

Efficient Statistical Inference for High Frequency Movement Data

Abdulmajeed Alharbi

School of Mathematical and Physical Sciences

The University of Sheffield

A thesis submitted in partial fulfillment for the degree of
Doctor of Philosophy

December 2025

Abstract

Recent years have seen a proliferation of high-frequency animal movement data, often exceeding 1 Hz, enabling much greater insight into behaviour than was possible with lower-resolution data. In particular, it is now possible to detect the precise moments at which animals make navigational decisions, placing the concept of movement as a sequence of "steps and turns" on a more rigorous footing.

Summarising high-frequency trajectories in terms of the timing and angles of discrete turning points is crucial for many downstream analyses. This is especially important given that widely used techniques in movement ecology, such as hidden Markov models (HMMs) and step selection analysis (SSA), were originally developed with coarser data in mind.

In this thesis, I propose a statistically principled method for identifying turning points in high-frequency movement data, with minimal reliance on arbitrary parameter choices. The method is formulated as a statistical changepoint detection problem, in which the animal's path is represented as a stream of heading data (typically derived from magnetometry) and denoised as a stepwise function over time.

Building on this segmentation, I develop a hidden Markov model framework for behavioural state inference based on steps defined between successive turning points. The model accommodates variable-duration steps and incorporates emission distributions for step angles, durations, and speeds. I further extend this framework by allowing transition probabilities to depend on environmental covariates, enabling state-switching

dynamics to be driven by external factors.

Together, these contributions provide a coherent and computationally efficient framework for analysing high-frequency animal movement data, with strong potential to improve behavioural inference in ecological studies.

Acknowledgements

First and foremost, all praise and gratitude are due to God, who has guided me, blessed me with the strength, patience and supportive surroundings to complete this journey, and opened the doors of opportunity that led me to pursue this PhD. I am deeply thankful for the unwavering love and support of my family, whose presence has been my greatest source of comfort and motivation throughout this endeavour.

My sincere thanks go to my supervisors, Professor Jonathan R. Potts and Professor Paul G. Blackwell, for their invaluable guidance, patience, and encouragement. Whenever I needed direction, they were always there generous with their time, wisdom, and insight. This work would not have been possible without their continued support and belief in me.

They say, "Behind every successful man, there is a great woman." In my case, I am blessed with two: my mother and my wife, who have always been by my side. First, to my mother, whose endless prayers and unconditional love have always been my shield and source of strength. And to my wife, Ghada, whose patience, understanding, and constant encouragement carried me through the toughest days of this journey. You both are truly the backbone of this achievement.

To my little son, Faisal, who joined us in the final months of this journey your presence brought me immeasurable joy and emotional strength when I needed it most. You reminded me of what truly matters and gave this accomplishment deeper meaning.

Finally, I extend my heartfelt gratitude to my beloved country, the Kingdom of Saudi

Arabia, for its continuous investment in education and for granting me the opportunity to pursue my studies abroad. I am proud to be part of a nation that believes in empowering its people to build a brighter and more prosperous future.

Contents

1 Introduction **1**

1.1 High-Frequency Movement Data 2

1.2 Ecological Implications 5

1.3 Statistical Change Point Analysis 7

1.4 Turning Points Analysis 11

1.5 Contributions 14

1.6 Thesis Structure 15

2 Turning Point Detection **17**

2.1 Method 18

2.1.1 Model Formulation 18

2.1.2 Maximising the Likelihood Function 20

2.1.3 Solving the Recursion 23

2.1.4 Estimating β, c, and κ 28

2.1.5 Simulation study 32

2.1.6 Case Study on Oryx Movement 36

2.2 Results 38

2.2.1 Simulation results 38

2.2.2 Oryx data 41

2.2.3 Run Time 44

2.3 Summary 44

3 Modelling Temporal Dependence in Turn Detection 48

3.1 Method 49

3.1.1 Model Formulation 49

3.1.2 Estimating β, ρ, and κ 52

3.1.3 Simulation study 54

3.2 Results 57

3.2.1 Simulation study 57

3.2.2 Application to Oryx Data 69

3.3 Summary 71

4 Improving HMMs with High-Frequency Data 74

4.1 Methods 75

4.1.1 Hidden Markov Model framework for constant-frequency data . 75

4.1.2 Model Formulation 77

4.1.3 Model Fitting 82

4.1.4 Empirical case study: alpine goats 83

4.1.5 Simulation Study 85

4.2 Results 89

4.2.1 Simulated Data 89

4.2.2 Goat Data 96

4.3 Summary 108

5 Discussion 112

List of Tables

2.1 Estimated number of turning points (m) and distribution parameters for the oryx data under the von Mises and robust models. 41

3.1 Estimated number of turning points (m) and distribution parameters for the oryx data under the angular AR(1) model 71

4.1 ΔAIC values for the different models across goats (Section 4.1.4). . . . 105

4.2 Emission parameter estimates for the models selected using ΔAIC and applied to the goat movement data. 107

List of Figures

1.1 A path segment extracted from a 1Hz trajectory of a free-ranging goat. The top left panel shows the original observed path, while the top right panel displays the demarcated path obtained using the detected turning points. The bottom left and bottom right panels show the trajectory subsampled at regular intervals of 30 and 60 minutes, respectively. . . . 3

2.1 Examples of different error distribution densities. Left: the density function of von Mises errors with $\kappa = 2$. Right: the density function of robust von Mises errors with $\kappa = 2$ and $c = 0.5$. 20

2.2 Top: simulated heading series with a single turning point assuming independent robust von Mises errors with $\kappa = 10$ and $c = 0.5$. Bottom: simulated series with a single turning point assuming an angular autoregressive process around the mean direction with $\kappa = 10$. 21

2.3 A simple example showing the function $F(t, \mu)$ for the first two iterations. The function $F(t, \mu)$ is plotted as a bold line. The colours indicate the time of the most recent turning points (red for $t = 1$ and black for $t = 2$). The dotted horizontal line indicates $F(1) + \frac{\beta}{\kappa}$. 25

2.4 The function $F(1500, \mu)$ under different error model choices in a simulation example. The function $F(1500, \mu)$ is shown as a bold line. Colours indicate the time of the most recent turning points. Top: when $c = -1$ (i.e., von Mises errors). Bottom: $c = 0.5$. The true last turning point is at 1401. 27

2.5 Simulated paths of headings under four scenarios. The first row shows scenario (A) with $\kappa = 5$ under a von Mises error model; the third row shows scenario (A) under a wrapped Cauchy error model with the corresponding concentration parameter. The second and fourth rows correspond to scenario (B) with the von Mises and wrapped Cauchy error models, respectively. In the first column, blue crosses indicate the true locations of the turning points. In the second column, red crosses show the turning-point estimates assuming an i.i.d. von Mises error model. The third column displays the estimates obtained using the i.i.d. robust von Mises model. 33

2.6 A short segment of one of the simulated trajectories under different error structures. **Left:** Errors generated using the von Mises distribution. **Right:** Errors generated using the wrapped Cauchy distribution. Each row represents a different level of concentration. For the von Mises distribution (left), the values of κ used are 3, 5, and 9. For the wrapped Cauchy distribution (right), the corresponding concentration values are computed using $I_1(\kappa)/I_0(\kappa)$. The vertical blue lines indicate the inferred turning points under the assumption of von Mises errors, red lines represent the turning points from the robust estimation method, and black lines indicate the ground truth. 35

2.7 F1-score (Equation 2.30) as a function of the concentration parameter κ, comparing the performance of three methods: the model assuming von Mises-distributed errors (black), the robust model (red), and the sliding-window approach (blue). The columns correspond to scenarios A, B, and C (see Section 2.1.5), while the rows represent different tolerance levels. All simulations were conducted in the absence of outliers. 39

2.8 F1-score (Equation 2.30) as a function of the concentration parameter κ, comparing the performance of three methods: the model assuming von Mises-distributed errors (black), the robust model (red), and the sliding-window approach (blue). The columns correspond to scenarios A, B, and C (see Section 2.1.5), while the rows represent different tolerance levels. All simulations were conducted under heavy-tailed errors distribution. 40

2.9 Reconstruction of two oryx paths from magnetometer data under the assumption of constant speed. The first row displays the observed paths. The second row shows reconstructions based on turning points inferred using the von Mises model. The third row corresponds to inference under the robust model, and the fourth row shows results from the sliding-window algorithm. 42

2.10 Error density functions for all oryx under the two error model choices. . 43

2.11 Timing comparisons between the three algorithms with different input value of c in case of the robust model. 45

3.1 Top: simulated heading series with a single turning point assuming independent von Mises errors ($\rho = 0$) with $\kappa = 15$. Bottom: simulated series with a single turning point assuming an angular autoregressive process around the mean direction with $\kappa = 15$ and $\rho = 0.80$. 51

3.2 Simulated paths of headings under different scenarios. The first row shows the simulated path for scenario (A), with $\kappa_\rho = 200$ and $\rho = 0.98$. The second, third, and fourth rows correspond to scenario (B), with $\kappa_\rho = 5$, 40, and 200, and $\rho = 0.75$, 0.75, and 0.98, respectively. Blue crosses in the first column indicate the ground-truth locations of turning points, while red crosses in the second column show the inferred turning points using the angular AR(1) model. The third column displays the turning points inferred using the IID model. 55

3.3 Error in estimation of ρ under the different scenarios from Section 3.1.3. Columns represent scenarios A, B, and C from left to right. Rows represent different κ_ρ values, with $\kappa_\rho = 5$, 15, 45, 100 , 200 and 400 from top to bottom. 59

3.4 Relative error in estimation of κ_ρ under different scenarios. The columns represent A, B, and C from left to right (see Section 3.1.3). Rows represent different ρ values, with $\rho = 0$, $\rho = 0.75$, and $\rho = 0.95$ from top to bottom. The black boxplots represent the errors made by the robust estimator, while the red boxplots represent the errors from the maximum likelihood estimator. 60

3.5 F1-score (Equation 2.30) as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and the sliding-window approach (blue). The results correspond to scenario A (see Section 3.1.3). Each column presents the tolerance levels 5, 10, and 20, while each row corresponds to one value of κ_ρ: 5, 15, 45, 100, 200, and 400, respectively. 62

3.6 F1-score (Equation 2.30) as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and the sliding-window approach (blue). The results correspond to scenario B (see Section 3.1.3). Each column presents the tolerance levels 5, 10, and 20, while each row corresponds to one value of κ_ρ: 5, 15, 45, 100, 200, and 400, respectively. 63

3.7 F1-score (Equation 2.30) as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and the sliding-window approach (blue). The results correspond to scenario C (see Section 3.1.3). Each column presents the tolerance levels 5, 10, and 20, while each row corresponds to one value of κ_ρ: 5, 15, 45, 100, 200, and 400, respectively. 64

3.8 The figure presents line plots of the F1-score with wrapped normal noise as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and sliding-window (blue). The first row presents the results for scenarios A, B, and C (see Section 3.1.3) with $\kappa_\rho = 5$, 15, 45, 100, 200, and 400 in each row, respectively. 65

3.9 The figure presents line plots of the F1-score with angular AR(3) noise as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and sliding-window (blue). The first row presents the results for scenarios A, B, and C (see Section 3.1.3) with $\kappa_\rho = 5$, 15, 45, 100, 200, and 400 in each row, respectively. 67

3.10 The figure presents line plots of the F1-score with AR(10) noise as a function of the autocorrelation parameter ρ, comparing the performance of three methods: AR(1) (red), IID (black), and sliding-window (blue). The first row presents the results for scenarios A, B, and C (see Section 3.1.3) with $\kappa_\rho = 5$, 15, 45, 100, 200, and 400 in each row, respectively. . 68

3.11 The figure presents line plots of the F1-score as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and sliding-window (blue) when κ_ρ is not constant. From left to right each presents the results for scenarios A, B, and C (see Section 3.1.3). 69

3.12 Inferring turning points from oryx data. (a): The observed bearing data (black) with the moving median estimate overlaid (red). (b): Centralized bearings based on the moving median. (c): Circular autocorrelation of the residuals. (d): Whitened centralized bearings after accounting for autocorrelation. (e): the path reconstructed from magnetometer data under the assumption of constant speed with red dots indicating the inferred turning points based on the angular AR(1) model. (f): presents the resulting model of the path as a sequence of steps and turns. 70

4.1 Example plots of Equation 4.7. 80

4.2 Conceptual framework for modelling behavioural states using high-frequency movement data. The first column shows the original observed path and the decomposed trajectory into steps and turns. The second column displays the distributions of key movement metrics, potentially combined with environmental covariates. The third column illustrates the Hidden Markov Model (HMM) structure used for inference. The last column presents various examples of post-hoc visualisations, explained in detail later. 81

4.3 Time series plots of simulated state sequences across six scenarios. The columns show the state sequences over time for Scenarios A, B, and C. The y-axis represents the discrete state values, and the x-axis denotes simulation steps. The first row corresponds to the two-state models, while the second row corresponds to the three-state models. 89

4.4 Comparison between the standard HMM and the proposed HMM-based approach applied to simulated movement data from Scenario A in a two-state case. Example decoded state sequences are shown under different subsampling levels, where blue represents State 1 and red represents State 2. 91

4.5 Results of the simulation analysis for the two-state model. The first row displays classification accuracy. Red boxplots represent results obtained using the TPs method under oracle settings, while blue boxplots correspond to results from regular subsampling at constant time steps. The second row shows the relative errors in parameter estimation for each state under oracle settings, and the third row shows the relative errors under TPs settings. Columns correspond to scenarios A, B, and C. . . 92

4.6 Results of the simulation analysis for the three-state model. The first row displays the classification accuracy. Red boxplots represent results obtained using the TPs method under oracle settings, while blue boxplots correspond to results from regular subsampling at constant time steps. The second row shows the relative errors in parameter estimation for each state under oracle settings, and the third row shows the relative errors under TPs method. The third Columns correspond to scenarios A, B, and C in the three-states case. 94

4.7 Comparison between the standard HMM and the proposed HMM-based approach applied to simulated movement data from Scenario A in a three-state case. Example decoded state sequences are shown under different subsampling levels, where blue represents State 1, red represents State 2 and orange represents State 3. 95

4.8 Results of the simulation analysis for the two-state model. The top row shows classification accuracy. Red boxplots represent results obtained using the turning-points (TPs) method under oracle settings, while blue boxplots correspond to regular subsampling at constant time steps. The second row shows relative errors in parameter estimation under oracle settings, and the third row shows relative errors under turning-point detection settings. Columns correspond to scenarios A, B, and C. All results assume turning angles follow a von Mises distribution, step times an exponential distribution, and step speeds a gamma distribution. . . 97

4.9 Results of the simulation analysis for the three-state model. The top row shows classification accuracy. Red boxplots represent results obtained using the turning-points (TPs) method under oracle settings, while blue boxplots correspond to regular subsampling at constant time steps. The second row shows relative errors in parameter estimation under oracle settings, and the third row shows relative errors under turning-point detection settings. Columns correspond to scenarios A, B, and C. All results assume turning angles follow a von Mises distribution, step times an exponential distribution, and step speeds a gamma distribution. . . 98

4.10 Comparison of the standard HMM with the proposed HMM-based approach, applied to one of the goats over a 1-day trajectory. The decoded state sequences are shown under different subsampling levels, where blue represents State 1 and red represents State 2. These results are based on the combined model. 99

4.11 Mean matching scores across all goats, comparing the state sequences inferred by the proposed HMM with those from a standard HMM, evaluated at various subsampling intervals. 100

4.12 Estimated transition probabilities between behavioural states in the two-state HMM as functions of distance to the pen and time of day for an individual goat. The left column shows the probability of transitioning from State 1 (active) to State 2 (inactive), while the right column shows the reverse transition from State 2 to State 1. 101

4.13 Post-hoc analysis of the distribution of State 1 and States 2 as functions of distance to the centre and time of day for all goats under the homogeneous model; red indicates the inactive state, while blue indicates the active state. 102

4.14 State-to-state transition probabilities under the homogeneous model, the distance-from-centre model, and the time-of-day model for the goat data. The left column shows transition probabilities for each individual from state 1 to state 2, the right column the transition probabilities from state 2 to state 1. The results for the homogeneous model, based on a post hoc inference on the effects of distance from centre and time of day, are shown in the first and second row, respectively. The third and fourth rows correspond to the estimated transition probabilities between behavioural states in the distance-based and time-of-day models, respectively, for each goat. 104

Nomenclature

Nomenclature

Physics Constants

AR(s) s-order Autoregressive process

ARGOS Advanced Research and Global Observation Satellite system

BIC Bayesian Information Criterion

CUSUM Cumulative Sum method

DP Douglas–Peucker algorithm

F1-score Harmonic mean of precision and recall (performance metric)

FPOP Functional Pruning Optimal Partitioning

GPS Global Positioning System

HMM Hidden Markov Model

IID Independent and Identically Distributed

LRT Likelihood Ratio Test

MA(s) Moving Average process of order s

MRPA Multi-Resolution Polygonal Approximation

OP Optimal Partitioning

PELT Pruned Exact Linear Time algorithm

SCSD Squared Circular Standard Deviation

SSA Step Selection Analysis

TPs Turning Points

VHF Very High Frequency (tracking system)

VW Visvalingam–Whyatt algorithm

WBS Wild Binary Segmentation

Chapter 1

Introduction

Movement lies at the core of nearly all animal life-history strategies (Nathan et al., 2008). Whether seeking food, avoiding predators, migrating, or finding mates, animals must continually make navigational decisions about where and when to move. Understanding the underlying reasons for these decisions is a central goal of behavioural ecology and has far-reaching implications in conservation biology, landscape ecology, and disease dynamics (Boulinier et al., 2016; Katzner and Arlettaz, 2020). However, inferring these decision-making processes from locational data is not trivial. The frequency at which movement data are collected can dramatically affect the behavioural inferences we can draw.

When data are collected at low frequencies, such as one GPS fix every few minutes or hours, key behavioural decisions may be entirely missed. For example, Figure 1.1 shows a goat's trajectory based on 1Hz observations. Subsampling at 30- or 60-minute intervals does not accurately represent the actual movement process. If a straight lines are fitted between consecutive turns and calculate the Euclidean distance of the observed trajectory from these lines, deviations can reach 0.212 km for 30-minute subsampling and 0.420 km for 60-minute subsampling. Moreover, several distinct turns and transitions between sites are entirely lost. Conversely, high-frequency data can cap-

ture movement decisions at the exact moment they occur, though they present unique analytical and computational challenges. Much of this data often reflects continued movement along previously established trajectories (see Figure 1.1), providing little new behavioural insight (Williams et al., 2020; Lewis et al., 2018). The so-called "big data problem" is particularly acute in modern movement ecology, where biologging devices now allow researchers to record animal movement multiple times per second (Bidder et al., 2015; Nathan et al., 2022). These advances create an urgent need for statistical tools capable of extracting behavioural meaning from vast, noisy, and highly autocorrelated data streams.

A promising strategy is to focus analysis on key locations where animals likely make active decisions, particularly, points at which they make navigational decisions. These turning points often reflect important behavioural decisions, and analysing their spatial and temporal patterns can reveal how animals interact with their environments (Gunner et al., 2024). For example, correlations between turning behaviour and environmental covariates can inform models of habitat selection and foraging strategy (Munden et al., 2021).

1.1 High-Frequency Movement Data

A major reason for tracking free-ranging animals is to provide data for accurate reconstruction of movement trajectories through time and space. This has driven decades of innovation in biologging technology, leading to a suite of tools that underpin modern movement ecology (Kays et al., 2015). Decades of technological developments have created vast possibilities in terms of data-gathering to study the movement of organisms, from early very high frequency (VHF), ARGOS system and Global Positioning System (GPS) to reverse-GPS, and integrated sensor networks (Williams et al., 2020).

At the core of these methods is the use of miniaturised tracking devices, typically attached via collars, harnesses, or adhesive mounts. These tools must balance data

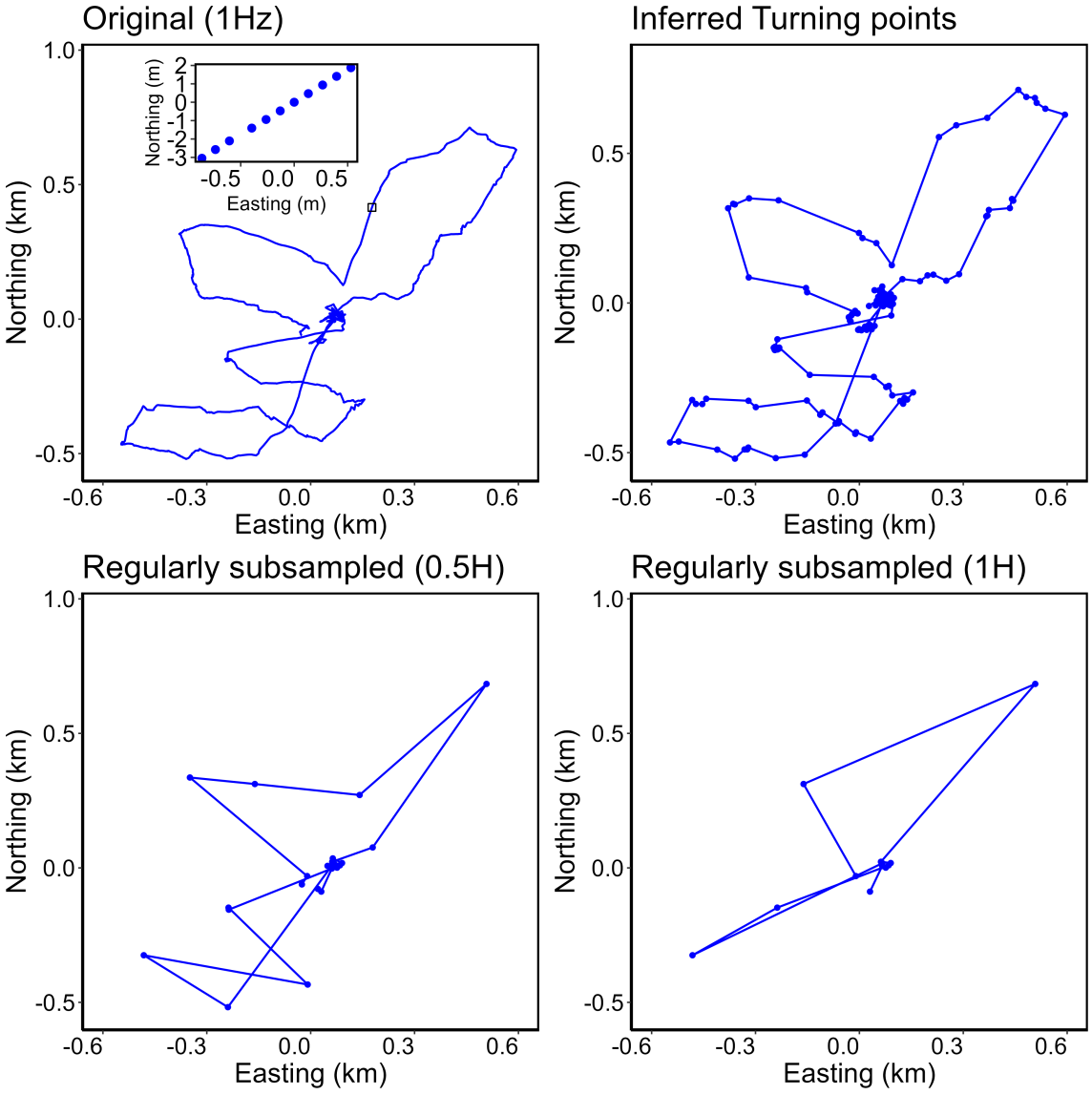


Figure 1.1: A path segment extracted from a 1Hz trajectory of a free-ranging goat. The top left panel shows the original observed path, while the top right panel displays the demarcated path obtained using the detected turning points. The bottom left and bottom right panels show the trajectory subsampled at regular intervals of 30 and 60 minutes, respectively.

quality, battery life, physical size, and durability under field conditions (Cooke et al., 2004; Patterson et al., 2017). The most common tracking technology is the GPS, which uses satellite triangulation to estimate geographic positions at different intervals. Tracking devices embedded with GPS usually provide highly accurate location estimates and can produce anything from sparse daily fixes to dense streams of coordinates at minute-level intervals (Tomkiewicz et al., 2010). However, transmitting data from GPS devices is energy-intensive and therefore costly, requiring high-power batteries that increase device weight and limit their use on smaller terrestrial species. Although not as accurate as GPS, the ARGOS systems represent another widely used satellite-based tracking technology, particularly suited to small terrestrial species because it requires less power and allows for lighter, smaller transmitters that can operate over extended periods (Hooten et al., 2017).

However, movement is more than a series of discrete locations. To gain a deeper understanding of how animals navigate their environment, researchers increasingly employ a variety of high-frequency motion sensors, moving away from satellite-based tracking. These include accelerometers, which detect changes in velocity; magnetometers, which record heading; and gyroscopes, which capture body orientation (Wilson et al., 2008). The data from these sensors can be processed to reconstruct a so-called dead-reckoned path, by continuously updating locations based on speed and direction (Wilson et al., 2008). However, dead-reckoning is prone to cumulative error and typically requires periodic GPS correction to maintain positional accuracy (Bidder et al., 2015; Gunner et al., 2021; Redcliffe et al., 2025). Recently, reverse-GPS systems have been constructed for gathering high frequency data, often 1Hz or more (Nathan et al., 2022). Finally, drone-based video tracking has also been used to generate high-frequency movement trajectories, by combining videos with computer vision technology (Koger et al., 2023).

These advances allow researchers to reconstruct near-continuous movement paths and explore fine-scale behaviours such as foraging bouts, predator evasion, and environmen-

tal exploration (Nathan et al., 2008; Munden et al., 2019, 2021; Gunner et al., 2023). This shift from discrete point data to almost-continuous trajectories opens new opportunities for behavioural analysis, especially when integrated with statistical models that handle temporal complexity.

1.2 Ecological Implications

High-frequency data represent a major advancement in studying animal behaviour and decision-making during the relocation process, as they allow positions to be tracked with fine temporal precision often at the scale of seconds or even fractions of a second (Potts et al., 2018). This level of detail provides new opportunities to understand how animals make navigational decisions. By capturing subtle movement patterns such as rapid bursts of speed or sharp turns, high-frequency data reveal the fine-scale dynamics of relocations. When combined with high-resolution environmental covariates, these data open up a wide range of analytical opportunities, for example, estimating sites of interest (Munden et al., 2019), understanding how and why animals position themselves in space (Munden et al., 2021), and making fine-scale behavioural inferences (Conners et al., 2021; Rautiainen et al., 2022).

In movement ecology, animal trajectories are typically represented as sequences of step lengths and turning angles (Bovet and Benhamou, 1988; Turchin, 1998; Turchin et al., 1991; Patterson et al., 2017). These descriptors have long formed the conceptual foundation of movement modelling underpinning approaches such as step selection analysis (Michelot et al., 2024), biased correlated random walks (Codling et al., 2008), a variety of state-space formulations (Patterson et al., 2017), and continuous-time movement models (Parton and Blackwell, 2017). The sequence of steps and turns is typically assumed to be regular or, at worst, irregular due to technological limitations. However, the timing between observations has generally been treated as arbitrary and determined by practical considerations, such as battery life, rather than by biological

relevance.

Nevertheless, applying traditional "step-and-turn" frameworks to modern high-frequency tracking data is often impractical. In such data, turning angles are typically close to zero for most consecutive observations, and step lengths become negligible when the displacement between fixes is smaller than the animal's body length. High-frequency data, however, allow researchers to interpret animal trajectories as sequences of discrete movement decisions and provide a basis for placing step-and-turn models on a more formal theoretical footing. Turning point analysis offers a systematic framework for studying animal movement by identifying the specific moments at which animals change direction or relocate (Gunner et al., 2024; Munden et al., 2021; Gunner et al., 2023).

From a behavioural perspective, turning points can serve as proxies for decision-making events such as those related to foraging, risk avoidance, vegetation attraction, or social interactions (Potts et al., 2018; Popp et al., 2025; Munden et al., 2021). For instance, a high density of turning points within a particular area may indicate zones of intensive search behaviour or local resource exploitation. In contrast, long, straight movements with few turning points are often indicative of transit or migratory behaviour. Moreover, when the timing and location of directional changes can be determined precisely, analysing how and why animals move between two locations can provide valuable insights into habitat selection processes, as shifts in direction are frequently triggered by environmental factors such as topography, vegetation structure, or the presence of conspecifics (Munden et al., 2021; Gunner et al., 2023). Therefore, turning points themselves are important structural features of an animal's trajectory that can be extracted to represent movement in a simplified yet behaviourally relevant form, closely reflecting real decision-making events and allowing further ecological analysis (Gunner et al., 2024).

1.3 Statistical Change Point Analysis

In the analysis of time series, a turning point is a particular example of a 'change point'. Analysis of change points has a rich background, summarised here to provide a general context for turning point analysis. A change point refers to a position within a data sequence where the underlying statistical properties of the generative model shift. For instance, consider a simple constant-signal model of the form $x_t = \mu + \epsilon_t$, which assumes a constant mean μ with random noise ϵ_t. A change point is defined by a shift in either the mean of the signal μ or change in the variance of the noise σ_ϵ^2, or both. This basic example can be generalized to accommodate more complex phenomena.

When the locations of change points are known, the data can be partitioned and analysed accordingly. However, identifying change points is computationally challenging, particularly when both their number and positions are unknown. For instance, detecting 10 change points within a sequence of 1,000 observations naively would require evaluating over $\binom{1000}{10} \approx 2.63 \times 10^{23}$ possible segmentations. The computational burden increases even further when the number of change points is itself unknown, as we must evaluate not only all combinations for a fixed number of changes but also for all possible counts $m = 0, 1, 2, \ldots, n$. In this case, the total number of segmentations grows exponentially with n, approximately on the order of 2^{n-1}, which for $n = 1000$ exceeds 10^{300} possible partitions. This immense combinatorial explosion underscores the need for efficient algorithms capable of accurately detecting unknown change points in time series data ideally achieving linear or near-linear computational complexity (Killick et al., 2012; Maidstone et al., 2017).

Change point detection has wide-ranging applications across several disciplines, including movement ecology, where it has been used to identify biologically meaningful events such as behavioural responses to capture (flee and return), parturition, stress reactions, mortality events, migration phases, and nesting transitions from animal biotelemetry data such as locations, heart rate, collar temperature, and overall dynamic body accel-

eration (Wolfson et al., 2022); financial time series analysis, where it has been applied to detect structural changes in the equity premium by identifying reductions in market volatility and expected returns (Kim et al., 2005); calcium imaging data, where it is used to analyse neuronal activity through changes in fluorescence intensity (Jewell and Witten, 2018; Jewell et al., 2020); and DNA sequence analysis, where it has been employed to identify abrupt shifts in mean signal levels along genomic sequences such as copy number variations in array comparative genomic hybridization data (Muggeo and Adelfio, 2011).

Page (1954) were motivated by a quality control problem to develop the first statistical method for estimating a single change point. Today, research and applications in statistical change point analysis are expanding. The CUSUM (Cumulative Sum) method, introduced by Page (1955), is designed to detect changes in the mean or variance of a stochastic process. It operates by cumulatively summing the residuals (or squared residuals, in the case of variance detection) and identifying points where the sum systematically deviates from zero.

Hawkins et al. (2003) introduces a Likelihood Ratio Test (LRT) approach to change-point analysis. This evaluates at each potential change point whether the data are better modelled by a single distribution (null hypothesis) or by two separate distributions before and after that point (alternative hypothesis). A significant increase in log-likelihood under the two-segment model indicates the presence of a change point. In contrast, the CUSUM method is more data-driven, relying on cumulative deviations from the mean rather than explicit likelihood comparisons, while the LRT provides a formal statistical test grounded in model-based likelihood inference.

The first attempt at multiple change-point detection was Binary Segmentation, introduced by Scott and Knott (1974). It begins by searching for a single change point in the entire dataset using any available method (e.g., CUSUM or LRT). If one is detected, the data are split into two segments at that point, and the process is repeated recursively on each segment until no further change points are found. While Binary

Segmentation is computationally efficient and simple to implement, however, it relies on a global search at each step, meaning that it always selects the most dominant change point within the current segment. As a result, it may fail to detect change points that are relatively weak or that occur close to stronger ones, since these can be masked during the global optimisation step (Fryzlewicz, 2014).

To overcome this limitation, Fryzlewicz (2014) proposed the Wild Binary Segmentation (WBS) method, which improves detection accuracy by repeatedly applying the segmentation procedure to many randomly selected subintervals of the data. Each subinterval is independently tested for potential changes and the strongest candidate across all intervals is chosen as the next change point. By localising the search within random subsegments, WBS is able to detect multiple and closely spaced change points more effectively than standard Binary Segmentation.

An alternative non-parametric method is a window-based approach (Truong et al., 2020). It works by sliding a small window along the signal and calculating a function that measures variability. The choice of variability measure depends on the underlying nature of the question for example, for data in the real line, variance for changes in mean, or absolute deviations for changes in variance. Sudden spikes in this function indicate a potential change point.

In addition, the performance of the previous methods depends on predetermined choices. For instance, the CUSUM and LRT approaches usually require a threshold to determine how large the CUSUM value or likelihood ratio must be to accept a change, while the window-based method strongly depends on the choice of window size: a window that is too large may smooth out short or closely spaced changes, whereas one that is too small may lead to excessive false detections due to noise. Nevertheless, neither Binary Segmentation nor sliding window approaches are guaranteed to recover the globally optimal segmentation, as they rely on greedy or local decision rules rather than explicitly minimising a global cost function over all possible segmentations (e.g., Jackson et al. (2005); Killick et al. (2012)).

Jackson et al. (2005) proposed a dynamic programming-based approach called *Optimal Partitioning* (OP), which finds the exact optimal segmentation of the data. The algorithm minimises a global cost function augmented with a penalty term to discourage overfitting:

$$\sum_{i=1}^{m+1} \Big[C\big(x_{(\tau_{i-1}+1):\tau_i}\big) + \beta \Big], \tag{1.1}$$

where $C(\cdot)$ denotes the segment cost, with the negative log-likelihood being a common choice; m is the number of change points; and β is a penalty term controlling model complexity, often set according to the Bayesian Information Criterion (BIC; Schwarz, 1978).

The key idea of OP is to solve this optimisation problem efficiently using dynamic programming. Rather than considering all possible segmentations simultaneously, which would be computationally infeasible, the algorithm builds the solution incrementally. Specifically, for each time point t, it computes the minimum cost of segmenting the data up to t by considering all possible locations $s < t$ of the most recent change point. The optimal cost up to time t, denoted $F(t)$, satisfies the recursion

$$F(t) = \min_{0 \le s < t} \big\{ F(s) + C\big(x_{(s+1):t}\big) + \beta \big\}, \tag{1.2}$$

At each time point, the algorithm determines the most recent change point, making the problem more tractable than attempting to estimate all change points simultaneously. Unlike Binary Segmentation and window-based methods, OP benefits from an exact and globally optimal segmentation, grounded in a model-driven framework with strong mathematical foundations. However, OP has a computational cost of $O(n^2)$, which limits its applicability for large datasets. Killick et al. (2012) and Maidstone et al. (2017) later demonstrated that, under certain regularity conditions, the computation time of OP can be reduced to approximately linear in the size of the data by pruning

unnecessary calculations, leading to the development of the Pruned Exact Linear Time (PELT) and Functional Pruning Optimal Partitioning (FPOP) algorithms (described in detail in Chapter 2).

These methods provide general frameworks for addressing a wide variety of research questions, such as detecting changes in the mean, median, variance, quantiles, or shifts in regression or autoregressive coefficients (Gombay, 2008; Bai, 1997); changes in the rate of a Poisson process (Ward et al., 2025); and transitions in the parameters of more complex models, including generalized linear models (Wang et al., 2023) and survival models (Lee and Lam, 2020). They can also be extended to online settings, where change-point inference is performed sequentially as new data arrive in real time (Adams and MacKay, 2007). That said, the present thesis focuses on developing methods for detecting shifts in the heading of animals, a variable that is inherently circular. This requires integrating circular statistics with change point analysis to construct a principled, computationally efficient model capable of detecting turning events in large, highly correlated datasets.

1.4 Turning Points Analysis

To infer animal behaviour from high-frequency movement data, various approaches have been employed. Conners et al. (2021), Rautiainen et al. (2022) and Chakravarty et al. (2019) applied sliding windows of fixed size to extract meaningful behavioural information for albatrosses, reindeers and meerkats, respectively, whereas Gupte et al. (2022); Bailey et al. (2021b,a) used subsampling, which essentially involves selecting every k-th observation, as applied to ground beetles and stickleback fish. However, neither the choice of window size nor the subsampling rate guarantees that the resulting data correspond to the temporal scale of the underlying behavioural dynamics. Building on this limitation, several studies have aimed to identify the most informative scale of movement, often defined by biologically significant turns.

The concept of analysing turning behaviour in movement trajectories predates the era of high-frequency tracking. One of the earliest formal approaches was introduced by Turchin et al. (1991), who simplified insect paths into a series of straight segments ("moves") separated by turns. That method operates directly on the sequence of recorded locations $x_0, \ldots, x_k$, assessing whether the segment connecting x_0 and x_k remains approximately linear. If all preceding points $(x_0, \ldots, x_{k-1})$ lie within a small distance ε from the line joining x_0 and x_{k-1}, but the inclusion of the next point x_k causes the deviation to exceed ε, then x_{k-1} is designated as a turning point.

Byrne et al. (2009) proposed a non-parametric permutation test for detecting change in movement direction. For each candidate point, the test computes the resultant lengths of summed movement vectors before and after the point, R_k and R_q, and compares their combined length $(R_k + R_q)$ with the resultant of the entire segment, R_{k+q}. A large difference between $(R_k + R_q)$ and R_{k+q} indicates a potential change in direction, and its statistical significance is assessed by comparing the observed test statistic

$$T = R_k + R_q - R_{k+q}$$

to a null distribution generated through random permutations of the vector order. In practice, the parameter q determines the number of consecutive movement vectors used to characterise the direction of travel before the candidate point, and thus influences the sensitivity of the change-point test

Extending this idea, Codling and Plank (2011) introduced two methods for identifying "turning events": the local and non-local approaches. In the non-local method, a turn is defined when the current movement direction deviates by more than a threshold angle θ from the direction at the previous turning event, whereas the local method examines only consecutive points, marking a turn whenever the angle between successive step vectors exceeds θ. In contrast to these angle-based methods, Humphries et al. (2013) proposed projecting the trajectory onto a single axis (e.g., latitude or longitude) and identifying turning points as the positions where movement along that axis changes

direction–that is, where successive one-dimensional displacements exceed a threshold distance.

However, the Turchin et al. (1991), Byrne et al. (2009), Codling and Plank (2011), and Humphries et al. (2013) methods were developed for low- or medium-resolution data, making them less suitable for the large data volumes produced by modern biologging technologies (Potts et al., 2018). To address this limitation, Potts et al. (2018) proposed an algorithm referred to here as the sliding-window algorithm for detecting turning points in high-frequency movement data. This algorithm moves a fixed-size window w along a sequence of headings and, within each window, computes the squared circular standard deviation (SCSD) as a measure of directional variability. Whenever the SCSD rises sharply above its mean level, the algorithm flags the corresponding position as a potential turning point. The midpoints of these peaks are then collected, and the resulting set of candidate turning points is further refined by examining the actual turning angle at each candidate. Any candidate for which the turning angle is smaller than a specified threshold θ is discarded.

Beyond methods explicitly developed for behavioural analysis, several polygonal simplification algorithms originating from fields such as cartography and computer graphics can also be adopted to identify turning points or simplify animal movement trajectories. Among the most widely used are the Douglas-Peucker (DP), Visvalingam–Whyatt (VW), and Multi-Resolution Polygonal Approximation (MRPA) algorithms. Although originally intended to reduce the number of vertices in a polyline while preserving its geometric form, these algorithms have been repurposed to delineate movement "segments" separated by substantial changes in direction.

The DP algorithm recursively connects the start and end points of a trajectory segment with a straight line and identifies the intermediate point that deviates the most from this line. If the perpendicular deviation exceeds a predefined distance threshold ε, that point is retained as a vertex (i.e., a turning point), and the procedure is repeated on the resulting sub-segments until all deviations fall below ε Douglas and Peucker (1973). In

contrast, the VW algorithm iteratively removes points according to the effective area each contributes to the local geometry: for every triplet of consecutive points, the area of the triangle they form is computed, and the point with the smallest area indicating minimal influence on the overall shape is removed first. Points are progressively eliminated until the smallest retained area exceeds a user-defined threshold Visvalingam and Whyatt (1993). While both DP and VW preserve the geometric shape of a trajectory without considering its temporal dimension, the MRPA method generalises these ideas by integrating an error criterion that explicitly accounts for time Chen et al. (2012).

Recently, Popp et al. (2025) demonstrated that the methods discussed above are highly sensitive to the choice of arbitrary parameters choices, with small adjustments in threshold values leading to markedly different sets of detected turning points. The degree of this sensitivity varies among algorithms. Because such thresholds lack clear statistical interpretation, the resulting segmentations provide no explicit probabilistic criterion to distinguish genuine turning points from random fluctuations or measurement noise.

1.5 Contributions

This thesis addresses two central challenges in high-resolution behavioural analysis:

1. How can we detect turning points in high-frequency movement data without relying on arbitrary parameters?

2. How can we use these turning points to infer switches in an animal's behavioural mode?

To address the first challenge, I develop a likelihood-based turning point detection method that treats observed headings (e.g., from magnetometers) as noisy measurements of latent directional intentions. Building on the framework of Jackson et al. (2005) and Maidstone et al. (2017), the method identifies directional shifts without re-

lying on pre-defined angle thresholds or smoothing windows. Simulation experiments demonstrate its robustness across varying levels of measurement error and autocorrelation. An application to magnetometer data from free-ranging oryx illustrates its empirical value.

To address the second challenge, I extend the Hidden Markov Model (HMM) framework of Morales et al. (2004); Langrock et al. (2012) to accommodate irregular time intervals defined by behaviourally inferred turning points. This allows direct inference of behavioural states from discrete path events rather than from evenly spaced samples. Simulated path trajectories composed of straight segments and turns are used to evaluate model performance, demonstrating high accuracy in both state classification and parameter recovery. The full pipeline is then applied to dead-reckoned data from free-ranging goats.

In summary, this thesis presents a unified framework for extracting, modelling, and interpreting behavioural decisions from high-resolution animal movement data. By integrating efficient turning point detection with an extended HMM framework, it provides a scalable framework for modern movement ecology. The approach enhances the temporal resolution of behavioural inference and opens new opportunities for studying animal-environment interactions at fine temporal scales.

1.6 Thesis Structure

The remainder of this thesis is organised as follows:

1. Chapter 2 introduces the proposed turning point detection method, which infers latent directional shifts under i.i.d. noise. Different noise distributions are evaluated through simulation studies, and the method is applied to magnetometer data from oryx.

2. Chapter 3 builds upon this method to incorporate temporal dependence in the

observed headings, allowing for more realistic modelling of autocorrelated directional data.

3. Chapter 4 extends the hidden Markov model framework to operate on irregular sequences of behavioural events (turns). The model is validated on simulated trajectories and applied to dead-reckoned data from goats, demonstrating its utility for behavioural-state inference in high-resolution movement ecology.

Chapter 2

Turning Point Detection

In this chapter, I introduce a model-based framework for identifying turning points in high-frequency animal movement data. The key idea is that an animal's true movement direction tends to remain stable for short periods and only changes when the animal turns. My model captures this behaviour by assuming that the true direction is piecewise constant through time, while the observed bearings are noisy measurements of this underlying direction. Directional changes are therefore interpreted as turning points, which I estimate using a penalised likelihood approach under the assumption of independently and identically distributed noise. I consider two forms of noise: from a von Mises distribution or from a heavy-tailed distribution.

To evaluate the method, I test it on simulated trajectories composed of straight segments connected by sharp turns, with varying turn densities and levels of noise added to mimic both measurement error and natural variation in movement. Finally, I illustrate the model's practical utility by applying it to magnetometer data from free-ranging Arabian oryx, demonstrating its effectiveness in detecting turning behaviour in empirical datasets. Part of this work has been published in Alharbi et al. (2026).

2.1 Method

2.1.1 Model Formulation

The aim of this algorithm is to take a high-frequency time series of measured animal headings (e.g. from a magnetometer) and demarcate the points where the animal made a decision to turn from those that reflect noise in the data-gathering process or small fluctuations arising from gait. The required dataset is thus a sequence of n observed headings, evenly spaced in time, denoted by $\{h_t\}_{t=1}^n$, and measured in radians. We assume that the animal's true headings form a piecewise constant function of time as follows

$$h_t = \mu_t + \epsilon_t \pmod{2\pi}, \tag{2.1}$$

with $\mu_t = \mu_{t-1}$ for all $t \notin \{\tau_1, ..., \tau_m\}$. Here, $\{\tau_1, ..., \tau_m\}$ is a set of m turning points, μ_t represents the animal's actual heading at time t and remains constant between any two turning points, and ϵ_t are i.i.d. random variables, defined on the circle, with zero circular mean. Since we anticipate shifts in the animal's direction, estimating the parameter μ_t in Equation (2.1) can be viewed as a change point detection problem, where the goal is to identify shifts in the mean of the underlying process. We estimate these shifts by identifying each time step t such that $\mu_t \neq \mu_{t-1}$.

I start with a simple model whereby the ϵ_t are assumed to be i.i.d. from a von Mises distribution with zero mean and concentration parameter κ. The probability density function of the von Mises distribution takes the form

$$f_V(h \mid \mu, \kappa) = \frac{e^{\kappa \cos(h-\mu)}}{2\pi I_0(\kappa)} \qquad \mu \in \mathbb{R}\ ,\ \kappa > 0. \tag{2.2}$$

Accordingly, the log-likelihood is given by

$$\mathcal{L}(\{\mu_t\}_{t=1}^n | \{h_t\}_{t=1}^n, \kappa) = \kappa \sum_{t=1}^{n} \cos(h_t - \mu_t) - n \log(2\pi I_0(\kappa)). \tag{2.3}$$

In high-frequency movement data, deviations from the underlying direction may arise not only from measurement error but also from biological processes. For example,

animals may exhibit rapid micro-adjustments in heading due to gait or small obstacle avoidance, while issues like sensor misalignment can further compound these variations. These combined effects can produce occasional large angular deviations that are inconsistent with the von Mises assumption of light-tailed noise and cannot be viewed as behaviourally significant turns. To account for these outliers, I propose a robust alternative to the von Mises distribution. This custom density function, f_{RV}, clips the influence of large deviations by introducing a threshold parameter

$$f_{RV}(h \mid \mu, \kappa, c) \propto \exp\left(\kappa \max\left\{\cos(h-\mu), c\right\}\right) \tag{2.4}$$

where $c \in [-1, 1]$ is a robustness parameter. This distribution serves as a robust alternative by limiting the influence of extreme angular deviations through the parameter c, which acts as a threshold below which observations are treated as equally uninformative. It operates in a manner similar to the biweight model in linear space (Fearnhead and Rigaill, 2019). These two distributions are illustrated in Figure 2.1. When $c = -1$, the distribution simplifies to the von Mises distribution, whereas setting $c = 1$ yields a uniform distribution over the circle.

Figure 2.2 presents example simulations from the two proposed models in a scenario involving a single turning point. Given the density function introduced in Equation (2.4), the log-likelihood takes the following form:

$$\mathcal{L}(\{\mu_t\}_{t=1}^n | \{h_t\}_{t=1}^n, \kappa, c) = \kappa \sum_{t=1}^{n} \max\{\cos(h_t - \mu_t), c\} + Q(\kappa, c). \tag{2.5}$$

Here, $Q(\kappa, c)$ stems from the normalising term and does not depend on ϵ; it is given by

$$Q(\kappa, c) = \int_{-\pi}^{\pi} e^{\kappa \max(\cos(x), c)}\, dx. \tag{2.6}$$

The normalising constant $Q(\kappa, c)$ does not have a closed-form expression and is therefore computed numerically. Neither of the likelihood functions in Equation (2.3) or Equation (2.5) impose any constraints on $\mu_t - \mu_{t-1}$ at turning points. In particular,

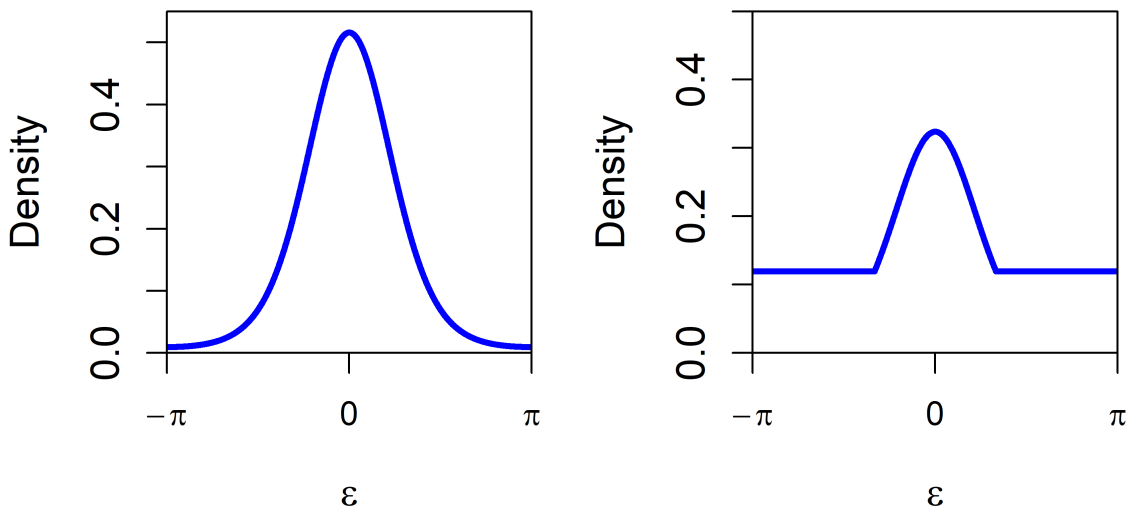


Figure 2.1: Examples of different error distribution densities. Left: the density function of von Mises errors with $\kappa = 2$. Right: the density function of robust von Mises errors with $\kappa = 2$ and $c = 0.5$.

directional changes are assumed to be isotropic, meaning that turns can occur with equal probability in any direction. This avoids modelling of the detailed motivation for a specific change in direction. However, in practice, very small angular changes may not correspond to sustained changes in movement direction and may instead arise from noise or minor adjustments. For this reason, in the simulation study in Section 2.1.5 I impose a minimum turning angle threshold, ensuring that only biologically significant turns are considered when evaluating performance. Importantly, this threshold is not part of the inference procedure itself, but is introduced solely for evaluation purposes.

2.1.2 Maximising the Likelihood Function

Since adding a new parameter to a model will always increase the likelihood, simply maximizing Equation (2.3) or Equation (2.5) is not sufficient; it is necessary to penalise for the number of turning points. To address this issue, I use an additional layer of fine-

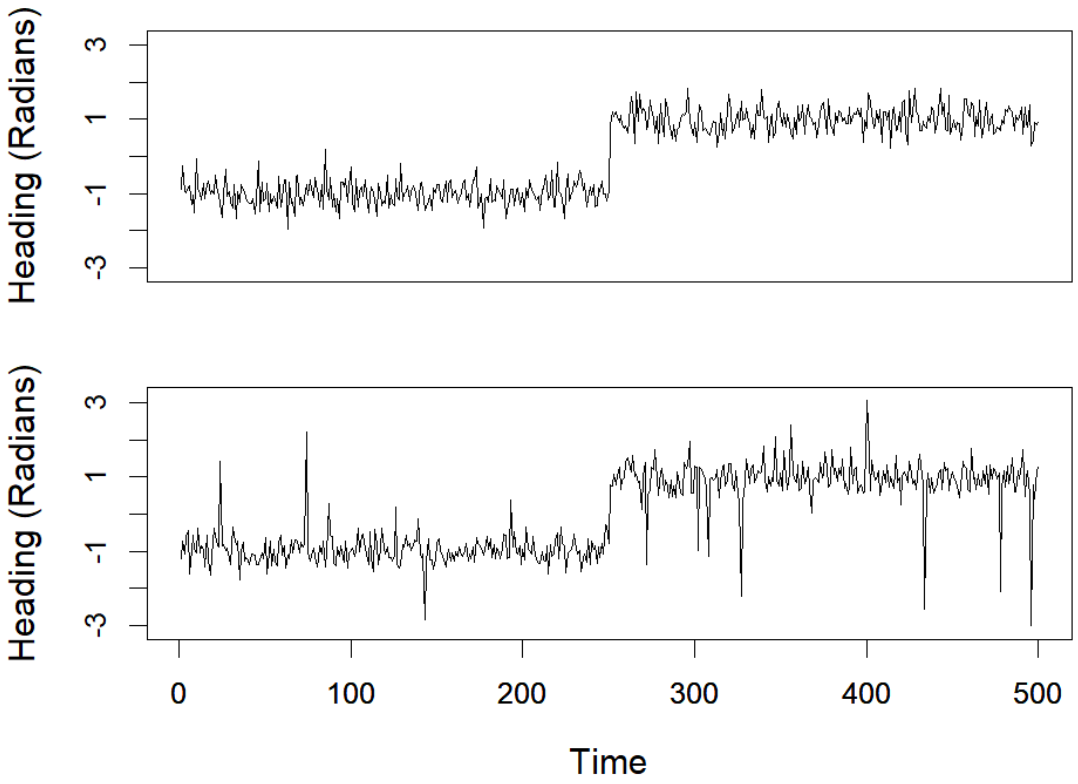


Figure 2.2: Top: simulated heading series with a single turning point assuming independent robust von Mises errors with $\kappa = 10$ and $c = 0.5$. Bottom: simulated series with a single turning point assuming an angular autoregressive process around the mean direction with $\kappa = 10$.

tuning to control the trade-off between introducing new turning points and achieving a good fit for the estimated headings. This enables the joint estimation of the true headings $\{\mu_t\}_{t=1}^n$, the number of turning points (m), and their locations. The resulting optimisation problem is

$$\hat{\mu}_1, \ldots, \hat{\mu}_n = \underset{\mu_1,\ldots,\mu_n}{\operatorname{argmax}} \left(\mathcal{L}(\{\mu_t\}_{t=1}^n | \{h_t\}_{t=1}^n, \kappa, c) + m\beta \right), \tag{2.7}$$

where $\beta < 0$ is the penalty term, which will be determined later.

To solve the optimisation problem in Equation (2.7), I build upon the Functional Pruning Optimal Partitioning algorithm (FPOP) introduced by Maidstone et al. (2017).

The optimal penalised log-likelihood of the parameters, given the data up to time s, is denoted as follows:

$$\begin{aligned} F(s) &= \max_{\mu_1,\ldots,\mu_s} \sum_{t=1}^{s} \left(\max\{\cos(h_t - \mu_t), c\} + \frac{\beta_t}{\kappa} \right) \\ &= \max_{\mu_1,\ldots,\mu_s} \left(\sum_{t=1}^{s-1} \left(\max\{\cos(h_t - \mu_t), c\} + \frac{\beta_t}{\kappa} \right) + \max\{\cos(h_s - \mu_s), c\} + \frac{\beta_s}{\kappa} \right) \\ &= \max_{\mu_s} \left(\max_{\mu_1,\ldots,\mu_{s-1}} \sum_{t=1}^{s-1} \left(\max\{\cos(h_t - \mu_t), c\} + \frac{\beta_t}{\kappa} \right) + \max\{\cos(h_s - \mu_s), c\} + \frac{\beta_s}{\kappa} \right) \\ &= \max_{\mu_s} \left(F(s-1) + \max\{\cos(h_s - \mu_s), c\} + \frac{\beta_s}{\kappa} \right), \end{aligned} \tag{2.8}$$

where $\beta_t = \beta$ if $\mu_t \neq \mu_{t-1}$, and $\beta_t = 0$ otherwise; additionally, $c = -1$ under the von Mises error model.

The function $F(s)$ can be reformulated to depend on the current heading $\mu_s = \mu$ as follows:

$$F(s, \mu) = \max \left\{ F(s-1, \mu), F(s-1) + \frac{\beta}{\kappa} \right\} + \max\{\cos(h_s - \mu), c\}. \tag{2.9}$$

In this formulation, $F(s, \mu)$ represents the maximum penalized log-likelihood for the data up to time s, conditional on the most recent actual heading being μ. In the first part of Equation (2.9), we assume the animal did not turn at time $s-1$ and continued moving in the same direction μ. In the second part, we assume that the last turning point occurred at time $s-1$ and that the animal has just started a new direction with bearing μ. By employing a pruned dynamic programming approach, the algorithm efficiently loops through the heading series to calculate the function $F(s, \mu)$. This process ultimately allows for finding the exact optimal sequence of headings $\{\mu_t\}_{t=1}^{n}$ as well as both the number and locations of turning points, subject to the penalised log-likelihood criterion. The dynamic programming recursion in Equation (2.9) can be solved for both model choices. Details are found in the next section.

2.1.3 Solving the Recursion

I begin with the case where errors are independent and identically distributed from the von Mises distribution (i.e. $c = -1$). For this scenario, Equation (2.9) is customised in the following recursive form:

$$F(s,\mu) = \max\left(F(s-1,\mu), F(s-1) + \frac{\beta}{\kappa}\right) + \cos(h_s - \mu). \tag{2.10}$$

A sum of phase-shifted cosine functions of the same frequency can be represented as a single scaled cosine with an adjusted phase and offset, that is

$$\sum_{i=1}^{n} (\cos(\mu - h_i) + z_i) = \sqrt{X^2 + Y^2}\cos(\mu - \alpha) + \sum_{i=1}^{n} z_i, \tag{2.11}$$

where

$$X = \sum_{i=1}^{n} \cos(h_i), \quad Y = \sum_{i=1}^{n} \sin(h_i), \quad \alpha = \tan_2^{-1}(Y, X).$$

The function $\tan_2^{-1}(y, x)$ computes the angle (bearing) from the origin to the point (x, y), measured anticlockwise from the positive x-axis (typically representing due east). Accordingly, $F(s,\mu)$ can be expressed as a piecewise cosine function provided $F(s-1,\mu)$ is of that form. Each segment of this function can be written in the form $R\cos(\mu-\alpha)+z$. This behaviour is illustrated by starting at time $s = 1$. At this point, the function is

$$F(1,\mu) = \cos(h_1 - \mu). \tag{2.12}$$

The function represents a single cosine wave defined over the interval $I_0^1 = [-\pi, \pi)$, with a maximum value of 1. Hence, $F(1) = 1$. Proceeding to time step $s = 2$, we obtain

$$\begin{aligned} F(2,\mu) &= \max\left\{\cos(h_1 - \mu), 1 + \frac{\beta}{\kappa}\right\} + \cos(h_2 - \mu) \\ &= \begin{cases} \cos(h_2 - \mu) + 1 + \frac{\beta}{\kappa} & \text{if } \cos(h_1 - \mu) \leq 1 + \frac{\beta}{\kappa} \\ \cos(h_1 - \mu) + \cos(h_2 - \mu) & \text{otherwise.} \end{cases} \end{aligned} \tag{2.13}$$

This yields two distinct cases, depending on whether the previous cosine contribution exceeds the penalized threshold.

1. If $\cos(h_1 - \mu) \leq 1 + \frac{\beta}{\kappa}$, the function becomes a shifted cosine wave by adding $1 + \frac{\beta}{\kappa}$ to the second cosine term.

2. Otherwise, it is the sum of two waves: $\cos(h_1 - \mu) + \cos(h_2 - \mu) = R_{1,2}\cos(\alpha_{1,2} - \mu)$,

where

$$R_{u,s} = \sqrt{\left(\sum_{i=u}^{s} \sin(h_i)\right)^2 + \left(\sum_{i=u}^{s} \cos(h_i)\right)^2}, \tag{2.14}$$

$$\alpha_{u,s} = \tan_2^{-1}\left(\sum_{i=u}^{s} \sin(h_i), \sum_{i=u}^{s} \cos(h_i)\right). \tag{2.15}$$

Using the trigonometric identity given in Equation (2.11), the maximum value at time $s = 2$, denoted by $F(2)$, is obtained by comparing the maxima of each segment (i.e., $R + z$). The key insight is that the function evolves as a series of cosine waves over distinct intervals, as illustrated in Figure 2.3.

As we proceed through the iterations, it is necessary to track the coefficients (z, R, and α) along with the intervals

$$I_{s-1,s} = \left\{\mu \mid F(s,\mu) \leq F(s) + \frac{\beta}{\kappa}\right\}, \tag{2.16}$$

$$I_{u,s} = I_{u,s-1} \cap (I_{s-1,s})^c \text{ for } u < s - 1. \tag{2.17}$$

Therefore, the most probable last turning point location is estimated by looking at τ where $\underset{\mu}{\operatorname{argmax}}\, F(s,\mu) \in I_{\tau,s}$.

By pruning empty intervals and tracking the most recent turning points, the optimal segmentation is efficiently determined. To achieve this, the process begins at the final time point n by identifying the last turning point, τ_m. From τ_m, the stored sequence of preceding turning points is traced back iteratively until the beginning of the sequence is reached at $\tau = 0$.

This algorithm is an instance of the class of functional pruning algorithms proposed by Maidstone et al. (2017). Pseudo-code for the algorithm is given in Algorithm (1). Using the robust distribution does not lead to substantial changes in the recursion. The

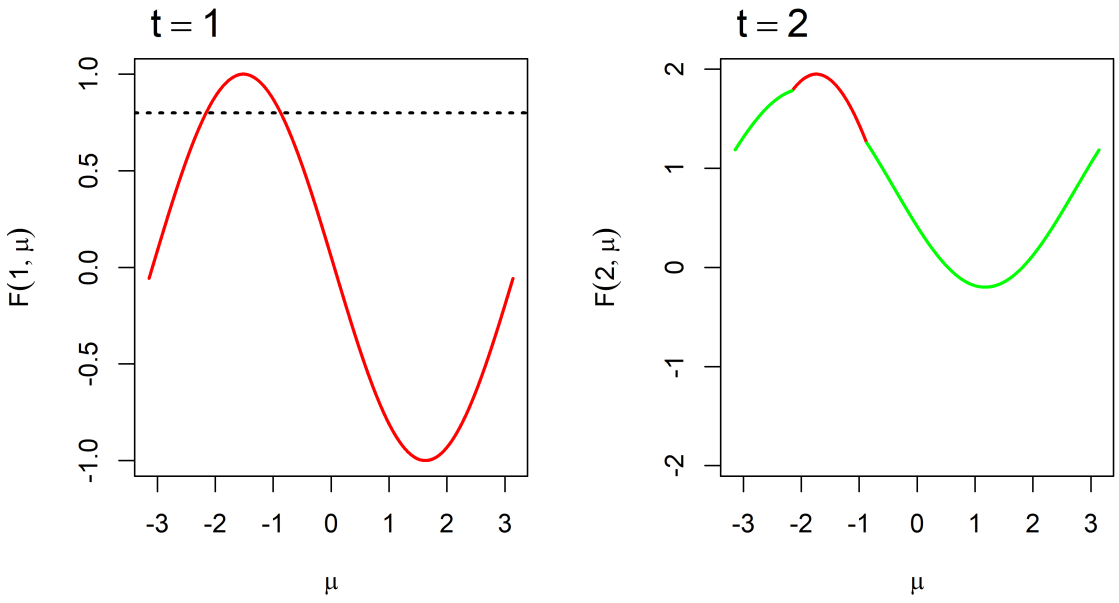


Figure 2.3: A simple example showing the function $F(t, \mu)$ for the first two iterations. The function $F(t, \mu)$ is plotted as a bold line. The colours indicate the time of the most recent turning points (red for $t = 1$ and black for $t = 2$). The dotted horizontal line indicates $F(1) + \frac{\beta}{\kappa}$.

function $F(s, \mu)$ can still be represented as a piecewise cosine function, although some regions may become flat segments where $R = 0$. Specifically, $F(1, \mu) = \max(\cos(h_1 - \mu),\ c)$ and step 7 of Algorithm 1 becomes

$$F(t,\mu) = \begin{cases} \max\left\{F(t-1,\mu), \max\limits_{\mu'} F(t-1,\mu') + \alpha\right\} + \cos(h_t - \mu), & \text{if } \cos(h_t - \mu) > c, \\ \max\left\{F(t-1,\mu), \max\limits_{\mu'} F(t-1,\mu') + \alpha\right\} + c, & \text{otherwise.} \end{cases} \tag{2.18}$$

Likewise, $I_{t-1,t}$ in step 8 becomes

$$I_{t-1,t} = \left\{\mu : F(t,\mu) = \max_{\mu'} F(t-1,\mu') + \alpha + \max(\cos(h_t - \mu),\ c)\right\}. \tag{2.19}$$

Figure 2.4 illustrates the differences in the calculation of $F(t, \mu)$ under the different error models.

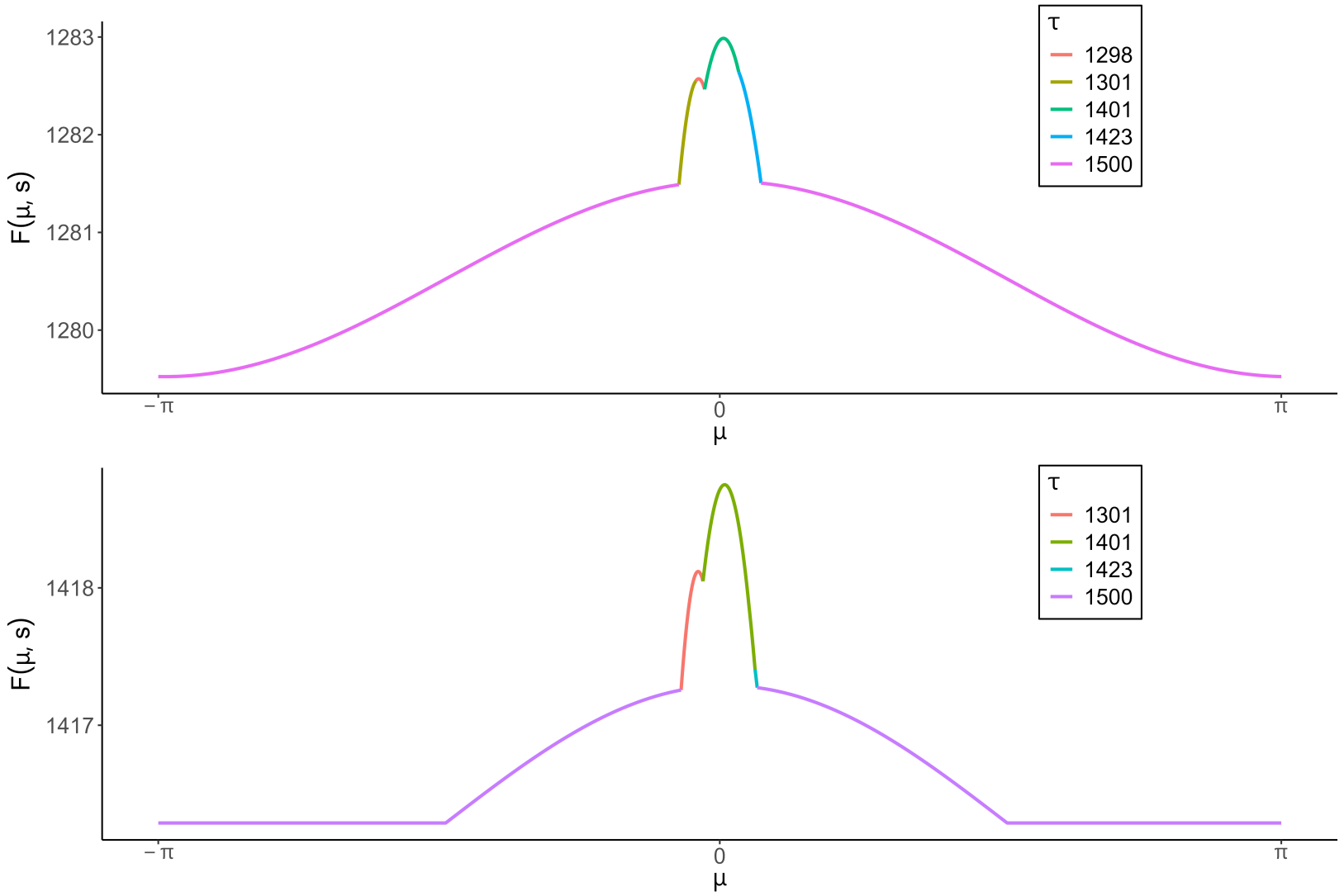


Figure 2.4: The function $F(1500, \mu)$ under different error model choices in a simulation example. The function $F(1500, \mu)$ is shown as a bold line. Colours indicate the time of the most recent turning points. Top: when $c = -1$ (i.e., von Mises errors). Bottom: $c = 0.5$. The true last turning point is at 1401.

2.1.4 Estimating β, c, and κ

The optimization problem presented in Equation (2.5) requires prior knowledge of three key parameters: the noise concentration (κ), the robustness parameter (c), and the penalty term (β). In this section, I outline approaches for determining these parameters from the data.

For the penalty parameter β, I adopt a form consistent with the Bayesian Information Criterion (BIC; Schwarz (1978)), a standard model selection method based on approximate Bayes factors which is widely used in changepoint detection problems (Fearnhead and Rigaill, 2019; Killick et al., 2012; Romano et al., 2022). The full BIC expression for this model is given by:

$$BIC = (2m + 3)\log(n) + n\log(2\pi I_0(\kappa)) - 2\kappa \sum_{t=1}^{n} \cos(\epsilon_t). \tag{2.20}$$

This can be simplified to the following form

$$BIC/2 = m\log(n) - \kappa \sum_{t=1}^{n} \cos(\epsilon_t) + \text{constant}, \tag{2.21}$$

where the final term does not involve the number or location of turning points. Comparing with Equations (2.7) and (2.5), I therefore take $\beta = -\log(n)$. This choice avoids the need for manual tuning and ensures that the number of detected turning points is selected in a principled, data-driven manner.

In principle, κ and c could be estimated jointly with the turning points; however, this would result in a highly non-convex optimization problem with substantial computational overhead. I therefore adopt a two-stage approach in which these parameters are estimated prior to segmentation. My primary objective is not exact parameter recovery, but rather to obtain values that sufficiently capture the scale and robustness of the noise to ensure reliable turning point detection. Below, I describe two distinct approaches for estimating κ and c.

Algorithm 1 Angular Functional Pruning Optimal Partitioning

Initialize:

$F(1, \mu) = \cos(h_1 - \mu)$

$I_{0,1} = (-\pi, \pi]$

$T = \{0\}$ ▷ Set of last turning points

$\alpha = \dfrac{\beta}{\kappa}$

for $t = 2$ to n **do**

Compute:

$$F(t, \mu) = \max \left\{ F(t-1, \mu), \max_{\mu'} F(t-1, \mu') + \alpha \right\} + \cos(h_t - \mu)$$

Compute intervals:

$$I_{t-1,t} = \left\{ \mu : F(t, \mu) = \max_{\mu'} F(t-1, \mu') + \alpha + \cos(h_t - \mu) \right\}$$

$$I_{i,t} = I_{i,t-1} \cap I^c_{t-1,t} \quad \text{for each } i < t-1$$

Update turning point candidates:

$$T \leftarrow T \cup \left\{ i : \operatorname*{argmax}_{\mu} F(t, \mu) \in I_{i,t} \right\}$$

Backtrack to find final turning points:

TPs $\leftarrow \{n\}$ ▷ Initialize with the end of the sequence

$\tau \leftarrow n$

while $\tau > 0$ **do**

$\tau \leftarrow T[\tau]$

TPs $\leftarrow$ TPs $\cup \{\tau\}$

return TPs

2.1.4.1 Robust estimators based on successive difference

Estimating κ and c is particularly challenging because the locations of the turning points are unknown. To address this, I exploit the fact that for sufficiently high concentration, the sum (and difference) of von Mises random variables is well approximated by another von Mises distribution. In particular, the circular successive differences with lag 1 are approximately von Mises distributed with zero mean. Under the small-angle approximation, the variance of a von Mises distribution satisfies $\mathrm{Var}(\theta) \approx 1/\kappa$; hence, taking differences of two independent variables doubles the variance, leading to an effective concentration parameter of approximately $\kappa/2$. This allows the estimation of noise characteristics without requiring estimation of the latent states μ_t.

The circular difference is defined as

$$d_t = h_t - h_{t-1} \pmod{2\pi}, \quad t = 2, \ldots, n. \tag{2.22}$$

For time points that are not adjacent to a turning point, the differences d_t can be treated as approximately independent random variables with zero mean and concentration parameter $\kappa/2$. In contrast, when a turning point occurs at time t, the corresponding d_t will typically exhibit a non-zero mean and therefore cannot be assumed to come from the described distribution.

Assuming that turning points are relatively rare, most values of d_t reflect the underlying noise process. This motivates the use of a robust estimator to reduce the influence of the turning points when estimating the concentration parameter. Accordingly, I estimate κ_d and relate it to κ via $\kappa = 2\kappa_d$, yielding

$$\hat{\kappa} = \frac{v}{1 - \tilde{R}}, \tag{2.23}$$

where

$$\tilde{R} = \sqrt{\overline{\sin(d)}^2 + \overline{\cos(d)}^2},$$

and $\overline{\sin(d)}$ and $\overline{\cos(d)}$ denote the medians of $\{\sin(d_2), \ldots, \sin(d_n)\}$ and $\{\cos(d_2), \ldots, \cos(d_n)\}$, respectively. The constant $v \approx 0.4592$ is a scale factor obtained via Monte Carlo simu-

lation, analogous to that used when employing the median absolute deviation (MAD) as an estimator of the standard deviation in linear data.

When assuming the robust von Mises model, κ_d and c_d are instead estimated jointly via maximum likelihood based on the differences d_t:

$$(\hat{\kappa}_d, \hat{c}_d) = \underset{\kappa_d, c_d}{\operatorname{argmax}} \left\{ \kappa_d \sum_{t=2}^{n} \max\left(\cos(h_t - h_{t-1}),\, c_d\right) - (n-1) \log\left(Q(\kappa_d, c_d)\right) \right\}. \quad (2.24)$$

These are then related to the original model parameters via

$$c = c_d \quad \text{and} \quad \kappa = 2\kappa_d. \quad (2.25)$$

Although this relationship is heuristic and does not guarantee full identifiability of (κ, c), it provides stable and practically useful estimates. Empirical results in Section 2.2.1 demonstrate that these estimates are sufficiently accurate for the purposes of segmentation.

2.1.4.2 Non-parametric approach

A non-parametric approach offers an alternative to the estimators described in the previous section. This method involves first obtaining a preliminary κ and c from the residuals relative to the estimated directions. One such non-parametric estimator is the moving circular median filter. Given a window of size w, the heading at time t is approximated as

$$\widetilde{\mu}_t = \tan_2^{-1}\left(\text{Median}\left(\sin(h_{(t-w/2):(t+w/2)})\right), \text{Median}\left(\cos(h_{(t-w/2):(t+w/2)})\right)\right). \quad (2.26)$$

The choice of window size is critical and, to some extent, arbitrary; however, it depends on the expected minimum segment length or on the specific biological question at hand (see Section 2.1.6 for an example).

Given the series $\widetilde{\mu}_t$ define the centralised bearing z_t as

$$z_t = h_t - \widetilde{\mu}_t \pmod{2\pi}. \quad (2.27)$$

Then κ and c can be easily be estimated from z_t by maximising the likelihood function which can be done numerically.

2.1.5 Simulation study

To evaluate the effectiveness of the proposed methods, I implemented a simulation study divided into three parts. I generated three sets of simulated trajectories, each containing 50,000 data points. This sample size was chosen to be large enough to mimic the structure of the oryx data (to be introduced and analysed in Section 2.1.6) while remaining computationally manageable for performing thousands of replications. The three scenarios varied by the number and distribution of turning points.

Scenario (A) contained 158 turning points, with segment lengths ranging from 17 to 1,602 time steps. Scenario (B) included 38 turning points, with segment lengths ranging from 167 to 4,542 time steps. In both scenarios, a minimum turning angle was enforced of $\pi/6$ radians by discarding any simulated bearings that did not meet this threshold. The number of turning points in scenarios A and B was drawn from a Poisson distribution with parameters 1/200 and 1/1000, respectively, while their locations were sampled from a discrete uniform distribution (see Figure 2.5). Scenario (C) was a straight trajectory with no turns. I then assessed each model's ability to detect turning points by introducing different error structures to the true path.

First, I simulated noise from the von Mises distribution across a range of concentration values, setting $\kappa \in \{1, 3, 5, \ldots, 19\}$. Second, to assess the effect of outliers on detection accuracy, I introduced noise generated from the wrapped Cauchy distribution. The scale parameter for the wrapped Cauchy was chosen as $I_1(\kappa)/I_0(\kappa)$, corresponding to each κ value considered, ensuring that its variance was comparable to that of the von Mises distribution.

The wrapped Cauchy distribution, known for its heavy tails, offers greater flexibility in generating outliers compared to the von Mises distribution. It was preferred over

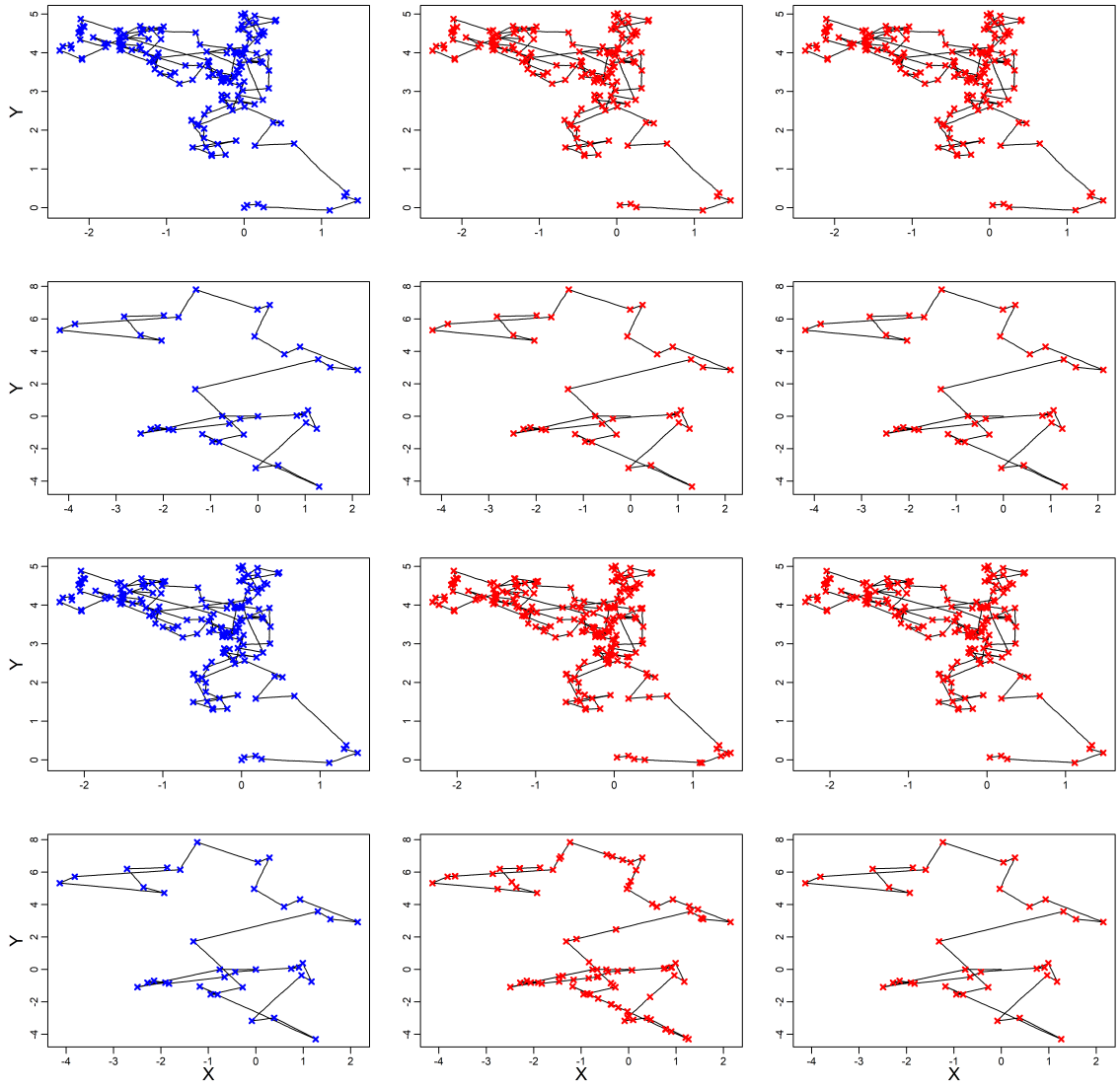


Figure 2.5: Simulated paths of headings under four scenarios. The first row shows scenario (A) with $\kappa = 5$ under a von Mises error model; the third row shows scenario (A) under a wrapped Cauchy error model with the corresponding concentration parameter. The second and fourth rows correspond to scenario (B) with the von Mises and wrapped Cauchy error models, respectively. In the first column, blue crosses indicate the true locations of the turning points. In the second column, red crosses show the turning-point estimates assuming an i.i.d. von Mises error model. The third column displays the estimates obtained using the i.i.d. robust von Mises model.

the robust von Mises distribution for simulating outlier-contaminated noise because the latter requires increasing the parameter c alongside κ to produce outliers thereby complicating control over the noise characteristics. In contrast, the wrapped Cauchy provides a direct and consistent mechanism for introducing outliers across different κ values. Nevertheless, in my inference procedure, I employ the robust von Mises distribution because it retains mathematical tractability, particularly enabling efficient optimization using the Functional Pruning Optimal Partitioning (FPOP) algorithm.

The difference between the two simulated error structures is visualised in Figure 2.6. Each simulation was repeated 50 times to ensure robust results. To assess the accuracy of the proposed method in estimating turning points, I define a turning point as correctly detected if it is inferred within w time steps of its true location. The choice of w is somewhat arbitrary but provides a practical and measurable criterion for evaluating model performance. Accordingly, Several tolerance levels were tested, $w \in \{5, 10, 20\}$.

An estimated turning point within w time steps of an actual one is classified as a true positive (TP); otherwise, it is a false positive (FP). Likewise, an actual turning point that is not matched by any estimated turning point within w time steps is considered a false negative (FN).

Based on these definitions, I compute the F1-score, a standard metric that balances precision and recall. The F1-score is the harmonic mean of precision (the proportion of detected changes that are correct) and recall (the proportion of true changes that are detected). It ranges from 0 to 1, with a score of 1 indicating perfect detection (Rijsbergen, 1979; Truong et al., 2020). The quantities are defined as follows

$$\text{Recall} = \frac{n(\text{TP})}{n(\text{TP}) + n(\text{FN})}, \tag{2.28}$$

$$\text{Precision} = \frac{n(\text{TP})}{n(\text{TP}) + n(\text{FP})}, \tag{2.29}$$

$$\text{F1-score} = 2 \cdot \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}, \tag{2.30}$$

where $n(\text{TP})$ denotes the number of true positives, and so on.

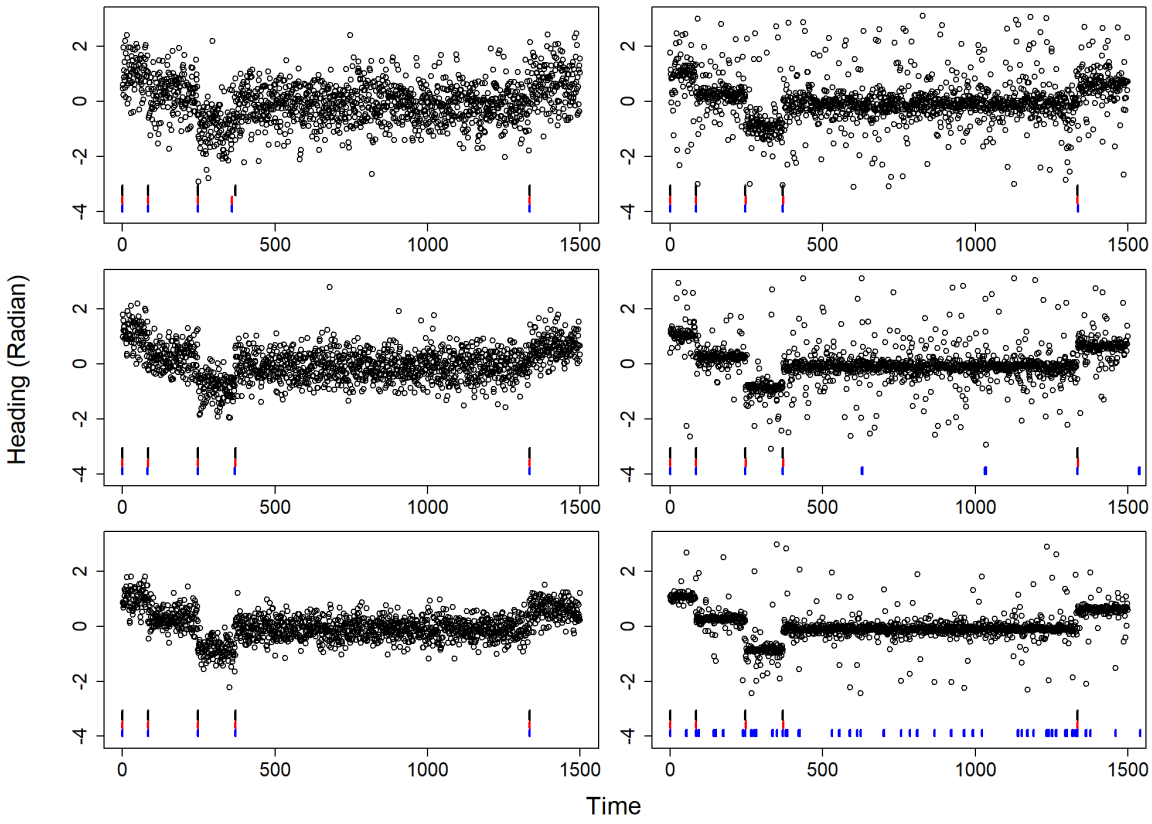


Figure 2.6: A short segment of one of the simulated trajectories under different error structures. **Left:** Errors generated using the von Mises distribution. **Right:** Errors generated using the wrapped Cauchy distribution. Each row represents a different level of concentration. For the von Mises distribution (left), the values of κ used are 3, 5, and 9. For the wrapped Cauchy distribution (right), the corresponding concentration values are computed using $I_1(\kappa)/I_0(\kappa)$. The vertical blue lines indicate the inferred turning points under the assumption of von Mises errors, red lines represent the turning points from the robust estimation method, and black lines indicate the ground truth.

To assess the proposed detection methods, I compare the results of the detections made by the two proposed models with the application of a previous turning point algorithm introduced by Potts et al. (2018). This algorithm uses a sliding-window approach, requiring specification of a window size, which I set to be $w = 40\tau$. This algorithm also post-processes the data to remove turns that are below a threshold angle θ, which I set as $\theta = \pi/6$ radians.

I conducted further investigations to evaluate the computational time required for inference under different scenarios. Specifically, I varied the data length n from 100,000 to 1,000,000 in increments of 100,000 points. For each value of n, a trajectory was simulated and ran the algorithms 10 times, taking the average computational time. I assessed two cases: one in which the error was assumed to follow a von Mises distribution, and another using the robust distribution with different values of the robustness parameter $c \in \{-0.5, 0, 0.5, 0.75\}$.

Our implementation of the algorithms is written in `C++` and interfaced with `R` via the `Rcpp` package (Eddelbuettel and François, 2011). All computations were carried out on a desktop computer equipped with an Intel Core i7-6700U CPU (2.40 GHz, 4 cores) and 32 GB of RAM. I then compared these results to those obtained using the previous turning point detection algorithm. For each scenario, the average runtime is reported over 10 independent runs.

2.1.6 Case Study on Oryx Movement

To demonstrate the effectiveness of the proposed algorithm on empirical data, I used six trajectories from 40Hz magnetometer data of oryx *(Oryx leucoryx)*, previously used and published in Potts et al. (2018). These oryx live in Mahazat as-Sayd, a protected desert steppe area in west-central Saudi Arabia ($28°15'\text{N}, 41°40'\text{E}$). Details of the study site and data-gathering techniques can be found in Potts et al. (2018).

For observed trajectories, analysing the precision and recall is infeasible because we

lack information on the "true" turning points. Instead, I use the method of Potts et al. (2018), whereby the quality of a proposed set of turning points is evaluated by creating a model path consisting of straight-line segments connecting each pair of consecutive inferred turning points, and comparing this with a path constructed directly from the data. The method computes a dimensionless quantity $\bar{\sigma} = \sigma/l$ where σ is a measure of the distance between the two paths and l is the average step length. Then $\bar{\sigma}$ is a measure of the model path's goodness of fit, with lower $\bar{\sigma}$ indicating a better fit (see Potts et al. (2018) for details).

I first conducted a preliminary analysis to understand the characteristics of the noise in the data. To do this, a rough approximation was obtained of the signal using a moving median filter, as described in Section 2.1.4. For this analysis, the window width was set to $w = 400$, which corresponds to using 5 seconds before and after the current time step to estimate the true heading. This window size was chosen based on the belief that an animal of the size of an oryx is unlikely to make multiple behaviourally meaningful turns within a 5-second interval.

With the moving median estimator in place, the residuals were then constructed, which will be used to analyse the structure of the noise. As the data are collected at high frequency, it is expected to exhibit a high degree of autocorrelation, which may lead to overestimating the number of turning points. To address this issue, the BIC value is inflated by a quantity that depends on the first-lag autocorrelation, in order to compensate for the autocorrelation present in the data. The inflation factor used is $\frac{1+\rho}{1-\rho}$, as justified by Lavielle and Moulines (2000); Bardwell et al. (2019). This adjustment compensates for the autocorrelation present in the data. Alternatively, autocorrelated errors are considered in detail in Chapter 3.

2.2 Results

2.2.1 Simulation results

Figure 2.7 illustrates the F1-score (defined in Equation 2.30) as a function of the error concentration parameter κ, comparing the performance of three methods for detecting turning points. All methods show improved performance as κ increases. The von Mises model and the robust model perform particularly well, maintaining high F1-scores across scenarios, especially for $\kappa \geq 3$. The sliding-window approach also shows steady improvement, achieving comparable performance in high-concentration regimes in some cases.

In contrast, Figure 2.8 reveals the impact of outlier contamination on model performance. The robust model demonstrates strong resistance to the presence of outliers, consistently achieving high F1-scores across all values of κ and scenarios, behaving exactly as in the absence of outliers. The sliding-window method shows moderate resilience, with performance improving gradually as κ increases, although it remains slightly below the robust method. Notably, the standard von Mises model suffers substantial degradation in performance under outlier conditions. In particular, its F1-score sharply declines as κ increases. This is because, at higher concentrations, the model becomes narrowly focused around the mean direction, making outliers appear increasingly incompatible with the assumed distribution, thereby highlighting its sensitivity to heavy-tailed noise.

Across all experiments, consistent patterns are observed across tolerance levels. Relaxing the tolerance (moving from top to bottom across rows) results in slight increases in F1-scores for all methods, with the most notable gains observed in the sliding-window approach. However, the relative ranking of methods remains largely unchanged.

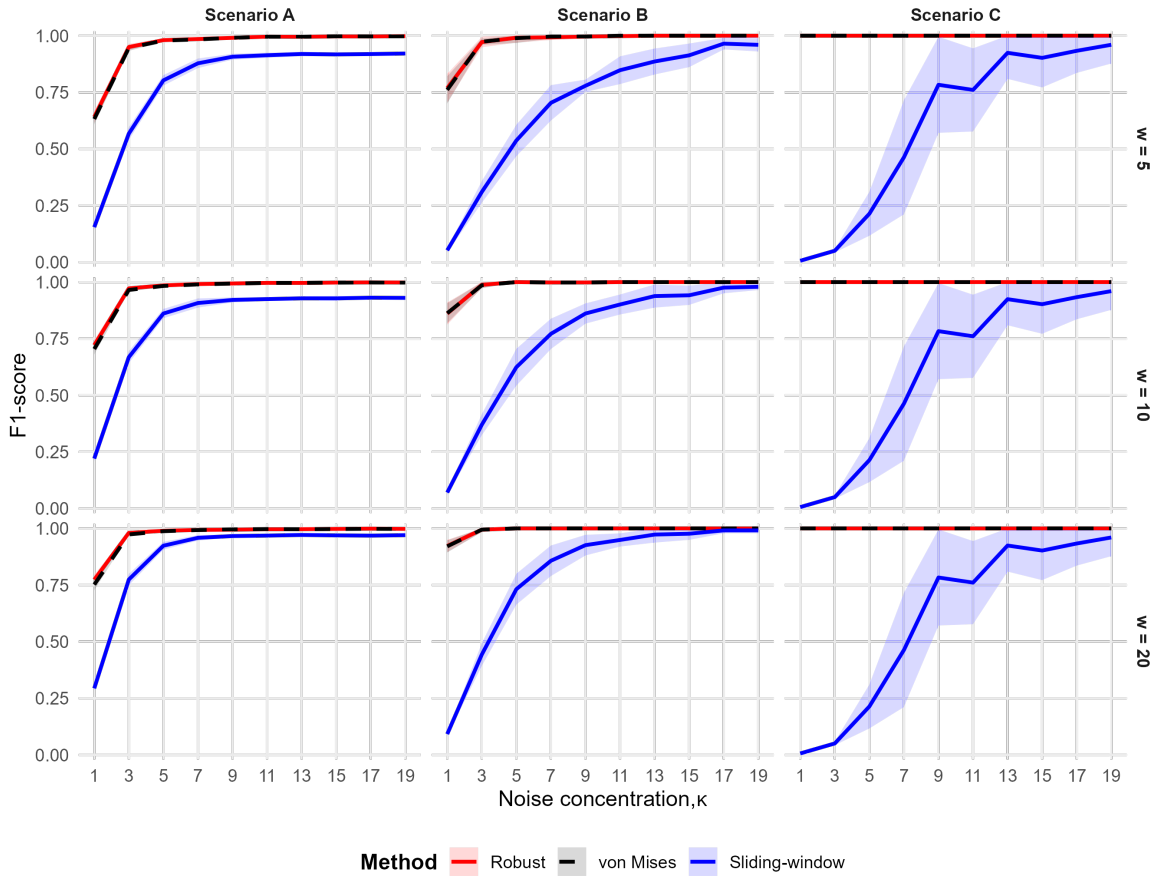


Figure 2.7: F1-score (Equation 2.30) as a function of the concentration parameter κ, comparing the performance of three methods: the model assuming von Mises-distributed errors (black), the robust model (red), and the sliding-window approach (blue). The columns correspond to scenarios A, B, and C (see Section 2.1.5), while the rows represent different tolerance levels. All simulations were conducted in the absence of outliers.

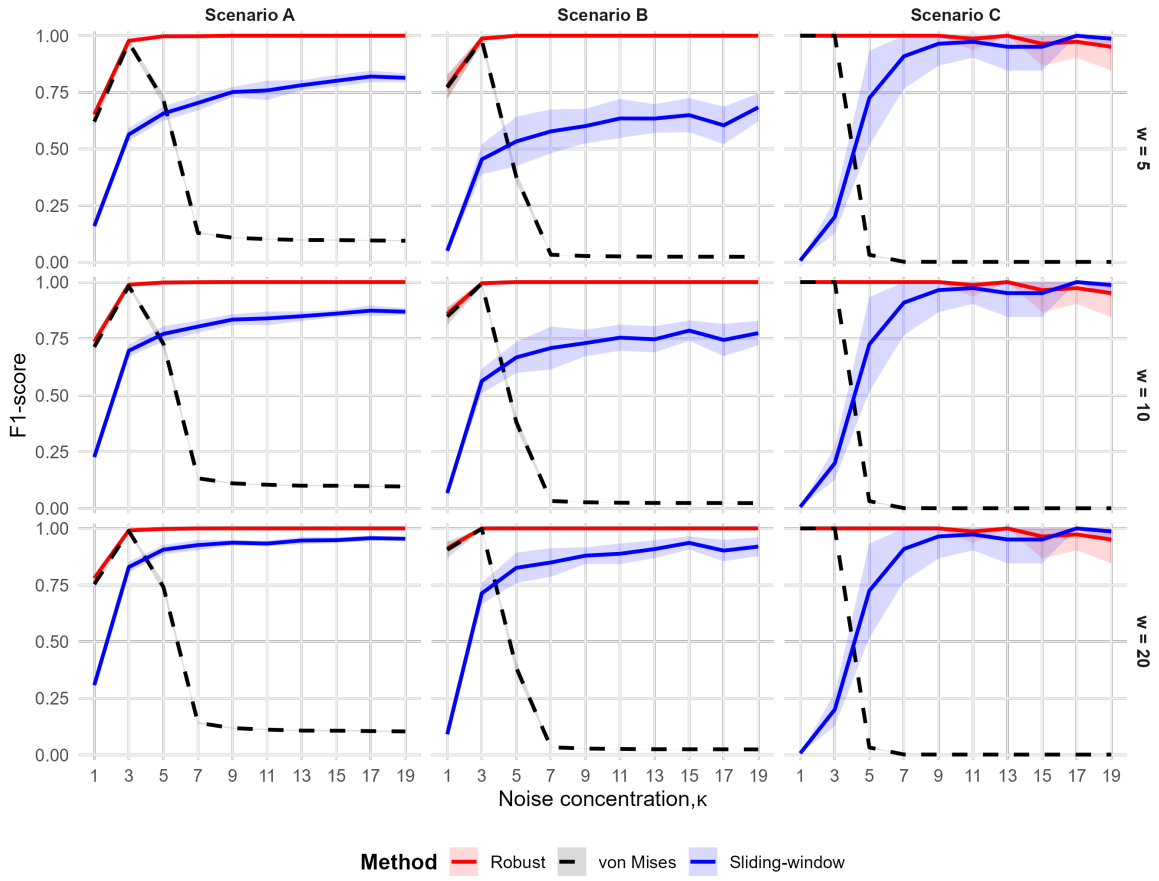


Figure 2.8: F1-score (Equation 2.30) as a function of the concentration parameter κ, comparing the performance of three methods: the model assuming von Mises-distributed errors (black), the robust model (red), and the sliding-window approach (blue). The columns correspond to scenarios A, B, and C (see Section 2.1.5), while the rows represent different tolerance levels. All simulations were conducted under heavy-tailed errors distribution.

2.2.2 Oryx data

Table 2.1 summarizes the parameter estimates obtained for each of the six oryx paths under both the von Mises and robust models. Specifically, it reports the estimated number of turning points (m), the fit measurement (σ), and the concentration parameter (κ) for each model. For the robust model, the robustness parameter c is also reported. The von Mises model often identifies more turning points than the robust model (see Figure 2.9). Figure 2.10 visualises the estimated error models obtained under different modelling settings across all individuals and illustrates systematic differences between individuals. While the estimated values differ from those reported by Potts et al. (2018, Table 1), the overall patterns remain consistent. All three algorithms—von Mises, robust, and sliding window—detect the most visually apparent turning points. However, they differ in their sensitivity to turn magnitude and segment length, which affects whether small directional changes are classified as true turning points or dismissed as noise.

Table 2.1: Estimated number of turning points (m) and distribution parameters for the oryx data under the von Mises and robust models.

ID	**von Mises**			**Robust**			
	m	σ	κ	m	σ	κ	c
1	137	0.0406	7.2627	102	0.0616	63.3604	0.9222
2	127	0.0660	5.1137	72	0.0786	35.3013	0.8827
3	97	0.0829	3.2316	118	0.2063	11.3526	0.6701
4	31	0.0845	4.1046	26	0.0920	5.1145	0.0822
5	70	0.1149	3.2726	57	0.1073	7.3332	0.5016
6	75	0.1063	2.6143	76	0.1567	5.0291	0.3254

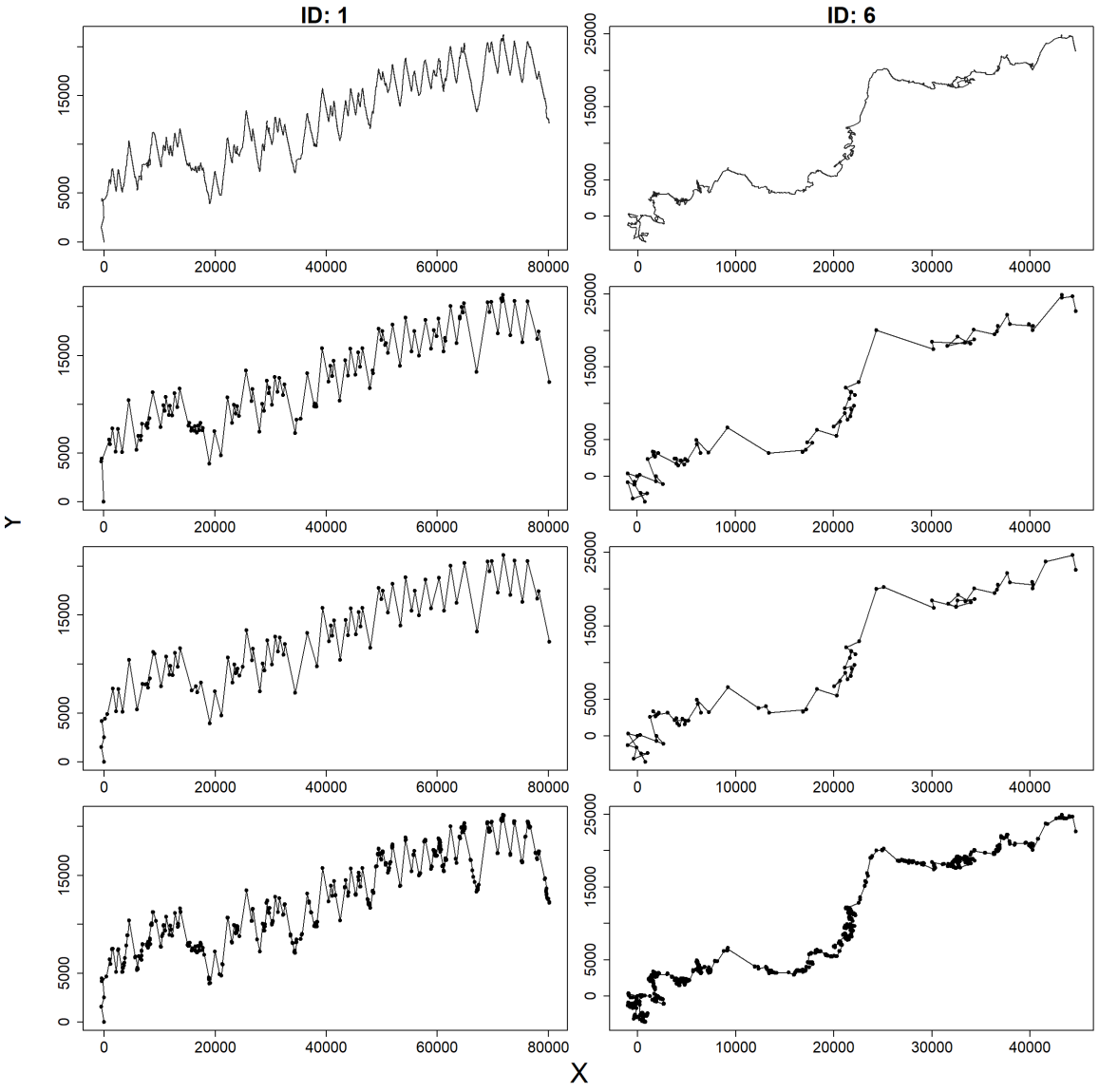


Figure 2.9: Reconstruction of two oryx paths from magnetometer data under the assumption of constant speed. The first row displays the observed paths. The second row shows reconstructions based on turning points inferred using the von Mises model. The third row corresponds to inference under the robust model, and the fourth row shows results from the sliding-window algorithm.

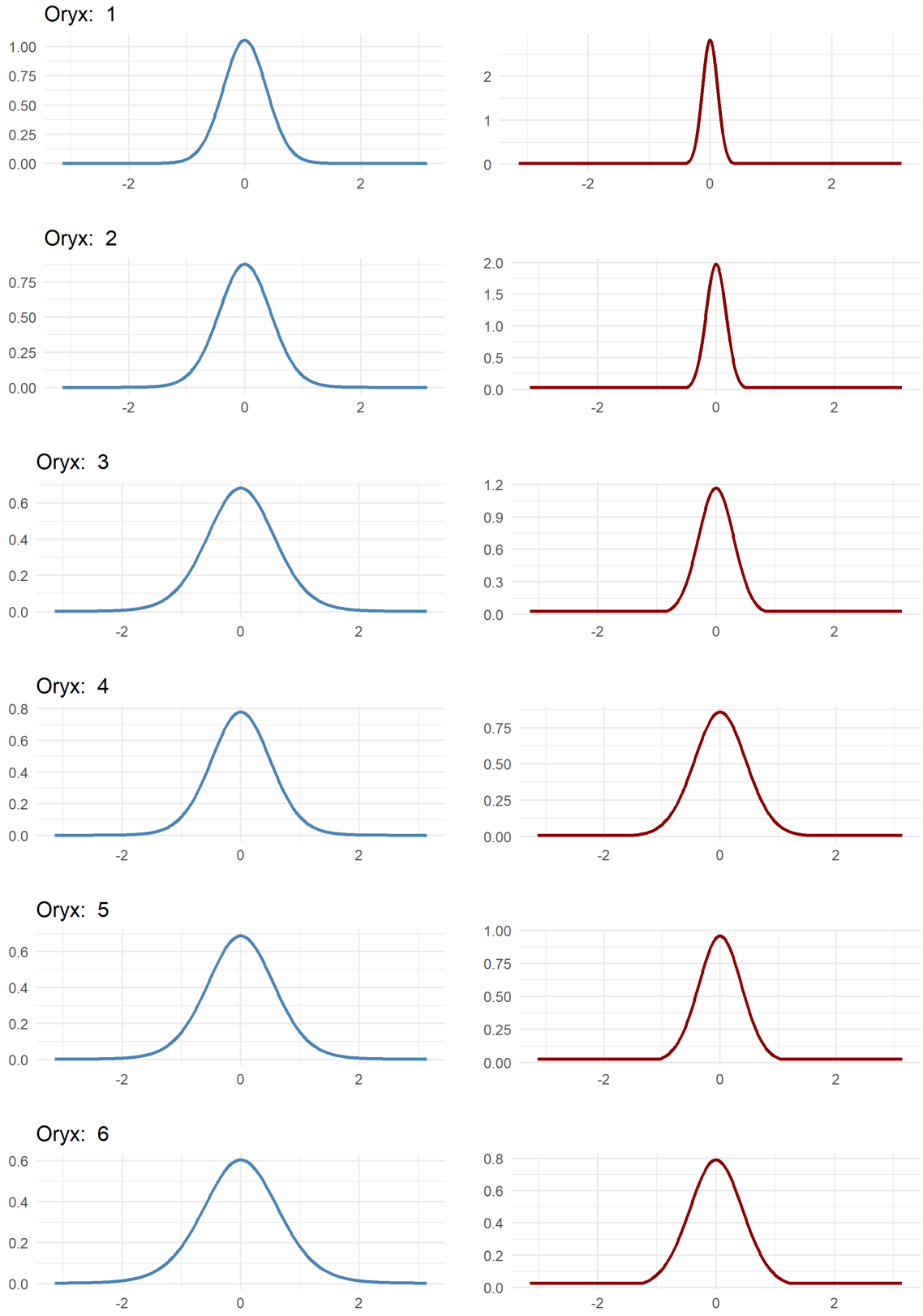


Figure 2.10: Error density functions for all oryx under the two error model choices.

2.2.3 Run Time

Figure 2.11 shows the runtime of the different algorithms across varying data sizes on a log-log scale. Each line represents the average runtime (in seconds), computed over multiple runs for trajectories of lengths ranging from $n = 10^5$ to $n = 10^6$. The von Mises model is the most computationally efficient across all scenarios. The robust models, which incorporate different values of the robustness parameter c, demonstrate increasing runtime as c increases. Notably, when $c = 0.75$, the runtime increases sharply, especially for large datasets. In the oryx data, c is large in most cases (see Table 2.1), which leads to inference times of several hours. This substantial increase in computational cost is primarily due to the added complexity in the shape of the cost function $F(t, \mu)$, which becomes more irregular and difficult to optimize as c approaches 1. Despite this, all methods exhibit approximately linear scalability with data size, highlighting their suitability for large-scale, high-frequency movement data.

2.3 Summary

I have described two model-based algorithms that provide a description of an animal's movement in terms of straight-line segments interspersed with turns. The approach consists of two main stages. In the first stage, estimators are constructed for the magnitude of the noise, given by $1/\kappa$, where κ is the concentration parameter of error. In the presence of outliers I defined an additional parameter c which controls the robustness in a biweight-like model.

Two methods are detailed for this purpose. The first method involves estimating κ and c using the angular successive differences of the observations at the first lag. The second method applies a simple moving median filter to estimate the mean steps, followed by estimating κ and c based on the residuals (Section 2.1.4).

An alternative approach might involve making an initial guess for c and κ, estimating the turning points, iteratively refining the parameters based on the fitted model,

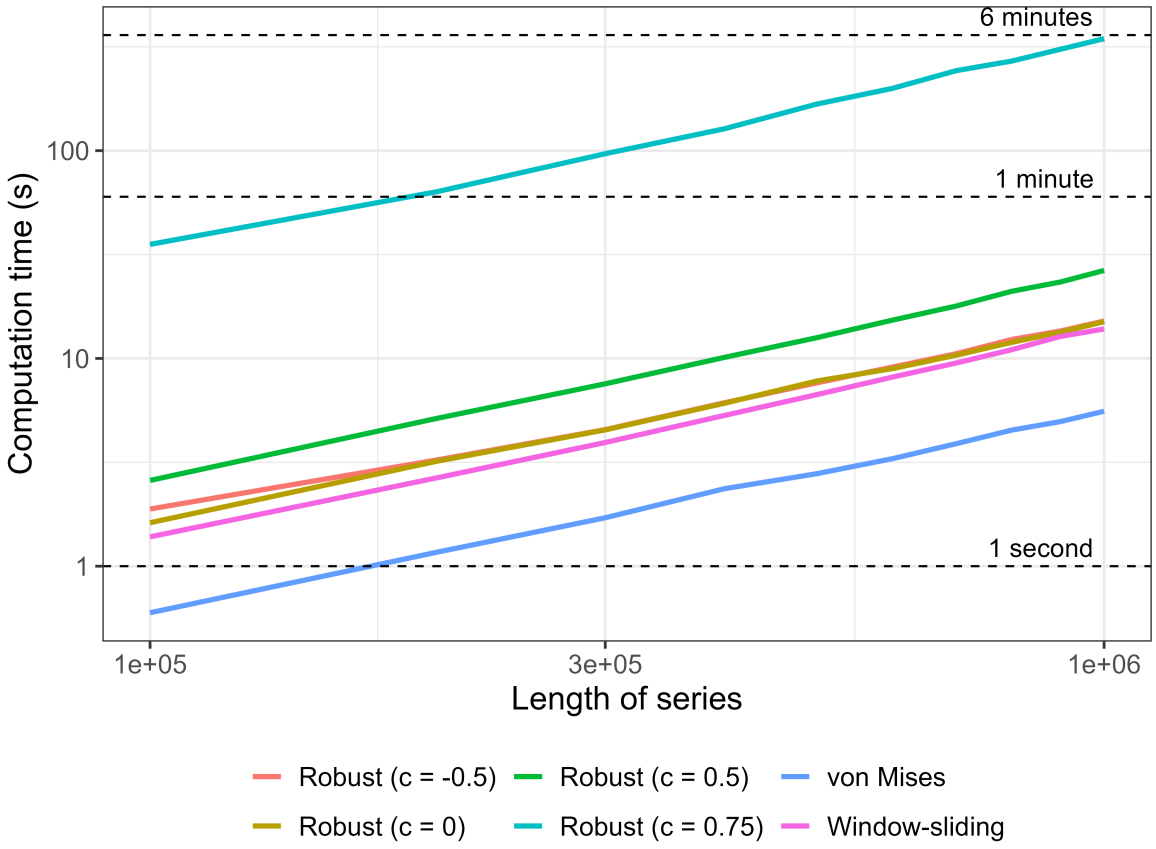


Figure 2.11: Timing comparisons between the three algorithms with different input value of c in case of the robust model.

and then updating them until the results stabilise. However, my preliminary analysis suggests that this method can be sensitive to the initial guess, potentially leading to an incorrect number of estimated turning points. Moreover, misspecification of the underlying model may adversely affect the accuracy of parameter estimates, while also being computationally less efficient.

In the second stage, I used the estimated values of κ and c to identify turning points by maximising a particular penalised likelihood. For this, I used a version of the functional pruning optimal partitioning (FPOP) algorithm, modified for use with circular statistics. I evaluated the proposed methods using both real and simulated data (Section 2.1.3).

Given a path consisting of straight moves and turns, where headings are measured with a standard deviation accuracy of $\pi/9$ radians and with the absence of outliers, the two algorithms successfully identified nearly 100% of the turning points. This level of accuracy compares favourably with the previous method of Potts et al. (2018). In the presence of outliers, modelling the errors as von Mises distribution mostly leads to overfitting the number of turning points.

Our proposed error models initially ignore sequential dependence. However, I compensated for the presence of autocorrelation by inflating the penalty value to account for temporal dependence aligning with the approach of Lavielle and Moulines (2000); Bardwell et al. (2019). In Chapter 3, I further increase model complexity by explicitly incorporating autocorrelation into the noise structure.

My method is designed for data collected using high-frequency magnetometers or modern reversed-GPS tracking technologies capable of sampling at 1 Hz over extended periods (Nathan et al., 2022). Specifically, it requires there to be enough data points between successive turns to be able to identify the turning points robustly. Whilst a detailed analysis was not performed to find the lower limit of number of data-points between turns, my simulated data had consecutive turns interspersed by no fewer than

17 datapoints. My procedure for assessing accuracy of inference relies on a tolerance of ± 10 datapoints, so to be consistent with this, it is suggested that data ought to be sufficiently high frequency that there is highly unlikely to be < 10 measured locations between successive turns. For lower frequency data, on a timescale closer to that of the turns, it is necessary to use an approach that is explicitly formulated in continuous time. Blackwell (2025) describes such a method, which reconstructs the precise times of turns between observations, but it is not computationally competitive with the methods introduced here.

Lastly, while the proposed method examines the data as a stream of headings at each time step, it is implicitly assumed that the animal is moving between each observation, which may not always hold in practice. If that is a concern and velocity information is available (e.g., from accelerometer or locational data), incorporating such data into the analysis or post-processing our inferences could ensure that the detected turns align with times when this assumption is satisfied (Gunner et al., 2024).

Chapter 3

Modelling Temporal Dependence in Turn Detection

In this chapter, I extend the methodology introduced in Chapter 2 to account for the autocorrelation between consecutive observations. The observed headings, typically obtained from magnetometer sensors, are treated as noisy observations of the true direction, modelled in Chapter 2 as a piecewise constant function with independent and identically distributed noise. Here, I refine that model by incorporating first-order autocorrelation to better capture temporal dependence in the data.

From a biological and observational perspective, the assumption of data being independent is often unrealistic. However, the deviations observed in the heading data should not be interpreted purely as measurement error. Animal movement is inherently smooth due to biomechanical constraints such as inertia, gait, and turning limitations, meaning that successive headings tend to be similar over time. This introduces temporal dependence that arises from the movement process itself rather than solely from observational noise.

In addition, magnetometer measurements may also exhibit temporal dependence due to internal filtering and environmental interference. As a result, the observed headings

reflect a combination of measurement error and intrinsic movement persistence. This motivates extending the model to explicitly incorporate temporal dependence in the observed headings (Sedlak, 1997).

I first evaluate the performance of the introduced method on simulated trajectories consisting of straight-line movements with autocorrelated noise, interspersed with sharp directional changes. I then apply the proposed algorithm to the same magnetometer data as in Chapter 2. Finally, I compare the performance of the proposed model with the model in the previous chapter and the method of Potts et al. (2018). This work has been published in Alharbi et al. (2026).

3.1 Method

3.1.1 Model Formulation

In Chapter 2 I developed a method for inferring turning points from a high-frequency time series of measured animal headings (e.g. from a magnetometer). I modelled the animal's true headings as an angular piecewise constant function of time with independent and identically distributed errors. However, errors in time series data are often autocorrelated, violating the i.i.d. assumption. Here, to address this, I reformulate the expected value of h_t as being dependent on the both the actual heading and the previous observation, as follows

$$E(h_t|\mu_t, h_{t-1}, \rho) = \tan_2^{-1}\left((1-\rho)\sin(\mu_t) + \rho\sin(h_{t-1}), (1-\rho)\cos(\mu_t) + \rho\cos(h_{t-1})\right), \tag{3.1}$$

where $\rho \in (-1, 1)$ is a correlation parameter that measures the influence of the observed heading at time $t-1$ on the expectation of the heading at time t. Notice that, in the absence of autocorrelation in error, i.e. $\rho = 0$, the expectation is simply $E(h_t|\mu_t, h_{t-1}, 0) = \mu_t$.

The formulation in Equation (3.1) can be interpreted as a circular analogue of a

first-order autoregressive (AR(1)) process, where the expected heading at time t is a weighted combination of the underlying behavioural direction μ_t and the previous observed heading h_{t-1}.

Importantly, this formulation introduces dependence at the level of the movement process itself. The contribution of h_{t-1} reflects persistence in the animal's movement, arising from biomechanical and behavioural constraints, rather than representing purely autocorrelated measurement error. In this sense, the model captures both intrinsic movement dynamics and observational noise.

In this angular AR(1) framework, the animal is assumed to move along a main direction between turns while exhibiting temporally dependent deviations around this direction. These deviations arise from a combination of intrinsic movement variability and observational noise. If these fluctuations are assumed to follow i.i.d. von Mises random variables with zero mean and a known concentration parameter κ, then the log-likelihood can be derived from the von Mises density function as

$$\mathcal{L}(\{\mu_t\}_{t=2}^{n} \mid \{h_t\}_{t=1}^{n-1}, \rho, \kappa) = \kappa \sum_{t=2}^{n} \cos(h_t - E(h_t|\mu_t, h_{t-1}, \rho)). \tag{3.2}$$

Here, the likelihood is constructed by modelling the conditional deviations of the observed headings from their expected values. These deviations capture both residual variability in the movement process and measurement noise. I condition on the first observation to avoid specifying an initial distribution; this reduces the effective sample size by one and is negligible as large n is expected, so it does not affect turn detection. Figure 3.1 gives example simulations of the model in a situation where there is a single turning point and compares that to the model assuming i.i.d. noise.

Using the penalty term ($\beta < 0$) to control the trade-off between the best fit model and reliable turning point estimation, therefore, the resulting optimisation problem is

$$\hat{\mu}_2, \ldots, \hat{\mu}_n = \underset{\mu_2,\ldots,\mu_n}{\operatorname{argmax}} \left(\mathcal{L}(\{\mu_t\}_{t=2}^{n} | \{h_t\}_{t=1}^{n}, \rho, \kappa) + m\beta\right). \tag{3.3}$$

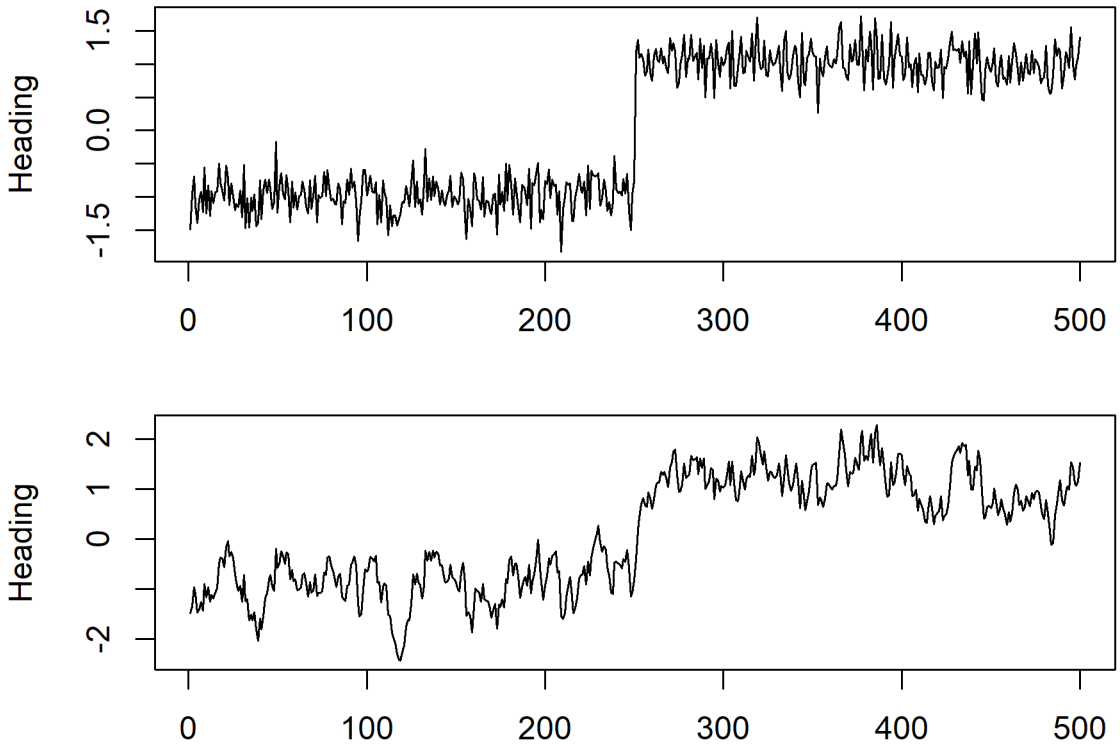


Figure 3.1: Top: simulated heading series with a single turning point assuming independent von Mises errors ($\rho = 0$) with $\kappa = 15$. Bottom: simulated series with a single turning point assuming an angular autoregressive process around the mean direction with $\kappa = 15$ and $\rho = 0.80$.

In Chapter 2, I showed that the optimal penalised log-likelihood of the parameters, given the data up to time s and the current direction μ, is expressed as

$$F(s,\mu) = \max\left\{F(s-1,\mu), F(s-1) + \frac{\beta}{\kappa}\right\} + \cos(h_s - E(h_s|\mu_t, h_{s-1}, \rho)). \tag{3.4}$$

Note that the function $F(s,\mu)$ cannot be expressed in closed form unless $\rho = 0$ (as in Chapter 2). This presents a challenge for directly applying the methods introduced in Chapter 2. However, given that the space of all possible values of μ is bounded within $[-\pi, \pi)$, we can still approximately solve the optimization problem by discretising the domain of μ into a finite set with reasonable resolution.

Discretising the space of μ involves dividing the continuous range $(-\pi, \pi]$ into a finite set of discrete values, making the problem computationally tractable. Once we have discretised the space of possible μ, we perform a forward iteration through the data to calculate $F(s,\mu)$ for each value of μ and s, while keeping track of a vector indicating the most recent turning points and a vector showing the origin of each point in the space of μ.

3.1.2 Estimating β, ρ, and κ

In this section, I build upon the foundations laid in Chapter 2 to present robust and data-driven approaches for estimating β, ρ, and κ. For the penalty parameter β, I set $\beta = -\log(n)$ when using the angular AR(1) model based on the BIC, as before. Alternatively, for the model assuming i.i.d. observations, an inflation factor is required in the estimation of β, as discussed in Section 2.1.5.

I then aim to estimate ρ and κ robustly in the absence of information about the heading steps. The autocorrelation parameter ρ is estimated as

$$\hat{\rho} = \frac{\text{Median}_{t>2}\,(1 - \cos(h_{t+2} - h_t))}{\text{Median}_{t\geq 2}\,(1 - \cos(h_{t+1} - h_t))} - 1. \tag{3.5}$$

The estimator in Equation (3.5) can be viewed as a circular analogue of the approach proposed by Chakar et al. (2017) for real-valued data x_t, which compares the medians of lag 1 and lag 2 differences to get a robust estimator of autocorrelation:

$$\hat{\rho} = \frac{\left(\text{Median}_{t>2}\ |x_{t+2} - x_t|\right)^2}{\left(\text{Median}_{t\geq 2}\ |x_{t+1} - x_t|\right)^2} - 1. \tag{3.6}$$

In practice, we avoid the case of negative autocorrelation due to the nature of bearings by setting $\rho = \max\{0, \hat{\rho}\}$. Additionally, there is a non-zero probability that the autocorrelation estimate exceeds one, resulting in an invalid measure of autocorrelation. This risk increases when the true autocorrelation is close to one. I address this issue by imposing an upper bound of 1 on ρ.

To clarify notation, I define κ_ρ as the concentration parameter conditional on ρ, corresponding to κ in Equation 3.2, and κ_μ as the marginal concentration parameter around the main headings μ, corresponding to κ in Section 2.1.1. Based on the estimate of ρ , I modify the estimate of κ in Section 2.1.1 to estimate κ_ρ and κ_μ as

$$\hat{\kappa}_\rho = \frac{v}{(1+\hat{\rho})(1-\tilde{R})}, \tag{3.7}$$

$$\hat{\kappa}_\mu = \frac{1-\hat{\rho}}{1-\tilde{R}}\, v, \tag{3.8}$$

where $v \approx 0.4592$ and $\tilde{R}$ is the median resultant length defined in Section 2.1.4. The factors $(1-\rho)$ and $(1+\rho)$ arise from the variance inflation/deflation induced by the AR(1)-type dependence, analogous to how autocorrelation rescales marginal and differenced variances in linear processes.

Alternatively, a non-parametric approach can be used, which consists of obtaining a rough approximation of the noise using an angular moving median filter (see Section 2.1.4), and subsequently estimating κ_ρ, κ_μ, and ρ from the estimated noise $\hat{\epsilon}_t$.

In this case, we estimate ρ as follows:

$$\bar{\rho} = \frac{\sum_{t=2}^{n} \sin(\hat{\epsilon}_t)\sin(\hat{\epsilon}_{t-1})}{\sqrt{\left(\sum_{t=2}^{n} \sin^2(\hat{\epsilon}_t)\right)\left(\sum_{t=1}^{n-1} \sin^2(\hat{\epsilon}_t)\right)}}. \tag{3.9}$$

This measure of association between two circular variables was proposed by Jammalamadaka and Sarma (1988). Given ρ, we estimate κ_ρ using the maximum likelihood estimate from the whitened series $\bar{\hat{\epsilon}}_t$:

$$\bar{\hat{\epsilon}}_t = \hat{\epsilon}_t - \tan_2^{-1}\big(\rho \sin(\hat{\epsilon}_{t-1}),\, 1 - \rho(1 - \cos(\hat{\epsilon}_{t-1}))\big) \pmod{2\pi}. \tag{3.10}$$

Finally, κ_μ is estimated as described in Section 2.1.4.

3.1.3 Simulation study

To evaluate the effectiveness of the introduced methods, I revisited the simulated trajectories presented in Chapter 2. These scenarios, denoted A,B and C, vary in the number of turns and are visualized in Figure 3.2. First, I assessed how the proposed methods in Section 3.1.2 performed in estimating the parameters ρ, κ_ρ. I then evaluated the model's ability to identify turning points using the F1-score. In this evaluation, I adopted a tolerance-based acceptance rule for detected turns, using a grid of tolerance values: 5, 10, and 20 time units. This evaluation was carried out by introducing different error structures into the true path. Finally, I explored various model mis-specification scenarios to assess the robustness of turning point detection.

For the first task, I ran simulations for the angular AR(1) model for various values of ρ and κ_ρ. I started by setting $\rho \in \{0.03, 0.08, ..., 0.98\}$ and $\kappa_\rho \in \{5, 15, 45, 100, 200, 400\}$. Each simulation was repeated 100 times, estimating ρ using Equation (3.5). Then, focusing on κ_ρ, I set $\kappa_\rho \in \{5, 10, \ldots, 400\}$ and $\rho \in \{0, 0.25, 0.50, 0.75, 0.85, 0.95\}$. Again, each combination is repeated 100 times, estimating κ_ρ using Equation (3.7). I compared each estimate to the one obtained using the maximum likelihood method. The maximum likelihood estimate is given by $\kappa_\rho = 2\kappa_d$, where κ_d is the estimated concentration parameter of the differenced data. Accordingly, $\kappa_\mu = 2\kappa_d(1 - \rho)$.

To assess the accuracy of the proposed method's estimation of the turning points, I used the F1-score that is described in Section 2.1.5. To assess the proposed detection methods, I compare the results of the detections made by the two proposed models

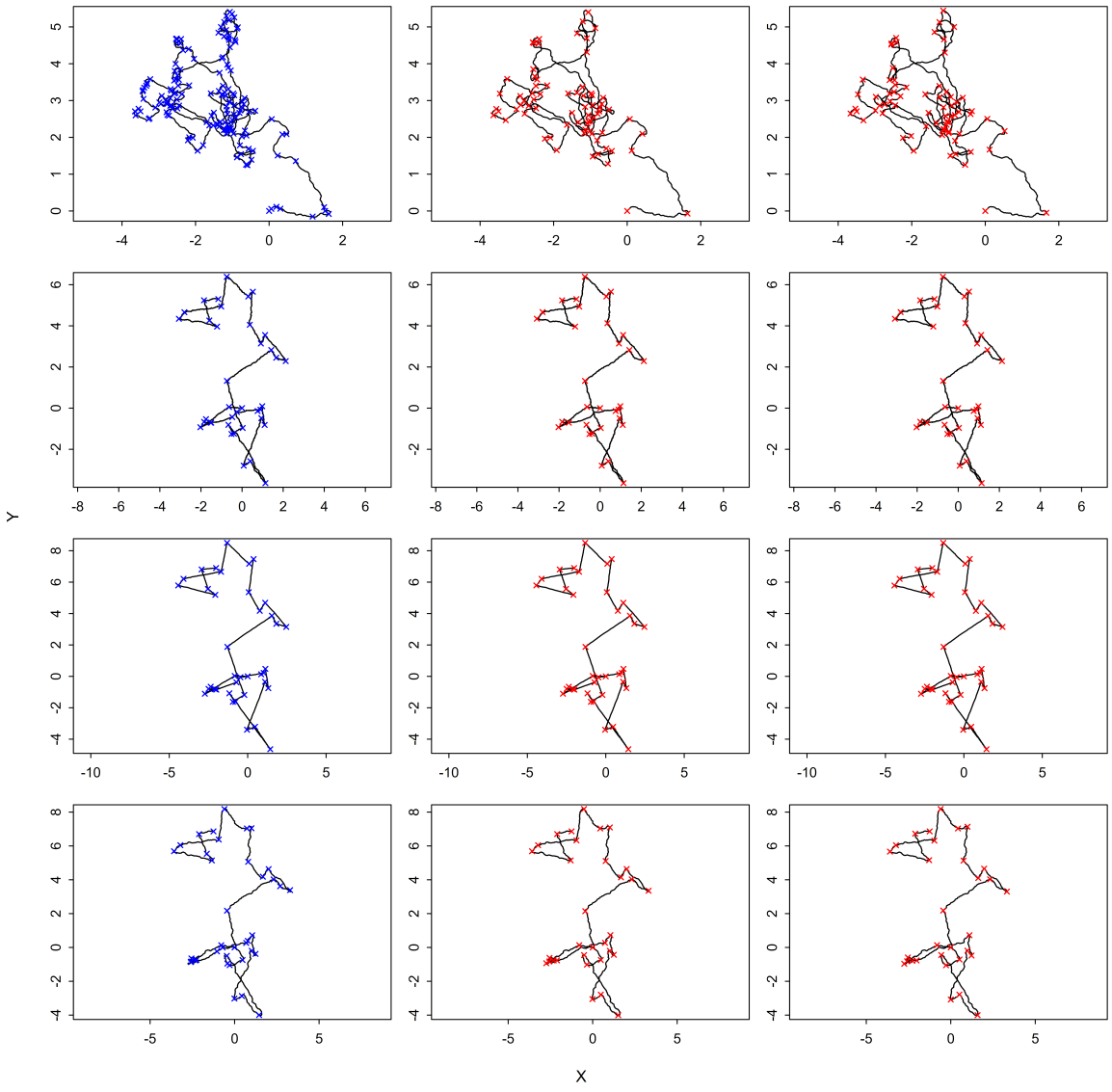


Figure 3.2: Simulated paths of headings under different scenarios. The first row shows the simulated path for scenario (A), with $\kappa_\rho = 200$ and $\rho = 0.98$. The second, third, and fourth rows correspond to scenario (B), with $\kappa_\rho = 5$, 40, and 200, and $\rho = 0.75$, 0.75, and 0.98, respectively. Blue crosses in the first column indicate the ground-truth locations of turning points, while red crosses in the second column show the inferred turning points using the angular AR(1) model. The third column displays the turning points inferred using the IID model.

with the application of a previous turning point algorithm introduced by Potts et al. (2018).

In addition to testing the inference procedure on data generated by the underlying model, I also examined its robustness to model mis-specification. First, I addressed mis-specification of the error distribution by considering errors from a wrapped normal distribution ($\epsilon_t \sim WN(\mu_t, \gamma)$). The wrapped normal density is given by

$$f_{\text{WN}}(x \mid \mu, \gamma) \;=\; \frac{1}{2\pi}\left[1 + 2\sum_{k=1}^{\infty} \gamma^{k^2} \cos\big(k(x-\mu)\big)\right], \quad x \in [0, 2\pi),$$

where μ is the mean direction and $\gamma = e^{-\sigma^2/2}$ is the concentration parameter (Kurz et al., 2014). The values $\gamma \in \{0.893, 0.966, 0.989, 0.995, 0.997, 0.999\}$ were chosen to closely mimic the concentration parameters used previously when it is assumed that the errors have a von Mises distribution with $\kappa \in \{5, 15, 45, 45, 100, 200, 400\}$. As before, the autocorrelation parameter $\rho \in \{0.03, 0.08, ..., 0.98\}$.

Next, I explored misspecification in the underlying dependence structure. While the proposed model assumes a first-order angular autoregressive dependence, real animal movement may exhibit memory extending beyond a single lag. In practice, movement decisions may be influenced not only by the most recent heading but also by short-term behavioural persistence over several previous steps.

To investigate the robustness of the proposed method under such misspecification, I consider higher-order autoregressive processes. As a first example, I examine an angular AR(3) process, as outlined in Equation (3.11), with coefficients $\phi_{1,2,3} = [0.5, 0.3, 0.2]$, allowing the expected heading to depend on a weighted combination of the three most recent observations.

$$\begin{aligned} E_1(h_t|\mu_t, h_{(t-1):(t-4)}, \rho) = \tan_2^{-1}\Bigg(& (1-\rho)\sin(\mu_t) + \rho \sum_{i=t-4}^{t-1} \phi_{t-i} \sin(h_i), \\ & (1-\rho)\cos(\mu_t) + \rho \sum_{i=t-4}^{t-1} \phi_{t-i} \cos(h_i) \Bigg). \end{aligned} \tag{3.11}$$

I also consider a longer-memory process, namely an angular AR(10), where $\phi_{1:10} = 0.1$, as described in Equation (3.12). This specification represents a more persistent dependence structure in which past headings contribute more evenly over a longer time window. Such higher-order models provide a useful way to assess whether the proposed AR(1)-based segmentation method remains effective when the true underlying process exhibits more complex temporal dependence.

$$E_2(h_t|\mu_t, h_{(t-1):(t-11)}, \rho) = \tan_2^{-1}\left((1-\rho)\sin(\mu_t) + \frac{\rho}{10}\sum_{i=t-11}^{t-1}\sin(h_i),\right.$$
$$\left.(1-\rho)\cos(\mu_t) + \frac{\rho}{10}\sum_{i=t-11}^{t-1}\cos(h_i)\right). \tag{3.12}$$

These models act as controlled departures from the angular AR(1) assumption, allowing systematic evaluation of sensitivity to the order of temporal dependence. These formulations allow us to capture the effects of misspecification in both autoregressive and moving average structures, to investigate whether the path demarcation procedure remains robust under various conditions.

Lastly, given the assumption of a constant concentration parameter across the entire trajectory, I investigated the performance of the proposed methods when this condition is not met. Specifically, I generated errors from different concentration parameters, $\kappa_\rho \in \{5, 15, 45\}$, where the first 100 observations were generated using $\kappa_\rho = 5$, followed by 100 observations from $\kappa_\rho = 15$, and then 100 observations from $\kappa_\rho = 45$. This pattern was repeated for the entire length of my artificial trajectories.

3.2 Results

3.2.1 Simulation study

Figure 3.3 compares the performance of the estimation of the autocorrelation parameter, ρ, under varying turning rates. For each value of κ_ρ (i.e. each row in Figure

3.3), the estimator performs similarly across the three different cases (i.e. the three columns). However, there is a notable bias as ρ increases in situations where κ_ρ is low, with the estimates tending to overestimate the true ρ for values of $\rho \gtrsim 0.5$. In other words, when the noise is large, this can exaggerate some of the autocorrelation. If the noise is reduced, i.e. κ_ρ is increased, this bias attenuates. This bias gradually diminishes and the estimates recover accuracy as ρ approaches a higher value, $\rho \gtrsim 0.9$.

Figure 3.4 presents simulation results comparing the robust estimator of the conditional concentration parameter, κ_ρ given in Equation (3.7), with the maximum likelihood estimator. In general, the robust estimator of κ_ρ is accurate for high input values of κ_ρ (low noise), but can overestimate κ_ρ by up to 8% for small values of κ_ρ (higher noise). This trend is consistent across the three scenarios and ρ-values. In contrast, the performance of the maximum likelihood estimator is distinctly worse as the number of turns increases and ρ decreases. The unconditional concentration parameter κ_μ depends on both ρ and κ_ρ through the relationship $\kappa_\mu = (1 - \rho^2)\kappa_\rho$.

Case 1: Errors generated from angular AR(1) process

Figures 3.5, 3.6 and 3.7 illustrate how the performance of the proposed segmentation methods change as I vary the number of turning points (m), the concentration parameter (κ_ρ), and the autocorrelation parameter (ρ). Performance is evaluated using the F1-score, and compared against the sliding-window approach of Potts et al. (2018). Several general patterns emerge across all scenarios.

First, the two proposed models angular AR(1) and IID with the adjusted penalty show almost identical performance when the noise autocorrelation is weak. For $\rho \leq 0.5$, both approaches achieve near-perfect F1-scores, indicating that segmentation accuracy is largely unaffected by mild autocorrelation.

Second, the F1-score declines as ρ increases, reflecting the increased difficulty of detecting turning points under stronger temporal dependence. This decline is not uniform across methods. The angular AR(1) model shows a slower deterioration than the IID

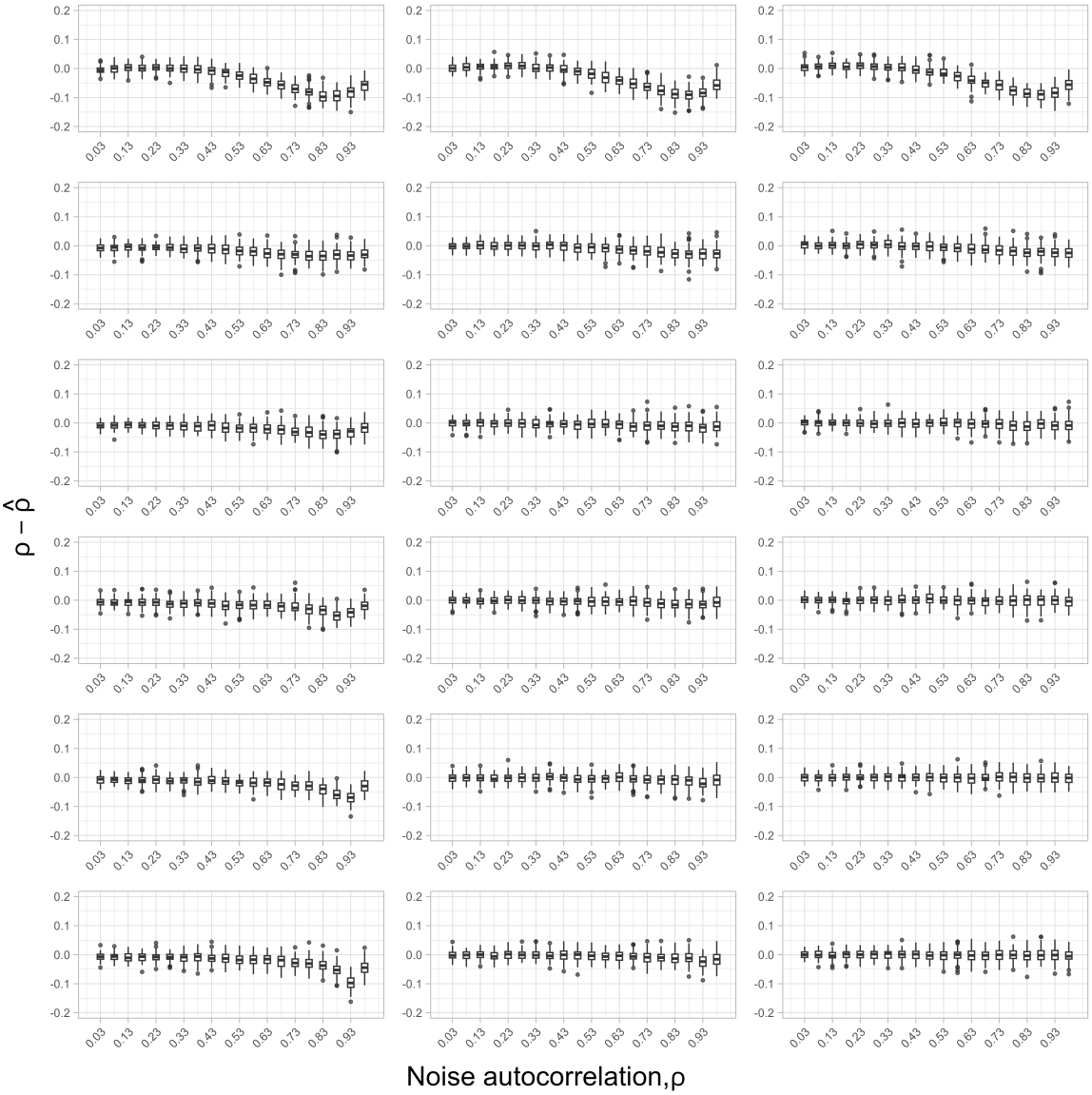


Figure 3.3: Error in estimation of ρ under the different scenarios from Section 3.1.3. Columns represent scenarios A, B, and C from left to right. Rows represent different κ_ρ values, with $\kappa_\rho = 5$, 15, 45, 100 , 200 and 400 from top to bottom.

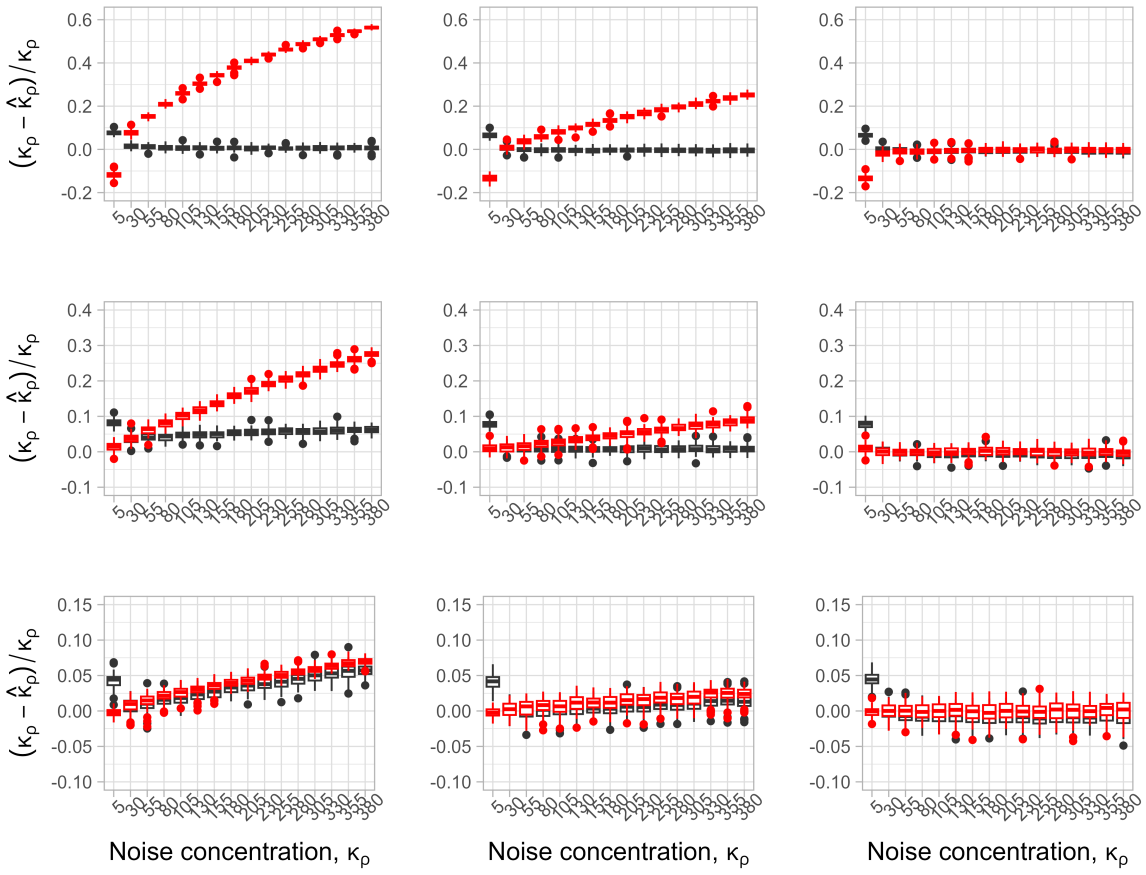


Figure 3.4: Relative error in estimation of κ_ρ under different scenarios. The columns represent A, B, and C from left to right (see Section 3.1.3). Rows represent different ρ values, with $\rho = 0$, $\rho = 0.75$, and $\rho = 0.95$ from top to bottom. The black boxplots represent the errors made by the robust estimator, while the red boxplots represent the errors from the maximum likelihood estimator.

model, particularly when ρ is large, demonstrating that the autoregressive structure provides some robustness under stronger correlation.

Third, the concentration parameter κ_ρ plays a major role in how sensitive the segmentation performance is to autocorrelation. For larger κ_ρ values (e.g., 100–400), both AR(1) and IID retain strong accuracy over a wider range of ρ. In contrast, for smaller κ_ρ (e.g., 5 or 15), the performance drops more noticeably as ρ increases. This indicates that the difficulty of the task at high autocorrelation depends jointly on ρ and κ_ρ, with weaker concentration amplifying the effect of autocorrelation.

Finally, the sliding-window method exhibits substantially lower accuracy in most cases, particularly when autocorrelation increases or when m is small. Even when tolerance is relaxed, its F1-scores remain consistently below those of the model-based approaches.

To examine how temporal tolerance influences the results, I recomputed the F1-score under three thresholds: 5, 10, and 20 time steps. Relaxing the tolerance slightly increases the F1-scores for all approaches, with the largest improvements observed for the sliding-window method. Nonetheless, the relative ordering of methods remains unchanged: AR(1) performs best, IID follows closely, and the sliding-window approach remains substantially less accurate. Based on these patterns, a moderate tolerance of 10 time steps offers a practical balance between strictness and stability, and I therefore adopt this threshold for the subsequent analyses.

Case 2: Errors generated from wrapped normal distribution

In contrast to the model's assumption of errors generated from a von Mises distribution, Figure 3.8 shows the simulation results of errors generated from a wrapped normal distribution with different levels of the variance and the turning points density. The results appear very similar to those in Figures 3.5,3.6 and 3.7, indicating robustness against misspecification of the underlying error-generating distribution.

Case 3: Angular AR(3) and angular AR(10) processes

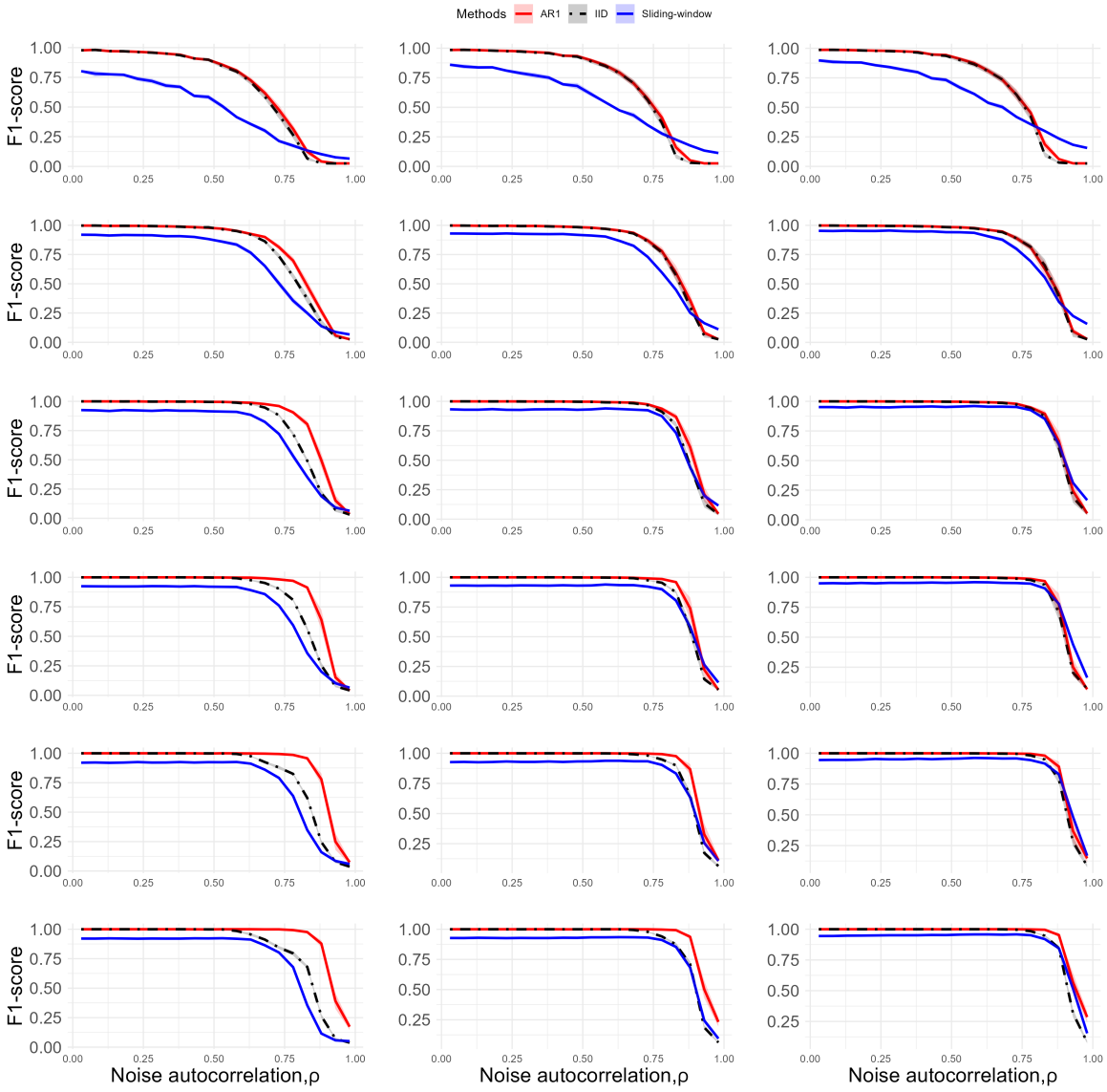


Figure 3.5: F1-score (Equation 2.30) as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and the sliding-window approach (blue). The results correspond to scenario A (see Section 3.1.3). Each column presents the tolerance levels 5, 10, and 20, while each row corresponds to one value of κ_ρ: 5, 15, 45, 100, 200, and 400, respectively.

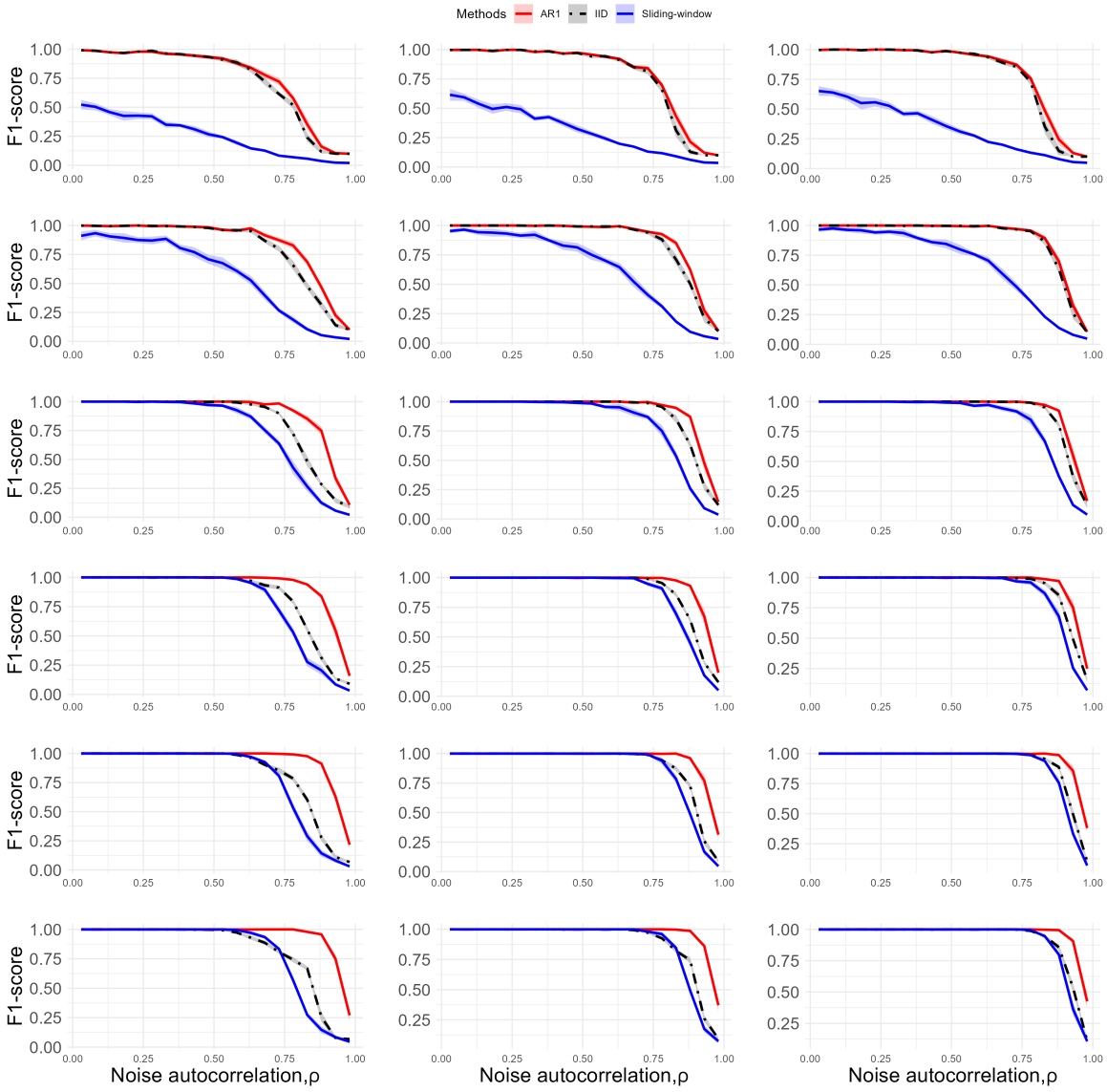


Figure 3.6: F1-score (Equation 2.30) as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and the sliding-window approach (blue). The results correspond to scenario B (see Section 3.1.3). Each column presents the tolerance levels 5, 10, and 20, while each row corresponds to one value of κ_ρ: 5, 15, 45, 100, 200, and 400, respectively.

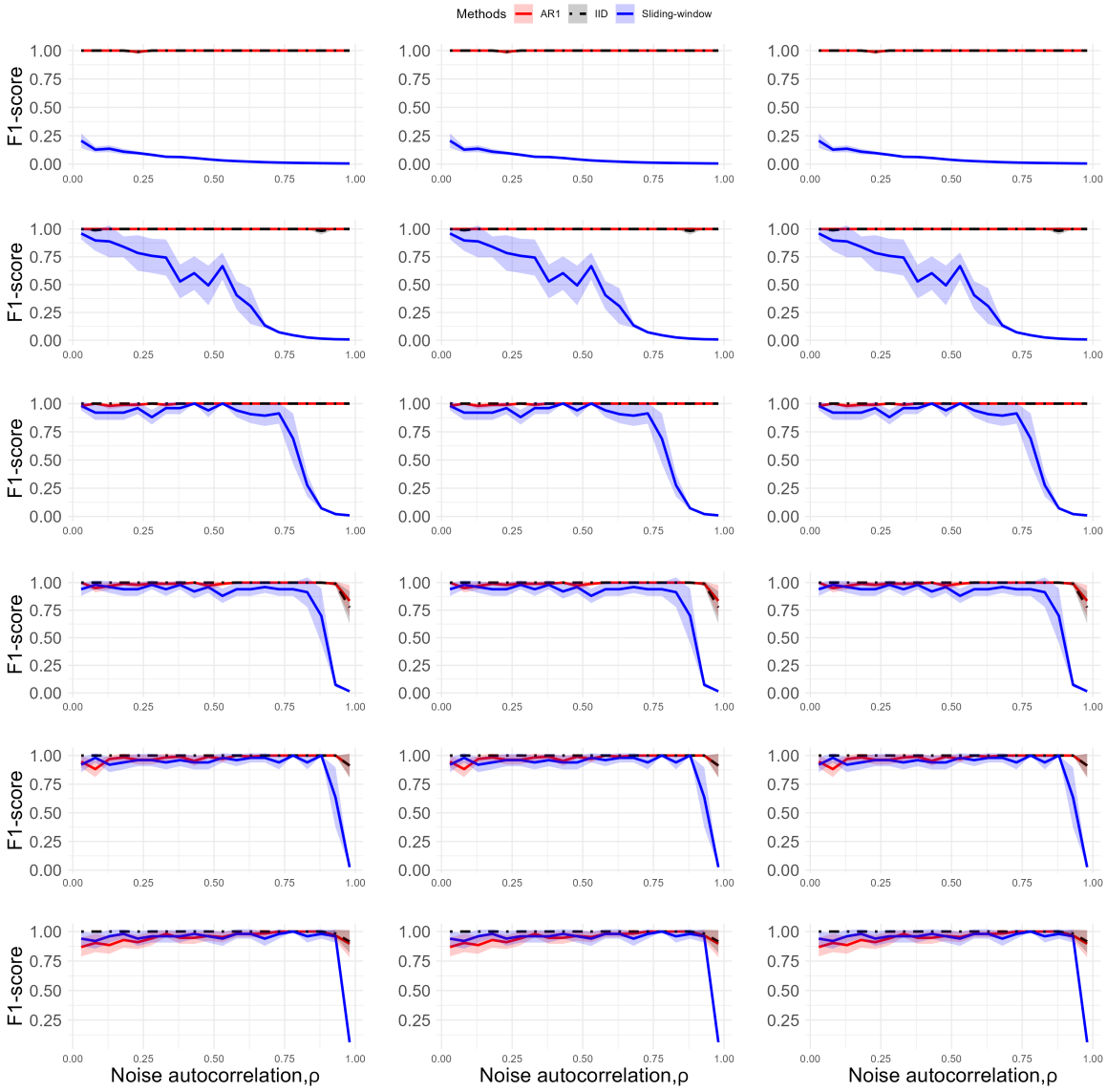


Figure 3.7: F1-score (Equation 2.30) as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and the sliding-window approach (blue). The results correspond to scenario C (see Section 3.1.3). Each column presents the tolerance levels 5, 10, and 20, while each row corresponds to one value of κ_ρ: 5, 15, 45, 100, 200, and 400, respectively.

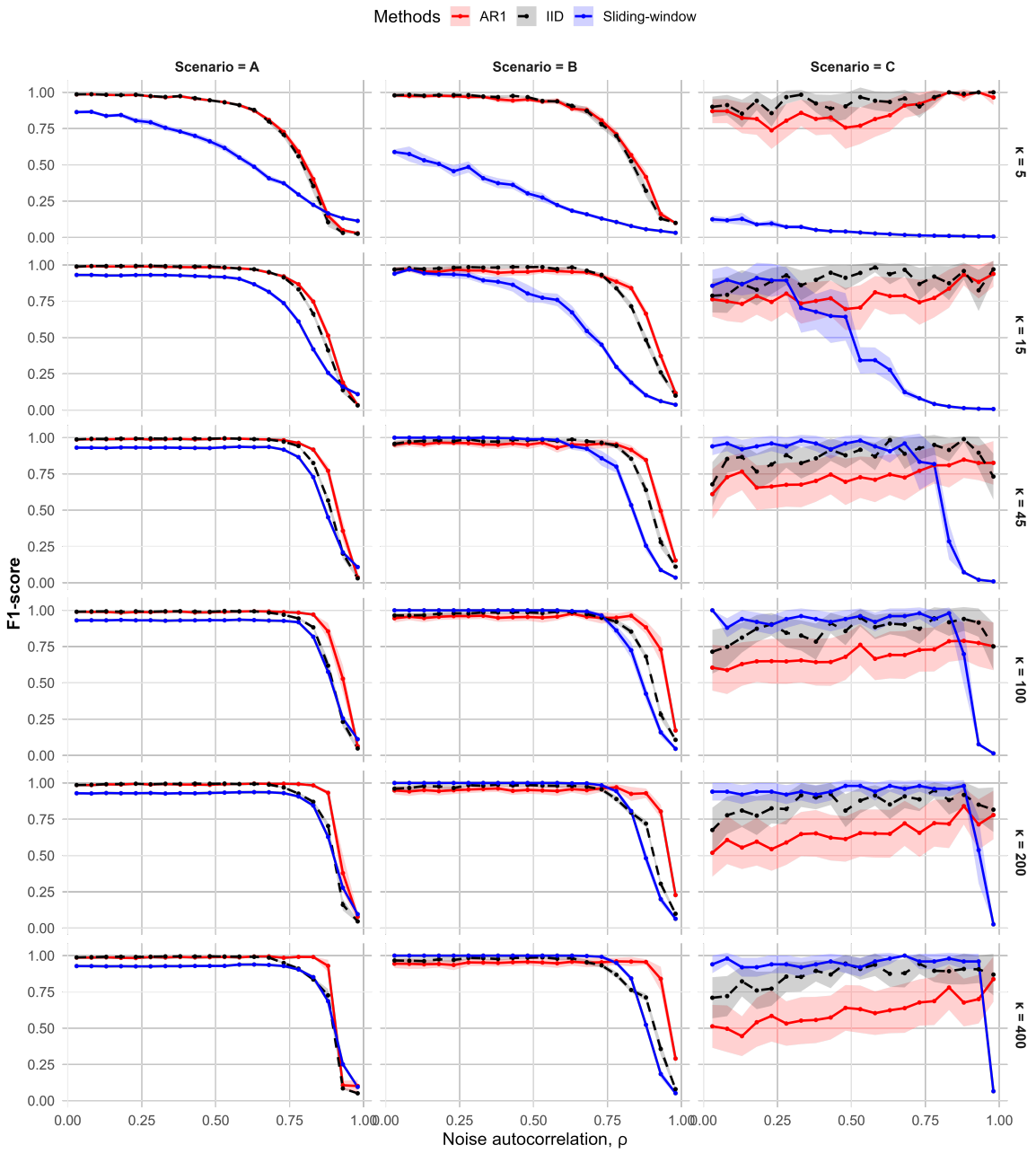


Figure 3.8: The figure presents line plots of the F1-score with wrapped normal noise as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and sliding-window (blue). The first row presents the results for scenarios A, B, and C (see Section 3.1.3) with $\kappa_\rho = 5$, 15, 45, 100, 200, and 400 in each row, respectively.

I examined the performance of the turning point identification techniques under scenarios where the underlying processes are misspecified. I first consider the case where the bearings are generated from an angular autoregressive process of order 3 (angular AR(3)), as shown in Figure 3.9. Here, the model assumes an angular AR(1) process, but the true data-generating process is angular AR(3). the proposed methods show very similar and accurate performance for small to moderate levels of autocorrelation. However, as the autocorrelation (ρ) increases, the F1-score begins to decline. This trend is consistent across different scenarios involving the number of turning points and concentration levels, suggesting that higher autocorrelation introduces challenges for turning point detection given this form of model misspecification. At very high concentration levels, the performance of the algorithm does not improve as opposed to the trends observed in Figures 3.5,3.6,3.7 and 3.8 making the sliding-window algorithm comparatively more accurate in these cases. Results from data generated under AR(10), shown in Figure 3.10, closely mirror those observed in the angular AR(3) case.

Case 4: Errors generated with non-constant κ

In the modelling setup, I assume that the concentration parameter is constant throughout the entire time series. However, in practice, this assumption may not always hold. To assess the robustness of the models under this condition, I analyse how well the methods perform when the concentration parameter varies. Figure 3.11 illustrates the accuracy of the demarcation methods across different scenarios. In Scenario A, both the AR(1) and IID models demonstrate nearly identical performance, achieving almost perfect F1-scores for values of $\rho \leq 0.60$. This indicates that both models are robust even when the concentration parameter is not constant, provided the autocorrelation is moderate. A similar trend is observed in Scenario B, where both methods maintain high accuracy as the autocorrelation decrease $\rho \leq 0.70$. In Scenario C, the two proposed methods achieve perfect accuracy across all levels of autocorrelation.

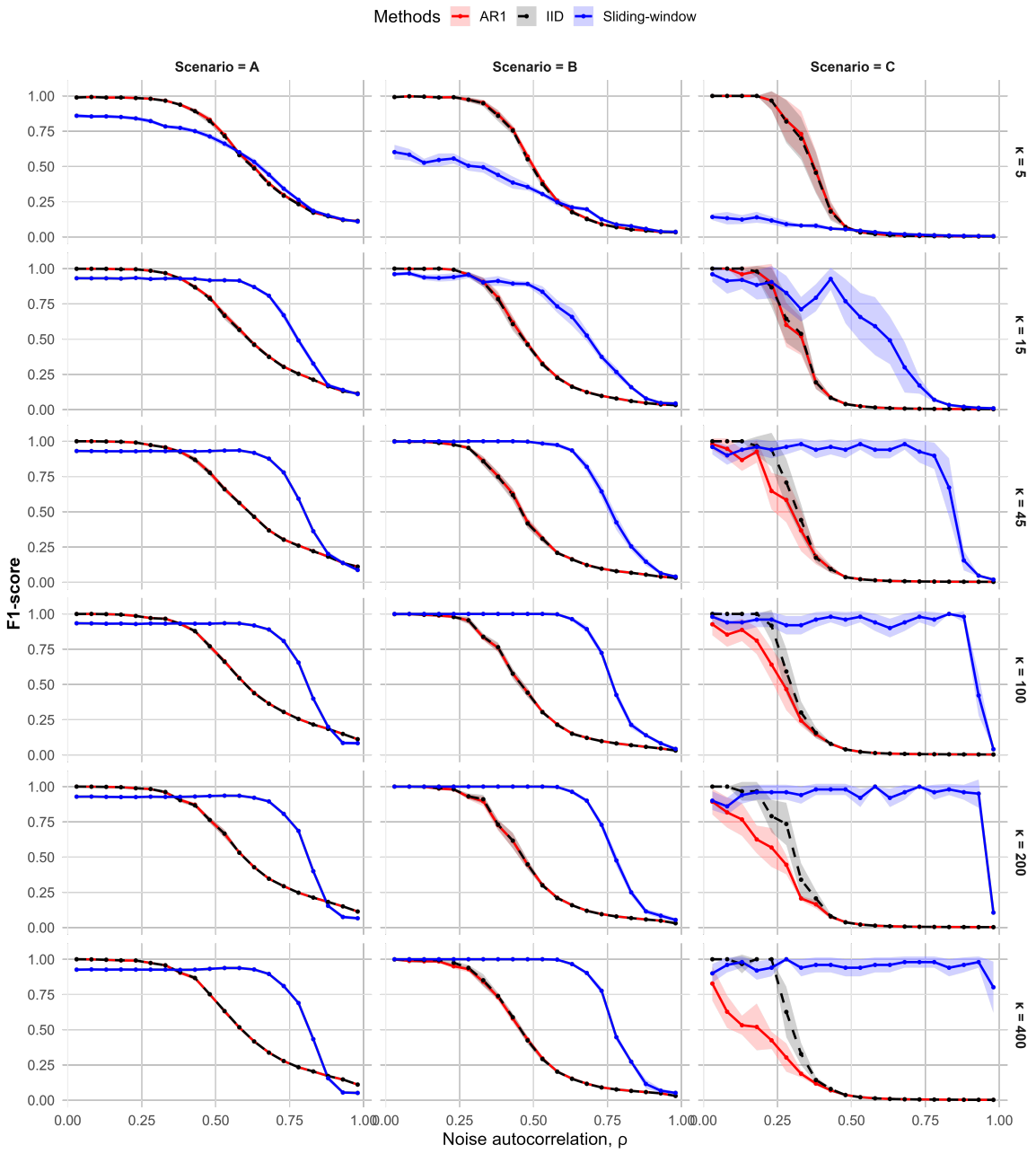


Figure 3.9: The figure presents line plots of the F1-score with angular AR(3) noise as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and sliding-window (blue). The first row presents the results for scenarios A, B, and C (see Section 3.1.3) with $\kappa_\rho = 5$, 15, 45, 100, 200, and 400 in each row, respectively.

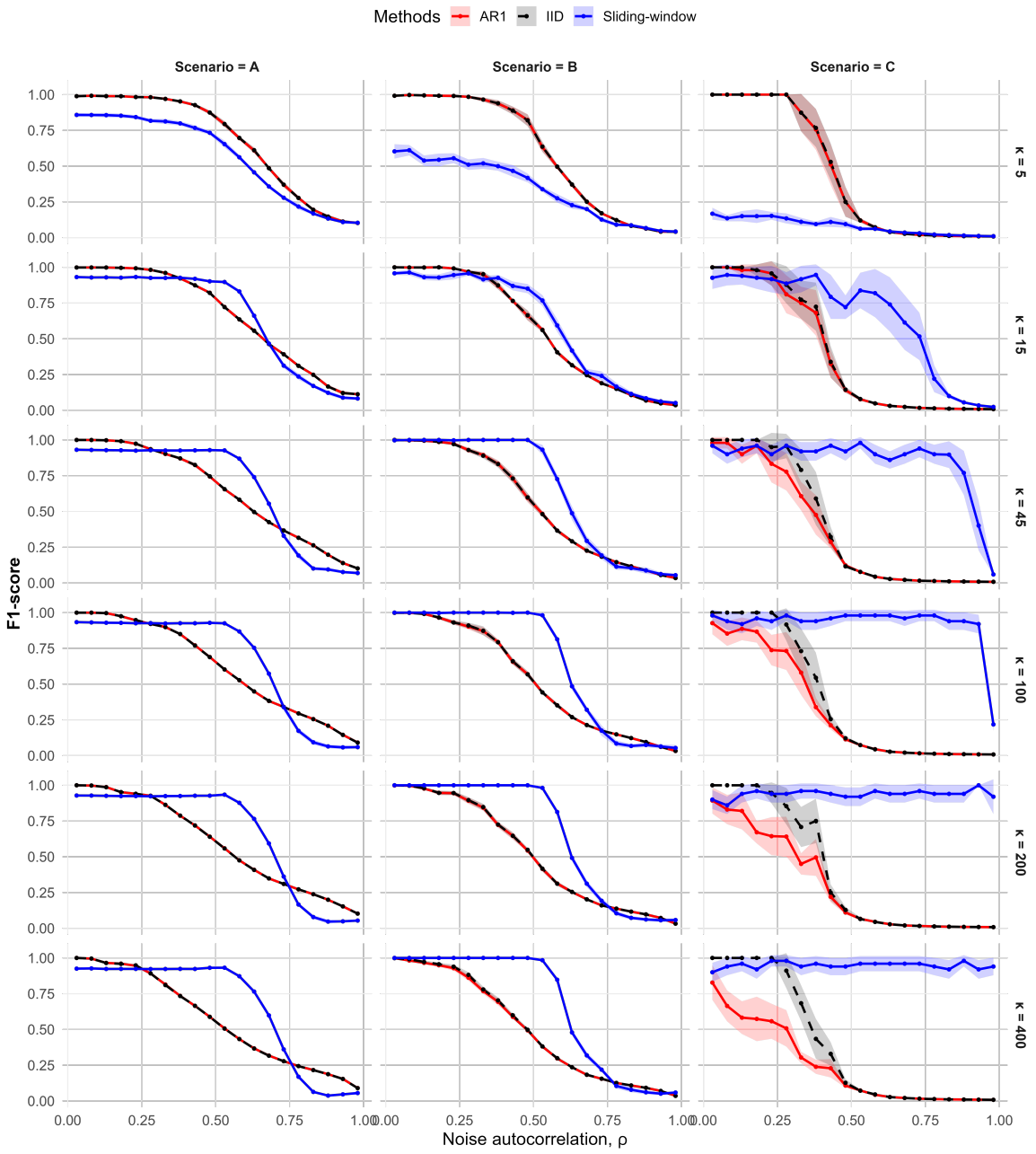


Figure 3.10: The figure presents line plots of the F1-score with AR(10) noise as a function of the autocorrelation parameter ρ, comparing the performance of three methods: AR(1) (red), IID (black), and sliding-window (blue). The first row presents the results for scenarios A, B, and C (see Section 3.1.3) with $\kappa_\rho = 5$, 15, 45, 100, 200, and 400 in each row, respectively.

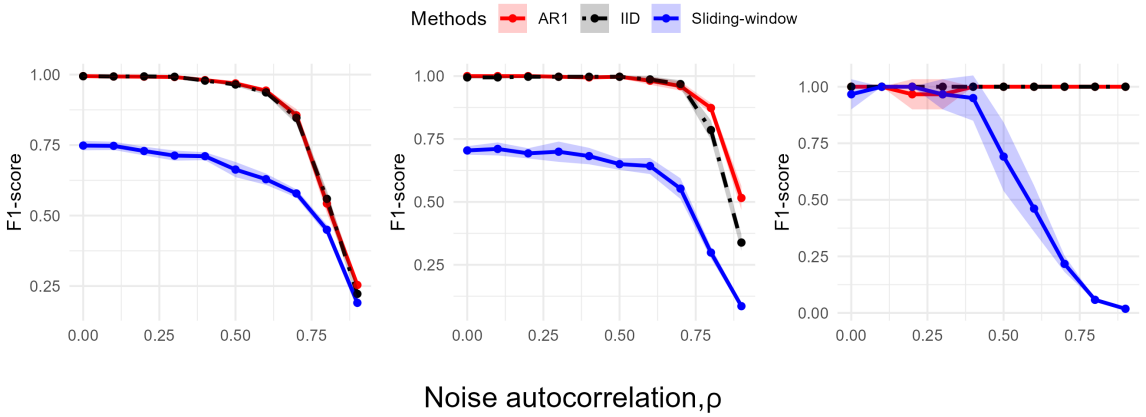


Figure 3.11: The figure presents line plots of the F1-score as a function of the autocorrelation parameter ρ, comparing the performance of three methods: angular AR(1) (red), IID (black), and sliding-window (blue) when κ_ρ is not constant. From left to right each presents the results for scenarios A, B, and C (see Section 3.1.3).

3.2.2 Application to Oryx Data

Here, I examine the ability of the proposed model using the same dataset as that I used in Chapter 2. I began with an initial analysis to examine the noise autocorrelation in the data. As in Section 2.1.6 I start by approximating the underlying signal by applying a moving median filter, as described in Section 3.1.2. Figure 3.12 shows the result of using the proposed algorithm to find turning points in an oryx path. The observed bearings suggest that the animal alternates between two main directions. The circular autocorrelation function of the residuals reveals a gradual, exponential-like decay in autocorrelation as the lag increases (Figure 3.12c). This pattern indicates strong autocorrelation starting at $\rho = 0.981$ at lag 1 and remaining significant over numerous lags before eventually approaching zero. Such behaviour is characteristic of an angular AR(1) process with a high autoregressive parameter. Additionally, the noise is highly concentrated, suggesting a large value of κ_ρ (Figure 3.12d).

Table 3.1 summarises the parameter estimates obtained for each of the six oryx paths using the angular AR(1) models. Compared to Table 2.1, The results of the two models

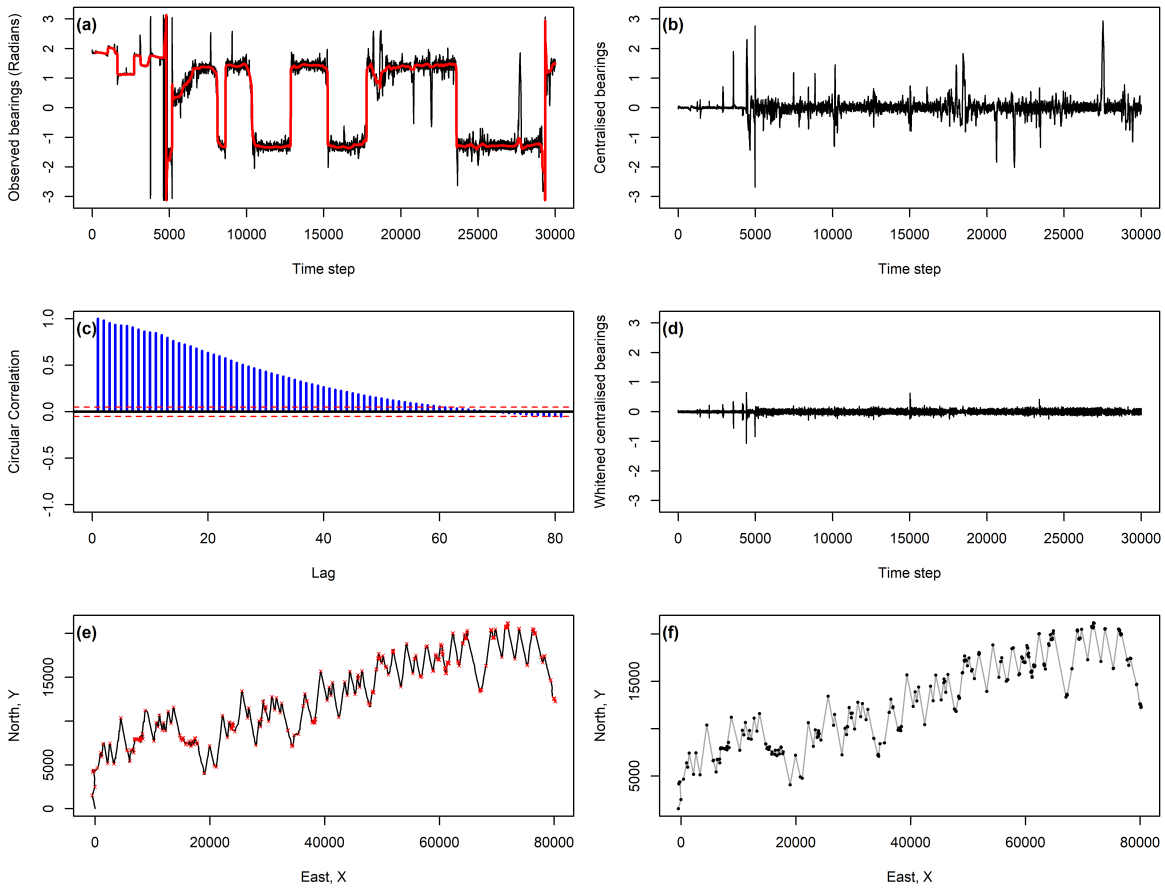


Figure 3.12: Inferring turning points from oryx data. (a): The observed bearing data (black) with the moving median estimate overlaid (red). (b): Centralized bearings based on the moving median. (c): Circular autocorrelation of the residuals. (d): Whitened centralized bearings after accounting for autocorrelation. (e): the path reconstructed from magnetometer data under the assumption of constant speed with red dots indicating the inferred turning points based on the angular AR(1) model. (f): presents the resulting model of the path as a sequence of steps and turns.

Table 3.1: Estimated number of turning points (m) and distribution parameters for the oryx data under the angular AR(1) model

ID	κ_ρ	ρ	m	σ
1	353.052	0.981	315	0.061
2	226.496	0.985	113	0.078
3	44.790	0.957	102	0.206
4	228.314	0.990	24	0.092
5	90.784	0.976	52	0.107
6	53.940	0.974	76	0.156

do not exhibit consistent patterns. Both sets of results differ from those reported by Potts et al. (2018, Table 1). Nevertheless, all algorithms identify the most visually obvious turning points, although they vary in their sensitivity to whether small apparent turns are classified as true turning points.

Our implementation of the algorithms is coded in `C++` and wrapped in `R` via `Rcpp` package (Eddelbuettel and François, 2011). The algorithms segmented a profile of length $n = 10^6$ in 7 seconds for angular AR(1) models, running on a desktop computer equipped with an Intel Core i7-6700U CPU (2.40 GHz, 4 cores) and 32 GB of RAM. The discretisation level was set to $\pi/360$ radians. These runtimes scaled linearly with n.

3.3 Summary

In this chapter, I build upon the work presented in Chapter 2, which describes an animal's movement as a sequence of straight-line segments interspersed with turns. I assume the heading data (usually obtained from a magnetometer sensor) are noisy realizations of the true direction, modelled as a piecewise constant function of time

with i.i.d. noise. However, here I modify the piecewise constant model to account for the first-lag autocorrelation in the data.

First, I construct estimators for the extent of autocorrelation in the noise, given by ρ, and the magnitude of the noise, given by $1/\kappa$, where κ is the concentration parameter of a von Mises distribution of error. I detail two methods for this purpose. The first method involves estimating κ and ρ using the angular successive differences of the observations at the first lag. The second method applies a simple moving median filter to estimate the mean steps, followed by estimating κ and ρ based on the residuals (Section 3.1.2).

In the second stage, I cannot directly apply the FPOP algorithm as in Chapter 2. Instead, I approximate the computation by discretising the space of possible headings (μ). An alternative approach for precisely identifying turning points under the angular AR model without discretising the space of μ is to use the Pruned Exact Linear Time (PELT) method (Killick et al., 2012).

The PELT algorithm optimises a global cost function for estimating changepoints, using a different pruning strategy compared to FPOP. While PELT may not surpass method proposed here in terms of speed, it offers linear time complexity under certain conditions, making it a competitive option. Since the fitting of the angular AR(1) model relies on discretisation, I recommend a discretisation level of $\pi/360$ radians, based on simulation results showing that higher resolution does not lead to improved segmentation accuracy.

I assess the performance of the proposed methods on both real-world and simulated datasets. In scenarios where movement paths consist of straight segments interspersed with turns, and the heading data are affected by noise with a standard deviation of $\pi/12$ radians and autocorrelation below 0.5, the proposed algorithms accurately detect nearly all turning points. This high detection rate marks a notable improvement over the commonly used approach introduced by Potts et al. (2018).

Estimating ρ using the robust method shows a small but consistent bias in some simulation scenarios, particularly when the noise concentration κ_ρ is low, with the worst-case bias not exceeding 0.15. This suggests that in situations with high noise and moderate autocorrelation, the method may slightly overestimate the true level of autocorrelation. Encouragingly, this bias becomes negligible as κ_ρ increases, with estimates becoming nearly unbiased for $\kappa_\rho \geq 45$ and $\rho = 0.98$. While the bias is relatively small, it is important to acknowledge it as a potential source of imprecision in downstream inference. In particular, such biases could cumulatively affect the estimation of the concentration parameters, and in turn, influence the accuracy of the turning points inference.

Chapter 4

Improving HMMs with High-Frequency Data

To what extent can the use of high-frequency location data (e.g., $\geq$ 1 Hz) enhance behavioural-state analysis compared to lower-frequency data? Behavioural state identification is often performed using hidden Markov models (HMMs) applied to step length and turning angle derived directly from GPS data sampled at fixed time intervals. However, this approach implicitly assumes that animals make movement decisions in lockstep with the times data are gathered, potentially missing key behavioural events that occur between recorded locations.

In contrast, high-frequency tracking data enable inference of the precise points at which an animal decides to turn (Potts et al., 2018; Alharbi et al., 2026). It is therefore natural to ask whether HMM-based behavioural state inference can be improved by incorporating these inferred turning points. Unlike traditional approaches where steps are defined between regularly spaced locations (e.g., Michelot et al., 2016), turning points inferred from high-frequency data are typically irregularly spaced in time, which necessitates an extension of the standard HMM framework. Therefore, I propose applying an HMM to turning points rather than raw data.

To this end, I build on the HMM formulation proposed by Morales et al. (2004) to accommodate irregular time intervals between events. I then conduct a simulation study to compare the performance of the proposed HMM approach, whereby turns are inferred using the turning-points algorithm developed in this thesis, against the previous HMM approach, whereby turns are taken from regularly-spaced low-frequency data. Finally, I apply the method here to high-resolution dead-reckoned movement data from free-ranging goats, demonstrating the algorithm's effectiveness in identifying behavioural states from animal movement trajectories. This work has been submitted for publication in *Methods in Ecology and Evolution*.

4.1 Methods

4.1.1 Hidden Markov Model framework for constant-frequency data

A HMM consists of two stochastic components: an observable process $\{\mathbf{z}_T\}_{T=1}^n$, which depends on the state, and a hidden process $\{S_T\}_{T=1}^n$, representing a finite set of latent states. The hidden process is assumed to follow a first-order Markov property, meaning that the state at time T, S_T, depends only on the previous state S_{T-1}. If all time steps have the same duration, as we are assuming throughout Section 4.1.1, then T represents the time and 1 unit is the duration between successive observations.

Additionally, the observation $\mathbf{z}_T$, $T = 1, \ldots, n$, is assumed to be conditionally independent of other observations and states, given the current state S_T. Thus, each state determines the distribution from which the corresponding observation is drawn. The transitions between states at time T are described by the state transition probability matrix $\Gamma_T = (\gamma_{ij}^{(T)})$, where the probability of transition from state i to state j at time T is

$$\gamma_{ij}^{(T)} = \Pr(S_T = j \mid S_{T-1} = i), \quad i, j = 1, \ldots, N. \tag{4.1}$$

A common parametrisation of $\gamma_{ij}^{(T)}$ is

$$\gamma_{ij}^{(T)} = \frac{\exp(\nu_{ij}^{(T)})}{\sum_{k=1}^{N} \exp(\nu_{ik}^{(T)})}. \tag{4.2}$$

Here, N denotes the total number of hidden states and $\nu_{ij}^{(T)}$ is a linear predictor with W covariates $\{x_1^{(T)}, \cdots, x_W^{(T)}\}$ given as

$$\nu_{ij}^{(T)} = \begin{cases} a_0^{(ij)} + \sum_{w=1}^{W} a_w^{(ij)} x_w^{(T)} & \text{if } i \neq j \\ 0, & \text{otherwise,} \end{cases} \tag{4.3}$$

where $a_k^{(ij)}$ is a constant denoting the strength of the effect of covariate $x_w^{(T)}$, for each $k = 0, ..., W$. The covariates $\{x_1^{(T)}, \cdots, x_W^{(T)}\}$ can be anything hypothesised to affect the probability of switching states. For example, green-up could trigger a transition from foraging to migrating for many terrestrial herbivores, or dusk could trigger the same animals to stop foraging and move back towards their resting site.

Let δ be a vector of the initial distribution of states so that the i-th element is $\delta_i = \Pr(S_1 = i)$, the probability of the first observation being in state i. Suppose our data consist of a time series $\mathbf{z}_T = (\phi_T, l_T)$, where l_T denotes the step length within the time interval $[T, T+1]$, and ϕ_T is the turning angle at time T. Then the conditional distribution of $\mathbf{z}_T$ given $S_T = i$ is

$$f(\mathbf{z}_T \mid S_T = i) = f_i(\mathbf{z}_T), \quad i = 1, \ldots, N. \tag{4.4}$$

Let $P(\mathbf{z}_T) = \text{diag}(f_1(\mathbf{z}_T), \ldots, f_N(\mathbf{z}_T))$. Then the likelihood of all the parameters given the data is

$$L(a_0^{(ij)}, ..., a_p^{(ij)} | \mathbf{z}_1, \ldots, \mathbf{z}_n) = \delta P(\mathbf{z}_1) \prod_{T=2}^{n} \Gamma_T P(\mathbf{z}_T) \mathbf{1}, \tag{4.5}$$

where $\mathbf{1}$ is a row vector of ones. We calculate this likelihood using the forward algorithm, a recursive procedure designed to compute the likelihood of an HMM efficiently (Zucchini and MacDonald, 2009).

The HMM framework is widely applied in animal movement research because it enables the identification of hidden behavioural states from observed movement data (Morales et al., 2004; Langrock et al., 2012; Li and Bolker, 2017). These states are commonly interpreted as proxies for behaviour (e.g., encamped or foraging).

4.1.2 Model Formulation

The aim here is to develop an HMM approach to infer behavioural state switching from high-frequency animal movement data (e.g. our goat data has one location per second). For this, I first use the method of Alharbi et al. (2026) to identify the locations where the animal has turned. This simplifies each high-frequency path into a sequence of turning points and the straight-line segments connecting them. These straight-line segments then become the "steps" in the subsequent analysis (Figure 4.2). Each step is characterized by both a step length (the distance travelled) and a step time (the time between successive turns). Unlike in Section 4.1.1, though, the resulting data are not necessarily regularly spaced in time.

To extend the framework of Section 4.1.1 to accommodate irregularly spaced turning points, I incorporated the distribution of step times (τ) alongside turning angles (ϕ) and step lengths (l). I redefined the emission probability as the joint probability density function of the observed variables which become $\mathbf{z}_t = (\phi_t, \tau_t, l_t)$, conditional on the animal's state $S_t = j$. Here, t indexes the turning points. Thus, unlike T (Section 4.1.1), t is not a unit of time, as the time between successive turns is not constant.

Figure 4.2 illustrates the model: the first column shows the original observed path (a) and its decomposition into steps and turns (b); the second column shows how key movement metrics are extracted and their distributions calculated (c, d, and e), potentially together with environmental covariates (f). The third column depicts the Hidden Markov Model (HMM) structure used for inference, which takes the movement metrics and environmental covariates as inputs and can subsequently be used to compute summary statistics describing behavioural processes.

As the turning points are being treated as known, I assumed that the step time τ_t is independent of the turning angle ϕ_t at turning point t, conditional on the state j. That is, the time between turns is not dependent on the size of the turns. This is in contrast with the application of HMMs to actual observations at different intervals, where there is likely to be a relationship between frequency of observation and the observed turning angles, due to unmodelled details of the movement in between observations.

I emphasise that this conditional independence assumption is a simplification rather than a direct consequence of the model construction. In the movement ecology literature based on regularly sampled data, dependence between turning angles and step characteristics is commonly observed, as both are influenced by unobserved movement between sampling times.

By contrast, in the proposed framework, steps are defined by detected turning points, so that each step represents an approximately straight movement segment between two behavioural decisions. This construction is intended to reduce the influence of within-step path variability, which is a key source of dependence under regular sampling. Conditional on the behavioural state, it is therefore reasonable to interpret τ_t as the duration of persistence in a movement mode, while ϕ_t reflects the reorientation at the end of that step.

Nevertheless, the independence between τ_t and ϕ_t remains a strong assumption. Some residual dependence may still arise, for example through averaging effects over longer step durations or through unmodelled heterogeneity within states. In this model, part of the dependence structure is instead captured through the latent behavioural states, which allow different joint regimes of movement to be represented.

Allowing for explicit dependence between variables within states would be a natural extension, for example via joint or conditional distributions, but this would substantially increase model complexity and is beyond the scope of the present work. Similar independence assumptions are commonly adopted in standard HMM formulations for

movement data, despite their known limitations. Additionally, I assumed that the step length l_t depends on both the step time τ_t and the state j. Under these assumptions, this results in the following decomposition

$$f_j(\mathbf{z}_t|S_t = j) = f_j((\phi_t, \tau_t, l_t)|S_t = j) = g_1(\phi_t|j) \; . \; g_2(\tau_t|j) \; . \; g_3(l_t|\tau_t, j). \tag{4.6}$$

The functions $g_1(\phi)$, $g_2(\tau)$, and $g_3(l)$ are the density functions of turning angles, step times, and step speeds (i.e. lengths given times), respectively.

I used gamma distributions for both step time and step speed. For the turning angles, I used the following distribution (Figure 4.1)

$$g_1(\phi|\kappa, w) \propto e^{\kappa \cos(\phi)}(1 - e^{-w(1-\cos(\phi))}) \qquad \phi \in \mathbb{R} \text{ and } \kappa, w > 0. \tag{4.7}$$

This is constructed from a von Mises distribution centred at 0, but with reduced density near zero to reduce the probability of very small turns. These small turns are usually not identified as turns by the turning points algorithm as they are often indistinguishable from noise (see the turning angles histogram in Figure 4.2.c).

I also examined a simpler set of models in which turning angles were modelled using a von Mises distribution with zero mean, step times are modelled using an exponential distribution, and step speeds were modelled using a gamma distribution. Under this formulation, the times at which movement decisions occur follow a Poisson process.

An important assumption in the proposed framework is that state switches occur precisely at the times of turns. Furthermore, I assumed that the transition probabilities are unaffected by this extension; that is, they remain constant and do not depend on the irregular time intervals between observations.

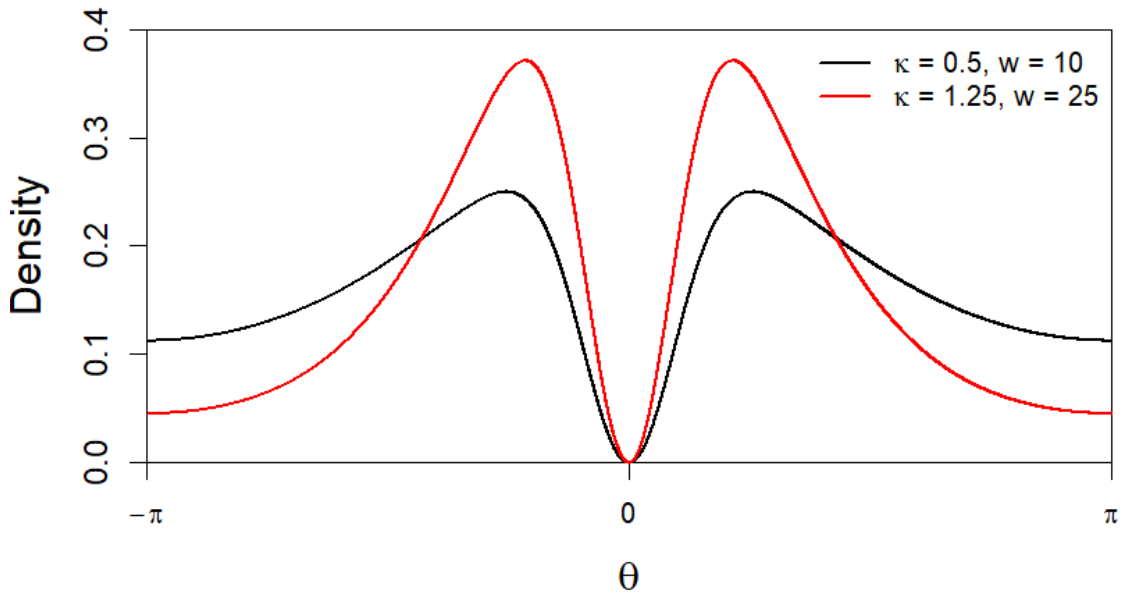


Figure 4.1: Example plots of Equation 4.7.

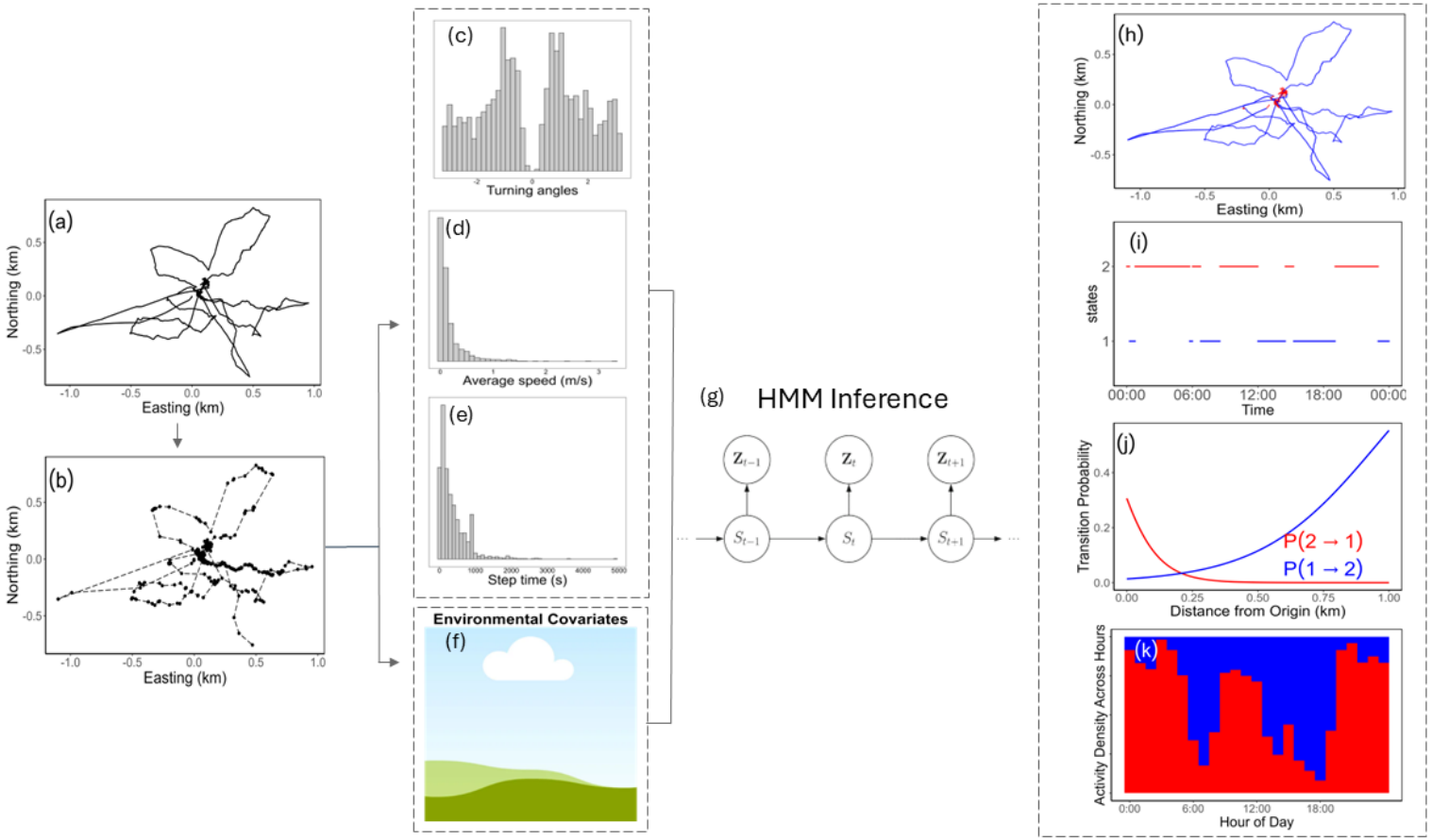


Figure 4.2: Conceptual framework for modelling behavioural states using high-frequency movement data. The first column shows the original observed path and the decomposed trajectory into steps and turns. The second column displays the distributions of key movement metrics, potentially combined with environmental covariates. The third column illustrates the Hidden Markov Model (HMM) structure used for inference. The last column presents various examples of post-hoc visualisations, explained in detail later.

4.1.3 Model Fitting

I fitted the proposed model using direct likelihood optimisation. The log-likelihood was computed using the forward algorithm (Zucchini and MacDonald, 2009), implemented in `C++`, and maximised using the `optim` function in `R` (Team, 2020). Given the sensitivity of numerical optimisation to initial values, Michelot et al. (2016) recommended performing multiple optimisation runs with randomly selected starting values. In this study, I developed a more systematic approach for selecting suitable initial values. Specifically, I applied k-means clustering to moving averages and variances computed from the movement variables: $\cos(\phi)$, $\sin(\phi)$, τ, and s. These statistics were calculated over a moving window of fixed size that was shifted sequentially across the time series, with the mean and variance recalculated at each step.

The number of clusters in the k-means algorithm was set equal to the number of hidden states in the HMM. This clustering procedure provided an initial approximation to the underlying state sequence. Emission distribution parameters were then initialized by computing empirical summaries of the observations assigned to each cluster. Likewise, the transition probability matrix was initialized from the estimated cluster sequence by calculating the empirical frequencies of transitions between consecutive clusters. When covariates were included in the transition probabilities, separate multinomial logistic regression models were fitted for each originating state using the preliminary cluster assignments as the response variable. More specifically, for a given state i, the observed transitions from state i to all possible destination states were treated as multinomial outcomes, and the corresponding covariates were used as predictors. The resulting regression coefficients were then used as initial values for the covariate-dependent transition probability parameters in the HMM. Additionally, the initial state distribution was initialized under the assumption of stationarity by solving $\delta\Gamma_1 = \delta$. These initial parameter values provided a coherent starting point, which was subsequently refined during likelihood maximisation.

Since the choice of window size for calculating the moving statistics influences the

clustering results, I explored a range of window sizes. The final model was selected as the one yielding the highest likelihood across all fitted models. After estimating the HMM parameters, the hidden state sequence was decoded using the *Viterbi algorithm* (see e.g. Zucchini and MacDonald, 2009). This algorithm determines the most probable sequence of hidden states, S_t, given the observed data and the estimated model parameters.

4.1.4 Empirical case study: alpine goats

To evaluate the performance of the proposed HMM framework on empirical data, I applied it to six trajectories derived from 1 Hz dead-reckoned location data of domestic goats (*Capra aegagrus hircus*), previously published in Munden et al. (2021). The data were collected in the Bauges Mountains (Massif des Bauges; 45.61°N, 6.19°E) in the French Alps. Further details on the study site, data collection, and dead-reckoning procedures can be found in Munden et al. (2021).

I fitted four different two-state HMMs to each trajectory:

- **Homogeneous Model (No Covariates):** Behavioural states were inferred without incorporating any environmental covariates, so that (recalling Equation 4.3)

$$\nu_{ij} = a_0^{(ij)}. \tag{4.8}$$

- **Distance-Based Model:** The distance to a known central location, denoted d_t, was included as a covariate. Goats typically spent much of their time in this central location, which contained a pen and nearby salt licks. The transition probability linear predictor was defined as

$$\nu_{ij}^{(t)} = \begin{cases} a_0^{(ij)} + a_1^{(ij)} d_t & \text{if } i \neq j \\ 0 & \text{otherwise.} \end{cases} \tag{4.9}$$

- **Time-of-Day Model:** The time of the day, H_t, measured in hours so that

$H_t \in [0, 24)$, was incorporated as a cyclic covariate as follows

$$\nu_{ij}^{(t)} = \begin{cases} a_0^{(ij)} + a_2^{(ij)} \cos\left(\frac{\pi H_t}{12}\right) + a_3^{(ij)} \sin\left(\frac{\pi H_t}{12}\right) & \text{if } i \neq j \\ 0 & \text{otherwise.} \end{cases} \tag{4.10}$$

- **Combined Model:** Both time of day and distance to the central location were included so that

$$\nu_{ij}^{(t)} = \begin{cases} a_0^{(ij)} + a_1^{(ij)} d_t + a_2^{(ij)} \cos\left(\frac{\pi H_t}{12}\right) + a_3^{(ij)} \sin\left(\frac{\pi H_t}{12}\right) & \text{if } i \neq j \\ 0 & \text{otherwise.} \end{cases} \tag{4.11}$$

In the homogeneous model, environmental effects can only be examined post hoc by relating inferred behavioural states to external variables after model fitting. In contrast, the remaining models incorporate environmental covariates directly into the transition dynamics, enabling formal inference about their influence on state-switching behaviour. This allows for the evaluation of whether including such covariates improves model fit and interpretability.

To benchmark the introduced approach, I also compared these models to a standard HMM applied to regularly subsampled trajectories, with step-time intervals ranging from 5 to 240 minutes. For this comparison, I modelled step lengths using a gamma distribution and turning angles using a von Mises distribution, following Patterson et al. (2017). All models were initialised using the k-means clustering procedure described in Section 4.1.3, testing rolling window sizes from 5 to 100 in increments of 5.

I examined two sets of movement models. In the first structure, the turning angles were modelled using Equation (4.7), while both step times and step speeds were modelled using gamma distributions; I refer to these models as MA1–MA4. As an alternative, I considered a second, simpler structure, MB1–MB4, in which turning angles followed a von Mises distribution and step times were modelled using an exponential distribution. Together, these two structures yielded a total of eight HMMs for each trajectory.

4.1.5 Simulation Study

To evaluate the accuracy of the proposed model, I conducted a simulation study. I fixed the number of turning points at $m = 2000$, which is similar to the number observed in the goat data in Section 4.1.4. I generated three two-state scenarios, labelled A, B, and C, as illustrated in Figure 4.3. These scenarios differ in their switching patterns. Scenario A features strong periodicity, which may reflect behaviours associated with day–night cycles. Scenario B is similar to Scenario A but with a reduced influence of periodic patterns. In contrast, Scenario C represents a purely random switching process, such as might arise from stochastic responses to unpredictable environmental cues.

Empirically, the mean durations of visits to states 1 and 2 were 136 and 150, respectively, in Scenario A; 103 and 97 in Scenario B; and 44 and 66 in Scenario C. These simulated state sequences were then used to generate synthetic turning angles, step durations, and step speeds.

For state 1, turning angles were modelled using Equation (4.7) with $\kappa_1 = 0.5$ and $w_1 = 10$. Step times were generated from a gamma distribution with shape parameter $\lambda_1 = 0.56$ and rate parameter $\delta_1 = 0.002$, corresponding to a mean duration of 280 seconds. Step speeds followed a gamma distribution with shape $\theta_1 = 0.34$ and rate $\alpha_1 = 2$. These parameters were chosen to reflect characteristics observed in the goat dataset.

For state 2, each parameter was increased by 150%. Increasing the parameters of the turning angle distribution for state 2 results in more directed movement, making large turning angles less probable compared to state 1. The resulting difference between the two turning angle distributions is illustrated in Figure (4.1).

Additionally, inflating both the shape and rate of the gamma distributions proportionally keeps the average step time and step speed unchanged but reduces their variability, making movements in state 2 more regular compared to state 1. Once the step times, turning angles, and step speed were generated, I conducted two main experiments to

assess how turning point estimation affects behavioural inference.

First, I supplied the HMM with the true step metrics derived directly from the simulated data. In other words, the model was given access to the exact locations of the true turning points. I term this the oracle setting. This setup serves as a reference benchmark, allowing us to evaluate the best-case performance of the HMM independently of any errors introduced by the turning point detection algorithm. Although such perfect knowledge is unattainable in real-world applications, the oracle scenario provides a valuable upper bound for assessing the accuracy and reliability of the full modelling framework. Second, I relaxed this assumption by reconstructing the animal's path from the same simulated data under the assumption of constant speed between turning points. Specifically, I generated high-frequency bearing and speed sequences as follows

$$\begin{aligned} h_k &= \left[\sum_{j=1}^{i} \phi_j\right] \mod 2\pi, \\ v_k &= \frac{s_i}{\tau_i} \qquad \text{for all } k \text{ such that } \sum_{j=1}^{i-1} \tau_j < k \leq \sum_{j=1}^{i} \tau_j \end{aligned} \tag{4.12}$$

Here, τ_i represents the duration (in steps) of the i-th segment. The x- and y-coordinates were then reconstructed using

$$x_t = \sum_{k=1}^{t} v_k \cos(h_k), \qquad y_t = \sum_{k=1}^{t} v_k \sin(h_k), \tag{4.13}$$

for $t = 1, \ldots, m$, where $m = \sum_{i=1}^{n} \tau_i$ denotes the total number of high-frequency steps.

To simulate realistic observation conditions, I added angular noise to the bearings using a first-order autoregressive von Mises process. This captures directional uncertainty typical in real trajectory data. The noise was simulated with concentration $\kappa_h = 30$ and autocorrelation $\rho_h = 0.6$, consistent with empirical estimates from the goat dataset. From the noisy high-frequency trajectories, I estimated turning point locations using the turning points detection algorithm (Alharbi et al., 2026). Step times, turning

angles, and step speeds were then reconstructed based on these estimated turning points and fed into the same HMM as in the oracle experiment.

To compare the proposed approach with an HMM approach based on regularly sampled data, I applied uniform subsampling using fixed time intervals ranging from 5 to 240 minutes. For the inference procedure, turning angles were modelled with a von Mises distribution, and step lengths with a gamma distribution, following the method of Patterson et al. (2017). For parameter initialization, I employed k-means clustering based on rolling means and variances, as described in Section 4.1.3. Window sizes ranged from 5 to 100 in steps of 5. Each configuration was repeated 20 times to ensure stability and robustness.

Model performance was evaluated in two ways. First, I assessed classification accuracy by measuring the percentage of high-frequency steps correctly assigned to their true states. Second, I evaluated the accuracy of parameter estimation by comparing the fitted parameters of the emission distributions against their true values. This design provides insight into both the state decoding performance and the robustness of parameter recovery under various sampling and inference conditions.

In a complementary analysis, I repeated the entire experiment using a simpler model specification. Here, turning angles followed a von Mises distribution with $\kappa_1 = 0.12$, step-times followed an exponential distribution with $\lambda_1 = 0.0037$ (again consistent with empirical estimates from the goat dataset), and step speed retained the same gamma distribution as before. As in the previous setup, parameters for state 2 were scaled by 150%.

Finally, I extended the simulation study to include three-state scenarios. As in the two-state case, three distinct sequences of states were simulated, denoted Scenarios A, B, and C, each with different transition dynamics (see Figure 4.3). Generally, three states might correspond to resting, grazing, and foraging behaviours, as observed in previous studies (Grainger and Blackwell, 2025; Pohle et al., 2017; Blackwell et al.,

2016).

Scenarios A and B were generated using semi-Markov chains, allowing transition probabilities to depend on the time since the last state switch. This approach enabled us to impose specific dwell-time patterns, such as longer persistence in certain states before switching (e.g., from state 1 to 2 and from state 2 to 3), despite not being explicitly accounted for in the fitted model.

Scenario C was simulated using a standard Markov chain, where transitions depend only on the current state. The mean durations of visits to states 1, 2, and 3 were 116, 109, and 107 in Scenario A; 75, 94, and 80 in Scenario B; and 60, 49, and 130 in Scenario C, with parameters for state 3 scaled by 300%.

A key consideration in the design of the simulation study is the choice of the data-generating process. In this work, the simulations are constructed such that behavioural changes occur at turning points, which aligns with the modelling framework developed in this thesis. As a result, the comparison with standard HMMs applied to regularly spaced data introduces an element of model misspecification, since those models implicitly assume that decisions occur at fixed time intervals.

However, this choice reflects a modelling philosophy rather than a bias in favour of the proposed method. In most biological systems, there is no clear justification for assuming that animals make decisions at regular time intervals. Instead, behavioural changes are more plausibly associated with specific movement events, such as turns.

To provide additional perspective, it is informative to consider the limiting case where step durations exhibit very low variability, for example under Gamma distributions with large shape parameters. In such settings, decision times become approximately regular, and the turning-point representation converges towards a regularly spaced process. In this case, both modelling approaches would be expected to yield similar results. Conversely, when variability in decision timing is high, regularly timed HMMs are more prone to misspecification, which motivates the development of the proposed

approach.

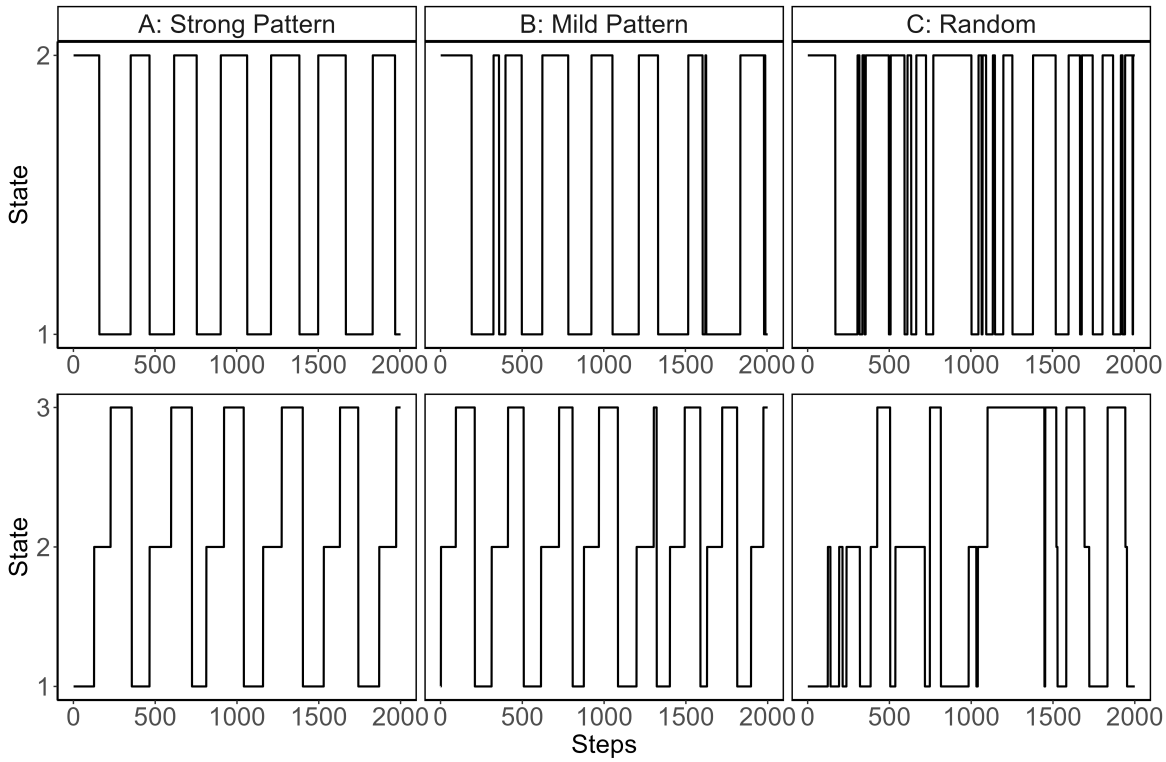


Figure 4.3: Time series plots of simulated state sequences across six scenarios. The columns show the state sequences over time for Scenarios A, B, and C. The y-axis represents the discrete state values, and the x-axis denotes simulation steps. The first row corresponds to the two-state models, while the second row corresponds to the three-state models.

4.2 Results

4.2.1 Simulated Data

4.2.1.1 The Default Emission Model

Two-state case. The performance of the extended HMM was evaluated against conventional constant-interval subsampling across three synthetic two-state movement

regimes (See first row in Figure 4.3). Turning angle series were generated from the distribution given in Equation (4.7), while step speed and step durations followed gamma distributions. Across all simulations, the mean trajectory length was $539{,}737.3 \pm 14{,}070.27$ (SD). The average counts of true turning point detections, false positives, and false negatives were $1{,}808 \pm 15.6$, 29 ± 5, and 156 ± 15.6 (SD), respectively, indicating that the turning point detection algorithm introduced a modest number of classification errors.

The top row of Figure 4.4 shows an example simulated trajectory with super-imposed turning points (Panel a), together with the corresponding time series of states (Panel b). The remaining rows show the results of the various different inference procedures. First, when I assumed perfect data about the path (Panels c-d), the inferred turning points and states are visually indistinguishable from the truth. When I relax this assumption, assuming some noise in the data-gathering process, and using the algorithm of Alharbi et al. (2026) to infer turning points, it still returns results that are visually indistinguishable from reality (Panels e-f). However, subsampling at regular time intervals prior to applying the HMM, inference of behavioural state can be quite far from correct (Panels g-l). Subsampling every 5 minutes leads to inferred states that switch back and forth between 1 and 2 at a much higher rate than in reality. Using 15-minute intervals produces results that are somewhat closer to reality but still far from accurate. In contrast, subsampling every 30 minutes results in an unrealistically slow switching rate, again leading to poor state classification.

The top row of Figure 4.5 demonstrates, via summary statistics, that similar conclusions occur across all three scenarios (A–C). The turning points (TPs) method nearly matches the oracle's accuracy in classifying states, whereas regular subsampling performs markedly worse as the sampling frequency decreases (top row). The second and third rows compare oracle-based estimates (row 1) with TPs detection (row 2). The former shows relative errors for all parameters clustered around zero, as expected.

For TPs detection (row 2) there is a slight bias for a few parameters, particularly in

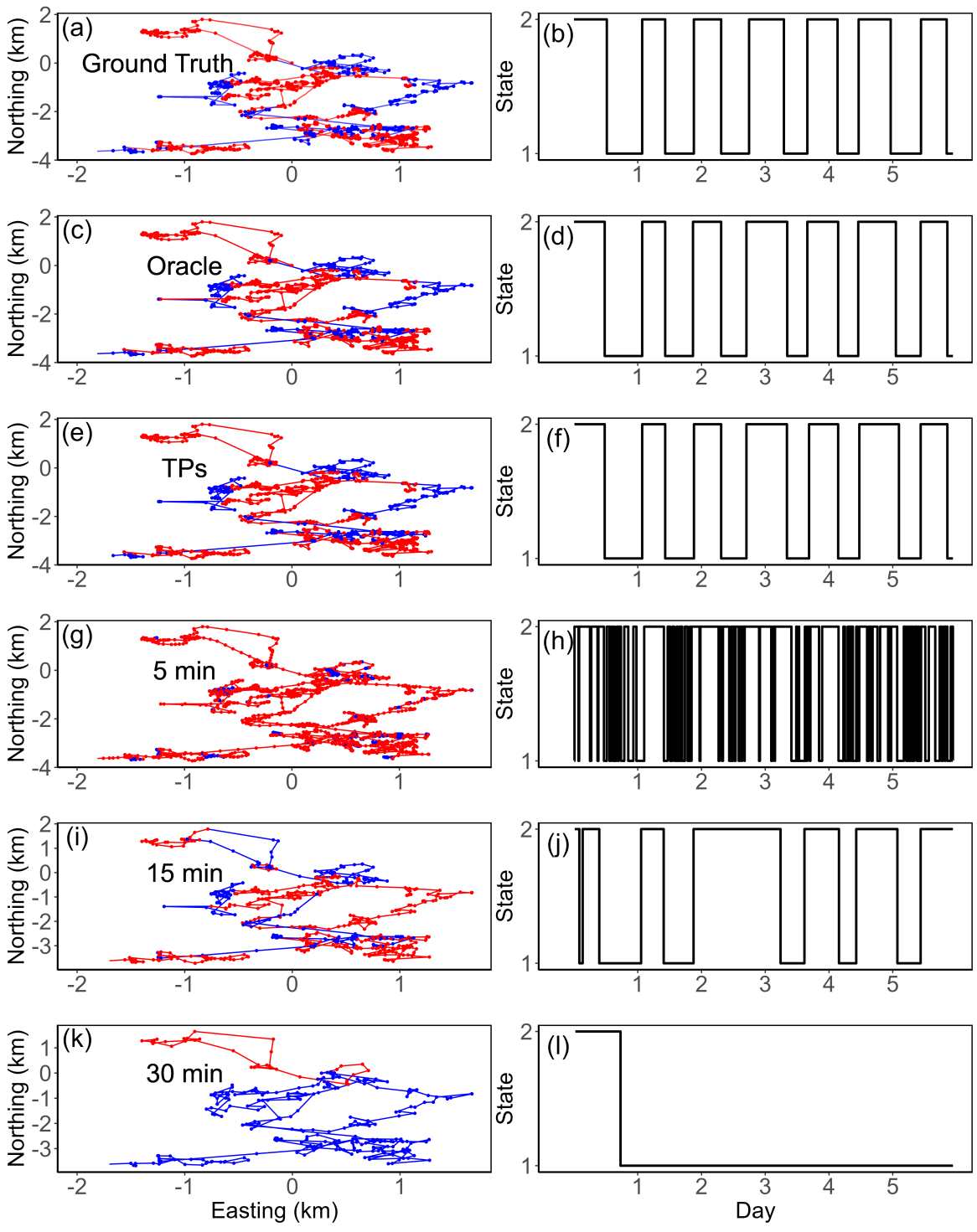


Figure 4.4: Comparison between the standard HMM and the proposed HMM-based approach applied to simulated movement data from Scenario A in a two-state case. Example decoded state sequences are shown under different subsampling levels, where blue represents State 1 and red represents State 2.

w_1 and w_2, and to a lesser extent in λ_1, reflecting misestimation of turn times. Despite these parameter-estimation errors, the TPs method achieves ∼95–99% correct labelling, nearly identical to the oracle. In the constant-step subsampling, accuracy peaks at only 75–85% for moderate intervals (10–15 min), while finer subsampling intervals perform much worse. Larger subsampling leads to a steady decline, approaching $\sim 50\%$ at the largest intervals (Figure 4.5, top row).

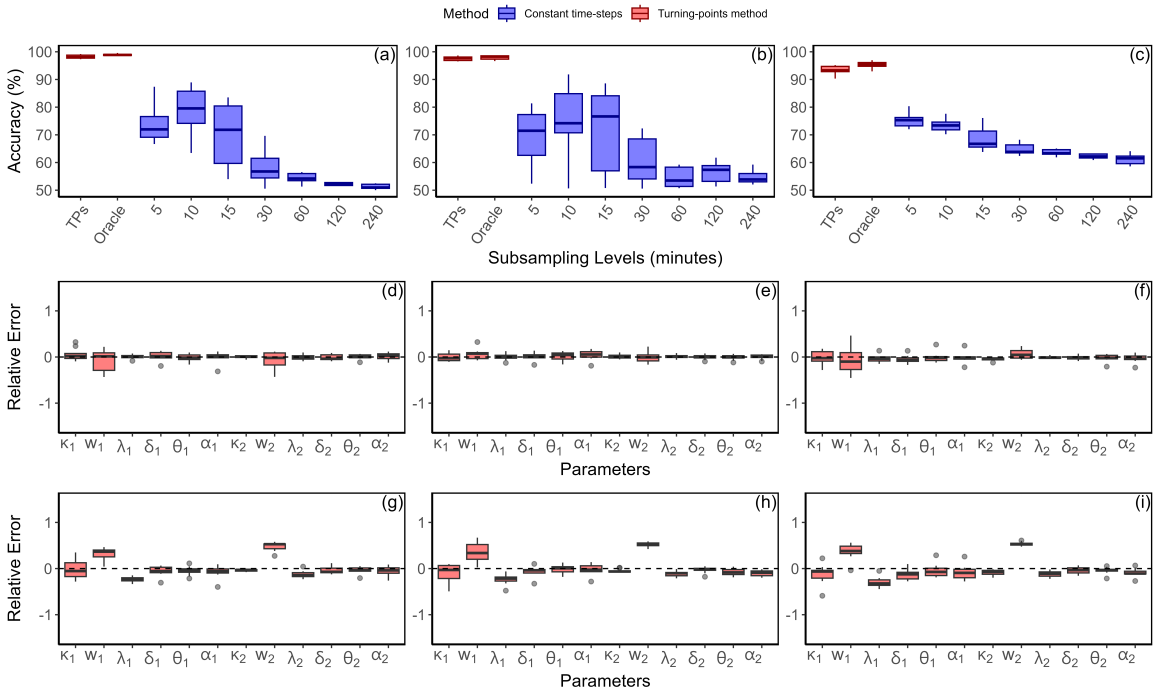


Figure 4.5: Results of the simulation analysis for the two-state model. The first row displays classification accuracy. Red boxplots represent results obtained using the TPs method under oracle settings, while blue boxplots correspond to results from regular subsampling at constant time steps. The second row shows the relative errors in parameter estimation for each state under oracle settings, and the third row shows the relative errors under TPs settings. Columns correspond to scenarios A, B, and C.

Three-states case. Across all three-state simulation scenarios (second row in Figure 4.3), the mean trajectory length was $539{,}014.2 \pm 14{,}361.6$ (SD). On average, the turning-point detection algorithm identified $1{,}812 \pm 20.48$ true turning points and pro-

duced only a modest number of errors: 44.43 ± 6.3 false positives and 188.93 ± 20.48 false negatives (SD).

Figure 4.6 summarises the parameter estimation results. Under the oracle scenario (row 2), mean errors remain close to zero; however, the error distributions—particularly for states 2 and 3 in Scenario B become broader and more skewed. When turning points are estimated (row 3), mean errors again remain near zero, but a similar systematic bias appears in the truncation parameter w.

Despite these biases in parameter estimation, the extended HMM still recovers the hidden states with high fidelity, achieving approximately 95–99% accuracy, virtually identical to the oracle results. In contrast, constant-interval subsampling performs much worse: accuracy peaks at only 50–60% for intermediate sampling rates (5–15 min) and drops to around 40% at the coarsest resolutions (row 1). An example simulated trajectory for the three-state case is shown in Figure 4.7.

It is important to interpret these results in light of the simulation design. Since the data-generating process is based on turning points, the proposed method is naturally well aligned with the underlying structure of the data. In contrast, regularly timed HMMs operate under a different set of assumptions, leading to model misspecification when decision times are irregular.

This does not imply that the turning-point method is universally superior, but rather that its advantages become most pronounced when the timing of behavioural decisions is inherently irregular. When decision times are approximately regular, the performance gap between the two approaches is expected to diminish.

For the case where I assumed that the turning angle data are drawn from a von Mises distribution, step speeds follow a gamma distribution, and step durations follow an exponential distribution:

- In the two-state model, across all simulations, the mean trajectory length was $329{,}774 \pm 19{,}426$ (SD). The average number of true turning-point detections,

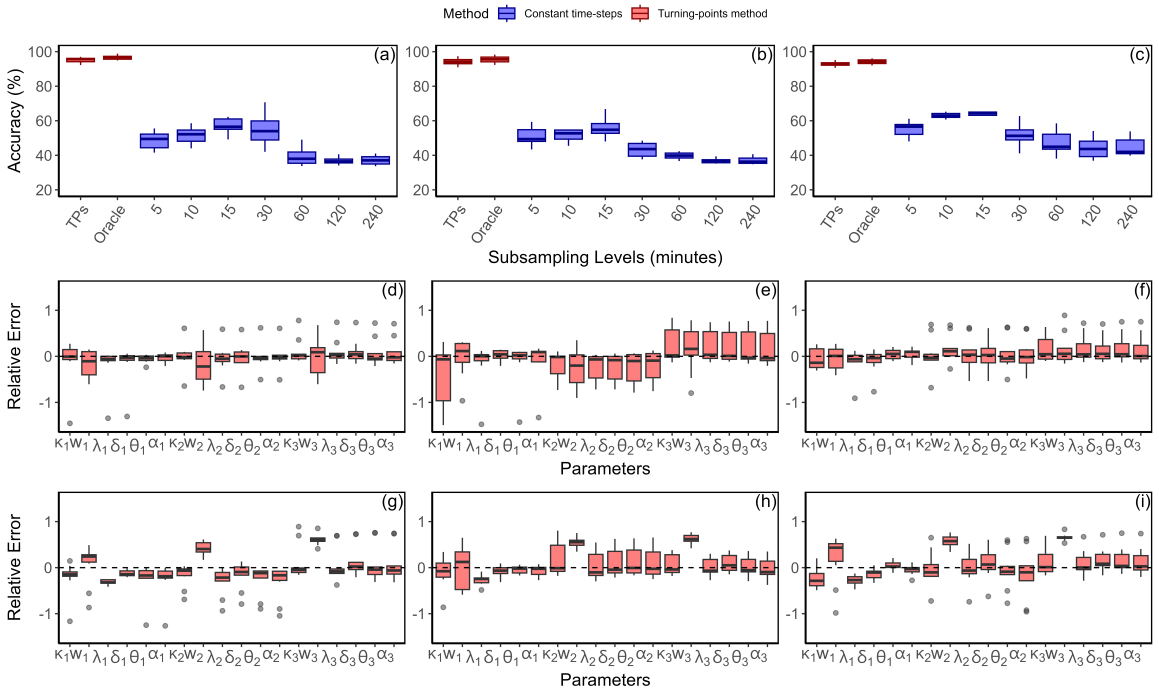


Figure 4.6: Results of the simulation analysis for the three-state model. The first row displays the classification accuracy. Red boxplots represent results obtained using the TPs method under oracle settings, while blue boxplots correspond to results from regular subsampling at constant time steps. The second row shows the relative errors in parameter estimation for each state under oracle settings, and the third row shows the relative errors under TPs method. The third Columns correspond to scenarios A, B, and C in the three-states case.

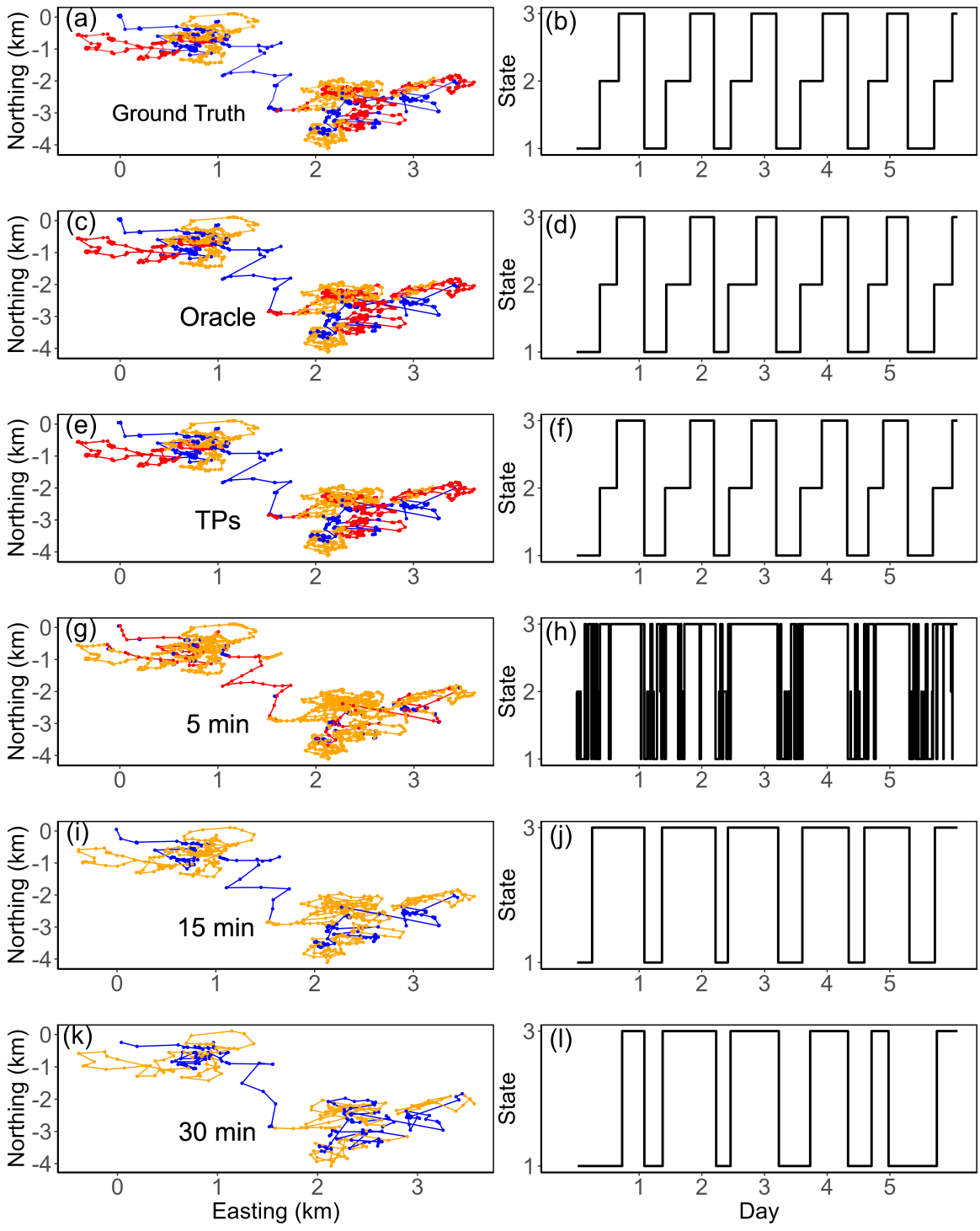


Figure 4.7: Comparison between the standard HMM and the proposed HMM-based approach applied to simulated movement data from Scenario A in a three-state case. Example decoded state sequences are shown under different subsampling levels, where blue represents State 1, red represents State 2 and orange represents State 3.

false positives, and false negatives were $1{,}600 \pm 21$, 8 ± 3, and 324 ± 21 (SD), respectively.

- In the three-state model, the mean trajectory length was $296{,}602 \pm 13{,}636$ (SD), with corresponding averages of $1{,}551 \pm 25$ true positives, 7 ± 2 false positives, and $1{,}623 \pm 25$ (SD) false negatives.

Figures 4.8 and 4.9 show the results for the two-state and three-state models, respectively. In both cases, state identification closely aligns with the results obtained under the more complex emission-probability formulation (see Section 4.2.1.1). However, parameter recovery reveals noticeable biases in the estimated parameters, particularly for the turning angle and step time distributions.

4.2.2 Goat Data

Figure 4.10 shows an illustrative comparison of the extended HMM (using TPs method) against fixed-interval subsampling for a single goat. In each row, the left panel depicts the reconstructed path, and the right panel shows the corresponding decoded state (blue = state 1; red = state 2) over a 24-hour period. Compared to my simulation experiments, the behavioural state classification is relatively robust to subsampling when applied to the goat data. Indeed, with 5-minute or 10-minute subsampling, we see only slight differences in behavioural classification compared with the turning points approach. However, longer subsampling intervals produce increasingly large discrepancies in the inferred behavioural states. This pattern is consistent across the different models (Figure 4.11): the highest agreement (92 %) between inferred state sequences occurred with 5-minute subsampling intervals, but matching accuracy declined rapidly as the subsampling window increased.

Although all four models produced almost identical segmentations (agreement $> 99.8\%$), incorporating environmental covariates into the transition–probability matrix allowed us to better interpret behavioural switches. Figure 4.12 shows an example of the

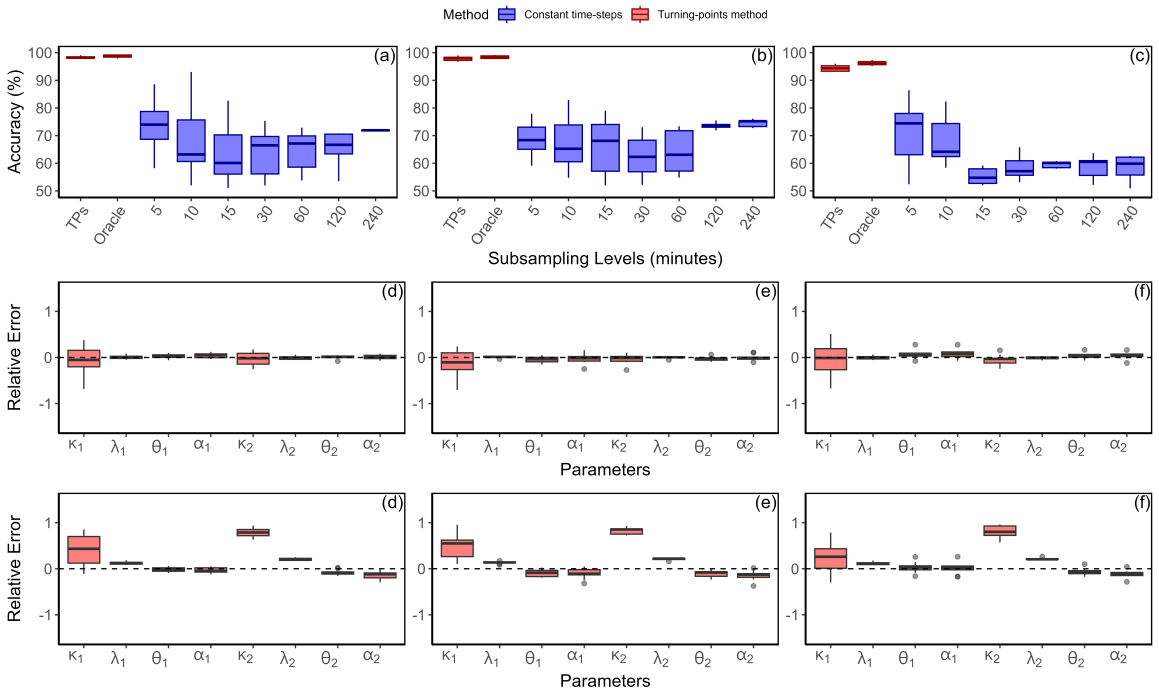


Figure 4.8: Results of the simulation analysis for the two-state model. The top row shows classification accuracy. Red boxplots represent results obtained using the turning-points (TPs) method under oracle settings, while blue boxplots correspond to regular subsampling at constant time steps. The second row shows relative errors in parameter estimation under oracle settings, and the third row shows relative errors under turning-point detection settings. Columns correspond to scenarios A, B, and C. All results assume turning angles follow a von Mises distribution, step times an exponential distribution, and step speeds a gamma distribution.

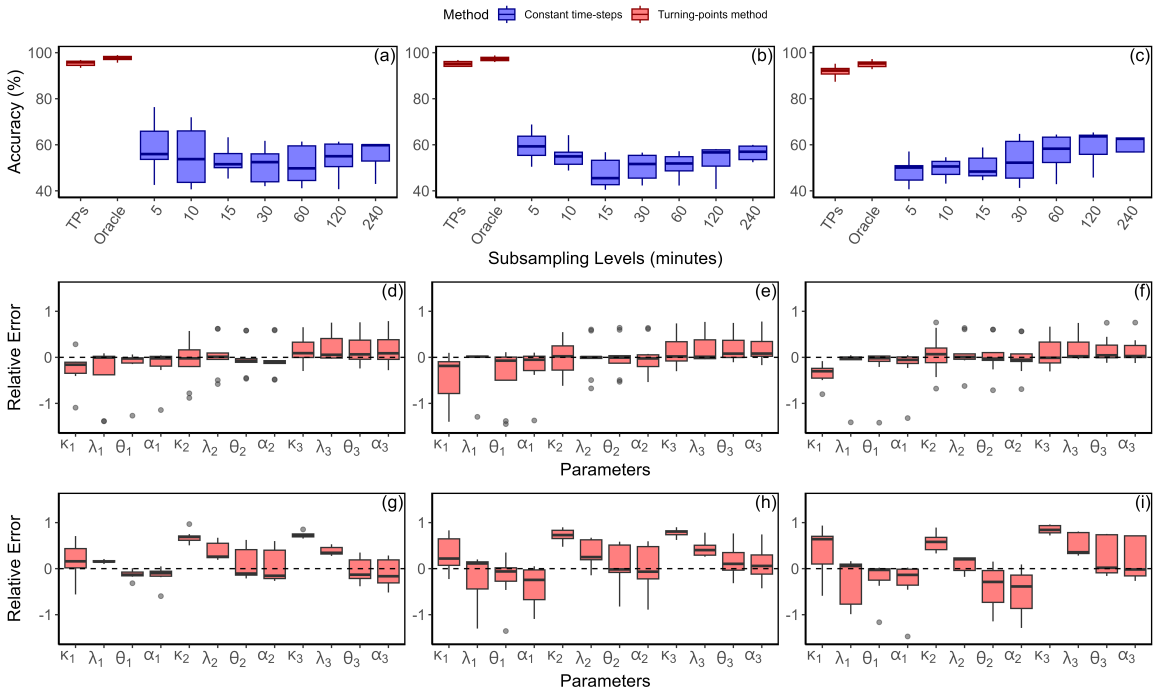


Figure 4.9: Results of the simulation analysis for the three-state model. The top row shows classification accuracy. Red boxplots represent results obtained using the turning-points (TPs) method under oracle settings, while blue boxplots correspond to regular subsampling at constant time steps. The second row shows relative errors in parameter estimation under oracle settings, and the third row shows relative errors under turning-point detection settings. Columns correspond to scenarios A, B, and C. All results assume turning angles follow a von Mises distribution, step times an exponential distribution, and step speeds a gamma distribution.

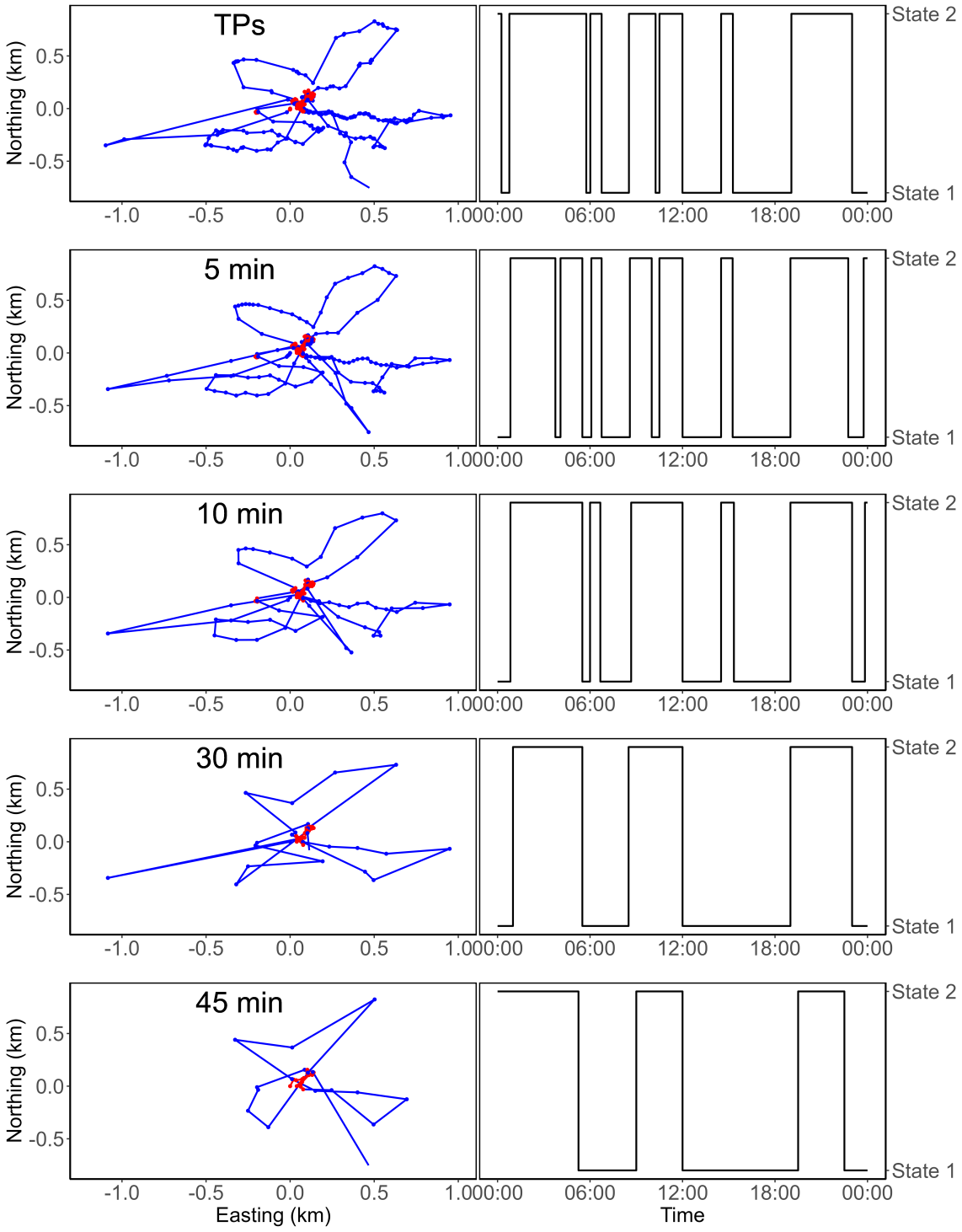


Figure 4.10: Comparison of the standard HMM with the proposed HMM-based approach, applied to one of the goats over a 1-day trajectory. The decoded state sequences are shown under different subsampling levels, where blue represents State 1 and red represents State 2. These results are based on the combined model.

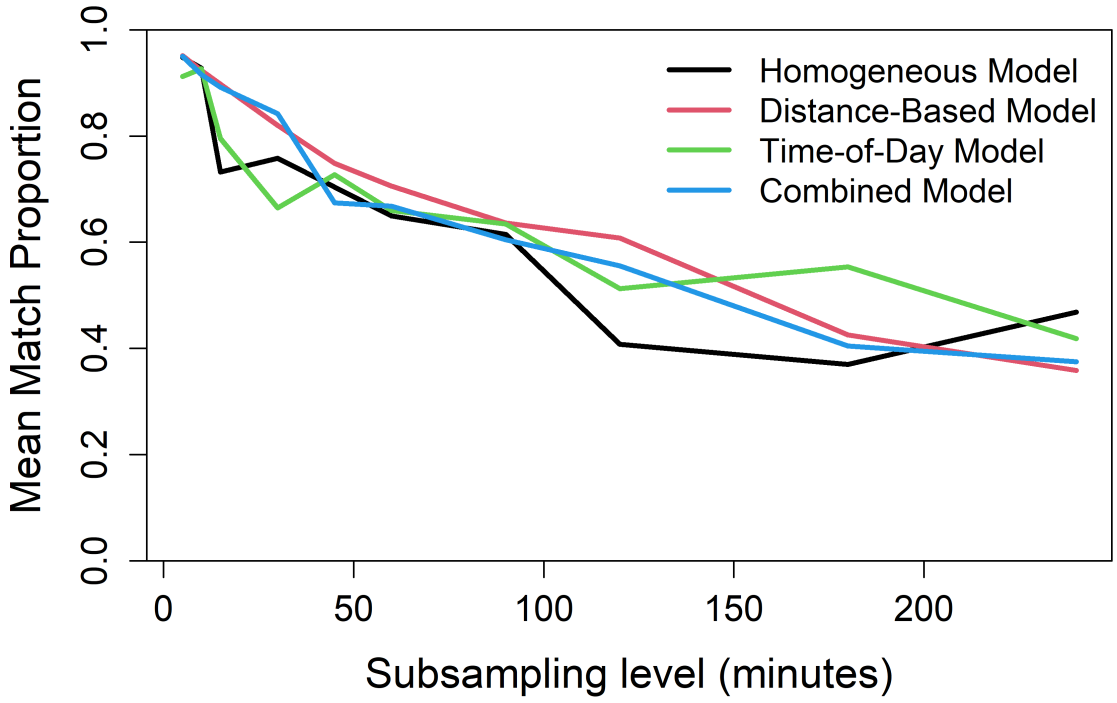


Figure 4.11: Mean matching scores across all goats, comparing the state sequences inferred by the proposed HMM with those from a standard HMM, evaluated at various subsampling intervals.

transition–probability surface for the combined model, in which both time of day and distance to the centre were included as covariates. The surface reveals an increased probability of switching from State 1 to State 2 during daytime and at greater distances from the centre, and a higher probability of switching from State 2 to State 1 during nighttime and when animals were closer to the centre (also see Figure 4.13). The HMMs indicated that goats spent just over half of their time (56% ± 6% SD) in State 2, with the remainder spent in State 1.

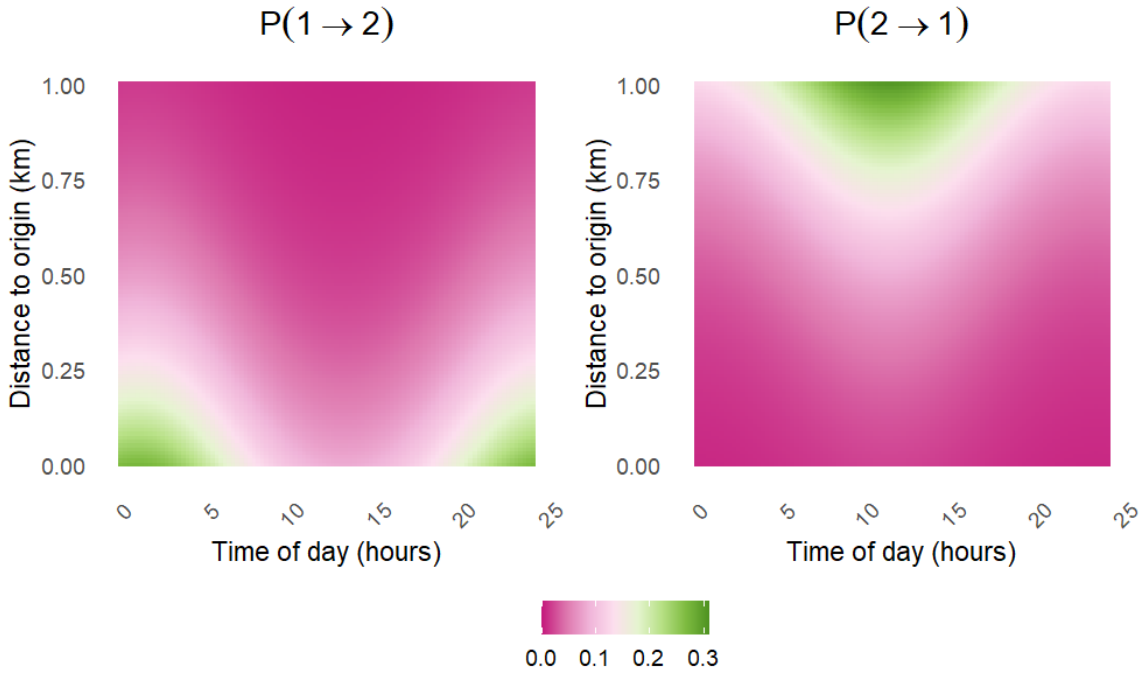


Figure 4.12: Estimated transition probabilities between behavioural states in the two-state HMM as functions of distance to the pen and time of day for an individual goat. The left column shows the probability of transitioning from State 1 (active) to State 2 (inactive), while the right column shows the reverse transition from State 2 to State 1.

Figure 4.13, the top row of narrow vertical panels one per goat depicts hourly state classification over a 24-hour cycle (00:00 at the bottom, 24:00 at the top), with red indicating the inactive state and blue indicating the active state. All goats exhibit prolonged inactivity during nighttime hours, extended bouts of activity during daylight,

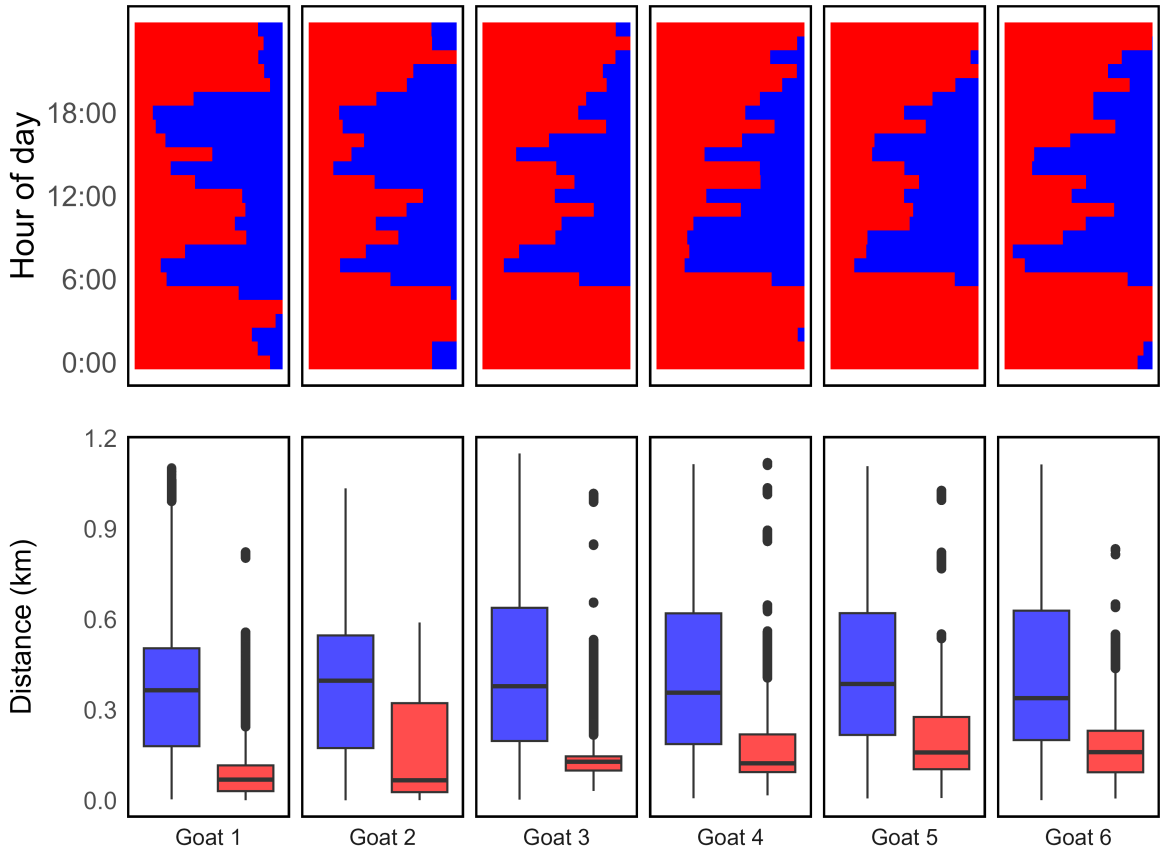


Figure 4.13: Post-hoc analysis of the distribution of State 1 and States 2 as functions of distance to the centre and time of day for all goats under the homogeneous model; red indicates the inactive state, while blue indicates the active state.

and a return to inactivity in the late afternoon or evening. Notably, goats 1 and 2 share very similar state distribution patterns, as do goats 3, 4, 5, and 6. The bottom row presents, for each goat, paired boxplots of distances from the centre by state. Active state distances have higher medians (approximately 0.3–0.6 km), greater variability, and more outliers reflecting excursions away from the centre whereas inactive state distances cluster near zero (medians around 0.1 km) with narrow interquartile ranges, indicating that goats remain close to the centre when inactive.

Figure 4.14 illustrates how state-to-state transition probabilities differ under the homogeneous model, the distance-from-centre model, and the time-of-day model. In the

homogeneous model, I estimated the per-turn switching probability in a post-hoc analysis. Across all models, the patterns are consistent among individuals: the probability of switching from State 1 to State 2 declines as animals move farther from the centre, whereas the probability of switching from State 2 to State 1 increases during daytime hours. Together, these results highlight a periodic interplay between foraging and resting cycles.

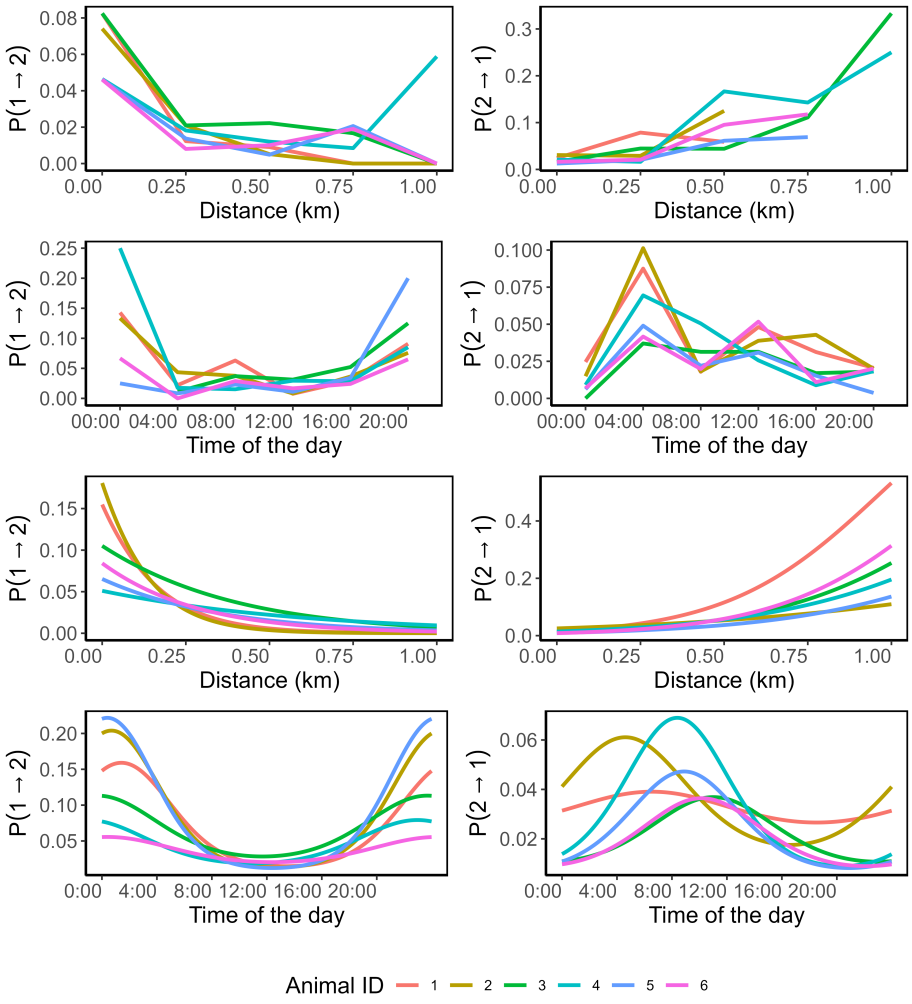


Figure 4.14: State-to-state transition probabilities under the homogeneous model, the distance-from-centre model, and the time-of-day model for the goat data. The left column shows transition probabilities for each individual from state 1 to state 2, the right column the transition probabilities from state 2 to state 1. The results for the homogeneous model, based on a post hoc inference on the effects of distance from centre and time of day, are shown in the first and second row, respectively. The third and fourth rows correspond to the estimated transition probabilities between behavioural states in the distance-based and time-of-day models, respectively, for each goat.

When turning angles were modelled using a von Mises distribution, step speeds using

gamma distributions, and step durations using an exponential distribution, the similarity in state classifications between this emission structure and the more complex emission setting remained remarkably high at 97%. This indicates that the behavioural state assignments produced under the two emission configurations are effectively indistinguishable.

Table 4.1 reports the ΔAIC values for the eight fitted models. Across all individuals, models using the first structure MA1-MA4 consistently achieved substantially lower ΔAIC values than their counterparts MB1-MB4, demonstrating clearly superior support from the data for each goat often with $\Delta\text{AIC} > 800$. The distance-based component was selected for most goats, except for ID 1, where the time-of-day model (MA3) provided the best fit. For three goats, the distance-based model (MA3) is preferred; for two goats, the combined model (MA4) performed best.

Table 4.1: ΔAIC values for the different models across goats (Section 4.1.4).

ID	MA1	MA2	MA3	MA4	MB1	MB2	MB3	MB4
1	16	18	0	3	1081	1085	1069	1076
2	31	0	27	6	864	832	860	839
3	18	0	20	4	1010	995	1016	998
4	8	1	5	0	1000	991	997	990
5	16	6	21	0	946	934	940	935
6	9	0	14	6	918	916	937	923

After aligning the time periods for which simultaneous recordings were available, I quantified the correspondence between the inferred states obtained across recording sessions. Individuals 3–6 which were recorded during the same respective periods, showed highly consistent state assignments across all pairwise comparisons, with matching accuracies ranging from 85–90%.

Emission parameter estimates for the selected two-state model are presented in Ta-

ble 4.2. The clearest separation between the two behavioural states is observed in the step-speed parameters (α_1, α_2): state 2 consistently exhibits a substantially higher gamma rate parameter than state 1, indicating faster movement speeds in state 2. A similar pattern is found in the step-time distribution, where both the rate and shape parameters show systematic differences between states across all individuals, further supporting a distinction between a slower, more tortuous state 1 and a faster, more directed state 2.

Differences between states are also evident in the turning-angle distributions for individuals 2–6, with state 1 generally characterised by higher concentration or lower truncation, indicating more directed movement relative to state 2. Overall, the parameter patterns are highly consistent across most goats, with state 1 corresponding to a less active, more tortuous movement mode, and state 2 reflecting a more active, faster, and more directed state.

ID	κ_1	κ_2	w_1	w_2	λ_1	λ_2	δ_1	δ_2	θ_1	θ_2	α_1	α_2	$\frac{\kappa_1}{\kappa_2}$	$\frac{w_1}{w_2}$	$\frac{\lambda_1}{\lambda_2}$	$\frac{\delta_1}{\delta_2}$	$\frac{\theta_1}{\theta_2}$	$\frac{\alpha_1}{\alpha_2}$
1	0.33638	1.3558	0.45931	9.11732	1.33502	0.85872	0.04097	0.00383	0.54	0.8	2.44	6.76	0.248	0.05	1.555	10.687	0.679	0.361
2	1.11119	0.3005	0.03519	4.79343	0.98472	0.75286	0.00791	0.00357	1.58	1.95	3.89	46.66	3.698	0.007	1.308	2.214	0.812	0.083
3	1.01852	0.78669	0.4866	2.29545	0.84454	0.65847	0.00885	0.00413	1.33	0.76	3.61	35.78	1.295	0.212	1.283	2.143	1.741	0.101
4	1.38432	0.18777	0.00276	14.67736	0.87484	0.58926	0.00978	0.0031	1.54	1.33	5.01	44.76	7.372	~ 0	1.485	3.157	1.162	0.112
5	1.21552	0.31355	0.04673	5.46109	0.93955	0.6492	0.00969	0.00423	1.61	1.69	4.47	53.12	3.877	0.009	1.447	2.294	0.953	0.084
6	0.57496	0.5237	1.92397	2.97158	0.95032	0.66109	0.0096	0.00384	1.19	1.6	3.13	72.75	1.098	0.647	1.438	2.498	0.742	0.043

Table 4.2: Emission parameter estimates for the models selected using ΔAIC and applied to the goat movement data.

Using a standard laptop equipped with an Intel Core i7–6700U CPU (2.40 GHz, 4 cores) and 32 GB of RAM, the full analysis of the goat data-including turning point estimation and HMM fitting takes approximately 4 seconds for a track of around half a million data points. This computation time corresponds to the homogeneous two-state model.

4.3 Summary

Hidden Markov Models (HMMs) have been widely used to model animal behaviour in recent years. These models were primarily developed under the assumption that the data are of low frequency, which aligns with earlier tracking technologies. However, with the advent of high-frequency tracking data, particularly from accelerometers and magnetometers, it is now possible to capture fine-scale animal movement. In such cases, classical HMMs are less applicable due to the minimal information between successive observations in high-frequency data and the substantial computational cost associated with processing such large volumes of data.

This study addresses a key question: how does incorporating high-frequency locational data into a behavioural model enhance analysis compared to using lower-frequency data? I have extended the HMM framework of Morales et al. (2004) and Langrock et al. (2012) to incorporate animal movement as a sequence of steps and turns, leveraging the algorithm of Alharbi et al. (2026) to identify places where the animal turned.

One challenge with this approach is that the turning points are not evenly spaced in time, unlike typical GPS data used for many HMM studies. I resolved this issue by incorporating the distribution of the time between two successive turns into the model. This adaptation allowed the model to bridge the gap between high-frequency movement data and HMM-based behavioural modelling, providing increased accuracy of behavioural state inference.

I evaluated the proposed methods using both simulated data and data from free-ranging

goats. I demonstrated that using turning points derived from high-frequency data gives much more accurate results than using regularly subsampled data (mimicking typical lower-frequency GPS data). In my analysis of goat data, I segmented the behavioural states of seven goats into two distinct states. I showed that these two state appear to be common across all individuals, even though fitted separately. All of them follow a specific state during the night when they are near a central area and switch to a more active state during the daytime when they are away from the central area.

In this contribution, I extended the standard step and turn Hidden Markov Models (HMMs) to better suit the characteristics of high-frequency animal movement data. Traditional implementations of HMMs, such as those in McClintock and Michelot (2018); Patterson et al. (2017); Langrock et al. (2012); Morales et al. (2004), typically assume that the observed locations are both informative and conditionally independent given the behavioural state, which limits their applicability for high-frequency data that are inherently temporally autocorrelated and computationally expensive to process. These assumptions limit their use for datasets that are massive, correlated and complex.

Importantly, the primary advantage of the proposed framework is not limited to improved classification accuracy. Rather, it lies in better alignment between the statistical model and the biological decision-making process, as well as improved estimation of transition dynamics and substantial reductions in computational cost when working with high-frequency data.

A few recent approaches have attempted to adapt HMMs to high-frequency data. For example, Conners et al. (2021) used summary statistics extracted from fixed-length windows (e.g., 30 seconds) of accelerometer and magnetometer data, while Rautiainen et al. (2022) applied a similar method using tri-axial accelerometer data. However, these window lengths are chosen arbitrarily and may not align with biologically meaningful behavioural transitions. In contrast, the proposed approach centres around the detection of turning points-points at which animals make explicit directional decisions-

as the primary behavioural events. Rather than summarising signals across time intervals, I segment the trajectory based on these detected decisions and model behaviour accordingly.

the proposed method also differs from recent extensions such as the autoregressive HMM (AR-HMM) proposed by Stoye et al. (2025), which incorporates within state autocorrelation by modelling step lengths and turning angles as autoregressive processes. While this offers a principled approach to handling autocorrelation in high-frequency data, it introduces greater model complexity, computational demands, and sensitivity to the choice of autoregressive order p. In contrast, the proposed method implicitly mitigates autocorrelation by aligning behavioural state transitions with actual movement decisions, inferred directly from the geometry of the path. This provides a simpler yet behaviourally meaningful alternative.

In many real-world scenarios, data are recorded at uneven time intervals due to limitations in tracking technology. Various methodologies have been developed to fit HMMs to such datasets. For instance, irregular spacing can be handled using multiple imputation techniques (McClintock, 2017) or continuous-time models (Blackwell et al., 2016; Grainger and Blackwell, 2025). These approaches are designed for situations where decisions or behaviours may occur between observations. However, if data are collected at a high enough frequency to extract behaviourally meaningful events directly, the proposed approach may be preferable, as it retains those advantages with significantly reduced computational complexity.

Selecting the appropriate number of states is beyond the focus of this study. However, the number of states is not estimated directly from the data using the HMM. In applied settings, it should be chosen based on expert knowledge, the specific research questions, and model checking, following the approach outlined by Pohle et al. (2017).

A further line of analysis could involve extending the model to examine how state transitions can be inferred when the sole data source is high-frequency magnetometer data,

as seen in Potts et al. (2018). In such cases, behavioural inference can be approached by omitting the speed component from the emission probability function and focusing instead on directional changes derived from heading estimates. This approach is particularly useful when speed information is unavailable or unreliable, while changes in heading still provide valuable insight into behavioural states.

Chapter 5

Discussion

This thesis adopts the idea that, rather than viewing movement as a sequence of regularly sampled locations, the focus should shift to behaviourally meaningful events an approach made possible by high frequency tracking. In particular, the turning points at which animals are likely to make navigational decisions (Potts et al., 2018; Gunner et al., 2024). This framework enables the development of models that are more biologically relevant for step and turn analyses compared to those based on arbitrarily subsampled locations. First, I developed new techniques for identifying turning events that requires minimal user input. The proposed method is statistically robust with clearly defined assumptions and properties, computationally efficient, with linear time complexity, and mathematically well grounded. It aligns more closely with the temporal scale of behavioural processes. I then demonstrate that the use of high frequency data, combined with the turning point technique, substantially improves the identification of behavioural states compared to traditional subsampling approaches.

In Chapter 2, I introduced a novel likelihood based approach for identifying turning points in high frequency heading data. Unlike existing methods (Potts et al., 2018; Codling and Plank, 2011; Turchin et al., 1991), the proposed algorithm avoids arbitrary user defined choices and instead derives optimal turning point sequences using a

statistically-grounded penalised likelihood framework. I showed that the method performs well across a wide range of simulated conditions, including different noise levels, noise structures, and turning frequencies. It also outperforms the sliding window algorithm of Potts et al. (2018), which itself was shown to outperform earlier approaches such as those of Codling and Plank (2011) and Turchin et al. (1991). Moreover, the method remains computationally efficient at scale, leveraging the Functional Pruning Optimal Partitioning (FPOP) algorithm to analyse datasets with up to one million points in only a few seconds. However, the method developed in Chapter 2 does not explicitly account for temporal dependence in the model residuals. To address this, I introduced an inflation of the penalty term based on the first lag autocorrelation. The simulation study, together with the analysis of 40 Hz oryx magnetometer data, demonstrated that this adjustment yields a robust procedure in practice, capable of identifying plausible turning points even in noisy empirical datasets.

In Chapter 3, I extended the framework proposed in Chapter 2 to account implicitly for temporal autocorrelation in heading data using a first lag angular autoregressive (angular AR(1)) model. High frequency data often exhibit strong autocorrelation, which can lead to spurious detection of behavioural events if ignored. I developed a method that robustly estimates both the noise autocorrelation and concentration parameters, alongside the turning points amount and locations. Simulation results demonstrated that accounting for temporal dependence improves detection accuracy. This extension further supports the adaptability of the proposed approach to the complex noise structures often found in real world biologging data.

Chapter 4 focused on integrating the output of the introduced turning point identification algorithm into a behavioural state modelling framework. Specifically, I modified the standard hidden Markov model (HMM) to operate on irregular sequences of turning points rather than on regularly spaced observations. This extension allowed us to decouple behavioural state inference from arbitrary sampling intervals and instead link it directly to behaviourally meaningful events. Through simulations and application

to 1Hz dead reckoned trajectories from goats, I showed that this approach achieves higher classification accuracy compared to standard HMMs.

Overall, these new methods progress movement ecology research in two ways. First, they provide objective, statistically-grounded ways to infer behaviourally meaningful turning points in high frequency animal movement data (Chapters 2 and 3). Second, they help improve behavioural state change identification, and their associated HMM algorithms, by leveraging these inferred turning points (Chapter 4). The proposed methods and results contribute to an ongoing effort to make use of high frequency data to improve ecological inference (Williams et al., 2020; Gunner et al., 2021; Redcliffe et al., 2025). More specifically, they advance recent efforts to ground analysis of movement data on a paradigm of rigorously inferred steps and turns(Potts et al., 2018; Munden et al., 2021; Gunner et al., 2024; Popp et al., 2025; Gunner et al., 2023), rather than the more artificial 'steps and turns' that arise as a by product of data gathering at lower frequencies (Avgar et al., 2016; Morales et al., 2004).

Although the methods in Chapters 2 and 3 are primarily motivated by ecological applications, they also hold potential for detecting change-points in angular data across other domains, such as changes in wind or wave direction (Mahrt, 2011; Klausner and Fattal, 2011). Thus, the proposed framework may be useful beyond ecology wherever identifying meaningful changes in directional time series is required.

Ecological Implications

Turning points in animal movement paths likely reflect behavioural decisions made by animals in response to their environment. For example, hilltopping butterfly repeatedly fly in straight lines between landing points before making distinct turns that structure their exploratory movements (Turchin et al., 1991). Similar patterns occur in fish, reptiles, and birds (Gunner et al., 2024). Likewise, mammals such as goats and oryx exhibit straight-line movement segments punctuated by discrete changes in direction

driven by environmental features (Munden et al., 2021; Gunner et al., 2023). There are, however, contexts in which external forces produce curved rather than straight line paths, making it essential to distinguish environmentally forced curvature from true behavioural turning decisions. For instance, wind drift can impose curved trajectories in soaring birds; water currents can bias the headings of marine animals such as sharks or turtles; and terrain curvature such as contour following on slopes, navigating river bends or moving around large obstacles can similarly induce smooth, non decision curvature (Gunner et al., 2024). In such cases, applying the turning point identification framework can still simplify movement paths by reducing the number of data points, but it remains important to consider and, where possible, correct for these external influences before interpreting detected turns as behavioural decisions.

Still, the goal of the proposed framework is to capture the structure of movement across a wide range of behaviours. In some contexts, movement is characterised by frequent turning at small spatial scales, such as the fine scale, tortuous paths of grazing herbivores (Schlecht et al., 2004). In other situations, turns may be far less frequent and interspersed with long, directed movement segments, for example when animals travel towards distant goals, as seen in the analysis in Chapter 4 and in studies of migratory (Nathan et al., 2008). There are also cases where animals turn at multiple spatial scales simultaneously, such as when navigating complex landscapes that contain barriers or obstacles (e.g., trees, rocks, fences, rivers), leading to nested or hierarchical patterns of movement decisions (Fryxell et al., 2008; Codling et al., 2008). Detecting turning points provides an essential first step in revealing these multi scale structures in the high frequency data, enabling researchers to distinguish between small scale tortuosity imposed by local conditions and large scale directional changes linked to broader behavioural goals like migratory.

Using turning points as the fundamental unit for behavioural analysis, rather than relying on regularly sampled locations, can provide more ecologically meaningful insights into spatial behaviour and resource use. For example, Munden et al. (2021) showed

that inference via step selection analysis improves markedly when movement is decomposed into behaviourally meaningful straight line steps separated by turning points, rather than by using arbitrarily sampled locations. Likewise, Gunner et al. (2024) demonstrated across more than 40 vertebrate species that animal movement paths are universally composed of straight segments punctuated by sharp turns, and that these turns represent decision points driven by sensory cues and environmental context. Indeed, the idea of using inferred turning points can be traced back to the foundational work of Turchin et al. (1991), who formalised how continuous, curved insect trajectories can be approximated by a series of straight line segments connected at turning points, thereby enabling quantitative behavioural inference. I also showed that behavioural state inference was significantly improved by using behaviourally meaningful turns rather than subsampled locations (see Chapter 4).

Also, behavioural changepoint analysis (BCPA) begins by reducing the movement path to a time series of summary statistics that capture key aspects of movement behaviour at a coarser temporal scale. For example, Gurarie et al. (2009) introduced an algorithm for detecting significant changes using persistence velocity and turning velocity, using these metrics to identify transitions between broad behavioural modes such as transiting, area-restricted search, and resting. However, such analyses were originally developed with relatively low-frequency tracking data in mind, where calculating summary statistics such as average speed or tortuosity over consecutive steps is informative, straightforward, and computationally stable.

Using turning points provides a way to infer specific summary statistics (i.e., step lengths, step duration and turning angles) from large, high-resolution datasets, thereby enabling existing behavioural changepoint analysis techniques to be applied to high-resolution movement data. I therefore anticipate that the output of the proposed algorithm could be used effectively as an input to BCPA and similar methods.

Methodological Limitations and Extensions

While the proposed methods offer several advantages, some limitations remain. First, although turning points are inferred as discrete events, the estimation process does not account for uncertainty in their timing and location. Future models could propagate this uncertainty into the inference, for example by integrating the method of Jewell et al. (2022), which was designed for testing changes in the mean following change point detection.

Second, I observed systematic bias in the estimates of the concentration (κ) and autocorrelation (ρ) parameters, which may affect subsequent inference when detecting turning points. Future work should quantify the extent to which these biases influence changepoint sensitivity particularly whether they cause underestimation or overestimation of the number of detected turns and develop appropriate correction strategies. I also note that, in some simulations, the estimated autocorrelation exceeded one, an outcome incompatible with a stationary angular AR(1) process. This limits the applicability of the current formulation for certain datasets and highlights the need for constrained estimation.

Third, although the AR(1) model improves robustness to autocorrelation, it assumes a first order dependence structure that may not fully capture the residual autocorrelation present in high frequency movement data. This limitation is reflected in the simulation experiments on misspecification of the underlying process: when data were generated from other ARMA models, performance deteriorated relative to the case where the data were truly generated from an angular AR(1) process. Therefore, more complex movement patterns may require higher order autoregressive models or modelling frameworks that allow for individual level heterogeneity.

Forth, in this study, I focused primarily on detecting turning points and using them to segment paths for behavioural state modelling. However, animals will not turn unless there is a beneficial reason for doing so, as turning incurs energetic costs that must be

offset to maximise net energy gain Wilson et al. (2013); Voigt and Holderied (2012); Gunner et al. (2024). Analysing the frequency, magnitude, and environmental context of turns could therefore provide deeper insights into how animals interact with their habitats and enhance our understanding of the decision-making process.

Finally, animal movement is also a hierarchical process, operating across multiple spatial and temporal scales Fryxell et al. (2008). Decisions occurring at fine scales may reflect short-term behavioural responses such as obstacle avoidance or micro-foraging movements, whereas larger-scale turning decisions may correspond to longer-term navigational goals, such as travelling between habitat patches, or shifting foraging areas. Future research should aim to classify different types of turns and to determine the scales at which they operate, particularly in relation to whether they represent short-term reactive adjustments or larger-scale strategic decisions. Disentangling these behavioural contexts will likely require the integration of additional data such as environmental covariates, landscape features, or resource distributions which could provide a more nuanced understanding of how animals process spatial information and adapt their movement strategies in dynamic environments.

Bibliography

R. P. Adams and D. J. MacKay. Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*, 2007.

A. F. Alharbi, P. G. Blackwell, A. Alagaili, N. C. Bennett, D. M. Scantlebury, and J. R. Potts. Efficient statistical inference of turning points in animal movement data. *Methods in Ecology and Evolution*, 17(3):889–902, 2026.

T. Avgar, J. R. Potts, M. A. Lewis, and M. S. Boyce. Integrated step selection analysis: bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution*, 7(5):619–630, 2016.

J. Bai. Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, 79(4):551–563, 1997.

J. D. Bailey, C. M. Benefer, R. P. Blackshaw, and E. A. Codling. Walking behaviour in the ground beetle, poecilus cupreus: dispersal potential, intermittency and individual variation. *Bulletin of Entomological Research*, 111(2):200–209, 2021a.

J. D. Bailey, A. J. King, E. A. Codling, A. M. Short, G. I. Johns, and I. Fürtbauer. "micropersonality" traits and their implications for behavioral and movement ecology research. *Ecology and Evolution*, 11(7):3264–3273, 2021b.

L. Bardwell, P. Fearnhead, I. A. Eckley, S. Smith, and M. Spott. Most recent changepoint detection in panel data. *Technometrics*, 61(1):88–98, 2019.

O. Bidder, J. Walker, M. Jones, M. Holton, P. Urge, D. Scantlebury, N. Marks,

E. Magowan, I. Maguire, and R. Wilson. Step by step: reconstruction of terrestrial animal movement paths by dead-reckoning. *Movement Ecology*, 3(1):1–16, 2015.

P. G. Blackwell. Bayesian inference for velocity-jump models for movement. *arXiv preprint 2509.21226*, 2025.

P. G. Blackwell, M. Niu, M. S. Lambert, and S. D. LaPoint. Exact Bayesian inference for animal movement in continuous time. *Methods in Ecology and Evolution*, 7(2): 184–195, 2016.

T. Boulinier, S. Kada, A. Ponchon, M. Dupraz, M. Dietrich, A. Gamble, V. Bourret, O. Duriez, R. Bazire, J. Tornos, et al. Migration, prospecting, dispersal? what host movement matters for infectious agent circulation? *Integrative and Comparative Biology*, 56(2):330–342, 2016.

P. Bovet and S. Benhamou. Spatial analysis of animals' movements using a correlated random walk model. *Journal of theoretical biology*, 131(4):419–433, 1988.

R. W. Byrne, R. Noser, L. Bates, and P. E. Jupp. How did they get here from there? detecting changes of direction in terrestrial ranging. *Animal Behaviour*, 77(3):619–631, 2009.

S. Chakar, E. Lebarbier, C. Lévy-Leduc, and S. Robin. A robust approach for estimating change-points in the mean of an ar(1) process. *Bernoulli*, 23(2):1408–1447, 2017. doi: 10.3150/15-BEJ781.

P. Chakravarty, M. Maalberg, G. Cozzi, A. Ozgul, and K. Aminian. Behavioural compass: animal behaviour recognition using magnetometers. *Movement Ecology*, 7 (1):28, 2019.

M. Chen, M. Xu, and P. Franti. A fast $o(n)$ multiresolution polygonal approximation algorithm for gps trajectory simplification. *IEEE Transactions on Image Processing*, 21(5):2770–2785, 2012.

E. A. Codling and M. J. Plank. Turn designation, sampling rate and the misidentifi-

cation of power laws in movement path data using maximum likelihood estimates. *Theoretical Ecology*, 4:397–406, 2011.

E. A. Codling, M. J. Plank, and S. Benhamou. Random walk models in biology. *Journal of the Royal society interface*, 5(25):813–834, 2008.

M. G. Conners, T. Michelot, E. I. Heywood, R. A. Orben, R. A. Phillips, A. L. Vyssotski, S. A. Shaffer, and L. H. Thorne. Hidden markov models identify major movement modes in accelerometer and magnetometer data from four albatross species. *Movement ecology*, 9(1):7, 2021.

S. J. Cooke, S. G. Hinch, M. Wikelski, R. D. Andrews, L. J. Kuchel, T. G. Wolcott, and P. J. Butler. Biotelemetry: a mechanistic approach to ecology. *Trends in Ecology & Evolution*, 19(6):334–343, 2004.

D. H. Douglas and T. K. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: the international journal for geographic information and geovisualization*, 10(2):112–122, 1973.

D. Eddelbuettel and R. François. Rcpp: Seamless r and c++ integration. *Journal of statistical software*, 40:1–18, 2011.

P. Fearnhead and G. Rigaill. Changepoint detection in the presence of outliers. *Journal of the American Statistical Association*, 114(525):169–183, 2019.

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

P. Fryzlewicz. Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, 42(6):2243–2281, 2014. ISSN 0090-5364.

E. Gombay. Change detection in autoregressive time series. *Journal of Multivariate Analysis*, 99(3):451–464, 2008.

D. Grainger and P. G. Blackwell. Finch: Fast statistical inference for continuous-time animal movement. *Methods in Ecology and Evolution*, 2025.

R. Gunner, R. Wilson, M. Lurgi, L. Borger, J. Redcliffe, E. Shepard, M. Holton, M. Crofoot, A. Alagaili, S. Andrzejaczek, et al. High resolution data reveal fundamental steps and turning points in animal movements. *Research Square*, 2024. doi: 10.21203/rs.3.rs-5559169/v1.

R. M. Gunner, M. D. Holton, D. M. Scantlebury, P. Hopkins, E. L. Shepard, A. J. Fell, B. Garde, F. Quintana, A. Gómez-Laich, K. Yoda, et al. How often should dead-reckoned animal movement paths be corrected for drift? *Animal Biotelemetry*, 9(1):43, 2021.

R. M. Gunner, R. P. Wilson, M. D. Holton, N. C. Bennett, A. N. Alagaili, M. F. Bertelsen, O. B. Mohammed, T. Wang, P. R. Manger, K. Ismael, et al. Examination of head versus body heading may help clarify the extent to which animal movement pathways are structured by environmental cues? *Movement Ecology*, 11(1):71, 2023.

P. R. Gupte, C. E. Beardsworth, O. Spiegel, E. Lourie, S. Toledo, R. Nathan, and A. I. Bijleveld. A guide to pre-processing high-throughput animal tracking data. *Journal of Animal Ecology*, 91(2):287–307, 2022.

E. Gurarie, R. D. Andrews, and K. L. Laidre. A novel method for identifying behavioural changes in animal movement data. *Ecology letters*, 12(5):395–408, 2009.

D. M. Hawkins, P. Qiu, and C. W. Kang. The changepoint model for statistical process control. *Journal of Quality Technology*, 35(4):355–366, 2003.

M. B. Hooten, D. S. Johnson, B. T. McClintock, and J. M. Morales. *Animal movement: statistical models for telemetry data*. CRC Press, 2017.

N. E. Humphries, H. Weimerskirch, and D. W. Sims. A new approach for objective

identification of turns and steps in organism movement data relevant to random walk modelling. *Methods in Ecology and Evolution*, 4(10):930–938, 2013.

B. Jackson, J. D. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Gioumousis, E. Gwin, P. Sangtrakulcharoen, L. Tan, and T. T. Tsai. An algorithm for optimal partitioning of data on an interval. *IEEE Signal Processing Letters*, 12(2):105–108, 2005.

S. R. Jammalamadaka and Y. R. Sarma. A correlation coefficient for angular variables. *Statistical Theory and Data Analysis II*, pages 349–364, 1988.

S. Jewell and D. Witten. Exact spike train inference via $\ell 0$ optimization. *The Annals of Applied Statistics*, 12(4):2457, 2018.

S. Jewell, P. Fearnhead, and D. Witten. Testing for a change in mean after changepoint detection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(4):1082–1104, 2022.

S. W. Jewell, T. D. Hocking, P. Fearnhead, and D. M. Witten. Fast nonconvex deconvolution of calcium imaging data. *Biostatistics*, 21(4):709–726, 2020.

T. Katzner and R. Arlettaz. Evaluating contributions of recent tracking-based animal movement ecology to conservation management. front ecol evol 24: 519, 2020.

R. Kays, M. C. Crofoot, W. Jetz, and M. Wikelski. Terrestrial animal tracking as an eye on life and planet. *Science*, 348(6240):aaa2478, 2015.

R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107 (500):1590–1598, 2012.

C.-J. Kim, J. C. Morley, and C. R. Nelson. The structural break in the equity premium. *Journal of Business & Economic Statistics*, 23(2):181–191, 2005.

Z. Klausner and E. Fattal. An objective and automatic method for identification of

pattern changes in wind direction time series. *International Journal of Climatology*, 31(5):783–790, 2011.

B. Koger, A. Deshpande, J. T. Kerby, J. M. Graving, B. R. Costelloe, and I. D. Couzin. Quantifying the movement, behaviour and environmental context of group-living animals using drones and computer vision. *Journal of Animal Ecology*, 2023.

G. Kurz, I. Gilitschenski, and U. D. Hanebeck. Efficient evaluation of the probability density function of a wrapped normal distribution. In *2014 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pages 1–5. IEEE, 2014.

R. Langrock, R. King, J. Matthiopoulos, L. Thomas, D. Fortin, and J. M. Morales. Flexible and practical modeling of animal telemetry data: hidden markov models and extensions. *Ecology*, 93(11):2336–2342, 2012.

M. Lavielle and E. Moulines. Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, 21(1):33–59, 2000.

C. Y. Lee and K. Lam. Survival analysis with change-points in covariate effects. *Statistical Methods in Medical Research*, 29(11):3235–3248, 2020.

K. P. Lewis, E. Vander Wal, and D. A. Fifield. Wildlife biology, big data, and reproducible research. *Wildlife Society Bulletin*, 42(1):172–179, 2018.

M. Li and B. M. Bolker. Incorporating periodic variability in hidden markov models for animal movement. *Movement ecology*, 5(1):1, 2017.

L. Mahrt. Surface wind direction variability. *Journal of Applied Meteorology and Climatology*, 50(1):144–152, 2011.

R. Maidstone, T. Hocking, G. Rigaill, and P. Fearnhead. On optimal multiple changepoint algorithms for large data. *Statistics and Computing*, 27:519–533, 2017.

B. T. McClintock. Incorporating telemetry error into hidden markov models of ani-

mal movement using multiple imputation. *Journal of Agricultural, Biological and Environmental Statistics*, 22(3):249–269, 2017.

B. T. McClintock and T. Michelot. momentuhmm: R package for generalized hidden markov models of animal movement. *Methods in Ecology and Evolution*, 9(6):1518–1530, 2018.

T. Michelot, R. Langrock, and T. A. Patterson. 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, 2016.

T. Michelot, N. J. Klappstein, J. R. Potts, and J. Fieberg. Understanding step selection analysis through numerical integration. *Methods in Ecology and Evolution*, 15(1): 24–35, 2024.

J. M. Morales, D. T. Haydon, J. Frair, K. E. Holsinger, and J. M. Fryxell. Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, 85(9):2436–2445, 2004.

V. M. Muggeo and G. Adelfio. Efficient change point detection for genomic sequences of continuous measurements. *Bioinformatics*, 27(2):161–166, 2011.

R. Munden, L. Börger, R. P. Wilson, J. Redcliffe, A. Loison, M. Garel, and J. R. Potts. Making sense of ultrahigh-resolution movement data: A new algorithm for inferring sites of interest. *Ecology and Evolution*, 9(1):265–274, 2019.

R. Munden, L. Börger, R. P. Wilson, J. Redcliffe, R. Brown, M. Garel, and J. R. Potts. Why did the animal turn? time-varying step selection analysis for inference between observed turning-points in high frequency data. *Methods in Ecology and Evolution*, 12(5):921–932, 2021.

R. Nathan, W. M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, and P. E. Smouse. A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, 105(49):19052–19059, 2008.

R. Nathan, C. T. Monk, R. Arlinghaus, T. Adam, J. Alós, M. Assaf, H. Baktoft, C. E. Beardsworth, M. G. Bertram, A. I. Bijleveld, et al. Big-data approaches lead to an increased understanding of the ecology of animal movement. *Science*, 375(6582): eabg1780, 2022.

E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

E. S. Page. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4):523–527, 1955.

A. Parton and P. G. Blackwell. Bayesian inference for multistate 'step and turn'animal movement in continuous time. *Journal of Agricultural, Biological and Environmental Statistics*, 22(3):373–392, 2017.

T. A. Patterson, A. Parton, R. Langrock, P. G. Blackwell, L. Thomas, and R. King. Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis*, 101: 399–438, 2017.

J. Pohle, R. Langrock, F. M. Van Beest, and N. M. Schmidt. Selecting the number of states in hidden markov models: pragmatic solutions illustrated using animal movement. *Journal of Agricultural, Biological and Environmental Statistics*, 22:270–293, 2017.

S. Popp, E. A. Codling, J. D. Bailey, and A. Dornhaus. A comparison of turn identification methods on high-frequency movement trajectories reveals potential comparability issues between studies of movement ecology. *bioRxiv*, pages 2025–04, 2025.

J. R. Potts, L. Börger, D. M. Scantlebury, N. C. Bennett, A. Alagaili, and R. P. Wilson. Finding turning-points in ultra-high-resolution animal movement data. *Methods in Ecology and Evolution*, 9(10):2091–2101, 2018.

H. Rautiainen, M. Alam, P. G. Blackwell, and A. Skarin. Identification of reindeer

fine-scale foraging behaviour using tri-axial accelerometer data. *Movement Ecology*, 10(1):40, 2022.

J. Redcliffe, J. Boulerice, I. Namir, R. Wilson, W. J. McShea, and H. Shamon. Using dead-reckoning to track movements and map burrows of fossorial species. *Animal Biotelemetry*, 13(1):11, 2025.

V. Rijsbergen. Information retrieval; ; butterworth, 1978. *Journal of Librarianship*, 11:237, 1979.

G. Romano, G. Rigaill, V. Runge, and P. Fearnhead. Detecting abrupt changes in the presence of local fluctuations and autocorrelated noise. *Journal of the American Statistical Association*, 117(540):2147–2162, 2022.

E. Schlecht, C. Hülsebusch, F. Mahler, and K. Becker. The use of differentially corrected global positioning system to monitor activities of cattle at pasture. *Applied Animal Behaviour Science*, 85(3-4):185–202, 2004.

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, pages 461–464, 1978.

A. J. Scott and M. Knott. A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, pages 507–512, 1974.

J. E. Sedlak. Improved spacecraft attitude filter using a sequentially correlated magnetometer noise model. In *16th DASC. AIAA/IEEE Digital Avionics Systems Conference. Reflections to the Future. Proceedings*, volume 2, pages 8–4. IEEE, 1997.

F. V. Stoye, A. Hoyer, and R. Langrock. Autoregressive hidden markov models for high-resolution animal movement data: Fv stoye et al. *Journal of Agricultural, Biological and Environmental Statistics*, pages 1–16, 2025.

R. C. Team. Ra language and environment for statistical computing, r foundation for statistical. *Computing*, 2020.

S. M. Tomkiewicz, M. R. Fuller, J. G. Kie, and K. C. Bates. Global positioning system and associated technologies in animal behaviour and ecological research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1550):2163–2176, 2010.

C. Truong, L. Oudre, and N. Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.

P. Turchin. *Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants*. Massachusets, Estados Unidos: Sinauer Associates,, 1998.

P. Turchin, F. Odendaal, and M. Rausher. Quantifying insect movement in the field. *Environmental Entomology*, 20(4):955–963, 1991.

M. Visvalingam and J. D. Whyatt. Line generalization by repeated elimination of points. *The Cartographic Journal*, 30(1):46–51, 1993. doi: 10.1179/000870493786962263. URL https://www.tandfonline.com/doi/abs/10.1179/000870493786962263.

C. C. Voigt and M. W. Holderied. High manoeuvring costs force narrow-winged molossid bats to forage in open space. *Journal of Comparative Physiology B*, 182(3): 415–424, 2012.

X. Wang, B. Liu, X. Zhang, Y. Liu, and the Alzheimer's Disease Neuroimaging Initiative. Efficient multiple change point detection for high-dimensional generalized linear models. *Canadian Journal of Statistics*, 51(2):596–629, 2023.

K. Ward, G. Dilillo, I. Eckley, and P. Fearnhead. Poisson-focus: An efficient online method for detecting count bursts with application to gamma ray burst detection. *Journal of the American Statistical Association*, 120(549):7–19, 2025.

H. J. Williams, L. A. Taylor, S. Benhamou, A. I. Bijleveld, T. A. Clay, S. de Grissac, U. Demšar, H. M. English, N. Franconi, A. Gómez-Laich, et al. Optimizing the

use of biologgers for movement ecology research. *Journal of Animal Ecology*, 89(1): 186–206, 2020.

R. Wilson, I. Griffiths, P. Legg, M. Friswell, O. Bidder, L. Halsey, S. A. Lambertucci, and E. Shepard. Turn costs change the value of animal search paths. *Ecology letters*, 16(9):1145–1150, 2013.

R. P. Wilson, E. Shepard, and N. Liebsch. Prying into the intimate details of animal lives: use of a daily diary on animals. *Endangered species research*, 4(1-2):123–137, 2008.

D. W. Wolfson, D. E. Andersen, and J. R. Fieberg. Using piecewise regression to identify biological phenomena in biotelemetry datasets. *Journal of Animal Ecology*, 91(9):1755–1769, 2022.

W. Zucchini and I. L. MacDonald. *Hidden Markov models for time series: an introduction using R*. Chapman and Hall/CRC, 2009.