

EXPLORING MODEL SELECTION IN ANIMAL MOVEMENT
HIDDEN MARKOV MODELS

by

Mark A. McTaggart

Submitted in partial fulfillment of the requirements
for the degree of Bachelor of Science (Honours) in Statistics

at

Dalhousie University
Halifax, Nova Scotia
April 2024

© Copyright by Mark A. McTaggart, 2024

Table of Contents

Abstract	iii
Acknowledgements	iv
Chapter 1 Introduction	1
1.1 Hidden Markov models	1
1.2 Animal movement ecology	3
1.3 Hidden Markov models for animal movement	3
1.4 About the wild boar data set and using moveHMM	7
Chapter 2 Comparing the two-state, three-state, and four-state models	12
2.1 Features of the two-state, three-state, and four-state models	12
2.2 Model checking	13
2.3 Choosing the number of states	15
Chapter 3 Examining environmental and temporal factors	23
3.1 Choosing which covariates to include	23
3.2 Examining the effects of the covariates on the stationary probabilities	24
Chapter 4 Conclusion	28
Bibliography	29

Abstract

In this paper, we provide a brief explanation of what hidden Markov models are, and their application to animal movement. Then, we present some hidden Markov models that were fitted using the R package `moveHMM`, based on location and time data of wild boars from the south of France. This data set was provided to us by Simon Chamaille-Jammes. We explore how to select the best number of states for an animal movement hidden Markov model using a combination of information criteria, state probability metrics, and pseudo-residual plots. Next, we identify which temporal and environmental covariates to include in our hidden Markov model. The model we identified as best for our data set was a two-state hidden Markov model with two temporal covariates and three environmental covariates. An analysis of this model indicates that wild boar movement can be described sufficiently using two states: a less active (or “resting” state) and a more active state. An analysis of the model’s covariates gives clear indication that the boars that were sampled are nocturnal. The model also indicates that the boars prefer to rest in forested areas and tend to be more active in non-forested areas.

Acknowledgements

I would like to thank Théo Michelot for taking me on as his first honours student, and for carving time out of his busy schedule every week to meet with me and help me with this project. I would like to thank him for loaning me his copy of Walter Zucchini, Iain L. MacDonald, and Roland Langrock's *Hidden Markov Models for Time Series*, which I cite extensively in this paper, for such a long period of time (almost a year?). I would also like to acknowledge the huge amount of time he has put into developing the moveHMM package in R and making it so user-friendly. I am beginning to get a sense of how of big a contribution this package is to the field of movement ecology.

Chapter 1

Introduction

1.1 Hidden Markov models

Hidden Markov Models are a type of statistical model used for time series data i.e. data that is recorded sequentially in time. They are a good tool to use when we believe that the variables we're interested in behave differently at different times. In particular, a hidden Markov model is apt when we believe these underlying differences in behaviour can be sorted into a discrete number of "states". By thinking about it in that way, we can say that the variable of interest behaves in such-and-such a way when its "state 1", and it behaves in a different way when its in "state 2", and it behaves in yet another way in "state 3", and so on. We can also estimate the probability that the underlying process will remain in the state its already in, or change states, at the next observation time. A hidden Markov model consists of two main parts: an "observed" process, and, as the name implies, a "hidden" process. The theory of hidden Markov models (HMMs), is that the observed process (the observation at a given time), is determined by what's going on in the "hidden" process, which is the underlying state that changes over time [9]. In HMMs, The number of different hidden states that are included in the model is something that must be decided on before fitting the model. For example, you could decide to have a 2-state model, a 5-state model, or a 20-state model. In theory, you can fit a 1-state model, but that would mean that each observation would come from the same distribution, so in that case it would not really make sense to think of it as fitting a hidden Markov model, because that would mean there is no heterogeneity in the model and no dependence in time, which are two of the main reasons for using the hidden Markov model in the first place.

The "Markov" part of a hidden Markov model comes from the fact that the hidden process is a Markov Chain, meaning that if C_t is the state at time t (we often use a whole number to refer to a state, i.e. C_t could be 1,2, or 3 at a given time for a

3-state HMM), then $Pr(C_t|C_{t-1}, C_{t-2}, C_{t-3}, \dots) = Pr(C_t|C_{t-1})$. This means that which state the model is in at time t only depends on what the state was at time $t-1$ [9].

The probabilities that the process will stay in the same state or change to another state can be summed up succinctly by a transition probability matrix, which is often denoted as Γ . The matrix below is an example of a transition probability matrix for a two-state hidden Markov model.

$$\Gamma = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$$

This matrix indicates that if the process is in state 1, then it has a probability of 0.8 of being in state 1 at the next observation and a probability of 0.2 of being in state 2 at the next observation. If the process is in state 2 at a given time, then it has a probability of 0.7 of staying in state 2 at the next observation and a probability of 0.3 of switching to state 1 at the next observation. For a transition probability matrix of a hidden Markov model with m states, the transition probability matrix will normally be of dimensions $m \times m$, with entry (i, j) indicating the probability that the process will switch into state j , given that it is currently in state i . Below is a table to help illustrate how transition probability matrices work.

	Future State (time t+1)	
Current State(time t)	1	2
1	0.8	0.2
2	0.3	0.7

The stationary probabilities for a hidden Markov model, at least for our purposes, can be thought of as the proportion of time that the hidden process spends in each state over the long run. The formula for the stationary probabilities of a m -state hidden Markov model is derived from the transition probability matrix [6]:

$$\delta = \mathbf{1}(\mathbf{I} - \Gamma + \mathbf{U})^{-1}$$

where δ is the vector of stationary probabilities, “ $\mathbf{1}$ is a row vector of ones, \mathbf{I} is an $m \times m$ identity matrix, and \mathbf{U} is the $m \times m$ vector of ones” [9].

1.2 Animal movement ecology

Humans have been interested in understanding the movement patterns of animals for thousands of years, dating back at least to Aristotle [7]. Technological advances, both in personal computing and GPS sensors have enabled the collection and analysis of animal location data in ways that were not possible until fairly recently. These developments allow researchers to gain new insight into the when, where, and why of animal movement [7].

Nathan et al. posit four components for animal movement: “the internal state (why move?), motion (how to move?), and navigation (when and where to move?)” [7]. The internal state of an animal will correspond to the goal (or goals) the animal is pursuing at a given time. Sometimes, an animal’s movement behaviour will accomplish multiple goals at once. Other times, the motivation behind an animal’s movement behaviour will change very rapidly, like for example, when an animal is foraging and then suddenly has to evade a predator [7]. Since we usually only have limited insight into the internal state of an animal whose movement we are tracking, it is natural to think of the internal or behavioural process as corresponding to the “hidden process” in a hidden Markov model.

1.3 Hidden Markov models for animal movement

Much can be learned about the movement behaviour of a species from fitting a hidden Markov model. The most common type of hidden Markov model for animal movement is called a “step-and-turn model”. This type of model features two response variables: step length, which is the distance traveled by the animal in between two successive observations, and turning angle, which is the angle at which the animal turned between two steps [5]. Often, when the animal is in a given state, step length is treated as having a gamma distribution and turning angle is treated as having a von Mises distribution. What changes from state to state are the specific observation parameters for the gamma and von Mises distributions. The gamma distribution has a range from $(0, \infty)$, as step lengths are non-directional and thus cannot be negative. The turning angle is typically measured in radians, thus the von Mises distribution, which is bounded from $(-\pi, \pi]$, is a natural fit. The step distribution and the turning

angle distribution in a given state are considered to be independent of one another [8].

The transition probabilities and the observation parameters for each state are estimated via maximum likelihood. The maximum likelihood for a step-and-turn hidden Markov model based on T observations is:

$$L_T = \delta \mathbf{P}(x_1) \Gamma \mathbf{P}(x_2) \Gamma \mathbf{P}(x_3) \dots \Gamma \mathbf{P}(x_T) \mathbf{1}'$$

Here, δ is the initial distribution of the Markov chain, which is a vector containing the probabilities that the process starts in a given state. $\mathbf{P}(x_t)$ is a diagonal matrix, where the i 'th diagonal is the multivariate density function of the i 'th state at time t [9]. Since the step and turning angle distributions are independent of one another, the multivariate probability density function is simply the product of the density function for the step, $p(X_{1t})$, and the density function for the turning angle, $p(X_{2t})$ [8]. The probability density function for the von Mises distribution is:

$$p(x) = c(\kappa \cos(x - \mu)) \quad , \quad x \in (-\pi, \pi]$$

where c is a constant, μ is the value where the density is highest and κ is a measure of concentration, which indicates the degree to which the density is centred around μ . Note that values close to π and $-\pi$ both essentially correspond to a reverse in direction in between steps.

The probability density function for the gamma distribution is:

$$p(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma \alpha} \quad , \quad x \in (0, \infty)$$

where α is the shape parameter and β is the rate parameter. The mean for the gamma distribution is $\frac{\alpha}{\beta}$ and the standard deviation is $\sqrt{\frac{\alpha}{\beta^2}}$. Usually when we describe the features of a gamma distribution in this paper we will refer to the mean and standard deviation, rather than referring to α and β .

There are two ways to include covariates in a hidden Markov model. The first way is to have the observation parameters be dependent on the covariates, while keeping the transition probabilities fixed. The second way is to keep the observation parameters fixed and have the transition probabilities be dependent on the covariates [9]. For this paper, we will focus on the second method. This method can be challenging

with more than two states, but is relatively straightforward when there are only two states. When there are two states, we can model the off-diagonal entries as following something like a logistic regression model:

$$\text{logit}_t \gamma_{i,3-i} = \beta_i y'_t$$

where logit_t is the logit function, $\gamma_{i,3-i}$ is one of the state-switching probabilities in our transition matrix for current state i , β_i is the set of regression coefficients, and y'_t is a list of covariates stored in a vector [9].

Because animal location data sets of interest often contain many observations (the data set we use in this paper contains 42197 rows), and because each additional state has its own parameters and requires a bigger transition matrix, fitting a hidden Markov model with a high number of states (> 4) can be very computationally expensive and time consuming. Furthermore, a high number of states increases the risk that the software will incorrectly identify the optimal solution. Fortunately, it will not be necessary to fit models with more than four states because when fitting animal movement HMMs, “biologists (...) typically expect 2-4 initial states to be present” [8]. This is intuitive because ideally we would like to think of each state as corresponding to a specific behavioural pattern in the animal [8]. While it is easy to imagine two distinct types of movement behaviour (i.e. active behaviour vs. resting behaviour), it becomes much harder to imagine that we can capture five or more distinct behavioural patterns in a particular animal simply by observing its movement.

When deciding on which statistical model to use, it is always helpful to have criteria that allow you to compare models. Two of the most common criteria for model selection are the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The formula for the Akaike information criterion is:

$$-2\log(L_T) + 2p$$

where $\log(L_T)$ is the natural logarithm of the likelihood, and p is the total number of parameters in the model [9].

The formula for the Bayesian information criterion is:

$$-2\log(L_T) + p\log(T)$$

where T is the number of observations [9]. For both the AIC and BIC, a lower value indicates a better model. For AIC and BIC, a higher likelihood results in a lower, and thus better, value. On the other hand, increasing p , the number of parameters results in a higher, and therefore worse, value. Thus, AIC and BIC are both functions of likelihood which are “penalized” by the number of parameters. The difference between the AIC and BIC lies in exactly how much of a penalty each parameter carries [9]. For the AIC the “penalty term” is always 2, whereas for the BIC the penalty term is the natural logarithm of the number of observations. If you have many observations ($T > 8$), the BIC carries a higher penalty for each added parameter [9]. The general idea is that AIC and BIC will favour a model that is neither underfitted nor overfitted. Ideally, the preferred model will capture the underlying pattern of the data without being unnecessarily complex. However, when choosing the number of states for animal movement hidden Markov models, Pohle et al. suggest that the AIC and BIC tend to favor models with more than four states, which, as stated earlier, is undesirable [8]. Unfortunately, it seems that “no one-size fits all objective and universally applicable criterion can be developed for order selection in HMMs” [8].

In this paper we will be looking at a few different things in addition to AIC and BIC that will help us determine the best number of states for our data. One thing we will be looking at are some functions of the state probabilities for each number of states. The state probabilities at time t are the probabilities that the hidden process is in each state at time t , given the observation process. The formula for the probability of the hidden process, C , being in state i at time t is given by:

$$Pr(C_t = i | \mathbf{X}^{(T)} = \mathbf{x}^{(T)}) = \frac{Pr(C_t = i, \mathbf{X}^{(T)} = \mathbf{x}^{(T)})}{Pr(\mathbf{X}^{(T)} = \mathbf{x}^{(T)})}$$

where $\mathbf{x}^{(T)}$ is a matrix of the observations at all times [9]. These will help indicate which model is best because it is preferable that all states are clearly distinguishable from one another. States that are significantly different from one another enable us to infer something about the animal’s behaviour in each state [8]. When each state has unique features, most of the time we should have a pretty good idea which state the hidden process is in at a given time, so the state probability for one state

will be close to 1 while the other states will have a state probability close to zero. When we looked at our fitted models in chapter 2, one set of metrics we used was the proportion of observations that had a state probability of at least .5, .7, or .95. The other state probability metric we looked at was mean information entropy per observation. The formula for information entropy is $E = -\sum p_i \log(p_i)$, where p_i is the state probability for state i at a given observation [2]. The entropy will equal zero when we are absolutely sure which state the hidden process is in, and the entropy will peak when all the state probabilities are equal [2]. Thus, for our purposes, a lower mean entropy for a given model is desirable.

Another way we can compare models with different numbers of states is through the use of pseudo-residual plots [9]. Technically, hidden Markov models do not have residuals, but we can essentially look at the same sorts of plots for pseudo-residuals as we would for regular residuals for other kinds of statistical models. I will not get into detail on the technical difference between residuals and pseudo-residuals here. If the pseudo-residual plots look good, they suggest that there are enough states to adequately capture the pattern in the data. If the pseudo-residual plots look much better with a higher number of states, that is potentially an indication that we should use the higher number of states [9].

1.4 About the wild boar data set and using moveHMM

The data set we will be using to explore model selection in animal movement hidden Markov models consists of wild boar location data that was collected in the south of France by Gustav Fradin and Simon Chamaillé-Jammes [1]. The subsection of their data that we used consisted of six wild boars, with their location recorded via a GPS collar once every half-hour, using Easting and Northing coordinates [1]. Across the six wild boars we had 42197 location observations. The observations also had four environmental covariates. The first covariate is distance to the nearest forest, the second and third covariates are the distance to the nearest large road and small road, respectively, and the fourth covariate is distance to the nearest village. Distance to forest took negative values when the animal was inside a forest and distance to village took negative values when the animal was inside a village. All of the covariates are in metres. The first piece of data preparation we did was breaking up the animal IDs

	ID	x	y	time	dist_forest	dist_road_large	dist_road_small	dist_village
1	41_1-1	602189.4	4862917	2018-02-03 00:00:13	-229.500427	320.61478	300.34421	534.3182
2	41_1-1	602109.4	4862911	2018-02-03 00:30:17	-166.240067	237.79482	222.43208	481.1548
3	41_1-1	602351.8	4862513	2018-02-03 01:00:38	-108.362190	206.80449	156.82024	952.0332
4	41_1-1	602274.9	4862499	2018-02-03 01:30:21	-77.961128	182.23061	123.75297	928.3124
5	41_1-1	602282.9	4862519	2018-02-03 02:00:17	-77.961128	182.23061	123.75297	928.3124
6	41_1-1	602285.0	4862504	2018-02-03 02:30:14	-77.961128	182.23061	123.75297	928.3124
7	41_1-1	602254.6	4862501	2018-02-03 03:00:13	-37.394051	157.65671	87.87125	906.3539
8	41_1-1	602289.3	4862518	2018-02-03 03:30:13	-77.961128	182.23061	123.75297	928.3124
9	41_1-1	602281.1	4862498	2018-02-03 04:00:12	-77.961128	182.23061	123.75297	928.3124
10	41_1-1	602291.2	4862519	2018-02-03 04:30:12	-77.961128	182.23061	123.75297	928.3124
11	41_1-1	602291.5	4862499	2018-02-03 05:00:13	-77.961128	182.23061	123.75297	928.3124
12	41_1-1	602294.7	4862472	2018-02-03 05:30:12	-38.265087	138.68610	86.43325	972.9650
13	41_1-1	602270.5	4862527	2018-02-03 06:00:14	-77.961128	182.23061	123.75297	928.3124
14	41_1-1	602284.5	4862519	2018-02-03 06:30:42	-77.961128	182.23061	123.75297	928.3124
15	41_1-1	602283.8	4862521	2018-02-03 07:00:17	-77.961128	182.23061	123.75297	928.3124

Figure 1.1: The first 15 observations of the data set, before additional data prep

into subsections when there was large time gaps using Michelot’s function “split at gaps” [3]. HMMs treat time as discrete, which means that inconsistent time intervals can be problematic. Breaking up the data into subsections when there are time gaps ensures that time can in fact be treated as discrete.

The tool we used to fit our HMMs was the R package “moveHMM”, which was developed by Théo Michelot, Roland Langrock, and Toby A. Patterson in 2016 specifically for the purpose of fitting animal movement hidden Markov models [5]. The first thing we did using moveHMM was derive the step lengths (in metres) and turning angles (in radians) from the Easting and Northing coordinates (listed as x and y in the data set) using the “prepData” function. Next, we identified the step lengths that were exactly zero, and replaced them with small non-zero values. Only 7 of the 42197 observations had a step length of exactly zero. We replace these values because the gamma distribution is undefined at zero, so fitting a gamma distribution would require an additional “zero-inflation” parameter for each state, which would make the models unnecessarily complex and less numerically stable. We then fitted our two-state, three-state and four-state models using the “fitHMM” function. To use fitHMM, you need to select initial values for each of the observation parameters in order for the function to work. Sometimes, the model can converge to a non-optimal solution which does not maximize the likelihood as intended [4]. Because of this, it is

good to fit the same model with many initial values and check if the likelihoods are the same for each one [4]. Fortunately, the likelihoods for our two-state, three-state, and four-state models seemed to be consistent regardless of the initial parameter values.

Once we fitted our models, we used the “plot” function to generate histograms of the step length and turning angles for each model, with coloured lines indicating the density functions for each state and a dotted line indicating the overall density function for the overall fitted model [6].

We used the “plotPR” function to print off plots of the pseudo-residuals. These graphs include a time-series plot of the pseudo-residuals, a Q-Q plot of the pseudo-residuals, and an autocorrelation plot of the pseudo-residuals [6].

We used the “stateProbs” function to obtain the state probabilities for each number of states, stored in a matrix [6]. This matrix allowed us to derive our state probability metrics.

Once we decided that two states are sufficient, we fitted some two-state models with covariates. For computational reasons, we converted the environmental covariates from metres to kilometres. We also derived a covariate for time of day as a decimal number between 0 and 24 using the “clock” package so we could use temporal covariates. Once we fitted our covariate models, we used the “plotStationary” function to generate graphs of how covariates affect the stationary probabilities.

Figure 1.1 shows the first 15 observations of the wild boar data set, after the data was split into tracks but before we performed additional preparations to make it suitable to fit our step-and-turn hidden Markov model. Figure 1.2 is the first 15 observations of the wild boar data set, after the data was fully prepared for us to fit our models with.

Figures 1.3 and 1.4 are histograms of the step length and turning angle values of the data set, respectively. The step lengths have a high frequency of observations near zero, with a long tail that is visible up to about 700 metres. The frequency of the turning angles has two peaks, one around 0, and one around $+\pi$.

ID	step	angle	x	y	time	dist_forest	dist_road_large	dist_road_small	dist_village	time_of_day
41_1-1	80.1645081	NA	602189.4	4862917	2018-02-03 00:00:13	-0.229500427	0.32061478	0.30034421	0.5343182	0.0000
41_1-1	466.4667325	2.04201086	602109.4	4862911	2018-02-03 00:30:17	-0.166240067	0.23779482	0.22243208	0.4811548	0.5000
41_1-1	78.0549890	-1.94542200	602351.8	4862513	2018-02-03 01:00:38	-0.108362190	0.20680449	0.15682024	0.9520332	1.0000
41_1-1	21.2626911	-2.12587536	602274.9	4862499	2018-02-03 01:30:21	-0.077961128	0.18223061	0.12375297	0.9283124	1.5000
41_1-1	14.6534361	-2.61258680	602282.9	4862519	2018-02-03 02:00:17	-0.077961128	0.18223061	0.12375297	0.9283124	2.0000
41_1-1	30.5690617	-1.60241868	602285.0	4862504	2018-02-03 02:30:14	-0.077961128	0.18223061	0.12375297	0.9283124	2.5000
41_1-1	38.5402018	-2.80367433	602254.6	4862501	2018-02-03 03:00:13	-0.037394051	0.15765671	0.08787125	0.9063539	3.0000
41_1-1	21.8907947	-2.40754389	602289.3	4862518	2018-02-03 03:30:13	-0.077961128	0.18223061	0.12375297	0.9283124	3.5000
41_1-1	23.7127415	3.08513795	602281.1	4862498	2018-02-03 04:00:12	-0.077961128	0.18223061	0.12375297	0.9283124	4.0000
41_1-1	20.2055152	-2.68573481	602291.2	4862519	2018-02-03 04:30:12	-0.077961128	0.18223061	0.12375297	0.9283124	4.5000
41_1-1	27.2583480	0.10350225	602291.5	4862499	2018-02-03 05:00:13	-0.077961128	0.18223061	0.12375297	0.9283124	5.0000
41_1-1	60.7205062	-2.85052378	602294.7	4862472	2018-02-03 05:30:12	-0.038265087	0.13868610	0.08643325	0.9729650	5.5000
41_1-1	16.4095184	-2.53845475	602270.5	4862527	2018-02-03 06:00:14	-0.077961128	0.18223061	0.12375297	0.9283124	6.0000
41_1-1	2.8576149	2.36567474	602284.5	4862519	2018-02-03 06:30:42	-0.077961128	0.18223061	0.12375297	0.9283124	6.5000
41_1-1	2.5030569	0.05187945	602283.8	4862521	2018-02-03 07:00:17	-0.077961128	0.18223061	0.12375297	0.9283124	7.0000

Figure 1.2: The first 15 observations of the data set, after additional data prep

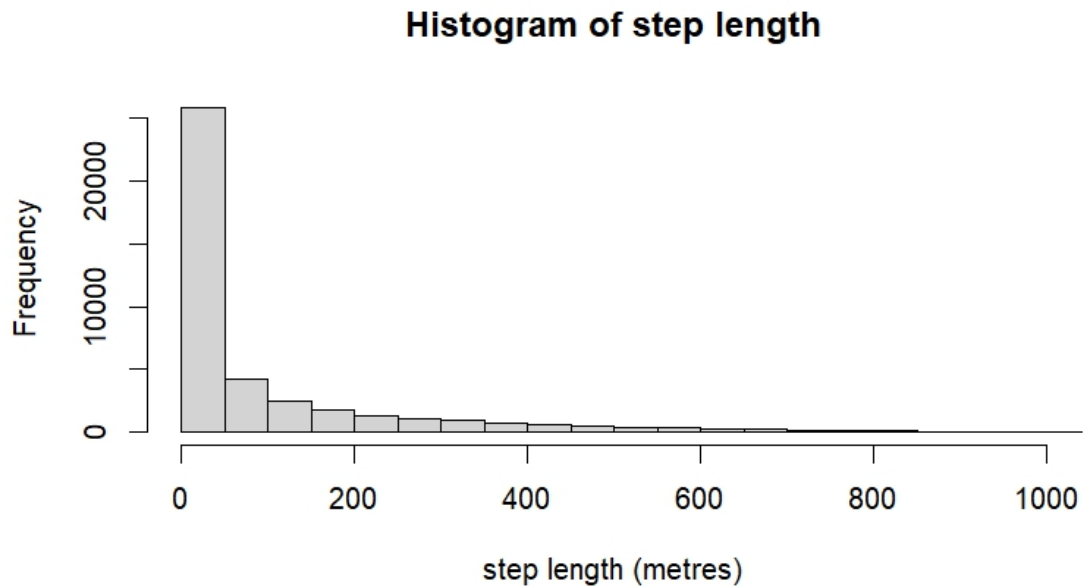


Figure 1.3: Histogram of the step lengths for our data set, derived from Easting and Northing location coordinates.

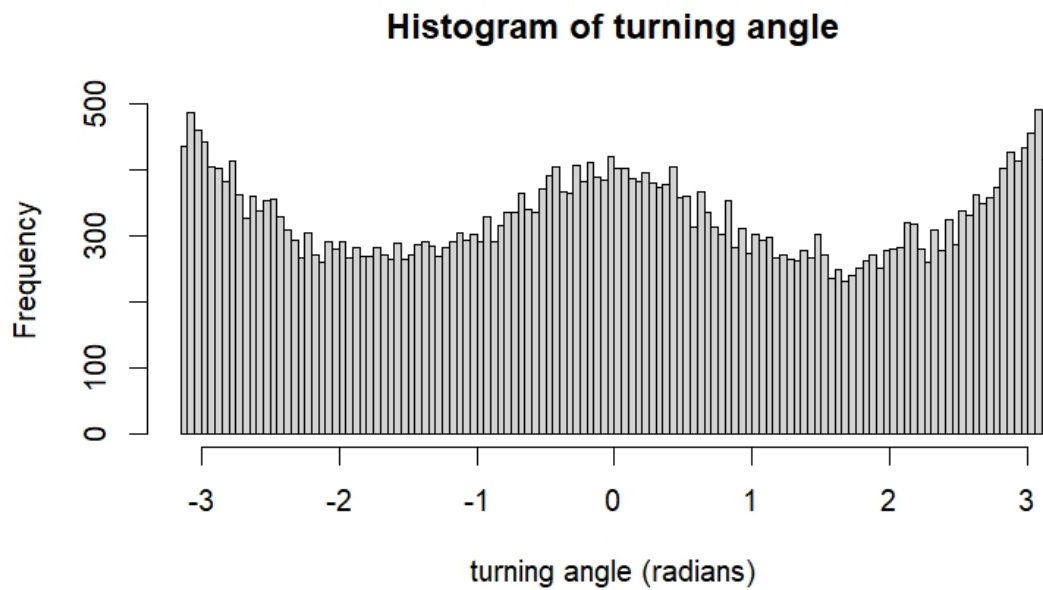


Figure 1.4: Histogram of the turning angles for our data set, derived from Easting and Northing location coordinates

Chapter 2

Comparing the two-state, three-state, and four-state models

2.1 Features of the two-state, three-state, and four-state models

Figures 2.1-2.3 are the R outputs showing the maximum likelihood estimates of the observation parameters and transition probabilities for the two-state, three-state, and four-state models, fitted using the “fitHMM” function in moveHMM.

Figure 2.1 shows the R output for the fitted two-state model. State 1 has a turning angle centered near $-\pi$, whereas state 2 has a turning angle distribution centered near zero. State 1 has a step length mean of around 18.5 metres and state 2 has a step length of around 242 metres. This indicates that state 1 clearly the less active state, and state 2 is clearly the more active state. A boar in state two has a higher chance of sustained movement in the same direction over multiple half-hour intervals. The diagonal entries of the transition matrix are .93 for state 1 and .90 for state 2. These values are quite high, and indicate that when an animal is in one state, it is very likely it will remain in that state at the next observation.

Figure 2.2 shows the R output for the fitted three-state model. State 1 and 2 have turning angles centered near $-\pi$, and state 3 has a turning angle distribution centered near zero. The step length means and standard deviations increase from state 1 to state 2 to state 3. There is a particularly large increase in mean step length from state 2 to state 3, going from roughly 83 metres to 369 metres. Interestingly, turning angle concentration is near 0 for state 2, meaning that the turning angle distribution is basically uniform in this state. The turning angle concentration for state 3 is by far the highest, indicating that in this state there is a lot of directional movement sustained over multiple half-hour periods in this state. The diagonal entries of the transition matrix show that if the animal is in state 1, it has a probability of .91 of being in state 1 at the next observation, whereas if the animal is in state 2, it only has a probability of .69 of being in state 2 at the next observation.

Figure 2.3 shows the R output for the fitted four-state model. States 1 and 2 have

turning angle distributions centered near $+\pi$, and states 3 and 4 have turning angle distributions centered near 0. Step length mean and standard deviation increases from state 1 to state 2 to state 3 to state 4. Interestingly, turning angle concentrations follow a different pattern, being lower for states 1 and 3 and higher for states 2 and 4. For the four-state model, none of the diagonal entries for the transition probability matrix are very high, ranging from about .68 to .77, indicating that state-switching occurs quite frequently.

Generally, states where the average distance travelled is higher have turning angle distributions centred at 0 and higher turning angle concentrations, although the four-state model subverted this pattern somewhat. Generally, as the number of states increases, the frequency of state-switching increases as well.

Figures 2.4-2.6 provide visualizations of the patterns discussed above, colour-coded based on state, with the black dotted line representing the overall fitted distribution. These visualizations were generated using the “plot” function in moveHMM.

2.2 Model checking

Table 2.1 shows the number of parameters, log-likelihood, AIC, and BIC for the two-state, three-state, and four-state models. As expected, AIC and BIC both favour the four-state model. Notably, the difference in AIC and BIC between three and four state models is not as pronounced as the difference between the two and three state models.

Tables 2.2 and 2.3 contain some information about the state probabilities. Table 2.2 shows the probability that a randomly chosen observation will have a state probability above a certain threshold. The thresholds we chose were .5, .7, and .95. Higher values for this probability indicate that we are more certain overall about what state the hidden process is in at a given time. Table 2.3 shows the mean information entropy for a given observation increase as the number of states increase. Mean entropy equals zero when we are certain about which state the process is in. The maximum entropy is reached when all the state probabilities are equal to each other. The maximum possible entropy is the natural logarithm of the number of states, so it increases as the number of states increases. Both tables indicate that as the number of states increase, we are less and less certain about what state the hidden

```

> hmm_2SNOU
Value of the maximum log-likelihood: -287767.9

Step length parameters:
-----
      state 1  state 2
mean 18.50750 242.1183
sd   15.49025 219.3747

Turning angle parameters:
-----
              state 1      state 2
mean          -3.0887166 -0.02421734
concentration  0.3925371  0.56200394

Regression coeffs for the transition probabilities:
-----
              1 -> 2    2 -> 1
intercept -2.615607 -2.168584

Transition probability matrix:
-----
              [,1]      [,2]
[1,] 0.9318593 0.0681407
[2,] 0.1026073 0.8973927

Initial distribution:
-----
[1] 0.5827695 0.4172305

```

Figure 2.1: R output for the fitted two-state model

process is in at a given time. As the number of states increases, the states become less and less distinguishable from one another.

Figures 2.7-2.9 show the pseudo-residual plots for the two-state, three-state, and four-state models, respectively. The time-series and Q-Q pseudo-residual plots look remarkably similar for all three models, and suggest that all three models capture the data quite well. The autocorrelation plots of the pseudo-residuals for every model indicate that there is some positive correlation between successive observations. The x axis is the sequential gap between observations. Since observations are every half hour, a gap of 24 corresponds to 12 hours and a gap of 48 corresponds to 24 hours. Interestingly, there is some negative correlation around the 12-hour mark, and some positive correlation around the 24-hour mark. This effect is most pronounced in the two-state model, although it is not pronounced enough to warrant serious concern.

```

> hmm_3sNo0 #4:00 mins mar 5
Value of the maximum log-likelihood: -285647.6

Step length parameters:
-----
      state 1  state 2  state 3
mean 15.17938 83.26936 368.691
sd   11.81658 68.08100 237.228

Turning angle parameters:
-----
              state 1      state 2      state 3
mean          -3.0728316 -3.11026418 -0.02073798
concentration  0.3758988  0.03500044  0.99234921

Regression coeffs for the transition probabilities:
-----
              1 -> 2    1 -> 3    2 -> 1    2 -> 3    3 -> 1    3 -> 2
intercept -2.402823 -4.722483 -1.574023 -1.433408 -2.87871 -1.426853

Transition probability matrix:
-----
              [,1]      [,2]      [,3]
[1,] 0.90962402 0.08228663 0.008089354
[2,] 0.14332794 0.69170417 0.164967886
[3,] 0.04336073 0.18519531 0.771443961

Initial distribution:
-----
[1] 4.873604e-01 5.125695e-01 7.014754e-05

```

Figure 2.2: R output for the fitted three-state model

Table 2.1: Log-likelihoods and information criteria for the two-state, three-state, and four-state models

Number of states	p	log-likelihood	AIC	Δ AIC	BIC	Δ BIC
2	11	-287770.5	575563	+6624	575658	+6451
3	20	-285651.7	571343	+2404	571516	+2309
4	31	-284438.3	568939	-	569207	-

2.3 Choosing the number of states

After careful considering, we have decided that the two-state model is best, despite the fact that the two-state model has the worst AIC and BIC. This is because the pseudo-residual plots in figure 2.7 tell us that the two-state model does a good job

```

> hmm_4sNo0
Value of the maximum log-likelihood: -284433

Step length parameters:
-----
      state 1 state 2 state 3 state 4
mean 9.652127 29.12655 121.3835 430.8554
sd   6.712641 17.12756 100.3735 245.8192

Turning angle parameters:
-----
      state 1 state 2 state 3 state 4
mean   -2.9399036 3.1262129 -0.05970967 -0.01319337
concentration 0.2091078 0.6972649 0.23609203 1.14737591

Regression coeffs for the transition probabilities:
-----
      1 -> 2  1 -> 3  1 -> 4  2 -> 1  2 -> 3  2 -> 4  3 -> 1  3 -> 2  3 -> 4
intercept -1.527656 -2.565437 -5.375532 -1.054272 -2.086005 -5.035851 -2.468823 -2.371201 -1.62158
      4 -> 1  4 -> 2  4 -> 3
intercept -3.186265 -3.522185 -1.191858

Transition probability matrix:
-----
      [,1] [,2] [,3] [,4]
[1,] 0.77008514 0.16714214 0.05920841 0.003564304
[2,] 0.23557513 0.67607362 0.08395633 0.004394928
[3,] 0.06156007 0.06787281 0.72693474 0.143632376
[4,] 0.03006580 0.02148743 0.22091859 0.727528178

Initial distribution:
-----
[1] 0.2019260 0.3386432 0.4487943 0.0106365

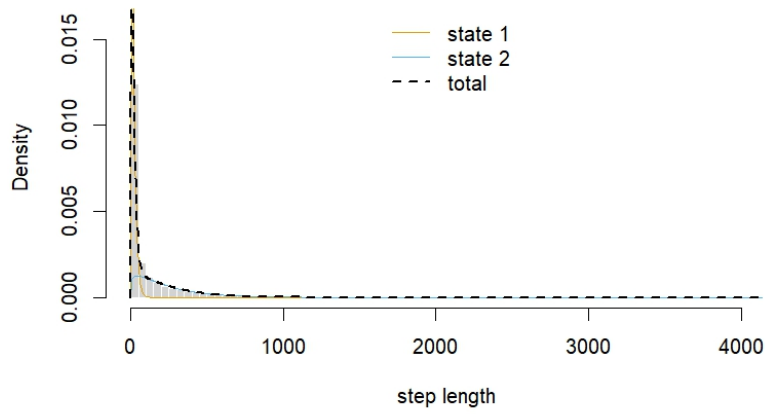
```

Figure 2.3: R output for the fitted four-state model

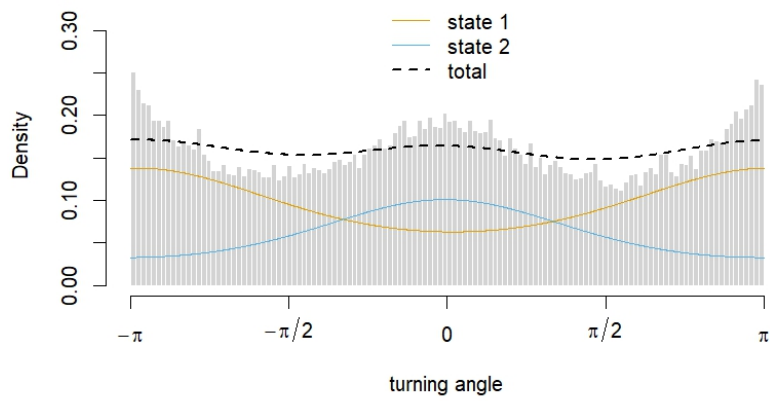
Table 2.2: State probability metrics

Number of states	Pr(max(stateProbs) > .5)	> .7	> .95
2	1	.9597	.8394
3	.9980	.8777	.5570
4	.9773	.7672	.2996

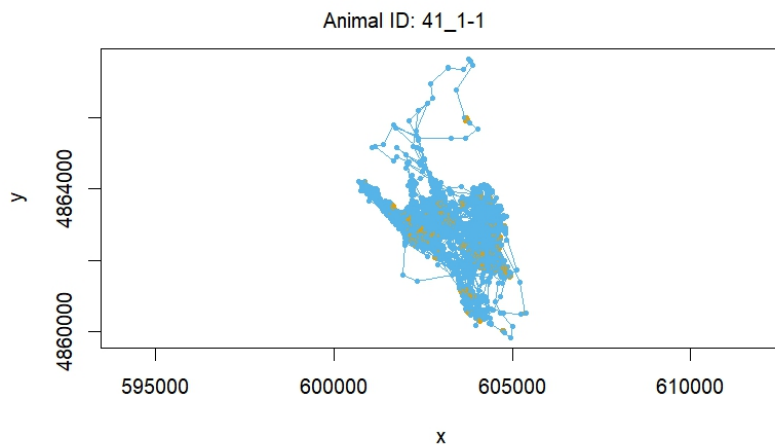
overall of fitting the data. Additionally, only having two states provides us with states that are the most clearly differentiable from one another, as evidenced by our tables of state probability metrics. Particularly striking was the fact that almost 84% of our observations had a state probability of at least .95. This is in sharp contrast to the three-state model, where only 55.7% of our observations fit this criteria, or the four-state model, where only 30% of our observations fit this criteria. While there was slightly more autocorrelation in the residuals for the two-state model, the differences between models were not drastic. Ultimately, since each model had their advantages and disadvantages, we opted to go with the simplest model.



(a) Step length probability density functions

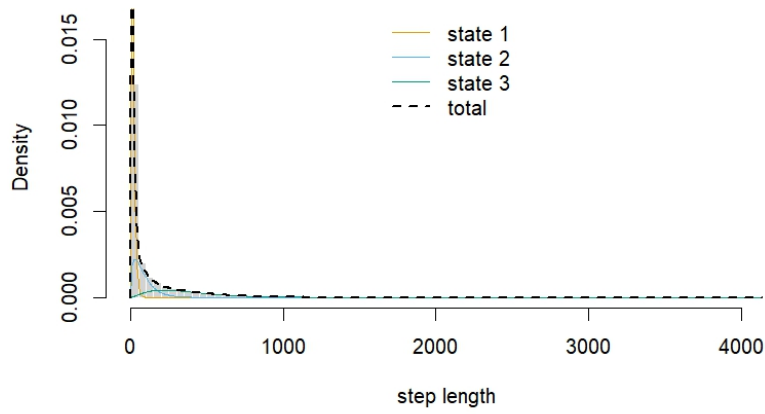


(b) Turning angle probability density functions

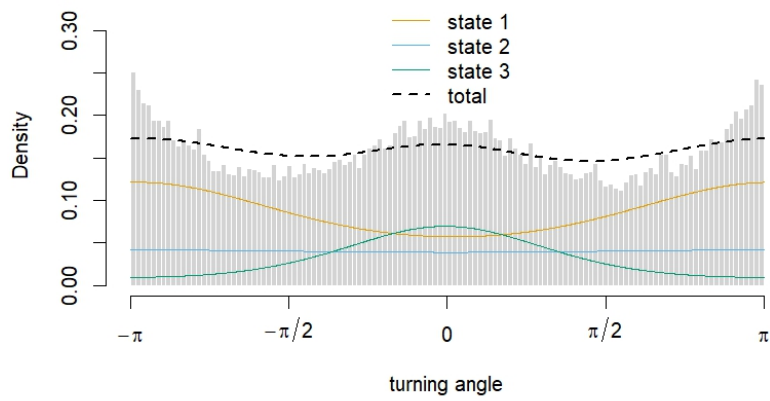


(c) Plotting animal 41's most likely state sequence

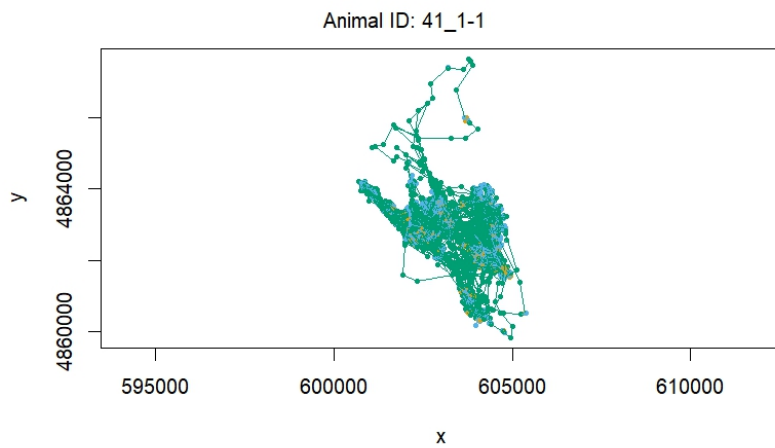
Figure 2.4: Two-state model results



(a) Step length probability density functions

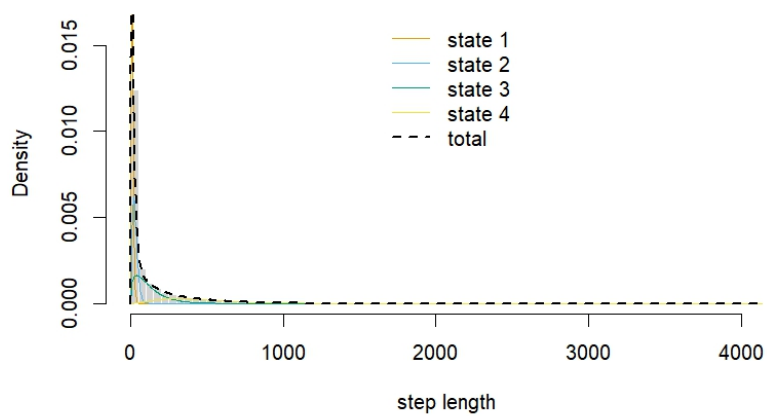


(b) Turning angle probability density functions

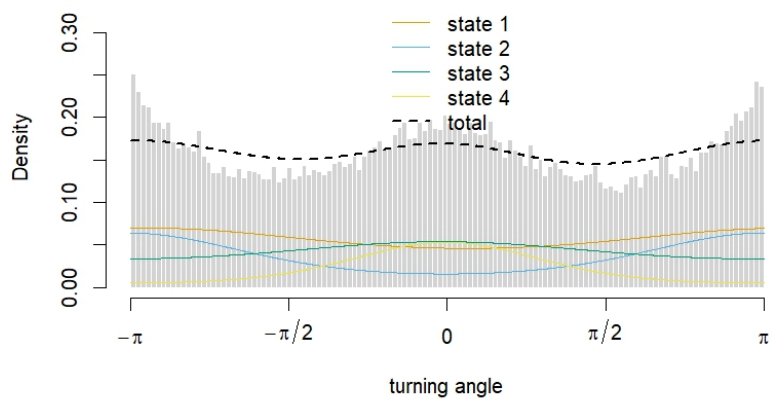


(c) Plotting animal 41's most likely state sequence

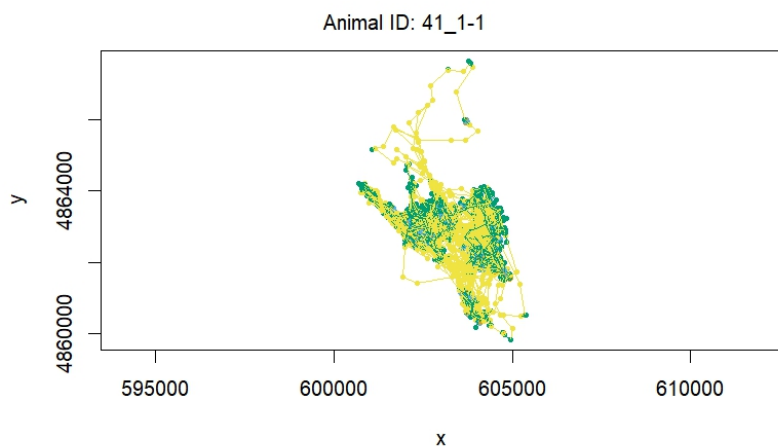
Figure 2.5: Three-state model results



(a) Step length probability density functions



(b) Turning angle probability density functions



(c) Plotting animal 41's most likely state sequence

Figure 2.6: Four-state model results

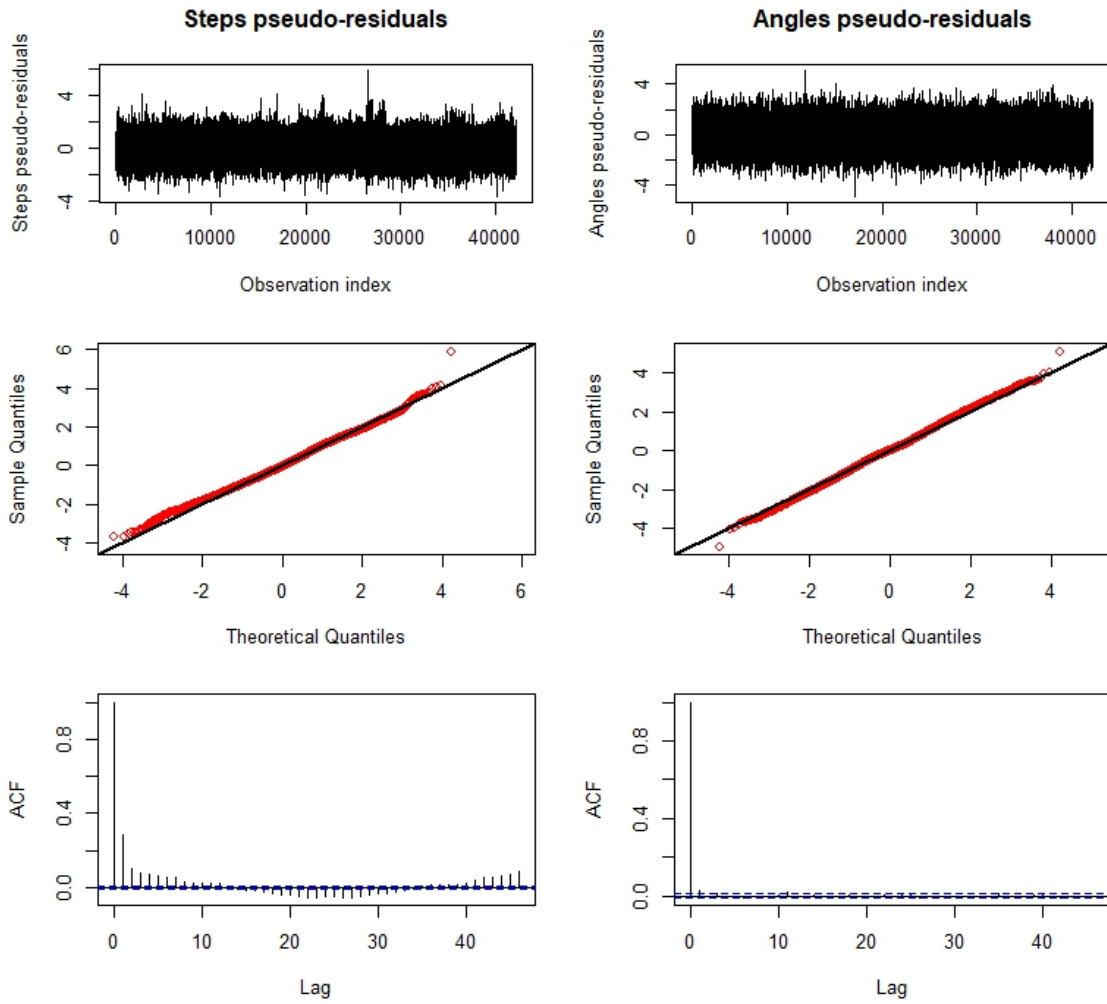


Figure 2.7: Pseudo-residuals for the two-state model

Table 2.3: Mean information entropy for the state probabilities

Number of states	mean entropy	maximum possible entropy
2	.0919	$\log(2) = .6931$
3	.2476	$\log(3) = 1.0986$
4	.4099	$\log(4) = 1.3863$

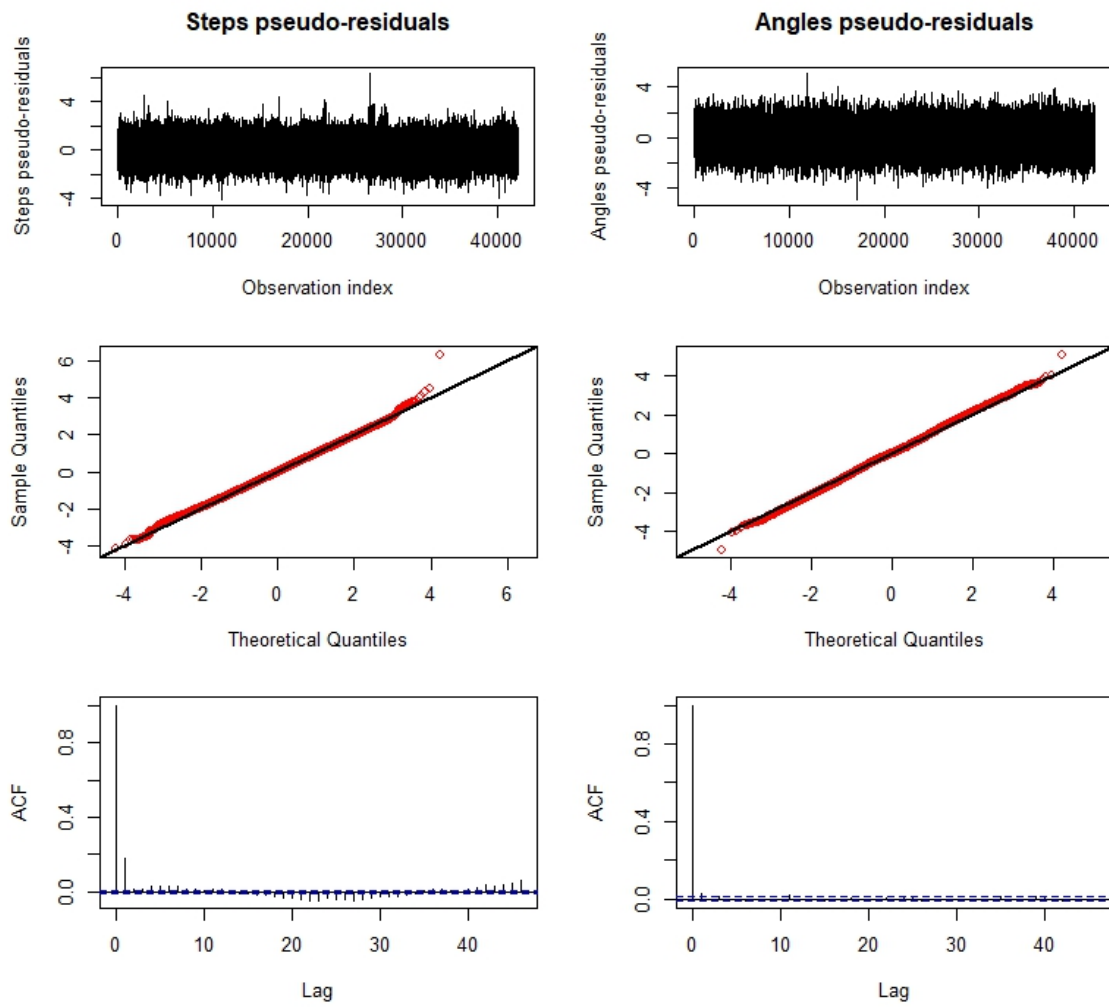


Figure 2.8: Pseudo-residuals for the three-state model

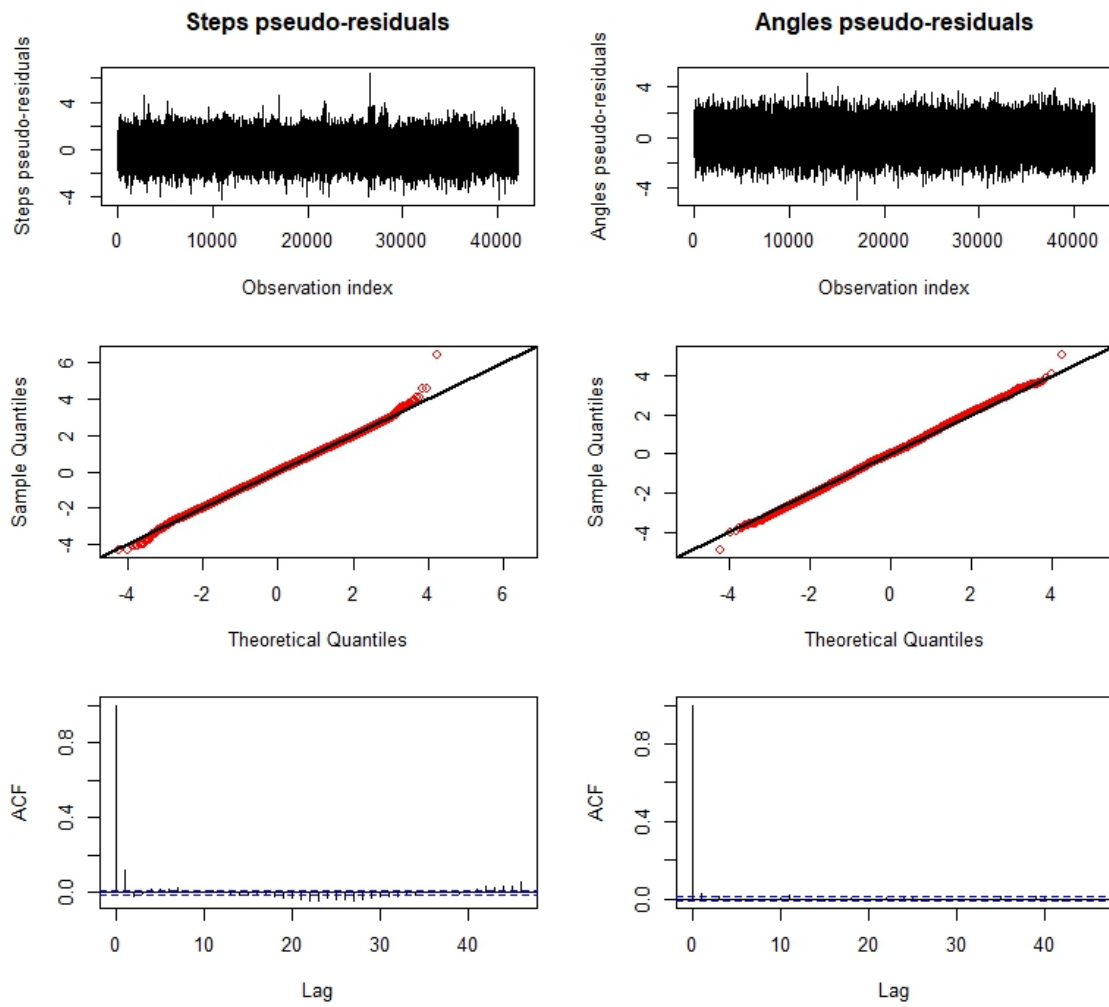


Figure 2.9: Pseudo-residuals for the four-state model

Chapter 3

Examining environmental and temporal factors

3.1 Choosing which covariates to include

Table 3.1: Covariate models, ranked based on AIC

Model	covariates	AIC	Δ AIC
1	sine time, cosine time, forest, large road, village	571778.2	-
2	forest, large road, village	575412.2	+3634.0
3	forest, large road, small road, village	575413.9	+3635.7
4	forest, small road, village	575466.9	+3688.7
5	forest, village	575468.5	+3690.3
6	forest	575470.7	+3692.5
7	forest, small road	575472.2	+3694.0
8	village	575522.7	+3744.5
9	small road	575561.3	+3783.1
10	no covariates	575562.9	+3784.7
11	large road	575565.9	+3787.7

While AIC and BIC are inadequate in determining the optimal number of states in a hidden Markov model, we believe that either criterion can be a decent indicator to look at when selecting covariates, after a number of states has been decided. As the table above indicates, we have fitted 11 two-state models with different combinations of 6 covariates, and then ranked them based on AIC. First, we identified the optimal combination of environmental covariates and then added time covariates to that model. The covariate model with the highest AIC is the model that includes the time covariates, and the forest, large road, and village environmental covariates, but does not include the small road covariate. The time covariates we included were sine and cosine functions of the time of day: $\sin(\text{time} * \frac{\pi}{12})$ and $\cos(\text{time} * \frac{\pi}{12})$. We chose the sine function because it equals 1 at 6am and -1 at 6pm. We chose the cosine function because it equals 1 at 12am and -1 at 12pm. These time covariates had a much more pronounced effect on the AIC and BIC than any of the environmental

```

Regression coeffs for the transition probabilities:
-----
                1 -> 2      2 -> 1
intercept      -2.419265096 -1.45786896
sin(time_of_day * pi/12) -0.467627619  0.68390525
cos(time_of_day * pi/12)  2.112145502 -1.23811667
dist_village    -0.007901360 -0.04198711
dist_road_large -0.003715833 -0.05592564
dist_forest     -0.044545613 -0.57661929

```

Figure 3.1: Covariate regression coefficients for the final model

covariates, which suggests that the movement behaviour of the wild boars we looked at follows strong circadian patterns. Thus, our final model is a two-state model with two time of day covariates, the forest covariate, the large road covariate, and the village covariate.

Figure 3.1 is the R output of the regression coefficients for the final model. The further the coefficient is from zero, the more pronounced the effect the corresponding covariate has. It is worth noting that since the regression contains a logit transformation, an increase of one covariate value by a certain amount will not always increase the transition probability by a directly proportional amount, although the general trend still holds. The R output confirms that the time covariates, particularly the cosine function, have the largest effect on the transition probabilities, although the forest covariate also has quite a considerable effect on the probability of switching from state 2 to state 1.

3.2 Examining the effects of the covariates on the stationary probabilities

Figures 3.2-3.5 are the plots of the effects of the covariates on the stationary probabilities, using the “plotStationary” function in moveHMM. Using the argument “plotCI”, we have included vertical 95% confidence intervals in the graphs to indicate the level of certainty of the estimates [6]. The time of day covariate has the most pronounced effect on the stationary distributions, with state 1 (the less active, or “resting” state) being much more prevalent during the day, and state 2 being much more prevalent at night. The stationary probabilities for a covariate are calculated and plotted assuming that all other covariates are at their mean value. Because of this, for the final

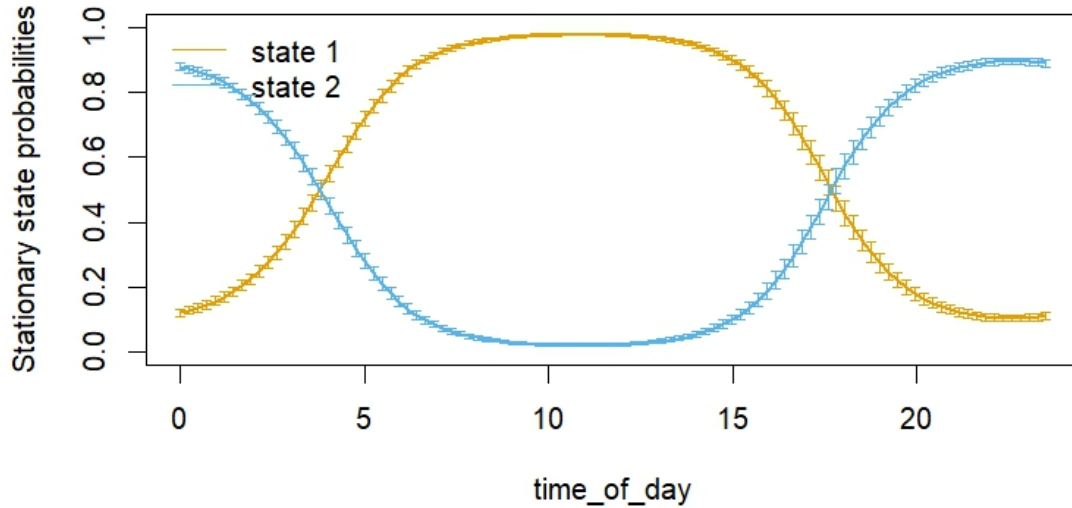


Figure 3.2: Plot of the stationary probabilities based on the hour of the day (in military time), from the final model.

model the plots for the environmental covariates were produced assuming that the time was sometime around noon, meaning that the prevalence of state 1 was very high regardless of what the value of the covariate of interest happened to be. It is for this reason that we present the stationary distribution plots of the environmental covariates from the model without time covariates to better illustrate their overall effects. Of the environmental covariates, the forest distance had the largest effect, with the more active state being much more prevalent outside the forest and the less active state being much more prevalent deep inside the forest.

Our findings concerning the effect of time on the stationary probabilities are a pretty good indication that these boars are nocturnal, which aligns nicely with the findings of Fradin and Chamaille-Jammes, who provided us with our data set [1].

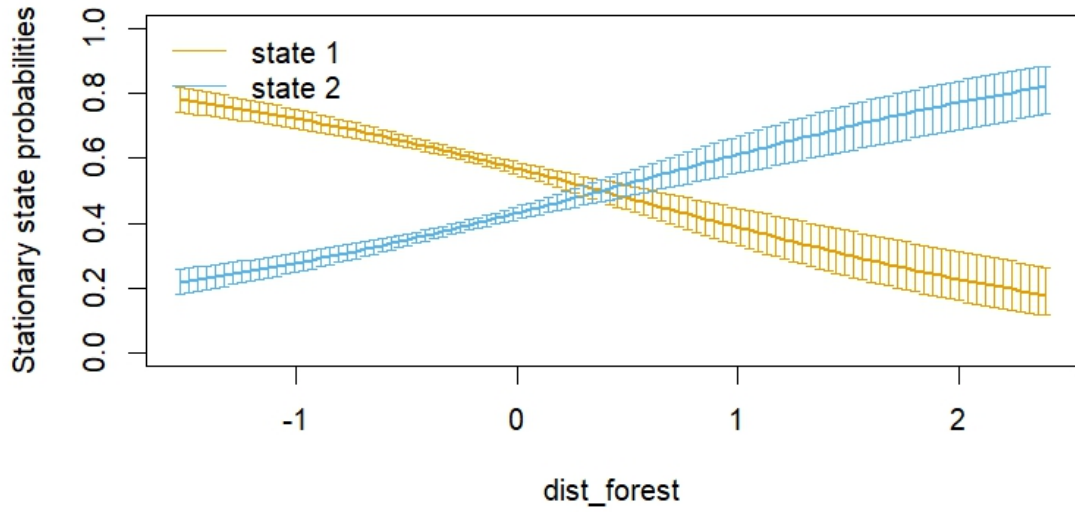


Figure 3.3: Plot of the stationary probabilities based on the distance to forest in kilometres, based on the model with no time covariates.

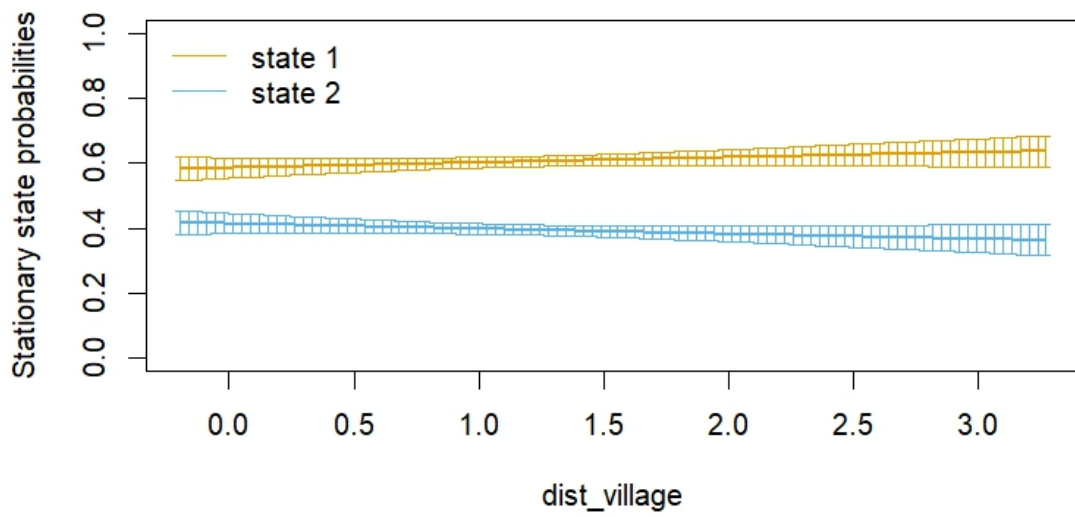


Figure 3.4: Plot of the stationary probabilities based on the distance to village in kilometres, based on the model with no time covariates.

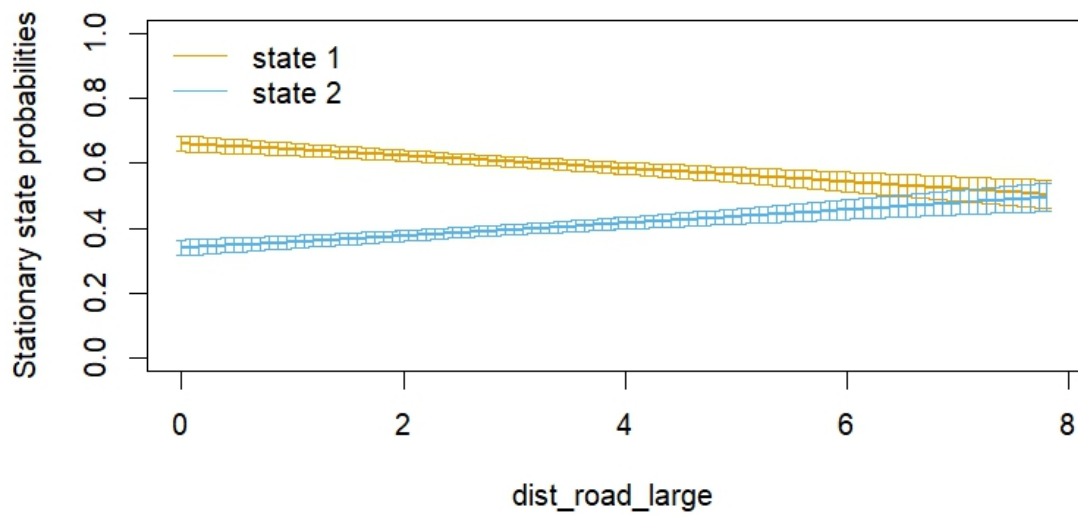


Figure 3.5: Plot of the stationary probabilities based on the distance to large road in kilometres, based on the model with no time covariates.

Chapter 4

Conclusion

In this paper, we have explored how to select the best hidden Markov model when investigating animal movement. As detailed in chapter two, we recommend considering a number of factors when selecting the number of states, including information criteria, pseudo-residual plots, and the state probabilities. This is by no means a comprehensive list of what factors to look at. As Pohle et al. mention, the right number of states may depend on the specific topic that is being investigated. In practice, it is best to consult with an ecologist who is knowledgeable about the species in question before making a decision [8].

In our analysis of the wild boar data set, we decided that the best model was a two-state model with both temporal and environmental covariates. The two temporal covariates and the distance to forest covariate had the largest impact on whether the boars were in the more active state or the less active state. Our findings indicate that these boars in the south of France tend to be more active at night, and tend to be more active when they are outside the forest rather than within it.

If the reader is interested in fitting their own animal movement hidden Markov models, we highly recommend looking at Théo Michelot’s website “<https://CRAN.R-project.org/package=moveHMM>”, which includes helpful vignettes on how to fit these kinds of hidden Markov models using moveHMM. We also recommend reading *Hidden Markov Models for Time Series* by Walter Zucchini, Iain L. MacDonald, and Roland Langrock [9] to develop a better understanding of the theory behind hidden Markov models.

Bibliography

- [1] Gustav Fradin and Simon Chammaille-Jammes. Hogs sleep like logs: Wild boars reduce the risk of anthropic disturbance by adjusting where they rest. *Ecology and Evolution*, 2023.
- [2] Robert M. Gray. *Entropy and Information Theory*. Springer, 2021.
- [3] Theo Michelot. Analyzing animal movement in r. <https://github.com/eco4cast/Statistical-Methods-Seminar-Series>, 2022.
- [4] Theo Michelot and Roland Langrock. A short guide to choosing initial parameter values for the estimation in movehmm. *cran.r-project.org*, 2023.
- [5] Theo Michelot, Roland Langrock, and Toby A. Patterson. movehmm: an r package for the statistical modelling of animal movement data using hidden markov models. *Methods in Ecology and Evolution*, 2016.
- [6] Theo Michelot, Roland Langrock, and Toby A. Patterson. movehmm: an r package for the analysis of animal movement data. *cran.r-project.org*, 2023.
- [7] Ran Nathan, Wayne M. Getz, Eloy Revilla, Marcel Holyoak, Ronen Kadmon, David Saltz, and Peter E. Smouse. A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences of the United States of America*, 2008.
- [8] Jennifer Pohle, Roland Langrock, Floris M. van Beest, and Niels Martin Schmidt. Selecting the number of states in hidden markov models: Pragmatic solutions illustrated using animal movement. *Journal of Agricultural, Biological and Environmental Statistics*, 2017.
- [9] Walter Zucchini, Iain L. MacDonald, and Roland Langrock. *Hidden Markov Models for Time Series*. CRC Press, 2016.