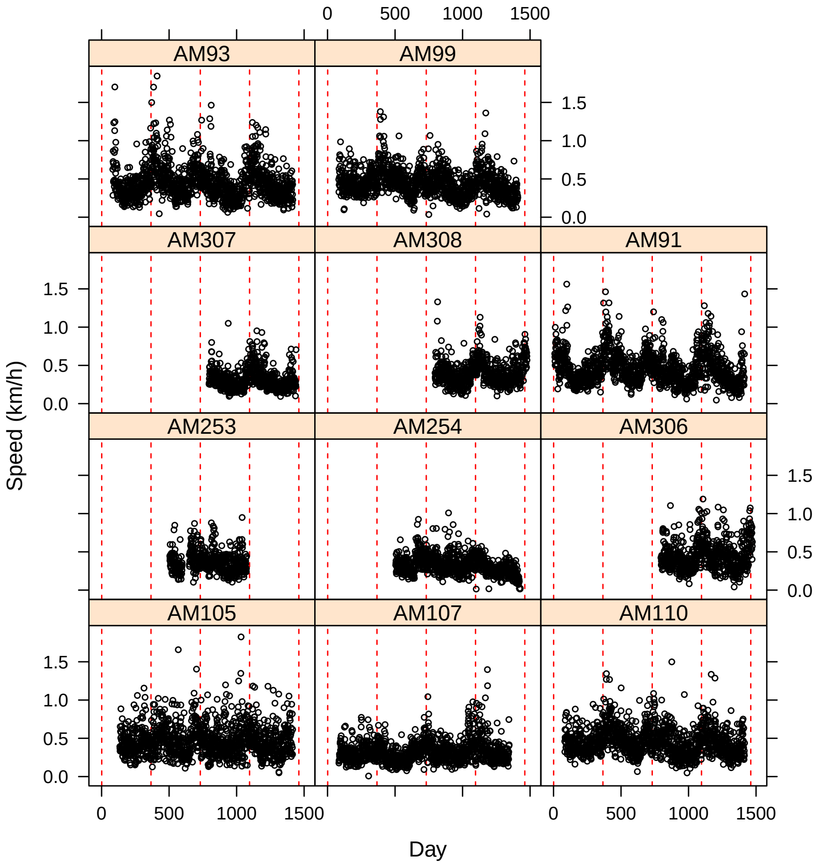

Figure 2.5: Daily time series of individual elephant movement from January 1, 2006 to January 17, 2010. The dashed red lines represent the years from 2006 to 2010.

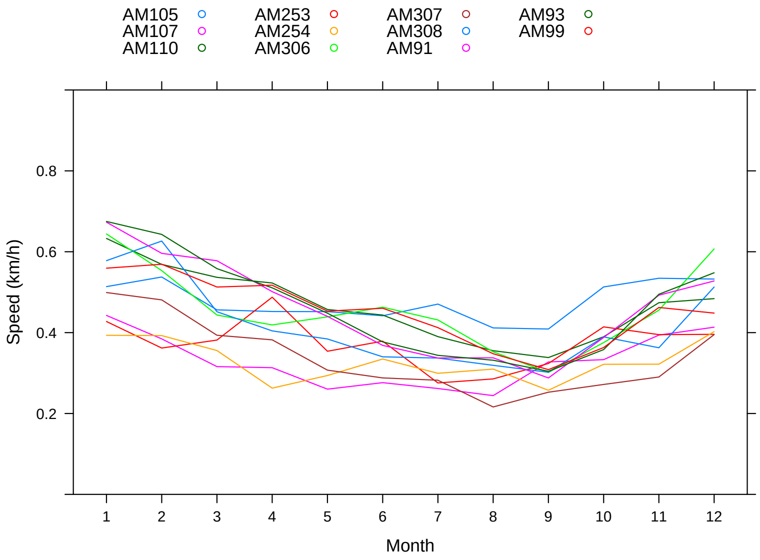

Figure 2.6: Individual elephant movement profiles versus month. Each elephant is represented by a different color.

2.3 Environmental data

As discussed in Chapter 1, the main goal of the thesis is to investigate the effects of some environmental variables on elephant movements. Environmental variables employed in this study are rainfall and temperature.

Rainfall values were averaged between two and three stations (Fig. 2.2, Tab. 2.3) located within the spatial range of individual elephants. These values represent ‘local’ rainfall for each elephant over the considered time. The rainfall stations and their used abbreviations are summarized in Table 2.3.

Table 2.3: Names of rainfall stations distributed in KNP.

	Station ID	Abbreviation
1	HOUTBOSCHRAND	HOU
2	KINGFISHERSPRUIT	KFI
3	LOWER-SABIE	OSA
4	NHLANGULENI	NHL
5	NWANETSI	NWA
6	PRETORIUSKOP	PRE
7	SATARA	SAT
8	SKUKUZA	SKZ
9	TALAMATI	TAL
10	TSHOKWANE	TSH

A measure of temperature was obtained from GPS-sensors attached to elephants which measured ambient temperature hourly data. The used average daily temperature for the analysis purposes is obatined as mean of the hourly temperature data, and hereafter we will refer to it as ‘temperature’. Local rainfall and temperature time series for each elephant, in the period under study, are shown respectively in Figure 2.7 and 2.8.

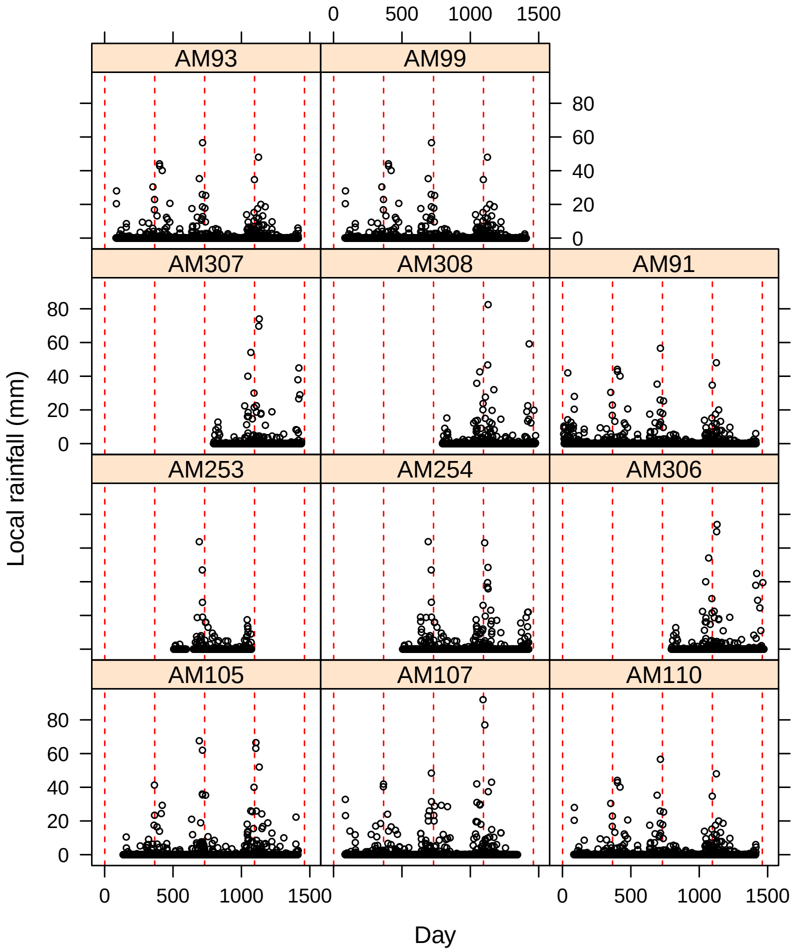

Figure 2.7: Observed local rainfall time series for each elephant across years. The dashed red lines represent the years from 2006 to 2010.

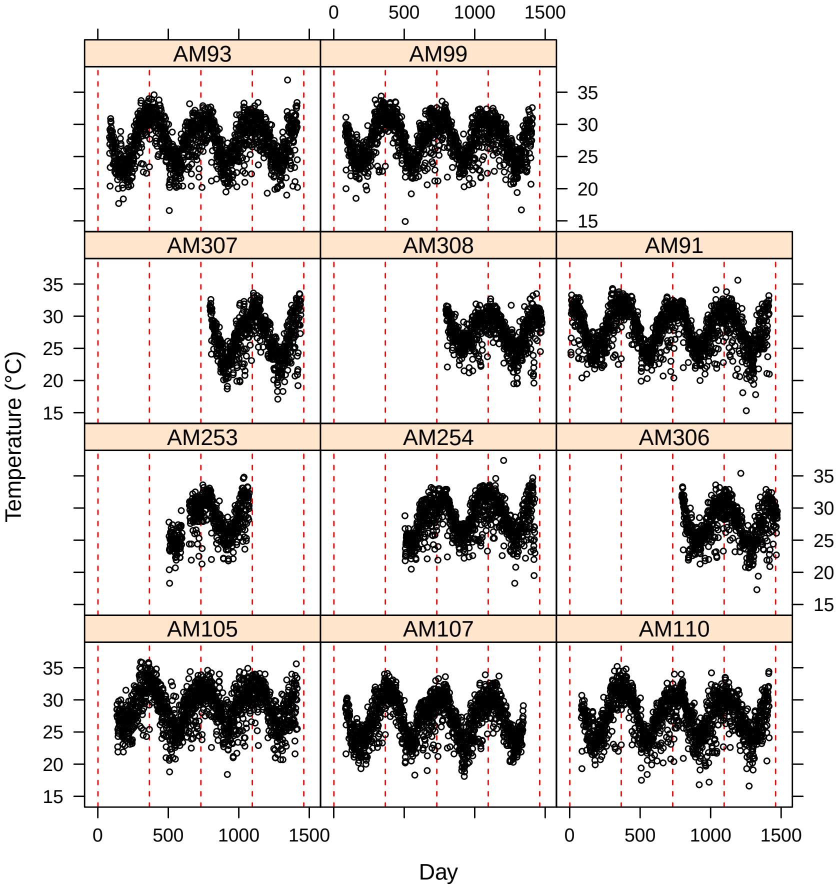

Figure 2.8: Observed temperature time series for each elephant across years. The dashed red lines represent the years from 2006 to 2010.

Table 2.4: List of explanatory variables and their used abbreviations.

Variable	Abbreviation
Date	`Date`
Elephant identification	`ID`
Years from 2006 to 2010	year
Average temperature (°C)	`avgTemp`
Local rainfall (mm)	`locRaif`
Day of year from 1 to 365	`day.year`
Average daily elephant location (longitude and latitude)	`long`, `lat`

A list with all available explanatory variables and their used abbreviations is given in Table 2.4.

Expolorative analysis was carried out also to investigate relationship between environmental variables, rainfall and temperature, and mean daily speed. Figures 2.9 and 2.10 show scatterplot of mean daily speed with rainfall and temperature, respectively, for each elephant under study. Looking at these two Figures we see that most likely the relationship with mean daily speed is not linear for both local rainfall (Fig. 2.9) and temperature (Fig. 2.10). Also a scatter plot between rainfall and temperature has been reported to investigate relationship between them. Plot is shown in Figure 2.11, where it is simple to see that no collinearity is present between local rainfall and temperature.

Exploratory analysis showed that there are some relevant features that need to be taken into account in the modelling process.

These are a strong seasonality effect, heterogeneity among elephants, and non-linear relationship between daily speed and environmental variables.

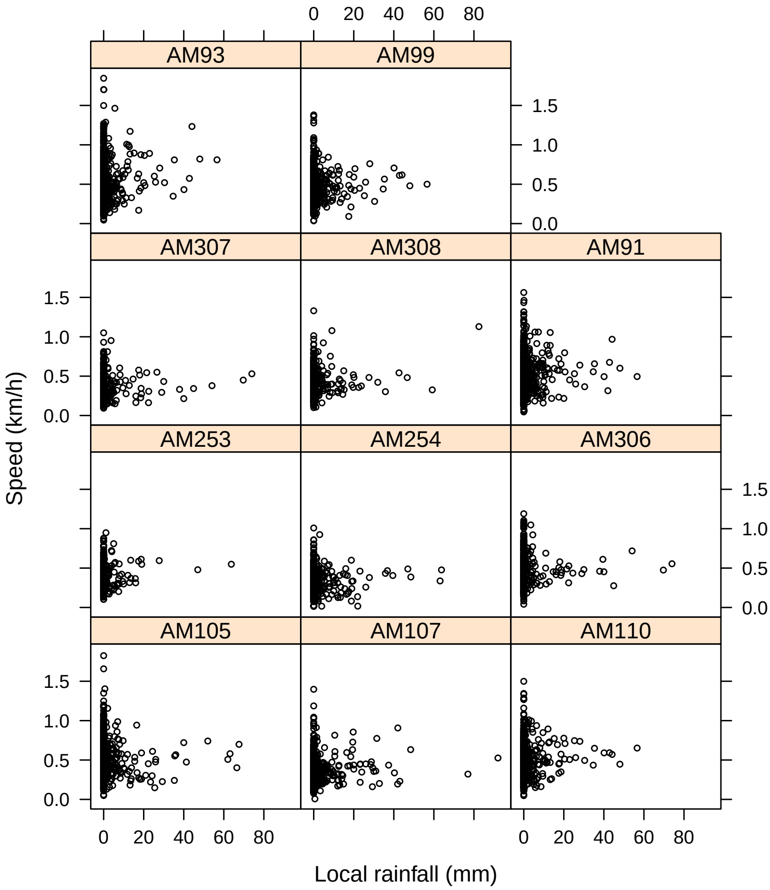

Figure 2.9: Scatterplot of mean daily speed and local rainfall for each elephant, from January 1, 2006 to January 17, 2010.

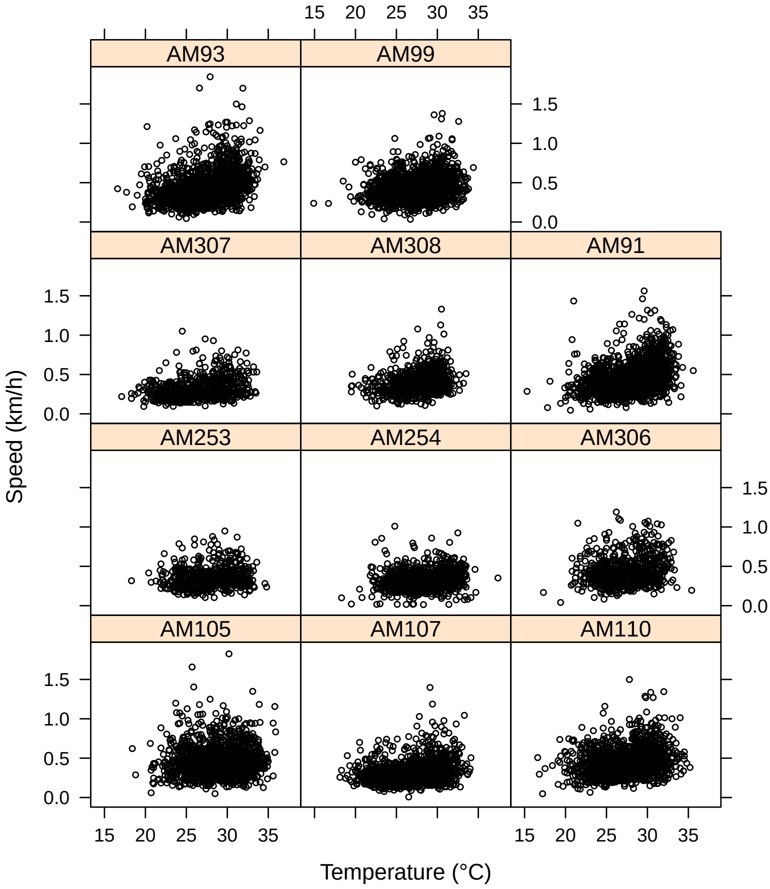

Figure 2.10: Scatterplot of mean daily speed and temperature for each elephant, from January 1, 2006 to January 17, 2010.

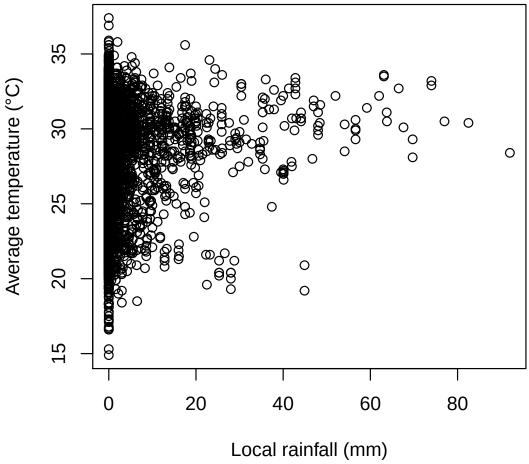

Figure 2.11: Scatter plot of local rainfall and temperature.

Chapter 3

Penalized Spline Smoothing

Smoothing methodology offers an extremely useful tool to handle non-linear relationships without the restrictions of parametric functional forms. It has become a widely used framework for data analysis and inference. Its integration into complex models and its use in applications are also becoming more and more pervasive. Smoothing is the basic "concept" underlying the GAMM employed in this thesis to modelling. We will give some concepts on smoothing and discuss GAMM in the next Chapter.

3.1 Idea of penalized smoothing

When interest lies in modelling how the covariate x affects the means of Y, it is usually assumed

$$y_i = f(x_i) + \epsilon_i \tag{3.1}$$

where y_i is a response variable, x_i a covariate, and f a smooth function of x_i and ϵ_i are i.i.d. $N(0, \sigma^2)$ random variables.

The idea is to build f out of a sum of J known functions B_js, scaled by coefficients β_j, namely

$$f(x) = \sum_{j=1}^{J} B_j(x_i)\beta_j \tag{3.2}$$

and then estimate these coefficients accordingly. Since the equation in (3.1) is in the form of a linear model, the problem is now to minimize the ordinary least squares (OLS) objective function,

$$||\boldsymbol{y} - \boldsymbol{X\beta}||^2 \tag{3.3}$$

where $\boldsymbol{X} = [B_1, \dots, B_J]$. Estimation of β is carried out by imposing a “wiggliness” penalty to prevent under smoothing,

$$||\boldsymbol{y} - \boldsymbol{X\beta}||^2 + \lambda \int_{-\infty}^{\infty} [f''(x)]^2 dx \tag{3.4}$$

where the integrated square of second derivative penalizes models that are too “wiggly”, and λ is the *smoothing parameter* which controls the trade-off between model fit and smoothness, leading to a too wiggly fitted curve when $\lambda = 0$.

The expression for the penalty given in (3.4) looks like it might require a rather large amount of integration and as such it would require a long time to compute, however it can be shown (see Wood (2006a), p. 126) that the integral of the penalty can always be written as a quadratic form in $\boldsymbol{\beta}$, since f is linear in the parameters β_j:

$$\int_{-\infty}^{\infty} [f''(x)]^2 dx = \boldsymbol{\beta}^T \boldsymbol{S\beta}, \tag{3.5}$$

where $\boldsymbol{S}$ is the penalty matrix of known coefficients. The penalized least squares estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X^T X} + \lambda \boldsymbol{S})^{-1} \boldsymbol{X^T y} \tag{3.6}$$

and the hat matrix, $\boldsymbol{H}$, for the model can be written as

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X^T}\boldsymbol{X} + \lambda \boldsymbol{S})^{-1}\boldsymbol{X^T}, \tag{3.7}$$

where $\hat{\boldsymbol{\mu}} = \boldsymbol{Hy}$. Generalizing also the definition of degrees of freedom from linear models as the trace of the hat matrix, we can define with df= $tr(\boldsymbol{H})$ the degrees of the smoother, corresponding to the smoothing parameter λ. It means the model complexity can be interpreted as the equivalent number of fitted parameters.

Giving a model with two explanatory variables x_1 and x_2,

$$y_i = f_1(x_{1i}) + f_2(x_{2i}) + \epsilon_i, \tag{3.8}$$

the model contains more than one function, $f_1(x_1)$ and $f_2(x_2)$, this introduces an identifiability problem: f_1 and f_2 are both only estimable within an additive constant. To see this, it is to notice that any constant could be simultaneously added to f_1 and subtracted from f_2, without changing the model predictions. Hence, identifiability constraints have to be imposed on the model in (3.8). Provided the identifiability issue, the additive model can be represented using penalized regression splines, estimated by penalized least squares, in the same way as the simple univariate model, and the degree of smoothing can be estimated by cross validation.

Each smooth function in (3.8) can be represented using a penalized regression spline basis as follows

$$f_1(x_{1i}) = \sum_{j_1=1}^{J_1} B_{1j_1}(x_{1i})\beta_{1j_1} \tag{3.9}$$

$$f_2(x_{2i}) = \sum_{j_2=1}^{J_2} B_{2j_2}(x_{2i})\beta_{2j_2}. \tag{3.10}$$

The identifiability problem with the additive model means that β_1 and β_2 are confounded. The simplest way to deal with this is to constrain one of them to zero, say $\beta_1 = 0$. Having done this, it is easy to see that the additive model can be written in the linear model form

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}, \tag{3.11}$$

where $\boldsymbol{X} = [B_1, \dots, B_{J_1}|B_1, \dots, B_{J_2}]$. As in the unidimensional case, the wiggliness of the functions can also be defined as

$$\int_{-\infty}^{\infty} [f_1''(x_1)]^2 dx_1 = \boldsymbol{\beta}_1^T \boldsymbol{S}_1 \boldsymbol{\beta}_1, \tag{3.12}$$

$$\int_{-\infty}^{\infty} [f_2''(x_2)]^2 dx_2 = \boldsymbol{\beta}_2^T \boldsymbol{S}_2 \boldsymbol{\beta}_2, \tag{3.13}$$

where $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are the penalty matrix, and assuming that $\boldsymbol{S} \equiv \lambda_1 \boldsymbol{S}_1 + \lambda_2 \boldsymbol{S}_2$, the parameters $\boldsymbol{\beta}$ of the model (3.11), where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)^T$, are obtained by minimizing the penalized least squares objective

$$||\boldsymbol{y} - \boldsymbol{X\beta}||^2 + \lambda \boldsymbol{\beta}^T \boldsymbol{S\beta}. \tag{3.14}$$

As discussed above, f can be decomposed into a series of basis functions. Among the several options, four important bases are: truncated power basis functions, thin plate regression splines, cubic splines, and cyclic cubic splines.

Since, in the next Chapter, we will use the *truncated power basis functions* (TPF) in order to explain the relationship between smooth terms and random effects, here we introduce this type of basis spline.

If we define K knots $k_1, \dots, k_K$, it is possible to define a TPF basis in the form

$$[1, x, \dots, (x - k_1)_+, \dots, (x - k_K)_+]$$

where $(x - k_1)_+ = (x - k_1)I(x > k_1)$.

Thus $f(\cdot)$ can be expressed as

$$f(x) = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} b_k (x_i - x_k)_+. \tag{3.15}$$

Now, we will give an introduction of *thin plate splines* (TPS) proposed by Duchon (1977), which offers a very elegant approach to estimate a smooth function of multiple predictor variables. Supposing that the problem of estimating the smooth function $f(\boldsymbol{x})$, from n observations $(y_i, \boldsymbol{x}_i)$ is such that

$$y_i = f(\boldsymbol{x}_i) + \epsilon_i, \tag{3.16}$$

where f is an unknown function on a fixed domain $D \subset \Re_d$, ϵ_i is a random error term, and $\boldsymbol{x}$ is a d-vector $(d \leq n)$. Thin-plate spline smoothing estimates f by finding the function g which minimizes the penalized sum of squares

$$||\boldsymbol{y} - \boldsymbol{g}||^2 + \lambda J_{m,d}(g), \tag{3.17}$$

where $\boldsymbol{y}$ is the vector of y_i data, and $\boldsymbol{g} = (\boldsymbol{g}(x_1), \boldsymbol{g}(x_2), \ldots, \boldsymbol{g}(x_n))'$, $J_{m,d}(g)$ is the penalty function measuring the wiggliness of g, and λ is the smoothing parameter, which controls the trade-off between data fitting and smoothness of g. Duchon (1977) in his work introduced the following penalty:

$$J_{m,d} = \int \cdots \int_{\Re^d} \sum_{\nu_1 + \cdots + \nu_d = m} \frac{m!}{\nu_1! \ldots \nu_d!} \left(\frac{\partial^m g}{\partial x_1^{\nu_1} \ldots \partial x_d^{\nu_d}} \right)^2 dx_1 \ldots dx_d \tag{3.18}$$

where m is the derivative order which can be any integer satisfying $2m > d$, d is the the number of covariates (in a spatial setting $d = 2$ for longitude and latitude coordinate data) and the $\nu_1, \ldots, \nu_d$ terms simply ensure that derivatives are taken with respect to all the parameters in all of the necessary combinations. Duchon (1977) showed that the minimizing function g in (3.17) has the following form:

$$g(\boldsymbol{x}) = \sum_{i=1}^{n} \delta_i \eta_{m,d}(\|\boldsymbol{x} - \boldsymbol{x}_i\|) + \sum_{j=1}^{M} \alpha_j \phi_j(\boldsymbol{x}) \tag{3.19}$$

where $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}$ are unknown parameters vectors to be estimated, subjected to the linear constraints that $\boldsymbol{T}^T\boldsymbol{\delta} = \boldsymbol{0}$, where $T_{ij} = \phi_j(\boldsymbol{x}_i)$. $M = \binom{m+d-1}{d}$ are ϕ_j functions, which are linearly independent polynomials of degree less than m which span the space of polynomials in $\Re^d$. All of the ϕ_js are unpenalized as they lie in the nullspace of the penalty. It is also important to notice that, to maintain continuity in f, $2m > d$; this means that the dimension of the nullspace increases rapidly with d. As in (3.2), f is decomposed into a sum of basis functions, however, for a thin plate spline this summation is split into two parts: M polynomials that act over the whole of the data (the ϕ_js) and a set of *radial basis functions*, one centred at each datum (the $\eta_{m,d}$). One can think of this as a global trend (in the 2-dimensional case, linear functions of the two coordinates) with extra flexibility provided by the radial basis functions. The remaining basis functions $\eta_{m,d}(r)$ in (3.19) are defined as:

$$\eta_{m,d}(r) = \begin{cases} \frac{(-1)^{m+1+d/2}}{2^{2m-1}\pi^{d/2}(m-1)!(m-d/2)!} r^{2m-d} \log(r) & d \text{ even} \\ \frac{\Gamma(d/2-m)}{2^{2m}\pi^{d/2}(m-1)!} r^{2m-d} & d \text{ odd} \end{cases} \tag{3.20}$$

Defining now a matrix $\boldsymbol{E}$, the thin plate spline fitting problem becomes

$$\text{minimize } \|\boldsymbol{y} - \boldsymbol{E\delta} - \boldsymbol{T\alpha}\|^2 + \lambda\boldsymbol{\delta}^T\boldsymbol{E\delta} \tag{3.21}$$

with respect to $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}$, and subject to $\boldsymbol{T}^T\boldsymbol{\delta} = \boldsymbol{0}$.

These TPS introduced by Duchon (1977) suffer from some limits. In TPS it is necessary to choose knot locations, in order to use each basis, introducing an extra degree of subjectivity into the model fits; furthermore the bases are only useful for representing smooths of one predictor variable and it is not clear to what extent the bases are better or worse than any other basis that might be used.

Wood (2003) proposed the use of iterative weighted fitting of reduced rank thin-plate splines for computational efficiency, named as *thin plate regression splines* (TPRS), which are knot free bases, and can be used for smooths of any number of predictors.

The idea is to truncate the space of the wiggly components of the thin plate spline (the components with parameters $\boldsymbol{\delta}$), while leaving the components of 'zero wiggliness' unchanged (the $\boldsymbol{\alpha}$ components). One way to reduce the size of $\boldsymbol{E}$ is by performing an eigen-decomposition, $\boldsymbol{E} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T$, where $\boldsymbol{D}$ is a diagonal matrix of eigenvalues decreasing in absolute value ($|D_{i,i}| \geq |D_{i-1,i-1}|$), and the columns of $\boldsymbol{U}$ are the corresponding orthogonal eigenvectors.

Now we define $\boldsymbol{U}_k$ and $\boldsymbol{D}_k$, where the former denotes the matrix consisting of the first k columns of $\boldsymbol{U}$ and the latter denotes the top right $k \times k$ submatrix of $\boldsymbol{D}$, and restrict $\boldsymbol{\delta}$ to the columns space of $\boldsymbol{U}_k$, by writing $\boldsymbol{\delta} = \boldsymbol{U}_k\boldsymbol{\delta}_k$. In this way we have $E_k = \boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{U}_k^T$, and (3.21) becomes

$$\text{minimize } \|\boldsymbol{y} - \boldsymbol{U}_k\boldsymbol{D}_k\boldsymbol{\delta}_k - \boldsymbol{T}\boldsymbol{\alpha}\|^2 + \lambda\boldsymbol{\delta}_k^T\boldsymbol{D}_k\boldsymbol{\delta}_k \tag{3.22}$$

with respect to $\boldsymbol{\delta}_k$ and $\boldsymbol{\alpha}$, where $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}$ are vectors of coefficients to be estimated, and δ is subject to $\boldsymbol{T}^T\boldsymbol{U}_k\delta_k = \boldsymbol{0}$. It can be shown that the reduced rank matrix $\boldsymbol{E}_k$ gives the best approximation to $\boldsymbol{E}$ (see Wood (2003) for details). In practice, k is set to be large enough and a further reduction in basis complexity is performed by penalization. So k indicates the "maximum basis size". TPRSs are particularly useful because they have the important property of the *isotropy* of the wiggliness penalty: wiggliness in all directions is treated equally, with the fitted spline entirely invariant to rotation of the co-ordinate system for the predictor variables, hence all directions have a common smoothing parameter so wigglyness in the x_1 direction has the same weight in the penalty as in the x_2 direction and so on through higher dimensions. This property is usually appropriate in a spatial setting, since there is nothing special about one geographical coordinate over another one when it comes to the smoothness of the function to be fitted. When two or more predictors, which are both arguments of the same smooth, are measured in a different scale, the TPRS is not appropriate to be use. In these situations a more satisfactory approach is to use tensor product smooths, that will be discussed later, in this Section.

Another low rank and efficient spline basis is the *Cubic spline*. CSs are univariate bases, which require the specification of knots. CS are made up of sections of cubic polynomials which are continuous (up to second derivatives) at the join points. The CS is the function which minimizes the objective function in (3.14).

A *cubic regression spline* basis (CRS) has many possible parametrizations. Here we present the parametrisation, which parameterizes the spline in terms of its values at the knots.

Considering a cubic spline function, $f(x)$, with k knots, $x_1, \ldots, x_k$, the conditions are that the spline has to be continuous to second derivative, at the x_j, and should have zero second derivative at x_1, and x_k. Letting $\beta_j = f(x_j)$ and $\delta_j = f''(x_j)$, the parametrization gives the following form for f:

$$
\begin{aligned}
f(x) = {} & \frac{x_{j+1} - x}{x_{j+1} - x_j}\beta_j + \frac{x - x_j}{x_{j+1} - x_j}\beta_{j+1} \\
& + \left\{ \frac{(x_{j+1} - x)^3}{x_{j+1} - x_j} - (x_{j+1} - x_j)(x_{j+1} - x) \right\} \frac{\delta_j}{6} \\
& + \left\{ \frac{(x - x_j)^3}{x_{j+1} - x_j} - (x_{j+1} - x_j)(x - x_j) \right\} \frac{\delta_{j+1}}{6} \text{ if } x_j \leq x \leq x_{j+1}.
\end{aligned} \tag{3.23}
$$

This setup leads to the spline having directly interpretable parameters, and this basis does not require any re-scaling of the predictor variables before it can be used to construct a GAM.

It is often appropriate for a model smooth function to be 'cyclic', for example when we have a smooth function of days of year that do not change discontinuously at the end of the year. In this case, the function would have the same value and first few derivatives at its upper and lower boundaries. The penalized cubic regression spline can be modified to produce such a smooth, by imposing the constraint that the spline must be continuous to second derivative at each knot, and that $\hat{f}(x_1) = \hat{f}(x_k)$ up to second derivative. This specifies that the spline must "join up" at each end. The form of f is the same as in (3.23), but there is one less coefficient to estimate, since the first and last ones are the same.

Frequently, there is the need to consider smooths of any number of predictors, and usually these predictors are expressed in different scale, hence, it is necessary to scale all predictors into the unit square, this is done thanks to tensor product smooths. This is made possible by thinking of each 1-dimensional basis as a marginal smooth, and then these marginal smooths are combined in a higher dimensional smooth of several variables by a tensor product construction. Wood (2006b) proposed a general method to use low rank tensor product smooths to represent smooth functions of several variables in GAMs and GAMMs, in which the smooth terms are represented using any relatively low rank basis, with an associated quadratic penalty, which measures the wiggliness of the smooth, and estimation is via penalized likelihood maximization. He shows how to form smooths of several variables from tensor products of any set of bases with quadratic penalties in a way that allows the smooth to be decomposed into fixed and random components suitable for the incorporation into a generalized linear mixed model, which produces smooths that are invariant to rescaling of their arguments and which produces smooths that are computationally efficient to work with, due to their relatively low rank. The tensor product smooth is here introduced starting from the construction of a smooth function of 3 covariates, x, z and v, the generalization is then trivial. The process starts by assuming that we have low rank bases available, for representing smooth functions f_x, f_z and f_v of each of the covariates. The basis functions for the marginal smooths of x, z and v are

$$f_x(x) = \sum_{i=1}^{I} \alpha_i A_i(x), \tag{3.24}$$

$$f_z(z) = \sum_{l=1}^{L} \delta_l D_l(z), \tag{3.25}$$

and

$$f_v(v) = \sum_{k=1}^{K} \beta_k B_k(v), \tag{3.26}$$

where α_i, δ_l and β_k are unknown coefficients, $A_i(x)$, $D_l(z)$ and $B_k(v)$ are known basis functions of the covariates x, z and v respectively, and they might be B-splines, thin plate regression splines, cubic regression splines, etc... To convert f_x into a smooth function of x and z, it is required for f_x to vary with z, and it is possible by allowing its parameters, α_i, to vary smoothly with z. The simplest way to do this would be defining:

$$\alpha_i(z) = \sum_{l=1}^{L} \delta_{il} D_l(z), \tag{3.27}$$

then $f_{x,z}$ would be defined as:

$$f_{x,z}(x,z) = \sum_{i=1}^{I} \sum_{l=1}^{L} \delta_{il} D_l(z) A_i(x). \tag{3.28}$$

Continuing in the same way, we could now create a smooth function of x, z and v by allowing $f_{x,z}$ to vary smoothly with v, and again, letting the parameters of $f_{x,z}$ vary smoothly with v, we get the following

$$f_{x,z,v}(x,z,v) = \sum_{i=1}^{I} \sum_{l=1}^{L} \sum_{k=1}^{K} \beta_{ilk} B_k(v) D_l(z) A_i(x). \tag{3.29}$$

For the set of observations of x, z and v, there is a simple relationship between the model matrix X, evaluating the tensor product smooth at these observations, and the model matrices $\boldsymbol{X}_x$, $\boldsymbol{X}_z$ and $\boldsymbol{X}_v$, that would evaluate the marginal smooths at the same observations. It is simple to show, given appropriate ordering of the β_{ilk} into a vector of β, that the i^{th} row of $\boldsymbol{X}$ is

$$\boldsymbol{X}_i = \boldsymbol{X}_{xi} \otimes \boldsymbol{X}_{zi} \otimes \boldsymbol{X}_{vi}, \tag{3.30}$$

where ⊗ is the usual Kronecker product. Having derived a tensor product basis in (5.6) to represent smooth function of (x, z, v) component, it is also necessary to determine a way to measure the wiggliness. To do this, it is possible to start from wiggliness measures associated with the marginal smooth functions. Let J_x, J_z and J_v be measures of the wiggliness of the functions f_x, f_z and f_v, respectively, given by

$$J_x(f_x) = \boldsymbol{\alpha}^T \boldsymbol{S}_x \boldsymbol{\alpha}, \quad J_z(f_z) = \boldsymbol{\delta}^T \boldsymbol{S}_z \boldsymbol{\delta}, \quad J_v(f_v) = \boldsymbol{\beta}^T \boldsymbol{S}_v \boldsymbol{\beta}, \tag{3.31}$$

where S_x, S_z and S_v are the matrices containing the known coefficients of the marginal smooth $\boldsymbol{\alpha}$, $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$. Given that $f_{x,z,v}(x, z, v)$ can be expressed also as $f_{x|zv}(x)$, i.e. as a function of x only, held z and v constant, and we similarly define $f_{z|xv}(z)$ and $f_{v|xz}(v)$, the penalty for such a smooth can be written as the sum of the three penalties weighted by the corresponding smoothness parameters λ_x, λ_z and λ_v, which control the tradeoff between wiggliness in different directions, and allowing the penalty to be invariant to the relative scaling of the covariates. Hence the wiggliness of $f_{x,z,v}(x, z, v)$, can be written as follows:

$$J(f_{xzv}) = \lambda_x \int_{z,v} J_x(f_{x|zv}(x))dzdv + \lambda_z \int_{x,v} J_z(f_{z|xv}(x))dxdv + \lambda_v \int_{x,z} J_v(f_{v|xz}(x))dxdz. \tag{3.32}$$

For further details see Wood (2006a). Although only a three-dimensional example is given here, tensor product splines provide an extremely useful tool, allowing for extra dimensions to be added to models using different bases. The use of a different smoothing parameter for each direction allows for anisotropic smoothing, so that covariates that are measured on different scales (for example temperature and rainfall) may be combined into one tensor product smooth, avoiding the assumption that the degree of smoothing required is the same in both directions. In particular this can be useful when constructing a spatio-temporal smooth: for example using a thin plate spline for the spatial part of the smooth (so the spatial part of the model is isotropic) then taking a tensor product of that with a cubic spline basis for the temporal part (so a different amount of smoothing can be used for each direction). This is the setup that will be used in Chapter 5 for the elephant movement data.

We have discussed the penalized likelihood maximization of $\boldsymbol{\beta}$ given $\boldsymbol{\lambda}$, and we have presented some basis functions. Now we discuss how to estimate $\boldsymbol{\lambda}$. Figure 3.1 shows how different values of $\boldsymbol{\lambda}$ affect the fitted smooth function. Hence, changing the smoothness parameter a variety of models of fitted regression functions of different smoothness can be obtained, but the question is how to select the optimal $\boldsymbol{\lambda}$; some approaches, based on the empirical measure of the mean square error (MSE), can be adopted. In particular, when scale parameter is known, the Mallow's C_p or UnBiased Risk Estimator (UBRE) (Craven and Wahba, 1978), is used, instead when the scale parameter is unknown, generalized cross validation (GCV) or its related criteria, such as Aikake Information Criteria (AIC) or generalized AIC, can be used.

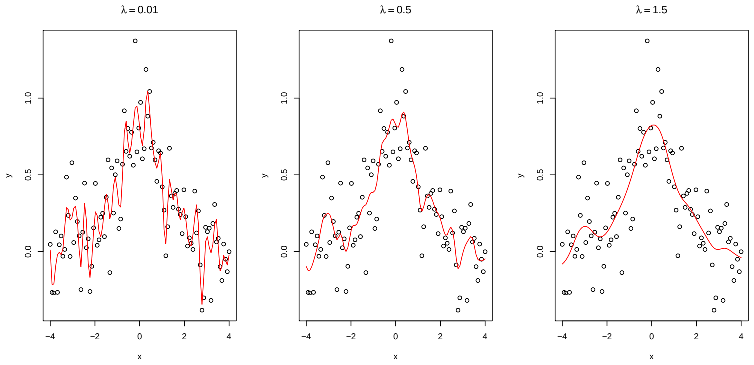

Figure 3.1: Penalized regression spline fits to simulated data using three different values for the smoothing parameter.

A simple and effective way to find $\hat{\boldsymbol{\lambda}}$ is to assess how well the model performs on data which were not in the sample, assessing the prediction error of the model using the *generalized cross validation* (GCV) score

$$V_g(\boldsymbol{\lambda}) = \frac{n\|\boldsymbol{y} - \hat{\boldsymbol{\mu}}_\lambda\|^2}{\{n - tr(\boldsymbol{H}_\lambda)\}^2}, \tag{3.33}$$

where tr($\boldsymbol{H}$) indicates the trace of $\boldsymbol{H}$, the hat (or influence) matrix for the smoother seen in (3.7), $\hat{\boldsymbol{\mu}}$ is the vector of fitted values for the model. Numerical minimization of V_g with respect to $\boldsymbol{\lambda}$ (which enters V_g via $\boldsymbol{H}$) gives the optimal smoothing parameter ($\hat{\boldsymbol{\lambda}}$); further details are given in Wood (2006a) (p. 134-137). The hat matrix $\boldsymbol{H}$ is the matrix such that $\hat{\boldsymbol{\mu}} = \boldsymbol{H}\boldsymbol{y}$. This matrix has the useful property which consists in the fact that its trace (tr($\boldsymbol{H}$)) gives the effective degrees of freedom (edf) of the model. The edf gives a measure of the complexity of the fitted model. The higher the edf is, the more complex the model is. Clearly, if the smoothing parameters are all set to zero then the degrees of freedom of the model are simply the length of $\boldsymbol{\beta}$ (minus the number of identifiability constraints) which is the case of the linear model. It has been shown that the maximum of tr($\boldsymbol{H}$) is just the number of parameters less the number of constraints, and similarly that the minimum value is the rank of $\boldsymbol{S}$ less than this. As the smoothing parameters vary, from zero to infinity, the effective degrees of freedom move smoothly between these limits.

In the next Chapter we will discuss details of how smoothing parameter selection and estimation of $\hat{\boldsymbol{\beta}}$ are combined into a fitting procedure (Section 4.3).

Chapter 4

Generalized Additive (Mixed) Models

Generalized additive mixed models (GAMM) (Laird and Ware, 1982) combine the flexible modelling of the relationship between a response and predictors embodied in generalized additive models (GAM), with the inclusion of random effects provided by generalized linear mixed models (GLMM) (Breslow and Clayton, 1993).

In this Chapter, firstly, we will present the connection between penalized smooth terms and random effects, then we will discuss GAMMs and finally we will focus on parameter estimation.

4.1 Penalized smooth term as Mixed model representation

Mixed models are an extension of regression models that allow inclusion of random effects. They also turn out to be closely related to smoothing, since

smoothing parameters can be viewed as quantities related to the random effect variances in a mixed model framework. This makes it possible to use mixed model methodology and software for penalized spline regression.
The linear mixed model can be defined as

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\epsilon}, \tag{4.1}$$

with

$$\boldsymbol{y}|\boldsymbol{b} \sim N(\boldsymbol{X\beta} + \boldsymbol{Zb}, \boldsymbol{R}), \qquad \boldsymbol{b} \sim N(\boldsymbol{0}, \boldsymbol{G}) \tag{4.2}$$

Thereby $\boldsymbol{y}$ is a response vector, $\boldsymbol{\beta}$ is the vector of fixed effects, and $\boldsymbol{b}$ is the vector of the random effects. $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the model matrices of the fixed and random effects respectively, and $\boldsymbol{\epsilon}$ is the error term. It is assumed

$$E\begin{bmatrix}\boldsymbol{b}\\ \boldsymbol{\epsilon}\end{bmatrix} = \begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}$$

and

$$Cov\begin{bmatrix}\boldsymbol{b}\\ \boldsymbol{\epsilon}\end{bmatrix} = \begin{bmatrix}\boldsymbol{G} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{R}\end{bmatrix},$$

where $\boldsymbol{G}$ and $\boldsymbol{R}$ are positive definite covariance matrices of $\boldsymbol{b}$ and $\boldsymbol{\epsilon}$, respectively. Tipically it is assumed that the random effects and the error terms are indipendent.
Estimation of the fixed effects $\boldsymbol{\beta}$ can be carried out via the marginal linear model corresponding to (4.1), namely

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}^*,$$

where $\boldsymbol{\epsilon}^* = \boldsymbol{Zb} + \boldsymbol{\epsilon}$ with $Cov(\boldsymbol{\epsilon}^*) = \boldsymbol{ZGZ}^T + \boldsymbol{R} \equiv \boldsymbol{V}$, for a given covariance matrix $\boldsymbol{V}$ depending on some variance parameters $\boldsymbol{\theta}$. Then the fixed effect estimates are (Ruppert *et al.*, 2003)

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}. \tag{4.3}$$

The estimate $\hat{\boldsymbol{\beta}}$ is referred to generalized least squares (GLS) and it is the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$.

Given the fixed effect estimate in (4.3), random effects $\boldsymbol{b}$ can now be predicted resulting in the best linear predictor (BLP) (Ruppert *et al.*, 2003)

$$\hat{\boldsymbol{b}} = \boldsymbol{G}\boldsymbol{Z}^T\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}). \tag{4.4}$$

In Section 3.1 we considered the ordinary nonparametric regression model and showed how f could be estimated by penalized splines in (3.4). The same penalized estimate can be obtained from a mixed model. For the sake of simplicity, we treat the linear case and suppose that the errors satisfy $Cov(\boldsymbol{\epsilon}) = \boldsymbol{R} = \sigma_\epsilon^2\boldsymbol{I}$. Using the truncated polynomial basis functions with K knots $(k_1, \ldots, k_K)$, the linear spline model for f is

$$f(x_i) = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} b_k(x_i - x_k)_+, \tag{4.5}$$

where $(x_i - x_k)_+$ are the truncated basis functions introduced in Section 3.1. Now, if we denote $\boldsymbol{X}$ as the model matrix with i*th* row $[1, x_i]$, $\boldsymbol{Z}$ as the model matrix with i*th* row $[(x_i - x_k)_+, \ldots, (x_i - x_K)_+]$, $\boldsymbol{\beta} = [\beta_0, \beta_1]^T$, and $\boldsymbol{b} = [b_1, \ldots, b_K]^T$, it is possible to rewrite (4.5) in a mixed framework as in (4.1) assuming $\boldsymbol{b} \sim N(0, \sigma_b^2)$. It was demonstrated that the ratio of variances $\frac{\sigma_\epsilon^2}{\sigma_b^2}$ in the mixed model framework plays the role of the smoothing parameter λ, that is $\lambda = \frac{\sigma_\epsilon^2}{\sigma_b^2}$. In this sense penalized spline smoothing is equivalent to the parameter estimation in a linear mixed model, thus estimation can be carried out by means of standard mixed model software.

It should be noted that the inverse of the penalty matrix imposed on spline coefficients has to be a proper covariance matrix that is symmetric and positive definite. While this is unproblematic for truncated polynomials as shown above (covariance matrix is just the identity matrix), other basis functions with corresponding penalties need to be adjusted in order to be represented by a linear mixed model.

4.2 Generalized Additive Mixed Models

A *generalized additive mixed model* (GAMM) is just a generalized linear mixed model (GLMM), in which part of the linear predictor is specified in terms of smooth functions of covariates

$$g(E(y_i|\boldsymbol{b})) = \boldsymbol{x}_i^T\boldsymbol{\beta} + f_1(x_{1i}) + f_2(x_{2i}, x_{3i}) + \cdots + \boldsymbol{z}_i^T\boldsymbol{b} \tag{4.6}$$

where y_i is a univariate response which has some exponential family distribution, $\boldsymbol{\beta}$ is a vector of fixed parameters, $\boldsymbol{x}_i^T$ is a row of a fixed effects model matrix, $f_j s$ are smooth functions of covariates, $\boldsymbol{z}_i^T$ is a row of a random effects model matrix, $\boldsymbol{b} \sim N(0, \boldsymbol{G})$ is a vector of random effects coefficients with unknown positive definite covariance matrix.

These models provide a unified likelihood framework for modelling response data as a function of linear and smooth terms with inclusion of random effects. The major difficulty in making inference is that a full likelihood analysis is burden by often intractable numerical integration. Due to the connection between penalized splines and random effects illustrated in the previous Section 4.1, it is possible to estimate GAMM via GLMM or via GAM. The following Section 4.3 discusses these alternative approaches.

4.3 Parameter Estimation

If the primary interest is in estimating the smooth relationships and if the random effects structure is simple and low dimensional, estimation via GAM (i.e penalized likelihood) has to be preferred. Alternatively, if the random effects structure is complex and high dimensional, estimation via GLMM represents a better choice.

4.3.1 GAMM estimation based on GLMMs settings

When we consider GAMM in a GLMM setting, we are treating smooth functions as random effects.

The likelihood for a GLMM is obtained by considering the joint distribution of the response conditional on the random effects. The model parameters in the model are the fixed effects $\boldsymbol{\beta}$ and the variance parameters $\boldsymbol{\theta}$, and the corresponding marginal likelihood is

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}) &= f(\boldsymbol{y}; \boldsymbol{\beta}, \boldsymbol{\theta}) \\ &= \int_{R^q} f(\boldsymbol{y}|\boldsymbol{b}) f(\boldsymbol{b}) d\boldsymbol{b} \\ &= (2\pi)^{-q/2} |\boldsymbol{G_\theta}|^{-1/2} \exp\{\boldsymbol{1^T} c(\boldsymbol{y})\} \boldsymbol{J}(\boldsymbol{\beta}, \boldsymbol{\theta}), \end{aligned} \tag{4.7}$$

where

$$\boldsymbol{J}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \int_{R^q} \exp\{\boldsymbol{y^T}(\boldsymbol{X\beta} + \boldsymbol{Zb}) - \boldsymbol{1^T} a(\boldsymbol{X\beta} + \boldsymbol{Zb}) - \frac{1}{2}\boldsymbol{b^T} \boldsymbol{G_\theta^{-1}} \boldsymbol{b}\} d\boldsymbol{b}. \tag{4.8}$$

Maximum likelihood estimation of $\boldsymbol{J}(\boldsymbol{\beta}, \boldsymbol{\theta})$ is complicated by the presence of this q-dimensional integral, where q, in the penalized fitting, is the number of knots. There has been a great deal of research, accelerating in the

1990s, on remedies to these computational problems. Usually, this complex integral is solved via Laplace approximation of PQL (Breslow and Clayton, 1993). PQL iteration is very time consuming and suffers from convergence problems. Furthermore, it is based on a quasi-likelihood method, and BIC or AIC cannot be calculated. The development of theory for BIC or AIC for these models is still an open research topic, and criteria based on full likelihood models are often preferred.

4.3.2 GAMM estimation based on GAMs settings

As alternative to GLMM-based estimation it is possible to treat the random effects as penalized regression terms (Wood, 2008, 2011b), i.e. via GAMs. In Chapter 3, it was shown how the problem of estimating an additive model becomes the problem of estimating model coefficients and smoothing parameters for a penalized likelihood maximization problem, once a basis for the smooth function has been chosen, together with associated measure of function wiggliness. In a generalized case, the penalized likelihood maximization problem is solved by penalized iteratively weighted least squares (P-IWLS), while the smoothing parameters, as for the linear case, can be estimated using cross validation (Wood, 2008) or likelihood criteria (Wood, 2011b). In a GAM setting, a GAMM can be expressed as

$$g(\mu_i) = \boldsymbol{x}_i^{*T}\boldsymbol{\beta}^* + \sum_j f_j, \tag{4.9}$$

where $g(\cdot)$ is a specific link function, $\mu_i = \mathbb{E}(Y_i)$ and Y_i is the response variable, which follows some exponential family distribution. $\boldsymbol{x}_i^{*T}$ is the ith row of the model matrix for any strictly parametric model components, $\boldsymbol{\beta}^*$ are the corresponding coefficients, some of which may be random, and f_js are smooth functions of some covariates. f_js are represented via regression spline bases (Section 3.1), with associated measures of function roughness which can be expressed as quadratic forms in the basis coefficients. By expressing f_js via basis functions and relevant coefficients, (4.9) can be re-written as a generalized linear model (GLM) (MacCullagh and Nelder, 1989)

$$g(\mu_i) = \boldsymbol{x}_i^T \boldsymbol{\beta}, \tag{4.10}$$

where $\boldsymbol{\beta}$ now includes $\boldsymbol{\beta}^*$ (fixed and random coefficients) and also the basis coefficients relevant to f_js, and $\boldsymbol{x}_i$ is the ith row of the model matrix, which includes the columns of $\boldsymbol{X}^*$ and columns representing the basis functions evaluated at the covariate values. If the spline bases dimensions are sufficiently large to ensure reasonably low bias, then maximum likelihood estimation of model (4.10) will almost certainly lead to overfitting. To avoid that, the model is estimated by penalized likelihood maximization via P-IWLS, where the penalties control overfit. The penalized likelihood for (4.10) is

$$l_p(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) - \frac{1}{2} \sum_j \lambda_j \boldsymbol{\beta}^T \boldsymbol{S}_j \boldsymbol{\beta}, \tag{4.11}$$

where $\boldsymbol{S}_j$ are positive semidefinite matrices and may also be components of more general random-effects precision matrices, and λ_j are positive smoothing parameters. Usually $\boldsymbol{\beta}^T \boldsymbol{S}_j \boldsymbol{\beta}$ measure the wiggliness of f_j, and λ_j control smoothness of f_j.

To select the most appropriate values for λ_j, it is possible to use either methods that minimize model prediction error (Akaike's information criterion (AIC), cross-validation or generalized cross-validation (GCV) (Wahba, 1975; Wood, 2004, 2008)), or maximum likelihood methods (Wood, 2011b). Wood (2011b) showed that the maximum likelihood methods (REML or ML) for $\boldsymbol{\lambda}$ selection perfom better than methods based on minimization of model prediction error. In this Section we briefly will present the REML method introduced by Wood (2011b) for $\boldsymbol{\lambda}$ estimation. The key for understanding REML or ML methods is that model (4.6) can be viewed as a generalized linear mixed model in the form shown in (4.10), and at this step smooth functions are viewed as random effects (Kimeldorf and Wahba, 1971), so that λ_j are treated as variance parameters which can be estimated by maximum (marginal) likelihood (Anderssen and Bloomfield, 1974), or restricted maximum likelihood. From a Bayesian perspective, it is well known that the penalized likelihood estimates of the coefficients $\hat{\boldsymbol{\beta}}$ are the posterior modes of the distribution of $\boldsymbol{\beta}|\boldsymbol{y}$ if $\boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{S}^{-}\phi)$ (Wahba, 1983), where $\boldsymbol{S} = \sum_j \lambda_j \boldsymbol{S}_j$ and $\boldsymbol{S}^{-}$ is the generalized inverse matrix of $\boldsymbol{S}$ and ϕ is the scale parameter. Given that in this way the elements of $\boldsymbol{\beta}$ are viewed as random effects, it is natural to try to estimate λ_j via ML or REML, where λ_j now control the dispersion of the priors, and hence the smoothness of f_j. To do this, Wood (2011b) takes the Laird and Ware's (1982) approach to REML, in which fixed effects are viewed as random effects with improper uniform priors and they are integrated out, and he uses the penalties to define (independent) improper priors on the wiggliness of each f_j, so that the improper prior density for $\boldsymbol{\beta}$ can be assumed

$$f_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \frac{|\boldsymbol{S}/\phi|_{+}^{0.5}}{\sqrt{(2\pi)^{n_b - M_p}}} \exp\{-\boldsymbol{\beta}^T \boldsymbol{S}\boldsymbol{\beta}/(2\phi)\}, \tag{4.12}$$

where n_b is the dimension of $\boldsymbol{\beta}$ and M_p is the dimension of the null space of $\boldsymbol{S}$. Integrating $\boldsymbol{\beta}$ out of $f(y, \boldsymbol{\beta}) = f_y(y|\boldsymbol{\beta}) f_{\boldsymbol{\beta}}(\boldsymbol{\beta})$, the marginal restricted likelihood depending only on $\boldsymbol{\lambda}$ is obtained from Laplace approximate REML criterion

$$L_R(\boldsymbol{\lambda}, \phi) = L(\hat{\boldsymbol{\beta}}) f_{\beta}(\hat{\boldsymbol{\beta}}) \frac{\sqrt{2\pi}^{n_b}}{|\boldsymbol{H} + \boldsymbol{S}/\phi|^{0.5}}. \tag{4.13}$$

In practice, estimation of the coefficients $\boldsymbol{\beta}$ and the smoothing parameters $\boldsymbol{\lambda}$ is carried out iteratively, where at each iteration, having fixed the λ value, the optimal $\boldsymbol{\beta}$ is found via P-IWLS. More specifically, an outer iteration updates the smoothness parameters by (4.13), and at each iteration step an inner P-IWLS iteration is carried out to find $\hat{\boldsymbol{\beta}}$.

The REML smoothness selection criteria of Wood (2011b) overcomes convergence problems of proposed single-iteration methods for REML or ML estimation of semiparametric GLMs (Wood, 2004; Breslow and Clayton, 1993).

Relative to PQL parameter estimation, the P-IWLS estimation using REML method for $\boldsymbol{\lambda}$ estimation offers two substantial advantages for GAMM estimation and smoothing parameter selection. The first advantage is that P-IWLS is computationally more reliable and much quicker than PQL. Since the smoothing parameters are based on optimizing a properly defined function, fitting does not suffer from the convergence problems of PQL. The second motivation is that it is possible to calculate the value of the optimized BIC or AIC useful for model comparisons, since we have a full likelihood.

Confidence intervals with GAMs

Various authors have proposed approximate Bayesian interval estimates for such models, based on extensions of the work of Wahba (1983) and Silverman (1985) on smoothing spline models of Gaussian data, but testing of such intervals has been rather limited and there is little supporting theory for the approximations used in the generalized case. Wood (2006c) in his work improved this situation by providing simulation tests and obtaining asymptotic results supporting the approximations employed for the generalized case. The simulation results suggested that while across-the-model performance was good, component-wise coverage probabilities were not so reliable. Since this was likely to result from the neglect of smoothing parameter variability, a simple and efficient simulation method was proposed to account for smoothing parameter uncertainty: this demonstrated a substantial improvement of the performance of component-wise intervals.

Bayesian approach (Wahba, 1983; Silverman, 1985) is preferred to frequentist approach (Wahba, 1980; Eilers and Marx, 1996) because generally, in a frequentist approach, inference with these models is complicated by the fact that, while the quadratic penalty term acts to limit estimator variance, it also biases the parameter estimators $\hat{\boldsymbol{\beta}}$, because of $E(\hat{\boldsymbol{\beta}}) \neq \boldsymbol{\beta}$, giving poor results in terms of realized coverage probabilities.

In a Bayesian approach it is necessary to specify a prior distribution on the parameters $\boldsymbol{\beta}$. Specifically let the improper prior for $\boldsymbol{\beta}$ be

$$f_\beta(\boldsymbol{\beta}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^T \left(\sum \mathbf{S}_j/\tau_j\right)\boldsymbol{\beta}\right\}, \tag{4.14}$$

where the τ_j parameters control the dispersion of the prior. Here, the prior is equivalent to assuming that each of the components of model wiggliness, $\boldsymbol{\beta}^T\mathbf{S}\boldsymbol{\beta}$,is an independent exponentially distributed random variable with expected value τ_j. The independence assumption is quite natural in situations in which the penalties are 'non-overlapping', for example when $\sum \mathbf{S}_j$ is block-diagonal, as in the case of GAMs constructed from penalized regression splines. The prior is appropriate since it makes explicit the fact that it is believed that smooth models are more likely than wiggly ones, but it gives equal probability density to all models of equal smoothness; the latter feature makes the prior improper. Considering the model specification

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{W}^{-1}\boldsymbol{\sigma}^2), \tag{4.15}$$

it is possible to write the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{\beta}$ as

$$f(\boldsymbol{y}|\boldsymbol{\beta}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})^T \boldsymbol{W}(\boldsymbol{y} - \boldsymbol{X\beta})/\sigma^2\right\}. \tag{4.16}$$

Using Bayes rule we have

$$\begin{aligned} f(\boldsymbol{\beta}|\boldsymbol{y}) &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{y}^T\boldsymbol{W}\boldsymbol{y}/\sigma^2 - 2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}/\sigma^2 + \boldsymbol{\beta}^T(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}/\sigma^2 + \sum \boldsymbol{S}_j/\tau_j)\boldsymbol{\beta})\right\} \\ &\propto \exp\left\{-\frac{1}{2}(-2\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}/\sigma^2 + \boldsymbol{\beta}^T(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}/\sigma^2 + \sum \boldsymbol{S}_j/\tau_j)\boldsymbol{\beta})\right\}. \end{aligned} \tag{4.17}$$

If now we consider an $\boldsymbol{\alpha} \sim N((\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + \sum \lambda_j\boldsymbol{S}_j)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}, (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + \sum \lambda_j\boldsymbol{S}_j)^{-1}\sigma^2)$, the probability density function for $\boldsymbol{\alpha}$ is

$$\begin{aligned} f_\alpha(\boldsymbol{\alpha}) &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\alpha} - (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + \sum \lambda_j\boldsymbol{S}_j)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y})^T(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + \sum \lambda_j\boldsymbol{S}_j)(\boldsymbol{\alpha} - (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + \sum \lambda_j\boldsymbol{S}_j)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y})/\sigma^2\right\} \\ &\propto \exp\left\{-\frac{1}{2}(-2\boldsymbol{\alpha}^T\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}/\sigma^2 + \boldsymbol{\alpha}^T(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}/\sigma^2 + \sum \lambda_i\boldsymbol{S}_j/\sigma^2)\boldsymbol{\alpha})\right\}. \end{aligned} \tag{4.18}$$

Comparing equation in (4.18) and in (4.17) it is clear that if we choose $\tau_j = \sigma^2/\lambda_j$, then

$$\boldsymbol{\beta}|\boldsymbol{y} \sim N\left(\hat{\boldsymbol{\beta}}, \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} + \sum \lambda_j\boldsymbol{S}_j\right)^{-1}\sigma^2\right), \tag{4.19}$$

where $\hat{\boldsymbol{\beta}}$ is the penalized least squares estimate. This result (4.19) yields a self consistent basis for constructing Bayesian confidence intervals for any quantity derived from $\boldsymbol{\beta}$. Such intervals should not suffer from the effects of estimator bias in the way that a more naive frequentist approach does. In the generalized case Wood (2006c) discusses that assuming the prior in (4.14) for $\boldsymbol{\beta}$, it can now be shown that, in the large sample limit, the posterior distribution for $\boldsymbol{\beta}$ is that in (4.19), where now $\hat{\boldsymbol{\beta}}$ is the maximizer of the penalized maximum likelihood for a penalized GLM estimated by the minimization of

$$-l(\boldsymbol{\beta}) + \frac{1}{2}\sum_j \lambda_j \boldsymbol{\beta}^T \boldsymbol{S}_j \boldsymbol{\beta}, \tag{4.20}$$

and the diagonal matrix $\boldsymbol{W}$ now has entries $W_{ii} = (g'(\mu_i)^2 V(\mu_i))^{-1}$, where V is the variance function, such that $V(\mu_i)\sigma^2$ is the variance of Y_i and σ^2 is the scale parameter. For many exponential family distributions the scale parameter σ^2 is known, but if an estimate is needed the Pearson estimator can be used.

P-values for the smooth components

Another important issue concerns the testing of smooth components of a GAM, namely whether some subset $\boldsymbol{\beta}_j$ of $\boldsymbol{\beta}$ is equal to zero. Wood (2013) proposed a Wald-type test for $\boldsymbol{f} = \boldsymbol{0}$, where $\boldsymbol{f}$ is the vector of evaluated values for the smooth component of interest. It was shown that confidence intervals for smooth components exhibit good across-the-function coverage probabilities if based on the approximate result

$$\hat{\boldsymbol{f}}_j(i) \sim N(\{\boldsymbol{f}_j(i), \boldsymbol{V}_{\boldsymbol{f}_j}(i,i)\}), \tag{4.21}$$

where $\boldsymbol{V}_{\boldsymbol{f}_j}$ is the covariance matrix for $\boldsymbol{f}$ according to the Bayesian view of the smoothing process, in which, as shown in the previous Section, the smoothing penalty is induced by an improper Gaussian prior on $\boldsymbol{\beta}$. The key idea is to base the test statistics on the same distributional result that yields well-calibrated confidence intervals for $\boldsymbol{f}$ in (4.21), and defining X_j

the matrix such that $\hat{f}_j = X_j\hat{\boldsymbol{\beta}}$, then $V_{f_j} = X_j V_\beta X_j^T$. The Wald statistics corresponding to (4.21) is

$$T_r = \hat{f}_j^T V_{f_j}^{r-} \hat{f}_j, \tag{4.22}$$

where $V_{f_j}^{r-}$ is a rank-r pseudo-inverse of V_{f_j}. The main problem is then to choose r appropriately. Naive choices lead to poor power or even to an incorrect null distribution for p-values, but investigation of the structure of T_r suggests a relatively simple choice of r (see (Wood, 2013)). (4.22) has null distribution χ_r^2, when r is integer. For noninteger r the distribution still has $E(T_r) = r$ and $var(T_r) = 2r$ under the null hypothesis, but $T_r \sim \chi_{k-2}^2 + \nu_1 chi_1^2 + \nu_2 chi_1^2$. For technical details on this case of the noninteger r see Wood (2013).

Chapter 5

Modelling spatio-temporal elephant movement data

5.1 Introduction

As discussed previously the aim of this thesis is to model the daily elephant speed as a function of environmental variables, rainfall and temperature, accounting for spatio-temporal trends in the study area over time. As seen in Chapter 2 the relationship between the response daily speed and environmental covariates cannot be assumed linear, and it is necessary to take into account heterogeneity among elephants. For this aim, we employed the generalized additive mixed model (GAMM) framework which provides a powerful tool to fulfil the model requirements as described in Section 1.4. This Chapter presents the first application of this approach to spatio-temporal smoothing in a complicated ecological system. All the analyses were performed by means of the `mgcv` (Wood, 2011a) package in R (R Core Team, 2013). The package gives a simple, extensible collection of fitting routines, basis functions and diagnostics.

5.2 The proposed model framework

The response daily speed (km/h) was modelled using a GAMM framework. More specifically, since the response has a positive asymmetric distribution (Figure 2.4), a Tweedie distribution (Candy, 2004) was assumed

$$\texttt{speed}_{it} \sim Tweedie\{E(\texttt{speed}_{it}), \phi E(\texttt{speed}_{it})^p\}. \tag{5.1}$$

The Tweedie distribution is an exponential family dispersion model with variance function given by the power function $\phi\mu^p$ with $1 < p < 2$. The class of Tweedie models includes most of the important distributions commonly associated with GLMs. When $1 < p < 2$ the distribution of Y is intermediate between a Poisson and a Gamma distribution with mass at zero but otherwise continuous on the positive reals and it is a Gamma when $p = 2$ (Jorgensen, 1987). To select the appropriate value for the index p a grid search over 100 values for p from 1.1 to 2 was performed. For each fixed value of p, the model (5.2) was fitted and the best value of p was the one minimizing the BIC.

We used the BIC rather than the penalized log-likelihood since the degrees of freedom of the considered models are different because of the penalization. The Tweedie index p was set to 1.83 as this led to a lower BIC for all considered models. The BIC values with respect to different p values are reported in Figure 5.1.

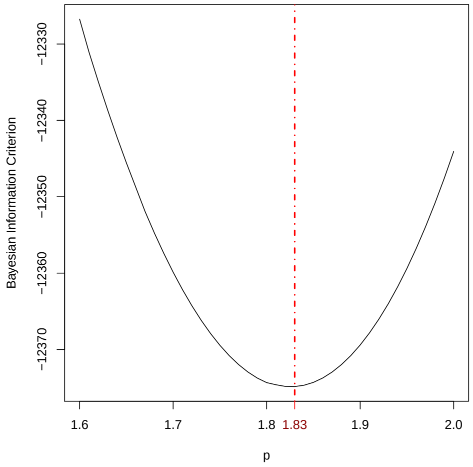

Figure 5.1: Baysian Information Criteria values for the Tweedie distribution from a grid over 100 values for the p index. The red dashed line represents the lowest BIC value at the p index for the Tweedie distribution ($p = 1.83$).

The regression equation of the proposed model is

$$\begin{aligned}\log\{E(speed_{it})\} &= \beta_0 + b_i + s(\texttt{day.year}_{it}) + r(\texttt{lon}_{it}, \texttt{lat}_{it}, \texttt{year}_{it})\\ &+ \sum_{l=0}^{3} f_l(\texttt{locRaif}_{i,t-l}) + \sum_{l=0}^{3} h_l(\texttt{avgTemp}_{i,t-l})\end{aligned} \tag{5.2}$$

for the elephant $i = 1, \ldots, m$, days of year $t = 1, \ldots, n_i$, where $\sum_{i=1}^{m} n_i = n$, and distributed lag (DL) $l = \{0, 1, 2, 3\}$. b_i is the random effect associated to elephant i, assumed i.i.d. $\boldsymbol{b} \sim N(0, \sigma_b^2)$, the function $s(\cdot)$ is a one dimensional smooth function of the seasonality effect using days of year represented from a cyclic cubic regression spline basis, $r(\cdot)$ is an invariant 3-dimensional tensor product of two bases: a two-dimensional smooth of thin plate spline basis for the space component (longitude and latitude), and a one-dimensional cubic regression spline basis for the time component given by years. Finally the $f_l(\cdot)s$, $h_l(\cdot)s$ functions are one dimensional DL smooth functions at lag $l = 0, 1, 2, 3$, respectively of rainfall and temperature values represented using cubic regression spline bases.

5.2.1 A three-dimensional spatio-temporal smoother

The term $r(\texttt{lon}, \texttt{lat}, \texttt{year})$ in (5.2) accounts for possible interaction effect of space and time. Similarly to Section 3.1 the model uses a tensor product between two bases: a two-dimensional isotropic spatial smooth (r_s) and a marginal one-dimensional smooth of time (r_t). The tensor product smooth presented in this Section refers to the general metology of Wood (2004, 2006b) for constructing scale invariant tensor product smooths of space-time dimension.

The spatial smooth (r_s) and the temporal smooth (r_t) can be written in terms of their basis decompositions as follows

$$r_s(\texttt{lon,lat}) = \sum_{q=1}^{Q} \delta_q D_q(\texttt{lon,lat}), \tag{5.3}$$

and

$$r_t(\texttt{year}) = \sum_{p=1}^{P} \alpha_p A_p(\texttt{year}). \tag{5.4}$$

$D_q(\texttt{lon,lat})$ and $A_p(\texttt{year})$ are thin plate spline and cubic spline basis functions (respectively), with corresponding parameters δ_q and α_p and spline dimensions Q and P.

In order to construct a three-dimensional tensor product smooth of space and time it is necessary for $r_t(\texttt{year})$ to vary smoothly within the spatial dimensions. This can be achieved by allowing the parameters α_p to vary with longitude ($\texttt{long}$) and latitude ($\texttt{lat}$). Using the spline structure for $r_s(\texttt{long},\texttt{lat})$ it is possible to write:

$$\alpha_p(\texttt{long},\texttt{lat}) = \sum_{q=1}^{Q} \delta_{pq} D_q(\texttt{long},\texttt{lat}). \tag{5.5}$$

By putting 5.5 into 5.4 we get

$$r_{s,t}(\texttt{long},\texttt{lat},\texttt{year}) = \sum_{q=1}^{Q}\sum_{p=1}^{P} \delta_{pq} D_q(\texttt{long},\texttt{lat}) A_p(\texttt{year}), \tag{5.6}$$

which emphasizes between the two marginal bases. To build a wiggliness measure relevant to 5.6 it is possible to start from wiggliness measures associated with the marginal smooth functions. Let J_s and J_t be measures of the wiggliness of the functions r_s and r_t respectively. The wiggliness for r_t is assumed to be the second order cubic spline penalty (Section 3.1)

$$J_t(r_t) = \int (\partial^2 r_t / \partial \texttt{year}^2)^2 d\texttt{year}. \tag{5.7}$$

An overall penalty for the tensor product smoother can be obtained by applying the penalties of the spatial smooth to the spatially varying coefficients of the marginal temporal smooth, $\alpha_p(\texttt{long},\texttt{lat})$,

$$\sum_{p=1}^{P} J_s\{\alpha_p(\texttt{long},\texttt{lat})\}, \tag{5.8}$$

and equivalently the penalties of the temporal smooth to the temporally varying coefficients of the marginal spatial smooth, $\delta_q(\texttt{year})$, are applied,

$$\sum_{q=1}^{Q} J_t\{\delta_q(\texttt{year})\}. \tag{5.9}$$

It follows that the roughness of $r_{s,t}$ can be written as the sum of the two penalties weighted by smoothing parameters for space λ_s and time λ_t:

$$J(r_{s,t}) = \lambda_s \sum_{p=1}^{P} J_s\{\alpha_p(\texttt{long,lat})\} + \lambda_t \sum_{q=1}^{Q} J_t\{\delta_q(\texttt{year})\}, \tag{5.10}$$

where as usual, λ_s and λ_t are the smoothing parameters for space and time respectively.

5.3 Results

In our model selection strategy we started with a model up to three lags (eq. (5.2)) and then we checked the possibility of simplifying it. The starting model was named M_1 and its simplified versions were named M_2, M_3, and M_4. Among the biologically plausible lag values (≤ 3), two lags were selected according to the BIC whose results are shown in Table 5.1. For model selection and comparison a generalized version of the BIC was used, that is $BIC = -2\log(\hat{L}) + \log(n)\mathrm{edf}$ (Kass and Raftery, 1995). The BIC is suitable for those situations in which there is a large sample size with respect to the number of parameters, which is the case of our application. In practice, due to the smaller penalty term, the AIC tends to keep more terms in the model than the BIC, hence BIC is preferred to AIC to avoid overfitting.

Table 5.1: Comparing models with different lag structures for rainfall and temperature.

Model	edf	BIC	lags in `locRaif`	lags in `avg.Temp`
M_1	175.5	-12273.4	3	3
M_2	173.3	-12275.9	3	2
M_3	173.9	-12281.4	2	3
M_4	171.7	-12283.9	2	2

The selected model M_4, was implemented using the basis function shown in Table 5.2 and the model results are reported in Table 5.3. The last column refers to the *p*-value testing for a zero effect of each term according to the methods described in Chapter 4.

Table 5.2: Type of basis and size per smooth term of the selected model.

Smooth term	Type of basis	Size of basis
`day.year`	`cc`	10
`space-time`	`tp` for space, `cr` for time	(50,5)
`locRaif` at lag 0	`cr`	10
`locRaif` at lag 1	`cr`	10
`locRaif` at lag 2	`cr`	10
`avg.Temp` at lag 0	`cr`	10
`avg.Temp` at lag 1	`cr`	10
`avg.Temp` at lag 2	`cr`	10

Table 5.3: Main results from the fitted model M_4.

Parametric coefficients			
	Estimate (SE)	*t*-value	*p*-value
Intercept	-0.92 (0.03)	-33.71	<0.0001
Approximate significance of smooth terms			
Smooth terms	edf	F	*p*-value
ID	9.09	22.32	<0.0001
day.year	7.23	122.09	<0.0001
space-time	138.67	11.06	<0.0001
locRaif, l = 0	1.81	0.91	0.4060
locRaif, l = 1	2.48	3.89	0.0082
locRaif, l = 2	1.09	4.59	0.0268
avgTemp, l = 0	3.14	12.70	<0.0001
avgTemp, l = 1	3.89	5.09	0.0001
avgTemp, l = 2	3.25	2.46	0.0410
BIC=-12283.9			

Figure 5.2 shows the estimated seasonality effect from the selected model M_4. As expected there is a strong seasonality effect of elephant movements. The estimated elephant speed has a significant non-linear down trend up to approximately day 245 (approx. September, 2), after this threshold elephants seem to increase their speed in the wet season until approximately day 30 (approx. January, 30).

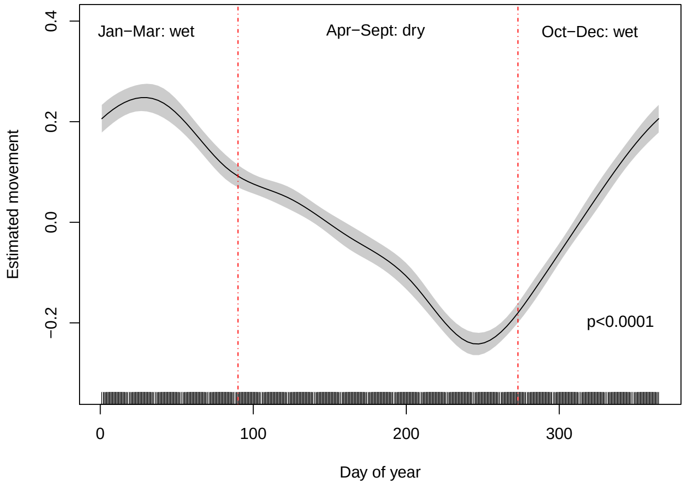

Figure 5.2: Estimated seasonality effect. The dashed red lines indicate the calendar seasonal thresholds (October-March, wet season; April-September, dry season).

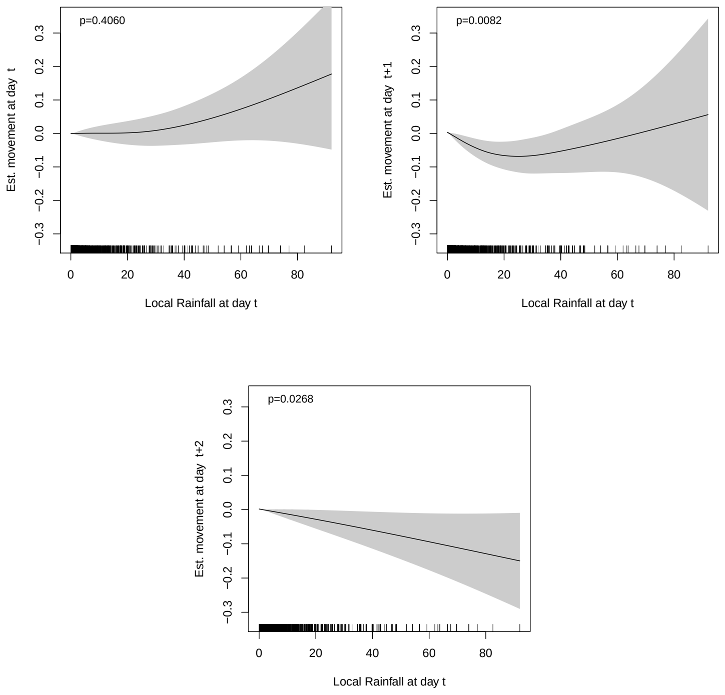

Figure 5.3: Estimated rainfall effect at lag 0, 1 and lag 2, with ± two standard errors (grey area).

As regards environmental variables there is a noteworthy effect of the local rainfall from lag 0 to lag 2, even if at lag 0 it is not statistically significant. Plots of these effects are shown in Figure 5.3. At lag 0, elephants seem to move almost constantly until a rainfall value of 20 mm, and after this value they seem to increase their speed. At lag 1 elephants seem to move less up to a rainfall value approximately of 15-20 mm, and then they increase their speed when rainfall increases, instead at lag 2 elephants seem to decrease their speed.

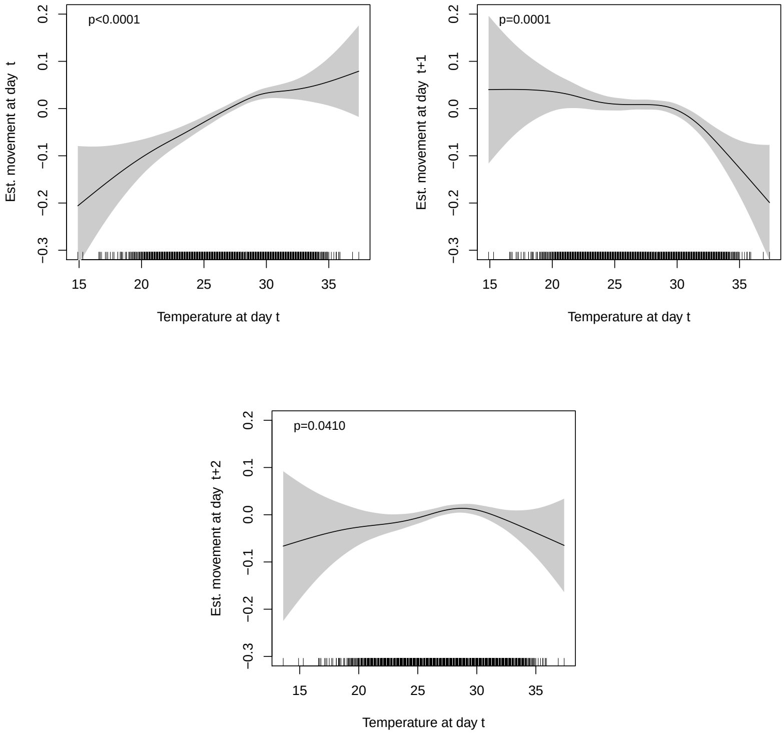

Figure 5.4: Estimated rainfall effect at lag 0, 1 and lag 2, with ± two standard errors (grey area).

We observe also a significant temperature effect up to lag 2 (Tab. 5.3). In Figure 5.4 at lag 0 it seems that elephants move faster up to a temperature value approximately of 30 °C, and in the following days, at lag 1 and lag 2, elephants seem to decrease their speed after this threshold of 30 °C. This suggests that elephants have a thermal limit of tolerance, beyond which they have to slow their movements.

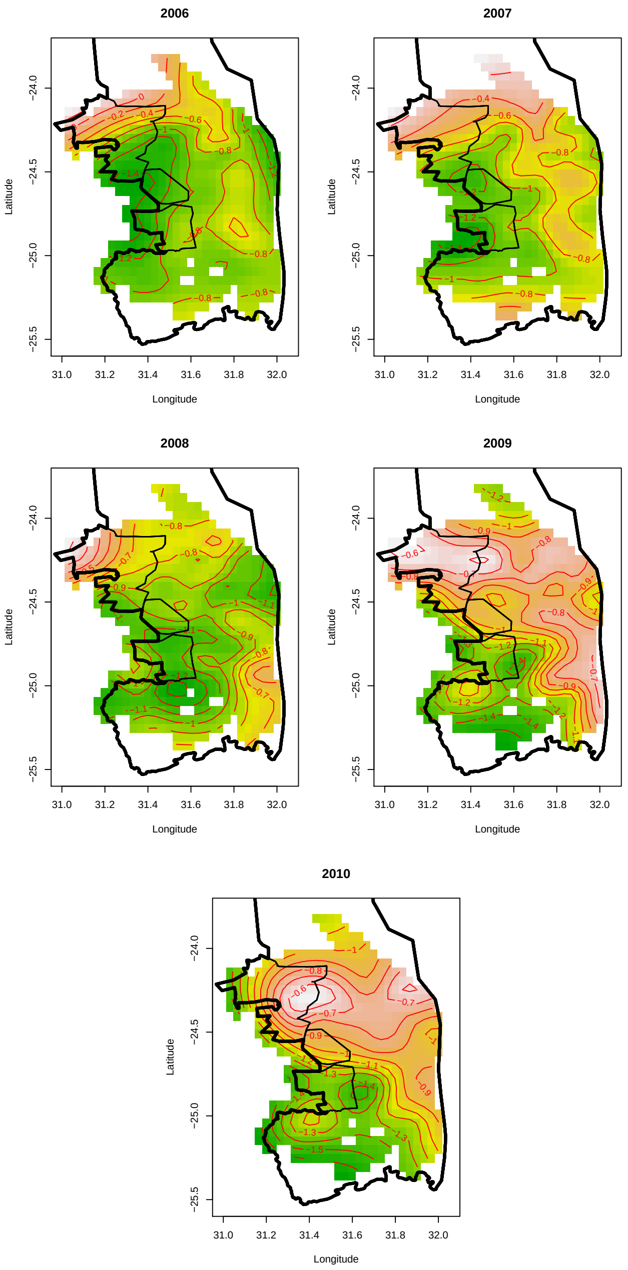

Figure 5.5: Estimated elephants movements over different years of study (2006-2010). Lighter colours represent areas with higher movements.

The selected model shows a statistically significant spatio-temporal interaction effect (Table 5.3). This means that there is a spatial effect on estimated elephant speed over years. Specifically, this suggests that in terms of speed of movement there are changes across the considered years in certain spatial areas. A possible explanation of why these changes in the elephant movement behaviour exist over years may be linked to changes in abiotic and biotic factors. Maps of estimated elephant speed over years are shown in Figure 5.5.

5.3.1 Model diagnostics

Model diagnostics was assessed using the qq-plot and the histogram of residuals for normality, residuals versus linear predictor for homogeneity, and response values versus fitted values for model fit.

Graphical diagnostics for the final model is shown in Figure 5.6. The qq-plot as well as the histogram of the residuals show no relevant anomalies. The histogram of residuals seems to be approximately symmetrical and qq-plot shows some outliers, especially in the right side. Plot of residuals versus linear predictor shows that variance is approximately constant with no clear violation of homogeneity. This indicates that the selected value of the index p for the Tweedie distribution is suitable.

Residual spatial and temporal autocorrelation was evaluated by means of variograms and partial autocorrelation functions for each elephant (see plots in Figures 5.7 and 5.8).

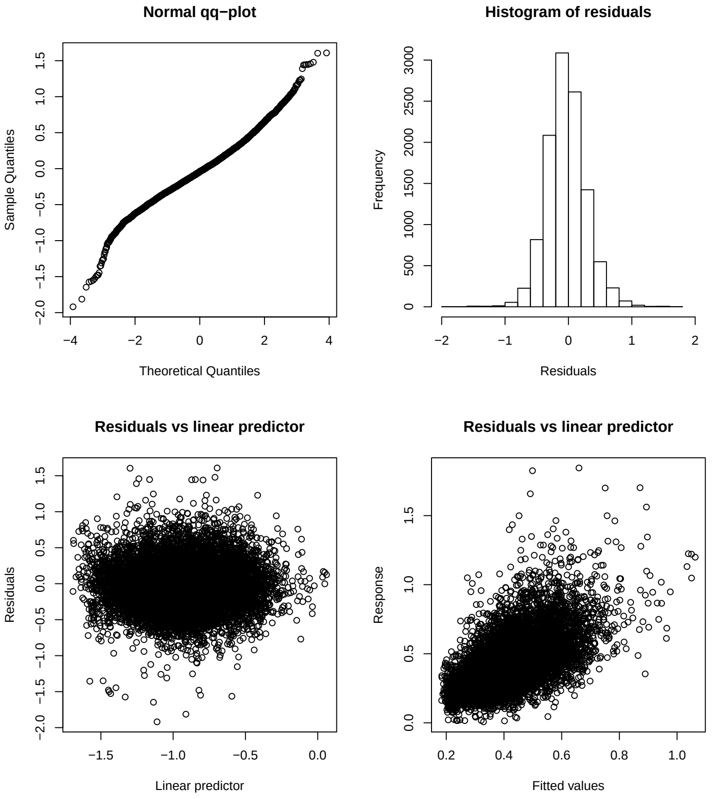

Figure 5.6: Model validation plots of the selected model.

The variograms also include envelops obtained by permutation (Ribeiro Jr and Diggle, 2001): values within the envelops suggest that no important spatial correlation affects the residuals from the final fitr. On the other hand residual temporal autocorrelation still persists. Especially, some elephants, AM107, AM253, AM254, AM91, and AM99 show a higher PACFs than other elephants, from 0.2 to 0.4 at lag 1, then at larger lags, PACF is very low, around 0.1.

Furthermore, we investigated the residual temporal correlation, assuming for the final model an autoregressive correlation structure of order 1 for the within-group correlation. Findings with respect to covariate effects are unchanged, and the residual temporal autocorrelation still remains, although to minor extent. Results are shown in Figure 5.9, where we can see that the values of the PACFs of standardized residuals are around 0.1, and the estimated correlation parameter is low ($\hat{\rho} = 0.26$).

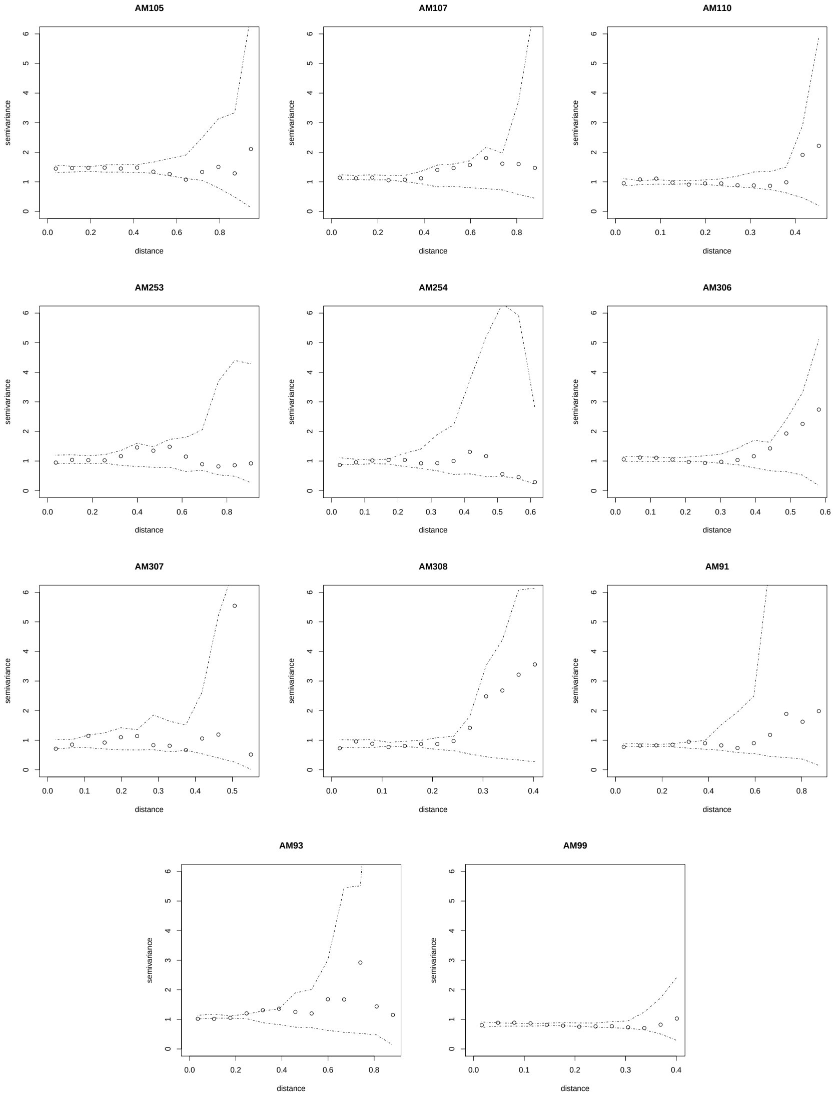

Figure 5.7: Variograms of standardized residuals for each elephant. The envelops represent minimum and maximum values derived by permuting data points on the spatial locations for each elephant. The variograms indicate model mis-specification if the empirical points lie outside the envelops.

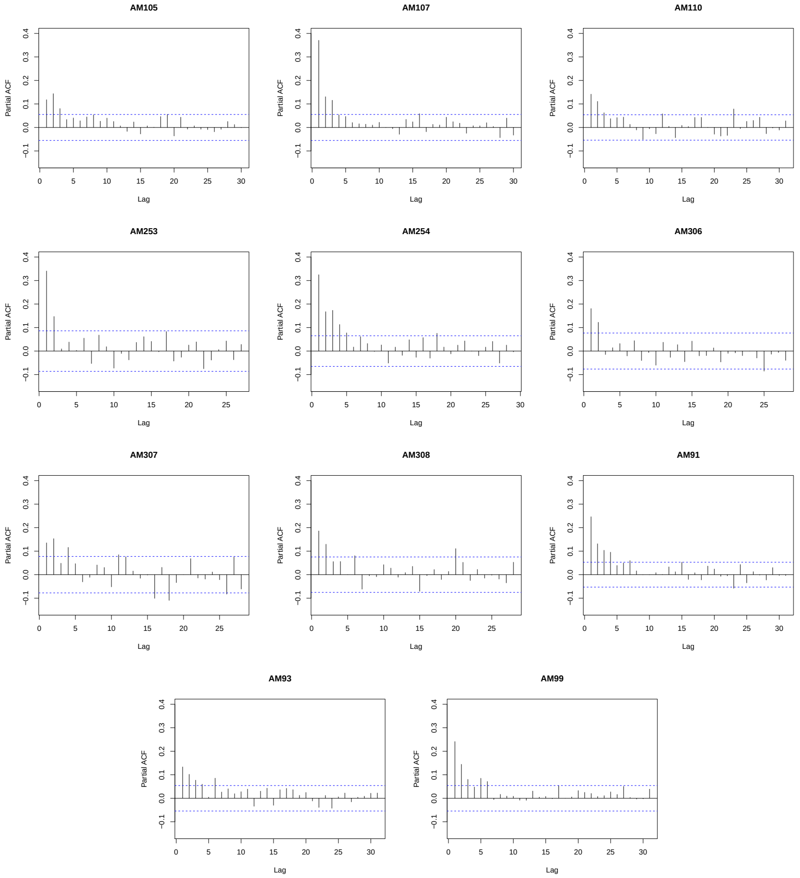

Figure 5.8: Partial autocorrelation functions of standardized residuals for each elephant.

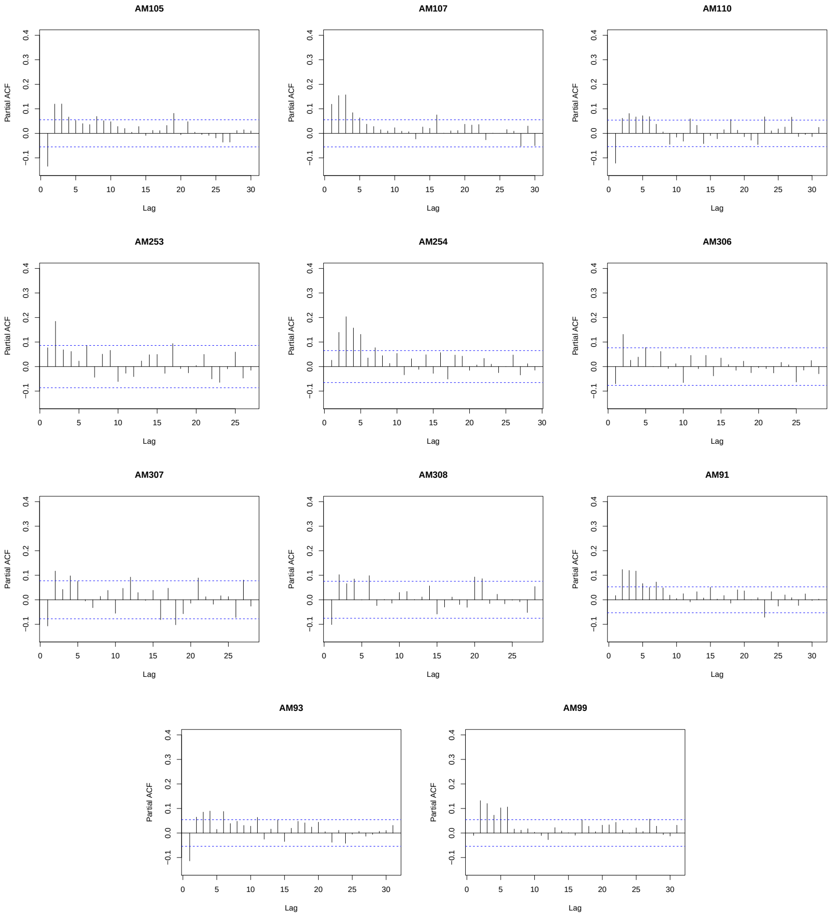

Figure 5.9: Partial autocorrelation functions of standardized residuals for each elephant for the selected model with a specified autoregressive structure of order 1 for the within-group error. The estimated correlation parameter is $\hat{\rho} = 0.26$.

Chapter 6

A possible improvement

In this Chapter we deal with some further advancements on the proposed model discussed in Chapter 5. Specifically, we discuss about including a new variable, the "cumulative rainfall", which can be considered as a 'wetness' measure.

6.1 Including *cumulative rainfall*

Another important goal of our research is to obtain a proxy of the 'wetness' to investigate the effect on elephant movements. A measure of 'cumulative rainfall' was obtained for each elephant, by means of a sum of local rainfall at each seasonal breakpoint obtained in Birkett *et al.* (2012). According to these 'seasonal breakpoints' the time axis of each elephant is into several intervals and within each interval the cumulative rainfall was computed. Birkett *et al.* (2012) used a piecewise regression model to obtain seasonal breakpoints separately for each considered elephant within each considered year. The framework of the model considered in this Section is the same of that considered for the selected model M_4 including the new variable

$$\begin{aligned}\log\{E(speed_{it})\} = \beta_0 + b_i + s(\texttt{day.year}_{it}) + r(\texttt{lon}_{it}, \texttt{lat}_{it}, \texttt{year}_{it}) \\ + \sum_{l=0}^{2} f_l(\texttt{locRaif}_{i,t-l}) + \sum_{l=0}^{2} h_l(\texttt{avgTemp}_{i,t-l}) + c(\texttt{cumRain}_t).\end{aligned} \tag{6.1}$$

The best value of the p index for the Tweedie distribution according to the BIC is substantially unchanged (p=1.84). The function `c(CumRain)` is a one-dimensional cubic regression spline of the cumulative rainfall effect. Model estimates and the estimated effects are reported in Table 6.1 and Figures 6.1 and 6.2, respectively.

We focus our discussion only on the cumulative rainfall effect, since all terms included in the model (eq. (6.1)) remained approximately the same. The estimated smooth effect of cumulative rainfall is shown at the bottom of Figure 6.2. Besides some unexplained seasonality captured by this term, the plot suggests that elephants increase their speed as cumulative rainfall increases at low levels, i.e. at the beginning of the wet season. As rainfall accumulates, namely the wet season goes on, elephants slow down possibly due to availability of forage in general.

Table 6.1: Main results from the fitted model in (6.1).

Parametric coefficients			
	Estimate (SE)	*t*-value	*p*-value
Intercept	-0.93 (0.02)	-32.43	<0.0001
Approximate significance of smooth terms			
Smooth terms	edf	F	*p*-value
ID	9.12	21.68	<0.0001
day.year	7.16	115.71	<0.0001
space-time	137.71	10.90	<0.0001
locRaif, l = 0	1.84	1.53	0.2081
locRaif, l = 1	2.64	3.73	0.0091
locRaif, l = 2	1.02	3.98	0.0444
avgTemp, l = 0	5.27	8.71	<0.0001
avgTemp, l = 1	3.89	5.02	0.0001
avgTemp, l = 2	3.39	2.64	0.0283
cum.rain	8.44	8.97	<0.0001
BIC=-12284.67			

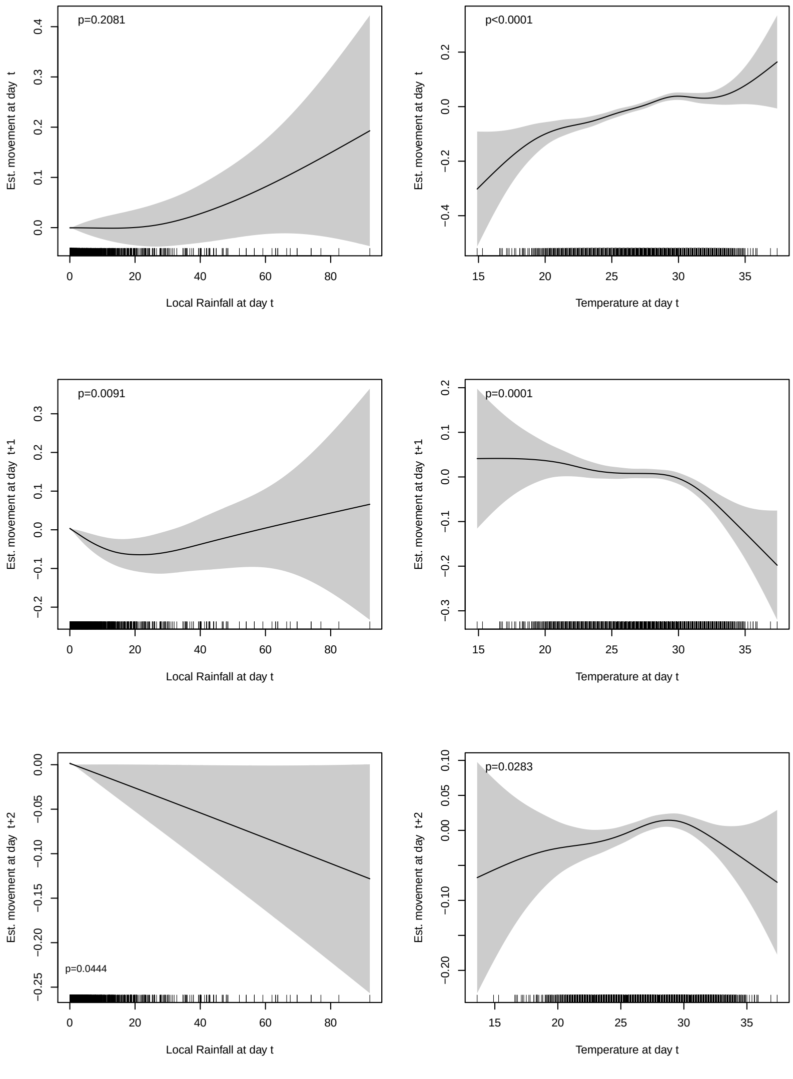

Figure 6.1: Estimated lagged effects of rainfall (left column) and temperature (right column) with ± two standard errors (shade grey area).

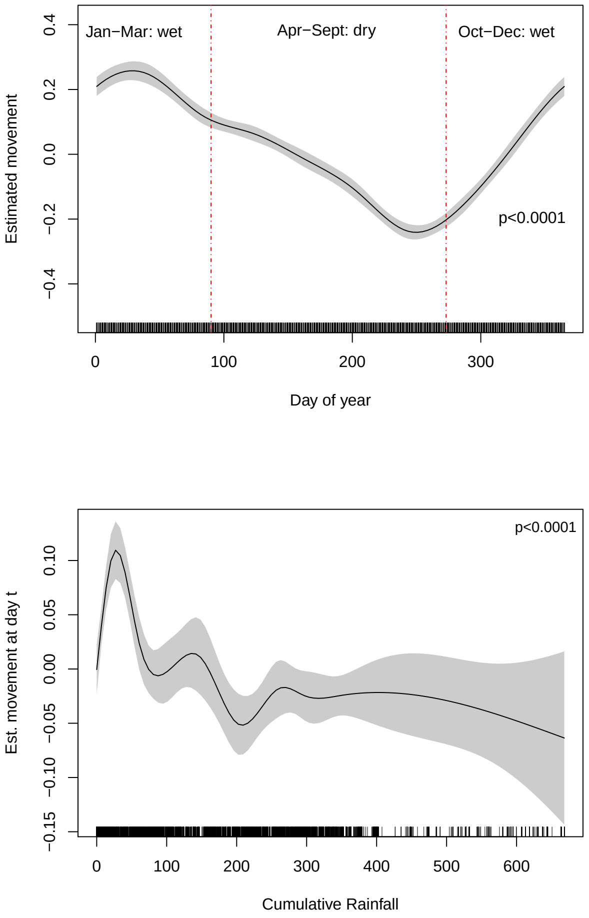

Figure 6.2: Estimated seasonality and cumulative rainfall effects with ± two standard errors (shade grey area). The dashed red lines indicate the calendar seasonal thresholds (October-March, wet season; April-September, dry season).

Chapter 7

Discussion

We used GAMMs to model spatio-temporal elephant movement data as a function of environmental variables. GAMMs are a flexible framework which takes into account random effects for heterogeneity, smooth terms for non linear effects of single or multiple covariates.

Our proposed model allowed an adequate assessment of environmental factors affecting elephant movements and of the spatial elephant movement trend across years. The proposed model included one-dimensional smooth functions of the seasonality effect (days of year), of distributed lag for rainfall and temperature, a three-dimensional smooth term accounting for space (longitude and latitude) and time (years), and the random effects associated to elephant i. The response was assumed to have a Tweedie distribution with a value of the p index equal to 1.83 (selected via BIC).

In terms of their biology, elephants are primarily driven by their needs for forage and water which are temporally and spatially variable, particularly in savanna systems. It is widely acknowledged that, within savanna environments, elephant movements are affected by seasonal changes (Young *et al.*, 2009a), and we found a strong seasonality effect of elephant movements. Precisely, the estimated elephant movement had a significant non-linear down trend up to approximately day 245 (approx. September, 1), after this threshold elephants seemed to increase their speed in the wet season until approximately day 30 (approx. January, 30) ($p<0.0001$). These seasonal shifts signal a response by elephants to a seasonal change in resources, e.g. forage and water. These results agree substantially with those obtained by Birkett *et al.* (2012).

The estimated spatial-temporal patterns of elephant movements in Figure 5.5 provided some indications of the broad scales changes across different years, in terms of speed of movement in certain spatial areas of the Kruger. These observed changes in spatial pattern support the hypothesis that the main choice factors in movement behaviour depend on local weather conditions, such as wet or dry season combined with Kruger-specific characteristics such as topography, and cumulative effects of abiotic and biotic factors.

It is known that African savanna systems typically experience a state of resource depletion during the dry season (Shrader *et al.*, 2006) followed by a release from these constraints when rainfall resumes (Owen-Smith, 1982). In these systems, species experience seasonal energetic bottlenecks, related to shifts in climatic variables that induce periods of restricted resources (Wiens, 1977; Owen-Smith, 1994). By examining variation in elephant speed across fine scale local rainfall, we allow the behaviour of elephants to reveal rainfall shift at a threshold approximately of 20 mm, where in general elephants increase their speed (Figure 5.3).

We found that temperature also affects significantly elephant movements up to lag 2 (Fig. 5.4). Temperature patterns suggested that elephants increase their speed when temperature gets higher, but when temperature reaches a value of approximately 30 °C, elephants decrease their speed. The temperature effect makes biological sense. Since rainfall and temperature both increase during the wet season, and both of these variables are associated with increases in biomass of forage production in savannas, elephants also move more when temperatures increase (Kinahan *et al.*, 2007). The interesting finding is that at a threshold temperature value of approximately 30 °C, elephants reduce their speed. This suggests that elephants have a thermal limit of tolerance, beyond which they have to slow their movements: this issue deserves major investigation.

Elephant movement data present several challenges in statistical modelling and data analysis due to heterogeneity, seasonal trends and non-linear effects of covariates. The proposed GAMM framework appears to provide a flexible and valuable tool to model this kind of data, making use of non-parametric uni- and multi-dimensional smooth functions combined with the Tweedie distribution for highly skewed data. While this GAMM framework was employed for elephants in Kruger National Park of South Africa, the present thesis may be applied to other herbivore species, within dynamically variable environments.

Bibliography

Anderssen, R. and Bloomfield, P. (1974). A time series approach to numerical differentiation. *Technometrics*, **16**(1), 69–75.

Archie, E. A., Morrison, T. A., Foley, C. A., Moss, C. J., and Alberts, S. C. (2006). Dominance rank relationships among wild female african elephants, *loxodonta africana*. *Animal Behaviour*, **71**(1), 117–127.

Augustin, N. H., Musio, M., von Wilpert, K., Kublin, E., Wood, S. N., and Schumacher, M. (2009). Modeling spatiotemporal forest health monitoring data. *Journal of the American Statistical Association*, **104**(487), 899–911.

Barraquand, F. and Benhamou, S. (2008). Animal movements in heterogeneous landscapes: identifying profitable places and homogeneous movement bouts. *Ecology*, **89**(12), 3336–3348.

Bartumeus, F., da Luz, M. G. E., Viswanathan, G., and Catalan, J. (2005). Animal search strategies: a quantitative random-walk analysis. *Ecology*, **86**(11), 3078–3087.

Beichelt, F. E. and Fatti, L. P. (2002). *Stochastic Processes and Their applications*. Boca Raton.

Ben-Shahar, R. (1998). Changes in structure of savanna woodlands in northern botswana following the impacts of elephants and fire. *Plant Ecology*, **136**(2), 189–189.

Bengtsson, G., Rydén, T., Öhrn, M. S., and Wiktorsson, M. (2002). Statistical analysis of the influence of conspecifics on the dispersal of a soil collembola. *Theoretical population biology*, **61**(2), 97–113.

Benhamou, S. (2007). How many animals really do the levy walk? *Ecology*, **88**(8), 1962–1969.

Birkett, P., Vanak, A., Muggeo, V., Ferreira, S., and Slotow, R. (2012). Animal perception of seasonal thresholds: Changes in elephant movement in relation to rainfall patterns. *PloS one*, **7**(6), e38363.

Blackwell, P. (1997). Random diffusion models for animal movement. *Ecological Modelling*, **100**(1-3), 87–102.

Blackwell, P. (2003). Bayesian inference for markov processes with diffusion and discrete components. *Biometrika*, **90**(3), 613–627.

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.

Bowland, J. and Yeaton, R. (1997). Impact of domesticated african elephants loxodonta africana on natal bushveld. *South African Journal of Wildlife Research*, **27**(2), 31–36.

Box, G. E. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, **65**(332), 1509–1526.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**(421), 9–25.

Calenge, C. (2011). Analysis of animal movements in r: the adehabitatlt package.

Candy, S. (2004). Modelling catch and effort data using generalised linear models, the tweedie distribution, random vessel effects and random stratum-by-year effects. *Ccamlr Science*, **11**, 59–80.

Clayton, D. G. (1996). Generalized linear mixed models. *Markov chain Monte Carlo in practice*, pages 275–301.

Codron, J., Lee-Thorp, J. A., Sponheimer, M., Codron, D., Grant, R. C., and de Ruiter, D. J. (2006). Elephant (loxodonta africana) diets in kruger national park, south africa: spatial and landscape differences. *Journal of Mammalogy*, **87**(1), 27–34.

Craven, P. and Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische Mathematik*, **31**(4), 377–403.

De Beer, Y., Kilian, W., Versfeld, W., and Van Aarde, R. J. (2006). Elephants and low rainfall alter woody vegetation in etosha national park, namibia. *Journal of Arid Environments*, **64**(3), 412–421.

Dempster, A. P., Rubin, D. B., and Tsutakawa, R. K. (1981). Estimation in covariance components models. *Journal of the American Statistical Association*, **76**(374), 341–353.

Diggle, P. J., Tawn, J., and Moyeed, R. (1998). Model-based geostatistics.

Journal of the Royal Statistical Society: Series C (Applied Statistics), **47**(3), 299–350.

Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in sobolev spaces. In *Constructive theory of functions of several variables*, pages 85–100. Springer.

Dunn, J. and Brisbin, I. (1985). Characterization of the multivariate ornstein-uhlenbeck diffusion process in the context of home range analysis. 181–205 in statistical theory and data analysis.(ed.) k. *Matusita. BV North-Holland: Elsevier*.

Dunn, J. E. and Gipson, P. S. (1977). Analysis of radio telemetry data in studies of home range. *Biometrics*, pages 85–101.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, pages 89–102.

Eilers, P. H. and Marx, B. D. (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and intelligent laboratory systems*, **66**(2), 159–174.

Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on markov random field priors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **50**(2), 201–220.

Franke, A., Caelli, T., and Hudson, R. J. (2004). Analysis of movements and behavior of caribou (¡ i¿ rangifer tarandus¡/i¿) using hidden markov models. *Ecological Modelling*, **173**(2), 259–270.

Fryxell, J. M., Hazell, M., Börger, L., Dalziel, B. D., Haydon, D. T., Morales, J. M., McIntosh, T., and Rosatte, R. C. (2008). Multiple movement modes by large herbivores at multiple spatiotemporal scales. *Proceedings of the National Academy of Sciences*, **105**(49), 19114–19119.

Gilmour, A., Anderson, R., and Rae, A. (1985). The analysis of binomial data by a generalized linear mixed model. *Biometrika*, **72**(3), 593–599.

Goldstein, H. and Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 505–513.

Gu, C. (2002). *Smoothing spline ANOVA models*. Springer.

Harris, G. M., Russell, G. J., van Aarde, R. I., and Pimm, S. L. (2008). Rules of habitat use by elephants loxodonta africana in southern africa: insights for regional management. *Oryx*, **42**(1), 66.

Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*, volume 43. Chapman & Hall/CRC.

Jonsen, I. D., Myers, R. A., and James, M. C. (2006). Robust hierarchical state–space models reveal diel variation in travel rates of migrating leatherback turtles. *Journal of Animal Ecology*, **75**(5), 1046–1057.

Jorgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 127–162.

Kammann, E. and Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **52**(1), 1–18.

Kareiva, P. and Shigesada, N. (1983). Analyzing insect movement as a correlated random walk. *Oecologia*, **56**(2-3), 234–238.

Kareiva, P. and Wennergren, U. (1995). Connecting landscape patterns to ecosystem and population processes. *Nature*, **373**(6512), 299–302.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, **90**(430), 773–795.

Kimeldorf, G. and Wahba, G. (1971). Some results on tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, **33**(1), 82–95.

Kinahan, A., Pimm, S. L., and Van Aarde, R. J. (2007). Ambient temperature as a determinant of landscape use in the savanna elephant, *loxodonta africana*. *Journal of Thermal Biology*, **32**(1), 47–58.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963–974.

Lang, S. and Brezger, A. (2004). Bayesian p-splines. *Journal of computational and graphical statistics*, **13**(1), 183–212.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 619–678.

Leggett, K. (2006). Effect of artificial water points on the movement and behaviour of desert-dwelling elephants of north-western namibia. *IUCN*, page 23.

Leuthold, W. (1977). Spatial organization and strategy of habitat utilization of elephants in tsavo national park, kenya. *Zeitschrift fur Saugetierkunde*, **42**, 358–379.

Lin, X. and Breslow, N. E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association*, **91**(435), 1007–1016.

Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **61**(2), 381–400.

Loarie, S., van Aarde, R., and Pimm, S. (2009). Elephant seasonal vegetation preferences across dry and wet savannas. *Biological Conservation*, **142**(12), 3099–3107.

MacCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*, volume 37. CRC press.

Marx, B. D. and Eilers, P. H. (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics & Data Analysis*, **28**(2), 193–209.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American statistical Association*, **92**(437), 162–170.

McCulloch, C. E. (2006). *Generalized linear mixed models*. Wiley Online Library.

Muggeo, V. M. (2008). Segmented: an r package to fit regression models with broken-line relationships. *R news*, **8**(1), 20–25.

Owen-Smith, N. (1982). Factors influencing the consumption of plant products by large herbivores. In *Ecology of tropical savannas*, pages 359–404. Springer.

Owen-Smith, N. (1994). Foraging responses of kudus to seasonal changes in food resources: elasticity in constraints. *Ecology*, pages 1050–1062.

Owen-Smith, R. N. (1988). Megaherbivores, the influence of very large body size on ecology. *Cambridge studies in ecology.*

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J. (2008). State–space models of individual animal movement. *Trends in ecology & evolution*, **23**(2), 87–94.

Pinheiro, J. and Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. Springer Verlag.

Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D. (2011). R development core team (2011) nlme: linear and nonlinear mixed effects models. r package version 3.1–98. *R Foundation for Statistical Computing, Vienna*.

Polansky, L. and Wittemyer, G. (2011). A framework for understanding the architecture of collective movements using pairwise analyses of animal movement data. *Journal of The Royal Society Interface*, **8**(56), 322–333.

Preisler, H. K. and Akers, R. P. (1995). Autoregressive-type models for the analysis of bark beetle tracks. *Biometrics*, pages 259–267.

Preisler, H. K., Brillinger, D. R., Ager, A. A., Kie, J. G., and Akers, R. P. (2001). Stochastic differential equations: a tool for studying animal movement. In *Proceedings of IUFRO4. 11 Conference:'Forest Biometry, Modelling and Information Science', University of Greenwich.*

Preisler, H. K., Ager, A. A., Johnson, B. K., and Kie, J. G. (2004). Modeling animal movements using stochastic differential equations. *Environmetrics*, **15**(7), 643–657.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ribeiro Jr, P. J. and Diggle, P. J. (2001). geor: A package for geostatistical analysis. *R news*, **1**(2), 14–18.

Robert, C. P., Casella, G., and Robert, C. P. (1999). *Monte Carlo statistical methods*, volume 58. Springer New York.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression*, volume 12. Cambridge University Press.

Sankaran, M., Hanan, N. P., Scholes, R. J., Ratnam, J., Augustine, D. J., Cade, B. S., Gignoux, J., Higgins, S. I., Le Roux, X., Ludwig, F., *et al.* (2005). Determinants of woody cover in african savannas. *Nature*, **438**(7069), 846–849.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, **78**(4), 719–727.

Schick, R. S., Loarie, S. R., Colchero, F., Best, B. D., Boustany, A., Conde, D. A., Halpin, P. N., Joppa, L. N., McClellan, C. M., and Clark, J. S. (2008). Understanding movement data and movement processes: current and emerging directions. *Ecology Letters*, **11**(12), 1338–1350.

Scholes, R. J., Bond, W. J., and Eckhardt, H. C. (2003). Vegetation dynamics in the kruger ecosystem. *The Kruger experience (J. DuToit, K.*

Rogers, and H. Biggs, eds.). Island Press, Washington, DC, pages 242–262.

Shrader, A., OWEN-SMITH, N., and Ogutu, J. (2006). How a megagrazer copes with the dry season: food and nutrient intake rates by white rhinoceros in the wild. *Functional ecology*, **20**(2), 376–384.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–52.

Steele, B. M. (1996). A modified em algorithm for estimation in generalized mixed models. *Biometrics*, pages 1295–1310.

Turchin, P. (1998). *Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants*. Sinauer Associates Sunderland.

Van Beest, F. M., Vander Wal, E., Stronen, A. V., and Brook, R. K. (2013). Factors driving variation in movement rate and seasonality of sympatric ungulates. *Journal of Mammalogy*.

Vanak, A. T., Thaker, M., and Slotow, R. (2010). Do fences create an edge-effect on the movement patterns of a highly mobile mega-herbivore? *Biological Conservation*, **143**(11), 2631–2637.

Venter, F. J., Scholes, R. J., and Eckhardt, H. (2003). The abiotic template and its associated vegetation pattern. *The Kruger experience: Ecology and management of savanna heterogeneity*, pages 83–129.

Viswanathan, G., Buldyrev, S. V., Havlin, S., Da Luz, M., Raposo, E., and

Stanley, H. E. (1999). Optimizing the success of random searches. *Nature*, **401**(6756), 911–914.

Wahba, G. (1975). Smoothing noisy data with spline functions. *Numerische Mathematik*, **24**(5), 383–393.

Wahba, G. (1980). Spline bases, regularization, and generalized cross-validation for solving approximation problems with large quantities of noisy data. *Approximation theory III*, **2**.

Wahba, G. (1983). Bayesian" confidence intervals" for the cross-validated smoothing spline. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 133–150.

Wahba, G. (1990). *Spline models for observational data*. Number 59. Siam.

Western, D. and Lindsay, W. (1984). Seasonal herd dynamics of a savanna elephant population. *African Journal of Ecology*, **22**(4), 229–244.

Wiens, J. A. (1977). On competition and variable environments: Populations may experience" ecological crunches" in variable climates, nullifying the assumptions of competition theory and limiting the usefulness of short-term studies of population patterns. *American Scientist*, **65**(5), 590–597.

Williams, R., Hedley, S. L., Branch, T. A., Bravington, M. V., Zerbini, A. N., and Findlay, K. P. (2011). Chilean blue whales as a case study to illustrate methods to estimate abundance and evaluate conservation status of rare species. *Conservation Biology*, **25**(3), 526–535.

Wittemyer, G. and Getz, W. (2007). Hierarchical dominance structure and social organization in african elephants, *loxodonta africana*. *Animal Behaviour*, **73**(4), 671–681.

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.

Wood, S. (2006a). *Generalized additive models: an introduction with R*, volume 66. Chapman & Hall/CRC.

Wood, S. (2011a). mgcv package for r, gams with gcv/aic/reml smoothness estimation and gamms by pql, version 1.7-6./h ttp. *cran. r-project. org/S*.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**(1), 95–114.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, **99**(467).

Wood, S. N. (2006b). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, **62**(4), 1025–1036.

Wood, S. N. (2006c). On confidence intervals for generalized additive models based on penalized regression splines. *Australian & New Zealand Journal of Statistics*, **48**(4), 445–464.

Wood, S. N. (2008). Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **70**(3), 495–518.

Wood, S. N. (2011b). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73**(1), 3–36.

Wood, S. N. (2013). On p-values for smooth components of an extended generalized additive model. *Biometrika*, **100**(1), 221–228.

Young, K., Ferreira, S., and Van Aarde, R. (2009a). Elephant spatial use in wet and dry savannas of southern africa. *Journal of Zoology*, **278**(3), 189–205.

Young, K. D., Ferreira, S. M., AARDE, V., and RUDOLPH, J. (2009b). The influence of increasing population size and vegetation productivity on elephant distribution in the kruger national park. *Austral Ecology*, **34**(3), 329–342.