

Universität Bielefeld

Fakultät für Wirtschaftswissenschaften

Masterarbeit

im Studiengang Wirtschaftswissenschaften

zum Thema:

Statistical modeling of tracking data – an empirical comparison of methods

vorgelegt von

Timo Meyer

Matrikel-Nr: 2729188

Anschrift: Nagelholz 11, 32139 Spenze

ausgeführt zum Zwecke der Erlangung des akademischen Grades Master der
Wirtschaftswissenschaften (M.Sc.)

1. Prüfer/in Prof. Dr. Roland Langrock
(Lehrstuhl Statistik und Datenanalyse)
2. Prüfer/in Prof. Dr. Dietmar Bauer
(Lehrstuhl Ökonometrie)

Bielefeld, im Mai 2022

Abstract

Investigating the movement behaviour of wild animals driven by physiological requirements and environmental factors is of growing interest to the ecological community. Due to modern GPS technology and increasing computing capacities, various statistical methods allow researchers to extensively infer mechanisms, patterns and latent state dynamics in animal movement. By using recently described time-varying stochastic differential equations, hidden Markov models, and generalised additive mixed models, savanna elephants' irregular spaced movement data will be analysed to empirically compare and evaluate the properties and suitability of these methods for tracking data. So far, the data of the savanna elephants have been analysed- neglecting the irregularity. Therefore, the data will be imputed using continuous-time correlated random walks, prior to the analysis. The present study shows that elephants' movement is highly influenced by temperature. Moreover, it also indicates that time of day is a significant behavioural factor and reveals a cyclic 24-hour movement pattern. Whereas generalised additive mixed models are rather an additional tool, hidden Markov models can categorise the cyclical movement pattern and response to temperature into three different ecological states: resting, foraging, and traveling. Although the three latent states are merely proxies for their true ecological state dynamics, these reveal different activity patterns and provide additional information about elephants' assumed shuttling behaviour, making hidden Markov models a favourable choice in movement ecology. Although there have been computational issues concerning the maximisation of the log-likelihood, time-varying stochastic differential equations uncover the influence of temperature and time of day in a more mechanistic sense. A combination of time-varying stochastic differential equations and hidden Markov models might advance inference of ecological state dynamics in a more technical manner.

Acknowledgements

To my family, and Madeleine.

Contents

List of Figures	iv
List of Tables	v
Nomenclature	vi
Acronyms	viii
1 Introduction	1
2 The Data	4
2.1 Exploratory Data Analysis	4
2.2 Continuous-Time Correlated Random Walk	7
3 Methods	14
3.1 Generalised Additive Mixed Models	14
3.2 Hidden Markov Models	21
3.3 Stochastic Differential Equations	29
4 Results	35
4.1 Generalised Additive Mixed Models	35
4.2 Hidden Markov Models	45
4.3 Stochastic Differential Equations	54
5 Summary and Discussion	62
6 Conclusion	67
Bibliography	73
Appendix	74

A1 R-Output - Generalised Additive Models	74
A2 Dimension checking for generalised additive models	78
A3 R-Output - Hidden Markov Models and animated state probabilities	79
A4 R-Output - Stochastic Differential Equations	81

List of Figures

2.1	Observations per animal	5
2.2	Time intervals between two consecutive observations	5
2.3	Actual distance traveled vs. calculated distance traveled	6
2.4	Ornstein-Uhlenbeck process - simulated velocity	9
2.5	AM105 imputed path	12
2.6	AM107 modelled step lengths using CTCRW	13
3.1	P-spline fitted regression function	19
3.2	3-state Markov process	22
3.3	Basic HMM structure	25
3.4	2-state HMM process	26
3.5	Estimated parameters for time-varying CTCRW	33
4.1	Fitted GAMM	37
4.2	ACF-plot GAMM	39
4.3	GAMMs using cyclic cubic regression splines	41
4.4	GAMMs using cyclic cubic regression splines - ACF-plots	42
4.5	GAMMs normal QQ-plots	43
4.6	GAMM - residuals	44
4.7	State-dependent densities for step lengths	49
4.8	State-dependent densities for turning angles	49
4.9	Stationary state probabilities by hour in wet season	50
4.10	Stationary state probabilities by hour in dry season	50
4.11	Stationary state probabilities by temperature in wet season	51
4.12	Stationary state probabilities by temperature in dry season	52
4.13	QQ- and ACF-plots of step lengths and turning angles	53
4.14	SDE parameters for AM254 by temperature	55
4.15	Fitted SDE AM254 by temperature	56

4.16	AM239 and AM253 SDE output by time of day	57
4.17	Fitted time-varying Brownian motion by time of day	59
4.18	QQ-plot for fitted time-varying Brownian motion model	60
4.19	ACF-plot for fitted time-varying Brownian motion model	61

List of Tables

4.1	Model selection - choice of k	36
4.2	Mean speed by season - differences	38
4.3	GAMM (Gaussian) - model selection by AIC	41
4.4	Model selection - number of states	47
4.5	Model selection - covariates in transition probabilities	48
4.6	SDE model selection - basis functions	58
4.7	Brownian motion - model selection by AIC	59

Nomenclature

Latin

a	state vector in state space models
b	random effects vector
$d_\Delta(t)$	movement of an animal over Δ time units
G	transformation matrix
$g(\cdot)$	smoothing function
H	variance matrix of measurement error in state space models
h	known, monotonic, twice differentiable link function
I	identity matrix
L	likelihood
l	log-likelihood
m	number of states in a Markov chain
O	O notation, upper bound on the growth rate of a function
P	matrix of (state-dependent) probability functions
p	probability mass or probability density function
Q	variance covariance matrix for error vector Ω
q	location quality covariate (e.g. ARGOS location data)
r	drift parameter
s	variability in velocity
U	(quadratic) unity matrix
u	probability of being in a certain state
$V(\cdot)$	variance function
$v(t)$	instantaneous velocity
W	matrix containing weights w_{ii}
X, x	regressor
Y, y	response variable

Z	model matrix for random effects
z	pseudodata

Greek

α	forward probabilities
β	coefficients for basis function
Γ	transition probability matrix
γ	transition probabilities
δ	stationary distribution
ε	measurement error
ζ	normal distributed random variable, second component of Ω
η	predictor
θ	time-varying parameter
ϑ	normal distributed error vector, first component of Ω
ι	influence term
κ	coordinates in discrete-time, location process in continuous-time
Λ	positive definite matrix to model correlation structures
λ	smoothing parameter
$\mu(\cdot)$	expected value; drift function of a stochastic differential equation
ν	estimated velocity
ξ	coefficients for fixed effects
ρ	(autocorrelation) parameter for stochastic differential equations
ϱ	concentration parameter in a von Mises distribution
$\sigma(\cdot)$	diffusion function in stochastic differential equations
τ	effective degrees of freedom
$\phi(\cdot)$	probability density function of a normal distribution
φ	backward probabilities
Ψ	variance covariance matrix for random effects
ψ	basis functions for smoothing function $g(\cdot)$
Ω	normal distributed error vector
ω	mean velocity in a velocity process

Acronyms

ACF	autocorrelation function
AIC	Akaike's information criterion
BIC	Bayesian information criterion
CTCRW	continuous-time correlated random walk
EDA	exploratory data analysis
EDF	effective degrees of freedom
EM	expectation-maximisation
GAM	generalised additive model
GAMM	generalised additive mixed model
GCV	generalised cross validation
HMM	hidden Markov model
LMM	linear mixed model
MCMC	Markov chain Monte Carlo
PIRLS	penalised iterative re-weighted least squares
PDF	probability density function
QQ-plot	quantile-quantile-plot
REML	restricted maximum likelihood estimation
SDE	stochastic differential equation
SSM	state-space model

1 Introduction

Investigating the movement behaviour of wild animals driven by physiological requirements and environmental factors is of growing interest to the ecological community. Increasing capabilities of GPS technology sensors provide movement data over more extended time periods and at a higher frequency (e.g. half-hourly data for several months or years). A recent study by Thaker et al. (2019) analysed the movement behaviour of 14 female savanna elephants (*Loxodonta africana*) from different herds, tracked via GPS with implemented temperature sensors, over a two-year period from 2007 to 2009 in Kruger National Park, South Africa. Using regression methods, such as linear mixed models (LMMs) and generalised additive mixed models (GAMMs), they found that elephants move faster at hotter temperatures, especially in the wet season. Moreover, the individuals showed a periodic shuttling pattern towards water sources (Thaker et al., 2019).

However, it is not clear if regression methods are an appropriate choice in analysing tracking data. In particular, Thaker et al. (2019) neglected the fact that the dataset is irregularly spaced. Since a relationship between the scale-dependent variable speed and environmental covariates was assumed, the irregular time intervals will most likely cause problems in statistical inference. Moreover, tracking data are highly correlated, which has not been considered in model formulation yet. Thus, the results might be biased and need to be reconsidered to gain further insights into the dynamics and drives of the movement behaviour of savanna elephants.

Given this circumstance, three different statistical methods will be compared in order to investigate the relationship between animal movement and their environment: GAMMs, hidden Markov models (HMMs) and a recently described extension to stochastic differential equations (SDEs). Due to high complexity and dynamic interdependencies, movement data is challenging to analyse and hence the aim is to find out which of the three methods is the most effective approach for this type of data. Prior to analysing the animal movement, the irregular

spaced dataset will be imputed using continuous-time correlated random walks (CTCRWs). The three methods will then be empirically compared based on suitability, comprehensibility, computational efficiency and output.

Methodologically, Thaker et al. (2019) analysed the behaviour of the respective elephants in the context of physiological requirements and revealed a shuttling pattern of the individuals towards water sources. Furthermore, concerning local and regional rainfall patterns, Birkett et al. (2012) analysed seasonal shifts in the movement behaviour of the 14 tracked elephants using piece-wise linear regression models (PRMs) and generalised linear models (GLMs). However, they also did not account for autocorrelation. Due to the irregular spaced observation times, they only used certain sequences of the movement paths, whereas Thaker et al. (2019) even fitted models with irregular data.

In contrast, different approaches than regression models gained popularity within the rapidly growing field of movement ecology. Derived from the nature of tracking data, most methods are time series tools. An overview of critical methods can be found in Patterson et al. (2017); however, it is impossible to discuss the wide variety of methods to analyse tracking data. Thus, the most common approaches will be briefly presented.

A prevalent approach is the Lévy walk proposed by Viswanathan et al. (1999). However, it is too simplistic to model critical key patterns of interest in the data because it does not account for autocorrelation. Therefore, Jonsen et al. (2003) proposed state-space models (SSMs) to analyse movement data. SSMs are doubly stochastic processes capturing movement dynamics using continuous-valued, unobserved states and can account for uncertainty in the location. HMMs can be used as a special case of SSMs with finite states, if the location error can be neglected. For example, Patterson et al. (2009) linked the movement behaviour of animals to certain activity modes using HMMs, whereas DeRuiter et al. (2017) investigated blue whale activity during sonar exposure. Besides the state-space approach, correlated random walks and SDEs gained popularity, as they can handle irregular spaced data. Morales et al. (2004) described a method to use mixtures of correlated random walks to further investigate movement behaviour. Moreover, Blackwell (1997) uses random diffusion models for the movement of animals switching between behavioural states. There are numerous extensions and applications in movement ecology based on these methods. As a more recent approach, Michelot et al. (2021) described time-varying SDEs, where the parameters

are functions of covariates and modelled via regression splines.

Due to this vast amount of methods, the three key concepts will be empirically compared; GAMMs, HMMs and (time-varying) SDEs. In the following, the movement behaviour of the savanna elephants' data will be investigated in chapter 2. Moreover, using CTCRWs, the data will be imputed to properly fit those methods to the data, requiring equally spaced observations. The theoretical concepts of the methods will then be described in chapter 3. In chapter 4, the methods will be applied to the data. Conclusively in chapter 5, the results will be discussed, whereas the methods will be compared in order to draw conclusions about what approach should be used when dealing with movement data.

2 The Data

This chapter is intended to provide an insight into the savanna elephants' data. In the following section, using exploratory data analysis (EDA), key aspects of the data will be presented. Then, based on EDA, further considerations are made in the context of the methods to be used.

2.1 Exploratory Data Analysis

The dataset includes 14 female elephants from 14 different herds. The elephants were tracked over a two-year period from 2007 to 2009, in Kruger National Park, South Africa (Slotow et al., 2019), by using GPS collars with integrated temperature sensors. GPS temperature sensors were intended to measure the ambient temperature to examine movement patterns. In addition, a timestamp including year, month, day, hour, and minute has been tracked for every observation.

The dataset contains about 280.000 observations, although they are not uniformly distributed among all animals. Differences in the number of observations result from the fact that some animals were tracked in shorter time periods, e.g. from 2008 to 2009, and therefore have fewer observations. Figure 2.1 shows the number of observations per animal. Each animal has its own ID, reaching from AM91 to AM308. As can be seen, IDs AM91 and AM99 have three times as many observations as IDs AM107 and 306. However, the absolute number of observations per animal is sufficient for the methods applied here and, thus, is not a limitation.

Another reason for varying observations per ID is justified in irregularly spaced observations. Most time intervals between two consecutive observations are about half an hour. However, the gap between two observations can be much larger. Even weeks or months can lie in between two consecutive observations. Figure 2.2 represents a distribution of different time intervals in minutes. About 6.000 consecutive observations result in time intervals longer than 150 minutes.

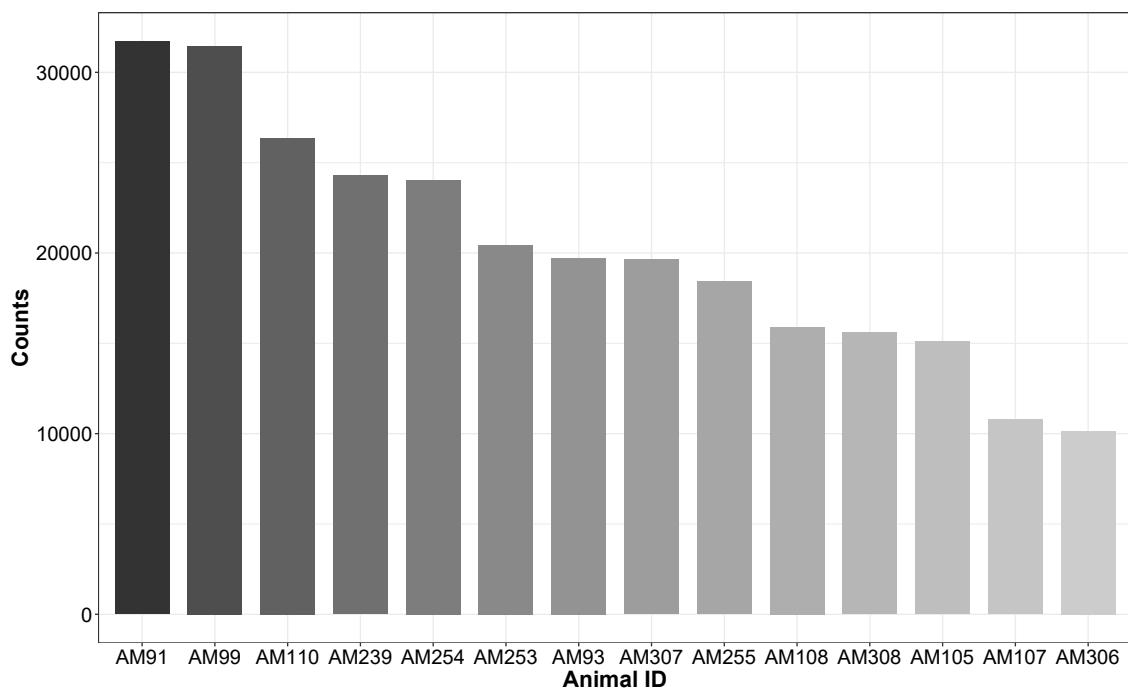


Figure 2.1: Observations per animal.

Source: own representation.

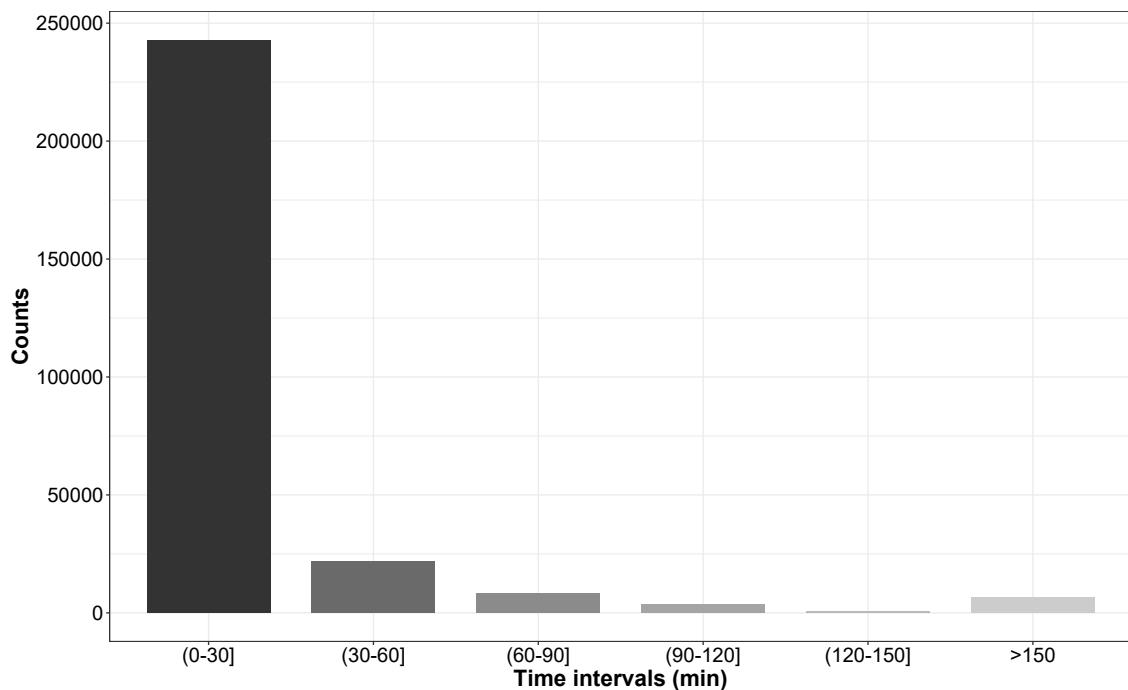


Figure 2.2: Time intervals between two consecutive observations.

Source: own representation.

This is, in fact, a drawback in the work of Thaker et al. (2019) since they did not account for this irregular spacing when calculating the speed of the animals, which depends on two consecutive observations. Ignoring this issue leads to biased results when modelling the speed of the animals using models that require equally spaced observations. The bias is illustrated in Figure 2.3. Starting from time t , the animal travels the distance marked in blue until time $t + 1$. The red dotted line indicates the distance between the two consecutive observations. As can be seen, the indicated distance is much smaller than the travelled distance. Speed is calculated as distance divided by time. If the time interval between two observations increases, the calculated speed decreases. As a result, the dependent variable, speed, is biased.

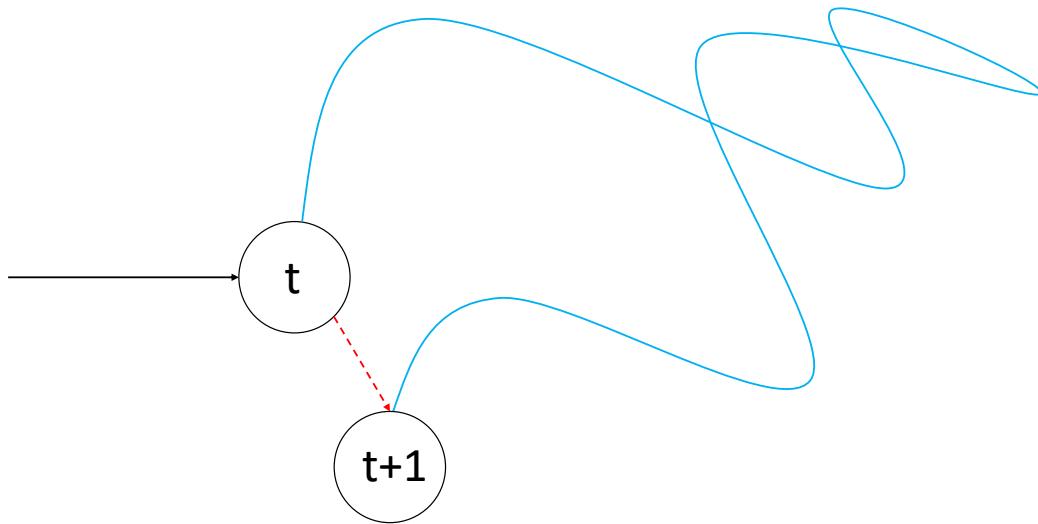


Figure 2.3: Actual distance traveled vs. calculated distance traveled.
Source: own representation.

When using the methods shown in Thaker et al. (2019), such as generalised additive models (GAMs), GAMMs, or LMMs and ignoring the irregularity in the data, this will result in biased estimators because these methods require equally spaced observations. The next section introduces a method to address this bias.

2.2 Continuous-Time Correlated Random Walk

The dataset must be imputed to obtain regularly spaced observations. A flexible method to accomplish this would be a continuous-time correlated random walk (CTCRW). A considerable advantage of a CTCRW is that it allows the modelling of unevenly sampled data without subsampling or interpolation techniques. Moreover, an Ornstein-Uhlenbeck process¹ can account for autocorrelation in that animals are assumed to have a certain kind of inertia, allowing for similar velocities for consecutive times. This is a fundamental feature because, statistically, most tracking data are highly spatially and temporally correlated, which must be considered in model formulation (Patterson et al., 2017). Since a correlated process does not depend only on the previous observation (which would then be called Markovian) and thus parameter estimation could be difficult, the model is formulated in a state-space framework. A state-space framework then allows the application of the Kalman filter to estimate the parameters via maximum likelihood, with predictions for unobserved location data as a byproduct (Johnson et al., 2008).

To derive the CTCRW, let $\kappa(t) = (\kappa_1(t), \kappa_2(t))'$ be the coordinates (longitude and latitude) of an animal location at time t . Calculating the difference $d_\Delta(t) = \kappa(t + \Delta) - \kappa(t)$ then describes the movement of an animal over Δ time units. The next step is to formulate this movement as a continuous-time process. If $\Delta \rightarrow 0$ and $\kappa(t)$ is a smooth and continuous path, then one obtains the differential equation

$$d\kappa(t) = v(t)dt, \quad (2.1)$$

with $v(t)$ representing the instantaneous velocity (Johnson et al., 2008).

In order to model this instantaneous velocity in continuous-time, an Ornstein-Uhlenbeck process will be used. In particular, for each coordinate axis $c = 1, 2$, the instantaneous velocity, $v_c(t)$, for each time unit Δ , is defined as the autoregressive equation

$$v_c(t + \Delta) = \omega_c + e^{-\rho\Delta}(v_c(t) - \omega_c) + \zeta_c(\Delta). \quad (2.2)$$

¹which is why the CTCRW is sometimes called velocity Ornstein-Uhlenbeck model

In (2.2), ω_c is the mean velocity (drift) rate, ρ describes an autocorrelation parameter, and $\zeta_c(\Delta)$ represents a normal random variable with distribution

$$\zeta_c(\Delta) \sim N\left(0, s^2 \frac{(1 - e^{-2\rho\Delta})}{2\rho}\right), \quad (2.3)$$

with s controlling the overall variability in velocity.² Equation (2.2) states that $v_c(t + \Delta)$ at time $t + \Delta$ equals the mean velocity ω_c , an adjustment for the difference between the velocity at time t and the mean velocity, plus a random variable whose variance increases with Δ , which is stated in Equation (2.3). The derived process can be used to model the velocity of a moving animal. Figure 2.4 shows one hundred simulated Ornstein-Uhlenbeck processes as an example of animal velocity. The Figure illustrates the so-called mean-reverting tendency because every process reverts to its mean velocity. However, most of the time, the centre of attraction, ω_c , will be set to zero to indicate that there is no systematic bias in the velocity (Michelot et al., 2021).

The bivariate velocity process $\mathbf{v}(t)$ must be integrated up to time t and added to the initial location $\boldsymbol{\kappa}(0)$, defining the continuous-time location process to obtain the location at time t in the continuous-time framework:

$$\boldsymbol{\kappa}(t) = \boldsymbol{\kappa}(0) + \int_0^t \mathbf{v}(u) du. \quad (2.4)$$

Thus, the combination of Equations (2.2) and (2.4) provides the CTCRW model. In principal, Equation (2.2) is an Ornstein-Uhlenbeck process that is supposed to capture autocorrelation in movement speed and direction, whereas the location process defined in (2.4) is just the integral of (2.2).

However, as mentioned earlier, since the location process relies on consecutive, correlated velocities (i.e., the process is not Markovian), estimating the parameters tends to be challenging. Therefore, the CTCRW will be set into a state-space framework (Johnson et al., 2008). The general framework of a Gaussian linear state-space model for a univariate observation comprises two equations, the ob-

²A strict derivation of the velocity process can be found in Johnson et al. (2008). In this work, Equation (2.2) already shows the derived velocity process.

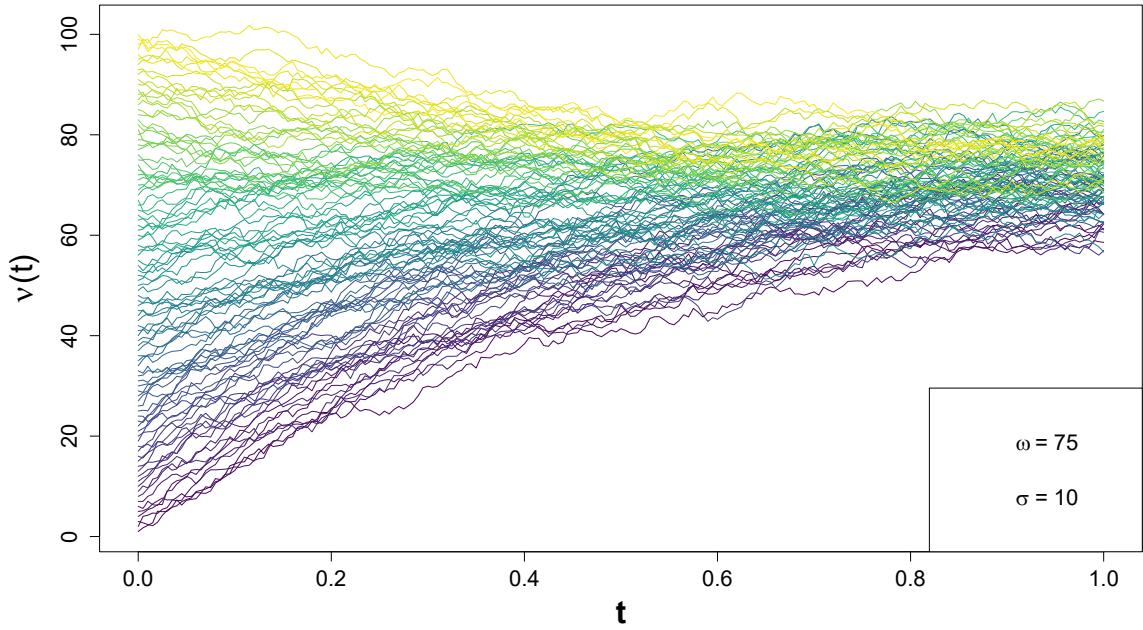


Figure 2.4: Ornstein-Uhlenbeck process - simulated velocity.

Source: own representation.

servation equation and the state equation:

$$y_i = \mathbf{G}'_i \mathbf{a}_i + \varepsilon_i \quad (2.5)$$

$$\mathbf{a}_{i+1} = \mathbf{K}_i \mathbf{a}_i + \boldsymbol{\Omega}_i. \quad (2.6)$$

In Equation (2.5), \mathbf{a}_i describes the current state vector, ε_i is a normal measurement error with variance H_i , \mathbf{G}_i is an appropriately sized transformation matrix for the state vector, and y_i is an observation at time i . In Equation (2.6), \mathbf{K}_i also is an appropriately sized transformation matrix for the state vector. Additionally, $\boldsymbol{\Omega}_i$ represents normal error vectors with variance-covariance matrix \mathbf{Q}_i (Johnson et al., 2008).

In the following, the CTCRW is reformulated into Equations (2.5) and (2.6) to obtain a state-space model version. Assuming that the locations $\mathbf{y}_i = (y_{1i}, y_{2i})'$ are observed at times t_1, \dots, t_n and, conditioning on the true location $\boldsymbol{\kappa}(t) = (\kappa_1(t), \kappa_2(t))'$, this will yield the observation equation for animal

movement:

$$y_{ci} = \kappa_{ci} + \varepsilon_{ci}, \quad \varepsilon \sim N(0, H_{ci}). \quad (2.7)$$

The variance H_{ci} of the error term ε_{ci} in Equation (2.7) could depend on external location quality covariates (for GPS data, the measurement error is often sufficiently small to be ignored).

The true location equation (corresponding to the state-space model's state equation) is somewhat more difficult to formulate since $\kappa(t)$ is not Markovian. However, regarding the velocity process described in (2.2) and the formulation of the location process given in Equation (2.4), the state a_i can be obtained by bundling the velocity process to the location process into a single state vector since the velocity process itself is Markovian (Johnson et al., 2008). Therefore, this yields the state equation

$$\kappa_{c,i+1} = \kappa_{ci} + v_{ci} \left(\frac{1 - e^{-\rho\Delta_i}}{\rho} \right) + \vartheta_{ci}. \quad (2.8)$$

As before, Δ_i is the difference between two consecutive times, and ϑ_{ci} represents normal error vectors (this is the first entry of Ω_i in the general formulation) with variance

$$\text{Var}(\vartheta_{ci}) = \frac{s^2}{\rho^2} \left(\Delta_i - \frac{2}{\rho}(1 - e^{-\rho\Delta_i}) + \frac{1}{2\rho}(1 - e^{-2\rho\Delta_i}) \right). \quad (2.9)$$

Furthermore, with covariance

$$\text{Cov}(\zeta_{ci}, \vartheta_{ci}) = \frac{s^2}{2\rho^2} \left(1 - 2e^{-\rho\Delta_i} + e^{-2\rho\Delta_i} \right) \quad (2.10)$$

between ζ_{ci} and ϑ_{ci} (consequently, ζ_{ci} is the second entry of Ω_i in the general formulation), the variance-covariance matrix of Ω_{ci} can be specified as

$$\mathbf{Q}_{ci} = \begin{pmatrix} \text{Var}(\vartheta_{ci}) & \text{Cov}(\zeta_{ci}, \vartheta_{ci}) \\ \text{Cov}(\zeta_{ci}, \vartheta_{ci}) & \text{Var}(\zeta_{ci}) \end{pmatrix}. \quad (2.11)$$

As a last step, defining $\mathbf{G}_i = (1 \ 0)'$, $\mathbf{a}_i = (\kappa_{ci} \ v_{ci})'$, $\Omega_{ci} = (\vartheta_{ci} \ \zeta_{ci})'$ and $H_{ci} = H(q_i)$,

where q_i is a known location quality covariate, and in combination with the matrix

$$\mathbf{K}_i = \begin{pmatrix} 1 & (1 - e^{-\rho\Delta_i})/\rho \\ 0 & e^{-\rho\Delta_i} \end{pmatrix}, \quad (2.12)$$

all components of a basic state-space formulation are specified. In particular, with Equations (2.2), (2.7), and (2.8), one obtains a CTCRW model in a state-space framework (Johnson et al., 2008).

This reformulation of the CTCRW allows the use of a Kalman filter to estimate the movement parameters $\hat{\theta} = (\hat{\rho}_1, \hat{\rho}_2, \hat{s}_1, \hat{s}_2)'$ via maximum likelihood. As mentioned earlier, the byproducts of estimation using a Kalman filter are (partly) unobserved, equally spaced locations $\hat{x}(t)$ (and velocities $\hat{v}(t)$) (Johnson et al., 2008).

Using the CTCRW, the dataset at hand has been imputed. The selected time interval is one hour, as this ensures sufficient granularity and, at the same time, a reasonable prediction quality of the CTCRW. Figure 2.5 graphically represents the concept of imputed hourly paths exemplified by individual AM105. The black dots represent the observed locations, and the blue coloured path corresponds to the predicted movement path, which is obtained by the CTCRW. For further modelling, the variables 'speed', 'step length' and 'turning angle' can be computed based on the regular movement path. The CTCRW was implemented with the statistical software R using the functions `crwFit` and `crwPredict` from the `crawl` package (Johnson and London, 2018). The step length and turning angle computations were accomplished using the function `prepData` via the package `moveHMM` (Michelot et al., 2016). Step lengths are defined as the distance between two consecutive observations, whereas turning angles correspond to the direction changes between two consecutive observations (Patterson et al., 2017). The computation time took only a few minutes on a 4.9 GHz Intel Core i7 9700k and 16GB DDR4 RAM.

However, as mentioned before, there are also time intervals of several months between two consecutive observations. This temporal discrepancy could not be modelled satisfactorily by a CTCRW. Temporal discrepancies that are too extensive to model do not affect all time series, but only a few animals. Animal AM107 exemplifies this. Figure 2.6 represents a plot of the imputed step lengths of AM107 depending on the tracking time.

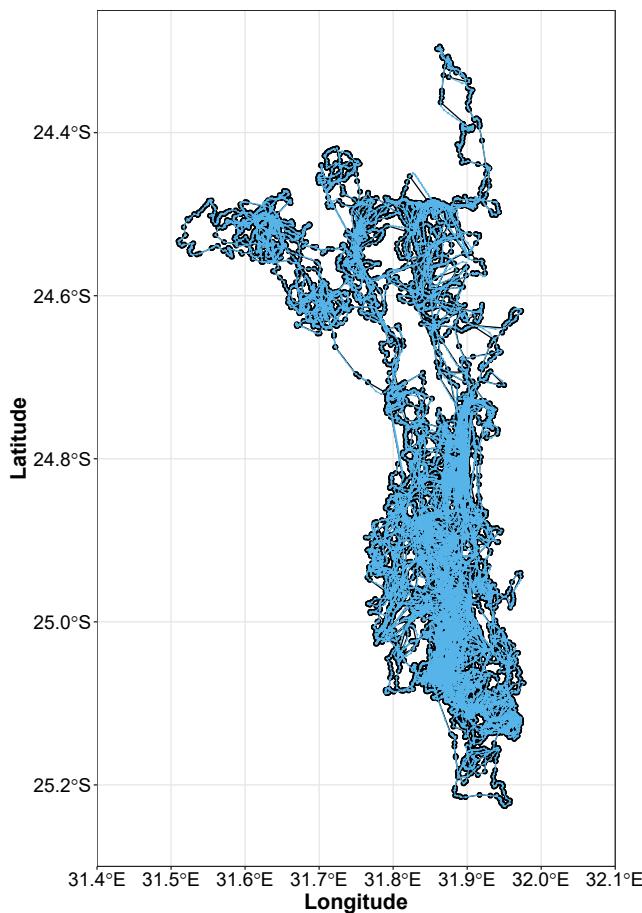


Figure 2.5: AM105 imputed path.
Source: own representation.

The diagram shows time intervals with a seemingly constant step length over weeks or even months. The (almost) constant step length results from the fact that the CTCRW can not model hourly time intervals satisfactorily because the temporal discrepancy between two consecutive observations in time is too large and results in infinitesimal changes in the hourly step length. As mentioned, this is the case for just a few animals. Therefore, instead of completely ignoring the animals' movement paths, sequences are used where there are no weeks or months between two consecutive observations. For example, the time series of AM107 was truncated to August 2007 to February 2008.

A CTCRW thus provides a new dataset with additional, useful information and movement paths of all 14 animals, forming the basis for the methods that will

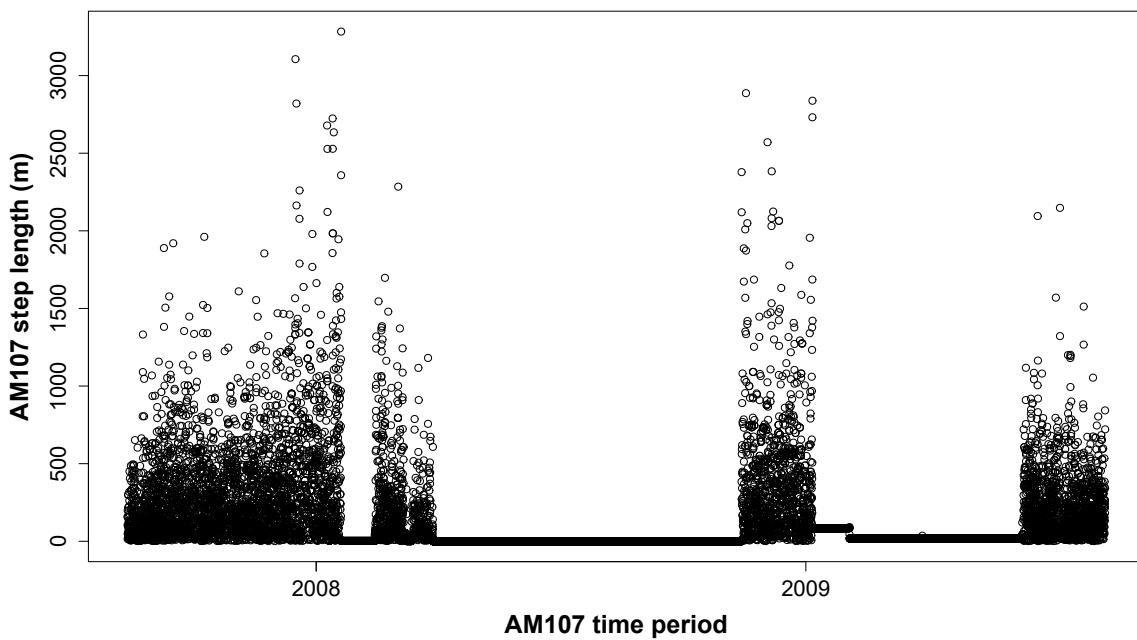


Figure 2.6: AM107 modelled step lengths using CTCRW.

Source: own representation.

be applied. In particular, the variables 'speed', 'step length' and 'turning angle' can now be modelled using methods that require equally spaced data in to make statistical inferences about the movement pattern of elephants. In particular, the goal is to further investigate the thesis by Thaker et al. (2019) that elephants show a shuttling behaviour between water sources and analyse the movement pattern of the individuals affected by environmental factors and physical requirements. Moreover, this can be considered a case study to compare appropriate methods to analyse tracking data empirically. In the following, the theoretical background of the methods that will be used will be introduced.

3 Methods

This chapter is intended to introduce the methods that will be applied in the next chapter. In particular, based on mixed models and an additive extension of the family of generalised linear models, called generalised additive models, generalised additive mixed models are described. The last two sections of this chapter introduce two concepts considered flexible tools for handling data collected over time: hidden Markov models and stochastic differential equations. Hidden Markov models will be covered first. Then, stochastic differential equations are slightly extended to a time-varying version of ordinary stochastic differential equations.

3.1 Generalised Additive Mixed Models

With the CTCRW model from the previous section, the data are now available at regular intervals expanding the possibilities of methods to be used. Over the past three decades, numerous extensions of the standard linear regression model have been developed. In the work of Thaker et al. (2019), LMMs, GAMs, and GAMMs were used. This section discusses mixed models as one component of GAMMs in general and GAMs as the basis model component. The two approaches will then be combined to formulate a GAMM.

Mixed Models

In their simplest form, mixed models are an extension of the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}\sigma^2) \tag{3.1}$$

to

$$\mathbf{y} = \mathbf{X}\boldsymbol{\xi} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}, \quad \mathbf{b} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}), \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}\sigma^2), \tag{3.2}$$

with \mathbf{b} as a vector containing so-called random effects with mean zero and covariance matrix Ψ (Wood, 2017, chap. 2). Random effects can model heterogeneity in the data, particularly if the data set has a hierarchical structure and many observations are available for several observational units (e.g., different animals). For instance, if one animal has traveled a greater distance in a month than others, this does not mean that it will consistently travel farther than other animals. That would rather be a random effect than a particular subject's effect. The first terms in (3.1) and (3.2) represent the regressor matrix and the corresponding coefficients ξ . The matrix Z is a simple model matrix for the random effects. In particular, Z contains fixed coefficients that describe how y depends on the random effects. Usually, Z is a subset of X because it does not make sense to assume a random deviation for every covariate effect in most applications. Moreover, the matrix Λ is a positive definite matrix that can be used to model autocorrelation. Consequently, the elements of the response vector y may no longer be assumed to be independent (Wood, 2017, chap. 2). However, for most regression models used in the literature, Λ corresponds to the identity matrix I .

Generalised Additive Models

GAMs are a generalisation of additive models. The linear predictor of a GAM contains one or more smoothing functions for the expected value of the response, while the response variable itself may follow any exponential family distribution:

$$h(\mathbb{E}(y_i)) = \mathbf{X}_i \xi + g_1(x_{1i}) + g_2(x_{2i}) + g_3(x_{3i}) + \dots \quad (3.3)$$

In Equation (3.3), \mathbf{X}_i is the i -th row of a model matrix corresponding to any strictly parametric model component, with parameter vector ξ . The expression $g_j(x_j)$ is a smoothing function of some covariate x_j . The response variable y follows any exponential family distribution and h is a known, monotonic, twice differentiable link function (Wood 2017, chap. 2; Hastie and Tibshirani 1990, chap. 6). The concept of link functions is also known from generalised linear models.¹

According to Wood (2017, chap. 4), the smoothing function $g_j(\cdot)$ can be represented as a linear combination of d_j (known) basis functions ψ_{jk} with parameters

¹For a detailed discussion of link functions, see McCulloch et al. (2008, chap. 5).

β_{jk} with

$$g_j(x) = \sum_{k=1}^{d_j} \beta_{jk} \psi_{jk}(x), \quad j = 1, \dots, J. \quad (3.4)$$

Equation (3.4) states that $g_j(x)$ is an element of the defined basis. The basis functions themselves are supposed to represent unknown functions. Consequently, choosing those basis functions suitable for approximating known functions makes sense. Splines are an appropriate option for smoothing functions; for example, cubic regression splines, P-splines, and thin plate regression splines could be used. In fact, most smoothing functions are spline-based. Nonetheless, it would not be appropriate to model variables like 'time of day' using ordinary splines because the fitted smooth would change discontinuously at the end of the day. Hence, splines can be modified to cyclic versions where, for any smoothing spline $\bar{g}_j(x)$, the spline function must be continuous up to the second derivative, whereas the derivatives of the endpoints must match, and $\beta_{j1} = \beta_{jd_j}$. The cyclic version of a spline ensures that it has the same value at its upper and lower boundaries (Wood, 2017, chap. 6).

For model fitting, the parameters ξ and β_{jk} must be estimated. For convenience, it will be assumed that vector β contains ξ , and the individual smooth term coefficient vectors β_j are stacked on end. The estimation process is called penalised likelihood maximisation because ordinary likelihood maximisation is prone to overfitting. The penalties are defined so that g_j is not overly wiggly, intended to account for the bias-variance trade-off.² To give an example of a spline-based smoother, and consistent to (3.4), penalised regression splines are used. The objective function

$$\|\mathbf{y} - \mathbf{X}\beta\|^2 + \sum_{j=1}^J \lambda_j \int_0^1 [g_j''(x)]^2 dx \quad (3.5)$$

has to be minimised, where the first term represents the sum of squared residuals, and the second term measures the overall curvature of the smoothers, weighted by a smoothing parameter λ_j (Wood, 2017, chap. 5) and represents the penalisation. It basically means that λ_j controls the trade-off between under- and overfit-

²The estimation methods are discussed in more detail in Wood (2017, chap. 6).

ting for every smoothing function $g_j(x)$. For $\lambda_j \rightarrow \infty$, the curvature of the smoothing function $g_j(x)$ will be as low as possible, resulting in a straight line. Vice versa, $\lambda_j = 0$ means that the curvature will not be penalised at all. Since $g_j(x)$ is linear in the parameters, β_{jk} , and $g_j''(x) = \sum_{k=1}^{d_j} \beta_{jk} \psi_{jk}''(x)$, the penalisation component can be written as a quadratic form

$$\int_0^1 [g_j''(x)]^2 dx = \boldsymbol{\beta}^T \mathbf{S}_j \boldsymbol{\beta}, \quad (3.6)$$

where \mathbf{S}_j is a positive definite matrix depending only on the basis functions, turning (3.5) into

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \sum_{j=1}^J \lambda_j \boldsymbol{\beta}^T \mathbf{S}_j \boldsymbol{\beta}. \quad (3.7)$$

The objective function (3.7) then must be minimised w.r.t. $\boldsymbol{\beta}$, given λ . In practice, this is often achieved by using penalised iterative re-weighted least squares (PIRLS) (Green 1984; Wood 2017, chap. 6).

PIRLS consists of the following three steps:

1. Initialise the estimates of the expected value of the exponential family distribution, μ_i , with $\hat{\mu}_i = y_i$, and the predictor with $\hat{\eta}_i = h(\hat{\mu}_i)$. The following two steps will then be iterated until convergence.
2. Based on the estimates of the previous step, compute pseudodata $z_i = h'(\hat{\mu}_i)(y_i - \hat{\mu}_i)/\iota(\hat{\mu}_i) + \hat{\eta}_i$ and iterative weights $w_i = \iota(\hat{\mu}_i)/\{h'(\hat{\mu}_i)^2 V(\hat{\mu}_i)\}$.
3. For fixed λ , find $\hat{\boldsymbol{\beta}}$ to minimise the weighted least squares objective

$$\|\mathbf{z} - \mathbf{X}\boldsymbol{\beta}\|_W^2 + \sum_{j=1}^J \lambda_j \boldsymbol{\beta}^T \mathbf{S}_j \boldsymbol{\beta} \quad (3.8)$$

to update $\hat{\eta} = \mathbf{X}\hat{\boldsymbol{\beta}}$ and $\hat{\mu}_i = h^{-1}(\hat{\eta}_i)$.

$V(\hat{\mu}_i)$ denotes the variance function determined by the exponential family distribution, while the influence term, $\iota(\hat{\mu}_i)$, is defined as $\iota(\hat{\mu}_i) = [1 + (y_i - \mu_i)\{V'(\mu_i)/V(\mu_i) + h''(\mu_i)/h'(\mu_i)\}]$ (Wood, 2017, chap. 6).

A measure for the flexibility of the model is represented by degrees of freedom. However, the parameters are fit in a restricted fashion controlled by λ . Con-

sequently, ordinary degrees of freedom would be misleading and thus, effective degrees of freedom (EDF) are computed. Using the last iteration of PIRLS and defining $\mathbf{F} = (\mathbf{X}^T \mathbf{W} \mathbf{X} + \mathbf{S}_\lambda)^{-1} \mathbf{X}^T \mathbf{W} \mathbf{X}$ with $\mathbf{W} = \text{diag}(w_i)$ and $\mathbf{S}_\lambda = \sum_{j=1}^J \lambda_j \mathbf{S}_j$, the EDF can be calculated as

$$\tau = \text{tr}(\mathbf{F}). \quad (3.9)$$

The matrix \mathbf{F} can be interpreted as mapping the non-penalised coefficient estimates to the penalised coefficient estimates. Hence, $\text{tr}(\mathbf{F})$ is basically the average shrinkage undergone by the coefficients, multiplied by the number of coefficients (Hastie et al. 2009, chap. 5; Wood 2017, chap. 6). To obtain the EDF of a specific smoothing spline $g_j(x)$, the values F_{ii} corresponding to the parameters β_j must be added up.

The penalisation parameter λ can be estimated by cross validation or maximum likelihood estimation. In favour of computational efficiency, generalised cross validation (GCV), as well as the so-called restricted maximum likelihood estimation (REML), are often used to estimate λ (Wood, 2004). REML is used instead of ordinary maximum likelihood estimation because ordinary maximum likelihood estimation underestimates variance components in this context³ (Wood, 2017, chap. 6), however, GCV and REML will not be discussed, because it would go beyond the scope of this work. Figure 3.1 graphically represents how splines work. An amount of d_j basis functions will be weighted by their estimated respective parameters β_j . The weighted values of the basis functions are then added up and thus, form a regression spline.

Generalised Additive Mixed Models

Now that GAMs and mixed models are introduced, the transition from GAMs to GAMMs is straightforward. Combining mixed models and GAMs results in a GAMM. A GAMM, thus, has the form

$$h(\mathbb{E}(y_i)) = \mathbf{X}_i \boldsymbol{\xi} + g_1(x_{1i}) + g_2(x_{2i}) + g_3(x_{3i}) + \dots + \mathbf{Z}_i \mathbf{b}, \quad \mathbf{b} \in \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}). \quad (3.10)$$

³For large sample sizes, the bias of the variance component gets smaller. Nonetheless, when working with high dimensional data, the variance estimators can be seriously biased.

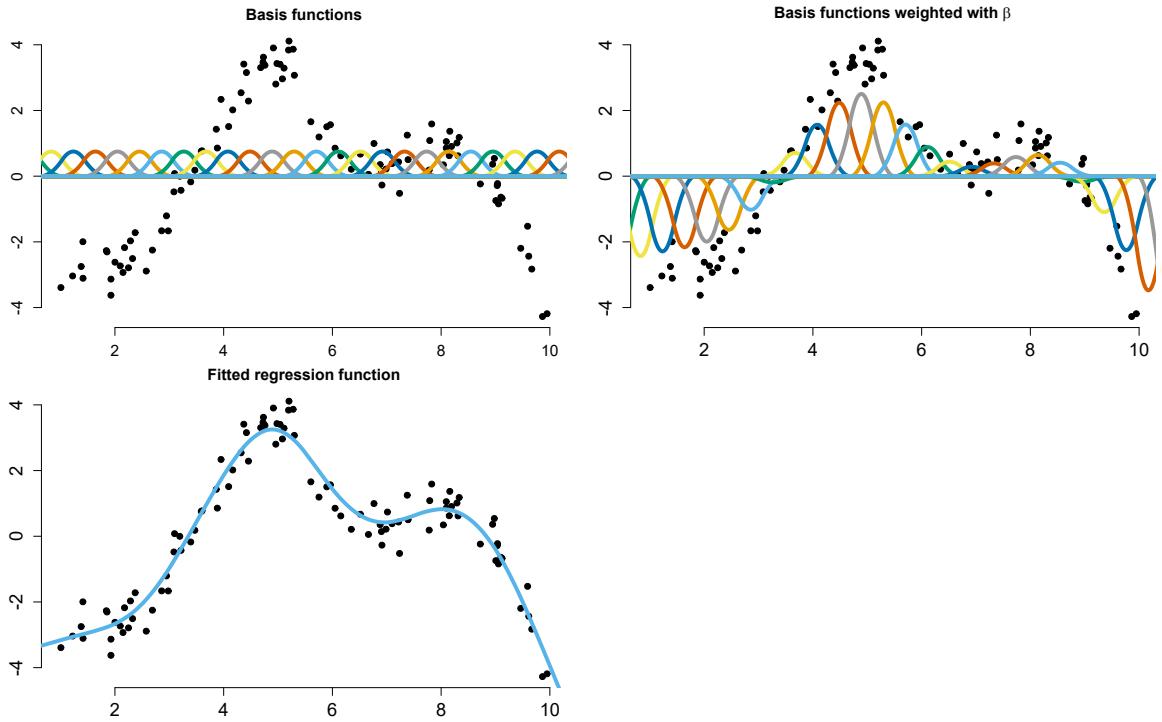


Figure 3.1: P-spline fitted regression function.
Source: own representation based on code by R. Langrock.

Consistent with (3.3), X_i again is the i -th row of a fixed-effects model matrix, with parameter vector ξ . Moreover, $g_j(x_j)$ is a smoothing function of some covariate x_j . As described in (3.2), Z_i and b is the i -th row of a random-effects model matrix and a vector of random effects coefficients with positive definite coefficient matrix Ψ , respectively. As specified in (3.3), y follows any exponential family distribution, and h is a known, monotonic, twice differentiable link function. Moreover, consistent with (3.2), an error term ϵ with i -th element ϵ_i can be incorporated with covariance matrix Λ to model correlation structures, which is an advantage compared to GAMs (Wood, 2006). From a Bayesian perspective, there is a duality between smoothing functions and random effects. It follows that simple Gaussian random effects can also be estimated as if they were smoothing terms in the model, or vice versa, smoothing functions, and GAMs can be estimated using mixed models estimation methods, which would be REML (Wahba 1983; Silverman 1985; Wood 2017, chap. 2, 6). Alternatively, PIRLS can be used to estimate β , and REML to obtain λ , as described for GAMs.

Turning to model selection, Akaike's information criterion (AIC) is a popular choice. In the context of GA(M)s, AIC requires some further attention because this part of model selection is distinct from (smoothing) parameter estimation. There are two main approaches for AIC: Marginal AIC and conditional AIC. Basically, they differ in the way they treat smooths. To obtain the marginal AIC, the model's (frequentist) marginal likelihood is used. Furthermore, smooths are treated as random effects and will not count for the number of parameters used. As mentioned, the frequentist marginal likelihood tends to underestimate variance components. However, replacing the ordinary likelihood estimation with REML does not solve the issue because the restricted likelihood of a model is only comparable to those models with the same fixed effects structure and hence cannot be compared in the sense of smooths (since the smooths are treated as random effects). Consequently, it cannot be evaluated if a smooth should be added to the model or not. In contrast, the conditional AIC is based on the (conditional) likelihood of all coefficients and uses EDF in place of the number of model parameters. In particular, smooths are treated as fixed effects (Wood, 2017, chap. 6). As Greven and Kneib (2010) have shown, the conditional version of AIC tends to select rather complex models, specifically when the model contains random effects. Thus, Wood et al. (2016) proposed a correction to the EDF to obtain a corrected version of AIC, accounting for smoothing parameter estimation uncertainty. In the remaining work, the corrected version of AIC will be used for GAM(M)s.

All in all, GAM(M)s provide a very accessible framework to model all kinds of relationships due to flexible smoothing functions. However, when modelling the observations in a time series context and assuming the observations are not independent of each other, the implementation of autocorrelations can be very time-consuming and computationally demanding. Moreover, in movement ecology, there are many approaches to interpreting the movement characteristics of animals by movement modes (see Patterson et al. (2009), for example). Thus, a popular method of modelling animal tracking data will be discussed in the next section: HMMs.

3.2 Hidden Markov Models

This section deals with hidden Markov models. Once the dataset has been imputed, HMMs are flexible tools for dealing with time series data, particularly for discrete-valued series. The main reason is that HMMs provide a convenient way to relax the assumption that the observed data are independent (as will be seen, they are conditionally independent) without being overly complex. HMMs can be conceptually divided into two blocks: the first block is a set of distributions that depend on the second block, an unobserved Markov chain consisting of different states. Then, depending on the active state, a particular distribution is chosen. That is, the state-dependent distribution then generates an observation Y_t for each $t = 1, \dots, T$. This section has been divided into three logical subsections to make it more concise and clear.

Markov Chains

Since the second block, the so-called Markov chain, is essential for understanding an HMM, a brief introduction is given. In general, a Markov chain is a sequence of discrete random variables $\{C_t : t \in \mathbb{N}\}$ if it satisfies the Markov property:

$$\Pr(C_{t+1} | C_t, \dots, C_1) = \Pr(C_{t+1} | C_t). \quad (3.11)$$

This property induces a certain dependency structure. Given the past C_{t-1}, \dots, C_1 of the process, the future value C_{t+1} only depends on C_t . Another major aspect of Markov chains is that they can take on m different values (states) in any countable set S with transition probabilities

$$\gamma_{ij}(t, s) = \Pr(C_{s+t} = j | C_s = i). \quad (3.12)$$

Equation (3.12) states that the probability of being in state j at time $s + t$, given the chain is in state i at time s , is denoted by $\gamma_{ij}(t, s)$. If the transition probability does not depend on s , the Markov chain is considered homogeneous (Zucchini et al., 2016, chap. 1). This will be denoted by $\gamma_{ij}(t)$ and shall be assumed in this work unless there is an explicit indication of the contrary. The row stochastic matrix $\Gamma(1)$, which will be abbreviated as Γ , contains the one-step transition probabilities

and is given by

$$\Gamma = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{pmatrix}. \quad (3.13)$$

Figure 3.2 provides a graphical representation of a basic three-state Markov process. The arrows indicate the probabilities of switching from state i to state j . The matrix $\Gamma(1)$ contains all one-step transition probabilities.

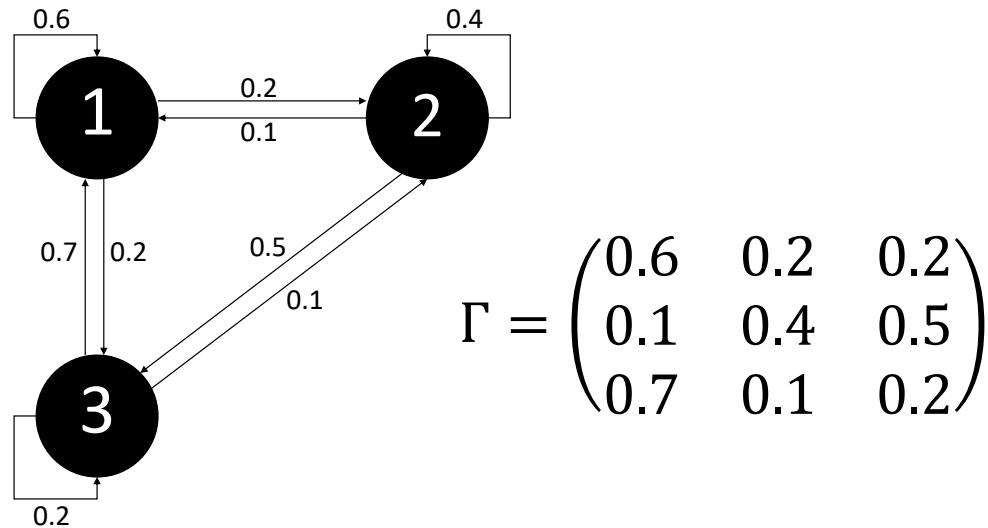


Figure 3.2: 3-state Markov process.
Source: own representation based on Zucchini et al. (2016, chap. 2).

In most cases, one is interested in the long-term behaviour of the Markov chain and whether there exists a limiting distribution of the chain. Such a limiting distribution is closely related to the existence of a so-called stationary distribution. A Markov chain is said to have a stationary distribution δ , where δ is a row vector

with entries δ_j , if

$$\delta_j \geq 0, \quad \forall j \quad (3.14)$$

$$\boldsymbol{\delta}\boldsymbol{\Gamma} = \boldsymbol{\delta} \quad (3.15)$$

$$\boldsymbol{\delta}\mathbf{1}' = 1. \quad (3.16)$$

Equation (3.15) indicates the stationarity, while (3.14) and (3.16) specify that $\boldsymbol{\delta}$ is a probability distribution since the entries add up to 1 and are non-negative. If one additionally assumes that the Markov chain is irreducible and aperiodic, it follows that a unique limiting distribution exists, which is the stationarity distribution (Zucchini et al., 2016, chap. 2). The vector $\boldsymbol{\delta}$ is a stationary distribution if, and only if

$$\boldsymbol{\delta}(\mathbf{I}_m - \boldsymbol{\Gamma} + \mathbf{U}) = \mathbf{1}, \quad (3.17)$$

where \mathbf{I}_m is the $m \times m$ identity matrix, and \mathbf{U} is the $m \times m$ matrix of ones. Thus, Equation (3.17) can be used to compute the stationary distribution if the assumptions hold (Zucchini et al., 2016, chap. 2). For the tracking data, it can be assumed that the underlying process is in its stationary distribution since the elephants existed before they were tracked, and thus the process has been running for some time. Consequently, $\boldsymbol{\delta}$ then gives the probabilities of encountering the process in any of the m states.

The properties of Markov chains form the core of the model structure of HMMs. The purpose of Markov chains in this context is that they can conveniently relax the independence assumption of the observations. In particular, they allow the observations to be independent conditional on the state being active, as shall be seen in the mathematical definition of HMMs.

Model Formulation

Formally HMMs consist of an unobserved parameter process $\{C_t : t \in \mathbb{N}\}$ and an observed state-dependent process $\{Y_t : t \in \mathbb{N}\}$ with state-dependent distributions. In particular, the underlying parameter process is a Markov chain and consists of m states. Most HMMs satisfy the Markov property given in Equation (3.11). This allows for serial dependence in the observations, such that a simple

HMM can be characterised by

$$\Pr(C_t | \mathbf{C}^{(t-1)}) = \Pr(C_t | C_{t-1}), \quad t = 2, 3, \dots \quad (3.18)$$

$$\Pr(Y_t | \mathbf{Y}^{(t-1)}, \mathbf{C}^{(t)}) = \Pr(Y_t | C_t), \quad t \in \mathbb{N}. \quad (3.19)$$

The above Equations (3.18) and (3.19) show that conditioning on the history of the Markov chain $\mathbf{C}^{(t-1)}$, C_t only depends on C_{t-1} . Moreover, the Markov property implies that $\{C_t : t \in \mathbb{N}\}$ is fully characterised by the stationary distribution δ and the transition probability matrix Γ . The same principle applies to the state-dependent process, namely that the distribution of Y_t depends only on C_t and neither on a previous state $\mathbf{C}^{(t-1)}$ nor on the history of observations $\mathbf{Y}^{(t-1)}$. Thus, Equation (3.19) defines the said conditional independence assumption. In particular, Equation (3.19) implies that $\{Y_t : t \in \mathbb{N}\}$ is fully characterised by the state-dependent distributions (Zucchini et al., 2016, chap. 2). Figure 3.3 illustrates the structure of a basic HMM. The unobserved parameter process C_t and the state-dependent process Y_t are shown.

The state-dependent distributions in the case of discrete observations can be written as

$$p_i(y) = \Pr(Y_t = y | C_t = i). \quad (3.20)$$

Equation (3.20) indicates that p_i is the probability mass function of Y_t when the Markov chain is in state i at time t (Zucchini et al., 2016, chap. 2). It is worth noting that the state-dependent distributions are usually chosen as a class of parametric distributions. However, sometimes it may be challenging to decide which parametric distribution to use. An alternative is to estimate a non-parametric state-dependent distribution, for example, by using P-splines, as shown in Langrock et al. (2015). Figure 3.4 shows the process that generates the observations in a two-state HMM. Each state corresponds to a state-dependent distribution, and the active distribution then generates the observation.

With the state-dependent distributions, one can obtain the marginal distribu-

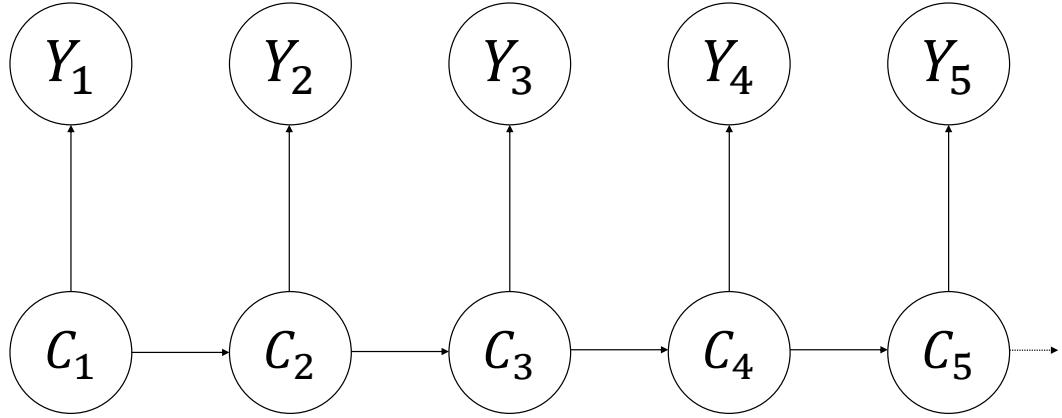


Figure 3.3: Basic HMM structure.

Source: own representation based on Zucchini et al. (2016, chap. 2).

tion of Y_t . By defining $u_i(t) = \Pr(C_t = i)$ for $t = 1, \dots, T$ it follows that

$$\Pr(Y_t = y) = \sum_{i=1}^m \Pr(C_t = i) \Pr(Y_t = y \mid C_t = i) \quad (3.21)$$

$$= \sum_{i=1}^m u_i(t) p_i(y), \quad (3.22)$$

or, in matrix notation:

$$\Pr(Y_t = y) = (u_1(t), \dots, u_m(t)) \begin{pmatrix} p_1(y) & & 0 \\ & \ddots & \\ 0 & & p_m(y) \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \quad (3.23)$$

$$= \mathbf{u}(t) \mathbf{P}(y) \mathbf{1}', \quad (3.24)$$

where $\mathbf{P}(y)$ is a diagonal matrix of all state-dependent distributions (Zucchini et al., 2016, chap. 2). Equations (3.22) and (3.24) show that the marginal distribution of an HMM is simply a mixture distribution. If the underlying Markov

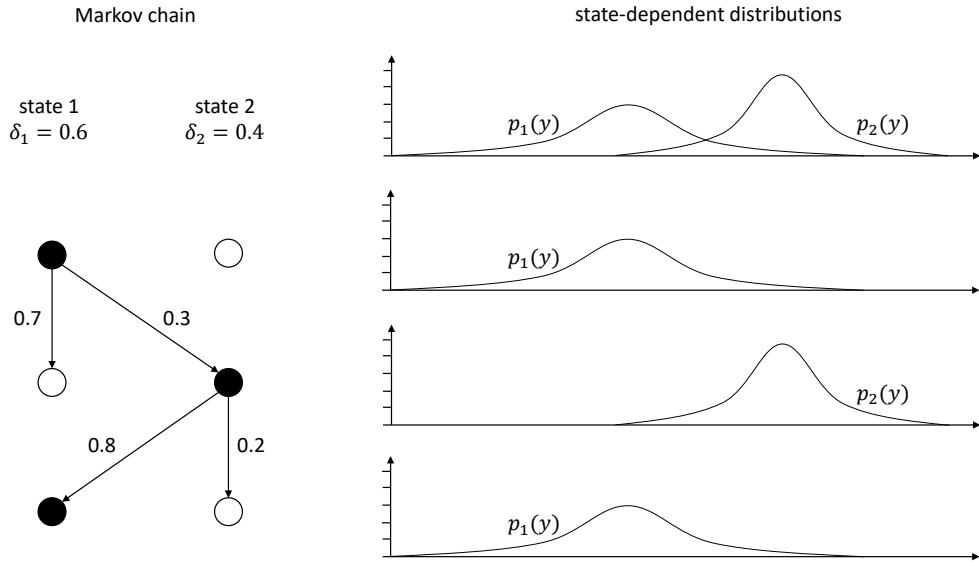


Figure 3.4: 2-state HMM process.
Source: own representation based on Zucchini et al. (2016, chap. 2).

chain is stationary with stationary distribution δ , Equation (3.24) even reduces to

$$\Pr(Y_t = y) = \delta \mathbf{P}(y) \mathbf{1}' . \quad (3.25)$$

The discussed framework above forms the basis for all HMMs. Thus, an HMM is a dependent mixture model where a Markov chain selects the state-dependent distributions, in its simplest form. On this basis, there are numerous extensions of HMMs.⁴ Some common examples are multivariate time series (assuming longitudinal and contemporaneous conditional independence), integration of covariates and seasonality, random effects, and higher-order Markov chains. As will be seen, some of these are used in later sections; particularly, the multivariate case and the integration of covariates and seasonality. First, however, before an HMM can be fitted to data, the likelihood must be calculated.

⁴And in fact, an HMM itself is a special case of SSMs. SSMs are also doubly stochastic processes, with the state process being continuous-valued, whereas HMMs take on a finite number of discrete states.

Likelihood

The likelihood can be used to estimate the parameters linked to an HMM (e.g. the parameters of the assumed state-dependent distributions and transition probabilities). However, before the likelihood can be maximised, it must be calculated. The likelihood L_T of an HMM for consecutive observations y_1, \dots, y_T is given by the general formula

$$L_T = \delta \mathbf{P}(y_1) \Gamma \mathbf{P}(y_2) \cdots \Gamma \mathbf{P}(y_T) \mathbf{1}'. \quad (3.26)$$

Equation (3.26) shows that the initial distribution δ of the Markov chain is multiplied by the diagonal matrix of the state-dependent distributions evaluated at y_1 and so on (Zucchini et al., 2016, chap. 3).

It has been (and still is) claimed in several statements in the literature that direct maximisation of the likelihood expressed in (3.26) is not feasible because m^T summands need to be computed (Albert, 1991; Morales et al., 2004). In terms of complexity, this leads to $O(Tm^T)$ operations and motivates complex estimation techniques such as the expectation-maximisation (EM) algorithm or Markov chain Monte Carlo (MCMC)-based methods. However, from more general discussions, it is well-known that recursive schemes can be used to compute the likelihood in a computationally much more efficient way. The likelihood described in Equation (3.26) clearly shows such a recursive scheme. In fact, defining a vector α_t for $t = 1, \dots, T$ leads to the so-called forward probabilities:

$$\alpha_t = \delta \mathbf{P}(y_1) \Gamma \mathbf{P}(y_2) \cdots \Gamma \mathbf{P}(y_t) = \delta \mathbf{P}(y_1) \prod_{s=2}^t \Gamma \mathbf{P}(y_s). \quad (3.27)$$

The forward probabilities defined in (3.27) contain information on the likelihood up to time t . In particular, the likelihood L_T can be computed using the forward algorithm:

$$\alpha_1 = \delta \mathbf{P}(y_1); \quad (3.28)$$

$$\alpha_t = \alpha_{t-1} \Gamma \mathbf{P}(y_t) \quad \text{for } t = 2, 3, \dots, T; \quad (3.29)$$

$$L_T = \alpha_T \mathbf{1}'. \quad (3.30)$$

Using Equations (3.28), (3.29), and (3.30), α_1 can be computed first, then α_2

based on α_1 and so on. As expressed in (3.30), with $\alpha_T = (\alpha_T(1), \dots, \alpha_T(m))$ the likelihood then is $\sum_{j=1}^m \alpha_T(j) = L_T$ (Zucchini et al., 2016, chap. 3).

By taking advantage of the forward algorithm, the number of operations involved is of order Tm^2 ($O(Tm^2)$, and just a fractal of $O(Tm^T)$). Thus, it can easily be maximised numerically in order to tailor an HMM to the data. The purpose of presenting the computation of the likelihood is to show that HMMs do not require substantially higher computational effort than the other methods discussed in this text. However, there are some issues to consider when performing direct numerical maximisation, such as numerical underflow and constrained parameters, but this is beyond the scope of this thesis (see Zucchini et al., 2016, chap. 3).

Considering again the forward probabilities α_t , they indeed can be identified as probabilities. The j -th component $\alpha_t(j)$ of α_t is the joint probability $\Pr(Y_1 = y_1, Y_2 = y_2, \dots, Y_t = y_t, C_t = j)$. Analogously, the conditional probability $\Pr(Y_{t+1} = y_{t+1}, Y_{t+2} = y_{t+2}, \dots, Y_T = y_t \mid C_t = j)$ can be defined using so-called backward probabilities. In particular, the vector of backward probabilities, $\varphi_t(j)$, is defined by

$$\varphi'_t = \mathbf{\Gamma P}(y_{t+1}) \mathbf{\Gamma P}(y_{t+2}) \cdots \mathbf{\Gamma P}(y_T) \mathbf{1}' = \left(\prod_{s=t+1}^T \mathbf{\Gamma P}(y_s) \right) \mathbf{1}', \quad t = 1, 2, \dots, T, \quad (3.31)$$

whereas the j -th component of φ_t corresponds to the conditional probability $\Pr(Y_{t+1} = y_{t+1}, Y_{t+2} = y_{t+2}, \dots, Y_T = y_t \mid C_t = j)$ (Zucchini et al., 2016, chap. 4). The forward and backward probabilities are helpful for the derivation of the most likely state of the Markov chain at time $t \in \{1, \dots, T\}$, which are state probabilities. The conditional distribution of C_t given the observations can be obtained, for $j = 1, 2, \dots, m$ as

$$\begin{aligned} \Pr(C_t = j \mid \mathbf{Y}^{(T)} = \mathbf{y}^{(T)}) &= \frac{\Pr(C_t = j, \mathbf{Y}^{(T)} = \mathbf{y}^{(T)})}{\Pr(\mathbf{Y}^{(T)} = \mathbf{y}^{(T)})} \\ &= \frac{\Pr(\mathbf{Y}^{(T)} = \mathbf{y}^{(T)} \mid C_t = j) \Pr(C_t = j)}{\Pr(\mathbf{Y}^{(T)} = \mathbf{y}^{(T)})} \\ &= \frac{\Pr(\mathbf{Y}^{(t)} = \mathbf{y}^{(t)} \mid C_t = j) \Pr(C_t = j) \Pr(\mathbf{Y}_{t+1}^{(T)} = \mathbf{y}_{t+1}^{(T)} \mid C_t = j)}{\Pr(\mathbf{Y}^{(T)} = \mathbf{y}^{(T)})} \\ &= \frac{\alpha_t(j) \varphi_t(j)}{L_T}, \end{aligned} \quad (3.32)$$

with $\mathbf{Y}_{t+1}^{(T)}$ denoting the history of \mathbf{Y} from $t + 1$ up to T . Consequently, given the observations $\mathbf{y}^{(T)}$, the (discrete) probability distribution of C_t for $t = 1, \dots, T$ can be determined using the forward and backward probabilities (Zucchini et al., 2016, chap. 5).

As mentioned at the beginning of this section, HMMs provide a flexible tool for time series analysis, especially in the discrete-time context. However, the HMMs described above, as well as GAMs and GAMMs, can only handle datasets with regularly spaced observation times. Since the given dataset has irregularly spaced observation times, the data need to be imputed before the mentioned concepts can be tailored to them. Given this circumstance, an alternative approach uses SDEs, as will be shown in the following section.

3.3 Stochastic Differential Equations

In addition to the HMM framework, SDEs are also a popular choice for modelling data collected over time. An essential aspect of this is that SDEs do not depend on regular-spaced observation times, are relatively simple, and have parameters that are easy to interpret. Consequently, no imputation is needed. However, SDEs are formulated in continuous-time, whereas HMMs are based on a discrete-time framework. Because SDEs are considered not flexible enough to model animal movements, this method will be extended in the following.

Formally the most common SDE is stated as

$$dY_t = \mu(Y_t, t)dt + \sigma(Y_t, t)dW_t, \quad Y_0 = y_0, \quad (3.33)$$

with W_t being a Wiener process⁵ and y_0 as the initial condition of the stochastic process Y_t . Equation (3.33) represents the evolution of a process Y_t with μ as the expected change in the process for an infinitesimal time interval. The function σ , frequently referred to as diffusion, is a measure for the variability (Michelot et al., 2021). A solution to (3.33) often results in parametric functions for μ and σ .

A more recent approach by Michelot et al. (2021) provides more flexibility within the structure of the basic SDE shown in (3.33) and allows for broad time-varying dynamics. In particular, μ and σ can be specified as basic penalty smooth-

⁵See Karatzas and Shreve (2012) for an in-depth discussion of Wiener processes.

ing splines. Splines have already been discussed in Section 4.1. In principle, the approach by Michelot et al. (2021) allows the integration of linear and nonlinear effects in covariates and random effects in SDEs. Unlike HMMs, where the estimated parameters of the assumed distributions are piece-wise constant (for each state), this approach allows smooth, time-varying parameters (Michelot et al., 2021).

Model Formulation

In this setting, a stochastic process Y_t is defined by

$$dY_t = \mu(Y_t, \theta_t)dt + \sigma(Y_t, \theta_t)dW_t, \quad Y_0 = y_0. \quad (3.34)$$

In contrast to (3.33), Equation (3.34) describes a dependency of μ and σ from a time-varying parameter vector θ_t , whereas μ and σ themselves determine the type of SDE (e.g., (geometric) Brownian Motion, Ornstein-Uhlenbeck process, CTCRW, etc.).

The parameter vector θ_t itself depends on J covariates x_{1t}, \dots, x_{Jt} and consists of components

$$h(\theta_t) = \beta_0 + g_1(x_{1t}) + \dots + g_J(x_{Jt}), \quad (3.35)$$

with a link function h and β_0 as an intercept parameter. The functions $g_j(\cdot)$ are not further specified as they could be linear or nonlinear effects of a covariate as well as random effects or smooth functions (Michelot et al., 2021). If the functions $g_j(\cdot)$ are smoothing functions, the definition of smoothing functions in (3.4) can be applied. Equations (3.34), (3.35), in combination with (3.4), form a model called varying-coefficient stochastic differential equation, which is based on varying-coefficient models described by Hastie and Tibshirani (1993) in the context of GAMs.

Likelihood

To fit this model formulation to data, the likelihood must be computed. In particular, the aim is to estimate the relationship between the time-varying parameter

vector θ_t , which determines the drift μ , and the diffusion σ of the stochastic process Y_t , as can be seen in Equation (3.34).

Given a sequence of n consecutive observations (with potentially irregular spaced observation times) y_1, \dots, y_n from the process Y_t with times t_1, \dots, t_n the goal is to maximise the likelihood L , given the data with parameter vector θ_t :

$$L(\boldsymbol{\xi}, \boldsymbol{\beta} | y_1, \dots, y_n) = p(Y_{t_1} = y_1) \prod_{i=1}^{n-1} p(Y_{t_{i+1}} = y_{i+1} | Y_{t_i} = y_i). \quad (3.36)$$

There are some aspects to note in Equation (3.36). First, stochastic processes are Markovian, leading to the same dependency structure imposed by Equation (3.11). The dependency structure implies that the likelihood of n consecutive observations can be obtained as the product of the likelihoods of the individual transitions (Michelot et al., 2021). Vector $\boldsymbol{\beta}$ contains all coefficients of Equation (3.4) (e.g., the parameters for the smooth or random effects functions $g_j(x)$) and the vector $\boldsymbol{\xi}$ respectively, contains all other coefficients of the model (that is, $\boldsymbol{\xi}$ contains the parameters for linear covariate effects) with $p(\cdot)$ as the probability density function (PDF).

For evaluating the likelihood, it is assumed that y_1 is deterministic, i.e. $p(Y_{t_1} = y_1) = 1$. It follows that only the conditional densities $p(Y_{t_{i+1}} | Y_{t_i})$ need to be evaluated. In this thesis, two processes are used; namely, Brownian motion and Ornstein-Uhlenbeck processes (and CTCRW, as it is based on Ornstein-Uhlenbeck processes). There are closed-form expressions for the densities of these two processes. However, for the general case, the PDF $p(\cdot)$ can be approximated by a PDF of a normal distribution using the Euler-Maruyama discretisation.

As mentioned before, μ and σ determine which type of SDE is considered. For time-varying Brownian motion and time-varying Ornstein-Uhlenbeck processes, μ and σ depend only on one time-varying parameter, leading to $\theta_t = (\theta_t^{(1)}, \theta_t^{(2)})$ (Michelot et al., 2021). In the following, $\theta_t^{(1)} = r_t$ and $\theta_t^{(2)} = s_t$ will be used for simplicity.

In fact, defining Equation (3.34) as a time-varying version of Brownian motion can be achieved by setting $\mu(Y_t, \theta_t) = r_t$ and $\sigma(Y_t, \theta_t) = s_t$ which yields

$$dY_t = r_t dt + s_t dW_t, \quad Y_0 = y_0. \quad (3.37)$$

In particular, using the Euler-Maruyama discretisation for time-varying Brownian motion processes, one can obtain the approximated density

$$p(Y_{t+\Delta} = y_{t+\Delta} \mid Y_t = y_t) = \phi(y_{t+\Delta}; y_t + r_t \Delta, s_t^2 \Delta), \quad (3.38)$$

with $\phi(y; b, k)$ being the PDF of a normal distribution with mean b and variance k . The approximated density described in Equation (3.38) can now be substituted into Equation (3.36) yielding the approximate likelihood for the time-varying Brownian motion process.

The same is true for Ornstein-Uhlenbeck processes. For this particular process, let $\theta_t^{(1)} = \rho_t$ and $\theta_t^{(2)} = s_t$. To obtain a time-varying version of the Ornstein-Uhlenbeck process, Equation (3.34) must be modified by setting $\mu(Y_t, \theta_t) = \rho_t(\omega - Y_t)$ and $\sigma(Y_t, \theta_t) = s_t$. This modification leads to

$$dY_t = \rho_t(\omega - Y_t)dt + s_t dW_t, \quad Y_0 = y_0. \quad (3.39)$$

In Equation (3.39), ω can be viewed as a mean-reverting parameter.⁶ Again, using the Euler-Maruyama discretisation for the time-varying Ornstein-Uhlenbeck process yields an approximate density

$$p(Y_{t+\Delta} = y_{t+\Delta} \mid Y_t = y_t) = \phi(y_{t+\Delta}; \omega + e^{-\rho_t \Delta}(y_t - \omega), \frac{s_t^2}{2\rho_t}(1 - e^{-2\rho_t \Delta})), \quad (3.40)$$

which can then again be substituted into Equation (3.36) in order to get the approximate likelihood for the time-varying Ornstein-Uhlenbeck process (Michelot et al., 2021). Additionally, one can note the exact analogy to Equations (2.2) and (2.3), which define an Ornstein-Uhlenbeck velocity process, meaning that Equation (3.40) is a time-varying version of a CTCRW. As mentioned earlier, the time-varying Ornstein-Uhlenbeck process can be used to model the velocity of an animal, when fixing ω to zero. Fixing ω to zero is intended to avoid any systematic bias in the velocity.

Turning back to the general likelihood as given in Equation (3.36), penalising

⁶And in fact, to be consistent with Section 2.2, ω , again, is a mean velocity, meaning that the defined process tends to revert to this centre of attraction.

the smoothing terms leads to the penalised log-likelihood

$$l_p(\boldsymbol{\xi}, \boldsymbol{\beta}, \boldsymbol{\lambda} | y_1, \dots, y_n) = \log(L(\boldsymbol{\xi}, \boldsymbol{\beta} | y_1, \dots, y_n)) - \sum_{j=1}^J \lambda_j \boldsymbol{\beta}_j^T \mathbf{S}_j \boldsymbol{\beta}_j. \quad (3.41)$$

As seen in Section 3.1, the last term $\sum_{j=1}^J \lambda_j \boldsymbol{\beta}_j^T \mathbf{S}_j \boldsymbol{\beta}_j$ in Equation (3.41) is the penalisation of the smoothing splines to account for the bias-variance trade-off in the relationship between the parameter vector $\boldsymbol{\theta}_t$ and the covariates in the model. In particular, λ_j is a penalisation parameter for the j -th smooth term in $\boldsymbol{\theta}_t$. The matrix \mathbf{S}_j is one of known covariates in the model, and $\boldsymbol{\beta}_j$ is the vector of the basis coefficients as stated in Equation (3.4). Thus, $\boldsymbol{\beta}_j^T \mathbf{S}_j \boldsymbol{\beta}_j$ is a measure for the roughness (wiggleness) of the j -th smoothing term. When the likelihood given in Equation (3.41) has been maximised⁷, the fitted model must be checked and validated in order to select the best model among several potential candidates (Michelot et al., 2021). Figure 3.5 provides an example of parameter estimation of a time-varying

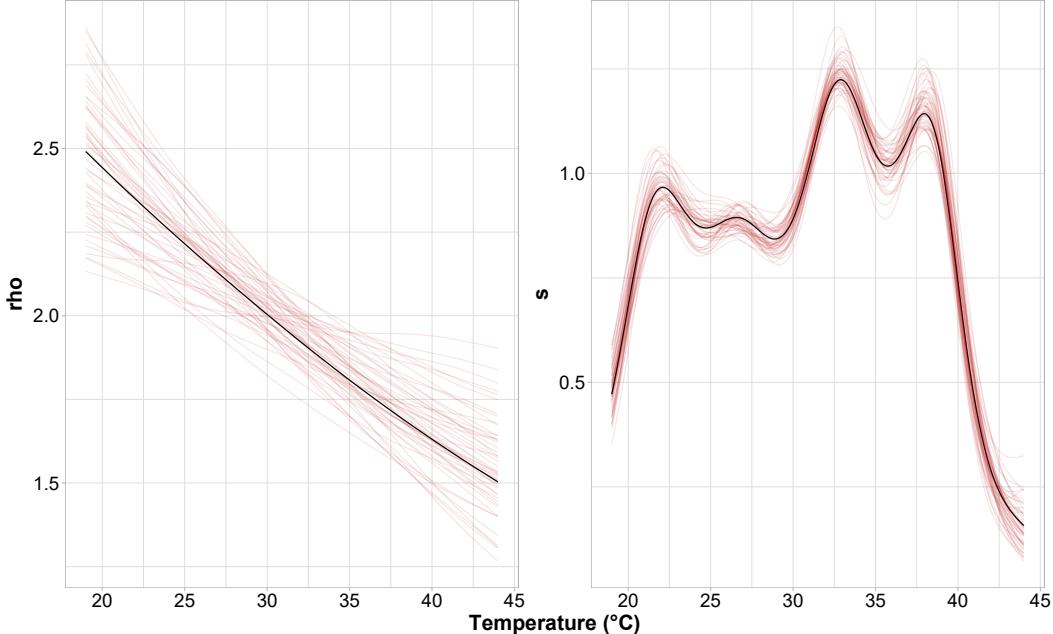


Figure 3.5: Estimated parameters for time-varying CTCRW.
Source: own representation.

⁷The likelihood itself is implemented using a Kalman filter with time-varying parameters, which is basically the same procedure as described in Section 2.2.

CTCRW, where ρ and s have been estimated. Both parameters depend on the covariate 'temp'. The red lines indicate 100 posterior draws, following Nychka's idea (1988), to generate confidence intervals.

At this point, all basics for the further procedure in this work have been introduced. In the following, the methods described above will be used to model the tracking data and gain more insights.

4 Results

This chapter provides the results of the analysis using the methods described in chapter 3. The results are divided into three logical blocks, as in the previous chapter. The results of the GAMMs are presented first, followed by those of the HMMs and finally, the outcome of the SDEs. All computations were performed on a 4.9 GHz Intel Core i7 9700k and 16 GB DDR4 RAM using the statistical software R (R Core Team, 2021). The focus in all three blocks is to model the relationship between animal movement characteristics, such as speed, step length, and turning angles, and environmental covariates like temperature and time of day.

4.1 Generalised Additive Mixed Models

The first method applied to the data is a GAMM to be directly compared to the GAMM fitted by Thaker et al. (2019). The GAMM original output by Thaker et al. (2019) and its model formulation can be seen in Appendix A1. In the first place, the GAMM in this work models the relationship between speed of movement (in km/h) and collar temperature, and hence, mimics the GAMM fitted by Thaker et al. (2019) as a reference measure. Speed of movement corresponds to the variable 'speed', whereas the collar temperature corresponds to the covariate 'temperature' (or 'temp', respectively). Additionally, 'season' was included as a categorical fixed effect as well as the animal ID and hour (time of day) as random effects. The season is divided into two categories: dry season (May to October) and wet season (November to April). Autocorrelation in the response variable is modeled via an AR(1) process, because consecutive observations of the movement path are correlated. In contrast to the GAMM by Thaker et al. (2019), woody density was not included because this information was not available. Apart from using an imputed dataset and including autocorrelation, this is the only difference in the model formulation. Thus, the first GAMM in this work can be compared to the GAMM by Thaker et al. (2019) to determine the impact of an imputed dataset

and incorporating autocorrelation.

Since this is a comparison of basically the same model, the choice of the best model candidate was limited to the choice of the spline basis parameter k , which equals the number of basis functions for a certain smoothing function g . If the number of basis functions is too small, the model will most likely oversmooth and tend not to capture any patterns in the data. Table 4.1 shows several possible choices for k . According to AIC and the adjusted R^2 , a choice of $k = 10$ leads to the best model and thus, is sufficient. However, a choice of $k = 15$ is not performing any worse.¹ The model itself was built using the function `bam` from the package `mgcv` (Wood, 2015). The GAMM was fitted using PIRLS, whereas the smoothing parameter estimation was done via REML. Furthermore, AIC was calculated using the conditional likelihood of the model (consequently, AIC is the conditional AIC) and the correction to the EDF proposed by Wood et al. (2016) was taken into account.

Table 4.1: Model selection - choice of k .

Choice of k	AIC	adj. R^2
$k = 5$	1499149	0.0677
$k = 10$	1499061	0.0687
$k = 15$	1499064	0.0687

Source: Own representation.

Figure 4.1 represents the fitted GAMM. The mean speed by temperature was calculated for each degree Celsius and classified by season, with triangles representing the wet season and circles for the dry season. The blue line represents a fitted temperature spline for the wet season, whereas the red dashed line represents a fitted spline for the dry season. It can be seen that elephants move faster when it gets hotter, while the increase per temperature decreases above 35° Celsius. Moreover, the individuals are faster in the wet season than in the dry season, which is consistent with the GAMM by Thaker et al. (2019). The output summary of the model can be found in Appendix A1.

However, Figure 4.1 indicates a more significant difference between the mean speed by season than indicated by the GAMM fitted by Thaker et al. (2019) which

¹As an alternative to specifying multiple GAMM candidate models and evaluating an appropriate choice of k by AIC, in Appendix A2, a more sophisticated method is shown.

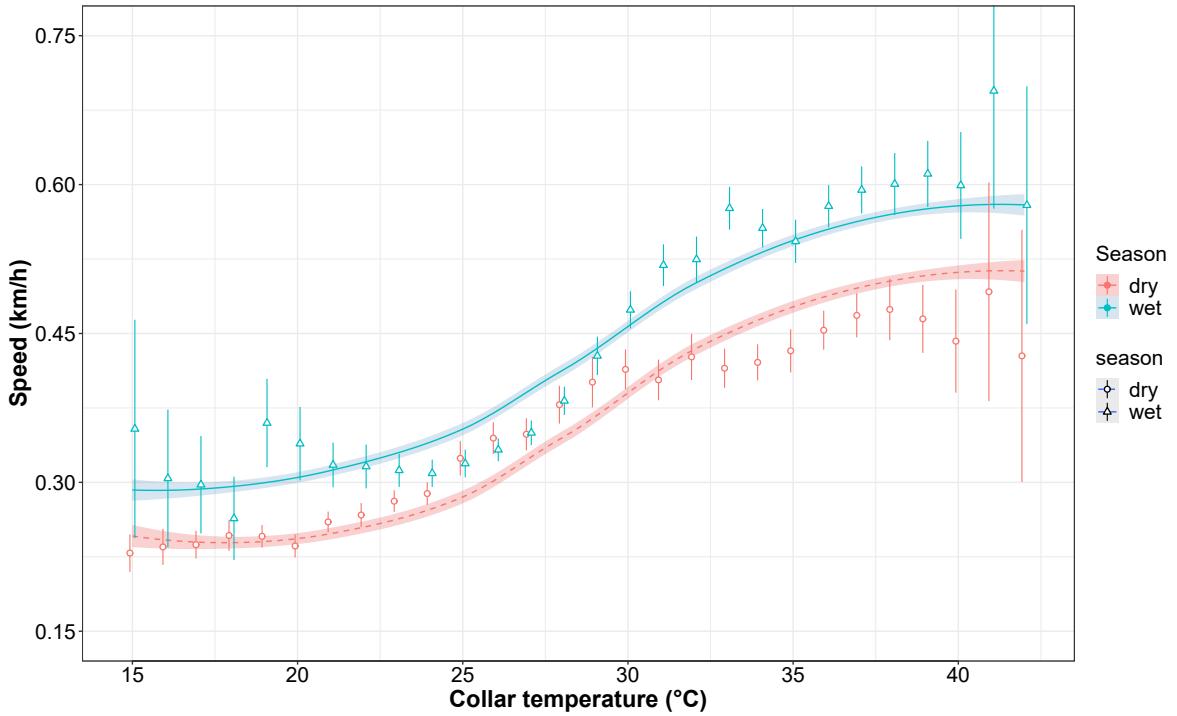


Figure 4.1: Fitted GAMM.
Source: own representation.

can be seen from the vertical distance between the splines. Also, the confidence intervals of the mean speed by temperature for both seasons and for the smoothing splines by season are wider. In fact, there are three reasons for the differences: The imputed dataset, the autocorrelation that is taken into account, and the omission of the covariate 'woody density'. Considering the first reason, Table 4.2 represents the difference in the mean speed by season. The calculation for the mean speed by season is independent of the GAMM fitted, and therefore, the difference can be related to the imputed dataset. While the mean movement speed in the wet season is faster, it is slower in the dry season compared to the mean speed computed by Thaker et al. (2019). The standard deviation of the mean speed in this work is about the same for the wet season and significantly lower in the dry season.

The reason for the deviance is that Thaker et al. (2019) neglected the irregularity when computing the mean speed. The imputed dataset in this work accounts for irregularity and hence shows a different mean speed because the calculated

Table 4.2: Mean speed by season - differences.

	Mean			Standard deviation		
	Wet	Dry	Difference	Wet	Dry	Difference
Mean speed Meyer	0.47	0.33	0.14	0.50	0.37	0.13
Mean speed Thaker et al. (2019)	0.42	0.39	0.03	0.49	0.46	0.03
Difference	0.05	-0.06	/	0.01	-0.09	/

Note: Wet and dry correspond to the possible categories of the covariate 'season'. Values are in km/h.

Source: Own representation.

speed is less biased by the distance traveled. Additionally, the imputation also influences the confidence intervals in the mean speed by temperature, shown in Figure 4.1.

Considering the GAMM splines in Figure 4.1, there are also wider confidence intervals than indicated in the model by Thaker et al. (2019). The reason (besides the mentioned speed differences in the datasets) is that autocorrelation is incorporated in the model of this work. If the observations are assumed to be uncorrelated when they are, in fact, correlated, the standard errors often are underestimated, and the confidence interval would be narrower (James et al. 2013, chap. 3; Wood 2017, chap. 5).

Autocorrelation itself is another aspect to note. Figure 4.2 shows the autocorrelation function (ACF) of the residuals of the fitted GAMM. Because mean, variance, and covariance of the process are not constant over time, non-stationarity, and, in particular, a pattern that repeats itself every 24 hours can be seen. The pattern indicates that there still is an effect not covered in the GAMM. Since this seems to be a daily pattern, an AR(24) process could solve this issue. The function `bam` can model autocorrelation via an AR(1) process; however, it cannot model the general AR(p) case. There are functions in R that are theoretically capable of modeling the AR(p) case², but AR(2) and AR(3) processes are already computationally very demanding to incorporate in GAMMs, and most computers do not have enough RAM to compute the autocorrelation matrices needed. Nevertheless, autocorrelation is a crucial aspect of movement data that Thaker et al. (2019)

²For example, the function `brm` from the package `brms` and the function `gamm` from the package `mgcv` are valuable alternatives.

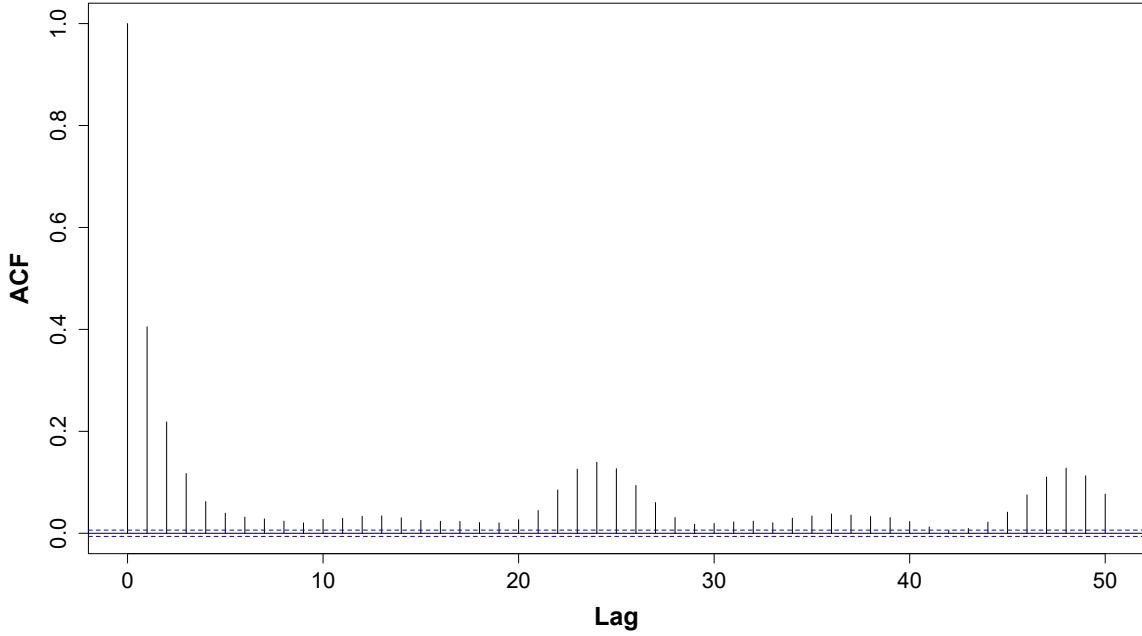


Figure 4.2: ACF-plot GAMM.

Source: own representation.

ignored when modeling the relationship between movement speed and temperature. Concerning the 'woody density', the effect of omitting this covariate cannot be measured since the data are not publicly available.

Up to this point, the two GAMMs were compared to determine the effect of the imputed dataset and the impact of accounting for autocorrelation in the data. That said, as can be seen in the model formulation, Thaker et al. (2019) included the covariate 'hour' as a random effect in the GAMM in order to model an hourly effect. Since there can still be seen a 24-hour cycle in the ACF-plot represented in Figure 4.2, it makes sense to reconsider the implementation of such a daily pattern. A cyclic cubic regression spline (or a cyclic penalised regression spline, respectively) to model the effect of a daily pattern might be advantageous in model fitting (Wood, 2017, chap. 7), and will be used in the following. Another issue is the assumed distribution of the response variable. Thaker et al. (2019) assumed that the response variable 'speed' is normally distributed. However, since the variable 'speed' does not take on negative values, a gamma distributed response is a con-

ceptually more appealing choice. Besides, a gamma distribution is suited to deal with heteroskedasticity in non-negative data because the variance is proportional to the mean squared (Ng and Cribbie, 2017). Therefore, in the following, two GAMMs will be fitted to further investigate the interaction between individuals' speed and their environmental system. A Gaussian distribution is assumed in the first GAMM, whereas in the second GAMM, a gamma distributed response variable is assumed. Moreover, h is a log link for the gamma GAMM, and an identity link function for the Gaussian GAMM.

For model selection, Table 4.3 shows multiple model variations of GAMMs, that assume a normally distributed response variable or a gamma distributed response variable, respectively. The first candidates both consist of covariates 'temperature' and 'hour', modelled as splines, while 'hour' is a cyclic cubic regression spline. Covariate 'season' is a categorical fixed effect, and 'ID' is a random effect. Additionally, the autocorrelation structure is, as before, incorporated in the models. The GAMM seen in Figure 4.1 results in a conditional AIC of 1499061, whereas the best candidates seen in Table 4.3 yield a conditional AIC of 1494092 and 1374824, respectively, whereas both are lower than the first GAMM. That means, instead of formulating time of day as a random effect, it is advantageous to use a cyclic regression spline to represent a daily pattern in the speed distribution. Likewise, assuming a gamma distributed response variable reduces AIC further. Not including the covariates 'season', 'temperature', 'id' and 'hour' results in a higher AIC. Omitting either the AR(1) process or the covariate time of day has the greatest effect, as, without them, there are considerably higher AIC values.

The resulting GAMMs are represented in Figure 4.3. The output summaries can be found in Appendix A1. Comparing both GAMMs to the GAMM by Thaker et al. (2019), the adjusted R^2 is significantly higher: 0.119 for the Gaussian, 0.125 for the gamma GAMM vs. 0.067 for the GAMM fitted by Thaker et al. (2019). According to the GAMM splines, the fit looks better than the previous GAMMs. Furthermore, the gamma distributed GAMM has slightly improved smoothing splines and the confidence intervals appear narrower in the dry season. The effect is mainly the same; when it gets hotter, the elephants tend to increase their movement speed up to 40° Celsius. However, at 40° Celsius and hotter, their movement speed stagnates in the wet season, whereas their speed increases in the dry season. The reason could be fewer water resources, which intuitively makes sense

Table 4.3: GAMM (Gaussian) - model selection by AIC.

Covariates/Structure omitted					AIC		
	Temperature	Season	ID	Time of day	AR(1)	Gaussian	Gamma
Candidate 1						1494092	1374824
Candidate 2	omitted					1494435	1375054
Candidate 3		omitted				1494792	1375666
Candidate 4			omitted			1495114	1376071
Candidate 5				omitted		1499142	1386532
Candidate 6					omitted	1500972	1381317

Note: For the calculation of the conditional AIC, the correction to the EDF proposed by Wood et al. (2016) was taken into account.

Source: Own representation.

and has already been mentioned by Thaker et al. (2019). Moreover, there is a steeper increase in speed in the wet season when the temperature rises. The high

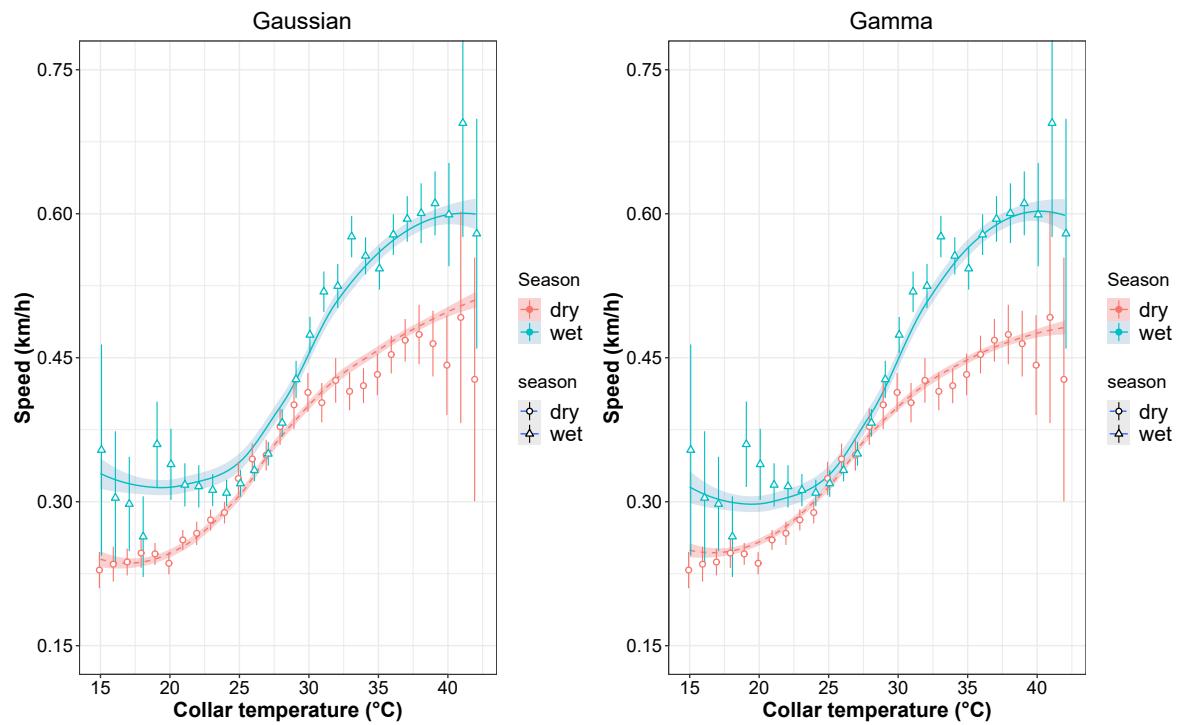


Figure 4.3: GAMMs using cyclic cubic regression splines.
Source: own representation.

velocity supports the thesis of Thaker et al. (2019) that elephants move faster

when they are near water sources due to predation risk.

When considering the ACF-plots represented in Figure 4.4, there is a little less autocorrelation left in the residuals, indicating that the cyclic implementation of the covariate ‘hour’ better explains the cyclical pattern than ‘hour’ implemented as a random effect. Besides, the process is not stationary and there is still some autocorrelation left in the residuals.

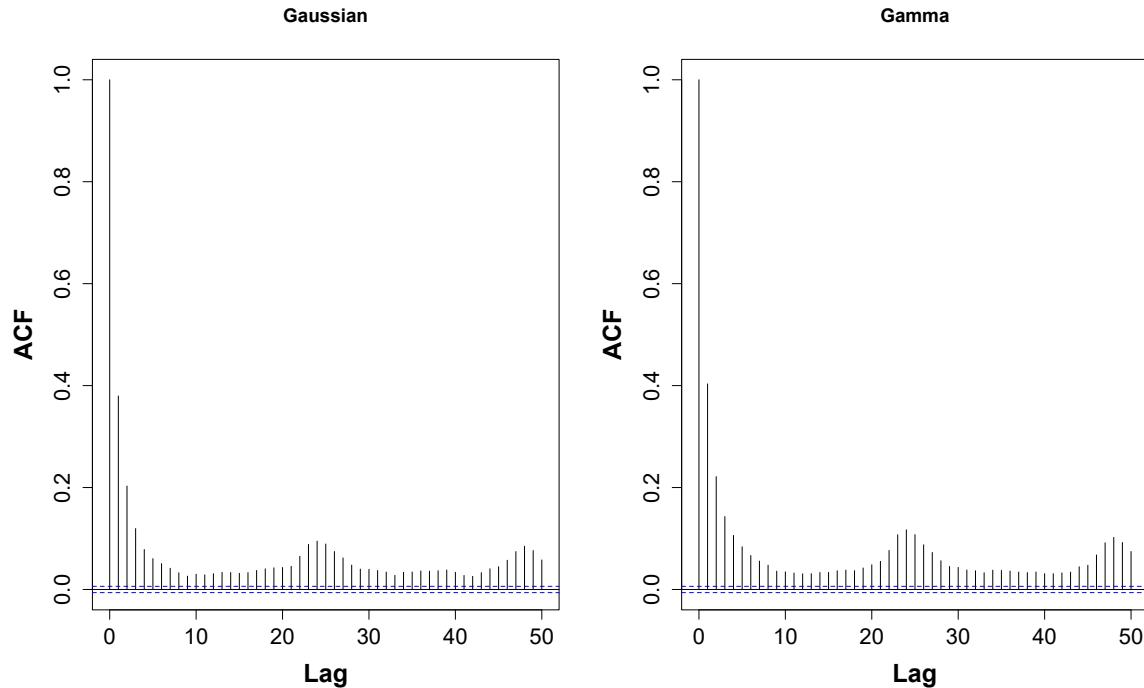


Figure 4.4: GAMMs using cyclic cubic regression splines - ACF-plots.
Source: own representation.

Another aspect that could help to further reduce the autocorrelation pattern still present in the ACFs would be to include the covariate ‘temperature’ in the CTCRW that is used to regularise the dataset. Nevertheless, the functions `crwFit` and `crwPredict` will not impute any missing covariate values. Hence, the correction must be done outside the package `crawl`. However, this procedure is beyond the scope of this work, and it is questionable whether it makes sense to impute covariate data that are missing over several days, weeks, or even months. Additionally, most sophisticated methods for regularisation, for example random forests, as described in Stekhoven and Bühlmann (2012), use other covariates in

the data to impute missing values. Beside computational demanding calculations, the only variables remaining are coordinates. Using location data to impute a covariate that is, in turn, used to model coordinates, can be misleading (moreover, as mentioned, for ‘temperature’, there is also an errors-in-variables problem).

Figure 4.5 shows quantile-quantile-plots (QQ-plots) for the fitted GAMMs. The function for generating the QQ-plot is called `qq_plot` from the package `gratia` (Simpson and Singmann, 2018). The deviance residuals are standardised and sorted for the QQ-plot illustrated in Figure 4.5. If the residuals were standard normally distributed, they would form a straight line, interfered by some random noise (Wood, 2017, chap. 7). The red line indicates this reference case. As can be seen from the deviance of the residuals, the overall fit of the GAMM is still not satisfying and improvable. However, there is a significant improvement when assuming a gamma distributed response variable. Figure 4.6 represents another

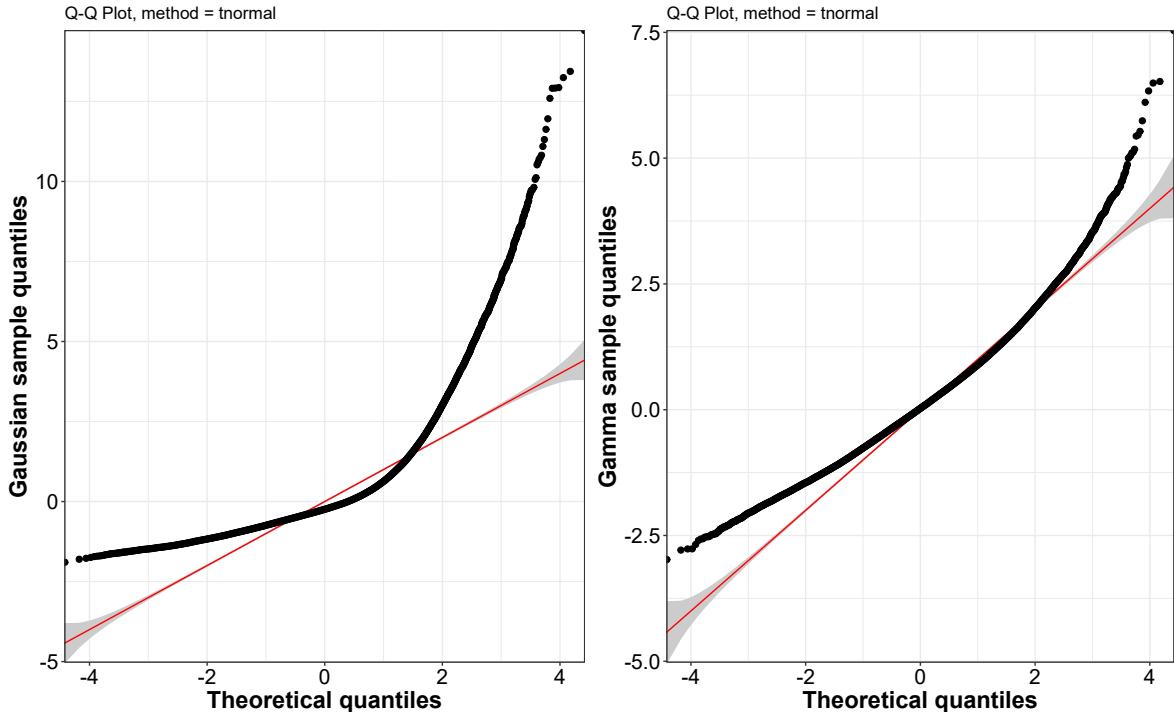


Figure 4.5: GAMMs normal QQ-plots. Gaussian GAMM left, gamma GAMM right.

Source: own representation.

view of the deviance residuals. As can be seen from the abscissa, the residuals for the gamma distributed GAMM are much less skewed than the residuals for the

normal distributed GAMM and, in fact, do not indicate a severely inadequate fit.

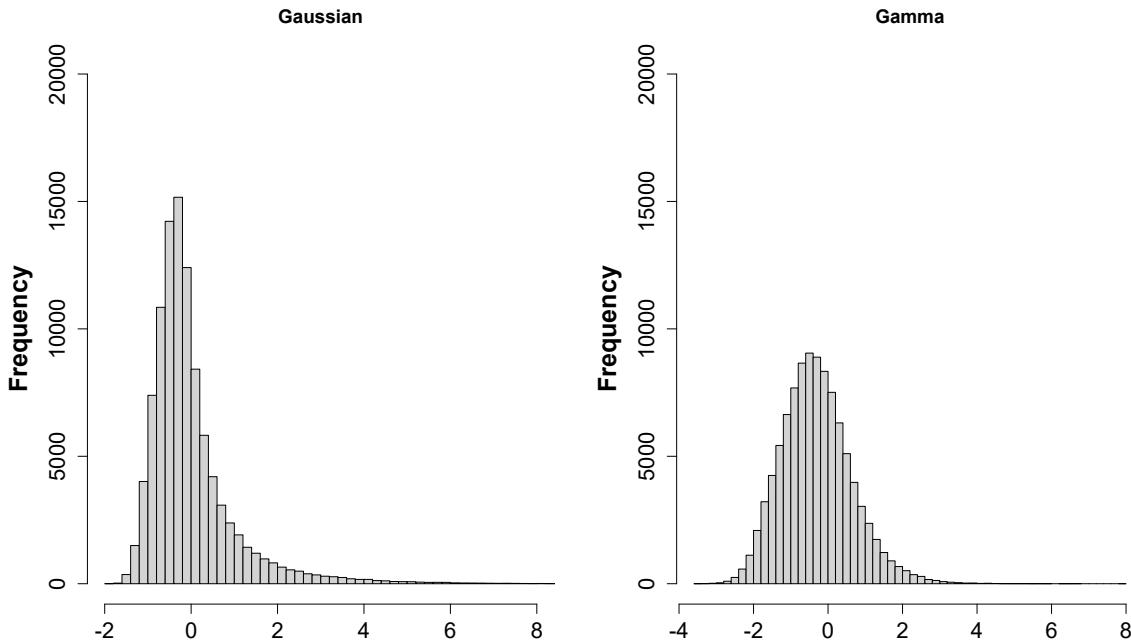


Figure 4.6: GAMM - residuals.
Source: own representation.

In summary, there are three aspects to note for the GAMM. The first aspect refers to the imputation of the dataset. Since there is a notable difference between the results of the GAMM fitted by Thaker et al. (2019) and the first GAMM in this work, it remains unclear which of the two datasets leads to better results in the sense of the real data-generating process modeled via GAMMs. The imputed dataset improves some irregularities; nonetheless, when the temporal distance between two consecutive observations becomes too large, it fails to model an appropriate path for this particular time interval and biases the calculated speed. To overcome this issue, SDEs can be used as an alternative approach, to avoid the need to impute missing data.

The second aspect concerns autocorrelation. ACF-plots might help to choose an appropriate correlation structure to be used in a GAMM. However, HMMs offer a more natural alternative to deal with autocorrelation due to their properties. Regarding the implementation of cyclic functions to model a daily pattern, this

procedure seems to improve the model quality and could also be used in HMMs.

The last aspect relates to the overall model formulation. GAM(M)s are very flexible and powerful tools, as they are relatively simple to interpret, and the regularisation helps to avoid overfitting (Larsen, 2015). Moreover, there is no need to specify any parametric distribution for the covariate effects, and in fact, movement processes are most often nonlinear (Patterson et al., 2017). In spite of that, this simultaneously implies that more caution is required when setting up a model. This includes choosing an appropriate distribution of the response variable, thoughtfully selecting splines (basis dimension, the correct choice of splines etc.) and checking for autocorrelation.

4.2 Hidden Markov Models

HMMs offer a more natural way to deal with autocorrelation since they constitute a flexible tool to handle time series data. A key issue concerning the GAMM formulation is that ecological systems are subject to specific changes among latent system states through time. The animals in ecological systems can be driven by these state dynamics and hence affect their behaviour, whereas state switching cannot be modelled using GAMMs. Using Markov chains, HMMs can accommodate complex structures that account for certain changes between latent system states and thus offer an appropriate choice for modeling animal tracking data (McClintock et al. 2020; Ephraim and Merhav 2002; Zucchini et al. 2016, chap. 1). Thus, the focus of this section is to gain further knowledge of the animals' interaction with their environment. In contrast to Section 4.1, the animals' step length and their turning angles are modeled by taking advantage of the properties of HMMs, which is to assume latent environmental states to drive the animals' behaviour. As already implied, the underlying assumptions for an HMM are regularly spaced observations in time, and negligible tracking error (Patterson et al., 2017). Both assumptions hold approximately true: The imputed dataset contains only equidistant observation times, and those are based on GPS technology, which is sufficiently accurate.

The fitted HMMs in this work are extensions of the basic HMM framework. Since the dataset contains step lengths (which can be considered a substitute for speed due to regular time intervals) and turning angles, the HMMs can be used

to model multivariate time series that will be driven by the same underlying state process. The first component series, step lengths, are assumed to follow a gamma distribution with mean μ and variance σ^2 because this has proven to be advantageous in the GAMM modeling. Moreover, step lengths cannot take on negative values. The second series component turning angles, is assumed to follow a von Mises distribution with mean μ and concentration parameter ϱ . The model formulation thus is given by

$$\begin{aligned} Y_{t1} \mid C_t = j &\sim \Gamma(\mu_j^{\text{step}}, \sigma_j^{\text{step}}), \\ Y_{t2} \mid C_t = j &\sim \text{von Mises}(\mu_j^{\text{turn}}, \varrho_j^{\text{turn}}), \quad Y_{t1} \perp Y_{t2} \mid C_j. \end{aligned} \quad (4.1)$$

Equation (4.1) describes a model that assumes contemporaneous conditional independence, meaning that the state-dependent distribution $p_j(y)$ is just the product of the corresponding marginal probabilities. However, the Markov chain still induces serial dependence and cross-dependence in the component series Y_{t1} and Y_{t2} , meaning that contemporaneous conditional independence neither implies serial nor mutual independence in the component series (Zucchini et al., 2016, chap. 9). Nonetheless, contemporaneous conditional independence makes it easier to find suitable distributions because one can select univariate distribution classes instead of finding multivariate distributions.

Additionally, the HMMs are modeled containing covariates in the state process, meaning that the transition probabilities are expressed as a function of covariates. The rationale for this extension is that it is usually of interest to relate the state-switching process to environmental covariates, in order to draw conclusions about how animals interact with their environment (Patterson et al., 2017). Among all HMMs fitted in this work, the transition probabilities are affected by the covariates 'temperature', 'hour', their interaction term, and 'season', to examine how the animals respond to these environmental factors (i.e. how state switching depends on these external factors). As shown in Figure 4.2, there is a within-day variation. In the HMM framework, within-day variation can be well modelled using trigonometric functions instead of relying on cyclic regression splines. A particular type of a so-called cosinor function will be applied to implement this (Cornelissen, 2014). The focus of the HMMs is to draw conclusions from latent state dynamics to refine the movement behaviour and reveal further insights into the shuttling behaviour of the individuals.

The first question that arises with HMMs is how many states should be chosen, when it comes to model fitting. Two HMMs were fitted with two and three states. For each candidate, the transition probabilities are functions of the covariates 'temperature', *cosinor*('hour'), their interaction term, and 'season'. All HMMs have been computed using the package `momentuHMM` (McClintock and Michelot, 2018). Table 4.4 shows AIC and Bayesian information criterion (BIC) of the corresponding number of states chosen. As can be seen, both AIC and BIC tend to select the three-state HMM. Nevertheless, caution should be taken, as AIC generally tends to favour models with numbers of states that are undesirably large in situations where states shall be meaningful entities. Most likely, this will be the case for this situation, however, the computation time for fitting candidates exceeding three states renders infeasible to appropriately check for local maxima in this work. In most cases, selecting the number of states tends to be very challenging (Pohle et al., 2017). The choice of three states is in agreement with most biologists' opinion who suggest that there should be three states: 'resting', 'foraging' and 'traveling' (Langrock et al., 2014; Patterson et al., 2008). Nonetheless, the states should be considered cautiously, as they are rather proxies for animals' true state processes (Patterson et al., 2017; Buderman et al., 2021).

Table 4.4: Model selection - number of states.

Number of states	AIC	BIC
$m = 2$	1654832	1654832
$m = 3$	1639946	1639834

Source: Own representation.

In addition, it is necessary to examine whether the transition probabilities should be affected by all available covariates or just a subset, including whether there should be any interaction term between temperature and time of day. In Table 4.5, five three-state HMMs with different functions for the transition probabilities are shown. Although it is the most complex model, according to AIC and BIC, the model with three covariates, including an interaction term for temperature and time of day fits the data best. Time of day contributes the most to model fit, as AIC is significantly higher when the covariate 'hour' is omitted. Furthermore, it is worth noting that the candidate including the covariates 'temperature' and *cosinor*('hour'), without any interaction term, fits worse than the candidate

Table 4.5: Model selection - covariates in transition probabilities.

Covariates in transition probabilities	AIC	BIC
cosinor('hour')	1641452	1641757
'temp'	1646528	1646775
'temp' + cosinor('hour')	1643667	1644029
'temp' + cosinor('hour') + interaction	1640022	1640498
'temp' + cosinor('hour') + interaction + 'season'	1639946	1639834

Source: Own representation.

including *cosinor('hour')* only, as it seems to be overly complex.

Turning to model checking, it must be evaluated how well the fitted HMM explains the data. Figure 4.7 represents the state-dependent gamma densities for the step lengths in metres for all individuals. The density displayed in orange corresponds to the 'resting' state and indicates that, given the animals are in the resting state, the expected step length per hour is about 82 metres. Likewise, the densities displayed in blue and green represent the 'foraging' and 'traveling' state. The density for the 'traveling' state is remarkably skewed due to possibly extensive step lengths. The expected step length per hour in the foraging state is 235 metres and 676 metres in the traveling state, respectively. The black dotted line shows the overall density. As the grey bars represent the observed step lengths of the animals, it can be seen that the total density fits the data quite well. The model output for the HMM can be seen in Appendix A3.

Similarly, Figure 4.8 represents the state-dependent densities for the turning angles. In the 'foraging' and 'traveling' states, the densities are concentrated around an angle of zero, whereas in the 'resting' state, the density rather tends to be uniformly distributed and less concentrated. Considering the relative size of the animals, it intuitively makes sense that the individuals tend not to turn that much when traveling more extensive distances but rather tend to turn more in the 'resting' state.

Using the definition of the state distributions in Equation (3.32), Figures 4.9 and 4.10 show the stationary state probabilities by hour, separated by season. As described in Section 3.2, the underlying state process can be assumed to be in its stationary distribution. Moreover, the state probabilities are functions of all

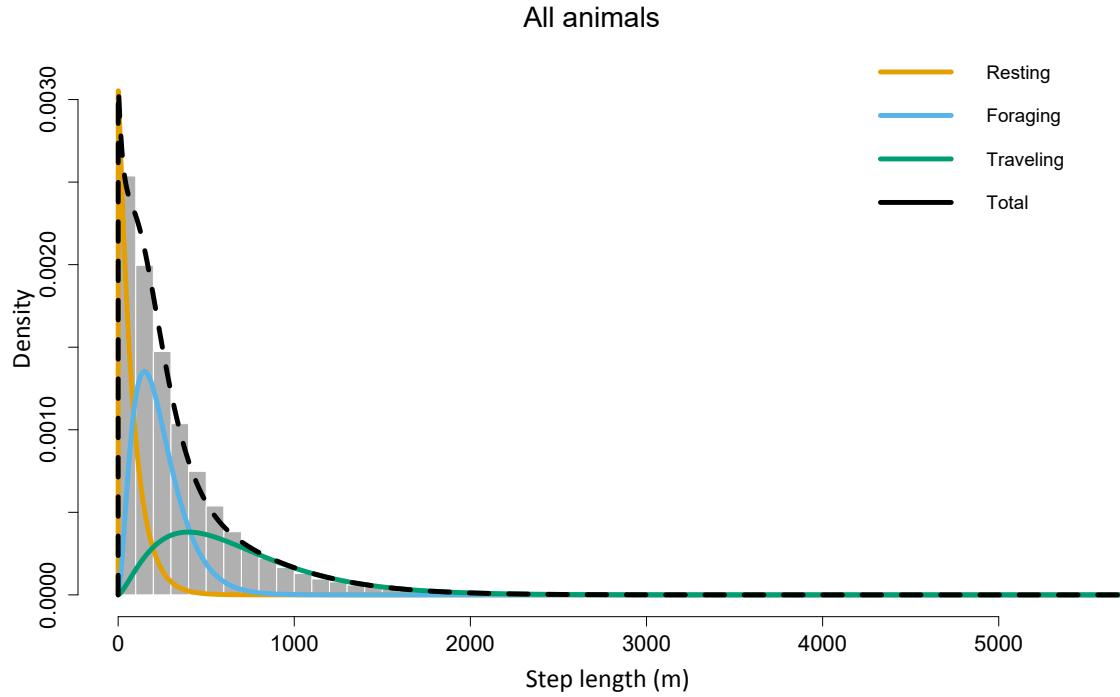


Figure 4.7: State-dependent densities for step lengths.
Source: own representation.

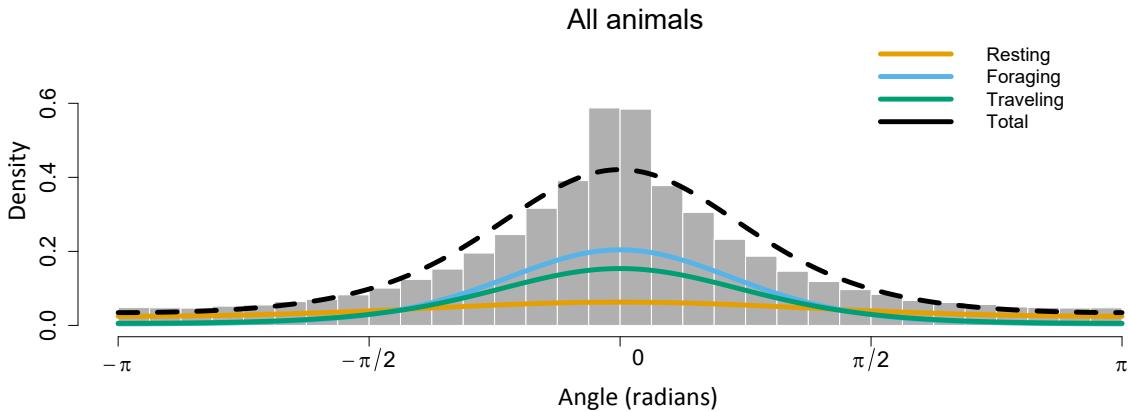


Figure 4.8: State-dependent densities for turning angles.
Source: own representation.

covariates, and thus, are multi-dimensional. However, the covariate 'temperature' is fixed to its mean value (27.48° Celsius), whereas the third dimension, 'season', separates the plots. Hence, Figure 4.9 represents the state probabilities per hour

in the wet season, for fixed temperatures. Likewise, Figure 4.10 shows the state probabilities per hour for the dry season. When considering the state

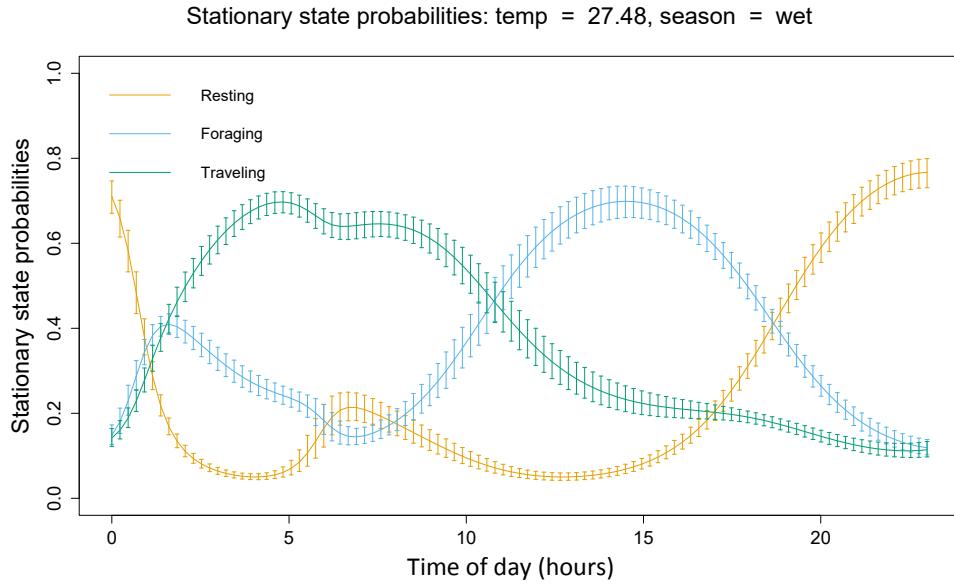


Figure 4.9: Stationary state probabilities by hour in wet season.
Source: own representation.

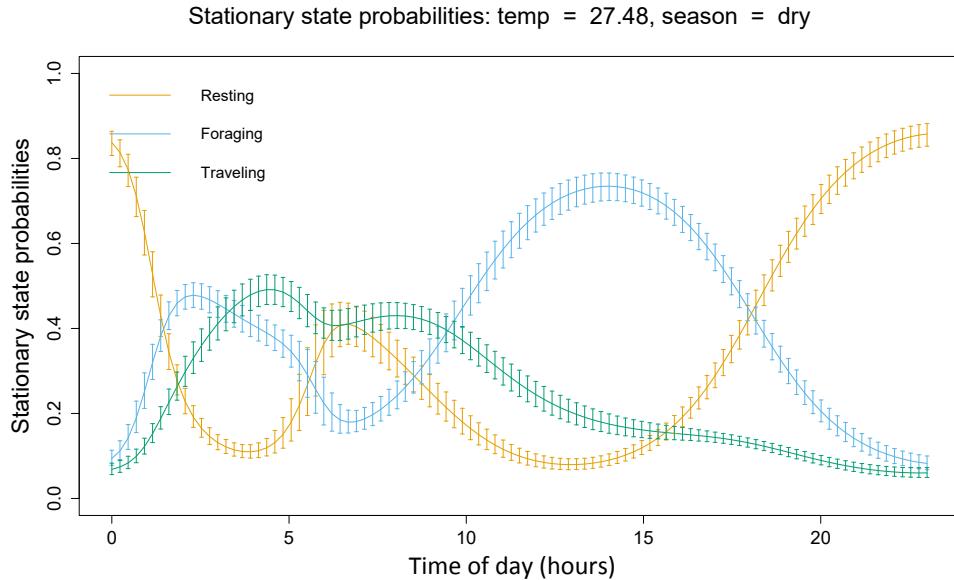


Figure 4.10: Stationary state probabilities by hour in dry season.
Source: own representation.

probabilities, a certain pattern emerges. The animals travel farther distances in the morning, forage in the evening and tend to rest at night, whereas in the wet season, the individuals are more likely to be in the ‘traveling’ state in the morning. As for the covariates ‘temperature’ and ‘hour’, they are most likely positively correlated to some extent. However, if collinearity was an aggravating issue, the transition probabilities would show wider confidence intervals as the standard errors grow.

As for the time of day, the state probabilities can also be shown by temperature, represented by Figures 4.11 and 4.12. This time, the dimension for the covariate ‘hour’ is held fixed. The plots show that the hotter it gets, the animals most likely occur to be either in the ‘foraging’ or ‘traveling’ state at noon and thus, travel farther distances. Moreover, in the wet season, the animals are even more likely to be in the ‘traveling’ state, the hotter it gets. In Appendix A3, a reference to an animated version of the stationary state probabilities can be found, allowing more profound insights into the state dynamics.

Considering the implementation of the covariate ‘season’, for reasons of complexity it has been assumed, that the season does not interact with the other covariates ‘temperature’ and ‘hour’. Nonetheless, from a biological point of view,

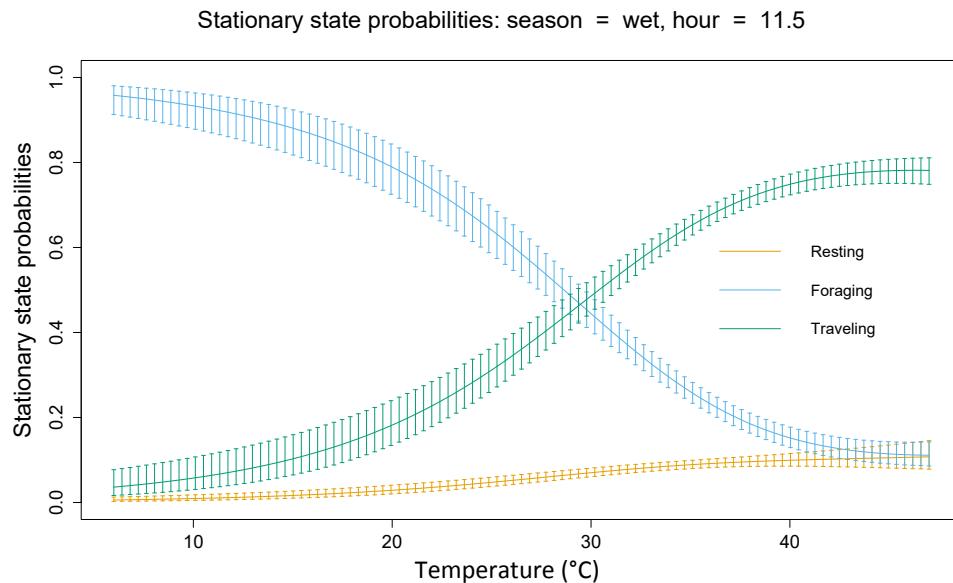


Figure 4.11: Stationary state probabilities by temperature in wet season.
Source: own representation.

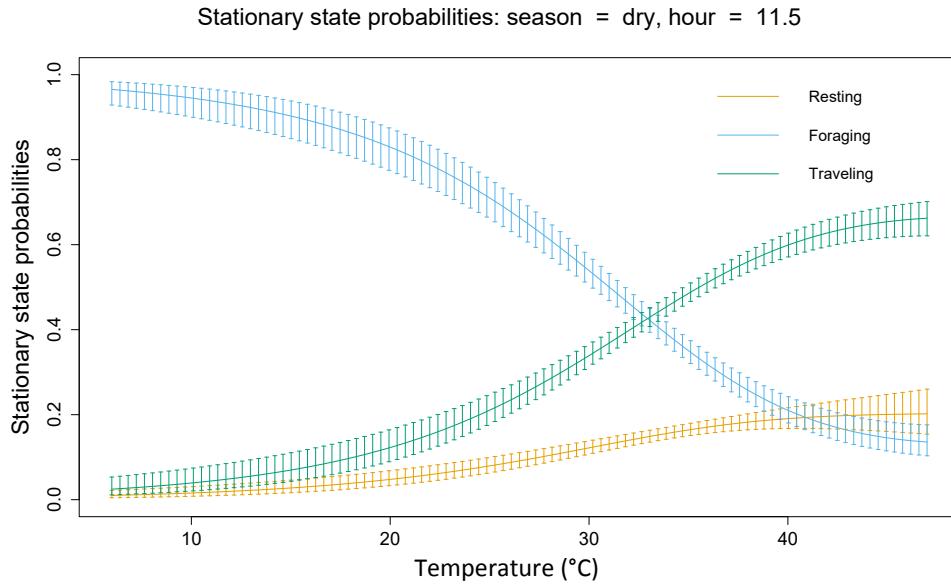


Figure 4.12: Stationary state probabilities by temperature in dry season.
Source: own representation.

it would certainly make sense to account for interactions with 'season', because otherwise it would suggest that the daily pattern and the effect of temperature do not change in both seasons. Additionally, instead of modeling 'season' as a dummy, it would be more appropriate to model it as a cyclical variable. This, however, would be beyond the scope of this work.

For a further step in model checking, Figure 4.13 graphically demonstrates the corresponding QQ-plots and the ACF-plots for the step lengths and the turning angles, respectively. The first row represents the output for the step lengths and the second row shows the output for the turning angles. As can be seen from the QQ-plot, the pseudo-residuals tend to be standard normally distributed. There seems to be a minor lack of fit in the tails, but overall, it appears to be a good approximation. Moreover, a substantial improvement can be seen, when comparing the QQ-plot with the QQ-plot of the fitted GAMM in Figure 4.5. While the ACF-plot for the turning angles does not show any autocorrelation, there is still some pattern left in the ACF for the step lengths. Comparing the ACF for the best-fitted GAMMs seen in Figure 4.4 and the step lengths ACF for the HMM, the HMM can handle the autocorrelation more suitably as there is less autocorrelation left in the residuals.

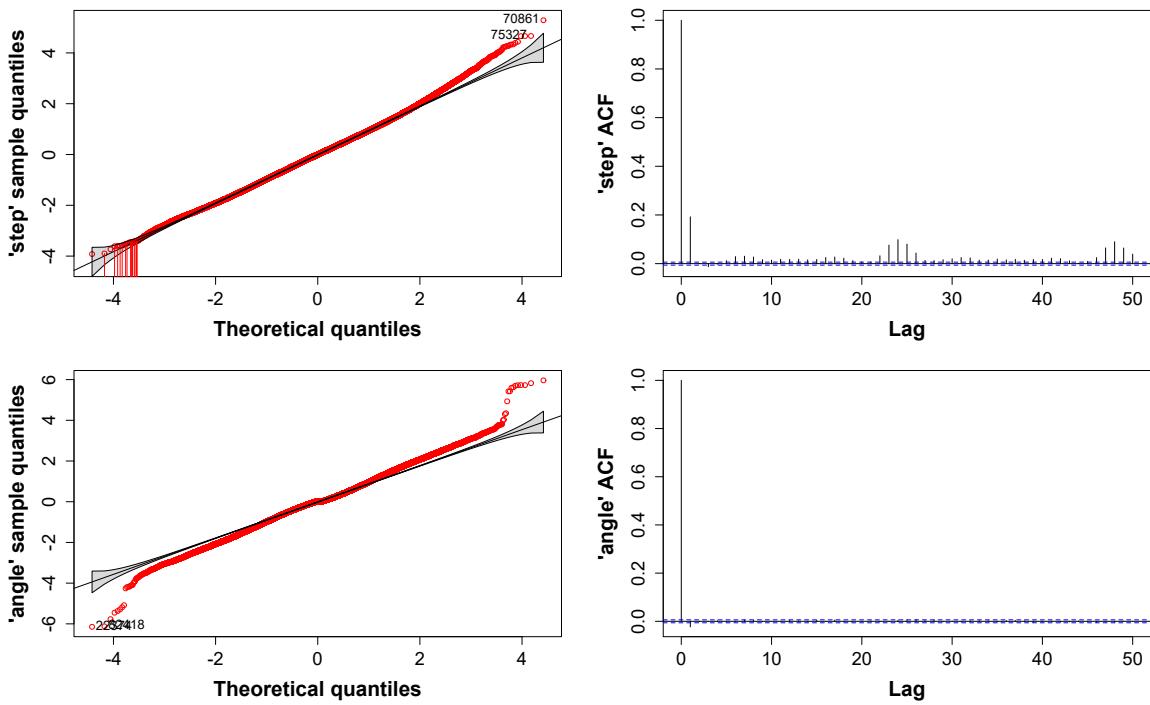


Figure 4.13: QQ- and ACF-plots of step lengths and turning angles.

Source: own representation.

All in all, the fitted HMM offers a broad variety of output and could even be tuned further. For example, random effects for the animals could be implemented, and more covariates (if accessible) could be added to the transition probabilities function. The cyclical pattern has been better captured by the HMM than the GAMM. Although there is still some room for improvement, the general fit of the HMM is better than that of the GAMM. Nonetheless, both methods, HMMs and GAMMs, have in common that they heavily rely on equidistant observations. For HMMs, the discrete-time Markov chains are not interpretable without any constant sampling unit because, based on irregular sampling, there is no way to formulate an appropriate distribution for the response variable that takes time passed into account (Patterson et al., 2017). As mentioned in Section 4.1, there still is some chance for the CTCRW to not catch every trend in the original dataset and thus, might bias the imputed dataset.

4.3 Stochastic Differential Equations

To model the relationship between movement speed and environmental covariates, an approach that does not require the imputation of data is the time-varying SDE method. Thus, a potential bias due to imputation can be excluded. Like HMMs, time-varying SDEs are models that can be applied to analyse data collected over time. The key conceptual difference is that SDEs are formulated as operating in continuous time. As already mentioned, the parameters μ and σ determine the type of SDE. The CTCRW is not only suitable for imputing movement paths, as has been seen in Section 2.2, but is also an option in this context because it takes certain inertia or autocorrelation between the increments into account. Thus, a CTCRW is used in the first place.

For the models shown in this chapter, μ and σ each depend on a time-varying parameter, i.e. $\mu(Y_t, \theta_t) = \mu(Y_t, \theta_t^{(1)})$ and $\sigma(Y_t, \theta_t) = \sigma(Y_t, \theta_t^{(2)})$, where $\theta_t = (\theta_t^{(1)}, \theta_t^{(2)})$ (see Equation (3.34) for the general reference of a time-varying SDE formulation). To be consistent with the CTCRW in Section 2.2, let $\rho_t = \theta_t^{(1)}$ and $s_t = \theta_t^{(2)}$. Since the animals move through space, there are two dimensions (longitude and latitude), and hence, the time-varying velocity process is bivariate. However, there is no need to specify a two-dimensional parameter vector because the velocity process itself is isotropic and consequently can be considered driven by the same underlying parameters (Michelot et al., 2021). In the CTCRW case, the speed, denoted as v , can be calculated via $v_t = \sqrt{\pi s_t} / (2\sqrt{\rho_t})$ (Michelot et al., 2021; Gurarie et al., 2017).³

As previous results show, the animals' movement paths are affected by temperature and time of day. Thus, in the first place, the parameters ρ and s both depend on the covariate 'temperature'. The relationship is modeled using thin-plate regression splines. All time-varying SDEs have been fitted using the `smoothSDE` package (Michelot, 2021). However, the negative log-likelihood was not finite. Consequently, it was not possible to fit a joint model for all individuals, meaning that the model failed to fit shared parameters ρ and s for the 14 individuals. Hence, adding the covariate 'hour' into the model to account for a cyclical pattern does not make sense. The failure could have several reasons. One reason might be that the elephants have too different movement patterns. Another related reason could be a too divergent reaction to temperature.

³For a detailed derivation of this claim, see Gurarie et al. (2017).

When no model fits all individuals at once, it makes sense to fit one for every 14 animals separately. Nevertheless, it was again not possible to fit a model with the entire movement path of each elephant due to the negative log-likelihood not being finite. In particular, only piecewise sequences could be fitted using the covariate 'temperature'. For example, this is shown in the following for the individual 'AM254'. The sequence is limited from May 2008 up to October 2008 and thus corresponds to the dry season.

Figure 4.14 shows the fitted parameters ρ and s . The red lines indicate 100 posterior samples. Following Nychka's (1988) idea, these posterior draws create a confidence interval for the fitted spline (black line). As can be seen, ρ increases when temperature increases. The same holds true for s , whereas a steep decrease can be observed when the temperature exceeds 35° Celsius. The decrease could be due to the fact that it might be too hot for the individual to move fast.

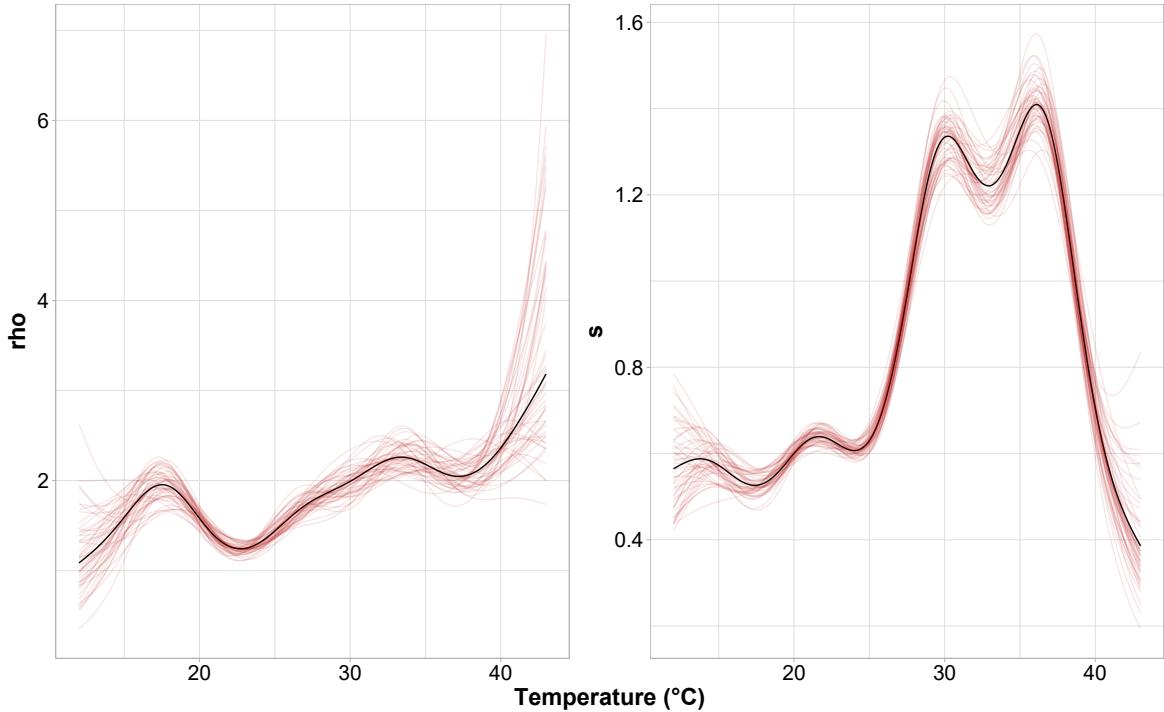


Figure 4.14: SDE parameters for AM254 by temperature.
Source: own representation.

Moreover, Figure 4.15 represents the relationship between speed and temperature. The speed, v_t , can be seen at the ordinate, whereas the temperature is shown

at the abscissa. Comparing the result of this SDE to the GAMM in Figure 4.3, the result is very similar. The hotter it gets, the faster the elephant moves. However, when it gets too hot, the elephant decreases its movement speed, as already observable in Figure 4.14.

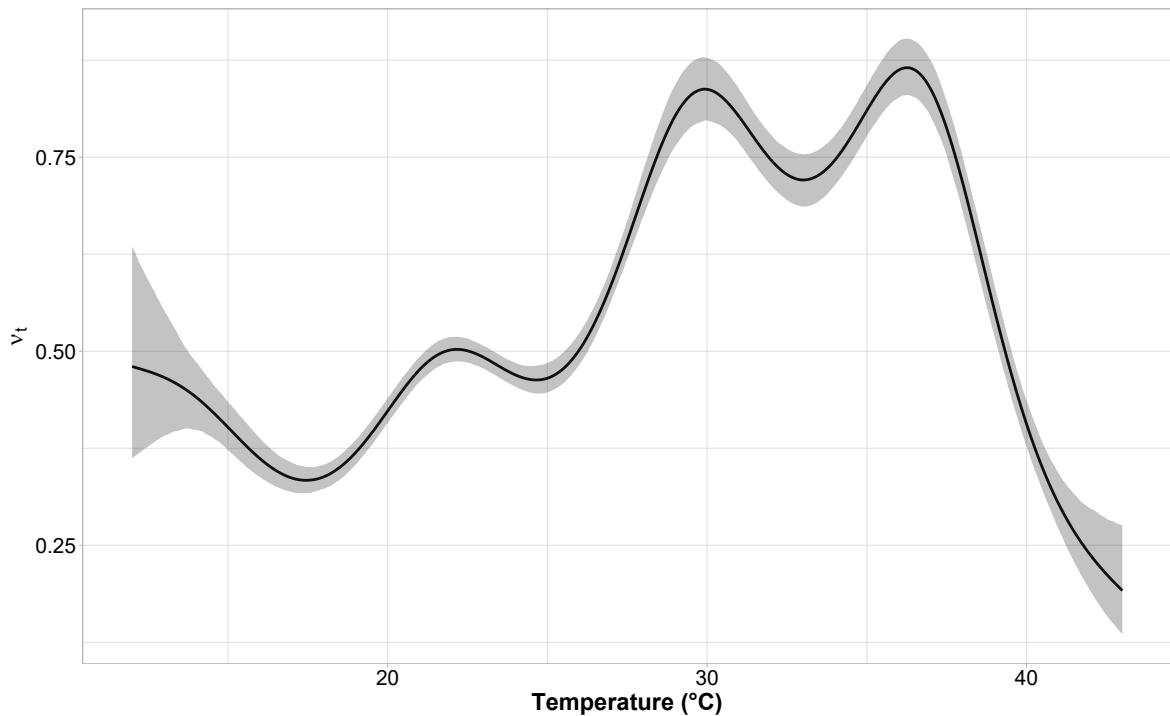


Figure 4.15: Fitted SDE AM254 by temperature.

Source: own representation.

Nonetheless, the overall result is not satisfying because the model cannot explain the entire movement path of the animal using the covariate 'temperature'. Thus, as a next step, the covariate 'hour' is used to capture the daily pattern of the animals. The parameters ρ and s are modeled using the covariate 'hour' via a cyclic P-spline. Again, a model for all 14 elephants with two shared parameters could not be fitted due to optimisation problems (negative log-likelihood is not finite). However, using this model formulation, it was possible to fit a model for each animal without limiting the movement path to a certain sequence. For example, Figure 4.16 shows the relation between speed and time of day for the animals 'AM239' and 'AM253'. The daily movement pattern is very similar to the movement pattern shown in the HMM, as can be seen in Figure 4.9. The animals

tend to be faster (and hence travel more extensively) in the early morning and around noon. All 14 fitted models for the time-varying SDEs by time of day can be seen in Appendix A4.

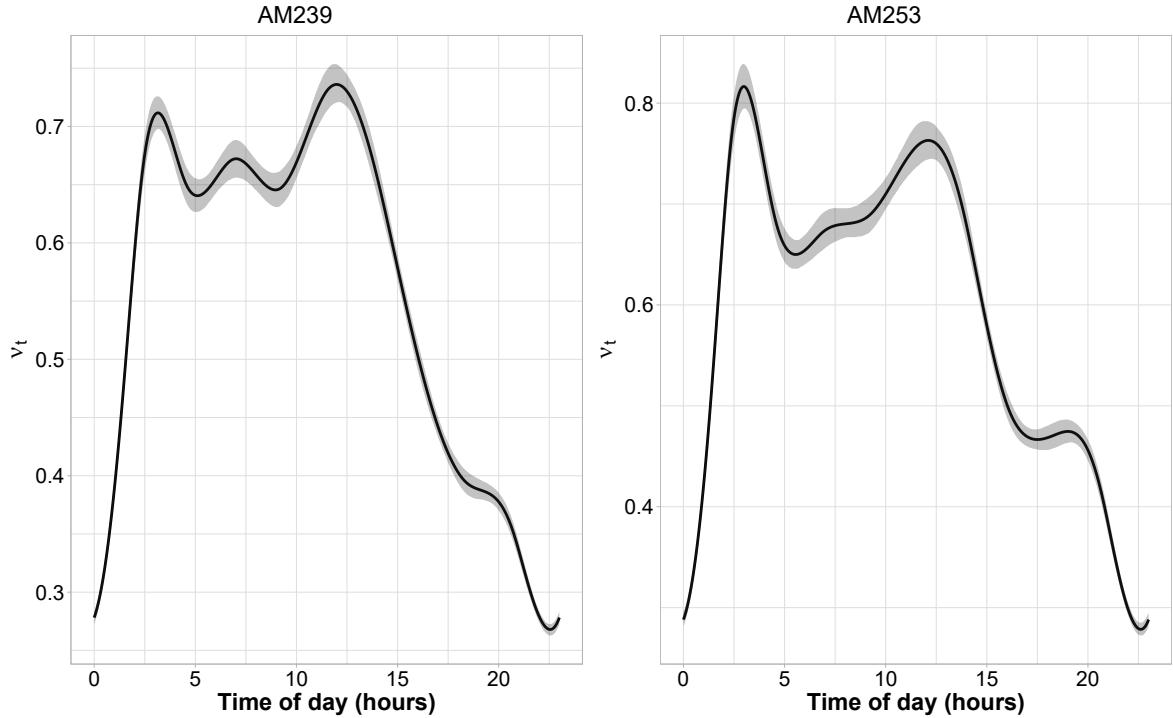


Figure 4.16: AM239 and AM253 SDE output by time of day.
Source: own representation.

Turning to model selection for the time-varying CTCRW, theoretically, there were three candidates: a time-varying SDE model with both parameters ρ and s depending on time of day and temperature, depending on temperature only, or on time of day only, respectively. Effectively, only one candidate remains, which is the parameters ρ and s , to depend on time of day only because it is the only model that could be fitted (which actually is not even one single model either). Nonetheless, because the covariate is modeled via a cyclic P-spline, a basis parameter k has to be chosen for each model. This can be achieved by formulating multiple time-varying SDEs with a varying number of basis dimensions for the P-spline. Consistent with the GAMM model selection, the conditional AIC will be used, following the correction by Wood et al. (2016) in the EDF.

Table 4.6 shows different parameter values of k for individual 'AM239' as an

example. As can be seen, according to the conditional AIC, $k = 10$ leads to the best fit. For $k = 15$, the negative log-likelihood is not finite, and thus, no parameter estimation took place.

Table 4.6: SDE model selection - basis functions.

Parameter value k	AIC
$k = 5$	-108340.0
$k = 10$	-109908.9
$k = 15$	NA

Source: Own representation.

Considering the results up to this point, due to correlated increments, a CTCRW offers the best characteristics for animal movement. However, the fitting is very problematic. An alternative would be to specify a time-varying Brownian motion to fit a model that combines all 14 individuals, although the increments of Brownian motion processes are independent. Again, to be consistent with the Brownian motion formulation in Section 3.3, let $r_t = \theta_t^{(1)}$ and $s_t = \theta_t^{(2)}$. The parameter r_t can often be fixed to zero, and s , the diffusion parameter, will be estimated as a smooth function of temperature or time of day, respectively. In this model, the distribution of $Y_{t+\Delta}$ is normal around Y_t , and s^2 is proportional to the variance of the distribution, as can be seen in Equation (3.38). Consequently, even though the interpretation is not as favourable as the CTCRW, s can be interpreted as a measure of speed, and so the effects of temperature and time of day can be analysed.

In fact, when using Brownian motion, a model including all 14 animals could be fitted. Table 4.7 basically shows the model selection for two approaches: one approach to model the relationship of speed using the covariate 'temperature' (again, via a thin-plate regression spline), the other approach using time of day (covariate 'hour', modeled using a cyclic regression spline). Using both covariates in one model did not work. According to the conditional AIC, using $k = 10$ basis functions and the covariate 'hour' turned out to be the best model fit.

Figure 4.17 graphically represents the fitted time-varying Brownian motion model, where the parameters depend on time of day.

Table 4.7: Brownian motion - model selection by AIC.

Number of basis functions	Conditional AIC		
	Parameter k	Temperature	Time of day
$k = 5$		-9231.2	-51536.5
$k = 10$		-9307.1	-59936.0
$k = 15$		NA	NA

Note: For the calculation of the conditional AIC, the correction to the EDF proposed by Wood et al. (2016) was taken into account.

Source: Own representation.

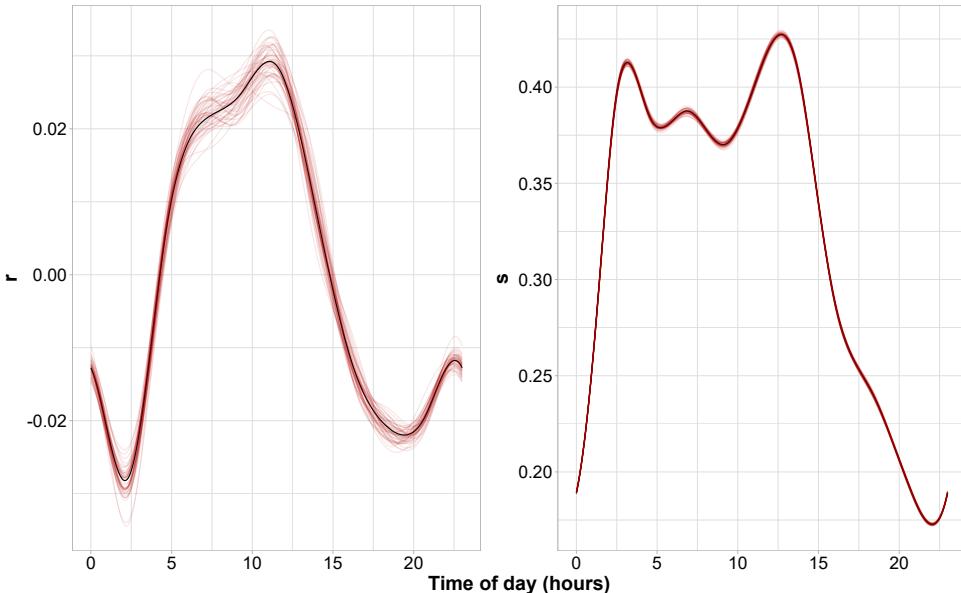


Figure 4.17: Fitted time-varying Brownian motion by time of day.
Source: own representation.

As in Figure 4.14, the red lines indicate a confidence interval via 100 posterior draws. As already mentioned, r does not have a significant influence, as it tends to be around zero, whereas s can be interpreted as a measure of speed. The result is very similar to the CTCRW single fits, indicating that the elephants tend to move faster in the morning and at noon. However, Brownian motion will not capture as many realistic features of animal movement as the CTCRW: There is a trade-off between realism and flexibility on the one side, and computational stability and speed on the other side.

For model validation, it still is an open question on finding suitable, general methods for model checking in this context. A simple diagnostic to investigate the goodness-of-fit in time-varying SDE models is based on the Euler-Maruyama discretisation to calculate the residuals:

$$\varepsilon_i = \frac{y_{i+1} - (y_i + \mu(y_i, \hat{\theta}_{t_i})\Delta_i)}{\sigma(y_i, \hat{\theta}_{t_i})\sqrt{\Delta_i}}, \quad (4.2)$$

with $i = 1, \dots, n-1$ and $\hat{\theta}_{t_i}$ as an estimate of θ_t over $[t_i, t_{i+1}]$. According to the model assumptions, the residuals should be independent and approximately follow a standard normal distribution (Michelot et al., 2021). Figure 4.18 shows a QQ-plot of the residuals against the standard normal distribution. The residuals were calculated using the formula represented in Equation (4.2). As can be seen, the residuals are severely skewed and hence, there is evidence that the fit is not sufficient. This can be confirmed when looking at the ACF of the residuals

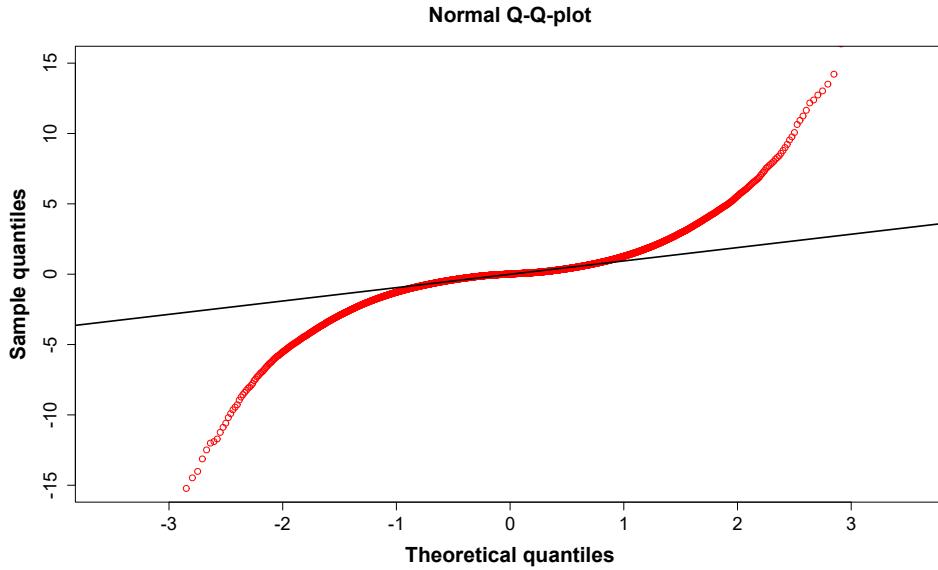


Figure 4.18: QQ-plot for fitted time-varying Brownian motion model.
Source: own representation.

represented in Figure 4.19. The Figure illustrates positive autocorrelation in the residuals for several hours, turning into a cyclical pattern. The autocorrelation indicates the model did not capture some degree of inertia. As already mentioned, per definition, the increments of a Brownian motion are independent. Conse-

quently, it would be adequate to use a CTCRW to account for correlation between the increments. However, a CTCRW could not be fitted containing all individuals in one model.

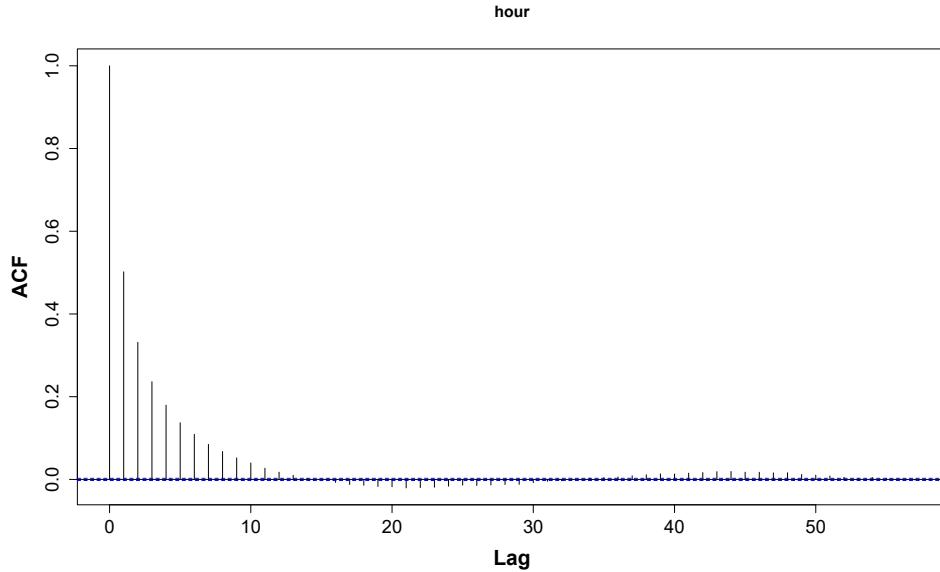


Figure 4.19: ACF-plot for fitted time-varying Brownian motion model.
Source: own representation.

Summarized, time-varying SDEs offer an intuitive way to analyse tracking data. In particular, a time-varying CTCRW has ideal characteristics for modeling animal movement. Moreover, time-varying SDEs do not depend on regular observation times, making them a natural choice for irregular tracking data while not being too simplistic, as ordinary SDEs are said to be (Michelot et al., 2021; Michelot and Blackwell, 2019). The option to implement splines and random effects in the parameters offers a compelling way to investigate the influence of certain covariates on the process.

5 Summary and Discussion

This work intended to empirically compare GAMMs, HMMs and time-varying SDEs in the context of analysing animal tracking data. In the previous chapter, the results of the analysis using the three methods were presented. In this chapter, a comparison of the methods will be given. To evaluate the methods, the comparison is based on specific criteria. For the purpose of an aggregated point of view, the main results of this work will briefly be summarized.

As a preliminary step, in Section 2.2, the irregular dataset of 14 tracked elephants was imputed to get equidistant observation times. This is critical when using methods such as GAMMs and HMMs, which Thaker et al. (2019) ignored when using GAMMs and LMMs. The GAMM fitted by Thaker et al. (2019) was modified in Section 4.1 as that the response variable (movement speed) is assumed to follow a gamma distribution instead of a Gaussian distribution, and time of day should preferably be modeled as a cyclical regression spline rather than a random effect, which significantly improves the model fit. Furthermore, autocorrelation was incorporated because movement patterns typically show strong evidence for autocorrelation (Patterson et al., 2017).

As a next step, in Section 4.2, HMMs were used as a more natural choice of incorporating autocorrelation in the model, as HMMs are flexible time series models. The HMMs modeled the step length and turning angles of the individuals depending on three ecological states (resting, foraging, and traveling), whereas the transition probabilities depend on the environmental factors, temperature and time of day. The HMM output revealed additional information on the behaviour of the individuals, such as a concentrated turning angle of zero degrees in the resting and foraging states indicating that the animals tend to move in straight directions, whereas, in the resting state, the individuals tend to have no attraction point in the sense of turning angles. Furthermore, the state dynamics depending on temperature and time of day reveal deep insights into the elephants' movement behaviour, such as traveling more extensive distances at night and early morning

and around noon, foraging in the early evening and resting in the early night.

As pointed out in Section 4.2, a drawback of the two previous methods is that both rely on equidistant observation times. Another tool that can be used to analyse data collected over time, and a natural choice for modeling irregular tracking data, are SDEs, which were used in Section 3.3. Because SDEs themselves are considered to be too simple to model environmental influences in movement behaviour, following the approach by Michelot et al. (2021) leads to an extension called time-varying SDEs. The most significant advantage of time-varying SDEs is that they do not need regular data and can handle irregularly spaced observations. Furthermore, the SDE parameters can depend on covariates and random effects, which are implemented using smoothing splines. Although the time-varying SDEs did not work quite as good as expected, they confirm the findings of the HMMs and GAMMs in a more mechanistic sense.

Comparing the three different methods in a statistical sense can be very expansive. Thus, in the following, the three methods will be compared according to key criteria for animal tracking data: comprehensibility, computational efficiency and suitability. Afterwards, the main results of the applied methods will be discussed.

Starting with GAMMs, this method is not a time series tool in the conventional sense. As has been pointed out, tracking data often include autocorrelation. As shown in Section 4.1, autocorrelation can be modeled using GAMMs; however, an explicit correlation structure must be provided because otherwise, the observations are assumed to be independent. Providing a correlation structure can be a challenging task. Regarding comprehensibility, GAMMs offer many options to model all kinds of characteristics and are very intuitive. In particular, a broad variety of different distribution assumptions for the response variable are available, covariate effects can be modeled via smoothing splines, and correlation structures can be used to investigate any patterns in the data.

Turning to computational efficiency, basic GAMMs are not very demanding due to clever optimisation routines. However, when the dataset gets larger, and AR(1) processes are insufficient for modeling autocorrelation, computations can be very challenging and demanding. In this work, even AR(2) processes took several hours, whereas AR(3) processes were not feasible with 16GB DDR4 RAM because the correlation matrices calculations demanded too much RAM.

Conclusively, in the sense of suitability, GAMMs are a very powerful tool but should be considered an additional, rather than a core tool for analysing tracking

data. In the first place, correlation must be explicitly modeled and can be computationally very demanding. This is reflected in the model checking of the fitted GAMMs, as the ACF and QQ plots suggest a lack of fit for AR(1) processes.

The second method, HMMs, can accommodate more complex structures in the data, such as latent system changes through time, which is achieved by Markov chains. The HMM framework offers lots of extensions to fit the data accurately. For instance, if the dwell time in the states might not appear geometrically distributed, semi-HMMs can be used to assume different dwell-time distributions, for example by assuming a shifted Poisson distribution. Moreover, random effects can be incorporated, and transition probabilities to be affected by covariates. The possibility to incorporate ecological states via Markov chains to analyse latent system changes, is the key advantage of HMMs. This feature adds further realism and is very intuitive, particularly when cooperating with biologists to study the movement behaviour of animals.

Contrary to some statements in the literature, direct maximisation of the likelihood expressed in (3.26) is indeed feasible, thanks to a recursive scheme using forward probabilities. Hence, there is need for complex estimation methods such as MCMC or EM-algorithms. Nonetheless, compared to GAMMs and SDEs, HMMs took the longest time to fit.

When considering suitability, compared to the other approaches in this work, HMMs are most likely the best method to analyse movement data for two reasons. The first reason concerns incorporating latent states via Markov chains, and the second reason is the great variety of possibilities to analyse tracking data. Although the data must be regularly spaced and possible tracking errors negligible, HMMs offer a natural choice for ecological analyses of animal movement behaviour.

The third method, time-varying SDEs, were recently described by Michelot et al. (2021) and can incorporate covariates and random effects to model SDE parameters via smoothing splines. In comparison to simply using GAMMs for modeling tracking data, the time-varying SDE approach is more mechanistic than the other methods, in the sense that it describes the mechanisms behind animal movement behaviour. Further, the variable of interest is modelled by the SDE, whereas splines and random effects model the parameters of the SDE. This is quite different from using GAMMs to model the variable of interest via splines and random effects directly, and thus, this is a more sophisticated method than simply using

GAMMs. However, time-varying SDEs do not require equally spaced observation times, making them a reasonable choice for modeling irregular tracking data.

Considering computational efficiency, model fitting for the time-varying SDEs only took a few minutes. However, as mentioned, there were difficulties in model fitting. Some model candidates, such as a temperature effect on the SDE parameters, could not be adequately fitted when using CTCRWs. Moreover, for CTCRWs, there is no model to include all 14 individuals in the dataset. Only Brownian motion leads to a model including all animals, whereas the increments of Brownian motion are not correlated, making this candidate too simplistic.

Concludingly, time-varying SDEs are suitable for animal tracking data with irregularly spaced observation times. Although they offer a more mechanistic way of investigating movement behaviour, there are still some open questions, such as finding a general method for model checking and validation, and computationally more stable estimation routines.

In terms of the output, all methods basically show similar results. The covariates temperature and time of day significantly influence the movement pattern of the individuals. Comparing the GAMMs to the time-varying SDEs, the output indicates that there is a cyclical movement behaviour and an increase in speed when temperature rises. However, the time-varying SDEs reveal a considerable decrease in speed when it gets too hot. For the GAMMs, the speed rather slightly decreases. Although it would be quite unlikely, this might be a bias caused by imputing the dataset. In contrast to time-varying SDEs and GAMMs, the HMM reveals a deeper insight into this movement pattern. Ecological states can explain many behaviours in their movement, as the transition probabilities in Figures 4.9, 4.10, 4.11, and 4.12, and the distributions for the step lengths and turning angles in Figures 4.7 and 4.8 illustrate. In particular, considering the transition probabilities in Figure 4.9, the covariate 'hour' shows, for a given temperature, how elephants switch their movement behaviour. In particular, the animals tend to travel farther distances in the morning, forage in the evening and rest at night. Figures 4.9, 4.10, 4.11, and 4.12 additionally show that the animals are more likely to be in the 'traveling' state in the wet season, meaning that, on average, they travel farther distances in the morning and around noon. This proposition supports the findings of the gamma distributed GAMM in Figure 4.3 and the thesis of Thaker et al. (2019). The reason is that there are more water sources in the wet season and elephants increase their speed when they are near water sources due to pre-

dation risk. Moreover, the turning angles reveal that the elephants tend to have a relatively concentrated movement direction when they are in the 'foraging' and 'traveling' state. Comparing the QQ-plots of the methods, according to Figures 4.5, 4.13, and 4.18, the HMM has the best fit of the three methods. Furthermore, the animated state probabilities in Appendix A3 show the dynamics of this shuttling behaviour.

6 Conclusion

To sum up, all three methods are valuable tools and can offer deep insights into tracking data. However, GAMMs are rather an additional method because they are no time series tool in the conventional sense and it is challenging to adequately account for autocorrelation, whereas HMMs lie at the heart of analysing tracking data. Further, time-varying SDEs revealed a more mechanistic pattern in the way the animals responded to environmental factors. In contrast, due to latent state changes, HMMs offered the most valuable output in this analysis and revealed major dynamics and dependencies in the movement behaviour of the tracked 14 individuals. This makes HMMs a favourable choice to investigate the movement pattern of animals using tracking data.

The analysis results show that elephants are strongly affected by temperature and time of day. As the GAMMs have shown, there are seasonable differences in speed distribution. Moreover, the fitted HMM offers a detailed insight into the state dynamics affected by temperature, hour, and season. The HMM reveals more profound insights into the shuttling behaviour of savanna elephants due to state transition dynamics induced by Markov chains. Conclusively, further statements about the elephants' habitat use and movement behaviour in Kruger National Park can be made. In particular, the park management can derive possible adjustments such as the need for water sources in a sufficient number for these animals. Individual-level effects of environmental factors can be further investigated by specifying random effects in the HMMs. However, this will most likely turn the HMM more complex for minor improvements only.

Furthermore, this work also opens up to discuss certain topics. Regarding the imputed dataset, the selected time interval was one hour. Based on the Nyquist-Shannon sampling theorem, a resolution of δt is adequate to model behaviour and interactions that last $2\delta t$ or longer (Shannon, 1949; Nathan et al., 2022). Although the frequency is sufficient to capture temperature effects and the described latent state dynamics, it could be investigated whether a finer resolution (e.g. 30 min

intervals) would lead to new insights. According to Nathan et al. (2022), potential dynamics and interactions with the environment can be found with finer resolution which could not be captured using a wider range of time intervals.

Considering time-varying SDEs, they are a recent and promising approach that needs some more discussion and attention. A general approach to validate the models needs to be implemented when it comes to model checking. This, in turn, allows for more in-depth comparisons of time-varying SDE methods. Nevertheless, the maximisation of the log-likelihood of the SDEs turned out to be quite unstable in certain circumstances, whereas the exact reasons need to be analysed before model checking in the conventional sense can take place. Another aspect concerns the combination of time-varying SDE methods and Hidden Markov models. A mixture of a more mechanistic approach on the one hand and, the possibility to incorporate state dynamics on the other hand, might be a reasonable model approach to advance the exciting topic of movement ecology further and to reveal new findings.

Bibliography

- Albert, P. S. (1991). A Two-State Markov Mixture Model for a Time Series of Epileptic Seizure Counts. *Biometrics*, pages 1371–1381.
- Birkett, P. J., Vanak, A. T., Muggeo, V. M., Ferreira, S. M., and Slotow, R. (2012). Animal Perception of Seasonal Thresholds: Changes in Elephant Movement in Relation to Rainfall Patterns. *PLoS one*, 7(6):e38363.
- Blackwell, P. (1997). Random Diffusion Models for Animal Movement. *Ecological Modelling*, 100(1-3):87–102.
- Buderman, F. E., Gingery, T. M., Diefenbach, D. R., Gigliotti, L. C., Begley-Miller, D., McDill, M. M., Wallingford, B. D., Rosenberry, C. S., and Drohan, P. J. (2021). Caution is Warranted When Using Animal Space-Use and Movement to Infer Behavioral States. *Movement Ecology*, 9(1):1–12.
- Cornelissen, G. (2014). Cosinor-Based Rhythmometry. *Theoretical Biology and Medical Modelling*, 11(1):1–24.
- DeRuiter, S. L., Langrock, R., Skirbutas, T., Goldbogen, J. A., Calambokidis, J., Friedlaender, A. S., and Southall, B. L. (2017). A Multivariate Mixed Hidden Markov Model for Blue Whale Behaviour and Responses to Sound Exposure. *The Annals of Applied Statistics*, 11(1):362–392.
- Ephraim, Y. and Merhav, N. (2002). Hidden Markov Processes. *IEEE Transactions on Information Theory*, 48(6):1518–1569.
- Green, P. J. (1984). Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(2):149–170.
- Greven, S. and Kneib, T. (2010). On the Behaviour of Marginal and Conditional AIC in Linear Mixed Models. *Biometrika*, 97(4):773–789.

- Gurarie, E., Fleming, C. H., Fagan, W. F., Laidre, K. L., Hernández-Pliego, J., and Ovaskainen, O. (2017). Correlated Velocity Models as a Fundamental Unit of Animal Movement: Synthesis and Applications. *Movement Ecology*, 5(1):1–18.
- Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Wiley Online Library, 1st edition.
- Hastie, T. and Tibshirani, R. (1993). Varying-Coefficient Models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4):757–779.
- Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd edition.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer, 2nd edition.
- Johnson, D. and London, J. (2018). *crawl: Fit Continuous-Time Correlated Random Walk Models to Animal Movement Data*. R Package Version 2.2.1. <https://cran.r-project.org/web/packages/crawl/index.html>, Accessed: 2022-06-06.
- Johnson, D. S., London, J. M., Lea, M.-A., and Durban, J. W. (2008). Continuous-Time Correlated Random Walk Model for Animal Telemetry Data. *Ecology*, 89(5):1208–1215.
- Jonsen, I. D., Myers, R. A., and Flemming, J. M. (2003). Meta-Analysis of Animal Movement Using State-Space Models. *Ecology*, 84(11):3055–3063.
- Karatzas, I. and Shreve, S. (2012). *Brownian Motion and Stochastic Calculus*. Springer Science & Business Media, 2nd edition.
- Langrock, R., Hopcraft, J. G. C., Blackwell, P. G., Goodall, V., King, R., Niu, M., Patterson, T. A., Pedersen, M. W., Skarin, A., and Schick, R. S. (2014). Modelling Group Dynamic Animal Movement. *Methods in Ecology and Evolution*, 5(2):190–199.
- Langrock, R., Kneib, T., Sohn, A., and DeRuiter, S. L. (2015). Nonparametric Inference in Hidden Markov Models Using P-Splines. *Biometrics*, 71(2):520–528.

- Larsen, K. (2015). *GAM: The Predictive Modeling Silver Bullet.* <https://multithreaded.stitchfix.com/assets/files/gam.pdf>, Accessed: 2022-06-06.
- McClintock, B. T., Langrock, R., Gimenez, O., Cam, E., Borchers, D. L., Glennie, R., and Patterson, T. A. (2020). Uncovering Ecological State Dynamics with Hidden Markov Models. *Ecology Letters*, 23(12):1878–1903.
- McClintock, B. T. and Michelot, T. (2018). momentuHMM: R Package for Generalized Hidden Markov Models of Animal Movement. *Methods in Ecology and Evolution*, 9(6):1518–1530.
- McCulloch, C. E., Searle, S. R., and Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models*. John Wiley & Sons, 2nd edition.
- Michelot, T. (2021). *smoothSDE: Varying-Coefficient Stochastic Differential Equations*. <https://github.com/TheoMichelot/smoothSDE/>, Accessed: 2022-06-06.
- Michelot, T. and Blackwell, P. G. (2019). State-Switching Continuous-Time Correlated Random Walks. *Methods in Ecology and Evolution*, 10(5):637–649.
- Michelot, T., Glennie, R., Harris, C., and Thomas, L. (2021). Varying-Coefficient Stochastic Differential Equations with Applications in Ecology. *Journal of Agricultural, Biological and Environmental Statistics*, 26(3):446–463.
- Michelot, T., Langrock, R., and Patterson, T. A. (2016). moveHMM: An R Package for the Statistical Modelling of Animal Movement Data Using Hidden Markov Models. *Methods in Ecology and Evolution*, 7(11):1308–1315.
- Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E., and Fryxell, J. M. (2004). Extracting More out of Relocation Data: Building Movement Models as Mixtures of Random Walks. *Ecology*, 85(9):2436–2445.
- Nathan, R., Monk, C. T., Arlinghaus, R., Adam, T., Alós, J., Assaf, M., Baktoft, H., Beardsworth, C. E., Bertram, M. G., Bijleveld, A. I., et al. (2022). Big-Data Approaches Lead to an Increased Understanding of the Ecology of Animal Movement. *Science*, 375(6582):eabg1780.
- Ng, V. K. and Cribbie, R. A. (2017). Using the Gamma Generalized Linear Model for Modeling Continuous, Skewed and Heteroscedastic Outcomes in Psychology. *Current Psychology*, 36(2):225–235.

- Nychka, D. (1988). Bayesian Confidence Intervals for Smoothing Splines. *Journal of the American Statistical Association*, 83(404):1134–1143.
- Patterson, T. A., Basson, M., Bravington, M. V., and Gunn, J. S. (2009). Classifying Movement Behaviour in Relation to Environmental Conditions Using Hidden Markov Models. *Journal of Animal Ecology*, 78(6):1113–1123.
- Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., and King, R. (2017). Statistical Modelling of Individual Animal Movement: An Overview of Key Methods and a Discussion of Practical Challenges. *AStA Advances in Statistical Analysis*, 101(4):399–438.
- 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.
- Pohle, J., Langrock, R., van Beest, F. M., and Schmidt, N. M. (2017). Selecting the Number of States in Hidden Markov Models: Pragmatic Solutions Illustrated Using Animal Movement. *Journal of Agricultural, Biological and Environmental Statistics*, 22(3):270–293.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Shannon, C. E. (1949). Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1):10–21.
- 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)*, 47(1):1–21.
- Simpson, G. L. and Singmann, H. (2018). *gratia: Ggplot-Based Graphics and Other Useful Functions for GAMs Fitted Using mgcv*. R Package Version 0.7.2. <https://cran.r-project.org/web/packages/gratia/gratia.pdf>, Accessed: 2022-06-06.
- Slotow, R., Thaker, M., and Vanak, A. T. (2019). Data from: Fine-Scale Tracking of Ambient Temperature and Movement Reveals Shuttling Behavior of Elephants to Water. Movebank Data Repository. <https://www.datarepository.movebank.org/handle/10255/move.980>. Accessed: 2022-06-06.

- Stekhoven, D. J. and Bühlmann, P. (2012). MissForest—Non-Parametric Missing Value Imputation for Mixed-Type Data. *Bioinformatics*, 28(1):112–118.
- Thaker, M., Gupte, P. R., Prins, H. H., Slotow, R., and Vanak, A. T. (2019). Fine-Scale Tracking of Ambient Temperature and Movement Reveals Shuttling Behavior of Elephants to Water. *Frontiers in Ecology and Evolution*, 7(11):1–12.
- Viswanathan, G. M., 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. (1983). Bayesian “Confidence Intervals” for the Cross-Validated Smoothing Spline. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(1):133–150.
- Wood, S. N. (2004). Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models. *Journal of the American Statistical Association*, 99(467):673–686.
- Wood, S. N. (2006). Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models. *Biometrics*, 62(4):1025–1036.
- Wood, S. N. (2015). *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. R Package Version 1.40.* <https://cran.r-project.org/web/packages/mgcv/index.html>, Accessed: 2022-06-06.
- Wood, S. N. (2017). *Generalized Additive Models*. Chapman and Hall/CRC, 2nd edition.
- Wood, S. N., Pya, N., and Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, 111(516):1548–1563.
- Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series*. Chapman and Hall/CRC, 2nd edition.

Appendix

A1 R-Output - Generalised Additive Models

GAMM (Family = gaussian) Output by Thaker et al. (2019):

```
Family: gaussian
Link function: identity

Formula:
v ~ s(temp, k = 4) + season + woody.density + s(id, bs = "re") +
  s(hour, bs = "re")

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  252.55538   5.20079  48.56 <2e-16 ***
seasonwet     16.48085   0.86423  19.07 <2e-16 ***
woody.density -1.60256   0.03278 -48.89 <2e-16 ***
---
Approximate significance of smooth terms:
        edf Ref.df      F p-value
s(temp)  2.995     3 4716.1 <2e-16 ***
s(id)    12.890    13 131.2 <2e-16 ***
s(hour)   0.991     1 112.6 <2e-16 ***
---
Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '*' 0.1 '.' 1

R-sq.(adj) =  0.0667  Deviance explained = 6.68%
fREML = 1.9511e+06  Scale est. = 52787      n = 284572
```

GAMM (Family = gaussian) Output (Meyer), hour modeled as random effect:

```

Family: gaussian
Link function: identity

Formula:
v ~ s(temp, k = 10, bs = "cp") + season + s(id, bs = "re") +
    s(hour, bs = "re")

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  363.582     12.807   28.39 <2e-16 ***
seasonwet     56.291      3.203   17.58 <2e-16 ***
---
Approximate significance of smooth terms:
          edf Ref.df      F p-value
s(temp)    7.2468      9 766.35 <2e-16 ***
s(id)      12.6282     13 70.97 <2e-16 ***
s(hour)    0.9691      1  32.20 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0687  Deviance explained = 6.89%
fREML = 7.4956e+05  Scale est. = 1.6732e+05 n = 100908

Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

          k'      edf k-index p-value
s(temp)  9.000  7.356     0.99     0.38
s(id)    14.000 12.674     NA       NA
s(hour)   1.000  0.979     0.95 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

GAMM (Family = gaussian) Output (Meyer), hour modeled as a cyclic effect:

```

Family: gaussian
Link function: identity

Formula:
v ~ s(temp, k = 10, bs = "ps") + season + s(id, bs = "re") +
    s(hour, bs = "cc")

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  338.017    12.967   26.07  <2e-16 ***
seasonwet     86.242      3.241   26.61  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
        edf Ref.df      F p-value
s(temp)  4.430  5.113 43.62 <2e-16 ***
s(id)    12.666 13.000 80.00 <2e-16 ***
s(hour)   7.959  8.000 706.19 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.118  Deviance explained = 11.9%
fREML = 7.4708e+05  Scale est. = 1.5927e+05 n = 100908

Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

          k'    edf k-index p-value
s(temp)  9.00  4.43     0.99    0.27
s(id)    14.00 12.67      NA      NA
s(hour)   8.00  7.96     1.02    0.90
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

GAMM (Family = gamma) Output (Meyer), hour modeled as a cyclic effect:

```

Family: Gamma
Link function: log

Formula:
v ~ s(temp, k = 40, bs = "ps") + season + s(id, bs = "re") +
    s(hour, k = 50, bs = "cp")

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.77989   0.03064 188.67  <2e-16 ***
seasonwet    0.19791   0.01013  19.53  <2e-16 ***
---
Approximate significance of smooth terms:
        edf Ref.df      F p-value
s(temp)  4.706  6.029 16.07  <2e-16 ***
s(id)    12.399 13.000 43.12  <2e-16 ***
s(hour) 20.534 22.000 326.12 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.125  Deviance explained = 16.2%
fREML = 1.4006e+05  Scale est. = 1.1531    n = 100908

Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

      k'     edf k-index p-value
s(temp) 39.00  4.71     0.89     0.19
s(id)   14.00 12.40      NA       NA
s(hour) 49.00 20.53     0.91     0.69
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

A2 Dimension checking for generalised additive models

As can be seen in the output files of the GAMMs in Appendix A1, the basis dimension (k) can be checked in a more sophisticated way than described in Section 4.1. An estimate of the residual variance based on differencing residuals is calculated to check whether k is appropriate. The residuals are (near) neighbours according to the (numeric) covariates of the smooth (Wood, 2017, chap. 5). When the calculated estimate of the residual variance is divided by the residual variance, one obtains the k -index as another measure. The more the k -index is below 1, the more likely the smoothing function has not captured some pattern in the data. This can be further quantified using a corresponding p-value. If the p-value indicates significance, k is most likely too low. However, the reference criterion for basis dimension checking is the EDF, a summary statistic that reflects the degree of non-linearity of a curve (Wood, 2017, chap. 6). In particular, the EDF should not be close to the basis dimensions.

For example, below is the output of the gamma GAMM fitted in Section 4.1. The k -index is less than 1, indicating that the smoothing function has not accurately captured some pattern. However, neither is the EDF close to k nor does the p-value indicate a significant underfit. Hence, the basis dimensions are appropriately chosen.

```
Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.
```

```
      k'    edf k-index p-value
s(temp) 39.00  4.71     0.89    0.19
s(id)   14.00 12.40      NA      NA
s(hour) 49.00 20.53     0.91    0.69
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A3 R-Output - Hidden Markov Models and animated state probabilities

HMM Output

```
Value of the maximum log-likelihood: -819917.2
step parameters:
-----
          Resting      Foraging     Traveling
mean      8.151080e+01 2.353738e+02 6.758789e+02
sd       8.100429e+01 1.423925e+02 4.331321e+02
zeromass 3.674101e-04 1.030025e-04 2.755791e-10

angle parameters:
-----
          Resting      Foraging     Traveling
mean      0.0000000 0.0000000 0.0000000
concentration 0.4693616 1.941718 1.643107

Regression coeffs for the transition probabilities:
-----
          1 -> 2      1 -> 3      2 -> 1      2 -> 3      3 -> 1      3 -> 2
(Intercept) -16.29656531 -3.0364978034 -4.11178174 -9.5064205 -1.92540038 -2.47442661
temp         0.04720145  0.0284744541  0.06670049  0.2170761 -0.07350824  0.04094871
seasonwet    0.93817709  0.4418381383 -0.01974100  0.5577786 -0.72698442 -0.40158076
cosinorCos(hour, 24) 5.79892789 -0.4758372824  4.10197774  4.3378357  2.39026188  1.38848208
cosinorSin(hour, 24) 17.93943142 -0.1741747251 -1.92562887  7.4514809 -1.35819064 -1.27475437
temp:cosinorCos(hour, 24) 0.15570393 -0.0311099860 -0.09302325 -0.1026706 -0.12178907 -0.01037393
temp:cosinorSin(hour, 24) -0.21762425  0.0002640021  0.06452041 -0.1945576  0.10356150  0.01639786

Transition probability matrix (based on mean covariate values):
-----
          Resting      Foraging     Traveling
Resting   0.62587722 1.073227e-10 0.37412278
Foraging  0.02052911 9.651448e-01 0.01432612
Traveling 0.02711921 4.827779e-02 0.92460300
Initial distribution:
-----
          Resting      Foraging     Traveling
9.999996e-01 7.325220e-09 4.409586e-07
```

Animated Versions of the Stationary State Probabilities

Stationary state probabilities by hour:

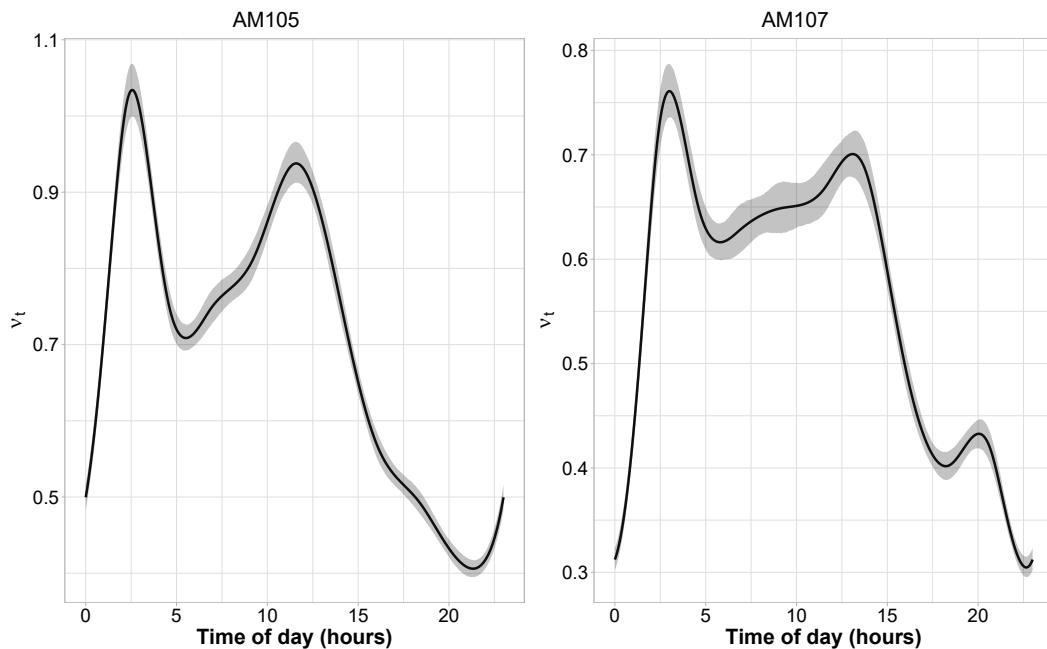
https://timom2110.github.io/Elephants-Moving-Data/stationary_state_probabilities_both_seasons_hour.html

Stationary state probabilities by temperature:

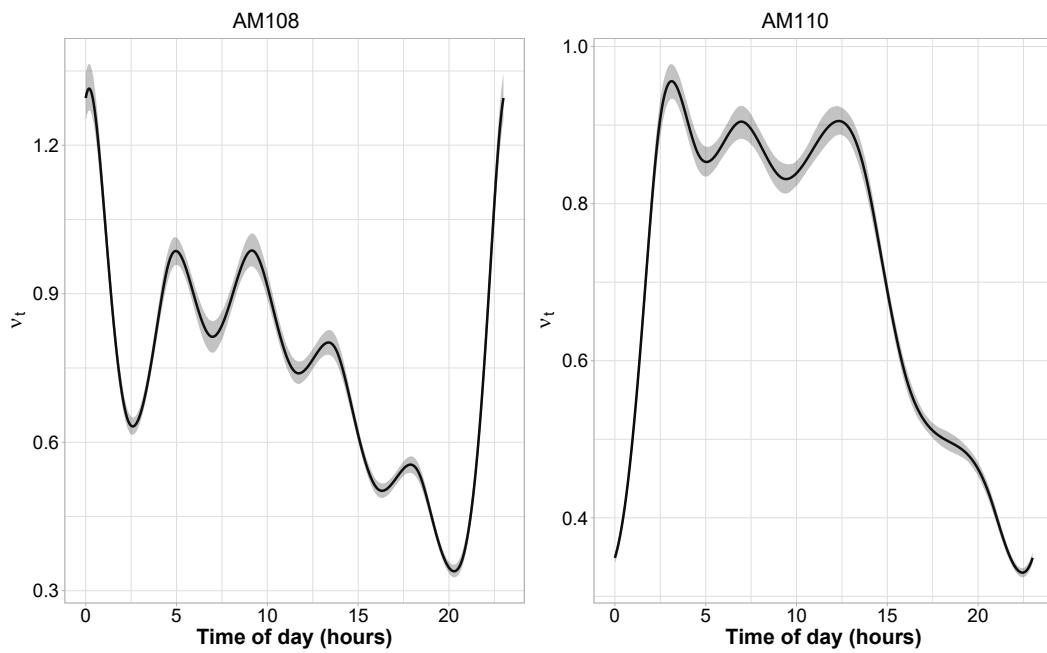
https://timom2110.github.io/Elephants-Moving-Data/stationary_state_probabilities_both_seasons_temperature.html

A4 R-Output - Stochastic Differential Equations

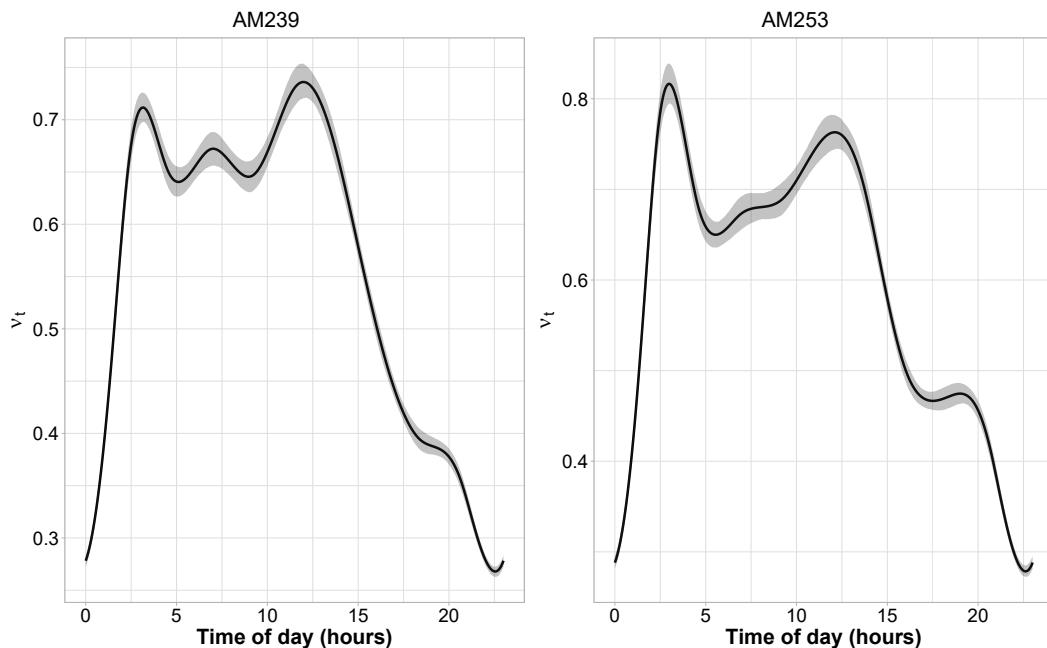
All time-varying SDE models. Parameters ρ and σ depending on covariate 'hour', fitted as a cyclic spline.



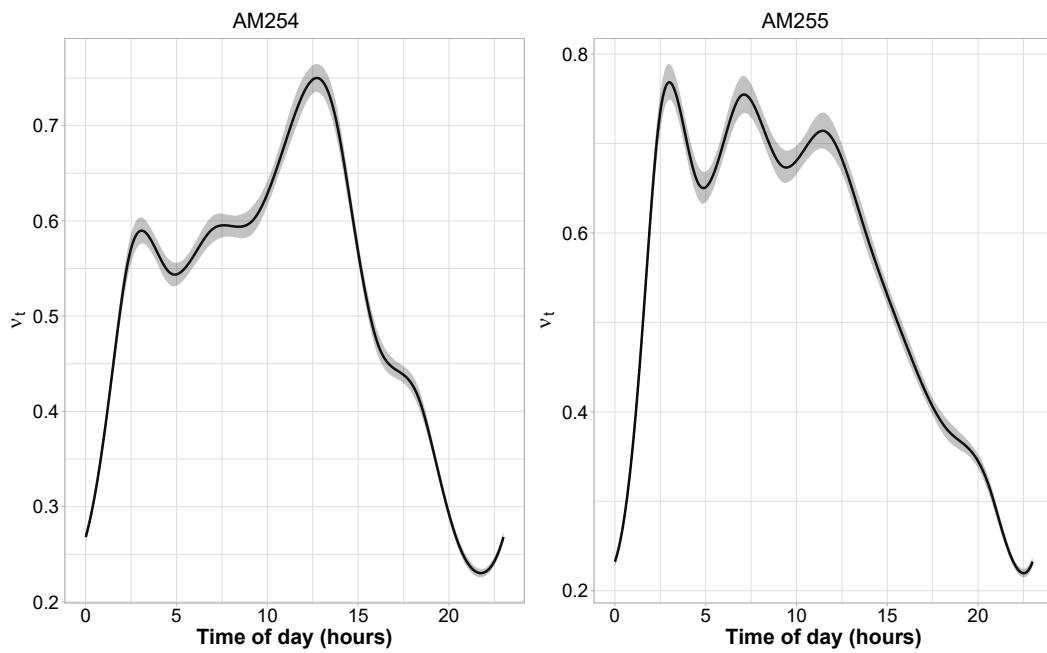
AM105 and AM107 SDE output by 'hour'
Source: own representation.



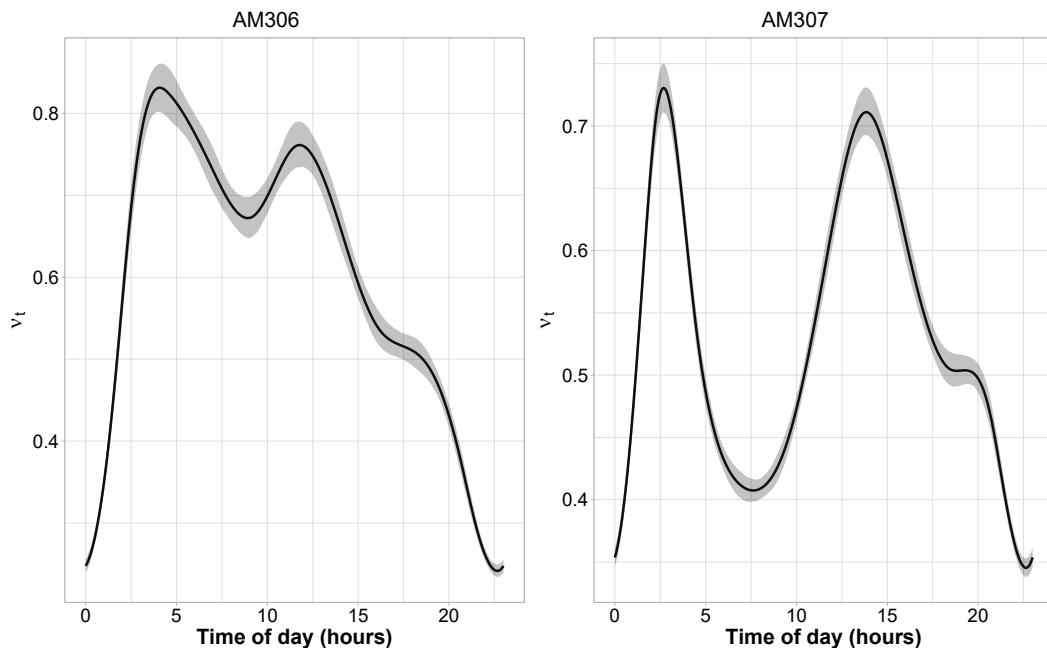
AM108 and AM110 SDE output by 'hour'
Source: own representation.



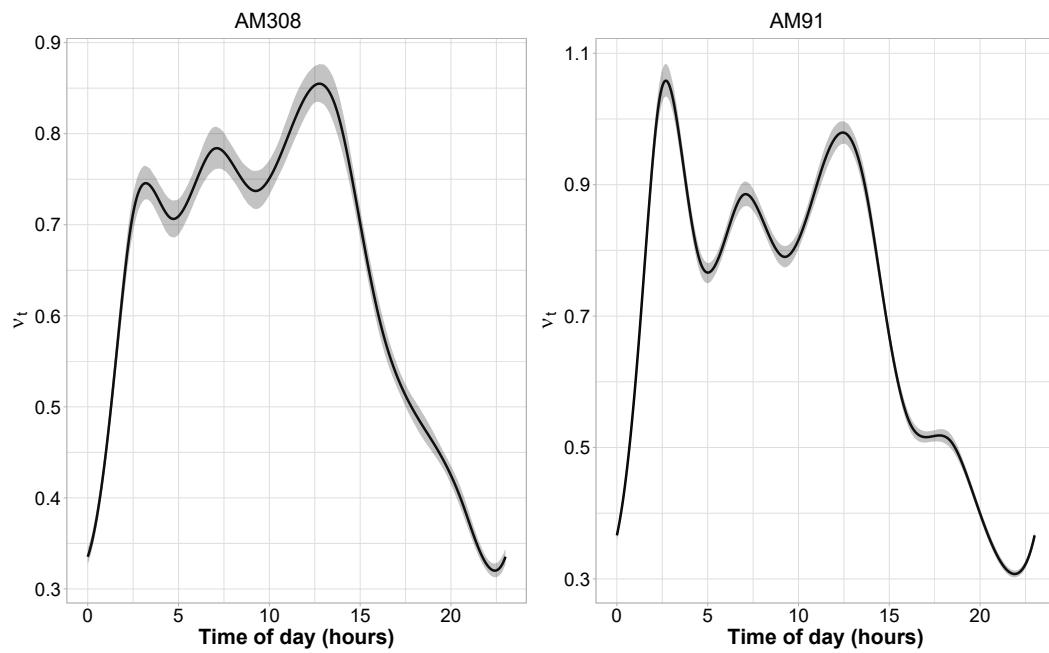
AM239 and AM253 SDE output by 'hour'
Source: own representation.



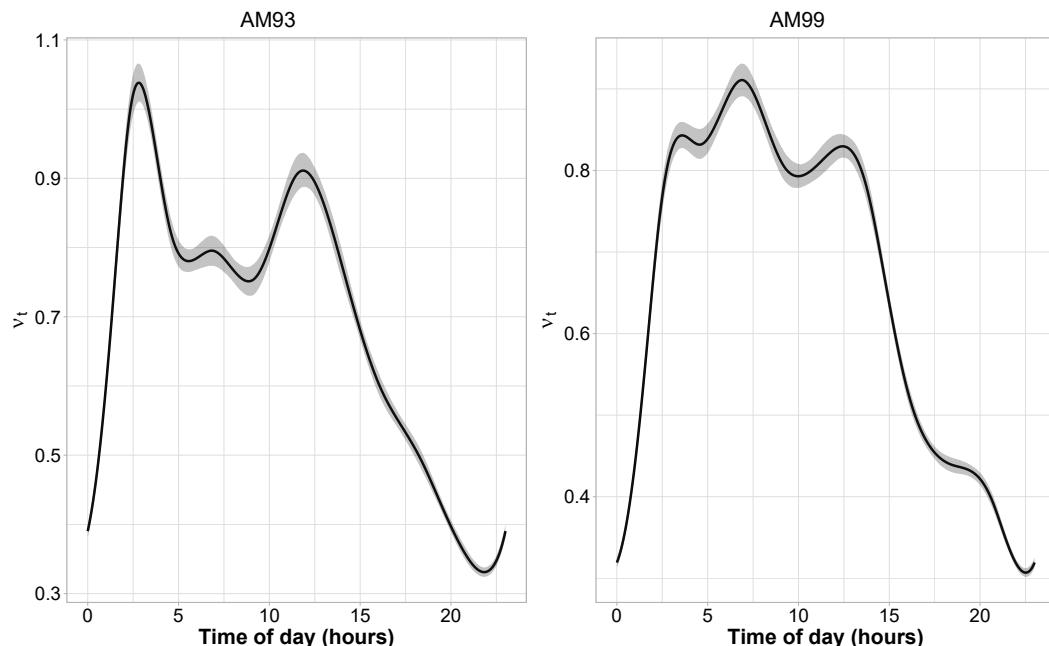
AM254 and AM255 SDE output by 'hour'
Source: own representation.



AM306 and AM307 SDE output by 'hour'
Source: own representation.



AM308 and AM91 SDE output by 'hour'
Source: own representation.



AM93 and AM99 SDE output by 'hour'
Source: own representation.

Versicherung

Name: Meyer

Vorname: Timo

Ich versichere, dass ich diese Masterarbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe.

Die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen habe ich als solche kenntlich gemacht.

Diese Versicherung gilt auch für alle gelieferten Datensätze, Zeichnungen, Skizzen oder grafischen Darstellungen.

Des Weiteren versichere ich, dass ich das Merkblatt zum Umgang mit Plagiaten (<http://phoenix.wiwi.uni-bielefeld.de/organisation/pamt/uploads/PlagiatInfo-BlattStudenten.pdf>) gelesen habe.

Bielefeld, den 29ten April, 2022

Unterschrift