

Bayesian inference for continuous-time step-and-turn movement models



The
University
Of
Sheffield.

Alison Parton

School of Mathematics and Statistics
Faculty of Science
The University of Sheffield

This dissertation is submitted for the degree of
Doctor of Philosophy

April 2018

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This thesis is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and the following acknowledgements.

Alison Parton

April 2018

Acknowledgements

I would like to thank the School of Mathematics and Statistics at the University of Sheffield for funding this PhD. Thanks to Professors Jeremy Oakley and Caitlin Buck for your inspiration and encouragement in pursuing an academic career. Thank you so much to my supervisor Professor Paul Blackwell for your expertise, guidance and infinite patience—I've thoroughly enjoyed my time skulking at your office door.

Chap. 2 features an extension to the edited contribution in the review Patterson et al. (2017). Thank you to my co-authors (Toby Patterson, Roland Langrock, Len Thomas and Ruth King), David Borchers and two anonymous referees on their useful feedback on this material. A proportion of the work in Chaps. 3 and 5 is an extended description of Parton et al. (2017), in particular the reindeer example of Sect. 5.3. Thank you to my co-author, Anna Skarin, for providing the reindeer data and feedback on this approach. The model framework and elk example featured in Chap. 4 are an expanded version of Parton and Blackwell (2017). Thanks to Théo Michelot and two anonymous referees for their comments on this work.

I am grateful to the many staff and students within the department and wider university who have supported me in pub locations across the city. Thanks to the Old House and the Notty for all the pies we have consumed and the f.r.i.e.n.d.s who forgive my empty promises of top-tier keg beer offerings.

Thank you to Adam for your constant support and (almost) never-ending ability to appear engrossed in yet another talk rehearsal.

Abstract

This thesis concerns the statistical modelling of animal movement paths given observed GPS locations. With observations being in discrete time, mechanistic models of movement are often formulated as such. This popularity remains despite an inability to compare analyses through scale invariance and common problems handling irregularly timed observations. A natural solution is to formulate in continuous time, yet uptake of this has been slow, often excused by a difficulty in interpreting the ‘instantaneous’ parameters associated with a continuous-time model.

The aim here was to bolster usage by developing a continuous-time model with interpretable parameters, similar to those of popular discrete-time models that use turning angles and step lengths to describe the movement process. Movement is defined by a continuous-time, joint bearing and speed process, the parameters of which are dependent on a continuous-time behavioural switching process, thus creating a flexible class of movement models. Further, we allow for the observed locations derived from this process to have unknown error. Markov chain Monte Carlo inference is presented for parameters given irregular, noisy observations. The approach involves augmenting the observed locations with a reconstruction of the underlying continuous-time process.

Example implementations showcasing this method are given featuring simulated and real datasets. Data from elk (*Cervus elaphus*), which have previously been modelled in discrete time, demonstrate the interpretable nature of the model, finding clear differences in behaviour over time and insights into short-term behaviour that could not have been obtained in discrete time. Observations from reindeer (*Rangifer tarandus*) reveal the effect observation error has on the identification of large turning angles—a feature often inferred in discrete-time modelling. Scalability to realistically large datasets is shown for lesser black-backed gull (*Larus fuscus*) data.

Table of contents

List of figures	xi
List of tables	xv
Nomenclature	xix
1 Introduction	1
1.1 Movement data	2
1.2 Discrete-time step-and-turn movement	4
1.3 Problems with discrete-time modelling	11
1.4 Aims of this thesis	12
2 Review of continuous-time modelling	15
2.1 Modelling movement with diffusion processes	15
2.2 Modelling movement with general SDEs	20
2.3 Modelling switching behaviour	21
2.4 Discussion	28
3 Single state movement	31
3.1 Analogue of discrete-time movement	31
3.2 Simulating movement	33
3.3 Fully Bayesian inference	35
3.4 Movement with correlated speed	45
3.5 Examples with simulated data	50
3.6 Discussion	72
4 Multistate movement	77
4.1 Model for multistate switching	77
4.2 Simulating movement	78

4.3	Extending the method for fully Bayesian inference	82
4.4	Simulated example	93
4.5	Two-state movement in elk	109
4.6	Discussion	119
5	Incorporation of observation error	123
5.1	Independent model for observation error	123
5.2	Extending the method for fully Bayesian inference	124
5.3	Noisy observations of single state reindeer movement	129
5.4	Noisy observations of two-state gull movement	139
5.5	Simulated example	146
5.6	Correlated error process	160
5.7	Discussion	164
6	Discussion and further work	167
6.1	Extending the model for behaviour	168
6.2	Identifiability in the presence of observation error	168
6.3	Efficiency of the inference algorithm	171
6.4	Comparison to discrete-time step-and-turn	172
	References	177
	Appendix A Additional details	187
A.1	Derivation of Gibbs samplers	187
A.2	Conditioning by Kriging	188
	Appendix B Additional figures	191
B.1	Single state independent steps simulation	191
B.2	Single state correlated steps simulation	194
B.3	Two-state simulation	198
B.4	Two-state movement in elk	201
B.5	Noisy observations of single state reindeer movement	204
B.6	Noisy observations of two-state gull movement	207
B.7	Observation error simulation example	211

List of figures

1.1	Parametrisation of movement in the discrete-time step-and-turn model . . .	4
1.2	Structure of a state-space model	8
1.3	Structure of a hidden Markov model	9
3.1	Single state simulated paths	35
3.2	Observations augmented with a single state refined path	36
3.3	DAG of single behaviour movement model	37
3.4	Section of the full single state refined path to update over.	41
3.5	Simulated path and observations (independent step model)	51
3.6	Initial path locations (independent step model)	53
3.7	Initial bearings and speeds (independent step model)	54
3.8	Sampled bearing variance against mean speed (independent step model) . .	56
3.9	Sampled speed mean against variance (independent step model)	57
3.10	Posterior credible intervals of parameters (independent step model)	59
3.11	Detailed path reconstructions (independent step model)	61
3.12	Full path reconstructions (independent step model)	62
3.13	Simulated path and observations (correlated step model)	63
3.14	Initial path locations (correlated step model)	65
3.15	Initial bearings and speeds (correlated step model)	66
3.16	Sampled bearing variance against mean speed (correlated step model) . . .	67
3.17	Sampled speed correlation against variance (correlated step model)	68
3.18	Posterior credible intervals of parameters (correlated step model)	70
3.19	Full path reconstructions (correlated step model)	73
3.20	Detailed path reconstructions (correlated step model)	74
4.1	Multistate simulated bearings and speeds	79
4.2	Multistate simulated locations	80
4.3	Observations augmented with a multistate refined path	83

4.4	DAG of multistate model	83
4.5	Section of the full multistate refined path to update over.	88
4.6	Simulated path and observations (two-state simulation)	95
4.7	Simulated bearings, speeds and behaviours (two-state simulation)	96
4.8	Initial configuration of bearings, speeds and behaviours (two-state simulation)	98
4.9	Sampled behavioural parameters (two-state simulation)	99
4.10	Sampled movement parameters (two-state simulation)	100
4.11	Parameter kernel density estimates (two-state simulation)	101
4.12	Short-term speed variance kernel density estimates (two-state simulation) .	103
4.13	Posterior behavioural state probabilities (two-state simulation)	104
4.14	Path reconstruction with high certainty (two-state simulation)	106
4.15	Uncertain path reconstruction (two-state simulation)	107
4.16	Path reconstruction with behavioural misclassification (two-state simulation)	108
4.17	Observations (two state elk)	110
4.18	Initial path (two-state elk)	112
4.19	Samples of the behavioural parameters (two-state elk)	113
4.20	Samples of the movement parameters (two-state elk)	115
4.21	Probability of residency in behavioural states (two-state elk)	117
4.22	Behavioural residency time density (two-state elk)	118
4.23	Path reconstructions (two-state elk)	119
4.24	Detailed section of path reconstructions (two-state elk)	120
5.1	Noisy observations augmented with a single state refined path	124
5.2	DAG of movement model with error	125
5.3	Section of the refined path to update over when observation error is present.	127
5.4	Observations (noisy single state reindeer)	130
5.5	Samples of the movement parameters (noisy single state reindeer)	133
5.6	Posterior densities of the movement parameters (noisy single state reindeer)	134
5.7	Path reconstructions (noisy single state reindeer)	136
5.8	Sampled observation error (noisy single state reindeer)	137
5.9	Detailed path reconstructions (noisy single state reindeer)	138
5.10	Observations (noisy two-state gull)	140
5.11	Samples of the behavioural parameters (noisy two-state gull)	142
5.12	Posterior behavioural state probabilities (noisy two-state gull)	143
5.13	Samples of the movement parameters (noisy two-state gull)	144
5.14	Error variance density (noisy two-state gull)	146
5.15	Simulated path and observations (observation error simulation)	147

5.16	Samples of the movement parameters (1) (observation error simulation) . . .	150
5.17	Samples of the movement parameters (2) (observation error simulation) . . .	151
5.18	Posterior density of the movement parameters (observation error simulation)	152
5.19	Posterior density of the movement parameters (observation error simulation)	153
5.20	Path reconstructions (observation error simulation)	156
5.21	True location reconstruction with small error (observation error simulation)	158
5.22	True location reconstruction with large error (1) (observation error simulation)	159
5.23	True location reconstruction with large error (2) (observation error simulation)	161
5.24	Section of the refined path to update over with correlated observation error .	163
6.1	Observed heterogeneity in reindeer movement	170
6.2	Low signal-to-noise observations of reindeer movement	170
6.3	Returns to the same site in gull movement	175
B.1	Posterior densities of movement parameters (independent step model) . . .	191
B.2	Movement parameter trace (independent step model)	192
B.3	Movement parameter ACF (independent step model)	193
B.4	Speed correlation trace at sampling time scale (correlated step model) . . .	194
B.5	Movement parameter trace (correlated step model)	195
B.6	Movement parameter ACF (corr step model)	196
B.7	Posterior densities of movement parameters (correlated step model)	197
B.8	Initial path (two-state simulation)	198
B.9	Parameter sample trace (two state simulation)	199
B.10	Short-term speed variance sample trace (two-state simulation)	200
B.11	Initial path as bearings and speeds (two-state elk)	201
B.12	Parameter trace (two-state elk)	202
B.13	Posterior densities of parameters (two-state elk)	203
B.14	Initial bearing and speeds (noisy single state reindeer)	204
B.15	Parameter trace (noisy single state reindeer)	205
B.16	Parameter ACF (noisy single state reindeer)	206
B.17	Behaviour parameter trace (noisy two-state gull)	207
B.18	Posterior densities of behaviour parameters (noisy two-state gull)	208
B.19	Movement parameter trace (noisy two-state gull)	209
B.20	Posterior densities of movement parameters (noisy two-state gull)	210
B.21	Sample trace of the movement parameters (observation error simulation) . .	211
B.22	Posterior distribution of movement parameters (observation error simulation)	212

List of tables

3.1	Initial parameters, perturbation variances and acceptance rates (independent step model)	52
3.2	Posterior credible intervals of movement parameters (independent step model)	60
3.3	Initial parameters, perturbation variances and acceptance rates (correlated step model)	64
3.4	Posterior credible intervals of movement parameters (correlated step model)	71
4.1	Posterior credible intervals of behavioural parameters (two-state simulation)	100
4.2	Posterior credible intervals of movement parameters (two-state simulation)	103
4.3	Posterior credible intervals of behavioural parameters (two-state elk)	114
4.4	Posterior credible intervals of movement parameters (two-state elk)	116
5.1	Posterior credible intervals of parameters (noisy single state reindeer) . . .	135
5.2	Posterior credible intervals of behavioural parameters (noisy two-state gull)	142
5.3	Posterior credible intervals of movement parameters (noisy two-state gull) .	145
5.4	Posterior credible intervals of the observation error variance (noisy two-state gull)	145
5.5	Posterior credible intervals of parameters (observation error simulation) . .	154
5.6	Posterior credible intervals of parameters (observation error simulation) . .	155

List of Algorithms

1	Simulating single state movement.	34
2	Sample conditional bearing process parameter	39
3	Sample conditional distance process parameters	40
4	Simulate bearing proposal	42
5	Simulate step proposal	44
6	Correlated step bridge distribution	49
7	Simulating multistate movement.	81
8	Sample conditional behavioural process parameters	85

Nomenclature

List of Abbreviations

ACF	Autocorrelation function
BB	Brownian bridge
BM	Brownian motion
BRW	Biased random walk
CRW	Correlated random walk
CTMC	Continuous-time Markov chain
DAG	Directed acyclic graph
DTMC	Discrete-time Markov chain
ESS	Effective sample size
GPS	Global Positioning System
HMM	Hidden Markov model
iid	Independent and identically distributed
MCMC	Markov chain Monte Carlo
MH	Metropolis-Hastings
OU	Ornstein-Uhlenbeck

PMCMC Particle Markov chain Monte Carlo

SDE Stochastic differential equation

SSM State-space model

SVF Semivariance function

VHF Very high frequency

Chapter 1

Introduction

Movement is fundamental to an animal's survival, yet the details of such a process are almost unknown. Wildlife trajectories are difficult to interpret, being a complex, noisy mixture of decisions that regularly exhibit randomness, non-linearity, and high spatial and temporal correlations. Internal behavioural states, physical constraints and memory regulate movement, whilst encounters with the environment, such as landscape, weather and other individuals, influence this decision-making process (Cagnacci et al., 2010).

Advances in tracking technologies have allowed data collection on individual animal movement at increasing precision and frequency. This has led the way for growth of research into movement ecology, concerned with questions of patterns in animal movements. Phenomena of interest include: the underlying mechanisms and causes for animals to move through space, the constraints affecting this movement (including internal and external influencing factors), and how these movements shape the animal's overall ecology. Although the focus here will be on the analysis of trajectories of individual animals, such research has the potential to impact other facets of movement ecology, such as: home-range analysis, resource use/selection and group movement.

The ability to collect observations at increased sampling frequencies has particularly steered research to study movement in the short-term, motivating the study of behaviours. Although real movement behaviour is highly complex and dynamic, much of the research in this area assumes that movement is driven by switches between 'behavioural modes' that allow for differing phases of the trajectory. The growing interest in this area involves identifying these "statistically detectable signatures" (Fleming et al., 2014a). Driving questions include the number of behavioural modes present in a trajectory, when/how often transitions between

these modes occur, and the characterisation of the underlying movement each behaviour represents.

Although the yearly number of publications on animal movement has doubled over the last 10 years, this remains dominated by the documentation of studies rather than addressing ecological questions (Holyoak et al., 2008). Large datasets and limiting computational power lead to a constant trade-off between models that are ecologically realistic and those that are feasible to implement. Overly complex models may capture the realism of individual movement but lack the machinery required to fit them and remain inaccessible to ecologists. In contrast, simplistic models are often employed in an attempt to avoid these complexities, ignoring the directionality and correlation present in movement (Brownian motion and Levy flights, for example (Pyke, 2015)). These single parameter models cannot describe the complex nature of movement, yet often make strong conclusions about it. There is a requirement for approaches that are able to capture enough realism of trajectories to address ecological questions on movement, whilst maintaining statistical robustness and interpretability.

1.1 Movement data

Movement data generally consists of location fixes of an animal (or group of animals) over a sequence of discrete points in time. For land-based animals such fixes are in the two-dimensional horizontal plane. Observations of aerial and marine animals are commonly collected in either the horizontal plane or the one-dimensional vertical direction, and only rarely in the full three dimensions. A variety of collection methods exist for movement observations, but Global Positioning System (GPS) will be the focus of the following work as it is the predominant collection method for modern studies. Historically, collection was by very high frequency (VHF) radio (Cagnacci et al., 2010; Hebblewhite and Haydon, 2010), and other collection methods for movement (direct or indirect) such as camera traps, accelerometers and magnetometers exist, but are not discussed further here (see evaluations in Cooke et al. (2004, 2013); Rutz and Hays (2009); Wilmers et al. (2015)).

The sampling scheme of movement observations vary considerably from sub-second to multiple days based on the particular focus of study and battery size/life. Sampling frequency affects the questions explorable from a set of observations; if migration is of interest then daily observations over multiple years may be sufficient whilst if foraging periods are of concern then observations may need to be at the short scale, such as minutes. It is important

that the sampling scheme is at a meaningful temporal scale with regards the animal dynamics being explored.

The common feature that sampling schemes are often irregular complicates statistical modelling of animal movement. A number of observations may be missing from the regular sequence of times either randomly or with structure through temporal or locational constraints. Small irregularities in sampling times are introduced when a measuring device attempts regular fixes but is delayed through either sensor, battery or memory limitations. Structured irregularities in observations may provide information on movement; such as the times at which seals are underwater making GPS fixes impossible.

A feature of movement data is the autocorrelation in the observed process—an animal's location in the near future will depend heavily on its current location. At high sampling frequencies this correlation is particularly strong and needs consideration when modelling movement (Cushman, 2010). It is undesirable to apply the informal, ad-hoc methods of some studies which remove observations until autocorrelation is assumed negligible (Nations and Anderson-Sprecher, 2006) or inflate parameter error ranges post-analysis (Fieberg et al., 2010); instead we believe robust techniques that incorporate correlation should be favoured.

Telemetry observations introduce spatial and temporal error in the true location of the animal. Due to the nature of GPS technology, the level of observation error is closely linked to the animal's environment and this introduces autocorrelation in the error process over time. For GPS devices, such errors are considered to be small at 10–28 m (Frair et al., 2010), but the level of effect this has on inference will depend on the movement scale of the animal in question. Further, the error structures obtained from some technologies (e.g. Argos, see McClintock et al. (2015)) are known to be complex and non-Gaussian.

After the collection of animal locations, movement modelling may be carried out based on a number of movement metrics (see e.g. Calenge et al. (2009)), including;

- the raw locations themselves,
- the increments in locations (as displacement or 'velocity'),
- the Euclidean distances between two consecutive locations ('steps'),
- the compass direction between two consecutive locations ('bearings'), and
- the change in direction between three consecutive locations ('turns').

Note that when considering steps and either bearings or turns together, such a bivariate process (given an initial location) defines the raw location process. This alternative parametrisation of movement has been modelled since the early days of telemetry (Marsh and Jones, 1988; Siniff

and Jessen, 1969) and has proved popular, described as the “intuitive approach” (McClintock et al., 2014). The development of modelling animal movements based on this metric is therefore outlined in the following section.

1.2 Discrete-time step-and-turn movement

Probably the earliest example of an approach to movement modelling that has become well-established by ecologists is the step-and-turn model. The characterisation of movement into a bivariate time series of turning angles and steps lengths was first used by Siniff and Jessen (1969) to gain insight into the movement of hares and foxes. The turning angle, ϕ , is the angle between three consecutive observed locations and the step length, r , is the straight line distance between two consecutive observed locations—Fig. 1.1 shows this parametrisation. Siniff and Jessen (1969), and later models building upon this, propose parametric distributions for these two variables so that the likelihood given observed locations $\{Z_1, \dots, Z_M\}$ is

$$p(\{Z_1, \dots, Z_M\} | \Omega) = \prod_{i=1}^{M-2} \Phi(\phi_i | \Omega_\phi) \prod_{i=1}^{M-1} R(r_i | \Omega_r), \quad (1.1)$$

where Φ and R denote the parametric distributions of the turning angles and steps, respectively, and Ω is the set of parameters defining such distributions. Note that for a set of M observed locations, there are $M - 1$ derived step lengths and $M - 2$ derived turning angles. Statistical inference is then concerned with learning about the parameters Ω , given observations $\{Z_1, \dots, Z_M\}$. Note from Eq. 1.1 that there is an assumption that the step and turn processes are independent from one another.

The distance an animal can travel between two locations is constrained to be positive, and so it is assumed that the step length, r , arises from some positive parametric distribution;

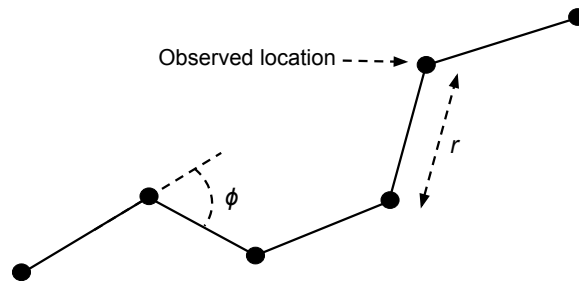


Fig. 1.1 Parametrisation of movement in the discrete-time step-and-turn model. Define movement between observed locations by the turning angle, ϕ , and the step length, r .

common choices for R in Eq. 1.1 include the gamma and Weibull distributions. The turn that an animal can achieve between two points in time is unconstrained, but is constrained to $[-\pi, \pi]$ when observed. The turns, ϕ , are therefore assumed to arise from a wrapped distribution with such boundaries; options for Φ in Eq. 1.1 include the von Mises (closely related to the wrapped normal) and the wrapped Cauchy distributions. Often, the underlying movement is assumed to be a correlated random walk (CRW), so that the turning angle distribution is centred at zero (Kareiva and Shigesada, 1983) and the animal is most likely to keep moving in the same direction over a short period of time.

The step-and-turn movement model requires a regular sampling frequency. Proposed ways to deal with irregularity are ad-hoc, including thinning and interpolation (Edelhoff et al., 2016). Although it may be feasible to assume the animal exhibits a constant speed between consecutive locations, making the step length scalable, it is unclear how to apply a similar assumption with regards the turning angle over differing time lags.

1.2.1 Incorporating behavioural switching

The single state movement model above was first extended to include behavioural switching in Morales and Ellner (2002), who highlighted the need for multiple behavioural states and proposed a simple method where movement in beetles switched modes at a single, fixed time after their release. Flexible multistate switching in a statistical setting was then introduced in Morales et al. (2004) to incorporate the idea that animals exhibit a number of distinct movement ‘behaviours’ over time. In this case, a movement behaviour relates to movement following the single state model of Eq. 1.1, but with a behaviour-specific set of parameters Ω_i (corresponding to behavioural state i) that govern the turning angle and step length distributions. Morales et al. (2004) use the Weibull and the wrapped Cauchy as the behaviour-specific distributions for the step lengths and turning angles, respectively. For example, a ‘foraging’ style behaviour may correspond to a step distribution with low mean whilst, in contrast, a ‘migratory’ style behaviour may have a high step mean.

The process by which the animal changes its behaviour is assumed to follow a discrete-time Markov chain (DTMC). This is the process $(X(t), t \in \mathbb{N})$ that takes values from a finite (or countable) set of states and obeys the ‘memoryless’ property; i.e. the future state of the process depends only on the current state and not the entire history (see e.g. Guttorp (1995)). Such a process is defined by a one-step transition matrix $P = \{p_{ij}\}$ for $i, j \in \{1, \dots, d\}$, which describes the probability of the state of the process at time t , given the state at time

$t - 1$, i.e.

$$p_{ij} = p(X(t) = j \mid X(t-1) = i).$$

Given a sequence of observations of the state of the process, $X(t)$, sufficient statistics for inference regarding the transition matrix P are given by $\{n_{ij}\}$, the number of observed transitions from states i to j in a single time step. The likelihood of the transition matrix is

$$p(\{n_{ij}\} \mid \{p_{ij}\}) = p(X(t_1)) \prod_{i=1}^d \prod_{j=1}^d p_{ij}^{n_{ij}},$$

where $p(X(t_1))$ is the probability of the initial observation. A possibility for $p(X(t_1))$ would be to assume that the process has reached its stationary distribution at the time of the initial observation. The stationary distribution is the row vector $\boldsymbol{\pi}$ that satisfies

$$\boldsymbol{\pi}P = \boldsymbol{\pi}, \quad \sum_{i=1}^d \pi_i = 1.$$

If such a distribution exists (see e.g. Guttorp (1995) for existence criteria) then $\boldsymbol{\pi}_{X(t_1)}$ can be assumed for $p(X(t_1))$.

The distribution of the steps and turns in this multistate model is a mixture of each behavioural-specific component, resulting in the computationally infeasible likelihood

$$\begin{aligned} & p(\{Z_1, \dots, Z_M\} \mid \Omega, \{p_{ij}\}) \\ &= \sum_{s_1=1}^d \cdots \sum_{s_{M-1}=1}^d p(\{Z_1, \dots, Z_M\} \mid \Omega, \{s_1, \dots, s_{M-1}\}) p(\{s_1, \dots, s_{M-1}\}) \\ &= \sum_{s_1=1}^d \cdots \sum_{s_{M-1}=1}^d \left\{ p(s_1) \prod_{i=1}^{M-2} \Phi(\phi_i \mid \Omega_{\phi, s_i}) \prod_{i=1}^{M-1} R(r_i \mid \Omega_{r, s_i}) \prod_{i=2}^{M-1} p_{s_{i-1} s_i} \right\}, \quad (1.2) \end{aligned}$$

where $\{s_1, \dots, s_M\}$ is the behavioural state sequence. Rather than integrating over all state process combinations, as in the likelihood above, inference for such a mixture model is carried out using Bayesian Monte Carlo techniques in standard software, such as WinBUGS (Lunn et al., 2000). This Bayesian approach makes inference computationally feasible by augmenting the unknown state process and using a complete data likelihood. This multistate movement model gained widespread use, with Beyer et al. (2013) assessing its effectiveness at decoding the behavioural sequence and estimating parameters with simulated data.

Morales et al. (2004) only assume movement follows a CRW, which was then relaxed in McClintock et al. (2012) to allow a range of movements such as biased and attractive walks.

Rather than modelling the turning angle explicitly, McClintock et al. (2012) use the bearing, which describes the angular direction the animal is facing (i.e. the turns are the increments of the bearing process). For a biased random walk (BRW), the animal is most likely to keep moving in a specific direction, and so the bearing process is centred at some non-zero value. In the case of an attractive walk, the animal is drawn to a specific location; the centre of the bearing distribution at any point in time is the bearing between the current location and the location of attraction. The ability to handle irregular sampling schemes was also addressed in McClintock et al. (2012), implementing an ad-hoc linear interpolation of locations to create a regularly timed set of ‘observations’. These models were further extended and applied in Roever et al. (2014) to include habitat covariates in the behavioural process.

1.2.2 Incorporating observation error

State-space models (SSM) extend basic step-and-turn movement models to allow for spatial observation error (McClintock et al. (2014); McClintock et al. (2012) and see review in Patterson et al. (2008)). The SSM is a class of models for time series where the process of interest may not be that observed, and with the additional complexity that the observed process may be noisy. An SSM assumes that the observed process is dependent only on the current unobserved value, which in turn is a Markov process (taking real values rather than a DTMC which takes finite values). Such a model has the structure shown in Fig. 1.2. The definition of an SSM therefore involves an observation and process model, defined as

$$\begin{aligned} \text{observation equation: } Z_t^* &= h(Z_t, \varepsilon_t), \\ \text{process equation: } Z_t &= g(Z_{t-1}, \zeta_t). \end{aligned}$$

The process $h(\cdot)$ describes the observation error model in terms of the parameters ε_t and the process $g(\cdot)$ describes the random process of movement in terms of the parameter ζ_t . In the step-and-turn model previously described, the process parameters are $\zeta_t = \Omega$ and the process $g(\cdot)$ is

$$Z_t = Z_{t-1} + r_{t-1} \begin{pmatrix} \cos(\phi_{t-1}) \\ \sin(\phi_{t-1}) \end{pmatrix}.$$

Often the error model $h(\cdot)$ is assumed to be Gaussian (McClintock et al., 2012) so that

$$Z_t^* = Z_t + e_t, \quad e_t \sim N_2(\mathbf{0}, \varepsilon_t I_2),$$

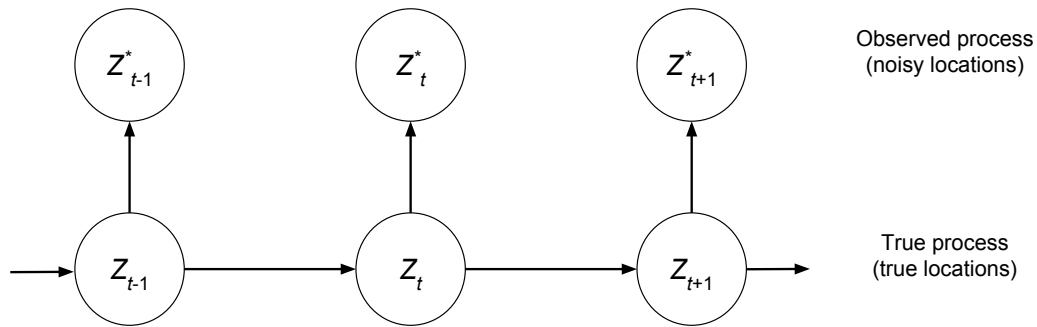


Fig. 1.2 Structure of an SSM. In this case, observations of the animal's location have error and its true location evolves in time based on the step-and-turn model.

where I_2 is the identity matrix, however more sophisticated error models for specific telemetry devices do exist (McClintock et al., 2015).

SSMs are used extensively in the movement modelling literature for the incorporation of observation error because of their modelling flexibility (Anderson-Sprecher and Ledolter, 1991; Breed et al., 2012; Jonsen et al., 2013, 2005, 2003; Patterson et al., 2010, 2008), however, fitting these models is not always straightforward. When the model specification is linear (i.e. $Z_t = A_t Z_{t-1} + B_t + \zeta_t$) and the error process is Gaussian then fast-fitting algorithms such as the Kalman filter are available (see Harvey (1990) for a detailed description). As long as the state of the process at the initial time is Gaussian, the likelihood of the whole process is also, allowing evaluation of the likelihood. The Kalman filter is a recursive algorithm to compute the optimal estimate of the true process that sweeps along the time series, predicting and updating the true state (Z_t) based on the observations up to and including that point in time. Given this optimal estimate and the assumption that the likelihood is Gaussian, estimates can be made for all unknown parameters in the model.

Although the Kalman filter provides efficient model fitting techniques, when the model specification is non-linear (as in the step-and-turn movement case given here) the above assumptions are invalid. In such a case, if the model is almost linear at the time scale of the filtering process (usually the observation time scale) then the extended Kalman filter can be implemented. This approach involves linearising the process around an estimate of the mean and covariance at the current step of the algorithm (Einicke and White, 1999). If the model specification is highly non-linear and the extended filter cannot be applied, unscented Kalman filtering can be used, in which a set of points are deterministically sampled around the mean, that allow for estimation of the covariance that is more robust than that of the extended filter (Julier and Uhlmann, 1997). Although these approaches exist, they are often difficult to implement and choose appropriate tuning values for. Alternative estimation techniques

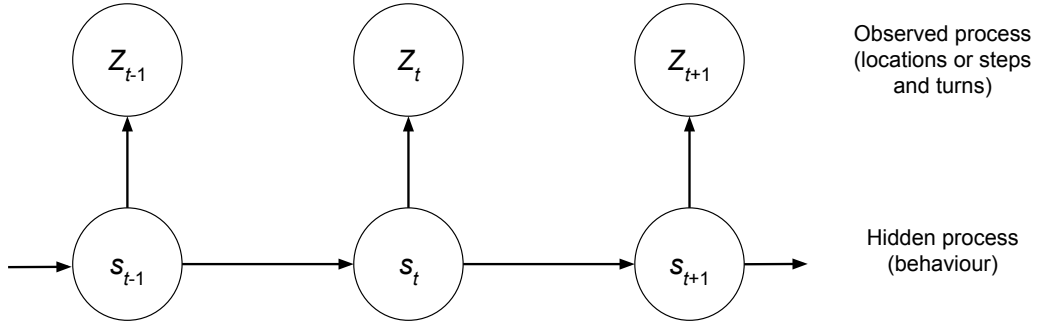


Fig. 1.3 Structure of an HMM. In this case, the animal's true location is observed, but is dependent upon a hidden behavioural process that follows a DTMC.

involve integrating over the entire state process, which is computationally infeasible in practice, or applying (computationally feasible but still demanding) Bayesian Monte Carlo methods to estimate the unobserved state process and parameters (as in McClintock et al. (2012)).

1.2.3 Efficient modelling with hidden Markov models

Formulating the mixture model of Morales et al. (2004) as a hidden Markov model (HMM) improves the efficiency of the inference approach. The HMM is a stochastic process comprising of an unobserved DTMC, with state-dependent observation process. As shown in Fig 1.3, an HMM has the same general structure as an SSM (Fig. 1.2), but models behaviour, not error. The observed process Z represents the locations (and therefore the steps r and turns ϕ), and the hidden process s is the unobserved behavioural state of the animal. The definition of these two processes are the same as that described in the Morales et al. (2004) model above.

In comparison with the SSM, in which the unobserved process has continuous states, the hidden process of the HMM is discrete and thus allows evaluating the infeasible likelihood in Eq. 1.2 in an efficient way based on the forward algorithm; a recursive algorithm similar to that of the Kalman filter (see Zucchini et al. (2016) for details). The likelihood is written as

$$p(\{Z_1, \dots, Z_M\} | \Omega, \{p_{ij}\}) = \boldsymbol{\pi} Q(Z_1) P Q(Z_2) \dots P Q(Z_M) \mathbf{1}^T,$$

where $\mathbf{1}$ is a row vector of ones, P is the one-step transition matrix between states, and $Q(Z_i)$ is the diagonal matrix with elements given by $\Phi(\phi_i | \Omega_{\phi,j}) R(r_i | \Omega_{r,j})$ for $j \in \{1, \dots, d\}$, where d is the number of states. Written in this form, the complexity of the likelihood is linear with regards the number of observations, and inference becomes feasible either as maximum

likelihood or Bayesian Monte Carlo. Along with parameter estimation, the unobserved behavioural process is reconstructed by applying the Viterbi algorithm (see Zucchini et al. (2016) for details), which is a recursive algorithm that constructs the set of optimal state sequences (typically used in the classical framework).

The HMM has been used for modelling the behaviours of animals in a number of ways, such as feeding (Schliehe-Diecks et al., 2012) and dive behaviours (Bagniewska et al., 2013). Employing an HMM for modelling step-and-turn movement was first introduced by Franke et al. (2004), but with the simple categorisation (such as slow, medium, fast) of the step lengths and turning angles rather than an underlying parametric model. Patterson et al. (2009) give a parametric version with only step lengths, and Langrock et al. (2012) give the same underlying movement model as Morales et al. (2004), extended by McKellar et al. (2015) to include environmental covariates in the behaviour process. In comparison, Forester et al. (2007) employ an SSM for behaviours, modelling this as a continuous-valued variable. Extensions to the standard HMM include Leos-Barajas et al. (2017); Li and Bolker (2017); Towner et al. (2016), who all use a behavioural process that is heterogeneous in time to account for periodicities, such as diurnal variations.

Although HMMs are able to account for missing observations, this is usually under the assumption that the process causing the missing data is random; often not the case with movement observations. Structure in missing observations leads to biased estimates (Nakagawa and Freckleton, 2008) unless correctly accounted for, which can be implemented through including the probability of recording an observation as a function of the system process. Although this extension is possible it has not been widely implemented regarding animal movement, despite it being common to have non-random missing observations. Further, the task of model selection, and in particular the choice of the number of behavioural states, is non-trivial. Pohle et al. (2017) and Li and Bolker (2017) found that information criterion often lead to models with a higher than expected number of states. It was found necessary to include additional states to account for observation errors and outliers, seasonality and heterogeneity between multiple individuals—particularly a problem in large datasets. Large numbers of states cause difficulties when interpreting behavioural classifications, being a construct of poor model fit rather than ecological processes.

1.3 Problems with discrete-time modelling

The step-and-turn movement models introduced in the previous section are formulated in discrete time, defining movement only on some pre-determined ‘grid’ of times. Described by McClintock et al. (2014) as the ‘intuitive’ choice, it is implicitly assumed that the discrete-time process represents regular observations from the underlying continuous-time movement process of the animal. The continuous and discrete formulations of movement, however, may not be fundamentally substitutable in this way (Nams, 2013).

The time scale of a step-and-turn model must be chosen prior to fitting, with model and inference not being time-scale invariant (McClintock et al., 2014). This places unwarranted importance on the choice of scale and sampling rates of paths have been shown to have a large effect on the inferred movement (Codling and Hill, 2005; Rowcliffe et al., 2012). In particular, discretising a CRW has been shown to result in a trajectory that is no longer such (Nams, 2013). Often, the times from the observed data are used, which is a potentially dangerous choice as this may be unrelated to any biologically important time scale for the animal—particularly one at which movement decisions are made (Harris and Blackwell, 2013). If the behavioural process of the animal is taken to represent regular observations from the underlying continuous-time behavioural process (as in Langrock et al. (2012); McClintock et al. (2012); Morales et al. (2004)), the existence of such a process and the effect of discretisation is not trivial to address. For example, not all DTMCs have a continuous-time counterpart and such a representation would therefore be invalid.

Even if the time scale is of biological importance, the lack of scale invariance makes combining sources of data or comparing analyses challenging (Harris and Blackwell, 2013). Irregularity of movement data therefore presents a challenge for discrete-time models. Further, the ability to think about an animal’s location between two observations is not trivial as the lack of scale invariance leads to this being undefined. A somewhat ad-hoc solution is often given by a simple linear interpolation of the locations at valid time points (Jonsen et al., 2005; McClintock et al., 2012).

Formulating a model in continuous time reflects the true mechanisms of animal movement and removes the need to choose an arbitrary time scale. Observations that are irregular and ‘gappy’ are applied to a continuous-time model with ease, thus offering a flexible and widely applicable framework resilient to the sampling scheme used (Gurarie et al., 2016). The model is defined for all real times and so locations at inter-observation times can be estimated. The practical application of continuous-time models, however, is limited; statistical inference is often more computationally demanding and non-statisticians have had

difficulty interpreting the parameters describing infinitesimal quantities (McClintock et al., 2014). There is therefore a need to develop robust and meaningful continuous-time models based on easily interpretable movement parameters.

1.4 Aims of this thesis

Through this thesis we aim to increase the accessibility to modelling animal movement in continuous time by developing a robust, statistical model with interpretable parameters. In particular, we sought an analogue to the popular discrete-time step-and-turn models with parameters similar to, e.g., a mean step length. In continuous time, we define this movement by a joint bearing and speed process.

As in discrete-time models, the following work intends to develop a movement model capable of behavioural switching, but one defined in continuous time to allow switches at more than just observation times. Not only is the aim to learn about the parameters defining this process, but interest also lies in estimating the unobserved behavioural process itself—similar to the ability to estimate the optimal state sequence via the Viterbi algorithm in an HMM.

Current approaches using steps and turns have often had to make the choice between including observation error or behavioural switching. Our aim here is to develop methods to incorporate both these features in a statistically robust way. Interest lies in developing techniques to learn about both the underlying error process and to estimate the true locations of the animal at observation times—an important feature when introducing environmental covariates to the model.

As part of the aim to make continuous-time movement models accessible to practitioners, the inference methods described in this thesis can be implemented using the R package `CTStepTurn`, which is available at <https://github.com/a-parton/CTStepTurn>. Scripts to reproduce many of the examples described here are included, covering a range of practical scenarios with the aim to provide an example base that practitioners can adapt to their situation.

1.4.1 Outline of thesis

Chap. 2 gives a review of continuous-time animal movement modelling. We focus on work relating to the aims listed above rather than the full breadth of the current state of the art.

In Sect. 2.4 we summarise how the literature on continuous time will relate to the models developed over Chaps. 3–5.

In Chap. 3 we introduce the proposed model for continuous-time single state movement based on step lengths and turning angles. This model formulates movement based on the popular step-and-turn varieties, using bearing and speed processes as analogues for these metrics in continuous time. This chapter begins by creating movement patterns comparable to the discrete-time Morales et al. (2004), but in Sect. 3.4 we propose an alternate model that includes correlation to the speed process in addition to the bearing. We outline a method for Bayesian inference in the form of an augmentation approach inspired by the techniques discussed in Chap. 2. These methods are then demonstrated with applications using simulated datasets. In these examples, inference is also carried out at the observation time-scale, highlighting the bias in estimation encountered in a ‘discrete-time style’ analysis.

Chaps. 4 and 5 extend the single state movement model to incorporate the aims listed above; multistate behavioural switching and the presence of observation error, respectively. We introduce behavioural switching through a continuous-time Markov chain (CTMC) based on current approaches in Chap. 2. Examples demonstrate the ability of the inference process using simulated data, and well-known data of elk movements compares these methods with popular discrete-time HMMs. In the observation error case, we first introduce an independent model for error and then extend this in Sect. 5.6 to incorporate correlation. We apply single state movement with error to data from reindeer, and two-state movement with error to observations of a lesser black-backed gull. A simulation example explores the ability to estimate both movement and error parameters simultaneously.

In Chap. 6 we discuss the methods presented in Chaps. 3–5 and how they fit within the current state of continuous-time animal movement in Chap. 2. We provide suggestions for future work.

Chapter 2

Review of continuous-time modelling

Continuous-time movement models can be seen as the ‘gold standard’ of movement modelling, avoiding the challenges of discrete time through being scale-invariant and respecting the continuous nature of an animal’s movement (see the discussion on the formulation of time in Sect. 1.3). Although a number of models formulated in continuous time exist, their uptake has been somewhat lagging in comparison to the discrete-time step-and-turn models of Sect. 1.2. Such limited application has mainly been rationalised by the often inaccessible parameter interpretation and computational demand of continuous-time methods. The following chapter outlines the existing continuous-time methods for modelling animal movement data that this thesis aims to build upon to provide intuitive and understandable methods. The following is an extension to the author’s edited contribution within the published review Patterson et al. (2017).

2.1 Modelling movement with diffusion processes

Animal movement models formulated in continuous time are usually based on diffusion processes—continuous-time Markov processes that are the solution to some set of stochastic differential equations (SDE). Such processes suit animal movement because their sample paths are (almost surely) continuous and can contain randomness. Linear, Gaussian diffusion processes are tractable, and can be ‘combined’ to produce a flexible range of possible trajectories. This section explores the current approaches to modelling movement in continuous time by focusing on the diffusion processes that form the ‘building blocks’ underpinning such models.

2.1.1 Brownian motion

The simplest diffusion process is Brownian motion (BM, or Wiener process), a Gaussian process with continuous paths and independent increments such that

$$W(t) \sim N(0, t), \quad (2.1)$$

with $W(0) = 0$ and $\text{Cov}(W(t), W(s)) = s$ for $s \leq t$. BM can be characterised as the continuous-time limit of a simple random walk (Iacus, 2008) and is an important component for more sophisticated diffusion processes, such as those in the following sections. Extending to d dimensions, BM is generalised as

$$\mathbf{W}(t) \sim N_d(\mathbf{0}, t\Sigma),$$

where Σ is the d -dimensional covariance matrix that scales standard BM in each dimension. Given a sequence of observations $\{\mathbf{x}(t_j)\}$ from a process following BM, the likelihood of such involves the product over each multivariate normal density describing the independent increments of the process;

$$p(\{\mathbf{x}(t_j)\} | \Sigma) = \prod_j (2\pi)^{-d/2} |\Delta t_j \Sigma|^{-1/2} \exp(-(\Delta \mathbf{x}_j' (\Delta t_j \Sigma)^{-1} \Delta \mathbf{x}_j)/2),$$

where $\Delta \mathbf{x}_j = \mathbf{x}(t_{j+1}) - \mathbf{x}(t_j)$ and $\Delta t_j = t_{j+1} - t_j$.

The d -dimensional process derived from BM that is constrained to start at \mathbf{x} at time s and end at \mathbf{y} at time u is referred to as the Brownian bridge (BB, Iacus (2008)), defined for $s \leq t \leq u$ in terms of BM as

$$\mathbf{W}_{s,\mathbf{x}}^{u,\mathbf{y}}(t) = \mathbf{x} + \mathbf{W}(t-s) - \frac{t-s}{u-s} (\mathbf{W}(u-s) - \mathbf{y} + \mathbf{x}),$$

where \mathbf{W} is Brownian motion so that

$$\begin{pmatrix} \mathbf{W}(t-s) \\ \mathbf{W}(u-s) \end{pmatrix} \sim N_{2d} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} t-s & t-s \\ t-s & u-s \end{pmatrix} \Sigma \right).$$

Hence $\mathbf{W}_{s,\mathbf{x}}^{u,\mathbf{y}}(t)$ is given by the multivariate normal distribution

$$\mathbf{W}_{s,\mathbf{x}}^{u,\mathbf{y}}(t) \sim N_d \left(\mathbf{x} + \frac{t-s}{u-s} (\mathbf{y} - \mathbf{x}), \frac{(u-t)(t-s)}{u-s} \Sigma \right). \quad (2.2)$$

2.1.1.1 Brownian motion as a model for position

As a model for animal movement, BM is simplistic, representing movement that is unbiased, undirected and has no interaction with the environment. For this reason, it is only of use either as a component within a switching model (implicitly in Blackwell (1997, 2003), see Sect. 2.3) or as a purely local model such as Horne et al. (2007). Movement between consecutive, known locations is estimated by the use of Brownian bridges, scaled by a volatility parameter.

Inference in Horne et al. (2007) involves learning about the variance parameter of BM, with observations having the likelihood given by Eq. 2.2. Given multiple observations, the product is taken over each disjoint BB formed between pairs of successive locations. Because the distribution of an animal's location at any time is normally distributed with the estimated variance, quantile regions can be calculated with ease. This kind of interpolation is informative about utilisation distributions—the marginal distribution of an animal's location—and hence habitat use.

Spatial observation error is accounted for in the BB model of Horne et al. (2007) by assuming independent, identically distributed (iid) normal perturbations with known variance. In this case inference uses a 'leave one out' approach, removing alternate observations and essentially using only part of the information provided by observations for inference. An extension enabling estimation of the observation error parameter is not provided and so sensitivity analysis based on the chosen value is important.

2.1.2 Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck (OU), or Vasicek, process (Uhlenbeck and Ornstein, 1930) is the solution to the SDE

$$dU(t) = \beta (\mu - U(t)) dt + \sigma dW(t),$$

where $W(t)$ is BM. This stationary, Gaussian process is mean-reverting (Dunn and Gipson, 1977), and so has a tendency to drift towards its long-term mean. The process is given in terms of BM as

$$U(t) = \mu + \frac{\sigma e^{-\beta t}}{\sqrt{2\beta}} W(e^{2\beta t}),$$

distributed (based on Eq. 2.1 and Iacus (2008))

$$U(t) \sim \text{N}\left(\mu, \frac{\sigma^2}{2\beta}\right).$$

The process at a future point in time, given the current value $U(0) = u_0$, is

$$U(t) = \mu + (u_0 - \mu)e^{-\beta t} + \frac{\sigma}{\sqrt{2\beta}}W\left(e^{2\beta t} - 1\right)e^{-\beta t},$$

having the conditional distribution (again, based on Eq. 2.1)

$$U(t) \mid U(0) = u_0 \sim N\left(\mu + (u_0 - \mu)e^{-\beta t}, \frac{\sigma^2}{2\beta}\left(1 - e^{-2\beta t}\right)\right).$$

Extending to d dimensions, the stationary and conditional distributions, respectively, of the OU process are

$$\begin{aligned} \mathbf{U}(t) &\sim N_d(\boldsymbol{\mu}, \Lambda), \\ \mathbf{U}(t) \mid \mathbf{U}(0) = \mathbf{u}_0 &\sim N_d\left(\boldsymbol{\mu} + (\mathbf{u}_0 - \boldsymbol{\mu})e^{Bt}, \Lambda - e^{Bt}\Lambda e^{B't}\right), \end{aligned} \quad (2.3)$$

where $\boldsymbol{\mu}$ is a d -dimensional vector, Λ is a $d \times d$ covariance matrix, and B is a stable $d \times d$ matrix—that is, $e^{Bt} \rightarrow 0$ as $t \rightarrow \infty$ (Blackwell, 1997; Dunn and Gipson, 1977). Hence $\boldsymbol{\mu}$ describes the centre of the process, with rate of attraction towards the centre controlled by B and with random variation governed by Λ .

In two dimensions, the distribution of the OU process is unimodal and elliptical. It is worth noting here the special, circular case in which the matrices B, Λ are isotropic, i.e. symmetric under rotation and reflection so that $B = \beta I_2, \Lambda = \frac{\sigma^2}{2\beta} I_2$, with $\beta < 0$. In this case the attraction towards the centre $\boldsymbol{\mu}$ is determined only by the current displacement from the centre rather than by direction and no significance is placed upon the dimensional system chosen (Blackwell, 1997). Further, this simplified case is equivalent to two independent, one-dimensional OU processes with parameters μ_i, β, σ^2 .

2.1.2.1 Ornstein-Uhlenbeck process as a model for position

Dunn and Gipson (1977) gave the first method for modelling animal positions in continuous time, describing two-dimensional positions by an OU process. This kind of process can be informative about the home range of an animal—the spatial range in which it performs its daily activities (see recent discussion/review by Börger et al. (2008)). This is often mathematically defined as the smallest geographical area in which the animal spends a fixed proportion of time and can be estimated by the equilibrium distribution of the OU process,

similar to the Gaussian home ranges estimated in Jennrich and Turner (1969). Inference for the OU parameters in Dunn and Gipson (1977) uses maximum likelihood, based on the animal beginning movement under its equilibrium distribution and successive observations arising from the conditional distribution.

The OU process addresses the problem of autocorrelation of position, meaning high frequency ‘bursts’ of observations can be modelled. The OU process, however, will always result in an estimate of home-range being elliptical and unimodal. For some animals and habitats this will not be an appropriate assumption (Blackwell, 1997). As with BM, this limits usefulness as a lone model for movement, but it is an important component in constructing more realistic models (see Sect. 2.3).

2.1.2.2 Ornstein-Uhlenbeck process as a model for velocity

The continuous-time model of Johnson et al. (2008a) takes the popular movement assumption of a CRW by directly modelling the velocity, rather than position, of the animal. The velocity describes the instantaneous rate of change of location, modelled as a bivariate OU process. In practice, Johnson et al. (2008a) use two independent one-dimensional OU processes (the circular case described above) as they argue that a non-zero correlation in the velocity process would lead to unrealistic movement patterns. The persistence assumption, that an animal will likely keep moving in a similar direction over a short period of time, is incorporated as a result of the autocorrelation of the OU process. The location of the animal at any time, t , is found by integrating the velocity process

$$\mathbf{z}(t) = \mathbf{z}_0 + \int_0^t \mathbf{v}(u) du.$$

This results in the location process no longer being Markovian—as in the OU position model above—as it depends on the entire velocity process prior to time t . However, the combined process of position and velocity is Markovian.

Observation error in position is incorporated into Johnson et al. (2008a) via a SSM (see Sect. 1.2.2) with Gaussian distributed errors. Unlike Horne et al. (2007), the parameters describing observation error in Johnson et al. (2008a) are not assumed known and are estimated as part of the inference process. In relation to the SSM definition of Sect. 1.2.2, the observation equation, Z_t^* , describes the observed location and the process equation, Z_t , describes the joint process consisting of the true location and the velocity (because this joint process is Markovian). Using the SSM framework, Johnson et al. (2008a) use the

Kalman filter (introduced in Sect. 1.2.2) to carry out statistical inference. The Gaussian error assumption and the linearity of the OU velocity process are necessary for the Kalman filter to be valid. The algorithm provides optimal estimates of the process Z_t (the unobserved true locations of the animal as well as the velocity at each observation time) and uses maximum likelihood to estimate the OU process parameters and the observation error variance. Further, the Kalman filter can also estimate the location and velocity at missing observation times.

As mentioned above, estimating the home range of an animal is a common metric in movement analyses. The true locations estimated in the Johnson et al. (2008a) method can be used to calculate the animal's home range, with a popular method for this being kernel density estimation. However, a key assumption of such a method is that observations are independent, yet the locations of an animal are highly correlated. Further, features such as irregularity in time that cause 'bunches' of similarly timed observations to exist bias density estimates. An extension to Johnson et al. (2008a) is given in Johnson et al. (2011) that allows estimation of movement summaries, such as home range, in a rigorous way. A Bayesian data augmentation approach is taken in which samples of the unobserved location process are drawn and used to gain a Monte Carlo estimate of the summary statistic of interest, allowing uncertainty to be properly quantified.

Applying the Kalman filter for statistical inference involves assuming that observation errors are normally distributed. This assumption is often questioned for some tracking technologies due to the presence of large outliers (e.g. marine animals observed inland) and is thought to be modelled more closely by a t -distribution that can allow for large errors through heavier tails (Jonsen et al., 2005). An extension to Johnson et al. (2008a) is presented in Albertsen et al. (2015) that uses t -distributed errors rather than Gaussian. In order to carry out statistical inference (because the Kalman filter can no longer be applied), the Laplace approximation is used to evaluate the integral of interest, in this case integrating over the unobserved true locations. Using simulations, Albertsen et al. (2015) show that this method improves upon true location estimation through the heavy-tailed error distribution.

2.2 Modelling movement with general SDEs

The diffusion processes of the previous section arise as solutions to particular cases of the general SDE

$$dX(t) = b(X(t)) dt + \sigma(X(t)) dW(t). \quad (2.4)$$

A more flexible modelling approach can be achieved by describing movement by an SDE but without the restriction of needing to find a solution to such. This flexibility is generally at the expense of computational, and hence statistical, tractability. A number of approaches consider the case where the SDE derives from basing $b(\cdot)$ in Eq. 2.4 on the gradient of a harmonic function (Brillinger et al., 2002; Brillinger and Stewart, 1998; Preisler et al., 2004, 2013). Such a function represents the animal's attraction/repulsion to a particular point, line or region in a general way. Brillinger and Stewart (1998) and Brillinger et al. (2002) use specific models that incorporate spherical geometry to allow for a natural representation of long-range migration along great circle routes. Preisler et al. (2004, 2013) present more general models, which also include the special, tractable cases of Sect. 2.1.

Some element of approximation is required to fit models of this general type. Typically, this involves a normal approximation to the movement at each observed time-step. An approximate likelihood can then be derived and maximised; the quality of the approximation obviously depends on the frequency of the data compared with the true rates at which $b(\cdot)$, $\sigma(\cdot)$ vary.

More sophisticated approaches to inference from SDEs are available, but are rarely used in a movement context because of the extra computational cost, particularly in the presence of measurement error. A recent advancement in this area is that of Scharf et al. (2017), where a range of movement types can be proposed by a general SDE. Although inference cannot be carried out directly, an augmentation technique is used via multiple imputation, in which the unobserved movement path is simulated multiple times from an approximate imputation distribution (such as Buderman et al. (2016) or Johnson et al. (2008a)). A Markov chain Monte Carlo (MCMC) sampler can be used to make inference about the movement parameters based on the imputed paths. This approach provides flexible modelling options, but relies heavily on the assumptions of the approximate imputation distribution.

2.3 Modelling switching behaviour

Animal movement cannot be homogeneous over an extended period of time, yet incorporating heterogeneity into a continuous-time framework has not been widespread. In the velocity model of Johnson et al. (2008a) behavioural switching is highly restricted, included only as a covariate that acts on movement by setting the velocity to zero to create a stationary state at known times (these times are based on additional tag information). Similarly limited scenarios include diffusive movement (e.g. BM or correlated velocity) where parameters are

dependent upon the discrete location of the animal (Ovaskainen, 2004; Ovaskainen et al., 2008), simplifying to a discrete-time behavioural process (Hanks et al., 2011; Kranstauber et al., 2014; McClintock et al., 2014) or movement process (Breed et al., 2012; Jonsen et al., 2005).

The following explores a number of different approaches to modelling that aim to incorporate heterogeneity in the movement process with time and (in a limited number of methods) environment. It is important to note that in all the following cases, while it is convenient to refer to the ‘behaviour process’, the behavioural state potentially has limitations; that is, the state may reflect a statistical description of movement rather than necessarily being ‘behaviour’ in a true biological sense.

2.3.1 Movement measures

A number of studies have used movement measures to describe and explore types of movement modes. A popular measure is the level of autocorrelation between the step lengths calculated from observations (Boyce et al., 2010; Gurarie and Ovaskainen, 2011). The sample autocorrelation function (ACF) is calculated for each time lag h as

$$\text{ACF}(h) = \frac{M(h)^{-1} \sum_{i=1}^{M(h)} (x_i - \bar{x})(x_{i+h} - \bar{x})}{M^{-1} \sum_{i=1}^M (x_i - \bar{x})^2},$$

where M is the total number of observations, $M(h)$ is the number of pairs of observations at time lag h and \bar{x} is the sample mean. By estimating the ACF over the range of time lags between the sampling frequency and period, this gives an indication of the properties of the time series, such as its order of autoregression. Other measures exist that attempt to classify movement over time, but are often overly simplistic. These include calculating a sinuosity/‘straightness’ index by taking the straight-line distances travelled between observed locations (Bovet and Benhamou, 1988; Postlethwaite et al., 2013), the first-passage/residence time calculated as the time required to cross a circle of given radius (Barraquand and Benhamou, 2008; Fauchald and Tveraa, 2003; Lavielle, 2005), and fractal analysis methods that estimate the degree with which an area has been covered by the animal (Tremblay et al., 2007). In all cases, these measures can only be used as exploratory tools and are sensitive to the observation sampling frequency because they often rely on derived quantities such as step lengths.

The ACF calculation above requires estimation of the mean and variance of the full time series, yet the usual estimators for these quantities are unreliable when data is highly correlated. In

contrast, the semivariance function (SVF) contains equivalent information to the ACF, but avoids estimation of the mean and variance. The SVF measures the variation in a series over all possible time lags spanning between the sampling frequency and the overall period, given as

$$\text{SVF}(h) = \frac{1}{2}M(h)^{-1} \sum_{i=1}^{M(h)} (x_{i+h} - x_i)^2,$$

for lag h . Further, the statistical properties of the SVF are more tractable than the ACF, being χ^2 -distributed for a given lag when the underlying process is Gaussian. Based on this, a robust statistical modelling approach is given in Fleming et al. (2014a,b) using measures of correlation to identify trends in movement at different time scales.

Fleming et al. (2014a) give the empirical form of the SVF in terms of a set of movement parameters for a number of popular models, including BM and the OU process, along with mixtures of models, such as the OU process interspersed with random bouts of foraging activity. Hence, a flexible range of movement models are considered. The non-stationarity of movement data due to cycles (such as seasonality) results in the SVF depending not only on the lag but also the absolute times of the observations. To handle this, Fleming et al. (2014a) adjust the process by taking a time-average that treats the non-stationarity as a nuisance factor and leaves only the lag dependence. Given observations, weighted least-squares regression is used to estimate the parameters in the SVF for each movement model and information criterion is used for model selection. This technique enables identifying and making inference on both the different movement behaviours associated with an animal's trajectory and the temporal scales over which these behaviours occur. The inference approach of Fleming et al. (2014a) is improved upon in Fleming et al. (2017) by using the Kalman filter (see Sect. 1.2.2) to speed up the model fitting process.

Fleming et al. (2014a) is limited by an inability to associate behaviours directly with environmental information or identify the specific behavioural state employed at any one point in time. Although this problem is still unsolved, a step towards this end is given in Fleming et al. (2016), presenting methods for inferring the full movement (but not behavioural) trajectory of the animal. This involves estimating the density of space use between observed locations. As described in Sect. 2.1.1.1, the BB is often used as a model for between-observation locations, but is overly simplistic. In Fleming et al. (2016), the more sophisticated movement models of Fleming et al. (2014a) are used to estimate the space use density. This approach relies on each model being Gaussian, enabling estimation of the location at unobserved times, conditional upon the observed locations and chosen model.

Another step towards identifying the behavioural changes in movement models of this form is Gurarie et al. (2017), concerned in particular with identifying movement behaviours with much longer time frames than home ranging and foraging (such as migration). This method attempts to improve upon the simple methods commonly used for this type of identification, such as spatial clustering. The location process is the sum of a mean process, $m(t)$, and an autocorrelated fluctuation process, $r(t)$. A ‘range shift’ is defined as altering the mean process, with a simple shift example being

$$m(t) = \begin{cases} m_1 & \text{if } t < t_1, \\ m_1 + \frac{(m_2 - m_1)(t - t_1)}{t_2 - t_1} & \text{if } t_1 \leq t \leq t_2, \\ m_2 & \text{if } t > t_2, \end{cases}$$

however more sophisticated models with features such as ‘stopovers’ could be used. In the case above, the animal has two centres of movement, m_1 and m_2 , with a linear process controlling the shift between these centres. The autocorrelated fluctuation process then defines the type of movement around the centre, with standard forms being given for all the movement models discussed in Fleming et al. (2014a). Inference in Gurarie et al. (2017) is by likelihood methods similar to Fleming et al. (2014a).

2.3.2 Change point analysis

Gurarie et al. (2009) describe a model that is able to identify heterogeneity and characterise movement within each ‘behaviour’. Their approach is then applied and compared with those including the first passage time of Fauchald and Tveraa (2003) and the multistate random walks of Morales et al. (2004). The underlying movement is parametrised by orthogonal quantities, coined as ‘persistent’ and ‘turning’ velocities, assumed to follow stationary, autoregressive Gaussian processes; each described by a mean, variance and autocorrelation parameter. The persistent and turning velocities are derived from speeds and turning angles, calculated as averages between observations, hence the reliability of this method will vary for irregular datasets.

Given a series of movement observations and the time of a single behavioural switch, the likelihood for the process parameters is simply the product of the two Gaussian functions. A single behavioural change-point is estimated by maximising this likelihood over all possible switching times. In turn, the parameters for each behavioural state are estimated by maximum likelihood using the observations from the respective behavioural state. The method is extended to any number of behavioural states and change-points by passing a

discrete moving window (the size of which needs to be chosen) over the observations and carrying out a single change-point analysis each time. Note that in this model behavioural switches can only occur at observation times, and hence the behavioural process is not continuous in time.

Although able to identify behavioural change-points in an animal's movement path, Gurarie et al. (2009) cannot group disjoint segments into specific behavioural states. Nams (2014) allow for this by taking a similar algorithmic approach, comparing an information criterion to find behavioural change-points. The difference here is that the optimal number of change-points and their positions are found given a set number of behavioural states. Each partition of the trajectory is then classified as being one of the behavioural states. This process is repeated with increasing numbers of behavioural states to find the optimum.

An extension to Horne et al. (2007) allowing for heterogeneity in the movement process over time is Kranstauber et al. (2012), similar to Gurarie et al. (2009). Although the underlying movement process, being BM, is continuous, the behavioural process defined by a changing variance parameter is discrete, being able to 'switch' only at observation times. As in Gurarie et al. (2009), inference involves a sliding window in which all possible switches in behaviour are considered and compared using information criterion. Other approaches to detecting behavioural change-points have also been proposed assuming more restrictive, discrete-time assumptions, based on classification trees (Madon and Hingrat, 2014), segmentation analysis of cumulative steps (Thiebault and Tremblay, 2013), collinearity of bearings (Byrne et al., 2009), wavelet analysis (Sur et al., 2014) and clustering techniques (Zhang et al., 2015).

2.3.3 Functional movement models

A number of approaches to modelling movement in continuous time are formulated using basis functions rather than mechanistic models based on diffusion processes (Brost et al., 2015; Buderman et al., 2016; Hooten and Johnson, 2017). Such models attempt to fit a smooth spline to the observed locations to reconstruct the underlying continuous-time movement path. The location $z(t)$ of the animal is described by the linear combination

$$z(t) = \sum_i \beta_i B_i(t),$$

where β_i are coefficients determining the weight of the basis functions $B_i(\cdot)$.

This modelling approach offers versatility, where the specification of the spline used for interpolating movement can be chosen to allow coefficients at varying ecologically relevant time scales (Buderman et al., 2016) and a temporal warping technique can be applied to allow heterogeneous dynamics in the process, mimicking switching movement behaviours (Hooten and Johnson, 2017). Observation error can be included as a simple Gaussian perturbation (Buderman et al., 2016; Hooten and Johnson, 2017) or as a complex process informed by auxiliary data such as environmental boundaries to the animal (Brost et al., 2015).

As the estimable quantities of these approaches are the coefficients of splines, rather than mechanistic parameters such as a ‘mean speed’, the interpretation of these quantities is unclear. The interpolated splines fit to movement data can post-hoc be used to derive informative quantities of the path such as speed, persistence and residency times (Buderman et al., 2016), but these will not have a covariance structure informed from statistical modelling choices and are only descriptive features. Further, the basis functions $B_i(\cdot)$ must be chosen, which requires knowledge of both the statistical properties of basis functions and their relation to ecological processes, possibly rendering such approaches inaccessible to practitioners (Hefley et al., 2017).

2.3.4 Continuous-time Markov chains

A Markov process is the stochastic process $(M(t), t \geq 0)$ that obeys the ‘memoryless’ Markov property that realisations of the process in the future depend only on the current value (Guttorp, 1995). If the process at time t takes realisations from a finite (or countable) set, and transitions between states are possible at any point in continuous time, then $M(t)$ is a CTMC.

The CTMC $M(t)$ taking values from a finite state-space with length d can be described by its generator matrix, $G = \{g_{ij}\}$ for $i, j \in \{1, \dots, d\}$. The value g_{ij} describes the infinitesimal transition rate from state i to state j . The process can be thought of as residing in state i for a length of time exponentially distributed with mean $\frac{1}{-g_{ii}}$, and then ‘switching’ to another state j with probability $\frac{g_{ij}}{-g_{ii}}$, for $i \neq j$. An alternative parametrisation (Guttorp, 1995) of the process is given by the transition rates out of each state, $\boldsymbol{\lambda} = \{\lambda_i\} = \{-g_{ii}\}$, and the set of jump probabilities $\boldsymbol{q} = \{q_{ij}\} = \frac{g_{ij}}{-g_{ii}}$, for $i \neq j$.

Given observed data of a CTMC, sufficient statistics for the process parameters $\boldsymbol{\lambda}$ and \boldsymbol{q} are given by $\boldsymbol{t} = \{t_i\}$, the total observed time spent in each state i , and $\boldsymbol{n} = \{n_{ij}\}$, the number of observed transitions from state i to state j (Guttorp, 1995). The likelihood for $\boldsymbol{\lambda}$ and \boldsymbol{q} is

then given by

$$p(\mathbf{t}, \mathbf{n} \mid \boldsymbol{\lambda}, \mathbf{q}) = \exp\left(-\sum_i \lambda_i t_i\right) \prod_{i \neq j} (q_{ij} \lambda_i)^{n_{ij}}.$$

2.3.4.1 Markov behaviour

Blackwell (1997) suggests an extension to the BM and OU position models in Sect. 2.1 to allow for behavioural ‘switching’. It is assumed that at any point in time the animal exhibits one of a finite set of behavioural states. The process regarding the behavioural state is assumed to follow a CTMC. The animal’s movement is then modelled as in Dunn and Gipson (1977) by an OU process (or a special limiting case being BM), but with the parameters of such a process being dependent on the behavioural process. When in behavioural state i , movement follows an OU process with parameters $\boldsymbol{\mu}_i, B_i, \Lambda_i$ —see Eq. 2.3.

Statistical inference for these switching models is given in Blackwell (2003), applied to positional data with known behavioural states at each observation time (however, observations of behaviour are not necessary, see Blackwell et al. (2015)). Inference is more complicated than in the Dunn and Gipson (1977) case as the conditional distribution of the animal’s position—given an earlier position in time—depends on the complete behavioural process between these two times. This entire behavioural process however, is unknown. The approach taken by Blackwell (2003) treats the behaviour process as ‘missing’ data and uses Bayesian Monte Carlo techniques. Quantities of interest are split into three groups and a hybrid MCMC is carried out, where posterior distributions are sampled from each group separately, using Gibbs sampling techniques. The three groups are the ‘missing’ complete behaviour process, the behaviour process parameters and the movement process parameters.

Blackwell (1997, 2003) assumes that the behavioural process is independent of the geographical position of the animal. Harris and Blackwell (2013) describe spatially heterogeneous extensions of these models, where movement and behaviour may depend on the discrete spatial region in which an animal is located at a given instant. Blackwell et al. (2015) give a method of Bayesian inference for models where switching probabilities may vary with spatial location, in either discrete or continuous form, and with time. Behaviour is generally taken to be unknown and is reconstructed as part of the MCMC algorithm. It is worth pointing out that a ‘behaviour’ here simply refers to a set of parameter values, and so different behavioural states may simply represent, for example, similar kinds of movement centred on different points of attraction. Combined with dependence of the switching probabilities

on location, this means that these models can represent varied interactions with spatially complex environments.

The inclusion of observation error is not discussed in the continuous-time OU position models (Blackwell, 1997; Blackwell et al., 2015; Dunn and Gipson, 1977). An attempt to address this is given in Nations and Anderson-Sprecher (2006) using an SSM, but at the cost of formulating the model as a discrete-time version of an OU process. Similarly, McClintock (2017) ultimately formulates their approach in discrete time. A single state movement model that is quick to fit but incorporates observation error, such as the velocity model in Johnson et al. (2008a), is applied and multiple realisations of trajectories are simulated. These are then used in a multiple imputation approach; repeatedly fitting an HMM with steps and turns to temporally regular, ‘true’ locations from the simulated trajectories and average over these fitted models. The behavioural process in this approach is defined in discrete time rather than the continuous-time process adopted in Blackwell (1997) and the ability to estimate these behavioural switches was affected by the sampling frequency.

A somewhat different approach to movement modelling is given by Hanks et al. (2015), in which movement is defined in discrete space, splitting the two-dimensional location space into a ‘grid’. A CTMC, with states given by the set of location grid cells, is used to model switches between locations. The movement process is dependent upon environment and changing behaviours of movement can be included by modelling a heterogeneous chain that has time-varying switching parameters. This method is promising, with the potential to describe a wide range of movement behaviours, but is currently limited by its inference approach. Multiple imputation is employed, reconstructing a continuous-time movement path via some other movement model, such as Buderman et al. (2016) or Johnson et al. (2008a), and so this method inherits the assumptions and limitations held by the chosen imputation model.

2.4 Discussion

This review covered a range of current approaches for modelling animal movement in continuous time, providing a base for the step-and-turn model developed over the following chapters. The BB movement model of Horne et al. (2007) is a useful method for the estimation of location between observations, providing a statistically robust approach in comparison to the simple linear interpolation assumed in discrete-time step-and-turn models. Not only can the location between observations be estimated for use in studies concerned

with environmental covariates at a fine spatial scale, but uncertainty is incorporated allowing evaluation of credible regions of location use. The model for movement itself, being BM described by a single parameter, is too simplistic to describe the complex movements of animals. The model described by Fleming et al. (2014a) does allow for complex movement patterns, but there is no way to interpolate location. Our model, developed in Chap. 3, aims to estimate between-observation locations as in Horne et al. (2007), but with more complex patterns of movement such as those in Fleming et al. (2014a). This includes the realistic assumption of correlation in the direction of movement (persistence) and the speed of the animal.

The basis function approaches of Brost et al. (2015); Buderman et al. (2016); Hooten and Johnson (2017) were designed with the same motivation as Horne et al. (2007); to interpolate the animal's location between observations. These methods provide a more adaptable approach than modelling movement with BM, but this flexibility sacrifices the interpretability of model parameters as intuitive descriptors of movement. The behavioural process created by 'warping' time creates a continuous-space, continuous-time process that is desirable in some respects, because it allows a gradual change in the behaviour that enforces correlation in the movement process. This does, however, hinder the ability to distinctly categorise the trajectory into decoded behaviours. The behavioural process introduced in Gurarie et al. (2009); Nams (2013) allows distinct behavioural states, but is formulated in terms of a persistent and turning velocity; metrics which are less ecologically interpretable than the standard steps and turns of discrete time. Further, the change point analysis approach relies heavily on interpolated estimates of these metrics from observations. In Chap. 4, the aim is to allow discrete behavioural states that can be grouped and directly 'attributed' to ecologically relevant activities. This builds upon the behavioural switching approaches of Blackwell et al. (2015), incorporating correlation within the movement process so that a gradual change in the movement is still achieved at behavioural switches.

The movement described by location models using OU processes, such as Blackwell et al. (2015), allow attraction to a centre but not direct correlation in the direction process (other than that enforced by the central attraction). This approach is good for describing the movement of sedentary animals, but not necessarily ranging animals such as reindeer or migrating birds. In contrast, the OU velocity model of Johnson et al. (2008a) does allow for persistent movement. The aim in the following is to create similar movement patterns to Johnson et al. (2008a), but to separate out the speed and direction, as in discrete-time step-and-turn models, to provide a more interpretable solution than the OU velocity approach (and other more general SDE models). Although post-processing of the fitted models of

Blackwell et al. (2015); Johnson et al. (2008a) could separate the inferred movement into these two quantities, the covariance structure of the speed and bearing would be inherited from the fitted process. There is a danger that this structure will not be ecologically relevant. In the following we aim to set the covariance directly for these two processes to ensure that they are based on ecologically realistic concepts. By including a behavioural process based on Blackwell et al. (2015) in Chap. 4, the restriction of single state movement in Johnson et al. (2008a) is overcome. In Chap. 5, the inclusion of observation error similar to Johnson et al. (2008a) is included, extending the ability of the behavioural switching model of Blackwell et al. (2015) that does not incorporate error.

Chapter 3

Single state movement

The following chapter introduces a single state movement model, defined in continuous time and based upon the ideas presented within discrete-time step-and-turn movement models such as Morales et al. (2004). We present a movement model most similar to such discrete-time models in Sect. 3.1, with simulation described in Sect. 3.2. An extension describing more realistic movement is then given in Sect. 3.4. We outline a fully Bayesian method for carrying out inference on the unknown parameters of the single state movement model in Sect. 3.3, adopting a data augmentation technique that aids parameter interpretation. Sect. 3.5 demonstrates both of the presented models for movement by applying the inference algorithm to simulated movement paths, comparing analyses at a number of different time scales. Note that the following (in part) provides an extended description to that in Parton et al. (2017).

3.1 Analogue of discrete-time movement

In continuous time the turning angles and step lengths of discrete-time models do not make sense. An equivalent idea is to model a bearing process in place of turning angles and a speed/distance process to represent the notion of steps.

3.1.1 Bearing process

The bearing process $\theta(t)$, for $t \geq 0$, represents the direction the animal is facing at time t . Comparing this with the turning angles featured in discrete-time models, this is an accumulation of instantaneous ‘turns’ up to t in continuous time. We model the bearing by the general

SDE

$$d\theta(t) = F_1(t, \theta(t)) dt + F_2(t, \theta(t)) dW(t),$$

where $W(t)$ is the Wiener process (see e.g. Guttorp (1995)) and F_i , for $i = \{1, 2\}$, are known functions. A common movement assumption in existing discrete-time models is that of ‘persistence’, implemented by a CRW (Morales et al., 2004). We replicate this assumption here by allowing $\theta(t)$ to evolve according to BM with variance σ_θ^2 , and so $F_1(t, \theta(t)) = 0$ and $F_2(t, \theta(t)) = \sigma_\theta$. Over a finite period of time, the change in direction the animal is facing will be normally distributed around a mean of zero. Uncertainty in bearing increases linearly with time, and correlation between points of the process decreases with increased separation in time.

The direction an animal is facing at any time is constrained to $[-\pi, \pi]$, however, the $\theta(t)$ modelled here is not constrained in this way and can take any real value. For example, given times $0 \leq t < s$, it may be that $\theta(t) = 0$ and $\theta(s) = 2\pi$. Although the animal was facing the same direction at both times, there is information about the behaviour of the process between these points, as the animal has performed an entire ‘loop’, with the process between these points being a BB.

3.1.2 Distance process

At any time $t \geq 0$, the animal has travelled a total distance $\varphi(t)$, which is an increasing quantity representing the accumulation of instantaneous ‘steps’ up to time t . We model the change in φ by the SDE

$$d\varphi(t) = F_3(t, \varphi(t)) dt + F_4(t, \varphi(t)) dW(t),$$

with F_i , for $i = \{3, 4\}$, known functions. We assume the distance process follows BM with drift to reflect discrete-time models, so that $F_3(t, \varphi(t)) = \mu$ and $F_4(t, \varphi(t)) = \sigma_\varphi$. Over the finite period of time t the animal travels a mean distance of μt with variance $\sigma_\varphi^2 t$. This choice reflects the realistic assumption that both the expected distance travelled by the animal, and the variance of such a distance, increases linearly with time. As with discrete-time models, the distances travelled over disjoint periods of time are independent.

The normality assumption of the BM leads to a positive probability of a negative travelling distance. However, introducing a constraint on the combination of μ and σ_φ^2 will ensure that this probability is negligible.

3.1.3 Joint bearing and distance process

GPS observations of an animal's movement occur as $\mathbf{Z}(t) = (X(t), Y(t))^T$ co-ordinate positions. The joint process given by the bearing and speed defines the location process with

$$dX(t) = d\varphi(t) \cos(\theta(t)),$$

$$dY(t) = d\varphi(t) \sin(\theta(t)).$$

3.2 Simulating movement

Realisations of the described movement model, given underlying movement parameters $\Phi = \{\sigma_\theta^2, \mu, \sigma_\varphi^2\}$, can be easily simulated, with an example of such shown in Fig. 3.1. Simulation of the movement path at an approximate time scale δt , which can be arbitrarily fine, is described in the following and summarised in Algorithm 1.

Given the bearing and distance processes at time $t \geq 0$,

$$\theta(t + \delta t) \mid \theta(t) \sim N\left(\theta(t), \sigma_\theta^2 \delta t\right), \quad (3.1)$$

$$\varphi(t + \delta t) \mid \varphi(t) \sim N\left(\varphi(t) + \mu \delta t, \sigma_\varphi^2 \delta t\right). \quad (3.2)$$

All parameters here represent 'per unit time' quantities and are scalable; as opposed to describing movement at a predetermined observed time scale. It becomes useful now to work with the differences of φ on this refined time scale, recovering the more familiar notion from discrete-time models of a 'step' \mathbf{v} , given as

$$\mathbf{v}(t) = \varphi(t + \delta t) - \varphi(t) \sim N\left(\mu \delta t, \sigma_\varphi^2 \delta t\right), \quad (3.3)$$

highlighting the independence of the 'steps' over time.

Given a realisation of the joint bearing and step process, locations in two-dimensional space are the cumulative sums

$$\begin{aligned} X(t + \delta t) \mid X(t), \theta(t), \mathbf{v}(t) &= X(t) + \mathbf{v}(t) \cos(\theta(t)), \\ Y(t + \delta t) \mid Y(t), \theta(t), \mathbf{v}(t) &= Y(t) + \mathbf{v}(t) \sin(\theta(t)). \end{aligned} \quad (3.4)$$

Algorithm 1 Simulating single state movement.

1: **procedure** SIMULATE SINGLE STATE MOVEMENT($\Phi, \{t_0, \dots, t_{m+1}\}$)

2: Set initial location

$$X(t_0), Y(t_0) \leftarrow 0$$

3: Simulate initial bearing (uniform) and step (increment of BM with drift)

$$\theta(t_0) \sim \text{U}(-\pi, \pi)$$

$$v(t_0) \sim \text{N}(\mu(t_1 - t_0), \sigma_\varphi^2(t_1 - t_0))$$

4: Initial step and bearing define next location

$$X(t_1) \leftarrow v(t_0) \cos(\theta(t_0))$$

$$Y(t_1) \leftarrow v(t_0) \sin(\theta(t_0))$$

5: **for** $i \in 1, \dots, m$ **do**

6: Simulate conditional bearing (BM) and step (increment of BM with drift)

$$\theta(t_i) \sim \text{N}(\theta(t_{i-1}), \sigma_\theta^2(t_i - t_{i-1}))$$

$$v(t_i) \sim \text{N}((t_{i+1} - t_i)\mu, (t_{i+1} - t_i)\sigma_\varphi^2)$$

7: Set next location

$$X(t_{i+1}) \leftarrow X(t_1) + v(t_i) \cos(\theta(t_i))$$

$$Y(t_{i+1}) \leftarrow Y(t_1) + v(t_i) \sin(\theta(t_i))$$

8: **end for**

9: **return** Simulated bearings and steps $\{\theta, v\}$ and corresponding locations $\{X, Y\}$

10: **end procedure**

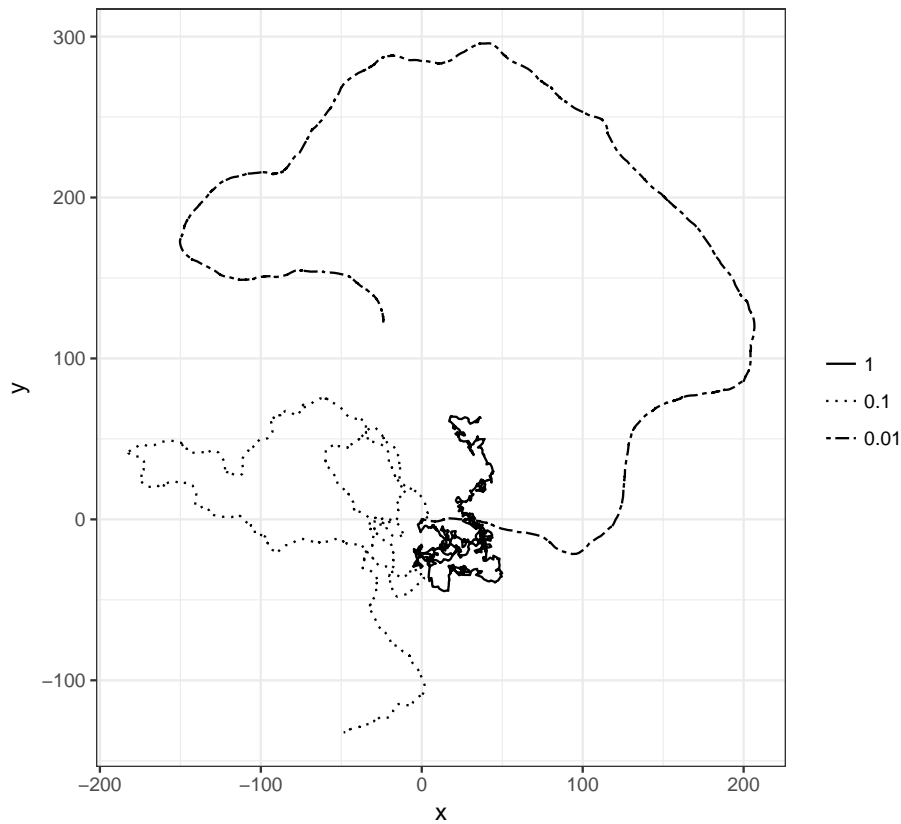


Fig. 3.1 Examples of single state path simulations. All simulations were for 1,000 time units, with identical distance parameters $\{\mu, \sigma_\phi^2\} = 1$ but with varying bearing parameters, given by $\sigma_\theta^2 = \{1, 0.1, 0.01\}$.

3.3 Fully Bayesian inference

Observations \mathbf{Z} of an animal's two-dimensional location are at a finite, but irregular, series of times \mathbf{T} . Direct inference on the movement parameters Φ is not possible as the likelihood of these observations is intractable. This is due to the non-linear relationship between the locations and the parameters when the bearing and distance processes are unobserved (see Eq. 3.4). The following describes the approach to carry out inference on the parameters, given observations, detailing the MCMC algorithm implemented.

3.3.1 Data augmentation approach

Similar to Blackwell (2003), a data augmentation approach is taken to simplify the relationship between the observations and the parameters of interest. In this approach, the augmentation is an approximation to the underlying bearing and distance processes on some (arbitrarily fine) time scale, displayed in Fig. 3.2. A directed acyclic graph (DAG) of the model with this augmented, approximate path is given in Fig. 3.3. The hybrid MCMC algorithm splits up the quantities of interest into three groups to update separately, in each case conditional on all other quantities. In the case where the full conditional distribution can be directly sampled from, Gibbs sampling is employed, and in all other scenarios the Metropolis-Hastings (MH) sampler (see below for introduction on sampling methods and Gelman et al. (2013) for full details). The groups to be separately sampled in sequence are:

- the movement parameters (split into those relating to the bearing and distance processes), and
- the unobserved refined path (consisting of bearings and distances).

In the following, Sect. 3.3.2 describes the sampling scheme used for the parameters of the movement process (bearing and distance). In both cases the sampling is standard, employing Gibbs and a random walk MH algorithm. Sect. 3.3.3 describes the MH sampling used for the reconstruction of the unobserved refined path. This includes a novel method of simulation to create independent proposals within this sampling scheme.

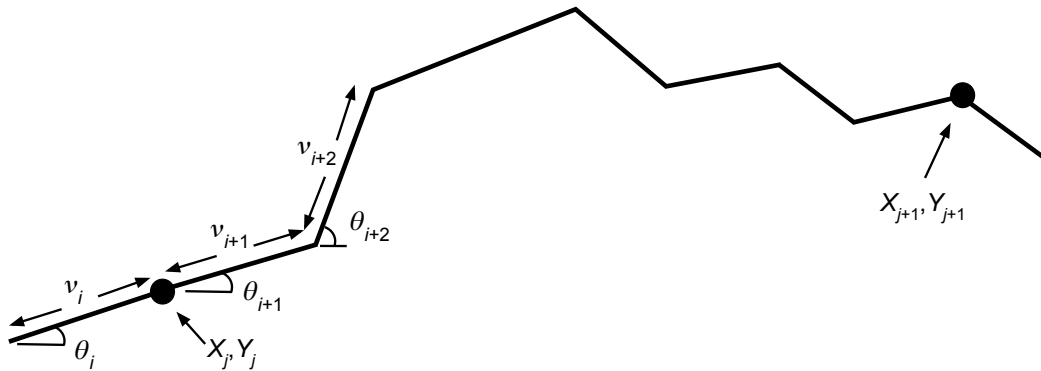


Fig. 3.2 Diagram of the augmentation approach to inference. Observed locations, (\mathbf{X}, \mathbf{Y}) , are augmented with a refined movement path, given by an approximation to the underlying path. The bearings and steps (given by the differences in distance process), $(\boldsymbol{\theta}, \mathbf{v})$, create the refined path at a time scale finer than that of the sampling scheme.

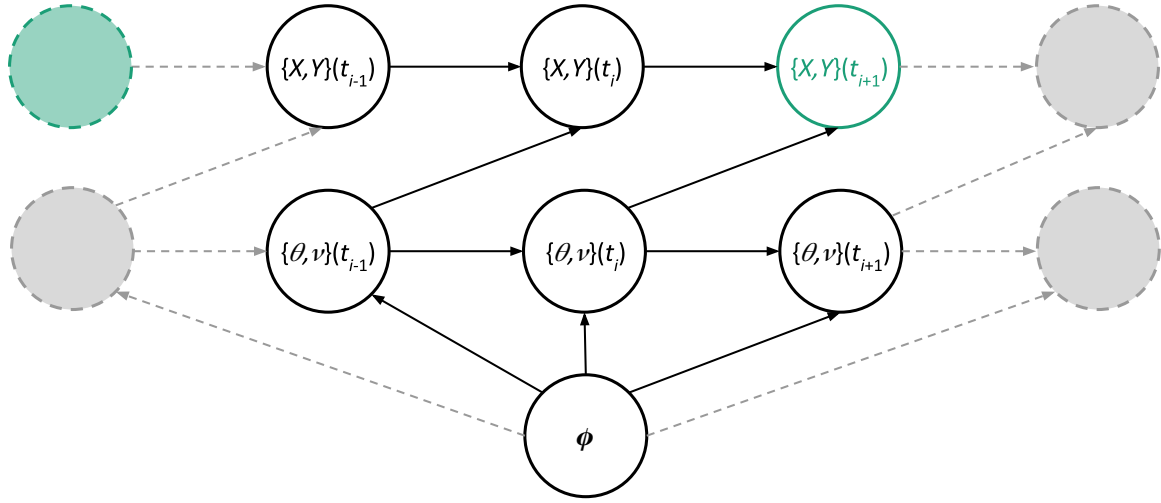


Fig. 3.3 DAG of the single state movement model with augmentation at an approximate time scale. In this representation, locations $\{X, Y\}$ are shown corresponding to the approximate time scale created for the augmentation of the bearings and steps $\{\theta, \nu\}$. Note that observed locations do not occur at each time shown here, but will be less frequently (for example, at the times shown in **green**) and the remaining, unobserved locations are augmented (deterministically through the augmented bearings and steps).

3.3.1.1 Markov chain Monte Carlo methods

Often in Bayesian statistics the posterior distribution of interest cannot be written in a closed form. However, Monte Carlo methods can be applied to samples from the posterior distribution to estimate quantities of interest, e.g. the sample mean can be used to estimate the posterior mean. MCMC methods are simulation-based techniques for estimating posterior distributions in this way by sampling from a probability distribution using a Markov chain with such an equilibrium distribution (Gelman et al., 2013).

3.3.1.1.1 The Metropolis-Hastings algorithm The MH algorithm is a popular MCMC method for sampling from a probability distribution whose density function is known up to some constant of proportionality. If $p(\mathbf{X})$ is the target distribution to sample from, and the current value of the Markov chain is given by \mathbf{X}_t , then a proposal \mathbf{Y} for the next sample in the chain \mathbf{X}_{t+1} is given by some proposal distribution $Q(\mathbf{Y}|\mathbf{X}_t)$. This proposal is then accepted with the probability given by

$$\min \left(1, \frac{p(\mathbf{Y})Q(\mathbf{X}_t|\mathbf{Y})}{p(\mathbf{X}_t)Q(\mathbf{Y}|\mathbf{X}_t)} \right). \quad (3.5)$$

If accepted, $\mathbf{X}_{t+1} = \mathbf{Y}$, otherwise $\mathbf{X}_{t+1} = \mathbf{X}_t$ (Gelman et al., 2013).

The choice of initial value and proposal distribution will affect the algorithm's efficiency at reaching the equilibrium distribution $p(\mathbf{X})$. Where possible, the proposal distribution will be chosen to be symmetric, so $Q(\mathbf{X}_t|\mathbf{Y}) = Q(\mathbf{Y}|\mathbf{X}_t)$, simplifying Eq. 3.5. This is achieved with a Normal random walk for proposals, i.e. $\mathbf{Y} \sim \text{MVN}(\mathbf{X}_t, \Sigma)$ for some covariance Σ .

3.3.1.1.2 The Gibbs sampler Gibbs sampling is the special case of the MH algorithm, in which the full conditional distribution of each dimension of \mathbf{X}_t , given by $p(X_{t,i}|\mathbf{X}_{t,-i})$, can be directly sampled from. Iterate over each dimension in turn, using the full conditional distribution as the proposal distribution, and the MH ratio reduces to 1 (accepted at every step (Gelman et al., 2013)).

The 'one dimension at a time' approach, proposing in turn from $Q_i(Y_i|\mathbf{X}_t)$ (not always equal to $p(Y_i|\mathbf{X}_{t,-i})$) and accepted/rejected based on the full conditional distribution $p(\cdot|\mathbf{X}_{t,-i})$ via the MH ratio above is known as the MH-within-Gibbs, or Metropolis-within-Gibbs, sampler.

3.3.2 Sampling the movement process parameters

Sampling the movement process parameters is carried out conditional on complete observation of the refined path (as part of the overall Gibbs sampler). The movement parameters are split into those relating to the bearing and distance processes and updated separately due to their conditional independence, described in the following.

3.3.2.1 Sampling the bearing process parameter

The bearing parameter describes the variance of BM, the conjugate distribution of which is the inverse gamma. Assuming such a conjugate prior allows direct sampling from the posterior conditional distribution as a Gibbs step, summarised in Algorithm 2. The bearings, distributed according to Eq. 3.1, can be normalised as

$$\frac{\delta\theta}{\delta t} \stackrel{\text{iid}}{\sim} \text{N}\left(0, \sigma_\theta^2\right),$$

where $\delta\theta_i = \theta_{i+1} - \theta_i$.

Assume a uniform distribution for the initial bearing. The posterior full conditional distribution is simplified as

$$p\left(\sigma_{\theta}^2 \mid \mu, \sigma_{\phi}^2, \boldsymbol{\theta}, \mathbf{v}, \mathbf{Z}\right) = p\left(\sigma_{\theta}^2 \mid \boldsymbol{\theta}\right),$$

defined as

$$\sigma_{\theta}^2 \mid \boldsymbol{\theta} \sim \text{IG}\left(a_{\theta} + \frac{m-1}{2}, b_{\theta} + \frac{1}{2} \sum_{i=1}^{m-1} \left(\frac{\delta\theta_i}{\delta t_i}\right)^2\right),$$

when assuming the conjugate prior $\sigma_{\theta}^2 \sim \text{IG}(a_{\theta}, b_{\theta})$ and where m is the number of points on the refined path. Appendix A.1.1 gives additional details showing the derivation of this posterior distribution.

Algorithm 2 Sample conditional bearing process parameter.

1: **procedure** SAMPLE BEARING PROCESS PARAMETER($\boldsymbol{\theta}, \mathbf{t}, a_{\theta}, b_{\theta}$)

2: Standardise the turns

$$\bar{\boldsymbol{\theta}} \leftarrow \frac{\delta\boldsymbol{\theta}}{\delta\mathbf{t}}$$

3: Calculate posterior shape

$$a_{\theta}^* \leftarrow a_{\theta} + \frac{m-1}{2}$$

4: Calculate posterior rate

$$b_{\theta}^* \leftarrow b_{\theta} + \frac{1}{2} \sum_{i=1}^{m-1} \bar{\theta}_i^2$$

5: Sample from conditional posterior

$$\sigma_{\theta}^2 \sim \text{IG}(a_{\theta}^*, b_{\theta}^*)$$

6: **return** σ_{θ}^2

7: **end procedure**

3.3.2.2 Sampling the distance process parameters

The parameters of the distance process are updated simultaneously using a random walk MH step with independent proposals for each parameter. Since both parameters are constrained to be positive, independent univariate Gaussians truncated below at zero are used as proposal distributions to generate the step in the random walk. As the distance process parameters are expected to be correlated with one another, dependent proposals can be used to improve performance of the algorithm through a multivariate Gaussian truncated below at zero.

The posterior conditional distribution, up to a constant, needed for the MH step is

$$\begin{aligned} p\left(\mu, \sigma_\varphi^2 \mid \sigma_\theta^2, \boldsymbol{\theta}, \mathbf{v}, \mathbf{Z}\right) &= p\left(\mu, \sigma_\varphi^2 \mid \mathbf{v}\right) \\ &\propto p\left(\mu, \sigma_\varphi^2\right) p\left(\mathbf{v} \mid \mu, \sigma_\varphi^2\right) \\ &= p\left(\mu, \sigma_\varphi^2\right) \prod_{i=1}^m p\left(v_i \mid \mu, \sigma_\varphi^2\right), \end{aligned}$$

where $p\left(\mu, \sigma_\varphi^2\right)$ is an appropriate prior probability and $p\left(v_i \mid \mu, \sigma_\varphi^2\right)$ is the probability of the refined steps, corresponding to the distribution given by Eq. 3.3. In a simultaneous update of the distance process parameters the above is calculated for both current and proposed sets and the standard MH acceptance ratio is then used to decide whether to accept the proposed parameters. Algorithm 3 summarises this.

Algorithm 3 Sample conditional distance process parameters.

- 1: **procedure** SAMPLE DISTANCE PROCESS PARAMETERS($\mu, \sigma_\varphi^2, \rho, \mathbf{v}, \mathbf{t}$)
- 2: Sample proposal distance parameters

$$\{\mu^*, \sigma_\varphi^{2*}\} \sim \text{TN}_0^\infty\left(\{\mu, \sigma_\varphi^2\}, \rho\right)$$

- 3: Evaluate proposal ratio (on the log scale)

$$Q \leftarrow \ell\left(\{\mu, \sigma_\varphi^2\} \mid \{\mu^*, \sigma_\varphi^{2*}\}, \rho\right) - \ell\left(\{\mu^*, \sigma_\varphi^{2*}\} \mid \{\mu, \sigma_\varphi^2\}, \rho\right)$$

- 4: Evaluate posterior ratio (on the log scale)

$$L \leftarrow \ell\left(\mu^*, \sigma_\varphi^{2*}\right) + \ell\left(\mathbf{v} \mid \mu^*, \sigma_\varphi^{2*}\right) - \ell\left(\mu, \sigma_\varphi^2\right) - \ell\left(\mathbf{v} \mid \mu, \sigma_\varphi^2\right)$$

- 5: **if** $U(0, 1) < Q + L$ **then**
 - 6: **return** Accept proposal, $\mu^*, \sigma_\varphi^{2*}$
 - 7: **else**
 - 8: **return** Reject proposal, μ, σ_φ^2
 - 9: **end if**
 - 10: **end procedure**
-

3.3.3 Reconstructing the unobserved refined path

The key step in this inference algorithm is to sample the unobserved refined path, given by the bearing and distance/step processes at a refined time scale, conditional on the movement

parameters. As the full movement path will be large, this step is carried out in short sections chosen randomly from the complete path. The aim is to simulate the refined path between two consecutive observation times j and k , conditional on the fixed path outside of these times and parameters. Such a scenario is displayed in Fig. 3.4. This method can be extended to span multiple observed locations, and to start/end at points on the refined path that are not observation times, but those cases are omitted here for simplicity in notation (though are implemented in all application examples).

The quantities to simulate are those displayed in **black** in Fig. 3.4, consisting of:

- the bearings $\{\theta_1, \dots, \theta_{n-1}\}$, and
- the steps $\{v_1, \dots, v_{n-1}\}$.

The set of refined times $\{t_1 = j, \dots, t_{n-1}\}$ that the bearing and step processes are approximated over can be chosen and fixed before inference and need not be regular. For a general discussion on the choice of δt in an analysis, see Sect. 3.6. The fixed values that are to be conditioned upon are displayed in **green** in Fig. 3.4, consisting of:

- the observed locations $\{\mathbf{Z}(j), \mathbf{Z}(k)\}$ (at the times $\{j, k\}$),
- the bearings $\{\theta_0, \theta_n\}$ at the times $\{t_0, t_n = k\}$, and
- the steps $\{v_0, v_n\}$ at the times $\{t_0, t_n\}$.

As the bearing and step processes are given by an approximation, the fixed points are the values at the refined point immediately before and after the path section of interest, as can be seen in Fig. 3.4.

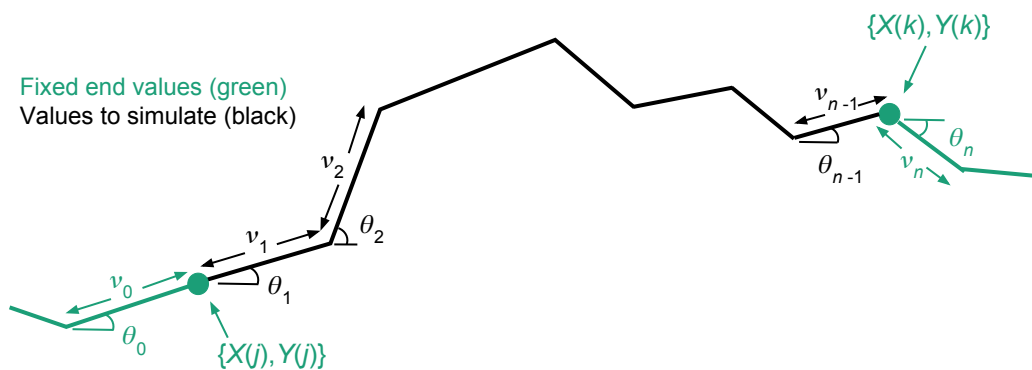


Fig. 3.4 Diagram of the section of path over which augmentation with a refined movement path will occur. Reconstruction is between two consecutive observations, at the times j and k (**points**). The bearing and step processes, $\{\theta_1, \dots, \theta_{n-1}, v_1, \dots, v_{n-1}\}$, between the observations (**black**) are simulated given fixed endpoints $\{\theta_0, \theta_n, v_0, v_n\}$ (**green**).

Simulating the quantities of interest from the full conditional distribution is not possible due to the non-linearity of the quantities conditioned upon (i.e. the fixed location $\mathbf{Z}(k)$ is a non-linear sum of the bearings and steps, as given in Eq. 3.4). A path section is proposed from a simpler distribution that is then accepted or rejected using an MH ratio. An independence sampler is employed using a novel simulation method to propose a new path section, described in the following. This method proposes a path section (in part) conditional on the observed locations, and so is more efficient than simulating purely based on parameters (such as the simulation of Sect. 3.2).

3.3.3.1 Simulating a refined path proposal

The path proposal begins by simulating the approximate bearing process $\boldsymbol{\theta}^*$ over the times $\{t_1, \dots, t_{n-1}\}$, conditional upon the fixed bearings $\{\boldsymbol{\theta}_0, \boldsymbol{\theta}_n\}$ at the times $\{t_0, t_n\}$ and the parameters $\boldsymbol{\Phi}$. This conditioned process is a BB with variance σ_θ^2 . The BB is simulated (Iacus, 2008) as

$$\theta_i^* = \theta_0 + M(t_i - t_0) - \frac{t_i - t_0}{t_n - t_0} (M(t_n - t_0) - \theta_n + \theta_0), \quad (3.6)$$

for $i \in \{1, \dots, n-1\}$, where $M(t)$ is BM with variance σ_θ^2 (see Algorithm 4).

The path proposal is completed by simulating an approximate step process \mathbf{v}^* . Unlike the bearings, this proposal will consider the observed locations to provide efficiency against the

Algorithm 4 Simulate bearing proposal.

1: **procedure** SIMULATE BEARING PROPOSAL($\{t_0, \dots, t_n\}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_n, \sigma_\theta^2$)

2: $M_1 \sim \text{N}\left(0, (t_1 - t_0)\sigma_\theta^2\right)$

3: **for** $i \in 2, \dots, n$ **do**

4: Sample BM

$$M_i \sim \text{N}\left(M_{i-1}, (t_i - t_{i-1})\sigma_\theta^2\right)$$

5: **end for**

6: **for** $i \in 1, \dots, (n-1)$ **do**

7: Correct to the BB

$$\theta_i^* \leftarrow \theta_0 + M_i - \frac{t_i - t_0}{t_n - t_0} (M_n - \theta_n + \theta_0)$$

8: **end for**

9: **return** $\boldsymbol{\theta}^*$

10: **end procedure**

simulation method of Sect. 3.2. First construct the joint distribution

$$\begin{pmatrix} \mathbf{v} \\ \mathbf{Z}(k) \end{pmatrix} \mid \Phi, \boldsymbol{\theta}^*, \mathcal{F} \sim \mathbf{N} \left(\begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2}^\top & \Sigma_2 \end{pmatrix} \right), \quad (3.7)$$

where $\mathcal{F} = \{\mathbf{Z}(j), \theta_0, \theta_n, v_0, v_n\}$. In the above, the marginal distribution of the steps is $\mathbf{v} \sim \mathbf{N}(\mathbf{m}_1, \Sigma_1)$. The distance process assumed here results in steps that are independent through time, and so the distribution of \mathbf{v} is unaffected by conditioning on $\{v_1, v_n\}$. Hence, $\boldsymbol{\mu}_1$ and Σ_1 are constructed from that in Eq. 3.3, with Σ_1 being a diagonal matrix.

In Eq. 3.7 the marginal distribution of the fixed end location is $\mathbf{Z}(k) \sim \mathbf{N}(\mathbf{m}_2, \Sigma_2)$, and the covariance between \mathbf{v} and $\mathbf{Z}(k)$ is $\Sigma_{1,2}$. The fixed end location can be written as

$$\begin{aligned} \mathbf{Z}(k) &= \mathbf{Z}(j) + \sum_{i=1}^{n-1} \begin{pmatrix} v_i \cos(\theta_i^*) \\ v_i \sin(\theta_i^*) \end{pmatrix} \\ &= \mathbf{Z}(j) + A\mathbf{v}, \end{aligned}$$

where

$$A = \begin{pmatrix} \cos(\theta_1^*) & \cdots & \cos(\theta_{n-1}^*) \\ \sin(\theta_1^*) & \cdots & \sin(\theta_{n-1}^*) \end{pmatrix}.$$

As $\mathbf{Z}(k)$ is a linear combination of the steps when the bearings are known, the means and covariances in Eq. 3.7 are

$$\begin{aligned} \mathbf{m}_2 &= \mathbf{Z}(j) + A\mathbf{m}_1, \\ \Sigma_{1,2} &= \Sigma_1 A^\top, \\ \Sigma_2 &= A\Sigma_1 A^\top. \end{aligned}$$

The step process proposal is simulated by further conditioning \mathbf{v} in Eq. 3.7 upon the location $\mathbf{Z}(k)$. This ensures that the proposed path section will ‘meet up’ with the fixed end location and ‘agree’ with the full movement path. The sought conditional distribution can be calculated by standard formulae for conditioning of a normal distribution (Eaton, 2007)

$$\begin{aligned} \mathbf{v} \mid \Phi, \boldsymbol{\theta}^*, \mathcal{F}, \mathbf{Z}(k) &\sim \mathbf{N} \left(\mathbf{m}_1 + \Sigma_{1,2} \Sigma_2^{-1} (\mathbf{Z}(k) - \mathbf{m}_2), \Sigma_1 - \Sigma_{1,2} \Sigma_2^{-1} \Sigma_{1,2}^\top \right) \\ &= \mathbf{N} \left(\hat{\boldsymbol{\mu}}, \hat{\Sigma} \right). \end{aligned} \quad (3.8)$$

This distribution is singular, with the rank of $\hat{\Sigma}$ reduced from full by two through conditioning upon both a known X and Y value. Simulation from a singular distribution can be carried out following the ‘conditioning by Kriging’ method of Rue and Held (2005) by first sampling from the marginal distribution

$$\mathbf{x}^* \sim \mathbf{N}(\mathbf{m}_1, \Sigma_1),$$

and then accounting for the linear constraint by setting

$$\mathbf{v}^* = \mathbf{x}^* - \Sigma_{1,2}\Sigma_2^{-1}(\mathbf{Z}(j) + A\mathbf{x}^* - \mathbf{Z}(k)). \quad (3.9)$$

Algorithm 5 summarises the procedure for proposing the step process. Appendix A.2.1 shows that the method for sampling \mathbf{v}^* is equivalent to sampling from the distribution in Eq. 3.8.

Algorithm 5 Simulate step proposal.

- 1: **procedure** SIMULATE STEP PROPOSAL($\{t_1, \dots, t_n\}, \mathbf{v}_0, \mathbf{v}_n, \Phi, \mathbf{Z}(j), \mathbf{Z}(k), \theta^*$)
- 2: **for** $i \in 1, \dots, (n-1)$ **do**
- 3: Construct unconditional step distribution

$$\begin{aligned} m_{1,i} &\leftarrow (t_{i+1} - t_i)\mu \\ \Sigma_{1,i,i} &\leftarrow (t_{i+1} - t_i)\sigma_\varphi^2 \\ \Sigma_{1,i,j \neq i} &\leftarrow 0 \end{aligned}$$

- 4: Construct linear constraint matrix for fixed end location

$$A_{\cdot,i} \leftarrow \begin{pmatrix} \cos(\theta_i^*) & \sin(\theta_i^*) \end{pmatrix}^\top$$

- 5: **end for**
- 6: Construct conditional distribution

$$\begin{aligned} \mathbf{m}_2 &\leftarrow \mathbf{Z}(j) + A\mathbf{m}_1 \\ \Sigma_2 &\leftarrow A\Sigma_1A^\top \\ \Sigma_{1,2} &\leftarrow \Sigma_1A^\top \end{aligned}$$

- 7: Sample unconditional step distribution $\mathbf{x}^* \sim \mathbf{N}(\mathbf{m}_1, \Sigma_1)$
 - 8: Adjust for constraint $\mathbf{v}^* \leftarrow \mathbf{x}^* - \Sigma_{1,2}\Sigma_2^{-1}(\mathbf{Z}(j) + A\mathbf{x}^* - \mathbf{Z}(k))$
 - 9: **return** \mathbf{v}^*
 - 10: **end procedure**
-

3.3.3.2 Simulating a proposal at the start/end of path

An exception to the given path proposal method occurs when the section is at the start or end of the full movement path. In such a case, there is no fixed bearing or step at the start or the end of the section, respectively. Rather than simulating a bearing proposal from a BB, the simulation is BM with a fixed starting point forwards/backwards in time when at the start/end of the path, respectively. As the step process is independent over disjoint periods of time, the lack of an endpoint does not affect the step proposal.

3.3.3.3 Accepting a refined path proposal

The method given for simulating a path section proposal does not take into account the fixed location at the end of the section when proposing the bearing process. An MH step is required to assess whether this proposal is accepted. The conditional distribution, up to a constant, is

$$\begin{aligned} p(\boldsymbol{\theta}^*, \mathbf{v}^* | \Phi, \mathcal{F}, \mathbf{Z}(k)) &= p(\boldsymbol{\theta}^* | \Phi, \mathcal{F}, \mathbf{Z}(k)) p(\mathbf{v}^* | \boldsymbol{\theta}^*, \Phi, \mathcal{F}, \mathbf{Z}(k)) \\ &\propto p(\boldsymbol{\theta}^* | \Phi, \mathcal{F}) p(\mathbf{Z}(k) | \boldsymbol{\theta}^*, \Phi, \mathcal{F}) p(\mathbf{v}^* | \boldsymbol{\theta}^*, \Phi, \mathcal{F}, \mathbf{Z}(k)). \end{aligned}$$

The proposal method has distribution proportional to

$$p(\boldsymbol{\theta}^* | \Phi, \mathcal{F}) p(\mathbf{v}^* | \boldsymbol{\theta}^*, \Phi, \mathcal{F}, \mathbf{Z}(k)).$$

The MH acceptance ratio is then

$$\frac{p(\mathbf{Z}(k) | \boldsymbol{\theta}^*, \Phi, \mathcal{F})}{p(\mathbf{Z}(k) | \boldsymbol{\theta}, \Phi, \mathcal{F})},$$

where

$$\mathbf{Z}(k) | \boldsymbol{\theta}^*, \Phi, \mathcal{F} \sim N(\mathbf{m}_2, \Sigma_2).$$

and \mathbf{m}_2, Σ_2 must be constructed in each case given the corresponding set of bearings (current and proposed).

3.4 Movement with correlated speed

The movement model described in Sect. 3.1 provides an analogue to discrete-time models, in which the distance an animal travels over disjoint time periods is uncorrelated. A number of

authors have suggested the need for correlation in movement processes (see e.g. Gurarie et al. (2016)) and so a more realistic model of movement would assume that these distances are correlated, with the magnitude of such dependence decaying over time. In such a case, rather than modelling the distance process of the animal, it is prudent to describe a speed process.

3.4.1 Speed process

At time $t \geq 0$ the speed process $\psi(t)$ is modelled by the SDE

$$d\psi(t) = F_5(t, \psi(t)) dt + F_6(t, \psi(t)) dW(t),$$

with F_i , for $i = \{5, 6\}$, known functions. This SDE is assumed to be a one-dimensional OU process (Iacus, 2008), with parameters μ , β , σ_ψ^2 , and

$$\begin{aligned} F_5(t, \psi(t)) &= \beta(\mu - \psi(t)), \\ F_6(t, \psi(t)) &= \sqrt{2\beta} \sigma_\psi. \end{aligned}$$

Note the σ_ψ^2 here is the $\frac{\sigma^2}{2\beta}$ in Sect. 2.1.2. The animal's speed is stochastic but correlated, with a long term average μ and long term variance σ_ψ^2 . As in the independent distance model, there is a positive probability of negative speeds, however this can again be made negligible by placing a constraint on the dependence between the set of speed parameters.

3.4.2 Joint bearing and speed process

Given the bearing and speed at time $t \geq 0$, the location process \mathbf{Z} is

$$\begin{aligned} dX(t) &= \psi(t) \cos(\theta(t)), \\ dY(t) &= \psi(t) \sin(\theta(t)). \end{aligned}$$

3.4.3 Simulating movement

The speed process can be simulated at a refined time scale δt as

$$\psi(t + \delta t) \mid \psi(t) \sim N\left(\mu + e^{-\beta\delta t} (\psi(t) - \mu), \sigma_\psi^2 (1 - e^{-2\beta\delta t})\right).$$

As in the independent distance model, it becomes useful to define the step process

$$\begin{aligned} v(t + \delta t) | v(t) = \psi(t + \delta t)\delta t | \psi(t)\delta t \\ \sim \text{N} \left(\delta t \left(\mu + e^{-\beta\delta t} \left(\frac{v(t)}{\delta t} - \mu \right) \right), \delta t^2 \sigma_\psi^2 \left(1 - e^{-2\beta\delta t} \right) \right), \end{aligned} \quad (3.10)$$

highlighting the correlation of this process.

Path simulation is given by that in Algorithm 1, but replacing line 4 with the equilibrium distribution

$$v(t_0) \sim (t_1 - t_0)\text{N}(\mu, \sigma_\psi^2),$$

and line 6 with $v(t_i)|v(t_{i-1})$ in Eq. 3.10.

3.4.4 Fully Bayesian inference

The following outlines changes to the inference algorithm of Sect. 3.3 to allow for correlated speed.

3.4.4.1 Sampling the movement process parameters

Inference for the set of movement parameters $\Phi = \{\sigma_\theta^2, \mu, \beta, \sigma_\psi^2\}$, given observed locations \mathbf{Z} , follows Sect. 3.3.2. Sampling the bearing process parameter remains identical to that of Sect. 3.3.2.1.

The parameters of the speed process are updated simultaneously using a random walk MH step, with independent proposals for each parameter, as in Sect. 3.3.2.2. All parameters are constrained to be positive, and so univariate Gaussians truncated below at zero are used as proposal distributions to generate the step in the random walk.

The posterior conditional distribution, up to a constant, is

$$p(\mu, \beta, \sigma_\psi^2 | \sigma_\theta^2, \boldsymbol{\theta}, \mathbf{v}, \mathbf{Z}) \propto p(\mu, \beta, \sigma_\psi^2) p(v_1 | \mu, \sigma_\psi^2) \prod_{i=2}^m p(v_i | v_{i-1}, \mu, \beta, \sigma_\psi^2),$$

where

$$v_1 | \mu, \sigma_\psi^2 \sim \delta_{t_1}\text{N}(\mu, \sigma_\psi^2),$$

and $v_i | v_{i-1}, \mu, \beta, \sigma_\psi^2$ is distributed as in Eq. 3.10.

3.4.4.2 Reconstructing the unobserved refined path

The method for sampling a section of refined path when speeds are correlated is that in Sect. 3.3.3. The only difference is the form of the marginal distribution of the steps (conditional on fixed end steps), introduced in Eq. 3.7. The marginal distribution of \mathbf{v} is no longer unaffected by the fixed endpoints once there is correlation. The steps are derived from the underlying speed process, which is an OU bridge, and the construction of this distribution is given in the following.

The fixed values v_0, v_n are transformed to the speeds $\psi_0 = v_0/\delta t_0$ and $\psi_n = v_n/\delta t_n$. In the following, assume (for brevity) that all expressions are also conditional upon

$$\{\Phi, \theta^*, \mathbf{Z}(j), \theta_0, \theta_n\}.$$

By noting that

$$\psi_i | \psi_{i-1} = e^{-\beta(t_i - t_{i-1})} \psi_{i-1} + \mathbf{N} \left(\mu \left(1 - e^{-\beta(t_i - t_{i-1})} \right), \sigma_\psi^2 \left(1 - e^{-2\beta(t_i - t_{i-1})} \right) \right),$$

the joint distribution $\{\psi_1, \dots, \psi_{n-1}, \psi_n\} | \psi_0$ can be constructed iteratively as

$$\begin{aligned} \mathbf{E}(\psi_i | \psi_0) &= e^{-\beta(t_i - t_{i-1})} \mathbf{E}(\psi_{i-1} | \psi_0) + \mu \left(1 - e^{-\beta(t_i - t_{i-1})} \right), \\ \mathbf{Var}(\psi_i | \psi_0) &= e^{-2\beta(t_i - t_{i-1})} \mathbf{Var}(\psi_{i-1} | \psi_0) + \sigma_\psi^2 \left(1 - e^{-2\beta(t_i - t_{i-1})} \right), \\ \mathbf{Cov}(\psi_i, \psi_{i+j} | \psi_0) &= e^{-\beta(t_{i+j} - t_{i+j-1})} \mathbf{Cov}(\psi_i, \psi_{i+j-1} | \psi_0). \end{aligned} \quad (3.11)$$

This joint distribution is partitioned into $\{\psi_1, \dots, \psi_{n-1}\}$ and ψ_n in order to condition upon the known value of ψ_n (by the same method for conditioning upon a normal distribution given in Eq. 3.8 by Eaton (2007)) to give the bridge distribution $\{\psi_1, \dots, \psi_{n-1}\} | \psi_0, \psi_n$. This can be transformed back to the steps $\{v_1, \dots, v_{n-1}\}$ to give \mathbf{m}_1, Σ_1 by multiplying the speeds $\{\psi_1, \dots, \psi_{n-1}\}$ by the time increments $\{\delta t_1, \dots, \delta t_{n-1}\}$. The conditional step proposal is then carried out as in Algorithm 5, but replacing line 3 with that of Algorithm 6.

When simulating a proposal for a section at the start or end of the path, rather than an OU bridge, the marginal distribution of \mathbf{v} is constructed by an OU process either forwards or backwards in time from the fixed point. This process is the same as that above, but without the need to condition upon the endpoint ψ_n .

Algorithm 6 Correlated step bridge distribution.

1: **procedure** CONSTRUCT OU BRIDGE($\{t_1, \dots, t_n\}, v_0, v_n$)

2: Transform end steps to speeds

$$\psi_0 \leftarrow v_0 / (t_1 - t_0)$$

$$\psi_n \leftarrow v_n / (t_{n+1} - t_n)$$

3: **for** $i \in 1, \dots, n$ **do**

4: Conditional mean and variance of OU process

$$E(\psi_i | \psi_0) \leftarrow \left(\mu + e^{-\beta(t_i - t_{i-1})} (E(\psi_{i-1} | \psi_0) - \mu) \right)$$

$$\text{Var}(\psi_i | \psi_0) \leftarrow \text{Var}(\psi_{i-1} | \psi_0) e^{-2\beta(t_i - t_{i-1})} + \sigma_\psi^2 \left(1 - e^{-2\beta(t_i - t_{i-1})} \right)$$

5: **for** $j \in 1, \dots, n$ **do**

6:

$$\text{Cov}(\psi_i, \psi_j | \psi_0) \leftarrow e^{-\beta(t_j - t_{j-1})} \text{Cov}(\psi_i, \psi_{j-1} | \psi_0)$$

7: **end for**8: **end for**

9: Condition to construct OU bridge distribution

$$E(\psi_{1:(n-1)} | \psi_0, \psi_n) \leftarrow E(\psi_{1:(n-1)} | \psi_0) + \text{Cov}(\psi_{1:(n-1)}, \psi_n | \psi_0) \text{Var}(\psi_n | \psi_0)^{-1} (\psi_n - E(\psi_n | \psi_0))$$

$$\text{Cov}(\psi_{1:(n-1)}, \psi_{1:(n-1)} | \psi_0, \psi_n) \leftarrow \text{Cov}(\psi_{1:(n-1)}, \psi_{1:(n-1)} | \psi_0) \text{Var}(\psi_n | \psi_0)^{-1} \text{Cov}(\psi_n, \psi_{1:(n-1)} | \psi_0)$$

10: Transform speed distribution to step

$$m_1 \leftarrow \delta t E(\psi_{1:(n-1)} | \psi_0, \psi_n)$$

$$\Sigma_1 \leftarrow \delta t \text{Cov}(\psi_{1:(n-1)}, \psi_{1:(n-1)} | \psi_0, \psi_n) \delta t^T$$

11: **return** m_1, Σ_1 12: **end procedure**

3.5 Examples with simulated data

Two example implementations of the inference algorithm are presented in this section, using data simulated with both independent and correlated steps. Both cases demonstrate the ability of the inference algorithm to correctly estimate the underlying movement parameters and compare such estimates with those obtained by using only the observed locations (as close as possible to a ‘discrete-time style’ analysis with this approach). A marked improvement in estimation is made by using the augmentation approach described in Sect. 3.3. In both examples the movement parameters were chosen to reflect the types of movement patterns observed in real data—in particular, the second example assumes a high level of volatility in the bearing and speed processes to demonstrate the presented methods in a scenario of high uncertainty in location between observations.

3.5.1 Simulated movement example with independent speed process

This first example provides a demonstration of the methods for inference presented in this chapter, using simulated data from the model with an independent distance process. The following results can be comparably reproduced with the `single_simulation` example within the repository `CTStepTurn` available at <https://github.com/a-parton/CTStepTurn>.

3.5.1.1 Underlying movement and observations

We simulated (following Sect. 3.2) an example movement path with the underlying model in Sect. 3.1. The movement parameters were

$$\Phi = \{\sigma_\theta^2 = 0.5, \mu = 50, \sigma_\varphi^2 = 30\}, \quad (3.12)$$

and the path simulated at the approximate time scale $\delta t = 0.01$, between the times 0–500. Such a time scale was chosen as a close approximation to the underlying continuous-time movement model.

We created observations of the movement path by sub-sampling the simulated locations at intervals of $500\delta t$, giving 101 ‘observed’ locations at a time scale of 5. The simulated path, with corresponding observations, is given in Fig. 3.5. The level of sub-sampling used here was chosen to create observations that were sufficiently often so that the overall movement

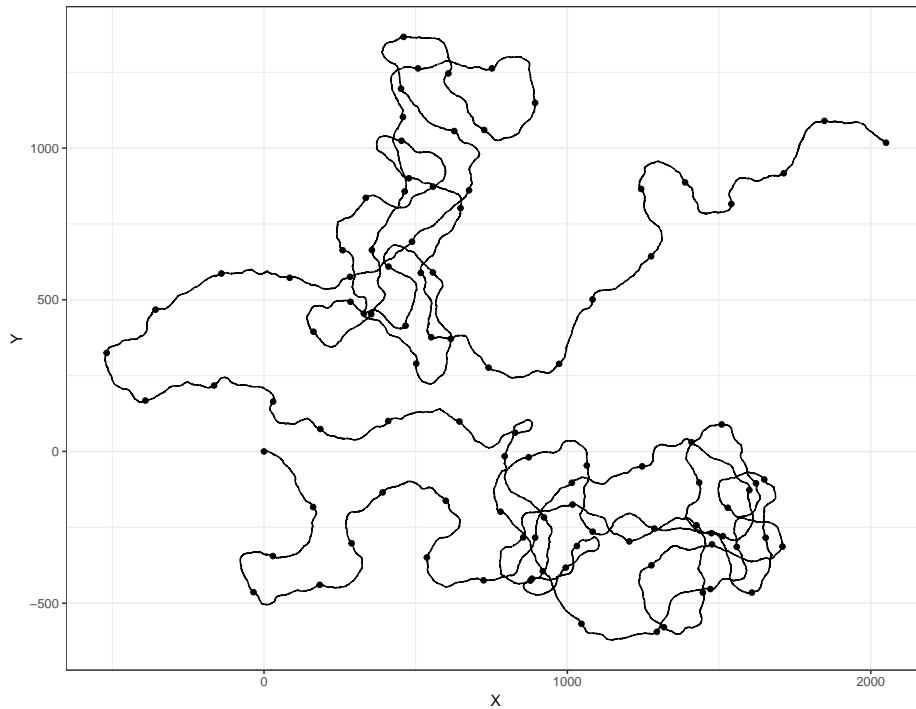


Fig. 3.5 The single state, independent step simulated movement path (**line**). Locations to be used as observations are included (**points**).

can be determined by the observations alone, whilst remaining sparse enough so as to allow a reasonable level of uncertainty as to the unobserved movement between observations.

3.5.1.2 Implementing the inference algorithm

To explore the ability of the inference algorithm presented in Sect. 3.3, this example includes a number of implementations at a range of time scales for the unobserved refined movement path.

- Although the true values of the movement parameters used to create the simulated path are known (Eq. 3.12), the posterior distribution of the parameters for the resulting simulated path is not. To learn about this distribution, a ‘baseline’ implementation was carried out assuming full observation of the simulated movement path at the refined time scale of 0.01. In this case, no augmentation is required and the movement parameters are sampled according to Sect. 3.3.2.
- To investigate the performance of the full inference algorithm, two example implementations were carried out with refined time scales of 0.5 and 1 (augmenting 9 and 4

locations between each pair of observations, respectively). For a general discussion on the choice of δt in an analysis, see Sect. 3.6.

- To compare the performance of the algorithm with cases where augmentation is not used, an implementation with a refined time scale of 5 (the same as the observations) was carried out. This case is similar to a discrete-time analysis, where the movement path is treated as fully observed and the observed locations are linearly interpolated to calculate turning angles and step lengths. As in the ‘baseline’ case above, inference only involves the movement parameters.

3.5.1.2.1 Initial values The two cases where full inference was implemented require an initial refined movement path, in both cases constructed by taking an interpolating cubic spline between the observed locations at the associated refined time scale. Bearing and step lengths were calculated from these interpolated locations. Using the spline to create the initial path is ideal as it requires no prior knowledge of the unobserved movement path and is in some sense an ‘extreme’ initial path because of its smoothness, showcasing the ability of the algorithm to infer the true movement parameters.

In all cases, initial movement parameters were set as maximum likelihood estimates from the fixed/initial movement path, given in Table 3.1. When the full simulated path is known, the initial movement parameters are similar to the true values, as expected. In all other cases the initial movement parameters underestimate the bearing variance and overestimate the speed variance.

Extracts of the movement paths (either fixed when fully observed or initial when reconstruction is required) are given in Fig. 3.6 as locations and in Fig. 3.7 as bearings and speeds. The similarity of the two initial paths created using interpolating cubic splines can be seen here,

Table 3.1 Initial parameters, perturbation variances and acceptance rates for the implementations carried out on the independent step simulation.

δt	$\Phi^{(0)}$			ρ		Acceptance (%)	
	$\sigma_{\theta}^{2(0)}$	$\mu^{(0)}$	$\sigma_{\varphi}^{2(0)}$	ρ_{μ}	$\rho_{\sigma_{\varphi}^2}$	$\{\mu, \sigma_{\varphi}^2\}$	$\{\theta, \mathbf{v}\}$
0.01 (simulation)	0.500	49.9	29.9	0.02	0.5	29	-
0.5	0.0563	42.4	39.4	0.2	5	37	3
1	0.107	42.1	78.2	0.2	5	40	1
5 (observation)	0.327	40.6	236	1	500	53	-

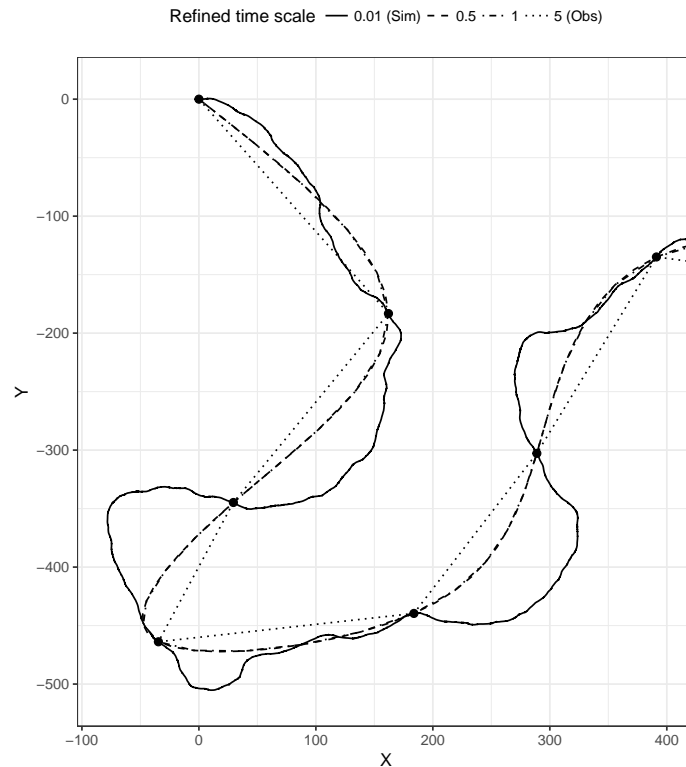


Fig. 3.6 Small section (between the times 0–30) of the initial movement path in the single state, independent steps simulation example. When the refined time scale was the same as the observations or simulations, the movement path was known and fixed. When the refined time scale was 1 or 0.5, the movement path was reconstructed, with initial path given here, created using an interpolating spline between observed locations. The smoothness of the interpolating spline results in two paths that are almost indistinguishable here.

as well as the ‘smoothing’ of the processes that occurs as the refined time scale becomes more coarse.

3.5.1.2.2 Prior information In all four implementations of the algorithm, the movement parameters were assumed to have flat, uninformative prior distributions. In the case of the bearing variance, the conjugate prior distribution was an inverse gamma with shape and rate 0.001.

3.5.1.2.3 Implementation For the two implementations in which path reconstruction was not needed (at the simulation and observation time scales), movement parameters were sampled 10^5 times and thinned by a factor of 100 to store 1,000 samples. For the two

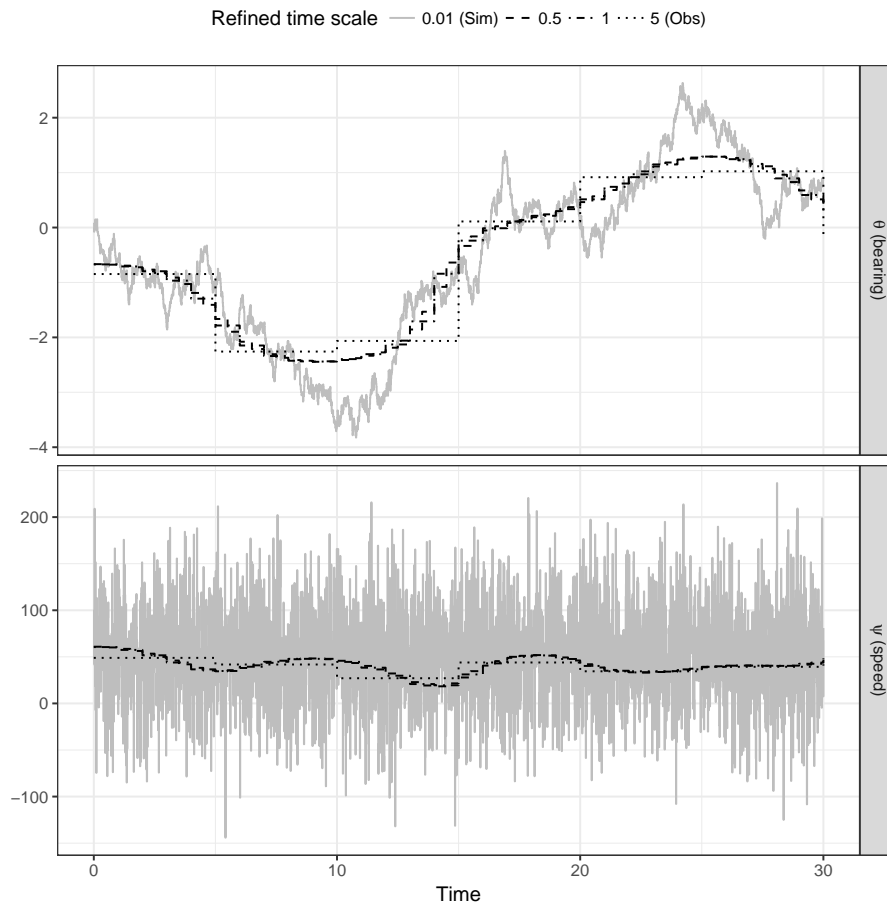


Fig. 3.7 Small section (between the times 0–30) of the initial movement path (as bearings and speeds) in the single state, independent steps simulation example. When the refined time scale was the same as the observations or simulation, the movement path was known and fixed. When the refined time scale was 1 or 0.5, the movement path was reconstructed, with initial path given here, created using an interpolating spline between observed locations. The smoothness of the interpolating spline results in two paths that are almost indistinguishable here.

implementations in which path reconstruction was necessary, movement parameters were sampled 4×10^6 times and thinned by a factor of 1,000 to store 4,000 samples. Note that thinning need only be done for memory storage purposes, improving the computational run-time. At the observation and simulation time scales, the computational run-time was at the order of minutes and hours, respectively. For the examples with path reconstruction, computational time took around 24 hours. Perturbation variances for the independent truncated Gaussian proposal in the random walk on the distance parameters were set based on pilot experiments, given in Table 3.1.

The method for path reconstruction in Sect. 3.3.3 describes reconstruction between two consecutive observations. In this example, the method employed is an extension to this, in which a reconstruction can span multiple observations, and can start and end at any point on the refined path. The reconstructions were carried out in random sections of lengths between 4–13 points on the refined time scale, with 100 reconstructions for each sample of the movement parameters. With regards the choice of length for path reconstructions, a long length will increase the proportion of the path that is being updated at a time and is obviously desirable. However, longer sub-path lengths incur computational costs through larger matrix calculations and are less likely to be accepted due to their high dimensionality. Having a mixture of short sub-path lengths that are easily accepted helps with the mixing of the algorithm. This idea follows on from such a discussion regarding reconstructing a behavioural process in Blackwell et al. (2015). The particular choice used here was based on the acceptance rates in pilot runs: lengths higher than 13 had too low acceptance to be feasible, and lengths of 4 allowed these short section updates that helped with mixing.

Acceptance rates of the distance parameters (note that the bearing parameter is updated via Gibbs sampling) and path reconstructions (where applicable) are given in Table 3.1. The implementations with full observations of the movement path (at time scales of 0.01 and 5) required fewer MCMC iterations because path reconstruction was not needed. In both these cases, the thinned samples of size 1,000 had an effective sample size (ESS) of 1,000 as autocorrelation was so low (the ACF is included in the Appendix, Fig. B.3). Longer MCMC samplers were required for the two implementations in which paths were reconstructed because of slow mixing. This is to be expected when such a Gibbs sampler is implemented due to the high correlation between a movement path and its associated parameters. In both these cases, the thinned samples of 4,000 parameters had ESSs of at least 100 (actual ESS for each parameter is given in Table 3.2 and the ACF is included in the Appendix, Fig. B.3). Acceptance rates of path reconstructions were 3% and under in both implementations, but due to the high dimensionality of the movement path, these rates were deemed acceptable for this example: in both implementations the ESS of the 100 stored reconstructed locations was 100 for 70% of the points on the path, with only 10 locations having an ESS less than 50. A discussion on extensions to the inference method that aim to increase path reconstruction acceptance is given in Sect. 3.6.

3.5.1.3 Movement parameter results

The sampled movement parameters for all implementations are given as trace plots in the Appendix, Fig. B.2, with the true value highlighted. The Heidelberger and Welch diagnostic

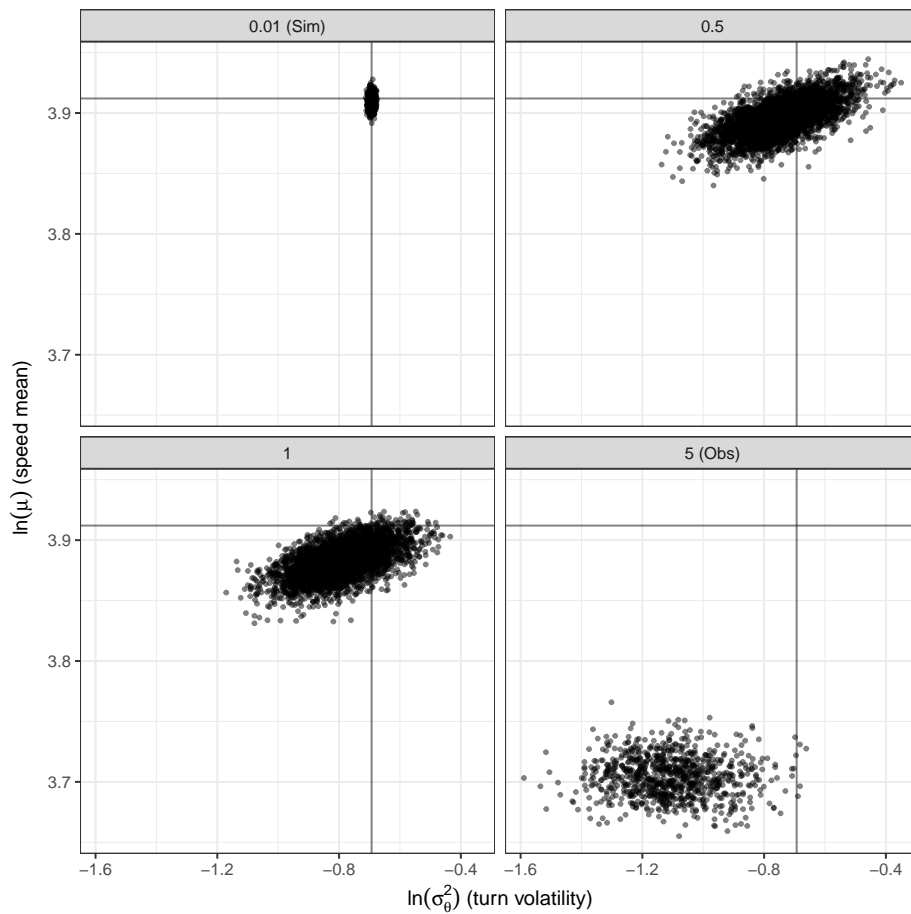


Fig. 3.8 Sampled (log) bearing variance against mean speed (thinned and with burn-in time omitted) for the single state, independent step simulation (**points**). True parameter values are highlighted (**lines**).

was used to evaluate convergence of the implementations (Heidelberger and Welch, 1983). This is a test that recursively discards 10% of the chain (until 50% has been discarded) and uses the Cramer-von-Mises criterion to test the null hypothesis that the chain is from a stationary process based on the mean of the process. In this example, all implementations passed Heidelberger and Welch convergence diagnostics (implemented using the coda package (Plummer et al., 2016)), and the ESS for each parameter is given in Table 3.2. The trace plots highlight the high autocorrelation in σ_φ^2 , which describes the variability of the distance process. Correlation between σ_θ^2 , the variability of the bearing process, and μ , the mean drift of the distance process, can also be seen. This is to be expected given a fixed set of observations; the longer a path is between two fixed locations the more tortuous that path must be.

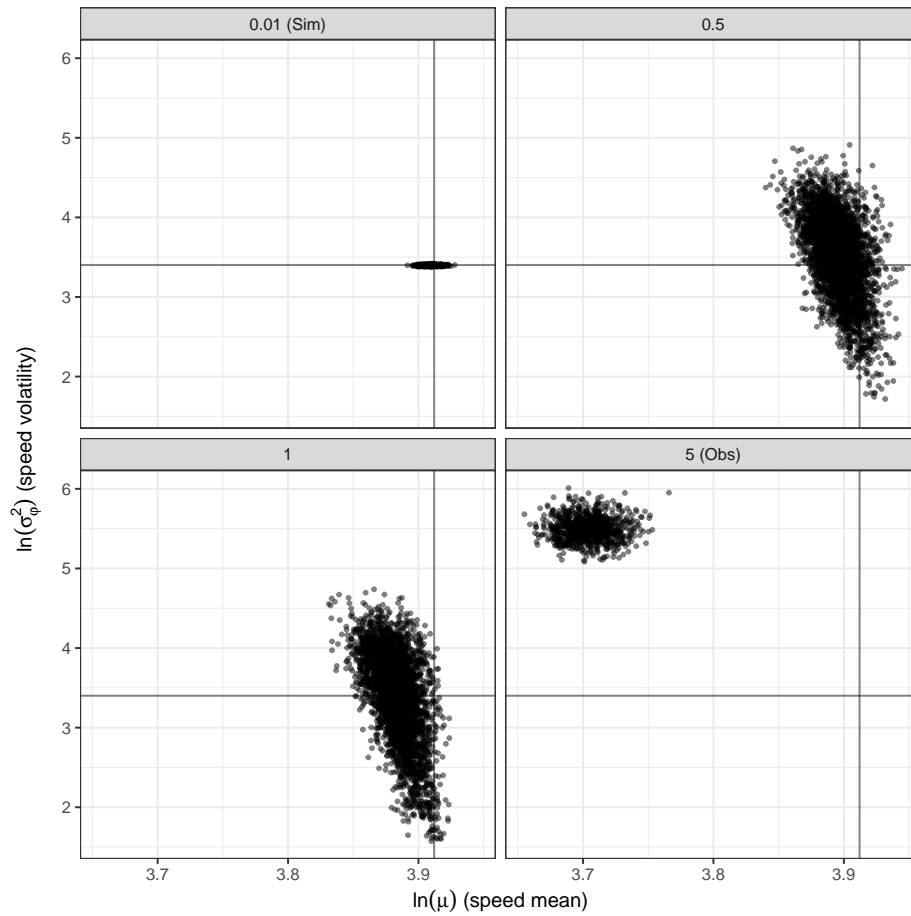


Fig. 3.9 Sampled (log) speed mean against variance (thinned and with burn-in time omitted) for the single state, independent step simulation (**points**). True parameter values are highlighted (**lines**).

Scatter plots of the posterior sampled movement parameters are given in Figs. 3.8–3.9 (on a log scale). The former compares the bearing variance and the speed mean and the latter compares the speed mean with variance. Each panel shows one of the four implementations to avoid overplotting and the true value is highlighted with lines. As expected, there is little error in parameter estimation when there is full observation of the simulated path (**top left**), however this implementation was carried out only as a ‘best-case’ comparison for the other scenarios. The two runs in which the path needed to be reconstructed (for $\delta t = 0.5$ **top right** and for $\delta t = 1$ **bottom left**) show similar results for the parameter estimates, with high correlation between the sampled values. The true parameter values have been captured well, with samples centred more closely on the true value when the refined time scale is smaller, providing a better approximation to the underlying process. The final run using only the observed locations without any refinement to the path (**bottom right**) has performed badly in parameter estimation. Both the bearing variance and the speed mean have been consistently underestimated and the speed variance overestimated.

Posterior credible intervals are given in Table 3.2 and Fig. 3.10. Kernel density estimates are given in the Appendix, Fig. B.1. All of these summaries use thinned samples, with burn-in (assumed to be the first eighth of each run) discarded. The ‘discrete-time style’ implementation does not contain the true parameter values in a 95% credible interval, underestimating the bearing variance and mean speed and overestimating the speed variance. The 95% credible intervals when the refined time scale was 0.5 are shown to be closer to the true simulation values than the coarser approximation at time scale 1, but in both cases contain the true value for all parameters. The credible intervals for the implementation with full observation of the simulated path capture the true values, as expected, however note that these are not centred on the true value for the distance parameters, instead favouring lower values.

These results demonstrate the ability of the presented method for inference to estimate parameters. Just a small refinement (four intermediate locations between each pair of observations) is shown to improve parameter estimation greatly over analysing the observations directly, particularly with regards to the distance parameters. Further refinement to the movement path provides a better approximation to the underlying movement process, and when doubling the level of reconstruction here the parameter estimates were improved upon.

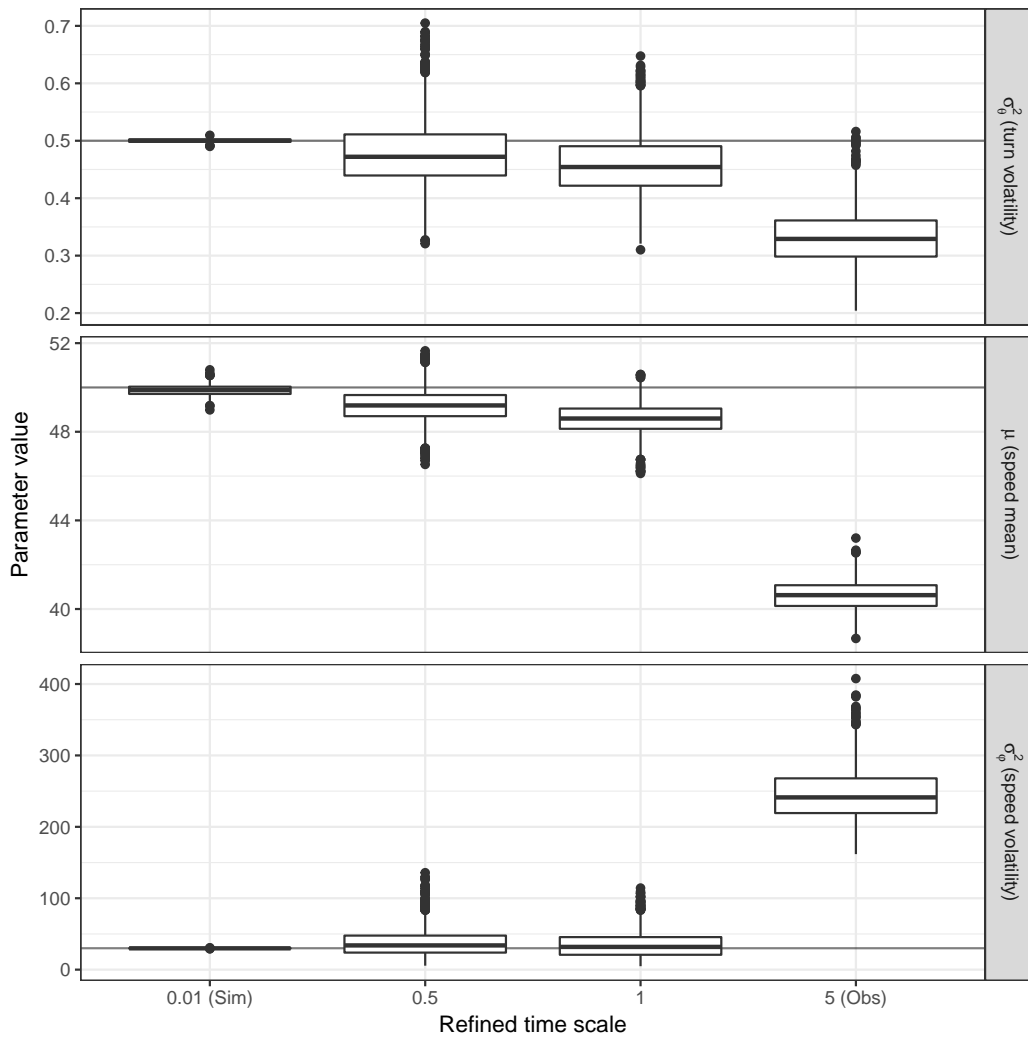


Fig. 3.10 Box plots of the movement parameters (using thinned samples and with burn-in time omitted) for the single state, independent steps simulation. True parameter values are highlighted (**horizontal lines**).

Table 3.2 Credible intervals for the movement parameters in the single state, independent step simulation. The true values used to simulate the observations were $\{0.5, 50, 30\}$, with the implementation highlighted if the true value is not included in the estimate. The ESS for each parameter is also given. When path reconstruction was not needed, this was taken from the 1,000 samples thinned from 10^5 iterations. When path reconstruction was implemented, this was taken from the 4,000 samples thinned from 4×10^6 iterations.

		2.5%	50%	97.5%	ESS
σ_θ^2	0.01 (Sim)	0.494	0.500	0.507	1000
	0.5	0.380	0.472	0.594	858
	1	0.366	0.454	0.565	612
	5 (Obs)	0.254	0.329	0.445	1000
μ	0.01 (Sim)	49.396	49.882	50.353	1000
	0.5	47.795	49.189	50.629	406
	1	47.244	48.598	49.907	242
	5 (Obs)	39.371	40.626	42.018	1000
σ_φ^2	0.01 (Sim)	29.525	29.879	30.239	1000
	0.5	11.026	33.957	80.392	214
	1	7.750	31.855	73.643	106
	5 (Obs)	183.460	241.229	334.539	1000

3.5.1.4 Path reconstruction results

Examples of 100 path reconstructions for the two implementations at time scales of 1 and 0.5 are given in detail in Fig. 3.11 and in full in Fig. 3.12. There is no uncertainty around the observed locations, and so each path reconstruction passes exactly through these locations. As expected, the locations at the times halfway between observations have the most uncertainty.

Fig. 3.12 demonstrates the ability to estimate the underlying movement between observed locations. When two observations are further apart from one another (such as the first two observations in the top left of Fig. 3.11), the reconstructed paths are more similar to one another, and uncertainty in location is generally low. However, when the true movement path involved a large curve (such as that in the bottom right of Fig. 3.11) there is more uncertainty in the intermediate locations, with reconstructed paths exploring the two possibilities of curving between the observations (albeit with one route having higher support).

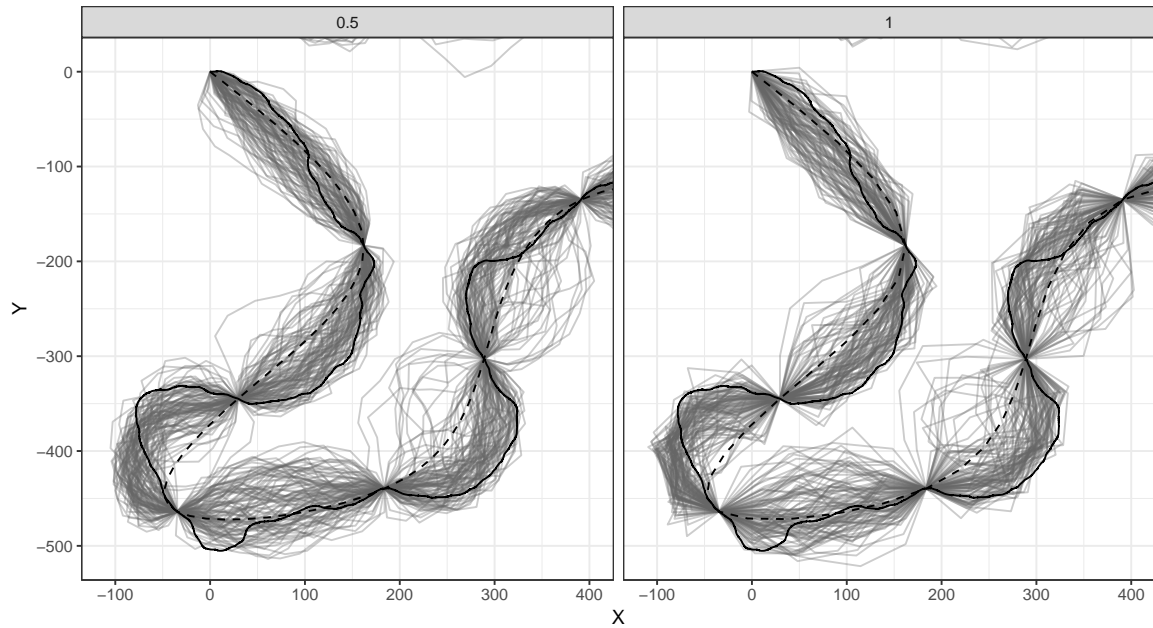


Fig. 3.11 Detailed path reconstructions between the times 0–30 (**grey**) for the single state, independent step example. Path reconstruction only occurred when the time scale was 0.5 or 1. The simulated path is given (**solid black**) along with the path used to initialise the MCMC sampler (**dashed black**).

3.5.2 Simulated movement example with correlated speed process

This second example demonstrates the extended movement model that includes correlation in the speed process and highlights the difficulty in estimating such a feature when only considering movement at the sampling time scale. This example features more volatile movement in both the bearing and speed processes, expected to lead to higher uncertainty in unobserved location.

3.5.2.1 Underlying movement and observations

Movement under the correlated speed process in Sect. 3.4 was simulated with movement parameters

$$\Phi = \{\sigma_{\theta}^2 = 0.7, \mu = 15, \beta = 0.1, \sigma_{\psi}^2 = 20\}, \quad (3.13)$$

at the approximate time scale $\delta t = 0.01$, between the times 0–500. Observations of the movement path were created as in the previous example, subsampling the simulation to give 101 ‘observed’ locations at a time scale of 5. The simulated path, with observations, is given in Fig. 3.13.

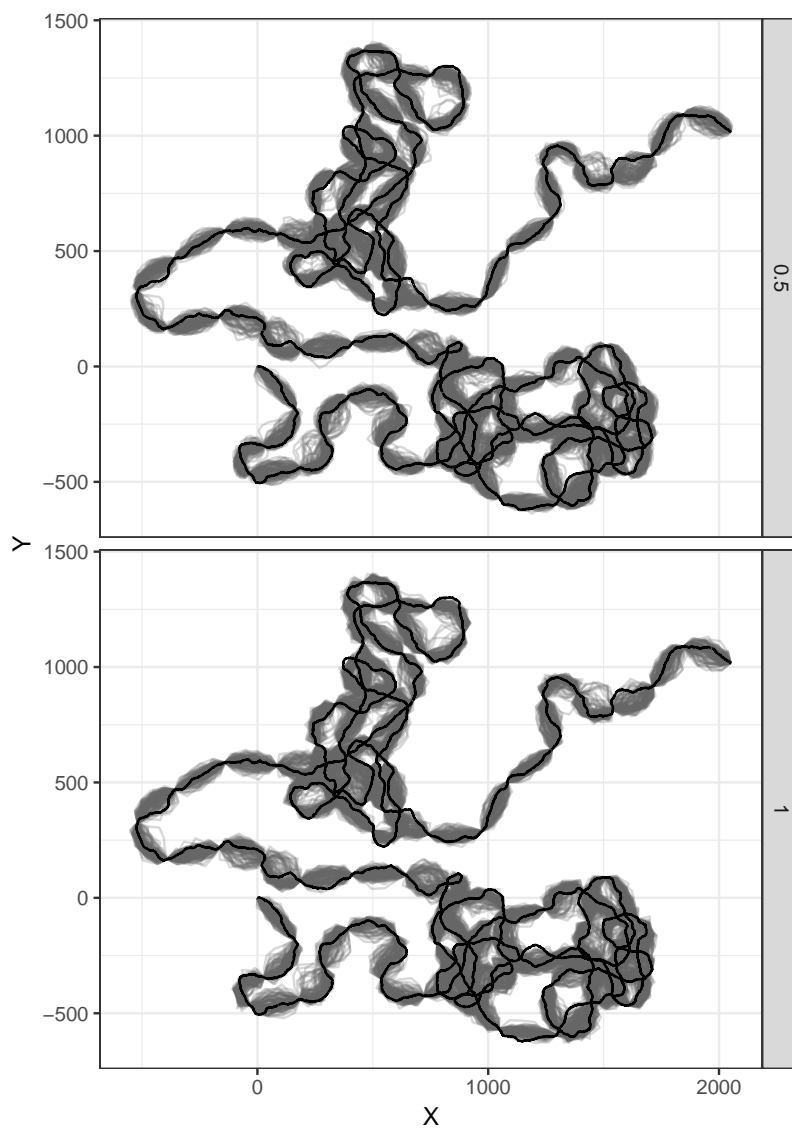


Fig. 3.12 Full path reconstructions (**grey**) for the single state, independent step simulation. Note that reconstruction only occurred for the implementations at time scales 0.5 and 1. The simulated path is also given (**black**).

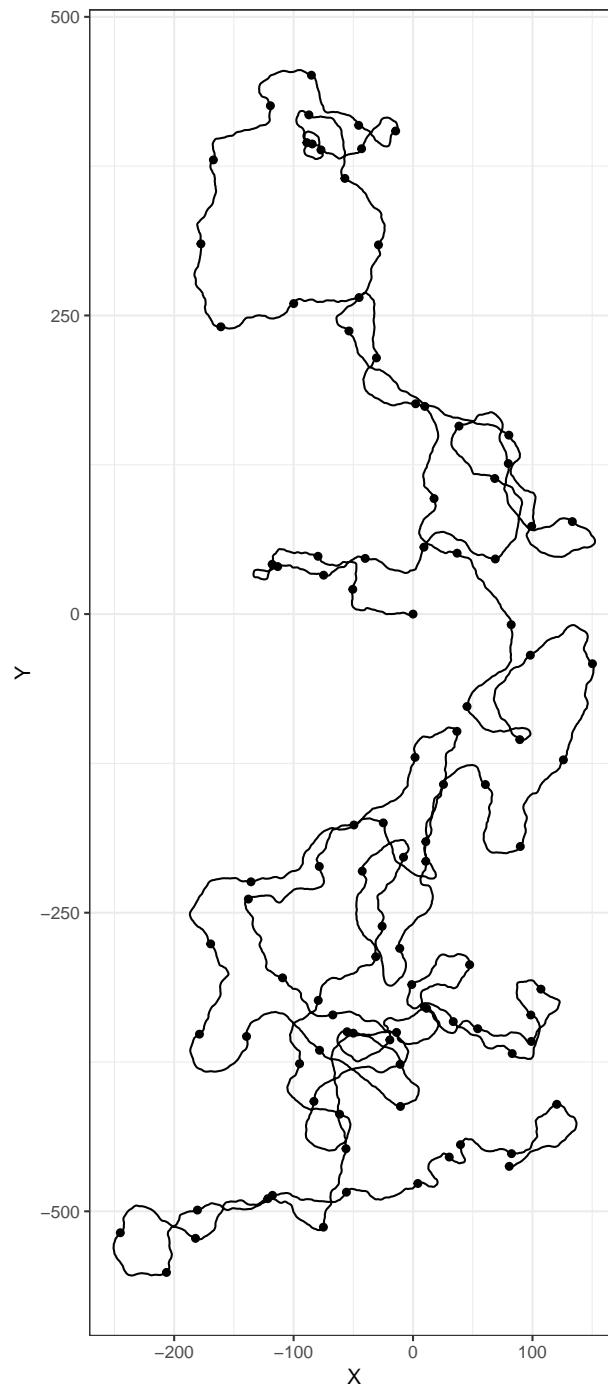


Fig. 3.13 The single state, correlated step simulated movement path (**line**). Observations are included (**points**).

Table 3.3 Initial parameters, perturbation variances and acceptance rates for the implementations carried out using the correlated step simulation.

δt	$\Phi^{(0)}$				ρ			Acceptance (%)	
	$\sigma_{\theta}^2{}^{(0)}$	$\mu^{(0)}$	$\beta^{(0)}$	$\sigma_{\psi}^2{}^{(0)}$	ρ_{μ}	ρ_{β}	$\rho_{\sigma_{\psi}^2}$	$\{\mu, \beta, \sigma_{\psi}^2\}$	$\{\theta, \mathbf{v}\}$
0.01 (sim.)	0.696	14.7	0.0673	14.9	0.5	0.001	10	2	-
0.5	0.138	11.4	0.0374	19.4	0.1	0.001	1	18	2
1	0.207	11.4	0.0697	19.2	0.1	0.001	1	25	1
5 (obs.)	0.427	11.0	0.0922	14.2	0.3	10^6	9	42	-

3.5.2.2 Implementing the inference algorithm

As in the previous example, four implementations were carried out at varying refined time scales. Using the simulation time scale (0.01) with full knowledge of the movement path provided a ‘baseline’ on the posterior distribution of the parameters. Two implementations of the full inference algorithm were carried out at time scales of 0.5 and 1, and finally at the time scale of the observations (5).

3.5.2.2.1 Initial values The initial refined movement path for the implementations at times 0.5 and 1 was created as in Sect. 3.5.1.2, extracts of which are shown in Figs. 3.14 and 3.15 along with the fixed paths for the implementations at times 0.01 and 5. The initial movement parameters are given in Table 3.3.

3.5.2.2.2 Prior information and implementation Prior distributions and sample run length were set as in the previous example. Perturbation variances for the speed parameters, and acceptance rates are given in Table 3.3. The ACF of the sampled parameters is included in the Appendix, Fig. B.6, and the ESS in Table 3.4.

3.5.2.3 Movement parameter results

Sampled movement parameters for all four implementations are given as trace plots in the Appendix, Fig. B.5, with the true value highlighted. Note that the trace for β in the case where $\delta t = 5$ is not included here, but given in Fig. B.4 because of its magnitude. All implementations passed Heidelberger and Welch convergence diagnostics apart from β when $\delta t = 5$, and the ESS is given in Table 3.4.

The ESS for the speed correlation parameter, β , in the case where the refined time scale was equal to the observation time scale was low (3), highlighted by the trace plot in the Appendix, Fig. B.4. This parameter could not be estimated with the information provided at this time scale, highlighting an important issue with analyses at such a coarse time scale. The correlation parameter features as an exponential term in the conditional distribution of the speed ($e^{-\delta t \beta}$), and so once β is above 5, this exponential term is below 10^{-10} (the true value would lead to 0.6). At the observation time scale the parameter could not be estimated, and instead sampling drifted around the space where β was large (and the exponential term approximately zero).

Scatter plots of the posterior sampled movement parameters for all implementations are given in Figs. 3.16–3.17 (on a log scale). The former compares the bearing variance and the speed mean while the latter compares the speed correlation with variance. Each panel shows one of the four implementations to avoid overplotting and the true values are highlighted with lines. As expected, there is little variability in parameter estimation when there is full observation of the simulated path (**top left**). However, the most variability can be seen in the estimation of the mean speed and there is a strong relationship between the speed correlation and variance.

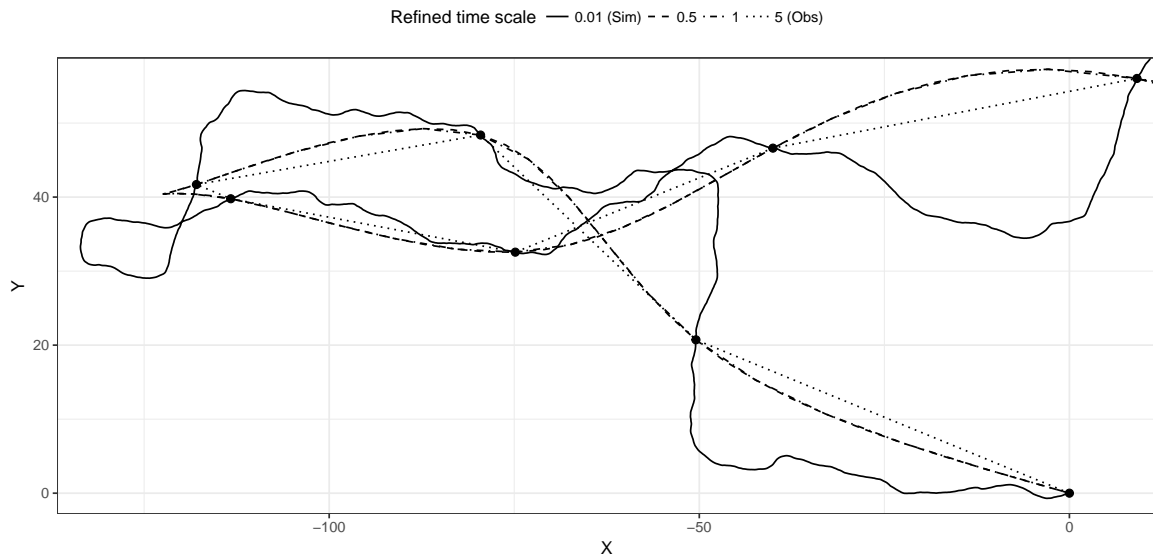


Fig. 3.14 Small section (between the times 0–40) of the initial movement path used within the single state correlated step simulation. When the refined time scale was the same as the observations, the movement path was fixed. When the refined time scale was the same as the simulation, the true simulated path was assumed known and fixed. When the refined time scale was 1 or 0.5, the movement path was reconstructed, with initial path given here, created using an interpolating spline between observed locations.

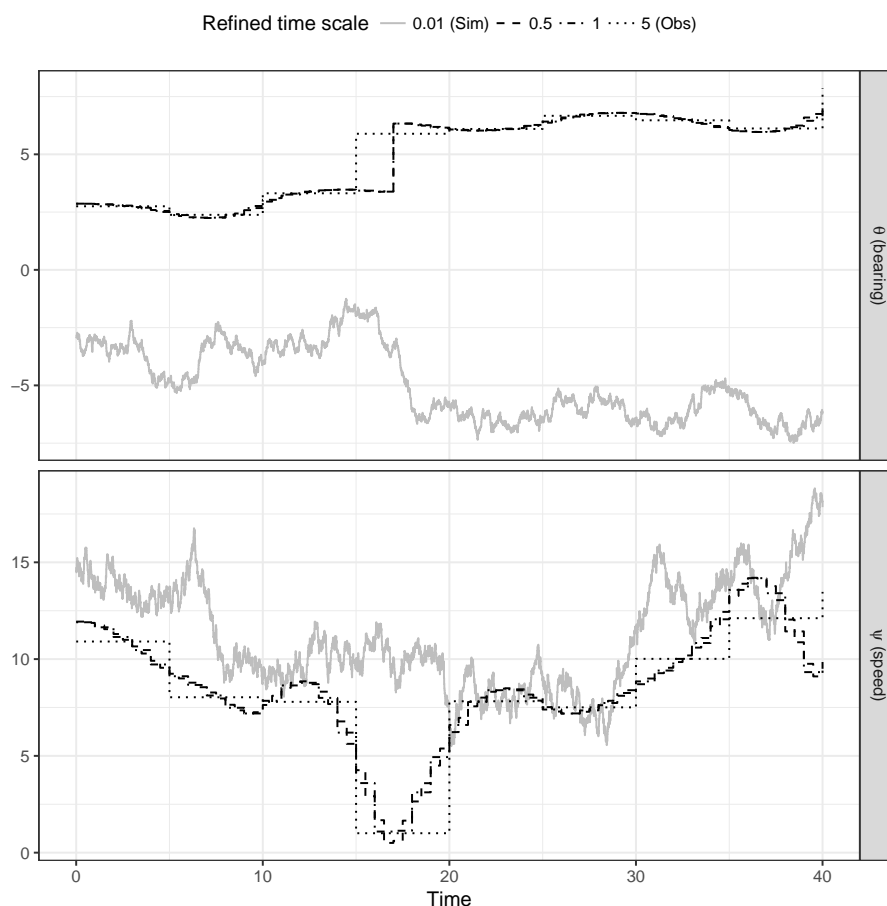


Fig. 3.15 Small section (between the times 0–40) of the initial movement path (as bearings and speeds) used within the single state correlated step simulation. When the refined time scale was the same as the observations, the movement path was fixed. When the refined time scale was the same as the simulation, the true simulated path was assumed known and fixed. When the refined time scale was 1 or 0.5, the movement path was reconstructed, with initial path given here, created using an interpolating spline between observed locations.

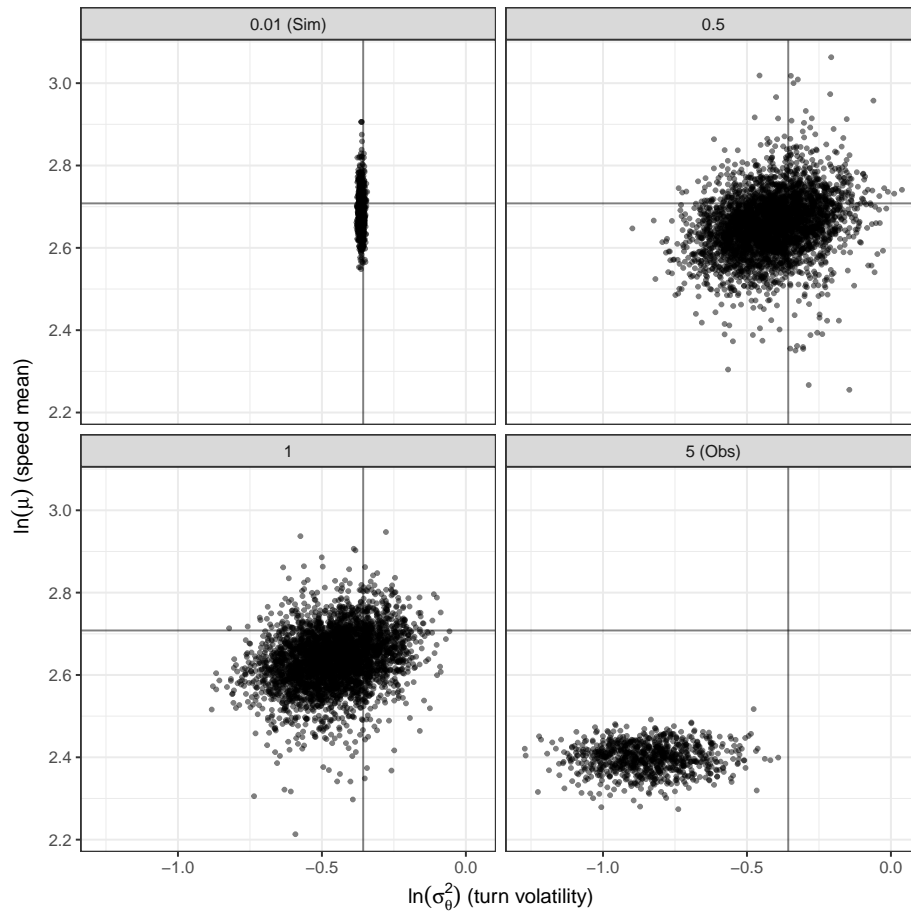


Fig. 3.16 Sampled (log) bearing variance against mean speed (thinned and with burn-in time omitted) for the single state, correlated step simulation (**points**). True parameter values are highlighted (**lines**).

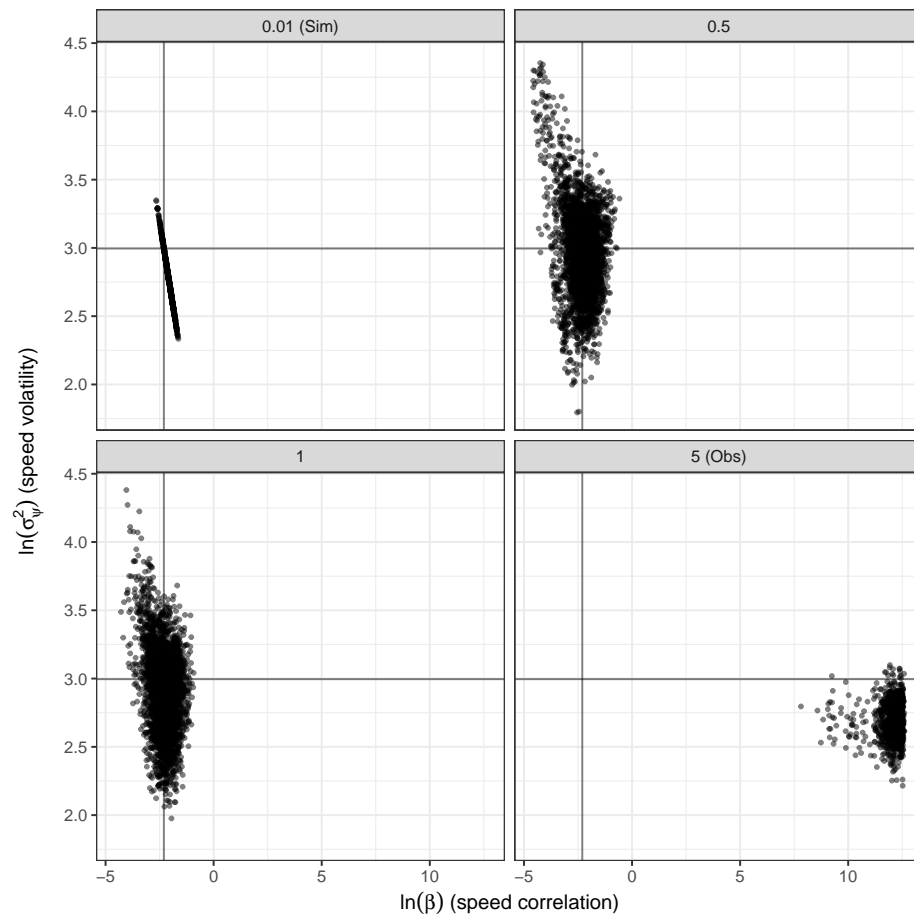


Fig. 3.17 Sampled (log) speed correlation against variance (thinned and with burn-in time omitted) for the single state, correlated step simulation (**points**). True parameter values are highlighted (**lines**).

The two runs in which the path needed to be reconstructed (for $\delta t = 0.5$ **top right** and for $\delta t = 1$ **bottom left**) show similar results for parameter estimates, with a relationship between the speed correlation and variance (as present when $\delta t = 0.01$). The true parameter values have been captured well, with little difference between the two implementations.

Using only the observed locations without refinement to the path (**bottom right**) performed poorly in parameter estimation. Both the bearing variance and speed mean have been consistently underestimated and the speed correlation parameter vastly overestimated (underestimating the level of correlation in the speed). Only the long-term variance of the speed process includes the true value within the posterior samples.

Posterior credible intervals are given in Table 3.4 and Fig. 3.18. Kernel density estimates are given in the Appendix, Fig. B.7. All of these summaries use thinned samples of the movement parameters, with burn-in (assumed to be the first eighth of each run) discarded. The implementation in which the refined time scale was equal to the sampling scheme only contains the true parameter values in a 95% credible interval for the speed variance, underestimating other parameters. The 95% credible intervals when the refined time scale was 0.5 are shown to be similar to the coarser approximation at time scale 1, containing the true value for all parameters.

The results here demonstrate the ability of the inference method presented for estimating movement parameters. Just a small refinement (four intermediate locations between each pair of observations) is shown to improve parameter estimation greatly over analysing the observations directly, particularly with regards to the speed correlation parameter which could not be estimated in the latter case.

3.5.2.4 Path reconstruction results

Examples of 100 path reconstructions for the implementations at time scales of 1 and 0.5 are given in Fig. 3.19, showing the full movement path. Movement overall has been captured well but there can be seen to be more uncertainty in location between observations than that seen in the previous example. This is to be expected due to the higher volatility of the simulation. Fig. 3.20 shows a detailed portion of the movement path with a smaller set of reconstructions (4) for clarity. This figure highlights a section of the path when there is a ‘loop’ in the simulation between two observed locations (at which there is no uncertainty). This kind of movement feature proved to be difficult to capture in the reconstructions because of the high uncertainty when such a feature exists. However, although there is high uncertainty as to

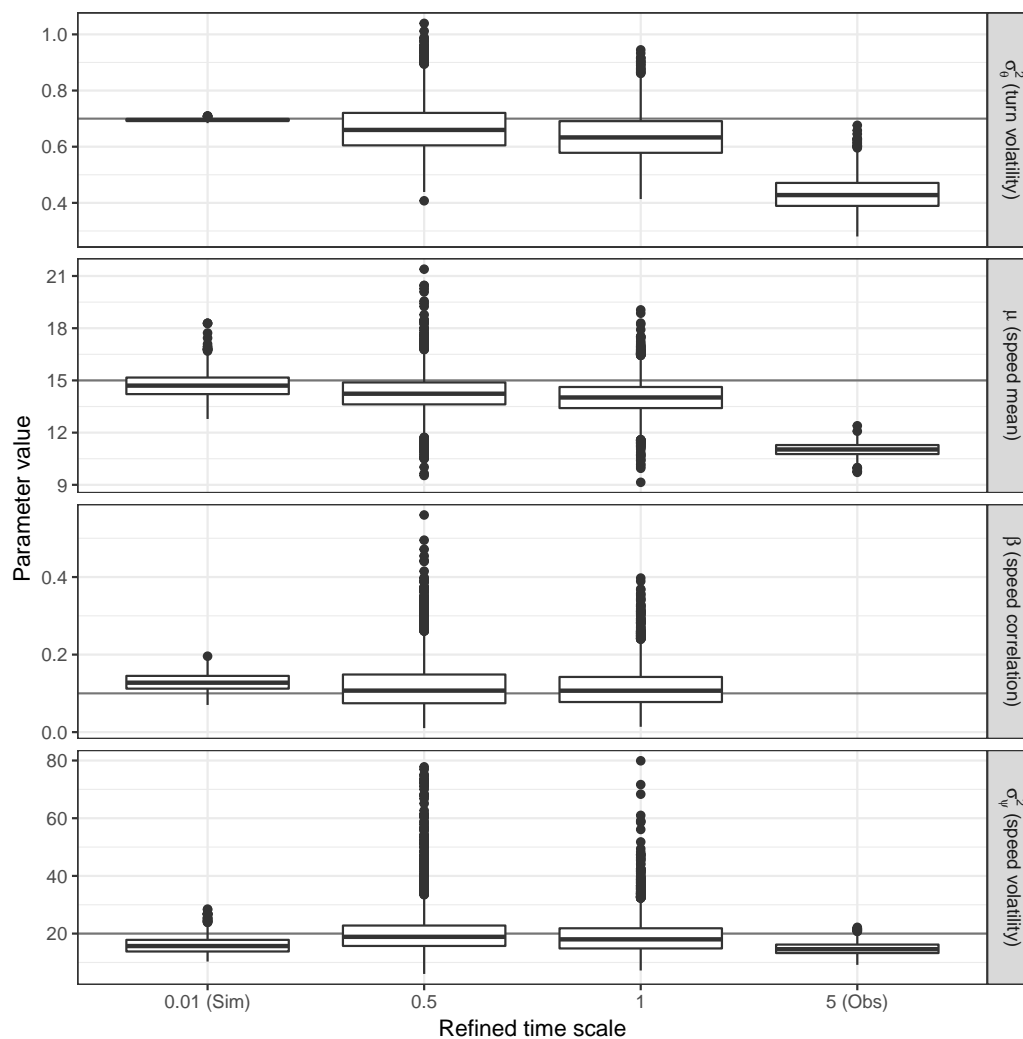


Fig. 3.18 Box plots of the movement parameters in the single state, correlated step simulation (using thinned samples and with burn-in time omitted). True parameter values are highlighted (**horizontal lines**). Note that β is not included here when $\delta t = 5$ as it could not be estimated.

Table 3.4 Credible intervals for the movement parameters in the single state, correlated step simulation. The true values were $\{0.7, 15, 0.1, 20\}$ with implementation highlighted if the true value is not included in the estimate. Note that summary statistics for β when $\delta t = 5$ are not included here as it could not be estimated. The ESS for each parameter is also given. When path reconstruction was not needed, this was taken from the 1,000 samples thinned from 10^5 iterations. When path reconstruction was implemented, this was taken from the 4,000 samples thinned from 4×10^6 iterations..

		2.5%	50%	97.5%	ESS
σ_θ^2	0.01 (Sim)	0.687	0.695	0.704	1000
	0.5	0.516	0.660	0.851	360
	1	0.494	0.633	0.805	600
	5 (Obs)	0.330	0.428	0.581	1000
μ	0.01 (Sim)	13.455	14.700	16.244	277
	0.5	12.281	14.236	16.447	943
	1	12.066	14.019	15.956	2332
	5 (Obs)	10.232	11.032	11.726	1000
β	0.01 (Sim)	0.084	0.128	0.177	200
	0.5	0.025	0.107	0.276	144
	1	0.036	0.107	0.242	321
	5 (Obs)	-	-	-	3
σ_ψ^2	0.01 (Sim)	11.326	15.685	23.828	183
	0.5	10.866	18.879	39.432	342
	1	10.263	18.012	34.056	814
	5 (Obs)	11.251	14.587	19.311	950

specific locations, the actual presence of a ‘loop’ appears in the majority of reconstructions, as shown.

3.6 Discussion

This chapter introduced a framework for modelling animal movement in continuous time based on the popular movement metrics of step lengths and turning angles. A method for statistical inference via the augmentation of a refined path that is assumed to approximate continuous time has been described and demonstrated on simulated observations.

The underlying model for movement described in Sect. 3.1 is based on the CRW ideas of discrete-time models such as Morales et al. (2004), assuming that the distance travelled over disjoint sections of times are independent from one another. The continuous-time formulation, however, allows missing and irregular observations to be handled with ease. The model introduced in Sect. 3.4 assumes that there is correlation in the distance process. Note that even if this model were to have a $\beta \gg 0$, hence virtually no correlation in speed, the two models would still differ in the accumulation of uncertainty with time. The correlated speed model has some similarities with the velocity-based continuous-time model of Johnson et al. (2008a) but is more intuitive, enabling a separation of speed and direction that matches empirical observations well. Parameter interpretation is simpler when separated in this way, describing aspects of movement such as a mean travelling speed and a volatility to the direction of movement. Although continuous-time models based on co-ordinate locations, such as Johnson et al. (2008a) or Blackwell et al. (2015), could be applied, with post-processing carried out to determine the distributions of speeds and bearings, the covariance structure of such distributions, and hence the implicit shape of the path, will not be the same as that defined by the model here. Ecological justification for such a covariance structure may be difficult or lacking, whereas the model here is directly defined by these quantities and motivated by ecological ideas.

For a given time interval, the distribution of the change in direction given by this continuous-time model will always be a wrapped Gaussian centred at zero. A von Mises distribution (often used to model turning angles in discrete-time models such as McClintock et al. (2012)) centred at zero is similar to this, but a von Mises (or other circular) distribution centred at $\pm\pi$ is not. In fact, no natural continuous-time process for change in direction would lead to such a distribution when observed at regular intervals. Such a distribution would require the expected rate of change of bearing to be non-zero, leading to paths that consistently

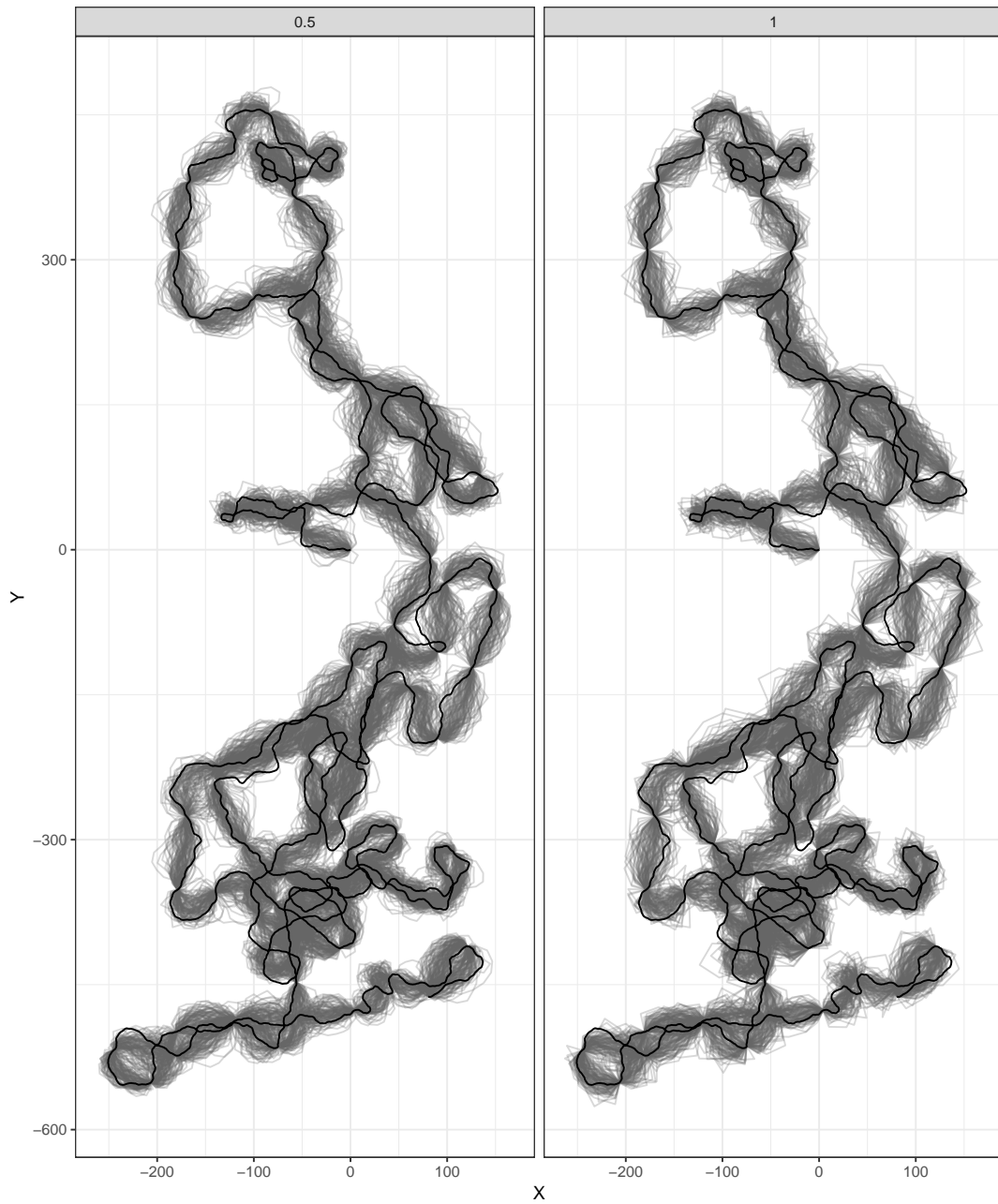


Fig. 3.19 Full path reconstructions (**grey**) for the single state, correlated step simulation. Note that reconstruction only occurred for the implementations at time scales 0.5 and 1. The simulated path is also given (**black**).

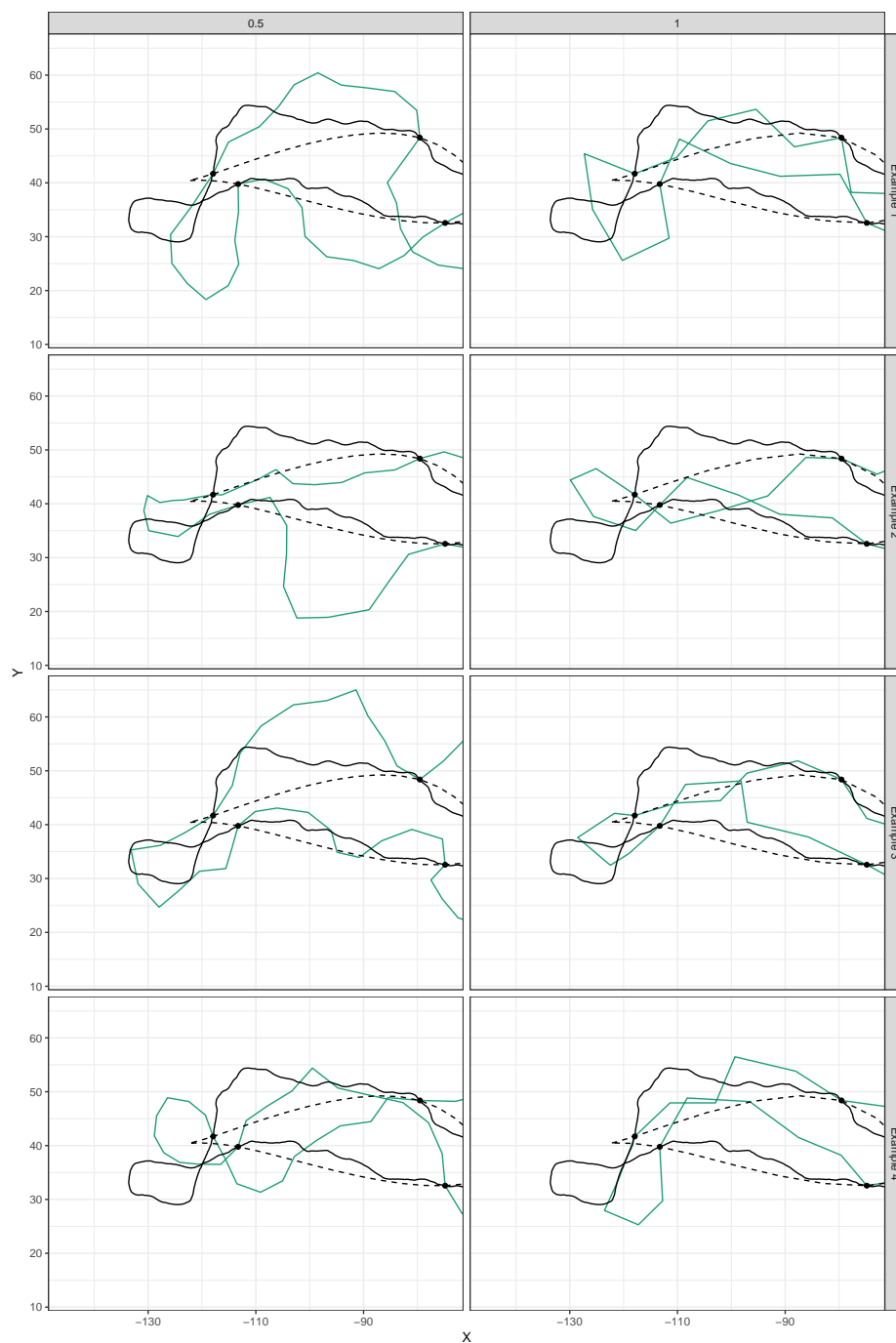


Fig. 3.20 Examples of a small section of the sampled refined path reconstructions (**green**) for the single state, correlated step simulation. The simulated path is given (**solid black**) along with initial paths (**dashed black**).

form loops. While this may be appropriate occasionally we do not feel it is realistic in most published applications. It seems more likely that such a distribution emerges only as an artefact of some other process e.g. ignored measurement error (Hurford, 2009) or attraction to a particular location in space. The classification of movement with a mean turning angle of $\pm\pi$ in a number of discrete-time applications (the elk example of Morales et al. (2004)) is questionable. The ecological interpretation of a ‘foraging’ style movement would be better modelled as having a uniform turning angle, such as $\sigma_\theta^2 \rightarrow \infty$ in the continuous-time model here. However, other modelling choices exist for the bearing process that could describe ecologically realistic movement. Rather than BM increments, an OU process centred at zero would create correlation in the increments of the bearing process, or directed movement could be achieved by considering the change in bearing required to reach an attractive centre.

Although the approach for inference in Sect. 3.3 involves an approximation to the underlying continuous-time model, advantages remain over discrete time because the parameters of the model are scalable (as they represent parameters of a continuous time model) rather than a ‘per observation time’ parameter. Reducing the time scale of the refined path will provide a ‘better’ approximation to the underlying model and improve upon discrete-time analyses, although this does come with a computational cost. A refined path with more points will be more costly to evaluate the likelihood of (when sampling movement parameters) and there will have to be more updates to the refined path to compensate for the increase in refined locations needing reconstruction. Simulation experiments on the effect of varying δt (Sect. 3.5) show that great improvements to parameter estimation could be made against using only observations by augmenting with as little as four locations between observation pairs. Improving the approximation with further refinement was found to increase accuracy of estimation further, but incurred the cost of additional computation time. Future analyses must make a decision on what is the acceptable trade off between the accuracy and computational cost required to address the relevant ecological questions.

The augmentation approach employed in the inference method furthers the aim for comprehensible inference. The ability to view examples of path reconstructions, such as those in Figure 3.11, aids in understanding the movement type associated with a given combination of parameters. Sampling a large number of reconstructions highlights the uncertainty in location, which can then be used to estimate the space/resource use of the animal at the local scale. With the resolution of environmental covariates increasing, this information can be correctly combined with local scale movement rather than assuming only the covariate values corresponding to directly observed locations are important. This gives an equivalent, for

step-and-turn models, of the BB approach of Horne et al. (2007). For discussion of the wider issues of linking movement and resource use, see for example Johnson et al. (2008b).

The inference method of Sect. 3.3 led to low path reconstruction acceptance rates in the examples Sect. 3.5. The acceptance of a path section is based on the likelihood of the known end location, given a set of bearings. The acceptance rate could potentially be improved by altering the proposal distribution for the bearings. Possibilities include a random walk perturbation of the current set of bearings, or (more favourable for mixing purposes) a proposal with mean given by a mixture of both the underlying BB and the current set of bearings.

The methods presented in this chapter assume that movement follows a single behavioural mode, which is unrealistic for animal tracks covering an extended period of time. Behavioural switching that builds on this model in continuous time is addressed in Chapter 4, based on the work of Blackwell et al. (2015) to include switching between a finite number of ‘behavioural states’ that represent quantitative or qualitative differences in movement. Further, the methods here assume that observation error is negligible. This is not the case in most GPS data sets, and so the ability to correctly account for this error is important. Chapter 5 extends the model here to a framework for incorporating observation error.

Chapter 4

Multistate movement

Chap. 3 introduced a continuous-time movement model based on similar quantities to those of popular discrete-time ‘step and turn’ models. This provides familiar descriptive parameters for estimation whilst respecting the inherent continuous-time characteristic of movement—hence has the ability to handle missing and irregular observations with ease. Describing only single state movement limits this model to applications with short term sampling periods (Morales and Ellner, 2002), and so the aim of this chapter is to introduce a statistical, multistate movement model in continuous time able to provide intuitive and easily interpretable parameters for the non-statistical user. Multistate switching movement is introduced by extending the model of Chap. 3 to include a CTMC behavioural process.

Sect. 4.1 introduces the proposed multistate model, with simulation described in Sect. 4.2. An extension to the inference method previously described, given observed telemetry data, that incorporates multistate movement is outlined in Sect. 4.3. This method assumes no direct observations of the behaviour itself. The use and interpretability of this method is then demonstrated on simulated (Sect. 4.4) and real (Sect. 4.5) datasets. Note that the following (in part) provides an extended description to that presented in Parton and Blackwell (2017).

4.1 Model for multistate switching

To reflect the changing behaviours of an animal over time, a switching model is employed for a ‘behavioural process’, where different movement characteristics are associated with each state/behaviour (Blackwell, 1997; Blackwell et al., 2015; McClintock et al., 2012; Morales et al., 2004). In the following, the behavioural process is taken to be a CTMC with switching

rates λ and probabilities q (see Sect. 2.3.4 and Guttorp (1995)). The animal will follow behavioural state i for a length of time exponentially distributed with rate λ_i , before switching to state j with probability $q_{i,j}$.

Within a behaviour there is a corresponding set of parameters describing movement as in either Sect. 3.1 or 3.4. With this extension in place the marginal joint process of bearing and speed is not Markovian, however the joint process of behaviour, bearing and speed is. The movement of the animal is parametrised by the set $\Phi = \{\Phi_B, \Phi_M\}$, with

$$\begin{aligned}\Phi_B &= \{\lambda_i, q_{i,j}\} \\ \Phi_M &= \{\sigma_{\theta,i}^2, \mu_i, \beta_i, \sigma_{\psi,i}^2\}\end{aligned}$$

for $i \neq j \in \{1, \dots, p\}$, where p is the number of behavioural states.

4.2 Simulating movement

Realisations of movement given parameters Φ can be easily simulated, summarised in Algorithm 7 and with an example in Figs. 4.1 and 4.2. The behavioural process is simulated according to a CTMC with generator matrix defined by Φ_B . Assume that the trajectory begins in equilibrium, so that the initial behavioural state is simulated from the equilibrium distribution corresponding to the generator matrix defined by Φ_B . This equilibrium distribution, π , is found by solving the global balance equations

$$-\pi_i \lambda_i = \sum_{j \neq i} \pi_j \lambda_j q_{j,i}, \quad (4.1)$$

for $i \in \{1, \dots, p\}$. The remainder of the behavioural process is then simulated by a conditional process. Given a current behaviour $B(t) = s$, this involves drawing the time until the next switch from an exponential distribution with rate λ_s and then choosing the new behaviour $j \neq s$ with probability $q_{s,j}$. Simulation ceases at the point where the time of the next switch exceeds the final time the simulation is to be carried out over.

Given a simulated realisation of the behavioural process, movement is simulated approximately at a time scale δt , which can be arbitrarily fine scale. Note however, that this approximation time scale must (at least) include all times of behavioural switches. If the

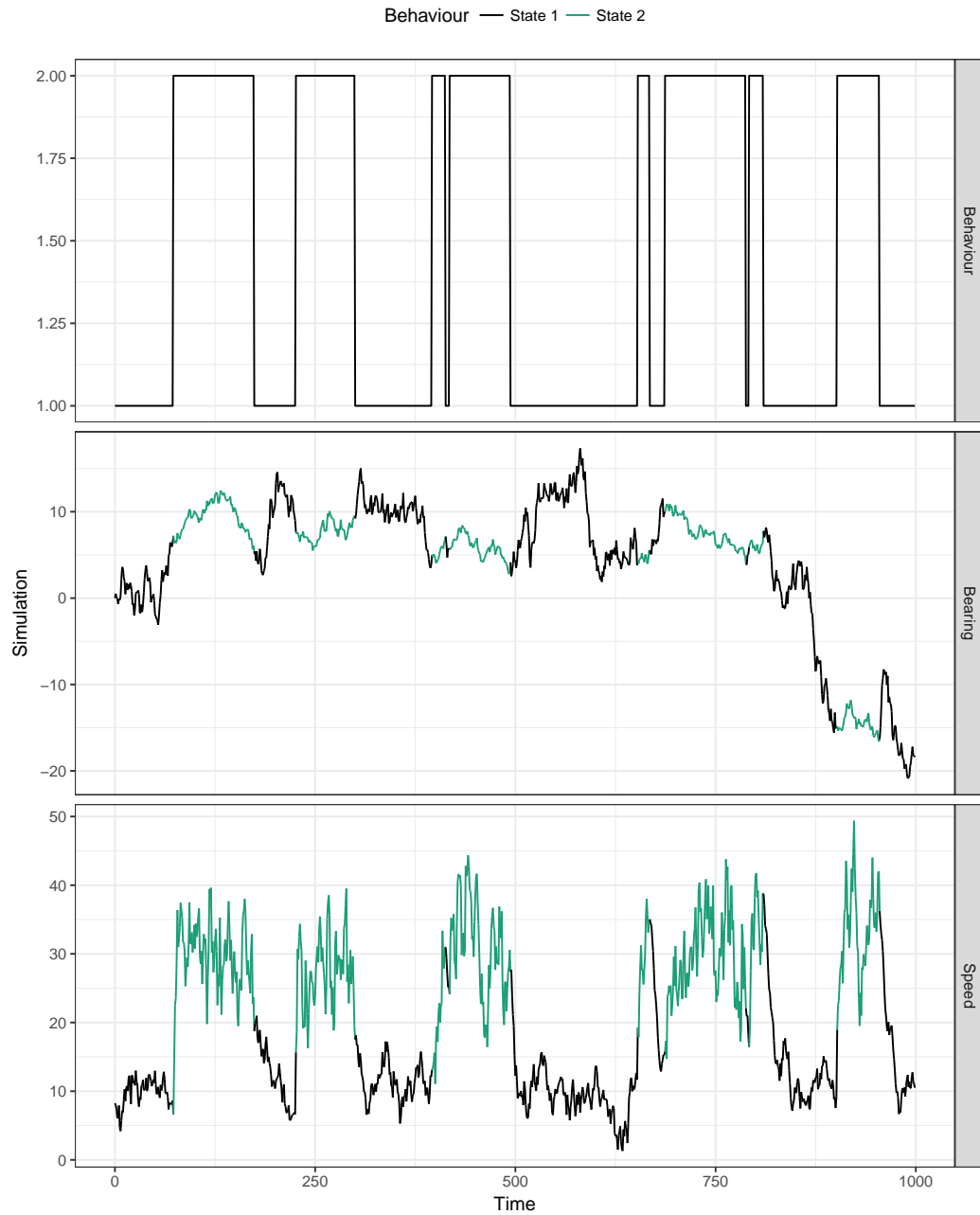


Fig. 4.1 An example of a simulated movement path with two behavioural states. The top panel shows the simulated behavioural process in time, and the bottom two panels show the bearing and speed processes. Each panel is additionally coloured by behavioural state; state 1 characterised by slow movement with high bearing variance and state 2 characterised by fast, directed movement. Corresponding locations are in Fig. 4.2.

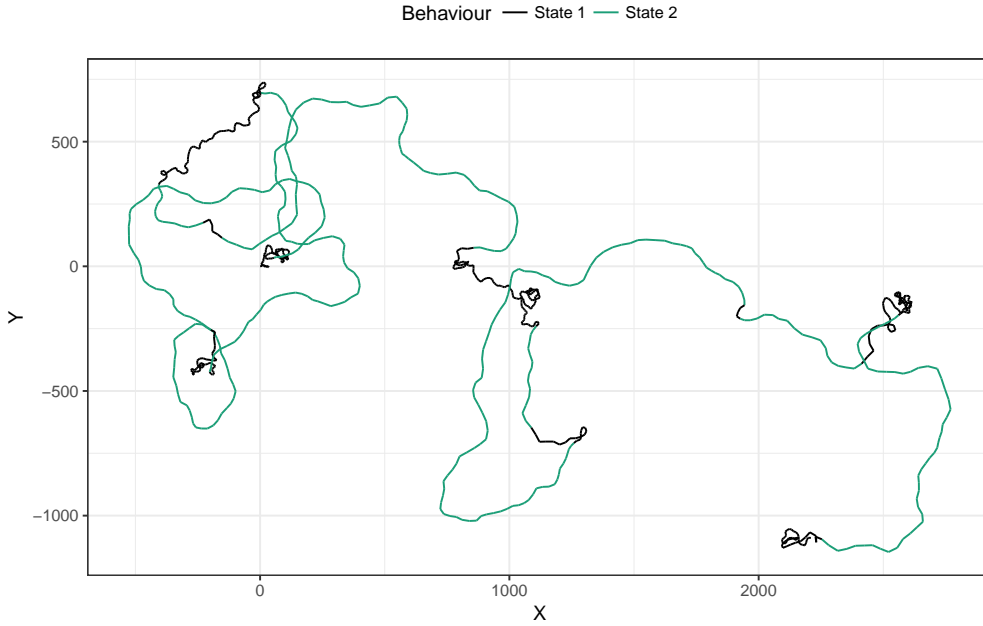


Fig. 4.2 The locations corresponding to the simulation of two-state movement in Fig. 4.1.

behaviour at time t is $B(t) = s$, then the bearing and speed are given as

$$\theta(t + \delta t) \mid \theta(t), B(t) = s \sim N\left(\theta(t), \sigma_{\theta,s}^2 \delta t\right), \quad (4.2)$$

$$\psi(t + \delta t) \mid \psi(t), B(t) = s \sim N\left(\mu_s + e^{-\beta_s \delta t} (\psi(t) - \mu_s), \sigma_{\psi,s}^2 (1 - e^{-2\beta_s \delta t})\right), \quad (4.3)$$

assuming the correlated speeds model of Sect. 3.4. For the remainder of this chapter it will be assumed that the speed process follows this correlated model, however, note that the independent model of Sect. 3.1 could equally be implemented. Again, the familiar notion of a ‘step’ is recovered by $v(t) = \psi(t) \delta t$.

As in single state movement, given the joint process of bearings and speeds, the Euler-Maruyama approximation of location in 2-dimensional space is given by the cumulative sums

$$\begin{aligned} X(t_i) &= X(t_0) + \sum_{j=1}^{i-1} v(t_j) \cos(\theta(t_j)), \\ Y(t_i) &= Y(t_0) + \sum_{j=1}^{i-1} v(t_j) \sin(\theta(t_j)). \end{aligned} \quad (4.4)$$

Algorithm 7 Simulating multistate movement.

1: **procedure** SIMULATE MULTISTATE MOVEMENT($\Phi, \{t_0, T\}$)

2: Simulate the initial behavioural state

$$B(t_0) \sim \text{Multinomial}(\boldsymbol{\pi})$$

$$t \leftarrow t_0$$

3: Simulate the time to the first behavioural switch

$$t^* \sim \text{Exp}(\lambda_{B(t)})$$

4: **while** $t + t^* < T$ **do**

5: Simulate the behaviour to switch into

$$B(t + t^*) \sim \text{Multinomial}(\mathbf{q}_{B(t), \cdot})$$

6: Update $t \leftarrow t + t^*$

7: Simulate the next behavioural switch time

$$t^* \sim \text{Exp}(\lambda_{B(t)})$$

8: **end while**

9: Create set of times $\{t_0, \dots, t_{m+1}\}$ for movement to be simulated over (containing at least all behavioural switch times)

10: Simulate movement according to Algorithm 1, with movement parameters at each time t corresponding to $B(t)$ (see Eq. 4.2).

11: **return** Simulated behaviour, bearings and steps $\{\mathbf{B}, \boldsymbol{\theta}, \mathbf{v}\}$ and corresponding locations $\{\mathbf{X}, \mathbf{Y}\}$

12: **end procedure**

4.3 Extending the method for fully Bayesian inference

As in Chap. 3, observed locations \mathbf{Z} are taken at a finite, but irregular, series of times \mathbf{T} . As before, the relationship between the locations and parameters when the bearing and speed processes are unobserved makes the likelihood intractable. This is further complicated by the unobserved behavioural process, where there is the possibility of multiple switches between observations. This section describes the extension to the Bayesian inference algorithm of Chap. 3 that can be used given observed locations and a pre-specified number of behavioural states, p .

4.3.1 Data augmentation approach

The data augmentation approach of Blackwell (2003) is taken by supplementing the observed locations with the times of all behavioural switches. Here, as in the single state case in Chap. 3, augmentation also includes an approximation to the underlying bearing and speed processes on some (arbitrarily fine) time scale, as shown in Fig. 4.3. The augmented model is also given as a DAG in Fig. 4.4. The hybrid MCMC algorithm splits the quantities of interest into groups to update separately, in each case conditional on all other quantities. In cases where the full conditional distribution can be directly sampled from, Gibbs sampling is employed, and in all other scenarios the MH sampler is used (Gelman et al., 2013). The groups to be separately sampled from are:

- the behavioural parameters (Φ_B),
- the movement parameters (Φ_M) (split into bearing and speed processes), and
- the unobserved refined path of behavioural switches, bearings and speeds ($\mathbf{B}, \boldsymbol{\theta}, \mathbf{v}$).

Sects. 4.3.2 and 4.3.3 describe the sampling schemes used for the behavioural and movement parameters, respectively. In both cases the sampling is standard, employing Gibbs sampling and a random walk MH algorithm. Sect. 4.3.4 describes the MH algorithm used for the reconstruction of the unobserved refined path, which now includes the behavioural process in addition to the bearings and speeds. Although reconstruction of the bearings and speeds is at an (arbitrarily fine) approximation, the behavioural process remains exact.

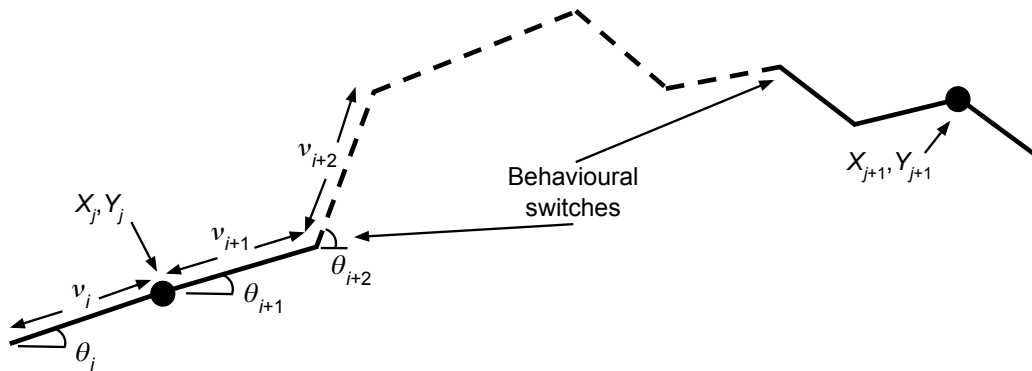


Fig. 4.3 Locations (X, Y) observed at the times T augmented with the unobserved refined movement path. The refined path consists of the behavioural process in continuous time (here shown as two behaviours: **solid** and **dashed lines**), and the bearings and steps (derived from the speed process) (θ, v) at the refined times t .

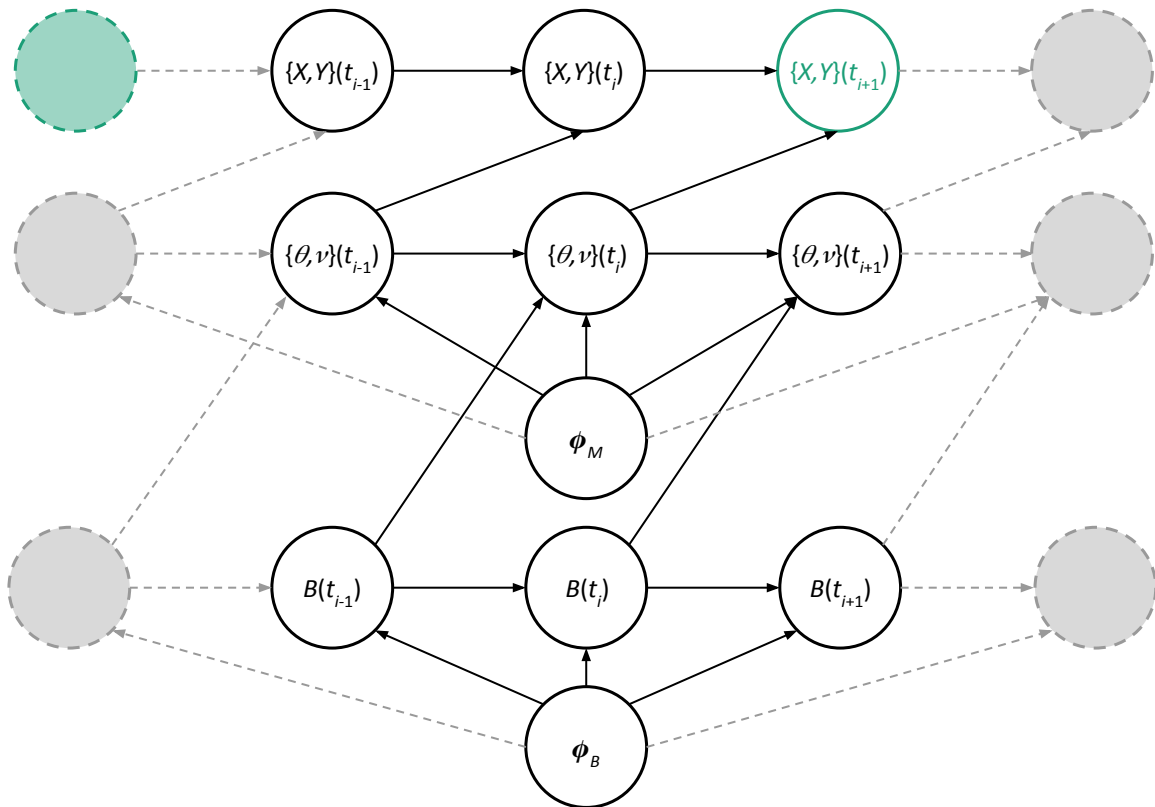


Fig. 4.4 DAG of the multistate movement model with augmentation at an approximate time scale. In this representation, locations $\{X, Y\}$ are shown corresponding to the approximate time scale created for the augmentation of the behaviours, bearings and steps $\{B, \theta, v\}$. Note that observed locations do not occur at each time shown here, but will be less frequently (for example, at the times shown in **green**) with the remaining locations augmented through the definition of the bearings and steps.

4.3.2 Sampling the behavioural process parameters

The behavioural process parameters are sampled conditional on complete observation of the behavioural process. Conjugate distributions for the switching rates ($\boldsymbol{\lambda}$) and probabilities (\boldsymbol{q}) of a CTMC are gamma and Dirichlet, respectively. Assuming such conjugate priors allows direct sampling from the posterior conditional as a Gibbs step (Blackwell, 2003), summarised below and in Algorithm 8.

The full conditional distribution is simplified as

$$p(\boldsymbol{\Phi}_B \mid \mathbf{B}, \boldsymbol{\theta}, \mathbf{v}, \mathbf{Z}, \boldsymbol{\Phi}_M) = p(\boldsymbol{\Phi}_B \mid \mathbf{B})$$

and is the posterior for a fully observed CTMC. Although \mathbf{B} describes the times and states of all switches, sufficient statistics for such a process are given by a_i , the total time spent in state i , and $b_{i,j}$, the number of transitions from state i to state j . Given independent prior distributions

$$\begin{aligned} \lambda_i &\sim \text{Gam}(c_i, d_i), \\ q_{i,1}, \dots, q_{i,p} &\sim \text{Dir}(\mathbf{f}_i), \end{aligned}$$

the posterior conditional distribution is

$$\begin{aligned} \lambda_i \mid \mathbf{B} &\sim \text{Gam}(c_i + \sum_{j=1}^p b_{i,j}, d_i + a_i), \\ q_{i,1}, \dots, q_{i,p} \mid \mathbf{B} &\sim \text{Dir}(\mathbf{f}_i + \mathbf{b}_i), \end{aligned}$$

for $i \in \{1, \dots, p\}$ and $\mathbf{b}_i = \{b_{i,1}, \dots, b_{i,p}\}$. Additional details showing the derivation of this posterior distribution are given in Appendix A.1.2.

4.3.3 Sampling the movement process parameters

The movement process parameters are sampled conditional on complete observation of the refined path (both behaviour and movement) and the behavioural parameters. The movement parameters are split into those relating to the bearing and speed processes and updated separately, described in the following.

Algorithm 8 Sample conditional behavioural process parameters.

-
- 1: **procedure** SAMPLE BEHAVIOUR PARAMETERS($\mathbf{B}, \{c_1, \dots, c_p\}, \{d_1, \dots, d_p\}, \{\mathbf{f}_1, \dots, \mathbf{f}_p\}$)
 - 2: Calculate sufficient statistics $\{a_1, \dots, a_p\}, \{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ from \mathbf{B}
 - 3: **for** $i \in 1, \dots, p$ **do**
 - 4: Calculate posterior parameters

$$c_i^* \leftarrow c_i + \sum_{j \neq i} b_{ij},$$

$$d_i^* \leftarrow d_i + a_i,$$

$$\mathbf{f}_i^* \leftarrow \mathbf{f}_i + \mathbf{b}_i.$$

- 5: Sample from the conditional posterior

$$\lambda_i \sim \text{Gam}(c_i^*, d_i^*),$$

$$q_{i1}, \dots, q_{ip} \sim \text{Dir}(\mathbf{f}_i^*).$$

- 6: **end for**
 - 7: **return** λ, \mathbf{q}
 - 8: **end procedure**
-

4.3.3.1 Sampling the bearing process parameters

There is a bearing process parameter describing the variance of the BM for each behavioural state. As in the single state method of Chap. 3, assuming an inverse gamma prior for this variance allows direct sampling from the posterior conditional distribution as a Gibbs step (as shown in Appendix A.1.1). In this case, the bearings parameters are independent and for each behavioural state will each have an inverse gamma prior, which need not be the same. Each variance parameter is then sampled sequentially by following the method in Sect. 3.3.2.1 and Algorithm 2. Assume the initial bearing, θ_1 , is uniformly distributed. For behavioural state i ,

$$p(\sigma_{\theta,i}^2 \mid \boldsymbol{\mu}, \boldsymbol{\beta}, \sigma_{\psi}^2, \boldsymbol{\Phi}_B, \mathbf{B}, \boldsymbol{\theta}, \mathbf{v}, \mathbf{Z}) = p(\sigma_{\theta,i}^2 \mid \boldsymbol{\theta}),$$

where

$$\sigma_{\theta,i}^2 \mid \boldsymbol{\theta} \sim \text{IG} \left(a_{\theta,i} + \frac{m_i}{2}, \quad b_{\theta,i} + \frac{1}{2} \sum_{\substack{j \\ B(t_j)=i}} \frac{\delta \theta_j}{\delta t_j} \right),$$

and where $\sigma_{\theta,i}^2 \sim \text{IG}(a_{\theta,i}, b_{\theta,i})$ and m_i is the number of points on the refined path following behavioural state i .

4.3.3.2 Sampling the speed process parameters

The speed process parameters for all behavioural states are updated simultaneously using a random walk MH step, with independent proposals for each parameter. Since all movement parameters are constrained to be positive, independent univariate Gaussians truncated below at zero are used as the proposal step in the random walk.

The posterior conditional distribution, up to a constant, is

$$\begin{aligned}
& p\left(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\sigma}_{\psi}^2 \mid \boldsymbol{\sigma}_{\theta}^2, \boldsymbol{\Phi}_B, \mathbf{B}, \boldsymbol{\theta}, \mathbf{v}, \mathbf{Z}\right) \\
&= p\left(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\sigma}_{\psi}^2 \mid \boldsymbol{\Phi}_B, \mathbf{B}, \mathbf{v}\right) \\
&\propto p\left(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\sigma}_{\psi}^2 \mid \boldsymbol{\Phi}_B, \mathbf{B}\right) p\left(\mathbf{v} \mid \boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\sigma}_{\psi}^2, \boldsymbol{\Phi}_B, \mathbf{B}\right) \\
&= p\left(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\sigma}_{\psi}^2\right) p\left(v_1 \mid \boldsymbol{\mu}, \boldsymbol{\sigma}_{\psi}^2, \boldsymbol{\Phi}_B\right) \prod_{i=2}^m p\left(v_i \mid v_{i-1}, B(t_{i-1}) = j, \mu_j, \beta_j, \sigma_{\psi,j}^2\right).
\end{aligned}$$

Here, $p\left(\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\sigma}_{\psi}^2\right)$ is the density of an appropriate prior distribution of the speed parameters, and

$$p\left(v_i \mid v_{i-1}, B(t_{i-1}) = j, \mu_j, \beta_j, \sigma_{\psi,j}^2\right),$$

is the conditional distribution of the steps, given by the distribution in Eq. 4.3. Assume the animal begins movement in equilibrium and that the initial step arises from the equilibrium of the OU speed process. The behaviour before the path begins is unknown (with no observed information relating to it). The density of the initial step is given by integrating over all possible behavioural states based on the equilibrium distribution of the Markov behavioural process,

$$p\left(v_1 \mid \boldsymbol{\mu}, \boldsymbol{\sigma}_{\psi}^2, \boldsymbol{\Phi}_B\right) = \sum_{j=1}^p p\left(v_1 \mid \mu_j, \sigma_{\psi,j}^2\right) p\left(j \mid \boldsymbol{\Phi}_B\right),$$

where

$$v_1 \mid \mu_j, \sigma_{\psi,j}^2 \sim (t_2 - t_1)N\left(\mu_j, \sigma_{\psi,j}^2\right),$$

and $p\left(j \mid \boldsymbol{\Phi}_B\right) = \pi_j$ is the probability of being in state j given that the CTMC with parameters $\boldsymbol{\Phi}_B$ is in equilibrium (see Eq. 4.1 for the equilibrium distribution of a CTMC).

In a simultaneous update of the speed parameters, the posterior conditional density is calculated for both the current and proposed values and combined with the proposal density (independent univariate truncated Gaussians in this case), similar to that summarised in Algorithm 3. The standard MH ratio is used to decide on the acceptance of the proposal.

4.3.4 Reconstructing the unobserved refined path

The central idea to this inference approach is the ability to reconstruct the unobserved path, conditional on the parameters. The approach of Chap. 3 is employed, but extended to also reconstruct the behavioural process. This section will outline reconstruction between three consecutive observations, but as in the previous chapter, this can be extended to span longer intervals if necessary. This scenario is displayed in Fig. 4.5, with three fixed, consecutive observations (**points**). The longer path section is required in this case because the behavioural process is reconstructed in continuous time, and so is fixed exactly at the endpoints on the chosen section. If only two consecutive observations were used, then the behaviour at observation times would remain constant.

The quantities simulated are those displayed in **black** in Fig. 4.5, consisting of:

- the behavioural process B between the times j and l ,
- the bearings $\{\theta_1, \dots, \theta_{n-1}\}$, and
- the steps $\{v_1, \dots, v_{n-1}\}$.

The set of refined times $\{t_1 = j, \dots, t_{n-1}\}$ that the bearings and step processes are approximated over are no longer fixed (as in Chap. 3) and will be discussed further below. Note that although an approximation is made to the bearing and step processes, the behavioural process remains exact. The set of $\{\theta_1, \dots, \theta_{\tau-1}\}$ and $\{v_1, \dots, v_{\tau-1}\}$ take the path to the observed location at the observation time $t_\tau = k$. Similarly, the set of $\{\theta_\tau, \dots, \theta_{n-1}\}$ and $\{v_\tau, \dots, v_{n-1}\}$ then take the path to the fixed location at the time $t_n = l$.

The fixed values conditioned upon are displayed in **green** in Fig. 4.5, consisting of:

- the observed locations at the ends of, and within, the path section $\{\mathbf{Z}(j), \mathbf{Z}(k), \mathbf{Z}(l)\}$
- the behaviours $\{B(j), B(l)\}$ at the times j, l ,
- the bearings $\{\theta_0, \theta_n\}$ at the times $\{t_0, t_n\}$, and
- the steps $\{v_0, v_n\}$ at the times $\{t_0, t_n\}$.

As the bearing and step processes are given by an approximation, the fixed points are the values of the respective process at the refined point immediately before and after the path section of interest, as in Figure 4.5. In contrast, because the behavioural process remains exact, the fixed points are at the times j and l .

The reconstruction of the refined path is sampled by an MH step, with proposal distribution described in the following.

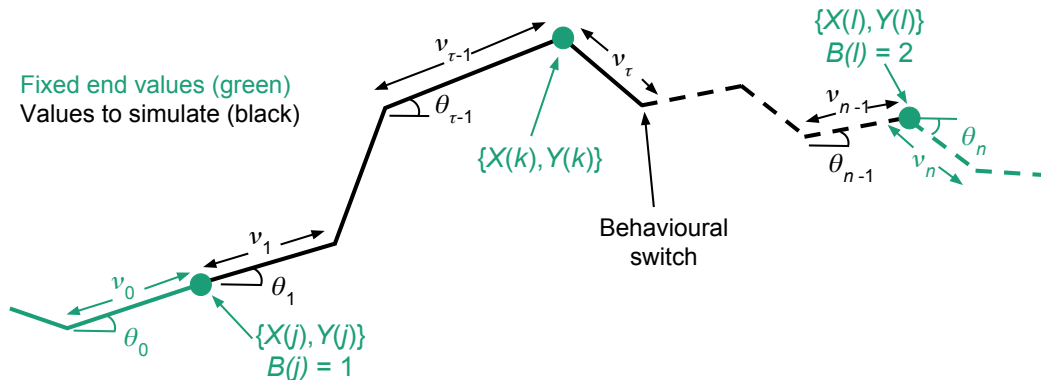


Fig. 4.5 Section of the full refined path to update over when there are multiple behavioural states. Fixed endpoint locations are given at the times j and l . The behavioural process (represented here by **solid** (state 1) and **dashed** (state 2)) is simulated with fixed endpoints $\{B(j), B(l)\}$. The bearing and step processes, $\{\theta_1, \dots, \theta_{n-1}, v_1, \dots, v_{n-1}\}$, are simulated given fixed endpoints $\{\theta_0, \theta_n, v_0, v_n\}$. All simulations are also conditioned upon the fixed observed location at the time k .

4.3.4.1 Simulating the behavioural proposal

A behavioural proposal B^* is simulated between the times j and l , given known values $\{B(j), B(l)\}$ and parameters Φ_B . The distribution of this process is a CTMC bridge, and is simulated here by a rejection method.

A CTMC with parameters Φ_B with state $B(j)$ at time j is simulated until time l (see method in Sect. 4.2 and lines 3–8 of Algorithm 7). If the final state is not equal to $B(l)$, then the proposal is instantly rejected. Otherwise, the path proposal continues (still with the possibility of rejection when considering the complete path proposal in the MH step). Less ‘naive’ approaches to this simulation, explicitly taking into account the endpoint, could be implemented (see e.g. Hobolth and Stone (2009); Rao and Teh (2013); Whitaker et al. (2017)), however this simple method performed well in our examples that follow.

4.3.4.2 Setting the refined time scale

Given the behavioural simulation, the set of refined times $\{t_1, \dots, t_{n-1}\}$ is created. This sequence can no longer be fixed before implementing the MCMC sampler, as in the previous chapter, because it must include the times of behavioural switches, which change (having just simulated a new realisation). There are numerous ways to choose such a set of times, as long as switch and observation times are included. In the examples that follow, the sequence is chosen to include all necessary times, whilst remaining approximately on some time scale

δt , the choice of which can be made prior to inference. The sequence $\{t_1, \dots, t_{n-1}\}$ forms the times to simulate the bearings and speed over, as in Fig. 4.5.

4.3.4.3 Simulating the bearing proposal

The bearing proposal $\boldsymbol{\theta}^*$ over the times $\{t_1, \dots, t_{n-1}\}$ is simulated conditional on the fixed bearings $\{\theta_0, \theta_n\}$ at the times $\{t_0, t_n = l\}$, the proposed behaviours \mathbf{B}^* and the parameters Φ . The distribution of this process is a BB with time-varying parameter, dependent on behaviour. To simulate from such a distribution, a transformation is implemented to account for the inhomogeneous variance parameter.

The times $\{t_1, \dots, t_{n-1}, t_n\}$ are transformed, weighted by the bearing variance for the behavioural state at each corresponding time, to give a process with a homogeneous variance of one. Therefore, $t'_0 = t_0$ and

$$t'_i = t'_{i-1} + \sigma_{\theta, B(t_{i-1})}^2 (t_i - t_{i-1}),$$

for $i \in \{1, \dots, n\}$. The BB (with variance one) is simulated on the set of transformed times, given the values $\{\theta_0, \theta_n\}$ at the end times $\{t'_0, t'_n\}$. Hence

$$\theta_i^* = \theta_0 + M(t'_i - t_0) - \frac{t'_i - t_0}{t'_n - t_0} (M(t'_n - t_0) - \theta_n + \theta_0),$$

for $i \in \{1, \dots, n-1\}$, where $M(t)$ is standard BM (Iacus, 2008).

4.3.4.4 Simulating the step proposal

To propose the steps \mathbf{v}^* a joint distribution similar to Eq. 3.7 is constructed:

$$\left(\begin{array}{c} \mathbf{v} \\ (\mathbf{Z}(k), \mathbf{Z}(l))^T \end{array} \right) \mid \Phi, \boldsymbol{\theta}^*, \mathbf{B}^*, \mathcal{F} \sim \mathcal{N} \left(\begin{array}{c} \mathbf{m}_1 \\ \mathbf{m}_2 \end{array}, \begin{array}{cc} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2}^T & \Sigma_2 \end{array} \right), \quad (4.5)$$

where $\mathcal{F} = \{\mathbf{Z}(j), B(j), B(l), \theta_0, \theta_n, v_0, v_n\}$.

The marginal distribution of the steps in Eq. 4.5 is given by $\mathbf{v} \sim \mathcal{N}(\mathbf{m}_1, \Sigma_1)$. The form of \mathbf{m}_1, Σ_1 arises from the speed process (from which \mathbf{v} is derived) being an OU bridge with inhomogeneous parameters. The method to calculate this follows the approach in Sect. 3.4.4.2 whilst accounting for the proposed behaviour at each time point. The fixed

values v_0, v_n are transformed to give speeds $\psi_0 = v_0/\delta t_0$ and $\psi_n = v_n/\delta t_n$. In the following, assume (for brevity) that all expressions are conditional upon

$$\{\Phi, \mathbf{B}^*, \boldsymbol{\theta}^*, \mathbf{Z}(j), B(j), B(l), \theta_0, \theta_n\}.$$

By noting that

$$\psi_i | \psi_{i-1} = Q_i \psi_{i-1} + N\left(\mu_{B(t_{i-1})} (1 - Q_i), \sigma_{\psi, B(t_{i-1})}^2 (1 - Q_i^2)\right),$$

where

$$Q_i = e^{-\beta_{B(t_{i-1})}(t_i - t_{i-1})},$$

the joint distribution $\{\psi_1, \dots, \psi_n\} | \psi_0$ can be constructed iteratively with

$$\begin{aligned} E(\psi_i | \psi_0) &= Q_i E(\psi_{i-1} | \psi_0) + \mu (1 - Q_i), \\ \text{Var}(\psi_i | \psi_0) &= Q_i^2 \text{Var}(\psi_{i-1} | \psi_0) + \sigma_{\psi}^2 (1 - Q_i^2), \\ \text{Cov}(\psi_i, \psi_{i+j} | \psi_0) &= Q_i \text{Cov}(\psi_i, \psi_{i+j-1} | \psi_0). \end{aligned}$$

This joint distribution is partitioned into $\{\psi_1, \dots, \psi_{n-1}\}$ and ψ_n to condition upon the known value of ψ_n using standard conditioning of a multivariate normal (Eaton, 2007) to give the joint distribution $\{\psi_1, \dots, \psi_{n-1}\} | \psi_0, \psi_n$. This distribution can be transformed back to steps $\{v_1, \dots, v_{n-1}\}$ (giving \mathbf{m}_1, Σ_1) by multiplying the speeds $\{\psi_1, \dots, \psi_{n-1}\}$ by the times $\{\delta t_1, \dots, \delta t_{n-1}\}$.

The locations $(\mathbf{Z}(k), \mathbf{Z}(l))^T$ can be written as

$$\begin{aligned} \begin{pmatrix} \mathbf{Z}(k) \\ \mathbf{Z}(l) \end{pmatrix} &= \begin{pmatrix} \mathbf{Z}(j) \\ \mathbf{Z}(j) \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^{\tau-1} \begin{pmatrix} v_i \cos(\theta_i^*) \\ v_i \sin(\theta_i^*) \end{pmatrix} \\ \sum_{i=1}^{n-1} \begin{pmatrix} v_i \cos(\theta_i^*) \\ v_i \sin(\theta_i^*) \end{pmatrix} \end{pmatrix} \\ &= (\mathbf{Z}(j), \mathbf{Z}(j))^T + A\mathbf{v}, \end{aligned} \tag{4.6}$$

where

$$A = \begin{pmatrix} \cos(\theta_1^*) & \cdots & \cos(\theta_{\tau-1}^*) & 0 & \cdots & 0 \\ \sin(\theta_1^*) & \cdots & \sin(\theta_{\tau-1}^*) & 0 & \cdots & 0 \\ \cos(\theta_1^*) & \cdots & \cos(\theta_{\tau-1}^*) & \cos(\theta_{\tau}^*) & \cdots & \cos(\theta_{n-1}^*) \\ \sin(\theta_1^*) & \cdots & \sin(\theta_{\tau-1}^*) & \sin(\theta_{\tau}^*) & \cdots & \sin(\theta_{n-1}^*) \end{pmatrix}.$$

As in Sect. 3.3.3.1, the marginal distribution of the locations in Eq. 4.5 is

$$(\mathbf{Z}(k), \mathbf{Z}(l))^T \sim \mathbf{N}(\mathbf{m}_2, \Sigma_2), \quad (4.7)$$

and the covariance between \mathbf{v} and $(\mathbf{Z}(k), \mathbf{Z}(l))^T$ is $\Sigma_{1,2}$. The means and covariances in Eq. 4.5 are calculated as

$$\mathbf{m}_2 = (\mathbf{Z}(j), \mathbf{Z}(j))^T + A\mathbf{m}_1, \quad (4.8a)$$

$$\Sigma_{1,2} = \Sigma_1 A^T, \quad (4.8b)$$

$$\Sigma_2 = A\Sigma_1 A^T. \quad (4.8c)$$

The step process proposal is simulated by further conditioning \mathbf{v} in Eq. 4.5 upon the locations $(\mathbf{Z}(k), \mathbf{Z}(l))^T$. This ensures that the proposed path section will pass through the observation at time k and meet up with the fixed end location at time l to agree with the full path. The sought conditional distribution (Eaton, 2007) is

$$\begin{aligned} \mathbf{v} \mid \Phi, \mathbf{B}^*, \boldsymbol{\theta}^*, \mathcal{F}, \mathbf{Z}(k), \mathbf{Z}(l) \\ \sim \mathbf{N} \left(\mathbf{m}_1 + \Sigma_{1,2} \Sigma_2^{-1} \left((\mathbf{Z}(k), \mathbf{Z}(l))^T - \mathbf{m}_2 \right), \Sigma_1 - \Sigma_{1,2} \Sigma_2^{-1} \Sigma_{1,2}^T \right) \\ = \mathbf{N} \left(\hat{\boldsymbol{\mu}}, \hat{\Sigma} \right). \end{aligned} \quad (4.9)$$

This distribution is singular, with the rank of $\hat{\Sigma}$ having been reduced by four through conditioning upon two known locations. Simulation can be carried out following the method in Sect. 3.3.3.1, by first sampling from the marginal distribution

$$\mathbf{x}^* \sim \mathbf{N}(\mathbf{m}_1, \Sigma_1),$$

and then accounting for the linear constraint by setting

$$\mathbf{v}^* = \mathbf{x}^* - \Sigma_{1,2} \Sigma_2^{-1} \left((\mathbf{Z}(j), \mathbf{Z}(j))^T + A\mathbf{x}^* - (\mathbf{Z}(k), \mathbf{Z}(l))^T \right).$$

4.3.4.5 Simulating a proposal at the start/end of path

An exception to the given path proposal method occurs when the section is at the end of the full movement path, for which there is no fixed behaviour, bearing or step at the end of the section (otherwise, those quantities would never change). In such a case, the simulation of

the behaviour proposal is a CTMC forwards in time but without the need to carry out the rejection step described above. The bearing proposal is simulated by transforming the set of times to create BM with constant variance of one, but then simulating merely from BM on these transformed times rather than a bridge, i.e. $\theta_i^* = \theta_0 + M(t_i' - t_0)$. Lastly, the marginal distribution of the steps is created as described above, but without the final step to condition upon the fixed ψ_n . The acceptance/rejection of the proposed path section is then the same as that in the following.

In a similar fashion, if the section is at the start of the full movement path, the simulation must be altered. The behavioural process is simulated backwards in time from the fixed end point of the section. Simulating from the reverse-time CTMC is equivalent to simulating from the chain with parameters $\{\hat{\lambda}, \hat{q}\}$, where

$$\hat{\lambda}_i = \lambda_i, \quad \hat{q}_{ij} = q_{ji} \frac{\pi_i}{\pi_j},$$

and π is the stationary distribution of the CTMC. Similarly, the bearings are simulated from BM on the transformed times but in reverse from the fixed end point, i.e. $\theta_i^* = \theta_n - M(t_n - t_i')$. The marginal step distribution is formed following the same approach as that described, but in which the Q_i is calculated by taking the OU process in reverse-time so that

$$\psi_i | \psi_{i+1} = Q_i \psi_{i+1} + N\left(\mu_{B(t_i)} (1 - Q_i), \sigma_{\psi, B(t_i)}^2 (1 - Q_i^2)\right),$$

where

$$Q_i = e^{-\beta_{B(t_i)}(t_{i+1} - t_i)},$$

and again with no fixed ψ_0 to condition upon in the final step. The acceptance/rejection of the proposal is the same as that below.

4.3.4.6 Accepting a refined path proposal

The path proposal method described does not take into account the fixed location at the end of the section when simulating the behaviours or the bearings. An MH step assesses whether

this proposal is accepted. The conditional distribution, up to a constant, is

$$\begin{aligned}
& p(\mathbf{B}^*, \boldsymbol{\theta}^*, \mathbf{v}^* \mid \Phi, \mathcal{F}, \mathbf{Z}(k), \mathbf{Z}(l)) \\
&= p(\mathbf{B}^*, \boldsymbol{\theta}^* \mid \Phi, \mathcal{F}, \mathbf{Z}(k), \mathbf{Z}(l)) p(\mathbf{v}^* \mid \mathbf{B}^*, \boldsymbol{\theta}^*, \Phi, \mathcal{F}, \mathbf{Z}(k), \mathbf{Z}(l)) \\
&\propto p(\mathbf{B}^*, \boldsymbol{\theta}^* \mid \Phi, \mathcal{F}) p(\mathbf{Z}(k), \mathbf{Z}(l) \mid \mathbf{B}^*, \boldsymbol{\theta}^*, \Phi, \mathcal{F}) p(\mathbf{v}^* \mid \mathbf{B}^*, \boldsymbol{\theta}^*, \Phi, \mathcal{F}, \mathbf{Z}(k), \mathbf{Z}(l)) \\
&= p(\mathbf{B}^* \mid \Phi, \mathcal{F}) p(\boldsymbol{\theta} \mid \mathbf{B}^*, \Phi, \mathcal{F}) p(\mathbf{Z}(k), \mathbf{Z}(l) \mid \mathbf{B}^*, \boldsymbol{\theta}^*, \Phi, \mathcal{F}) p(\mathbf{v}^* \mid \mathbf{B}^*, \boldsymbol{\theta}^*, \Phi, \mathcal{F}, \mathbf{Z}(k), \mathbf{Z}(l)).
\end{aligned}$$

The simulation method employed to create the proposal has density proportional to

$$p(\mathbf{B}^* \mid \Phi, \mathcal{F}) p(\boldsymbol{\theta} \mid \mathbf{B}^*, \Phi, \mathcal{F}) p(\mathbf{v}^* \mid \mathbf{B}^*, \boldsymbol{\theta}^*, \Phi, \mathcal{F}, \mathbf{Z}(k), \mathbf{Z}(l)).$$

The MH acceptance ratio involves the probability of the observed and fixed end location, given behaviour and bearings, requiring the distribution of $(\mathbf{Z}(k), \mathbf{Z}(l))^T$ in Eq. 4.7,

$$\frac{p\left((\mathbf{Z}(k), \mathbf{Z}(l))^T \mid \mathbf{B}^*, \boldsymbol{\theta}^*, \Phi, \mathcal{F}\right)}{p\left((\mathbf{Z}(k), \mathbf{Z}(l))^T \mid \mathbf{B}, \boldsymbol{\theta}, \Phi, \mathcal{F}\right)},$$

where \mathbf{m}_2, Σ_2 must be constructed in each case given the corresponding set of behaviours and bearings (current and proposed).

4.4 Simulated example

The following example applies the methods of this chapter to a simulated dataset with two behavioural states. This example demonstrates the ability of the algorithm to estimate the underlying parameters and reconstruct the unobserved behavioural process. The following results can be reproduced with the `multi_simulation` example within the repository `CTStepTurn` available at <https://github.com/a-parton/CTStepTurn>.

4.4.1 Underlying movement and observations

Two-state movement following Sect. 4.1 was simulated according to Sect. 4.2 with parameters

$$\begin{aligned}\Phi_B &= \{\lambda_1 = 0.01, \lambda_2 = 0.015\}, \\ \Phi_M &= \{\sigma_{\theta,1}^2 = 0.8, \mu_1 = 15, \beta_1 = 0.07, \sigma_{\psi,1}^2 = 40, \\ &\quad \sigma_{\theta,2}^2 = 0.05, \mu_2 = 40, \beta_2 = 0.02, \sigma_{\psi,2}^2 = 400\},\end{aligned}$$

between the times 0–1000 at an approximate time scale of around 0.01. Note that the actual time scale in this case was not always exactly 0.01 because behavioural switching occurs continuously in time and any switch time is included in the simulated time scale. The set of parameters chosen were based on the results of pilot runs carried out on movement data of a lesser black-backed gull (for details on this data, see the example of Sect. 5.4 and Garthe et al. (2016b)), with behavioural state 1 describing tortuous movement with low speed and state 2 describing directed movement with high speed that is highly variable in the long term but strongly correlated in the short term.

Observations of the movement path were created by sub-sampling the simulation at intervals of 500, giving 200 observations at a time scale of 5. The simulated path, with corresponding observations, is given in Figs. 4.6 and 4.7 as locations and bearings/speeds, respectively. Note that although the observations are shown to be coloured by underlying behavioural state, the only information used in this example is location, with the behaviour left unobserved. However, it is assumed known that there are two behavioural states present.

From Fig. 4.7, although the mean speed in state 2 is much higher than state 1, the high volatility of the speed process leads to bouts of simulated speeds that are of the same magnitude as the slower state 1. The behavioural parameters lead to longer average residency times in state 1 than state 2, including a short residencies in state 1 (times 0–11.3) and state 2 (times 576.3–582.3), which will be informative about the limitations of the algorithm when reconstructing the unobserved behavioural process. Because behavioural switches can occur between observation times, the uncertainty around the exact time of switches will be investigated in the following.

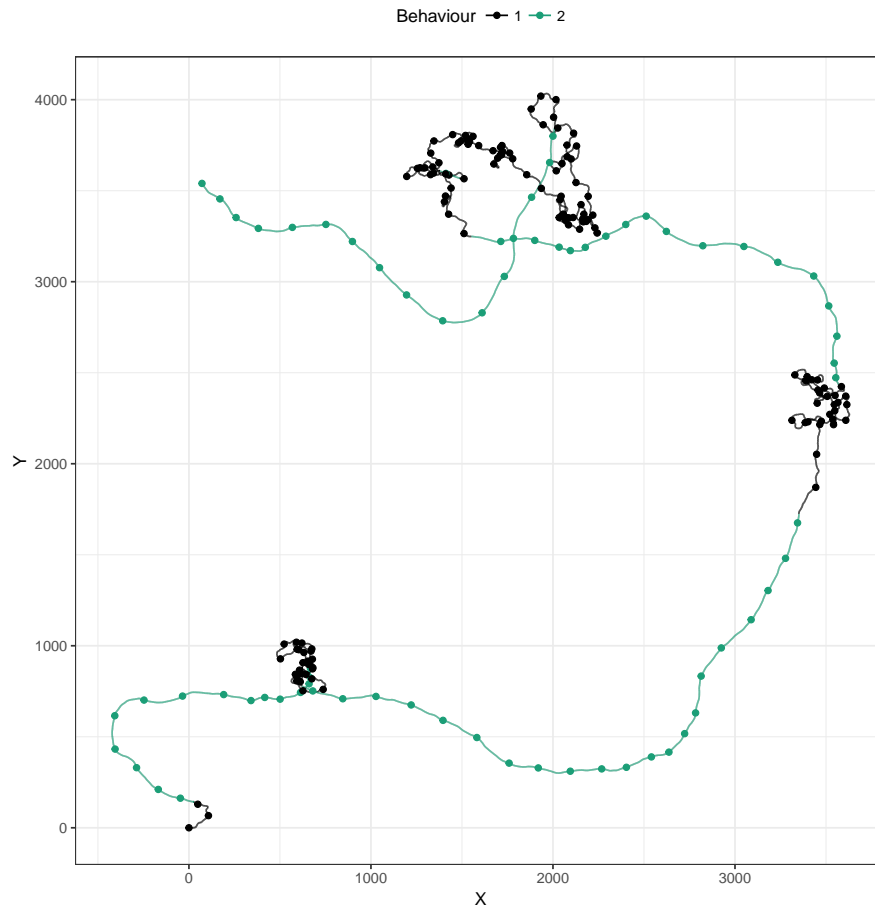


Fig. 4.6 Simulated two-state movement path, coloured by behaviour (**line**). Observed locations are included (**points**), coloured here by corresponding behavioural state. Note that the observations used in this example are only locations, and the behaviour is unknown.

4.4.2 Implementing the inference algorithm

4.4.2.1 Initial values

An initial movement path was created at a refined time scale of one by taking an interpolating cubic spline between observed locations. This gave a reconstruction of around four unknown locations between each pair of observations. Although this is a small refinement to the set of observations, it was shown in Sect. 3.5.2 that further refinement improved parameter estimation marginally but at the cost of a significant increase in computational speed. The reconstruction scale implemented here was chosen to provide computational speed with the assumption that a better approximation to the underlying continuous-time process would

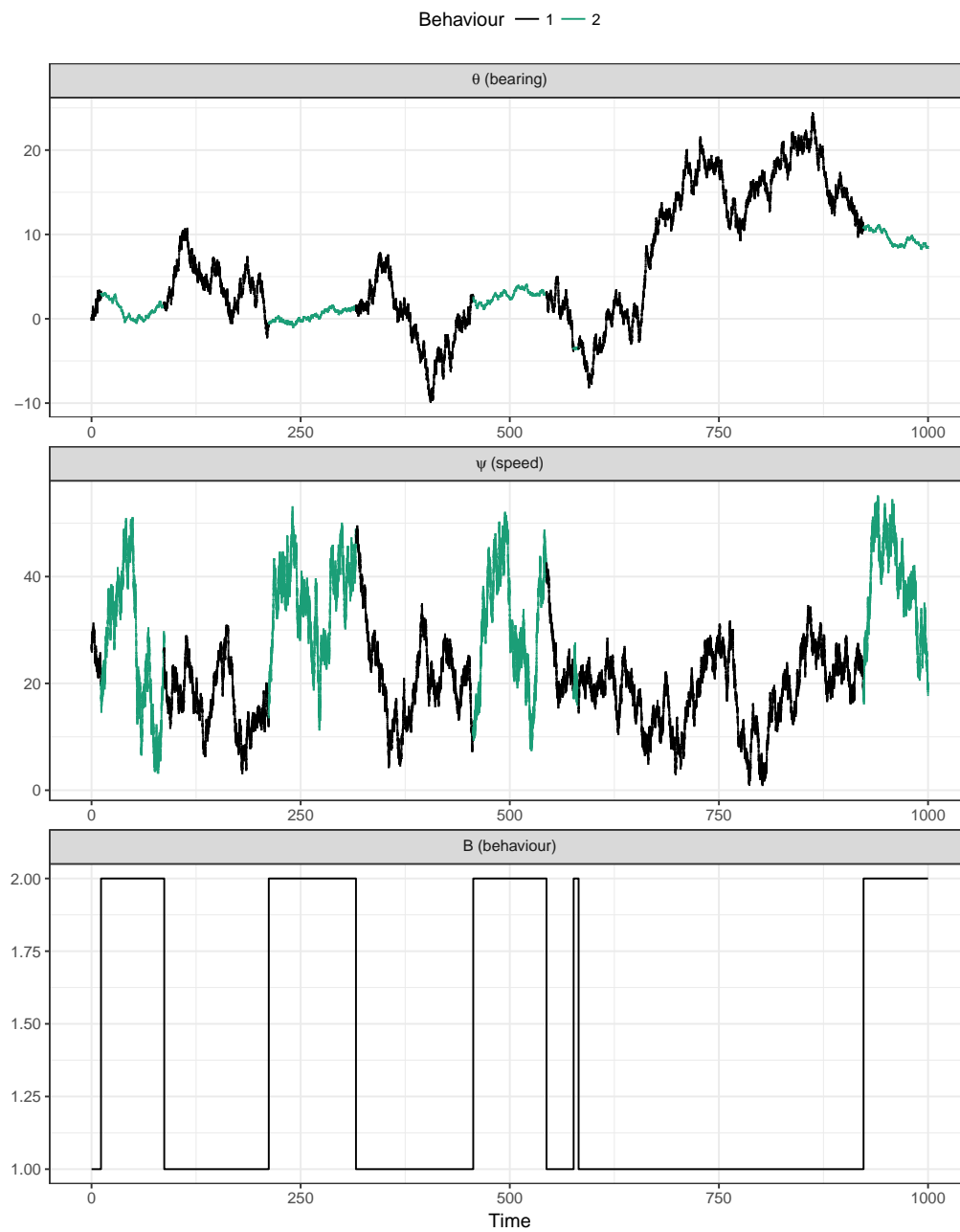


Fig. 4.7 Simulated two-state movement path. **Top and middle:** Simulated bearings and speeds, respectively, coloured by behavioural state. **Bottom:** Simulated behavioural process.

not make significant improvement to parameter estimation. The choice of refinement in the multistate case is explored further in the example of Sect. 5.4.

The initial behavioural process at corresponding times was set by identifying any points on the movement path with speed above 20 and setting these as behavioural state 2. This initial configuration is shown in Fig. 4.8 as bearings/speeds and as locations in the Appendix, Fig. B.8. Comparing Figs. 4.7 and 4.8, the initial speed process can be seen to replicate the overall trends of the true simulation, but the bearing and behavioural configurations are different to the true simulation, with a much higher switching rate.

Initial movement parameters were set as estimates from this initial configuration, given as

$$\begin{aligned} &\{\sigma_{\theta,1}^2{}^{(0)} = 0.271, \mu_1^{(0)} = 11.6, \beta_1^{(0)} = 0.109, \sigma_{\psi,1}^2{}^{(0)} = 23.9\}, \\ &\{\sigma_{\theta,2}^2{}^{(0)} = 0.0138, \mu_2^{(0)} = 31.9, \beta_2^{(0)} = 0.0151, \sigma_{\psi,2}^2{}^{(0)} = 75.7\}. \end{aligned}$$

Note that initial behavioural parameters need not be guessed as they can be sampled as a Gibbs step based on the initial path configuration.

4.4.2.2 Prior information

To avoid state label switching, a prior was placed on the movement parameters specifying that $\mu_1 < \mu_2$, so that state 1 would remain as the slower state. Uninformative priors were set for the behavioural switching parameters, λ_1, λ_2 , and the bearing parameters, $\sigma_{\theta,1}^2, \sigma_{\theta,2}^2$, given by gamma distributions with shape and rate both equal to 0.001. The speed parameters all had flat prior distributions.

4.4.2.3 Implementation

The algorithm of Sect. 4.3 was applied for 10^7 iterations, with each iteration consisting of a single parameter update and 100 refined path updates on random sections of the path with lengths between 4–14 points. Perturbation standard deviations for the speed parameters were based on pilot runs as

$$\begin{aligned} &\{\rho_{\mu,1} = 0.13, \rho_{\beta,1} = 0.017, \rho_{\sigma_{\psi,1}^2} = 2.7\}, \\ &\{\rho_{\mu,2} = 0.18, \rho_{\beta,2} = 0.0144, \rho_{\sigma_{\psi,2}^2} = 2.1\}. \end{aligned}$$

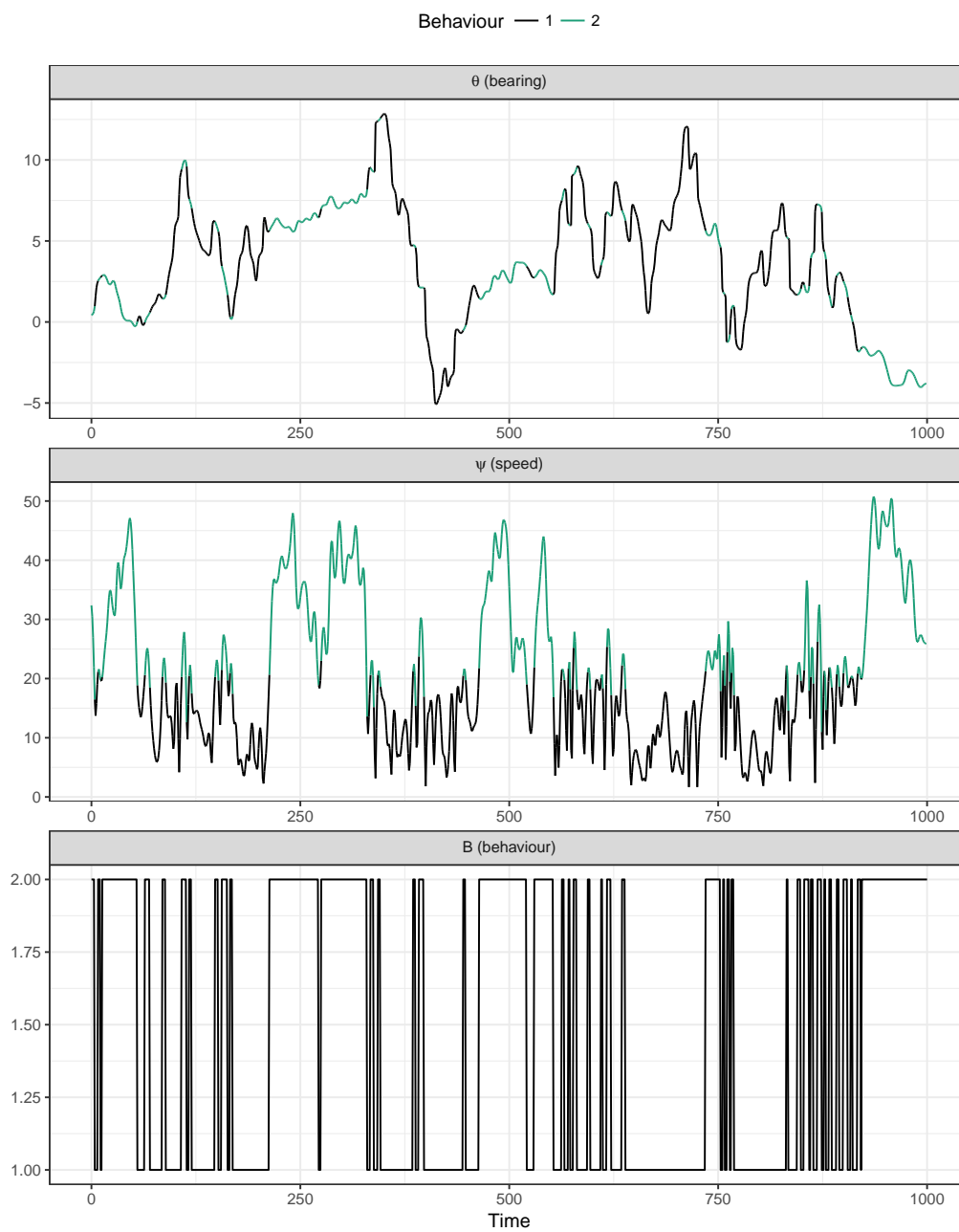


Fig. 4.8 Initial configuration of the two-state movement path. **Top and middle:** Initial bearings and speeds, respectively, coloured by behavioural state. **Bottom:** Initial behavioural process.

The resulting acceptance rates were 30% and 3.2% for the speed parameters and refined movement path, respectively. The rejection sampling of the behavioural process proposal resulting in less than 20% of path proposals being instantly rejected. Samples were thinned (for memory storage purposes) by a factor of 1,000 with the first 4,000 samples treated as burn-in, leaving 6,000 samples for estimation with ESS given in Tables 4.1 and 4.2. Computational run-time for this implementation was between 3–4 days.

4.4.3 Behavioural parameter results

Posterior samples of the (log) behavioural parameters are in Fig. 4.9, kernel density estimates are given in the **bottom panels** of Fig. 4.11, and the sample trace in the Appendix, Fig. B.9. Posterior summary statistics are given in Table 4.1. For both states, the posterior mode is an underestimate of the true switching rate, but the true value is captured within the posterior estimates (contained within the 95% credible intervals in both cases).

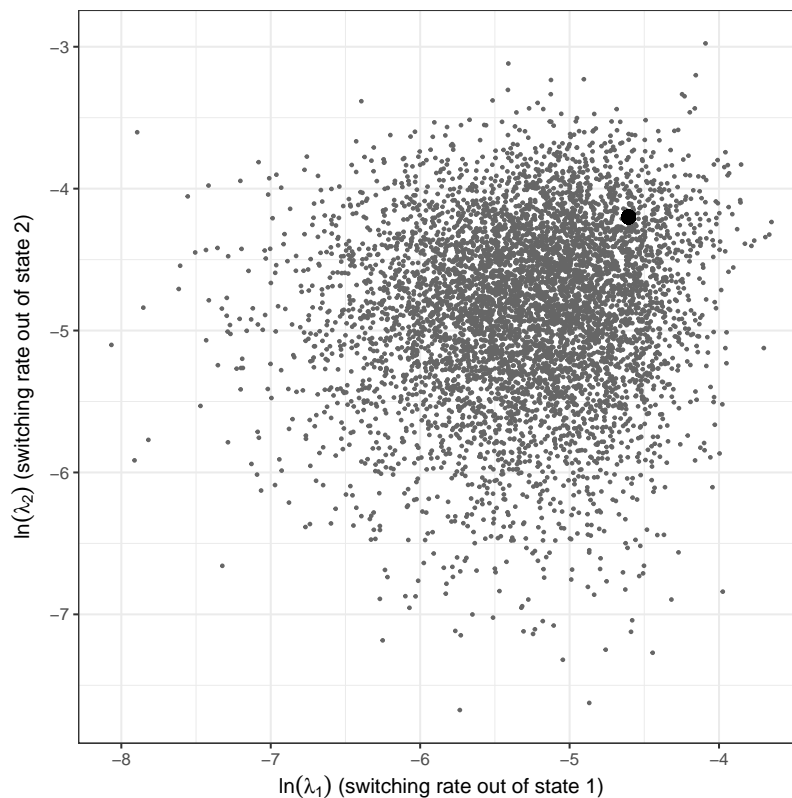


Fig. 4.9 Posterior samples of the behavioural parameters (on a log scale) for the two-state simulation example (**grey points**); λ_1 is the switching rate out of state 1 and λ_2 is the switching rate out of state 2. True values are highlighted (**black point**).

Table 4.1 Posterior credible intervals (2.5%, 50%, 97.5% quantiles) and ESS of the sampled behavioural parameters in the two-state simulation example. True values were $\{0.01, 0.015\}$. The ESS is taken from the 6,000 samples thinned from 6×10^6 iterations.

	2.5%	50%	97.5%	ESS
λ_1	0.00134	0.00531	0.0140	583
λ_2	0.00189	0.00822	0.0224	840

4.4.4 Movement parameter results

Posterior samples of the (log) movement parameters are in Fig. 4.10, kernel density estimates are given in the **top four rows** of Fig. 4.11, and the sample trace in the Appendix, Fig. B.9. The **left panel** of Fig. 4.10 shows the bearing parameter against the mean speed, both of which are estimated well in each state.

The **right panel** of Fig. 4.10 shows the speed correlation against long term variance, with highly correlated samples. These lie along a linear relationship between the two parameters, with the true values being reasonably captured for state 1 (**grey**), but biased for state 2 (**green**). The linear relationship present corresponds to the short term variance of the speed

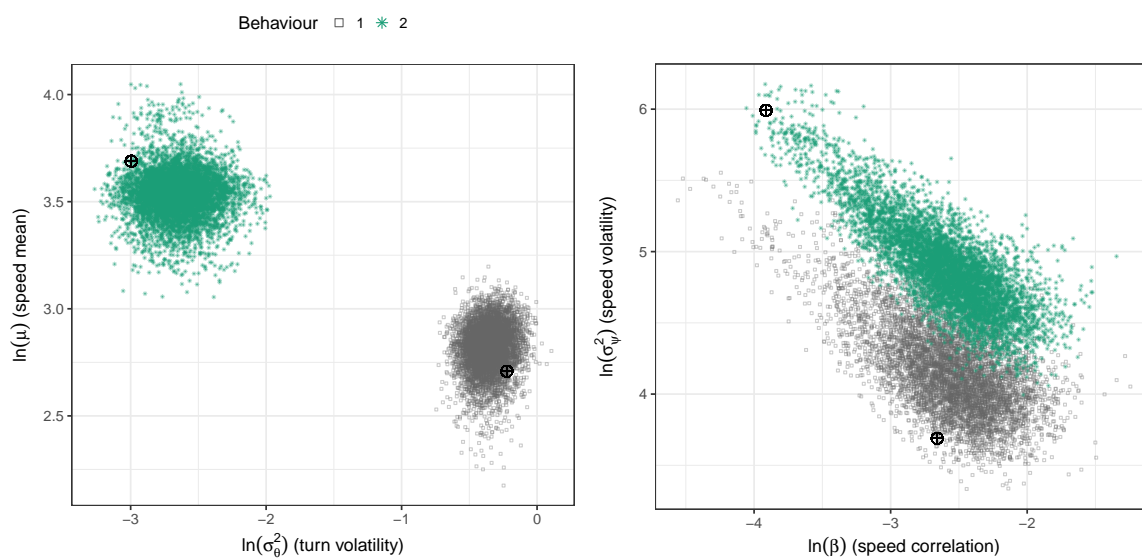


Fig. 4.10 Posterior samples of the movement parameters (on a log scale) for the two-state simulation example, coloured by behavioural state. True values are highlighted (**black target**).

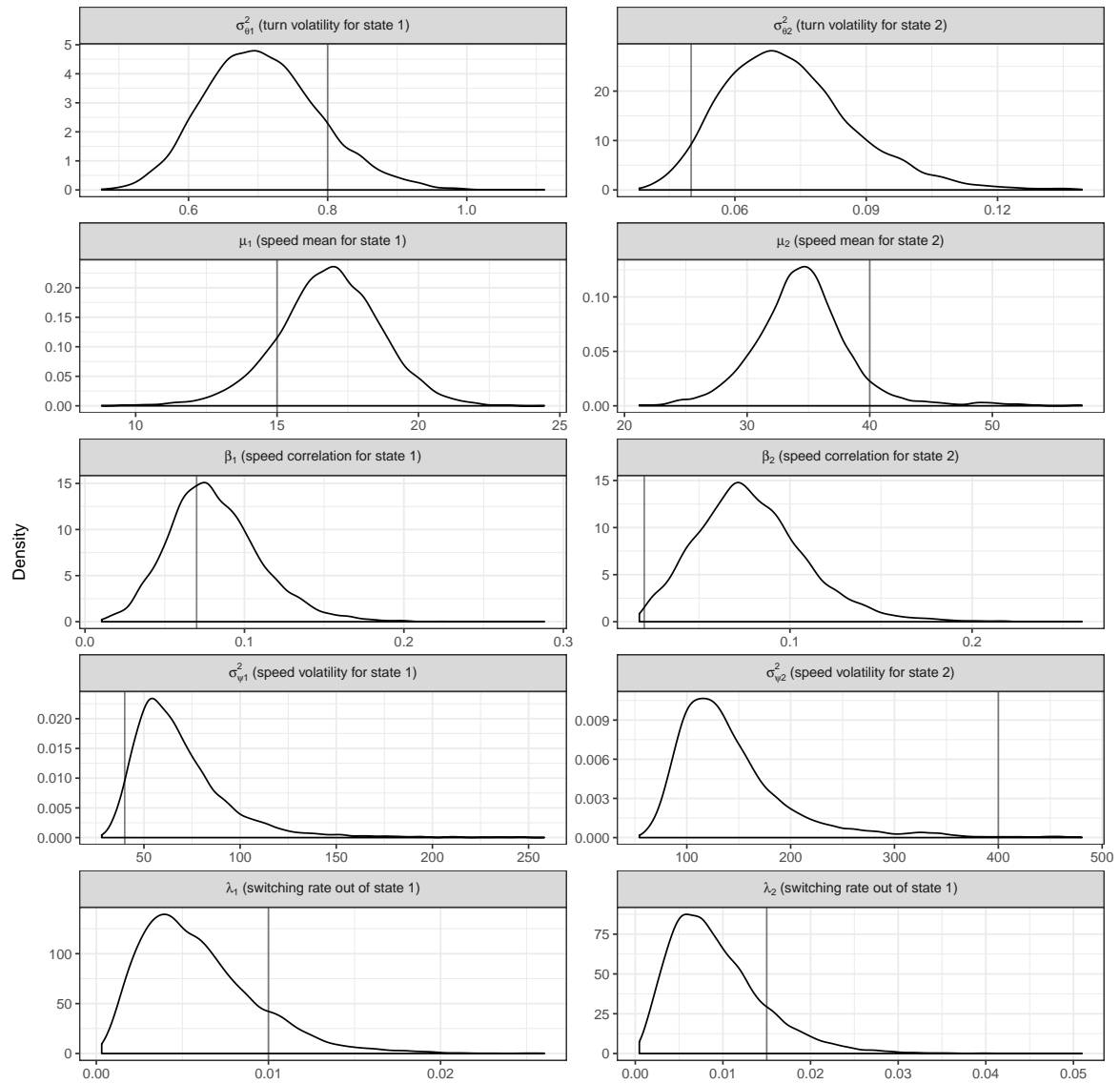


Fig. 4.11 Posterior kernel density estimates of the parameters for the two-state simulation example (**black**). True values are highlighted (**grey**).

process, where both the correlation and variance parameters feature; given as

$$\sigma_{\psi}^2 (1 - e^{-10\beta}),$$

for the observation time scale. The kernel density estimate for this movement measure is given in Fig. 4.12 and the sample trace in the Appendix, Fig. B.10, showing estimation for this feature to be much better than the long term speed variance for state 2. To understand why this is the case, consider the true value for the effect the correlation parameter has on the speed variance at the observation time scale ($1 - e^{-10\beta}$); given as 0.503 and 0.181 for the two behavioural states, respectively. For this quantity to be almost 1 (taken here to be over 0.98), and hence no correlation present because the short term speed variance is equal to the long term variance, the time scale would have to be 50 and 200, for the two states respectively. The residency times in state 1 from the true behavioural process were

$$\{11.3, 125, 140, 32.4, 341\}.$$

There are bouts of time spent in state 1 that are longer than this time period (50) so that correlation in the speed process has been able to deteriorate. For state 2, the true residency times were

$$\{75.6, 104, 87.6, 5.94, 77.1\},$$

and so there are no sojourns in state 2 that are longer than the time period required for that state (200). It is likely that this is the reason that both the long and short term speed variance can be estimated for state 1 (and β , σ_{ψ}^2 and $\sigma_{\psi}^2(1 - e^{-10\beta})$ can all be estimated), but for state 2 only the short term speed variance can be estimated ($\sigma_{\psi}^2(1 - e^{-10\beta})$) and not the long term variance (σ_{ψ}^2 , and hence separate β and σ_{ψ}^2).

Posterior summary statistics are given in Table 4.2 for the movement parameters, split by behavioural state. As discussed above, β and σ_{ψ}^2 cannot be estimated separately for the second state, and so the true values are not included in the credible intervals given. Additionally, the short term speed variance for state 1 is over-estimated and does not include the true value; this could be due to over-dispersed sampled steps in state 1 at uncertain behavioural switching times (i.e. larger steps being misclassified). The remaining parameters are estimated well and included in the credible intervals given.

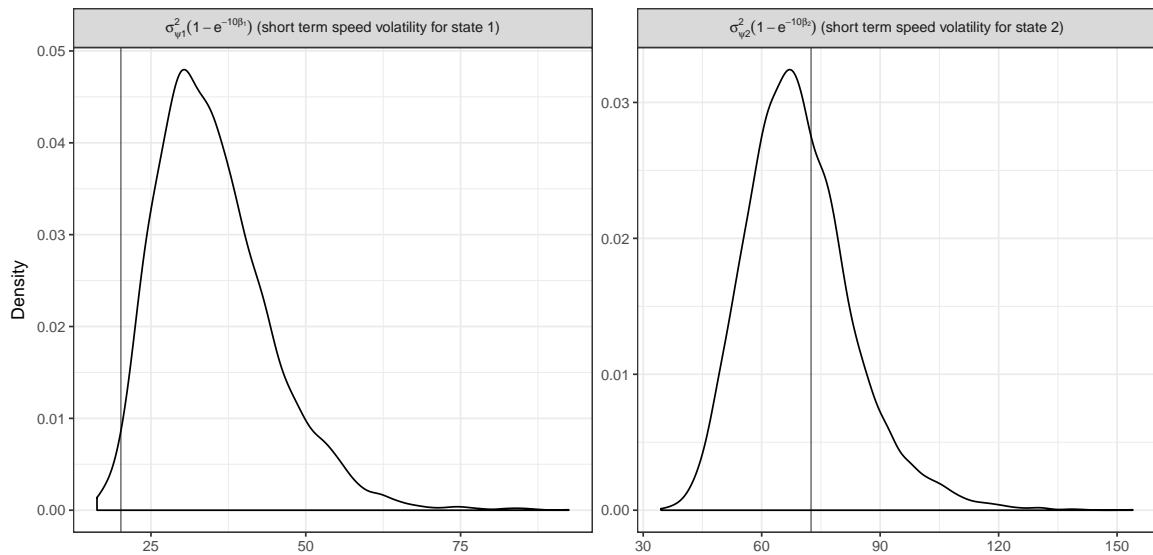


Fig. 4.12 Posterior kernel density estimates of the short-term speed variance (at the observation time scale) for the two-state simulation example (**black**). True values are highlighted (**grey**).

Table 4.2 Posterior credible intervals (2.5%, 50%, 97.5% quantiles) and ESS for the sampled movement parameters, split by state, in the two-state simulation example. True values are included, highlighted in **bold** for parameters not included in the 95% credible interval. The ESS is taken from the 6,000 samples thinned from 6×10^6 iterations.

	2.5%	50%	97.5%	True	ESS	
State 1	$\sigma_{\theta,1}^2$	0.561	0.702	0.884	0.8	584
	μ_1	13.1	16.9	20.3	15	711
	β_1	0.0330	0.0782	0.146	0.07	238
	$\sigma_{\psi,1}^2$	38.1	63.1	135	40	135
	$\sigma_{\psi,1}^2 (1 - e^{-10\beta_1})$	21.3	33.8	56.4	20.1	74
State 2	$\sigma_{\theta,2}^2$	0.0482	0.0706	0.107	0.05	584
	μ_2	27.0	34.4	43.4	40	193
	β_2	0.0279	0.0760	0.145	0.02	189
	$\sigma_{\psi,2}^2$	77.3	130	307	400	74
	$\sigma_{\psi,2}^2 (1 - e^{-10\beta_2})$	47.7	68.2	101	72.5	547

4.4.5 Path reconstruction results

Calculated from posterior samples of the reconstructed behavioural process, Fig. 4.13 shows the probability of being in each behaviour throughout the course of the sampling period. The algorithm has successfully identified the majority of the true behavioural process correctly. The behavioural process between the times 87–544, which starts in state 1 and includes two bouts of movement in state 2 before ending once again in state 1, has been estimated well. The posterior behavioural probabilities closely match the switching times in the true behavioural process, with little uncertainty surrounding them.

As an example of successful reconstruction, take the part of the movement path between the times 300–475, shown in Fig. 4.14. At the start of the period shown, the posterior probability of being in state 2 was correctly identified as approximately 1 (2/6,000 reconstructions were classified incorrectly as state 1). As time reached the true switch time, increasing numbers of the reconstructions identified a switch; at the true time of the switch around 40% of the reconstructions had switched into state 1 and over 95% of the reconstructions had switched by time 328 (2 observations after the true switch). For the full period between the times

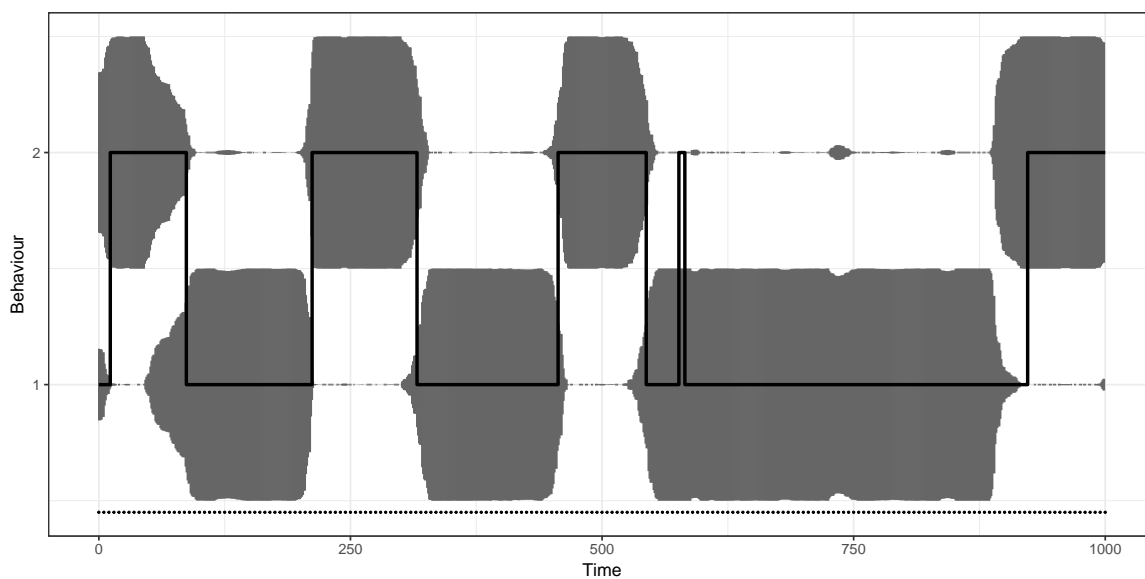


Fig. 4.13 Posterior probability of residing in each behavioural state over time, calculated from the sampled reconstructions of the behavioural process (**grey bars**). The taller a bar at a given time, the higher the proportion of sampled reconstructions classified as that behaviour (so each bar for state 1 and 2 sum to a constant value). The true, simulated behavioural process is included (**black line**) and **points** along the horizontal axis highlight the times/frequency of observations.

328–442 (the next switch occurred at time 456) the posterior probability of correctly residing in state 1 was over 98%. The probability of being in state 2 then begins to increase from this point, with 45% of the reconstructions having switched into state 2 by the time of the true switch, until over 95% of reconstructions had switched by time 466 (again, 2 observations after the true switch). This section of the full path demonstrates the ability of the algorithm to identify the underlying behavioural process, with uncertainty highest at the actual switch times, and spanning a period around the switch time of less than 20 in both cases.

From Fig. 4.13, the times of the switches for the two bouts in state 2 at the start and end of the path (starting at times 11 and 923), have not been estimated with as much reliability as that previously discussed. Although the presence of a switch is correctly identified there is a large amount of uncertainty as to when the switches occur. Further, the times with the highest uncertainty around the state of process are earlier than the true switch time in both cases; 50% of the reconstructions had switched to state 1 by time 75 when the true switch occurred at time 86.9 and, similarly, 50% had switched to state 2 by time 894 when the true switch occurred at time 922.9.

An example of a scenario in which the algorithm struggles to correctly identify the true behavioural process is given at the beginning of the movement path, as shown in Fig. 4.15. The true behavioural process started in state 1 but switched to state 2 at time 11, with only 3 observations being within the initial bout in state 1. Fig. 4.15 shows that the algorithm had difficulty correctly identifying this part of the behavioural process. At time 0, the posterior probability of correctly being in state 1 was only 31% and once the true switch occurred at time 11, the posterior probability of being in state 2 was over 90%. The presence of this bout in state 1 was considered, but with low posterior probability; this could be due to the residence time in state 1 only being for a short amount of time and also that this occurred at the start of the path where it is expected there would be most uncertainty about the initial state of the process.

A further example of an inability to correctly reconstruct the behavioural process is the even shorter residence in state 2 at time 576, shown in Fig. 4.16. This behavioural sojourn lasts a period of 5.94 with only a single observation included within that time. As a result, this sojourn is missed entirely by the algorithm with only 2/6,000 reconstructions considering a behavioural bout over this period. This section highlights a limitation of the algorithm and reconstruction process, but one that is expected to be unavoidable for such a short sojourn. It is important to note, however, that at least these methods do have the capacity to allow for the estimation of unobserved sojourns, whereas discrete-time models would not.

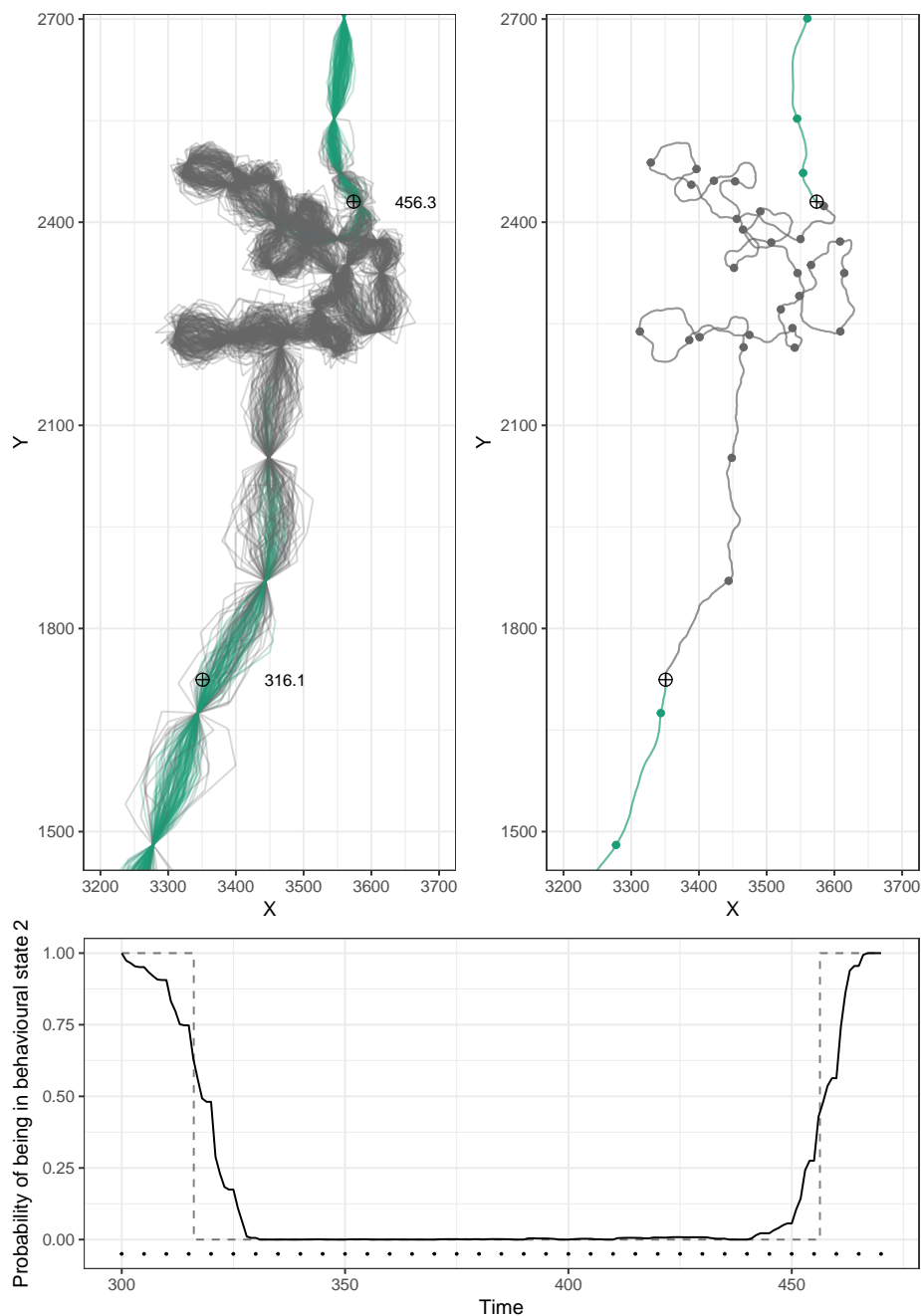


Fig. 4.14 A section of the path reconstruction with high certainty. **Top left:** Posterior samples of the reconstructed, refined movement path between the times 310–470, coloured by behavioural state (state 1 **grey**, state 2 **green**). **Black targets** highlight the true location of a switch into state 1 at time 316 and back into state 2 at time 456. **Top right:** Simulated movement path over the same period of time (**line**) with observations highlighted (**points**). **Bottom:** Posterior probability of residing in behavioural state 2 over the period of time 300–475 (**solid, black**), with true behavioural process included (**dashed, grey**).

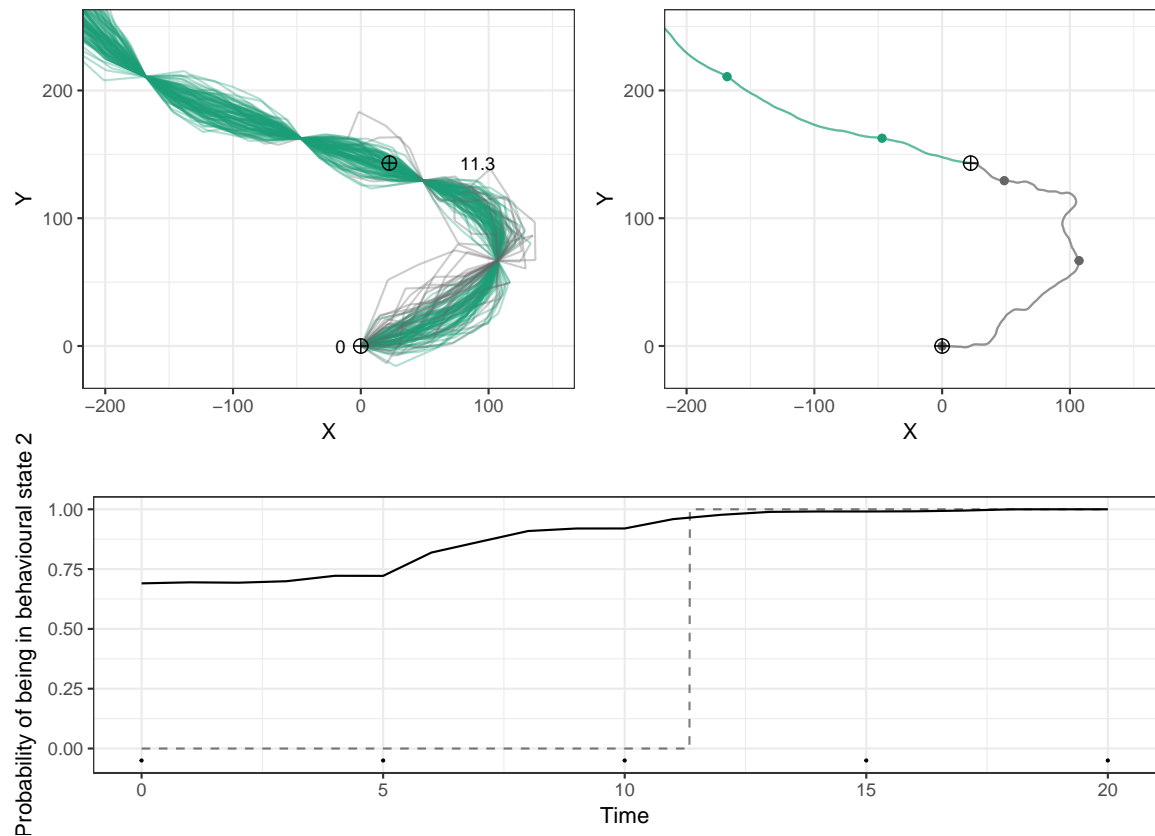


Fig. 4.15 An uncertain section of the path reconstruction. **Top left:** Posterior samples of the reconstructed, refined movement path between the times 0–20, coloured by behavioural state (state 1 **grey**, state 2 **green**). **Black targets** highlight the true location of a switch, starting in state 1 at time 0 and switching into state 2 at time 11. **Top right:** Simulated movement path over the same period of time (**line**) with observations highlighted (**points**). **Bottom:** Posterior probability of residing in behavioural state 2 over the period of time 0–20 (**solid, black**), with true behavioural process included (**dashed, grey**).

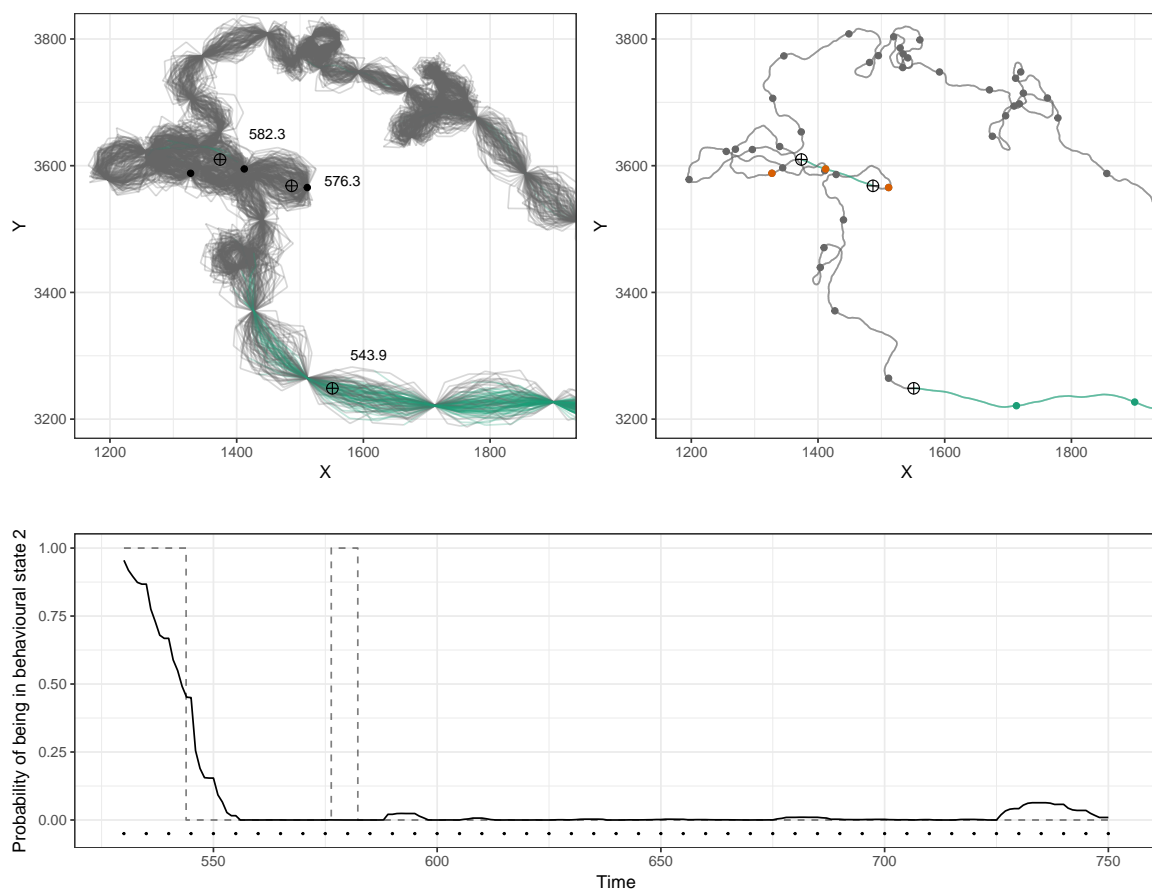


Fig. 4.16 A section of the path reconstruction with misclassification of a behavioural sojourn. **Top left:** Posterior samples of the reconstructed, refined movement path between the times 535–750, coloured by behavioural state (state 1 **grey**, state 2 **green**). **Black targets** highlight the true location of a switch into state 1 at time 544, into state 2 at time 582 and back into state 1 at time 576. **Black points** highlight the three observations; immediately before, during and after the sojourn into state 2. **Top right:** Simulated movement path over the same period of time (**line**) with observations highlighted (**points**). **Bottom:** Posterior probability of residing in behavioural state 2 over the period of time 510–750 (**solid, black**), with true behavioural process included (**dashed, grey**).

4.5 Two-state movement in elk

The following example applies the methods of this chapter to well known GPS data from elk (*Cervus elaphus*), which have previously been modelled in discrete time. This example demonstrates the interpretable nature of the multistate, continuous-time model by finding clear differences in behaviour over time and insights into short-term behaviour that could not have been obtained in discrete time.

4.5.1 Elk-115 observations

A set of 194 ‘daily’ GPS observations from the elk tagged as ‘elk-115’ are used in this example (see https://bitbucket.org/a_parton/elk_example for data and to reproduce this work). These observations were introduced and modelled as part of a larger set consisting of four elk in the discrete-time, step-and-turn model of Morales et al. (2004), and more recently modelled in the vignette of the R package `moveHMM` (Michelot et al., 2016) applying the HMM of Langrock et al. (2012). The observations are shown in Fig. 4.17, indicating two distinct movement modes: slow, volatile movement where observations are over-plotted, and fast, directed movement where observations are sparse.

Morales et al. (2004) fit a number of models to the larger dataset containing the observations from elk-115, with the model most similar to that described in this chapter being what they describe as the ‘double switch’ model. This is the model described in section 1.2.1, where there are fixed switching probabilities between states. Two states are implemented, which in turn govern a mixture of CRWs describing the movement process. In the vignette of `moveHMM` the larger dataset is used to demonstrate a two-state HMM with switching dependent on environment. For comparison with the methods here, the reproduced analysis shown in Fig. 4.21 does not include this environmental information and so is the same underlying movement model as the ‘double switch’ in Morales et al. (2004). In both these discrete-time applications, ‘exploratory’ and ‘encamped’ states were identified as having mean daily turning angles of close to zero and π , respectively. The implications of turn distributions not centred at zero was discussed in Sect. 3.6.

In this example, the model of Sect. 4.1 with two behaviours is applied to the elk-115 observations. The original analysis in Morales et al. (2004) described the majority of observations as being daily, but with some taken at 22 and 26 hour intervals. In order to handle this irregularity, they divided the observed straight line step lengths by the sampling time frame to approximate daily steps. A method transforming the observed turning angles

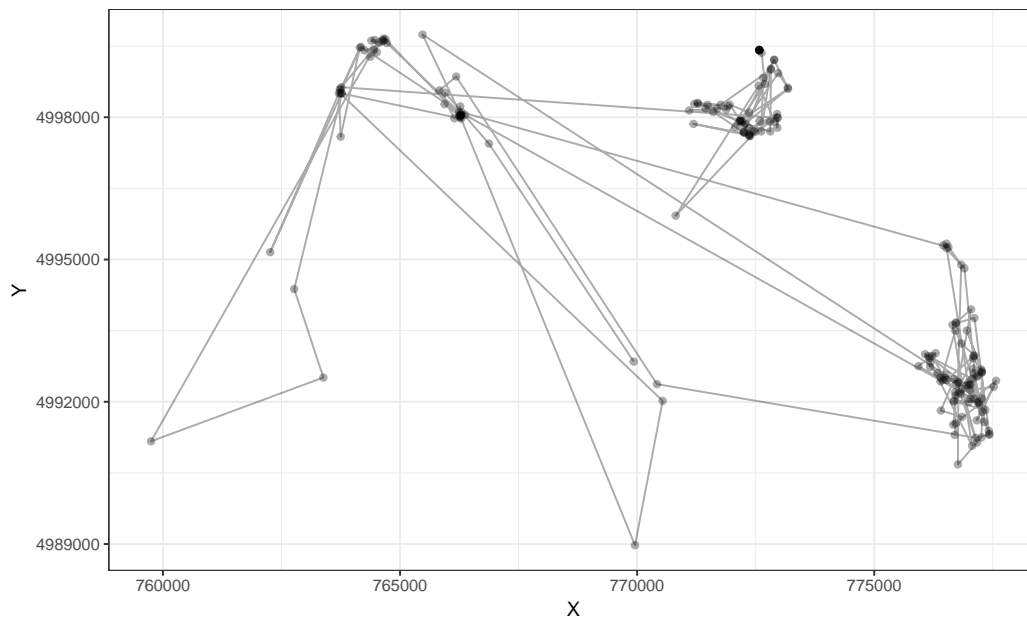


Fig. 4.17 Observed ‘daily’ locations of Elk-115 (**points** linked chronologically with **lines**). Note that observed points are displayed here with transparency to highlight the times where multiple observations were captured in the same/similar location.

to some daily approximation is unclear, and so these remained as the observed values in their analysis. The open-access version of the elk data does not include the times of the observations, and rounding of the Morales et al. (2004) ‘daily step lengths’ meant that the original observation times could not be ascertained. The analysis carried out here follows that in the vignette of `moveHMM`, using the observed locations, but assuming that these were all at 24 hour intervals. The continuous-time formulation of the model presented in this chapter would easily allow for these irregularly timed observations (and missing observations, if applicable) to be handled if exact observation times were known.

Applying the presented methodology to multiple animals in the same way as `moveHMM`, by pooling information across individuals and estimating a set of population parameters, could be implemented by a simple extension to the current R code but is not attempted here for simplicity. Following Morales et al. (2004) and the vignette of `moveHMM`, observation error is assumed to be negligible here (though see Chap. 5). Interest thus involves inference on the eight movement parameters, consisting of a bearing variance and three speed parameters for each of the two states.

Using daily observations leaves large portions of the elk’s movement unobserved, and so it is expected that the reconstructed movement paths, and thus parameters, for this example will

be uncertain in comparison to previous examples. Rather than a full ecological analysis, this example is included as a proof of concept for the presented methods and to highlight some of the possible dangers when analysing daily observations in discrete time. For comparison, an example of single state movement on a dataset with a sampling scheme of 2 minutes is given in Sect. 5.3, with uncertainty in location being much lower.

4.5.2 Implementing the inference algorithm

4.5.2.1 Initial values

An initial movement path was created at a time scale of 2 hours by taking an interpolating cubic spline between observations. The choice of a 2 hour time scale gives around 11 unknown locations for reconstruction between each pair of observations, thought to provide an acceptable trade-off between computational cost and approximation to continuous time (see Sect. 3.6 for further discussion of δt). The corresponding initial behavioural configuration was set by identifying any points on this path with speed above 100 metres per hour. The initial path is shown as both locations and bearings/speeds in Fig. 4.18 and the Appendix, Fig. B.11, respectively. Initial parameters were set as estimates from this initial path configuration, given as

$$\begin{aligned} & \{\sigma_{\theta,1}^2{}^{(0)} = 2, \mu_1^{(0)} = 100, \beta_1^{(0)} = 0.115, \sigma_{\psi,1}^2{}^{(0)} = 160\}, \\ & \{\sigma_{\theta,2}^2{}^{(0)} = 0.1, \mu_2^{(0)} = 500, \beta_2^{(0)} = 0.08, \sigma_{\psi,2}^2{}^{(0)} = 1800\}. \end{aligned}$$

4.5.2.2 Prior information

A prior distribution specifying an upper bound on the ratio of the speed parameters to avoid the presence of negative speeds in both states was applied. To define state 2 as ‘exploratory’, a Gaussian prior with mean 0.05 and standard deviation of 0.1 was placed on the bearing variance. All remaining movement parameters had flat priors.

The same prior was placed on both switching rates, being a gamma distribution with rate 4 and shape 0.1. This was chosen to limit the rate of behavioural switching, strongly discouraging switching occurring at time frames much shorter than the sampling scheme. Such a prior leads to the 90% prior credible interval for the residency time being $(6, 7 \times 10^{13})$ hours. It

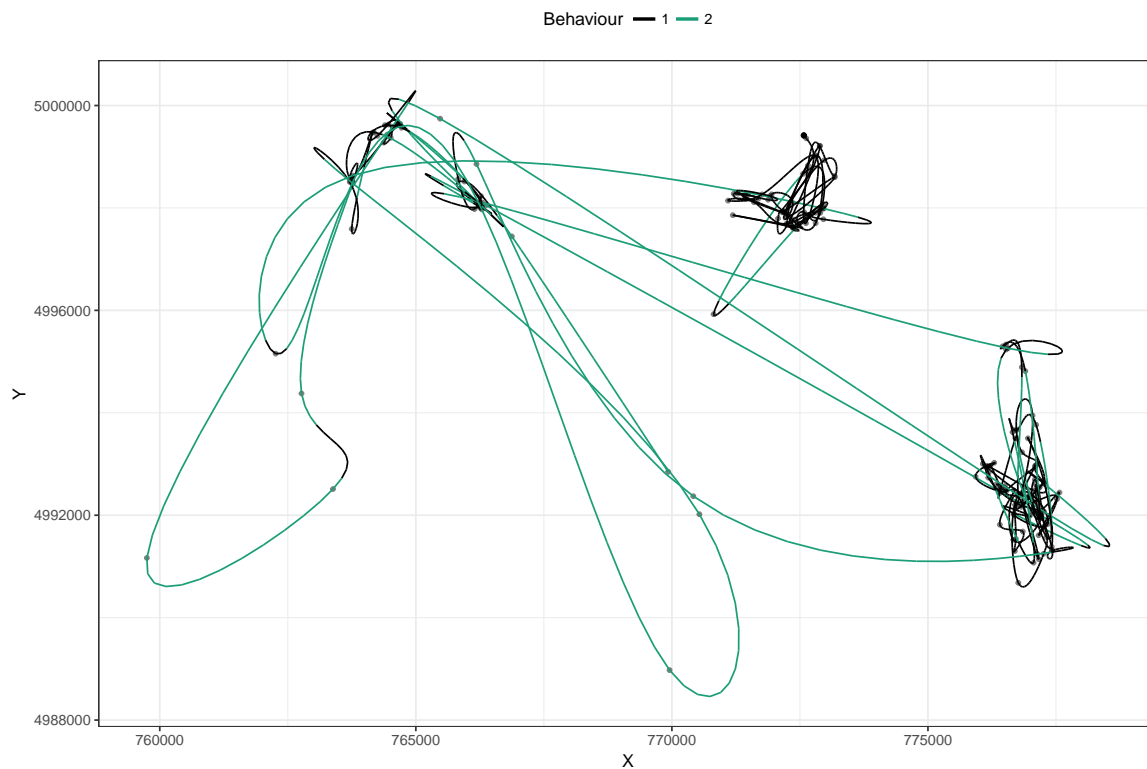


Fig. 4.18 Initial path for the elk-115 example, coloured by the initial behavioural configuration.

is believed this prior is vague when compared with the posterior credible intervals for the switching rates (see below).

4.5.2.3 Implementation

The algorithm of Sect. 4.3 was applied for 4.8×10^6 iterations, with each iteration consisting of a single parameter update and 100 refined path updates on random sections of path with lengths ranging 4–24 points (i.e. 8–48 hours, see Sect. 3.5.1.2 for a discussion on the choice of sub-path length). The choice here was based on acceptance rates in pilot runs: lengths higher than 24 had too low acceptance to be feasible, and lengths of 4 allowed these short section updates that helped with mixing. Perturbation variances for the speed parameters were influenced by a pilot run. Samples were thinned (for memory storage purposes) by a factor of 1,000 and the first quarter were treated as a ‘burn-in’ period, leaving 3,600 stored samples of parameters and reconstructed paths. Computational run-time for this implementation was around 5 days.

4.5.3 Behavioural parameter results

A scatter plot of the posterior (log) samples for the two switching rates that define the behavioural process is shown in Fig. 4.19. The trace of the sampled parameters and kernel density estimates are provided in the Appendix, Figs. B.12–B.13. The behavioural parameters pass standard convergence diagnostics (Heidelberger and Welch), with ESS of over 125, taken from the 3,600 samples thinned from 3.6×10^6 iterations. The sample trace and scatter plots show high positive correlation between the switching rates. A clear difference in residency time was inferred for the two behavioural states, differing in magnitude by a factor of 10, with state 1 (the ‘encamped’ state) having the longer residence time.

Posterior summary statistics for the switching rates are given in Table 4.3. The sampled mode of the mean switching rate leads to a residency time of 6.4 days for state 1 and 19.2 hours for state 2. The given 90% credible intervals correspond to a mean residence time in state 1 being between 4–11 days and in state 2 between 10–36 hours.

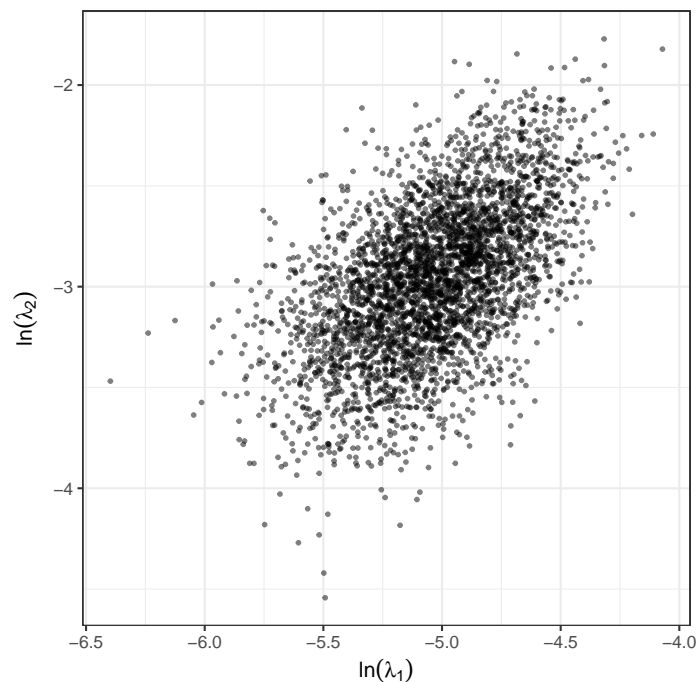


Fig. 4.19 Posterior samples of the behavioural parameters (on a log scale) for elk-115: λ_1 is the switching rate out of the ‘encamped’ state and λ_2 is the switching rate out of the ‘exploratory’ state.

Table 4.3 Posterior credible intervals (5%, 50%, 95% quantiles) of the sampled behavioural parameters in the elk-115 example.

	5%	50%	95%
λ_1	0.00391	0.00651	0.0105
λ_2	0.0275	0.0520	0.0959

4.5.4 Movement parameter results

The posterior samples of the movement parameters are given in Fig. 4.20, coloured by behavioural state. The trace of the sampled parameters and kernel density estimates are provided in the Appendix, Figs. B.12–B.13. These show clear differences in inferred movement types between the two behavioural states.

The movement parameters for state 1 have a low ESS and do not pass standard convergence diagnostics, due to the parameters σ_θ^2 and β . The bearing variance for the ‘encamped’ state is so high as to produce uniform turns at the observation time scale. Hence, samples of this parameter move slowly around high values that lie above this threshold; Fig. 4.20 shows that σ_θ^2 is not only high, but highly spread for state 1 (**top panel, black**). As discussed previously, the correlation parameter appears as an exponential term in the speed likelihood (see Eq. 4.3), and values greater than 1 lead to an exponential of less than 10^{-10} at the observation time scale. This parameter is ‘drifting’ around the sample space and cannot be estimated, see Fig. 4.20 (**bottom panel, black**), it can only be inferred that this parameter has a high value. The movement parameters for state 2 pass standard convergence diagnostics (Heidelberger and Welch) with ESS of over 75 (taken from the 3,600 samples thinned from 3.6×10^6 iterations.).

Posterior summary statistics of the parameters are given in Table 4.4. Behavioural state 1 has high σ_θ^2 and low μ —defining volatile, slow movement categorised here as ‘encamped’. The uniform turns at the observation time scale (due to $\sigma_{\theta,1}^2$ having median of 5.61 squared radians per hour) means the bearing of the elk is effectively unknown at these times. The median for long term exploratory speed is given by 77.3 metres per hour. A high β and low σ_ψ^2 , describes speeds that are effectively uncorrelated in the short term (the mean expression of the speed process in Eq. 4.3 is dominated by the first term involving the ‘mean speed’ parameter rather than the second term involving the ‘current speed’) and have low variation in the long term.

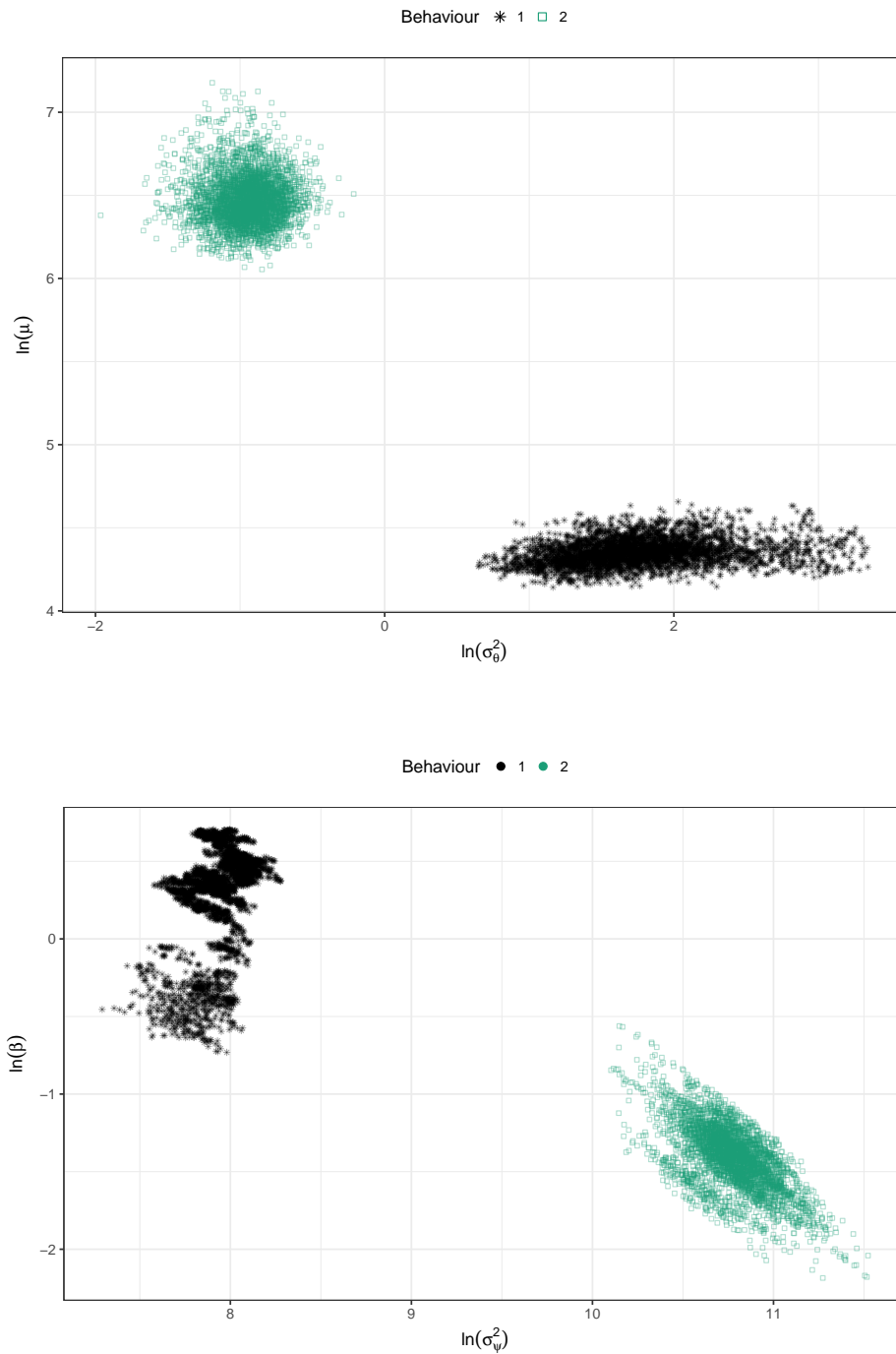


Fig. 4.20 Posterior samples of the state-dependent movement parameters (on log scale) for elk-115. **Top:** joint sample space between bearing parameter and the mean speed. **Bottom:** joint sample space between the speed correlation and the long-term speed variance.

Table 4.4 Posterior credible intervals (5%, 50%, 95% quantiles) for the sampled movement parameters, split by state, in the elk-115 example.

		5%	50%	95%
Behaviour 1 (‘encamped’)	σ_{θ}^2	2.87	5.61	16.4
	μ	68.8	77.3	90.2
	β	0.627	1.45	1.94
	σ_{ψ}^2	2,160	2,820	3,390
Behaviour 2 (‘exploratory’)	σ_{θ}^2	0.274	0.389	0.521
	μ	519	638	855
	β	0.170	0.245	0.340
	σ_{ψ}^2	34,300	47,600	66,400

Behavioural state 2, the ‘exploratory’ state, has low σ_{θ}^2 and high μ , reflecting fast, straight movement. The median long term exploratory speed for is 638 metres per hour, with speeds that are highly correlated in the short term (through a low β) but with high variation in the long term (through a high σ_{ψ}^2).

4.5.5 Path reconstruction results

Based on the posterior samples of the path reconstruction, Fig. 4.21 (**top panel**) shows the probability of being in behavioural state 2 throughout the course of the sampling period. Additionally, the corresponding state probabilities estimated by fitting an HMM, as in the vignette of moveHMM (but using the larger dataset of tracks from four elk), are shown (**bottom panel**). The **points** along the horizontal axis indicate the times at which observations were taken. The two models can be seen to identify the same areas of the movement path as being in the ‘exploratory’ state, however the residency times in this state differ, with the HMM classifying three long stays in state 2 in the middle of the observation period whilst the continuous-time model includes more frequent switching over this period.

The estimated residency of the exploratory state (90% credible interval leading to a mean residence time between 10–36 hours) and the probabilities of residency shown in Fig. 4.21 suggest that there are parts of the movement path where short sojourns of fast movement occur. This can be further seen in Fig. 4.22, showing the sampled residency times calculated from the reconstructed paths. In fact, from the sampled behavioural reconstructions, 73% of the residencies in behavioural state 2 were less than the 24 hour sampling scheme (respective proportion for state 1 was 17%).

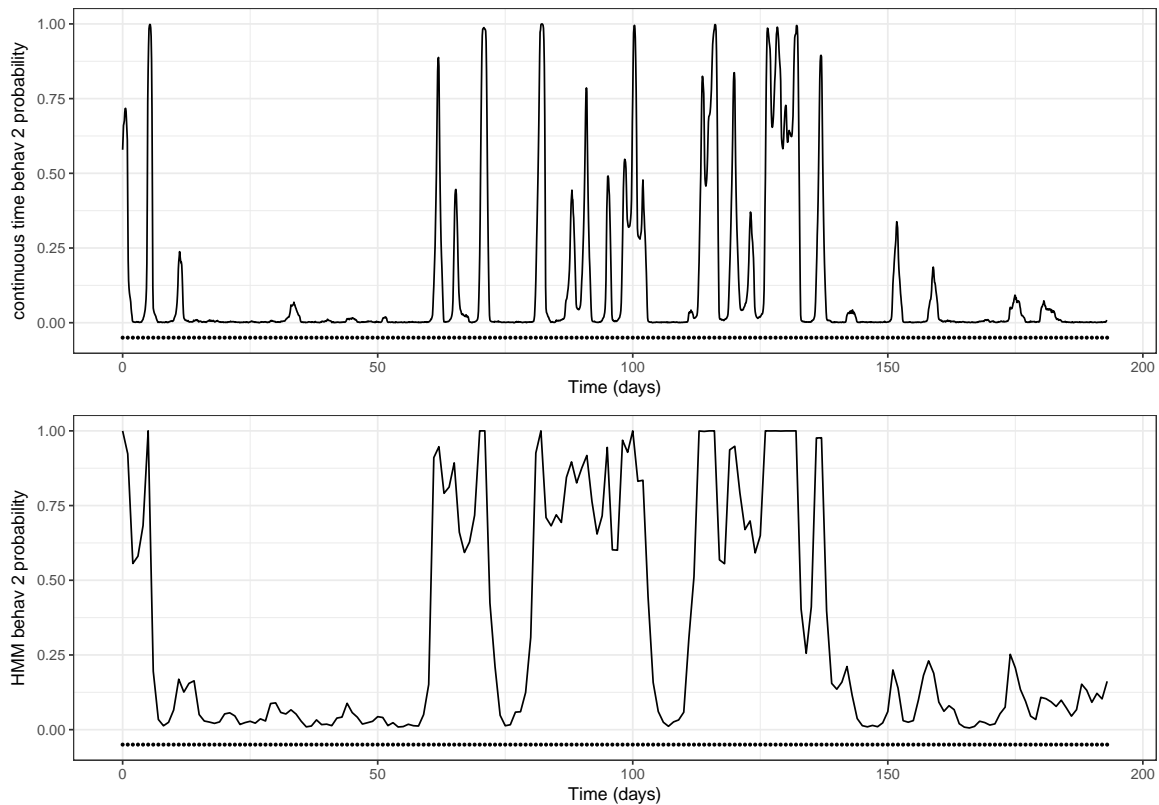


Fig. 4.21 **Top**: probability of residing in behaviour 2 ('exploratory') over time, estimating from the sampled path reconstructions. **Bottom**: probability of residing in behaviour 2 using the R package moveHMM (Michelot et al., 2016). **Points** are included along the horizontal axes to highlight the times/frequency of observations.

Fig. 4.23 shows four examples (one per plot) of the reconstructed refined movement path, coloured by corresponding behavioural state. The difference in movement types between the two identified states is highlighted here, showing the same interpretation of movement as Morales et al. (2004) and the vignette of moveHMM: a slow 'encamped' state (**black**) and fast 'exploratory' state (**green**). These path reconstructions aid in the interpretation of the movement parameters and give insight into the space use of the animal between observations.

Fig. 4.24 shows a detailed portion of the movement path between 3 consecutive observations with 40 reconstructed path examples displayed. Plotting a number of reconstructions in this way highlights the extent of uncertainty between observations, and the areas that have a high density of residency. The area covered by the reconstructions in the **left panel** of Fig. 4.24 is displayed within the full area of observations by the shaded rectangle in the **top right panel** for reference. Additionally, the probability of being in state 1 over the time period displayed is given in the **bottom right panel**. Starting at observation 114, the probability of moving

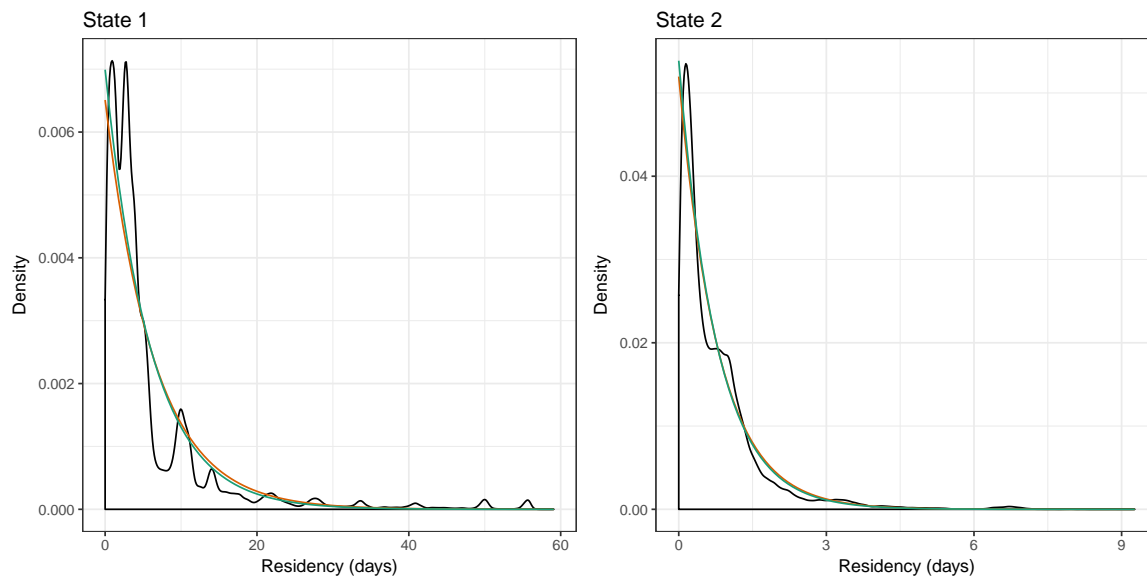


Fig. 4.22 Kernel density estimates of the sampled residence time in each behavioural state, based on reconstructed path samples (**black**). Additional lines give the exponential distribution with rate given by the maximum likelihood of the sampled residencies (**green**) and with rate given by the median of the sampled behavioural parameters (**orange**).

into the ‘encamped’ state increases (as also seen by the density of **black** sections in the path reconstructions) until it is equally uncertain which state the animal is in. As the animal moves towards the location of observation 116, the probability of being in the ‘encamped’ state decreases rapidly.

In Figs. 4.23 and 4.24, it can be seen in a number of places that the reconstruction involves a switch into and back out of state 1 between two consecutive observations. The exact time when these short (between observation) switches occur vary over the sampled reconstructions, but their presence has high probability. For example, in Fig. 4.24, there can be seen to be a number of short sections of **black**, ‘encamped’ sojourns between observations 115–6, but their location and the time that they occur between the two observations varies. There is information in the observations indicating a sojourn has occurred, but when is uncertain. Being able to extract such qualitative information on short term behaviours from observations, albeit with uncertainty, gives extra insight into the movement that is not possible when switches can only occur at the observation time-scale.

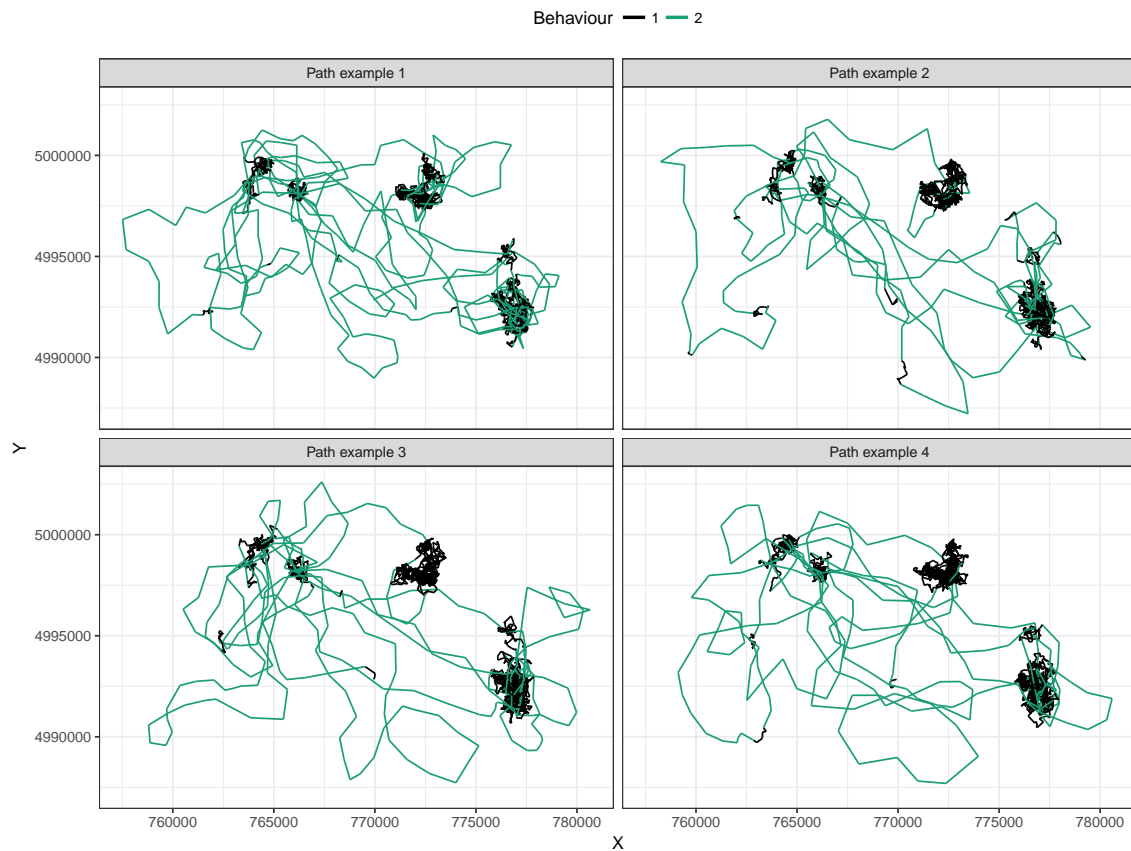


Fig. 4.23 Four examples of reconstructed refined movement paths for elk-115. For each example, the reconstructed refined path is displayed as a linearly interpolated path. The path is coloured in each case by the corresponding behaviour, highlighting the difference in movement characteristics resulting from the parameters associated with each state.

4.6 Discussion

This chapter extended the continuous-time step-and-turn movement model of Chap. 3 to include multistate, behavioural switching. The movements of animals are complex and in order to model this for an extended period of time, the different kinds of movement modes must be taken into account (Morales and Ellner, 2002). Not considering the behavioural process of the animal biases estimates of movement metrics (Nakagawa and Freckleton, 2008) and resource selection analyses (Roever et al., 2014). The model for behavioural switching implemented here follows the lead of popular movement models (Blackwell, 1997; Blackwell et al., 2015; Langrock et al., 2012; McClintock et al., 2012; Morales et al., 2004), using a Markov chain to describe the switching process, which is autoregressive with order one. Alternative extensions to this work that may be applicable to a variety of scenarios are

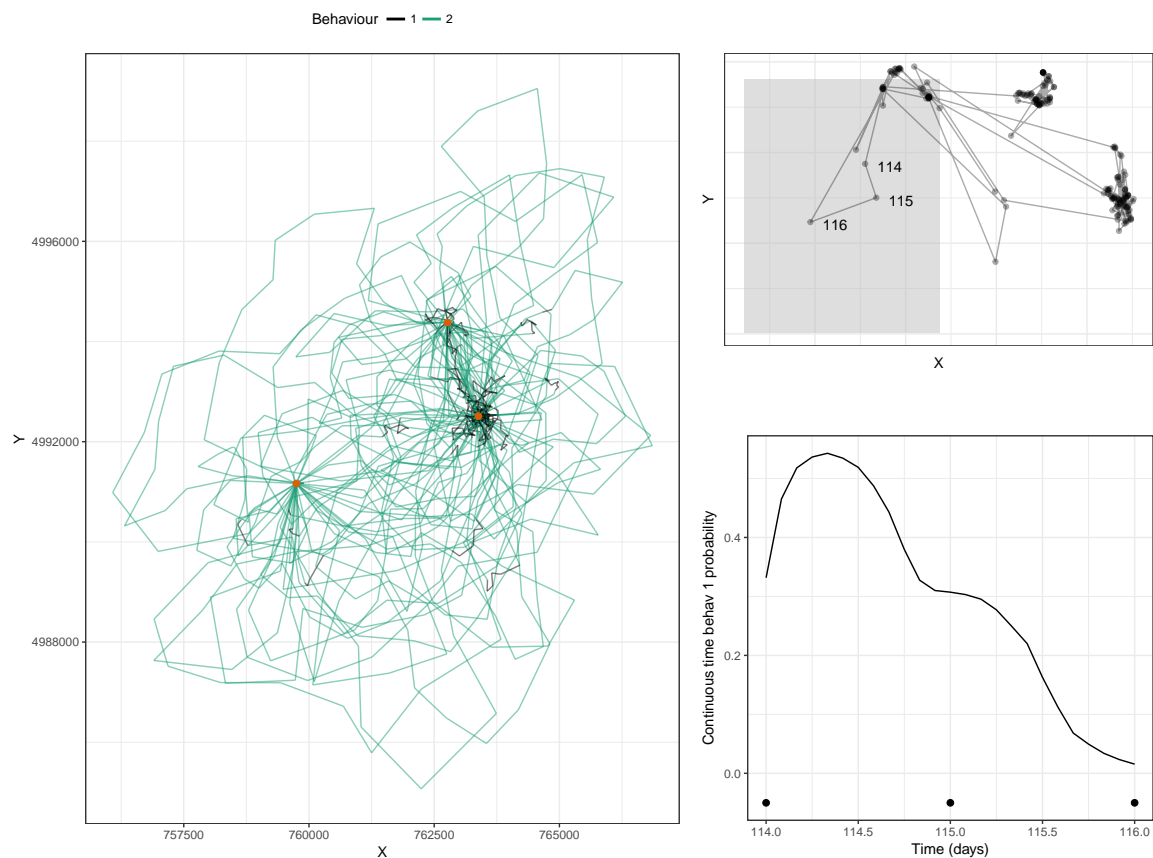


Fig. 4.24 The uncertainty in location and behavioural configuration for a short section of the elk-115 observations. **Left:** 40 example refined path reconstructions (**lines**) between three consecutive locations (**orange points**), coloured by inferred behavioural state. **Top right:** full set of observations, with the three observations 114–6 highlighted and the area covered by the reconstructions in the left panel shaded in **grey**. **Bottom right:** probability of residing in state 1, calculated from posterior path reconstructions.

the semi-Markov model—as suggested in Langrock et al. (2012)—or a stochastic process with a slower decaying memory—as discussed in Fleming et al. (2014a).

In contrast to discrete-time models such as Langrock et al. (2012); McClintock et al. (2012); Morales et al. (2004), the Markov chain applied here is continuous in time. As in Blackwell et al. (2015), this allows switching between observation times. This lessens the ecological importance placed upon the discrete-time scale used for analysis (often the sampling time scale), and allows for the consideration of whole behavioural sojourns between successive observations. Such short behavioural bouts are identified in the elk example of Sect. 4.5, in which there was a high posterior probability of remaining in the fast, travelling state for less than the 24 hour sampling period. There is, however, a limit to the behavioural

information that can be extracted between successive observations; this is highlighted in the simulation example of Sect. 4.4 in which a short behavioural sojourn containing only a single observation is not identified by the inference algorithm.

Although the method for inference presented here uses a refined approximation to the continuous-time movement process, the behavioural process remains exact throughout—as in Blackwell et al. (2015). Advantages remain with this approach even when the refined path is at the same time scale as a comparative discrete-time analysis. Benefits arise from the augmentation approach used, based on Blackwell et al. (2015), because posterior reconstructions of the behavioural process can be used to investigate the uncertainty around switching times. An example analysing this uncertainty is presented in the simulation example of Sect. 4.4.

A rejection method is used in the inference approach of Sect. 4.3 to simulate a behavioural reconstruction proposal between two fixed behaviours. In the naive approach used, a proposal is simulated from an unconstrained Markov chain, and instantly rejected if this does not agree with the fixed endpoint. This simple method was found to work reasonably well in the two examples within this chapter (20% and 18% of proposals were rejected as a result of this method in the simulation and elk examples, respectively), however, in applications with higher numbers of behavioural states this rejection rate is expected to become infeasible. A number of other simulation methods exist that could be used in place of this simple rejection method (for example, see Hobolth and Stone (2009); Rao and Teh (2013); Whitaker et al. (2017)). In particular, implementation of the methods of Hobolth and Stone (2009) have recently been released through the R package `ECctmc` (Fintzi, 2017).

The examples of this chapter have only explored two-state implementations. Although examples with higher numbers of states can be implemented with the methods described in this chapter, the feasibility of this on a computational front is questionable; a larger sample space for the unknown behavioural process will lead to slower mixing and longer computational run times. Avenues to make the inference approach more efficient to allow for implementations with three or more behavioural states are discussed further in Chapter 6.

Throughout this chapter it has been assumed that the number of behavioural states present in a trajectory are known, and that the transition rates between behaviours are constant. Fixed numbers of behaviours are common in other movement models (Blackwell et al., 2015; Morales et al., 2004) and are not thought to be a restrictive assumption because the states should have ecological relevance to be interpreted, and so should be loosely identifiable (such as the ‘encamped’ and ‘exploratory’ states in the elk example of Sect. 4.5). Although the methods here could potentially be extended to allow for an unknown number of behavioural states using a reversible jump MCMC algorithm, we feel this is an unnecessary complication,

with a danger of identifying ‘behavioural states’ that describe constructs such as observation error rather than differing movement behaviours—as identified previously in similar discrete-time methods (Li and Bolker, 2017; Pohle et al., 2017). Although it may not be necessary to allow varying numbers of states, it would be desirable to allow the switching rates between them to depend on spatial covariates (Morales et al., 2004) or on location itself. Depending on the duration of study, it may also be useful to allow varying rates with time, perhaps periodically to reflect daily or annual cycles. Both these extensions could be addressed, without any additional approximation, using the framework in Blackwell et al. (2015), applied there to movement models directly based on location (rather than velocity or steps and turns) with heterogeneity in both space and time. More generally, some more of the complexity of behaviour could be captured by including an additional ‘resting’ state, likely to occur at particular times of the day, with low or zero speed and perhaps a high variance to represent the ‘forgetting’ of bearing while resting.

Chapter 5

Incorporation of observation error

The movement models of Chaps. 3 and 4 do not take into account the error present in GPS observations of animal location, despite such error being known to occur. The following describes a simple model for observation error (Sect. 5.1) and how the approach for Bayesian inference described in the previous chapters can be extended to accommodate the presence of noisy observations (Sect. 5.2). Note that the following (in part) provides an extended description to that presented in Parton et al. (2017).

The methods of Sect. 5.2 are demonstrated through an example on a small subset of noisy observations from a single reindeer in Sect. 5.3, showing how inferred movement paths are more persistent when error is introduced. The effect the refined path approximation has on parameter estimation is explored in an example on gull movement in Sect. 5.4, identifying possible limitations in the identifiability of both the speed parameters and the observation error. This limitation is investigated further using a simulated data example in Sect. 5.5 by comparing inference at a number of levels of assumed error.

Methods to extend the independent model for observation error are given in Sect. 5.6 to incorporate the somewhat more realistic assumption of correlation in the error process. However, note that this has not been applied in practice.

5.1 Independent model for observation error

Observations of location, \mathbf{Z}^* , are assumed to have been taken with error from the true, underlying movement path that is described in Sects. 3.1 and 3.4. Hence, the true locations,

\mathbf{Z} , are those in Eq. 4.4 and the observed locations are

$$\mathbf{Z}^*(t_j) = \begin{pmatrix} X^*(t_j) \\ Y^*(t_j) \end{pmatrix} = \begin{pmatrix} X(t_j) \\ Y(t_j) \end{pmatrix} + \begin{pmatrix} \varepsilon_{j,x} \\ \varepsilon_{j,y} \end{pmatrix}. \quad (5.1)$$

The observation error, ε , is assumed to be a circular, bivariate Gaussian, that is independent and identically distributed in space and time, hence $\varepsilon_{j,\cdot} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. The relationship between the unobserved path and the observations is given in Fig. 5.1. Note that the underlying movement process remains the same, and we are essentially generalising the special case given in previous chapters where $\mathbf{Z}^* = \mathbf{Z}$, i.e. $\varepsilon = 0$. Interest lies in making inference about the parameters Φ and the observation error variance σ_ε^2 . Note that Fig. 5.1 displays a single behavioural state for simplicity, but could equally be applied to the multistate model of Chap. 4.

5.2 Extending the method for fully Bayesian inference

Given a set of noisy observations, Bayesian inference on the parameters Φ and observation error variance σ_ε^2 follows the same approach as Sect. 4.3, but with the addition of error. The observations are augmented with an approximate reconstruction of the underlying movement path, and a hybrid Gibbs sampler is used to conditionally sample in turn from:

- the behavioural parameters (if applicable),
- the movement parameters,
- the observation error variance, and

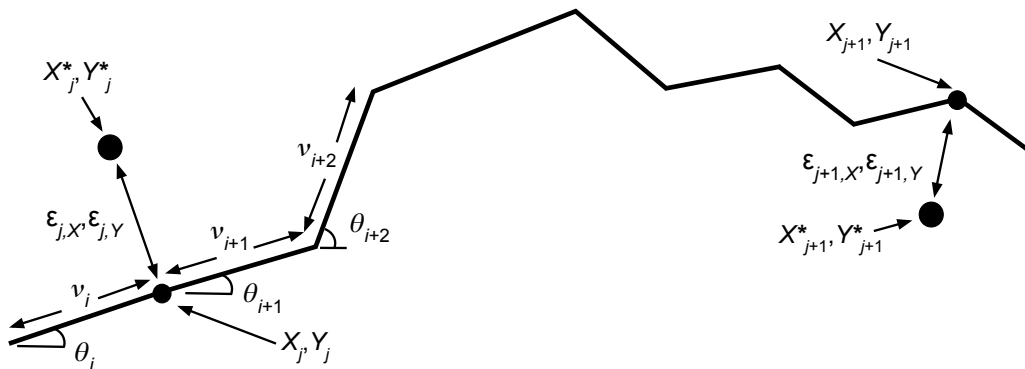


Fig. 5.1 Relationship between the observed locations (**large points**), the ‘true’ underlying locations at observation times (**small points**), and the unobserved refined path (**line**).

- the unobserved refined path.

A DAG of the movement model with observation error is given in Fig. 5.2. Sampling the parameters Φ is identical to that of Sects. 4.3.2 and 4.3.3—both are unaffected by the presence of error because they are independent of this when conditioned upon full observation of the behaviour and movement processes. The following describes the (standard) method for sampling the observation error parameter. The extension to the previously described method for reconstructing the refined movement path is then described, generalising to allow for independent Gaussian errors. The following methods supersede that of Chaps. 3 and 4, with such being the special case where σ_ε^2 has a known value of zero.

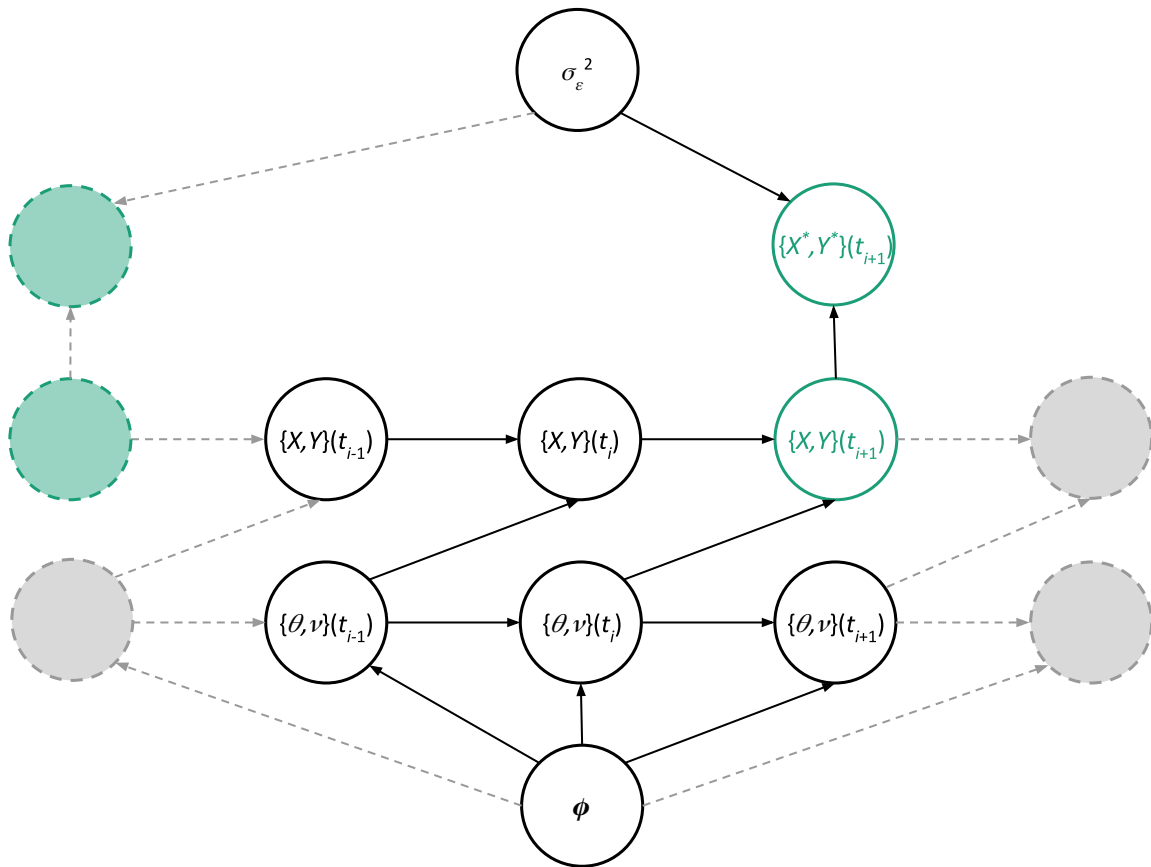


Fig. 5.2 DAG of the movement model with augmentation at an approximate time scale when observation error is present. In this representation, locations $\{X, Y\}$ are shown corresponding to the approximate time scale created for the augmentation of the behaviours, bearings and steps $\{B, \theta, \nu\}$. Observed locations $\{X^*, Y^*\}$ do not occur at each point on the augmented time scale, but will be less frequently (for example, at the times shown in **green**).

5.2.1 Sampling the observation error parameter

The observation error parameter describes the variance of a Gaussian distribution. The conjugate distribution for this is the inverse gamma; assuming such a prior allows direct sampling from the posterior conditional distribution as a Gibbs step (as in the case of sampling the bearing parameter, shown in Appendix A.1.1). Given the observed locations \mathbf{Z}^* and the reconstruction of the refined path $\{\boldsymbol{\theta}, \mathbf{v}\}$, the observation error $\boldsymbol{\varepsilon}$ can be calculated using Eqs. 4.4 and 5.1. The posterior full conditional distribution simplifies to

$$p\left(\sigma_{\boldsymbol{\varepsilon}}^2 \mid \boldsymbol{\Phi}, \mathbf{B}, \boldsymbol{\theta}, \mathbf{v}, \mathbf{Z}^*\right) = p\left(\sigma_{\boldsymbol{\varepsilon}}^2 \mid \boldsymbol{\theta}, \mathbf{v}, \mathbf{Z}^*\right),$$

where

$$\sigma_{\boldsymbol{\varepsilon}}^2 \mid \boldsymbol{\theta}, \mathbf{v}, \mathbf{Z}^* \sim \text{IG}\left(a_{\boldsymbol{\varepsilon}} + M, b_{\boldsymbol{\varepsilon}} + \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^2 \varepsilon_{ij}\right),$$

when the prior for the error variance is given by the conjugate $\sigma_{\boldsymbol{\varepsilon}}^2 \sim \text{IG}(a_{\boldsymbol{\varepsilon}}, b_{\boldsymbol{\varepsilon}})$ and M is the number of observations.

5.2.2 Reconstructing the unobserved refined path

The general approach for sampling the unobserved path is the same as that described in Sects. 3.3.3 and 4.3.4. As in the latter, reconstruction will be outlined between three consecutive observations as this is the minimum necessary to update the level of observation error. This scenario is displayed in Fig. 5.3, again omitting the presence of multiple behavioural states for simplicity but this does not affect the approach. Here, both observations (**large points**) and ‘true’ locations (**small points**) are shown.

The quantities to be simulated are those displayed in **black** in Fig. 5.3, consisting of:

- (if applicable) the behavioural process B between the times j and l ,
- the bearings $\{\theta_1, \dots, \theta_{n-1}\}$,
- the steps $\{\mathbf{v}_1, \dots, \mathbf{v}_{n-1}\}$, and
- the true location $\mathbf{Z}(k)$ (implicitly through the simulation of the bearings and steps).

The set of $\{\theta_1, \dots, \theta_{\tau-1}\}$ and $\{\mathbf{v}_1, \dots, \mathbf{v}_{\tau-1}\}$ takes the path to the true (rather than observed) location $\mathbf{Z}(k)$ at the observation time $t_{\tau} = k$. Similarly, the set of $\{\theta_{\tau}, \dots, \theta_{n-1}\}$ and $\{\mathbf{v}_{\tau}, \dots, \mathbf{v}_{n-1}\}$ takes the path to the fixed, true location $\mathbf{Z}(l)$ at the time $t_n = l$.

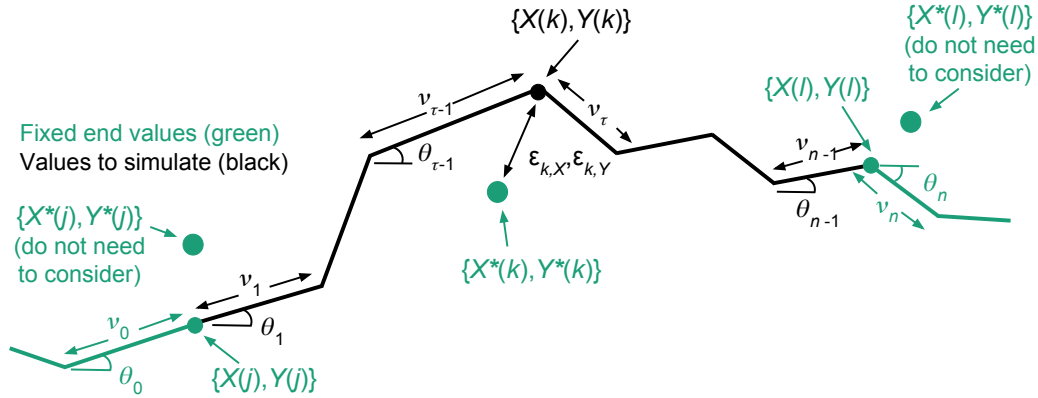


Fig. 5.3 Section of the full refined path to update over when observation error is present. Fixed endpoint locations are given at the times j and l . There is observation error on these locations, but only the ‘true’ locations are required for the reconstruction. The bearing and step processes, $\{\theta_1, \dots, \theta_{n-1}, v_1, \dots, v_{n-1}\}$, are simulated given fixed endpoints $\{\theta_0, \theta_n, v_0, v_n\}$ and the fixed observed location at the time k . Note that the reconstruction will allow error on the observation at time k so will not pass directly through it as in Chap. 4.

The fixed values that are to be conditioned upon are displayed in **green** in Fig. 5.3, consisting of:

- the true locations at the ends of the path section $\{\mathbf{Z}(j), \mathbf{Z}(l)\}$ (hence the observed locations $\{\mathbf{Z}^*(j), \mathbf{Z}^*(l)\}$ do not affect the reconstruction and are not considered from this point onwards),
- the observed location within the path section $\mathbf{Z}^*(k)$,
- (if applicable) the behaviours $\{B(j), B(l)\}$ at the times j, l ,
- the bearings $\{\theta_0, \theta_n\}$ at the times $\{t_0, t_n = l\}$, and
- the steps $\{v_0, v_n\}$ at the times $\{t_0, t_n\}$.

The reconstruction of the refined path is sampled by an MH step, with the simulation of the refined path proposal following Sects. 3.3.3 and 4.3.4, but having been extended to allow for the error associated with the observation $\mathbf{Z}^*(k)$.

5.2.2.1 Simulating a refined path proposal

The behavioural (if applicable) and approximate bearing process proposals, $\{\mathbf{B}^*, \boldsymbol{\theta}^*\}$, are carried out identically to that in Sects. 4.3.4.1 and 4.3.4.3 as the proposal is independent of the observed locations.

Consider the joint distribution of the steps and true locations given in Eq. 4.5, but where $\mathcal{F} = \{\Phi, \sigma_\varepsilon^2, \mathbf{Z}(j), B(j), B(l), \theta_0, \theta_n, v_0, v_n\}$, with values for $\mathbf{m}_1, \mathbf{m}_2, \Sigma_1, \Sigma_{1,2}, \Sigma_2$ remaining the same as in Sect. 4.3.4.4. In this scenario the locations $(\mathbf{Z}(k), \mathbf{Z}(l))^T$ have not been fully observed. What has been observed is $(\mathbf{Z}^*(k), \mathbf{Z}(l))^T$, where

$$(\mathbf{Z}^*(k), \mathbf{Z}(l))^T \mid (\mathbf{Z}(k), \mathbf{Z}(l))^T, \mathbf{B}^*, \boldsymbol{\theta}^*, \mathcal{F} \sim \mathbf{N}\left((\mathbf{Z}(k), \mathbf{Z}(l))^T, \Sigma_\varepsilon\right), \quad (5.2)$$

with $\Sigma_\varepsilon = \text{diag}(\sigma_\varepsilon^2, \sigma_\varepsilon^2, 0, 0)$. The approximate step proposal is simulated from the conditional distribution

$$\begin{aligned} \mathbf{v} \mid \mathbf{B}^*, \boldsymbol{\theta}^*, \mathcal{F}, \mathbf{Z}^*(k), \mathbf{Z}(l) \\ \sim \mathbf{N}\left(\mathbf{m}_1 + \Sigma_{1,2}(\Sigma_2 + \Sigma_\varepsilon)^{-1}\left((\mathbf{Z}^*(k), \mathbf{Z}(l))^T - \mathbf{m}_2\right), \Sigma_1 - \Sigma_{1,2}(\Sigma_2 + \Sigma_\varepsilon)^{-1}\Sigma_{1,2}^T\right) \\ = \mathbf{N}\left(\hat{\boldsymbol{\mu}}, \hat{\Sigma}\right), \end{aligned} \quad (5.3)$$

based on Eqs. 5.2 and 4.9.

An extended version of the ‘conditioning by Kriging’ method of Rue and Held (2005) is used to simulate from Eq. 5.3, involving first sampling a realisation of unconditional steps and a realisation of the true location

$$\begin{aligned} \mathbf{x}^* &\sim \mathbf{N}(\mathbf{m}_1, \Sigma_1), \\ \mathbf{y}^* &\sim \mathbf{N}\left((\mathbf{Z}^*(k), \mathbf{Z}(l))^T, \Sigma_\varepsilon\right), \end{aligned}$$

and then accounting for the linear constraint by setting

$$\mathbf{v}^* = \mathbf{x}^* - \Sigma_{1,2}(\Sigma_2 + \Sigma_\varepsilon)^{-1}\left((\mathbf{Z}(j), \mathbf{Z}(j))^T + \mathbf{A}\mathbf{x}^* - \mathbf{y}^*\right). \quad (5.4)$$

It can be shown (see Appendix A.2.2) that the distribution of \mathbf{v}^* is given by that in Eq. 5.3.

5.2.2.2 Simulating a proposal at the start/end of path

When simulating a proposed path section that lies at the start of the complete movement path, the location at the time j must also include observation error (as there is no need to ‘match’ it up with some remainder of the movement path). Similarly, when at the end of the complete movement path, the location at the time l must include observation error. In both these cases,

in addition to the changes to the algorithm described for these special cases in Sect. 4.3.4.5, the observation error covariance is altered to $\Sigma_\varepsilon = \sigma_\varepsilon^2 I_4$ (rather than $\text{diag}(\sigma_\varepsilon^2, \sigma_\varepsilon^2, 0, 0)$).

5.2.2.3 Accepting a refined path proposal

The conditional distribution, up to a constant, is

$$\begin{aligned} & p(\mathbf{B}^*, \boldsymbol{\theta}^*, \mathbf{v}^* \mid \mathbf{Z}^*(k), \mathbf{Z}(l)) \\ &= p(\mathbf{B}^*, \boldsymbol{\theta}^* \mid \mathbf{Z}^*(k), \mathbf{Z}(l)) p(\mathbf{v}^* \mid \mathbf{B}^*, \boldsymbol{\theta}^*, \mathbf{Z}^*(k), \mathbf{Z}(l)) \\ &\propto p(\mathbf{B}^*, \boldsymbol{\theta}^*) p(\mathbf{Z}^*(k), \mathbf{Z}(l) \mid \mathbf{B}^*, \boldsymbol{\theta}^*) p(\mathbf{v}^* \mid \mathbf{B}^*, \boldsymbol{\theta}^*, \mathbf{Z}^*(k), \mathbf{Z}(l)), \end{aligned}$$

where every term in the above is also conditioned upon \mathcal{F} . The simulation method employed to create the proposal has density proportional to

$$p(\mathbf{B}^*, \boldsymbol{\theta}^* \mid \mathcal{F}) p(\mathbf{v}^* \mid \mathbf{B}^*, \boldsymbol{\theta}^*, \mathcal{F}, \mathbf{Z}^*(k), \mathbf{Z}(l)).$$

The MH acceptance ratio involves the marginal probability of the locations $\mathbf{Z}^*(k)$ and $\mathbf{Z}(l)$, given a set of behaviours and bearings. This is given by

$$(\mathbf{Z}^*(k), \mathbf{Z}(l))^T \mid \mathbf{B}^*, \boldsymbol{\theta}^*, \mathcal{F} \sim \text{N}(\mathbf{m}_2, \Sigma_2 + \Sigma_\varepsilon),$$

noting the addition of the observation error covariance required for the marginal distribution of the noisy location $\mathbf{Z}^*(k)$ (from combining Eq. 5.2 with the marginal distribution of $(\mathbf{Z}(k), \mathbf{Z}(l))^T$).

5.3 Noisy observations of single state reindeer movement

The following example applies the methods of this chapter to a small subset of GPS data from reindeer (*Rangifer tarandus*). This example aims to demonstrate the impact of incorporating observation error into analysis rather than providing an informative ecological analysis and, for simplicity, implements only single-state movement. The inclusion of observation error is shown to have the effect of ‘smoothing’ out the movement process between a number of the observed locations, driven by the overall persistence that the full set of observations suggests. The following results can be reproduced with the `single_reindeer_error` example within the repository `CTStepTurn` available at <https://github.com/a-parton/CTStepTurn>.

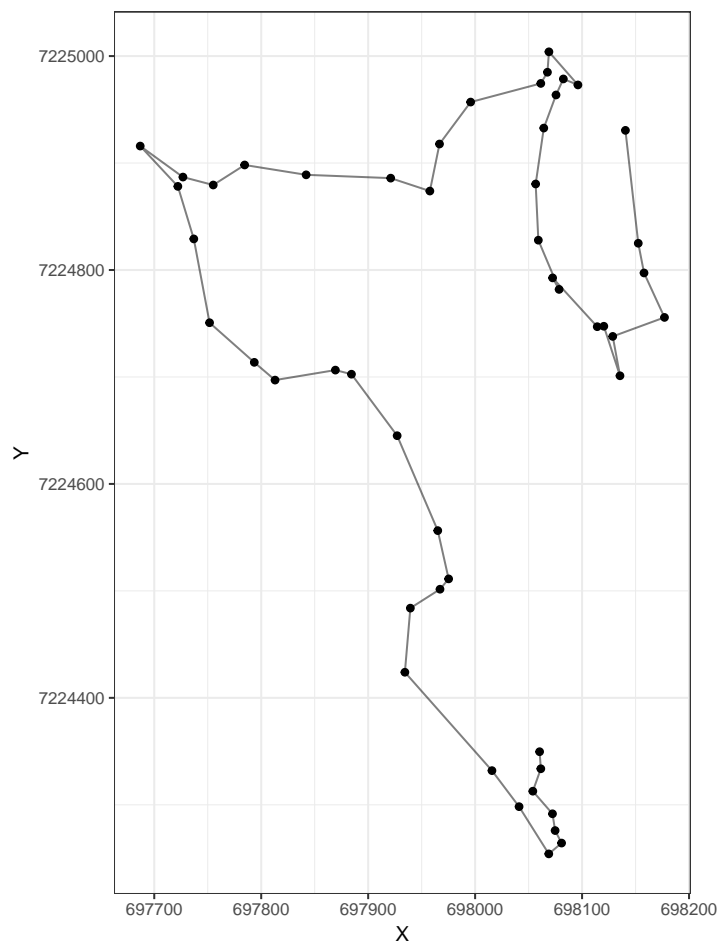


Fig. 5.4 Observed 2 minute locations of reindeer ‘b53.10’ (points, linked chronologically with lines).

5.3.1 Reindeer-b53.10 observations

A set of 50 GPS observations, taken at 2 minute intervals from the reindeer tagged as ‘b53.10’ were used. These observations form part of a larger dataset spanning a week, taken of multiple reindeer walking in the Malå herding community in northern Sweden, and collected by Anna Skarin. A short sample of the full dataset was chosen as multistate movement is obviously present in sampling periods spanning multiple hours. The observations used here are shown in Fig. 5.4, suggesting movement that is generally persistent but with a number of observations closely separated with ‘large turns’ between them.

In this example the single state model with correlated speed process of Sect. 3.4 is applied to the b53.10 observations, with an assumed model for observation error following the independent Gaussian perturbation described in Sect. 5.1. Interest thus involves learning

about the four movement parameters (consisting of a bearing variance and a speed mean, correlation and variance) along with the additional parameter describing the extent of observation error. Due to the observations here being at a fine time scale (2 minutes), it is expected that observation error will have a large effect on inference regarding the movement parameters. This is due to the high signal-to-noise ratio resulting from the level of expected observation error from GPS and the small distances being travelled over 2 minutes.

5.3.2 Implementing the inference algorithm

5.3.2.1 Initial values

The refined time scale for this example was chosen as 0.25 minutes, providing reconstruction at 7 locations between each pair of successive observations. Having observations at a frequent time scale suggests that only a small amount of additional reconstruction should be required as little uncertainty is expected. The approximation chosen was thought to provide a reasonable trade off between computational efficiency and flexibility of the underlying movement path (through providing a good approximation to continuous time).

The initial movement path at the refined time scale of 0.25 minutes was created by first perturbing the observed locations to create an initial set of ‘true’ locations. This followed the assumed Gaussian model for observation error, with variance of 100. The initial path was then created by taking an interpolating cubic spline between the initial ‘true’ locations. The initial bearing and speed process are shown in the Appendix, Fig. B.14 and, similarly, the initial location process is shown by the **black line** in Fig. 5.7. Initial movement parameters were set as estimates from this path configuration, given as

$$\Phi_M^{(0)} = \{\sigma_\theta^{2(0)} = 0.422, \mu^{(0)} = 22.3, \beta^{(0)} = 0.100, \sigma_\psi^{2(0)} = 175\},$$

$$\sigma_\varepsilon^{2(0)} = 100.$$

5.3.2.2 Prior information

A prior distribution specifying an upper bound on the ratio of the speed parameters to avoid the presence of negative speeds was applied (see Eq. 3.10). Additionally, a half normal prior was placed on the speed correlation parameter β with scale 2.25. Because β is featured within a decaying exponential term in the conditional distribution of the OU process (Eq. 3.10), once it becomes large (what value this is depends on the refined time scale in use) such a

term becomes zero and its value cannot be determined. At such a stage an MH sampler will drift around the parameter space above this threshold, and an example of where such an event occurred is given in the single state example at the observation time scale in Sect. 3.5.2. The placement of the prior on β discourages large values (90% quantile at 2.46 and median at 1.01) so that if there is little correlation in the speed process the MH sampler does not drift unnecessarily. However, for values of β that are within an estimable range, this prior is uninformative.

The conjugate prior for the bearing parameter σ_θ^2 is inverse gamma. In this example, this prior was defined with shape and scale both equal to 0.5. Such a prior is uninformative, being heavy-tailed with 90% prior credible interval being (0.260, 254) and prior median 2.20.

The conjugate inverse gamma prior for the observation error variance parameter was defined with shape 2 and rate 200. A 90% prior credible interval is (42.2, 563) and the prior median is 119 and the mean is 200. This prior was chosen to reflect the belief that GPS observation error would be present, discouraging negligible levels of error, and also that it is expected from previous studies that levels of error are as high as 20 m. The prior is therefore concentrated around the region of values lead to a non-negligible probability of producing errors upto 20 m (i.e. looking at within two standard deviations). All remaining parameters (speed mean and long term variance) had flat priors.

5.3.2.3 Implementation

Perturbation variances in the random walk sampler for the speed parameters were assigned based on a pilot run

$$\boldsymbol{\rho} = \{\rho_\mu = 2, \rho_\beta = 0.55, \rho_{\sigma_\psi^2} = 10\}.$$

The algorithm of Sect. 5.2 was applied for 10^7 iterations, with each iteration consisting of a single parameter update and 50 refined path updates on random sections of path with lengths ranging 4–15 points (i.e 1–3.75 minutes). Samples were thinned (for memory storage purposes) by a factor of 2,000 and the first fifth were treated as a ‘burn-in’ period, leaving 4,000 stored samples of parameters and reconstructed refined paths. Computational run-time for this example was under 24 hours.

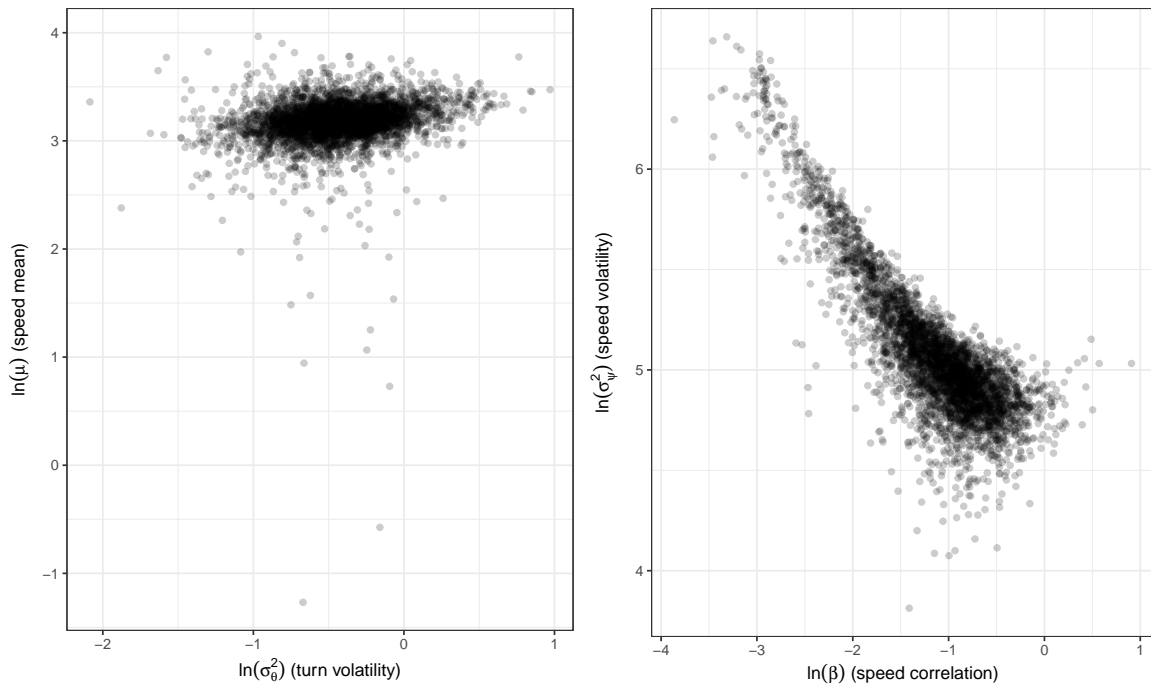


Fig. 5.5 Posterior samples of the movement parameters (on log scale) for the reindeer b53.10 example.

5.3.3 Parameter results

The posterior samples of the movement parameters are given in Fig. 5.5 and shown as kernel density estimates (which also include the observation error variance parameter) in Fig. 5.6. For the kernel density estimates, in the cases where non-flat prior distributions were specified, the prior density is also included. The trace and ACF of all sampled parameters are given in the Appendix, Figs. B.15 and B.16. These summaries show there to be some positive correlation (0.34) between the bearing variance and the mean speed, which is to be expected as a more tortuous path between observations must be longer. There is negative correlation (0.53) between the correlation and variance of the speed process, i.e. if there is less correlation in the speed process and it is instead more strongly attracted to the long term mean then the long term variability in speed is also lower. All other pairs of parameters had correlation weaker than ± 0.2 .

Posterior summary statistics of the parameters are given in Table 5.1. The bearing parameter describes movement that is persistent/directed, with the 95% quantile of the posterior reconstructed turns (differences in the bearing process) being less than 0.726 radians (~ 42 de-

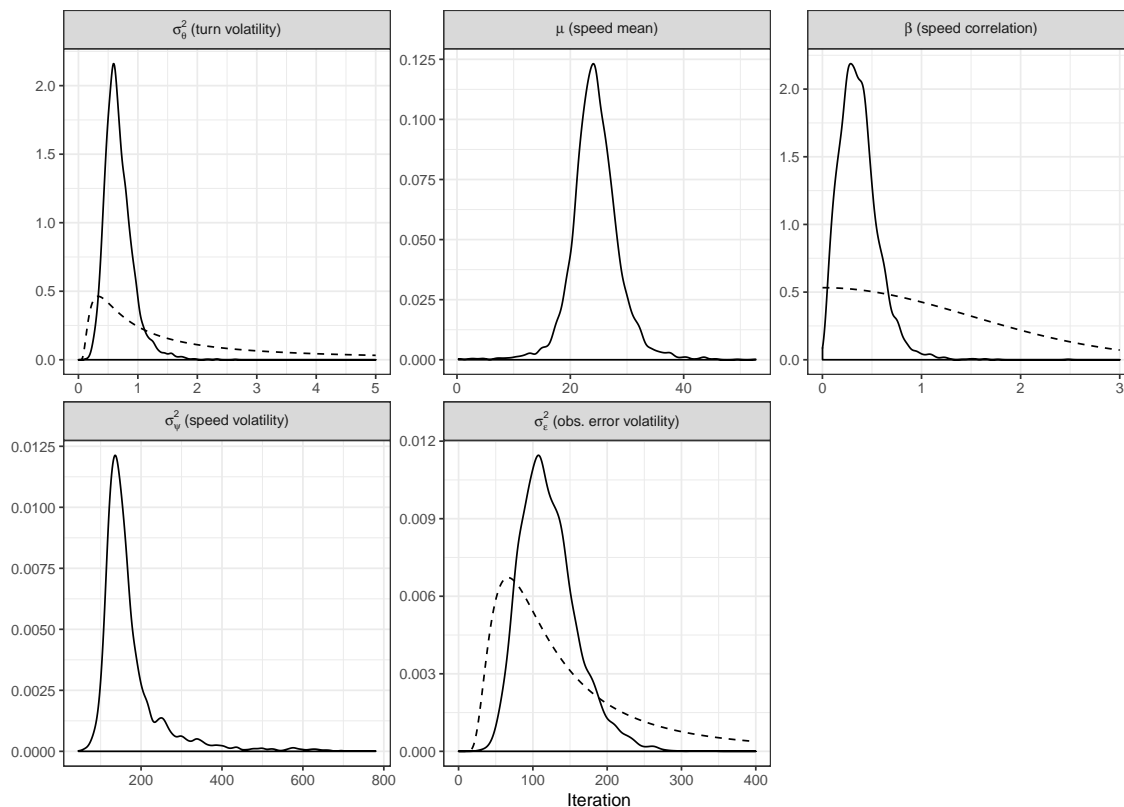


Fig. 5.6 Kernel density estimates (**solid line**) of the posterior samples of the movement and observation error parameters in the reindeer b53.10 example. Estimates are based on thinned samples with burn-in period discarded. In the cases where non-flat prior distributions were specified, such a prior density is also included (**dashed line**).

grees). The 95% posterior credible interval can be seen to be concentrated around a median lower than that of the prior, being positively skewed but with a non-heavy tail (see Fig. 5.6).

The posterior of the long term mean speed is symmetrical, with median 24.3 metres/minute. The posterior of β , the correlation parameter of the speed, is concentrated at low values in contrast to the prior distribution (see Fig. 5.6). The speed process is highly correlated—for example, the 95% posterior credible interval for the exponential term in the OU process of Eq. 3.10 at the refined time scale is (0.665, 0.963) and the median is 0.842. The posterior of σ_ψ^2 leads to the equilibrium distribution of the speeds having high uncertainty, but because of the high speed correlation (through β) the short term speed distribution has low variance.

The posterior and prior distributions of the observation error variance have similar medians (prior 119, posterior 116), however the posterior mode is higher than the prior (prior 66.7, posterior 107) and the posterior is more concentrated around its mode than the prior.

Table 5.1 Posterior summary statistics (2.5%, 50%, 97.5% quantiles) for the sampled movement and observation error parameters in the single state reindeer b53.10 example. ESS is also included, taken from The ESS is taken from the 4,000 samples thinned from 8×10^6 iterations.

	2.5%	50%	97.5%	ESS
σ_θ^2 (bearing variance)	0.327	0.638	1.28	191
μ (long term speed mean)	16.7	24.3	33.1	878
β (speed correlation)	0.0740	0.343	0.815	968
σ_ψ^2 (long term speed variance)	97.9	150	412	192
σ_ϵ^2 (observation error variance)	62.1	116	212	2378

5.3.4 Path reconstruction results

The posterior distribution of path reconstruction at the refined time scale of 0.25 minutes is shown in Fig. 5.7. Each **green line** represents a single path reconstruction, the **black line** highlights the path that the sampler was initialised at, and the observations are displayed by **points**. Presenting a number of path reconstructions when error is present, as in Fig. 5.7 which features 80 distinct reconstructions, may not be useful for aiding in parameter interpretation but is useful in determining the space-use of the animal and the level of observation error. Uncertainty in location over time and space use can be informally judged by the ‘width’ and density of the band of sampled locations running between observations, also highlighting the observations with high noise.

The posterior reconstructed paths aid in quantifying the error associated with observed locations. Although the assumed model for observation error is Gaussian with centre at zero, the posterior distribution of the error at an observation time does not necessarily have such a distribution. This is demonstrated in Fig. 5.8, showing a section of the movement path with posterior samples of the ‘true’ locations at two observation points (labelled **A** and **B**). Such errors are then plotted in more detail in the **right panels**. In both cases, the majority of errors in the x and y directions are within 20 m, which agrees with the prior expectation. The shape of the errors, however, differs between the two examples. In the case of the observation labelled **A (top right panel)** the distribution of the posterior observation error appears Gaussian with centre being the observed location, i.e. no error. In contrast, the observation labelled **B (bottom right panel)** is skewed with the observed location non-central to the distribution of the ‘true’ location. Such an event arises when the full set of observations suggest a movement path that is consistently biased in one direction to the observation; in the case of location **B**, the reconstructions form a smooth curve rather than

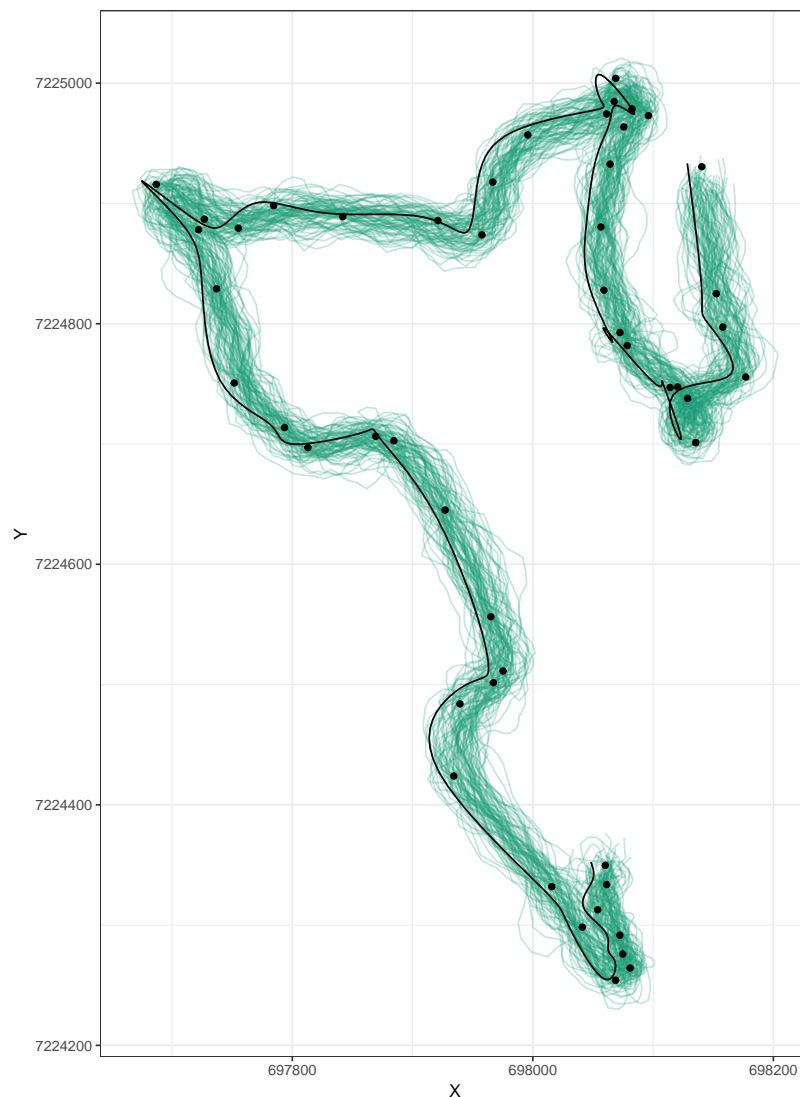


Fig. 5.7 Example sampled refined path reconstructions (**green lines**) for the reindeer b53.10 example with noisy observations. Observations (**points**) and initial path are highlighted (**black line**).

the abrupt ‘corner’ created by linear interpolation and hence the paths consistently lie to the west of the observed location. Such information on the ‘true’ locations of the path enable more reliable determination of the animal’s space use.

Although large samples of path reconstructions aid in the estimation of space use and observation error, when they are considered together it is difficult to identify fine-scale features of the movement path or interpret the inferred parameter estimates. Fine-scale interpreta-

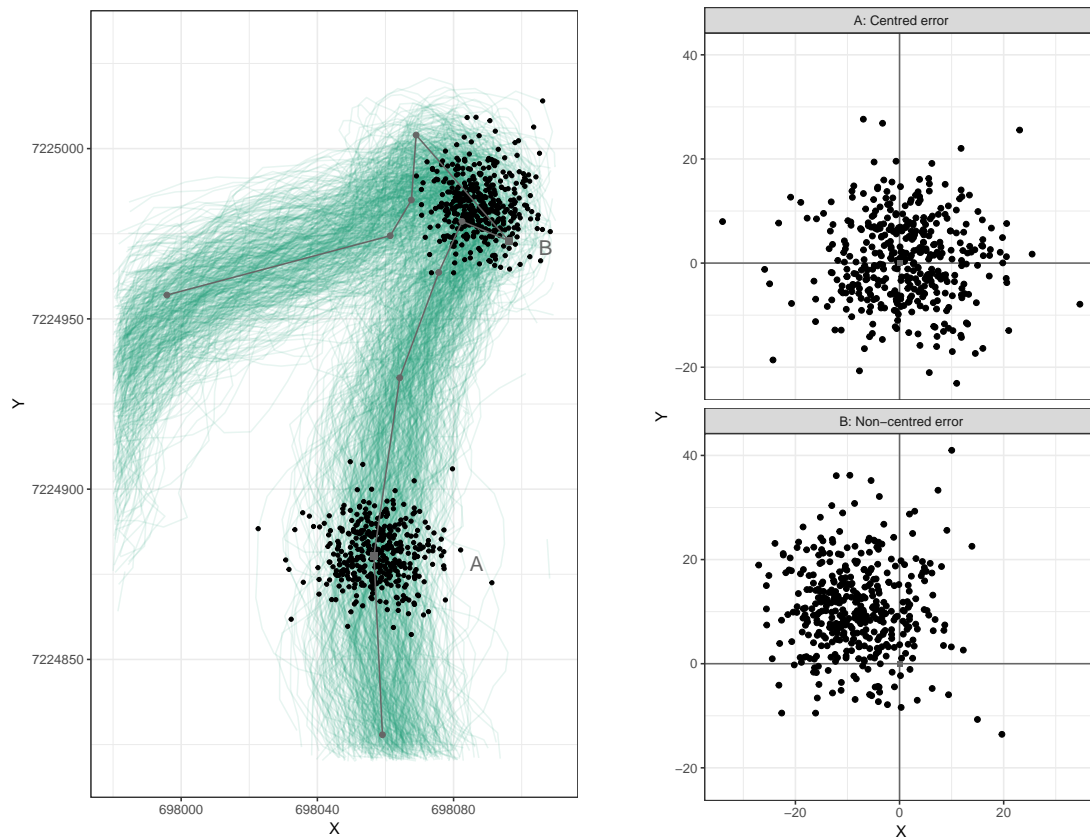


Fig. 5.8 The presence of centred and non-centred observation error in the single state reindeer example. **Left:** section of the movement path, displaying observations (**grey points** and linearly interpolated **line**), posterior samples of reconstructed paths (**green lines**), with corresponding locations at two observation times (**black points** at **A** and **B**). **Top right:** posterior samples of error at the time of observation **A** (**black points**) with no observation error highlighted (**grey lines**). **Bottom right:** posterior samples of error at the time of observation **B**.

tion can be carried out on small numbers of reconstructed paths, such as the five examples shown in Figure 5.9, taken from equally spaced iterations within the MCMC sampler. The full movement paths are shown in the **left panel**, with observations highlighted, allowing instant correspondence between the inferred parameters and the types of movement path they define. In this case, the reindeer's path is persistent, with smooth curves/loops. The marked difference in the four example reconstructions between pairs of observations suggests that the linear interpolation employed by discrete time models could be ignoring important characteristics of movement.

The two highlighted regions in the **left panel** of Fig. 5.9 are detailed in the **right panels**. Between observations 21 and 25 there is a 'sharp turn' and the two observations 41 and 42

create a sequence of ‘double-back’ moves. In a discrete time analysis this would amount to multiple turns of $\pm\pi$ radians, leading to large estimates of the bearing variance. In this example, however, the path is reconstructed given the information from all observations, which in this case suggests persistent movement. To allow for this, the reconstructions ‘smooth out’ the sharp turning behaviour, often with an increase in error around the affected observations. This results in a smoother corner being reconstructed in all examples between observations 21 and 25 than a linear interpolation (**top right panel** of Fig. 5.9). For the observations 41 and 42 (**bottom right panel** of Fig. 5.9) the example reconstructions feature two types of movement; a loop that allows for the ‘double-back’ movement but at a scale less

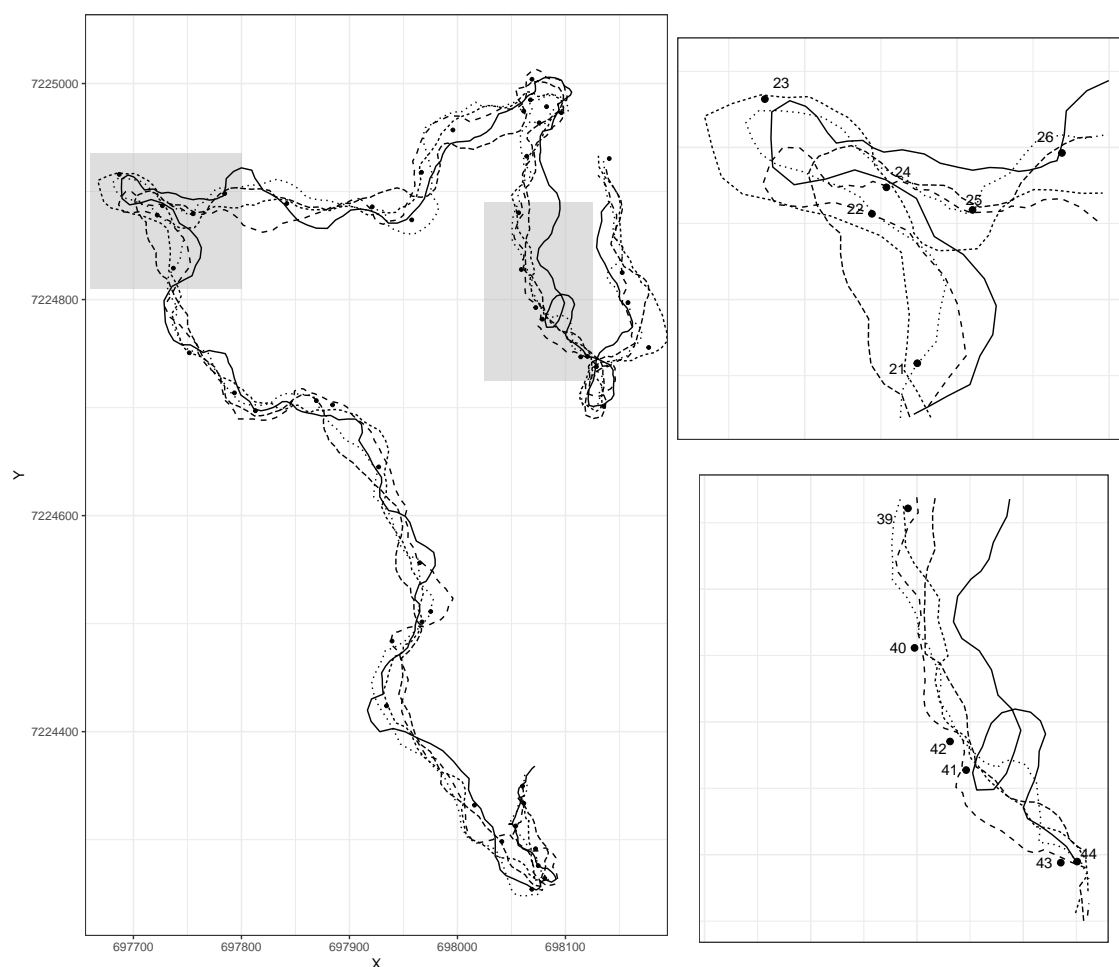


Fig. 5.9 Five examples of posterior path reconstructions for the reindeer b53.10 example. **Left:** The movement path (**line** for each reconstruction) with observations (**points**). The two **grey areas** highlight detailed sections featured in the **right panels**. In both cases, observations have been numerically labelled by their position in time.

volatile than linear interpolation, and straight paths that ignore the ‘double-back’ movement by increasing the level of error around the observations.

5.4 Noisy observations of two-state gull movement

This example applies a two-state movement model with the presence of observation error to a subset of GPS data from lesser black-backed gulls (*Larus fuscus*). For further information on the full data, which is available at Garthe et al. (2016a), see Garthe et al. (2016b). Two implementations were carried out on this data with differing approximations to the refined movement path to investigate the effect this has on inferred parameters. In this case, the differing time approximations had little effect on the inferred behavioural parameters, but led to differences in the inferred speed parameters and level of observation error.

5.4.1 Gull-3 observations

A subset of 1,001 GPS observations, taken at roughly 3 minute intervals from the gull tagged as individual 3 (incorporating a sampling period of over 52 hours) are shown in Fig. 5.10. Unlike the previous examples with elk and reindeer, the gull observations do not have a regular sampling interval, ranging between 3.05–7.7 minutes and with a mean interval of 3.13. Thus, applying a discrete-time model would require some form of interpolation of the observations at a regular time scale. A two-state switching model was applied with the independent Gaussian assumption of observation error.

The two-state model was implemented with two refined time scales, referred to throughout the following as 0.5 and 1. The actual refined time scale for a given path reconstruction, however, is dependent upon the times of observations (which are measured to within 0.01 minute accuracy) and the inferred behavioural switches. For each reconstruction the refined time scale was chosen as being the sequence of times that included at least all observation and behavioural switch times and had points spaced no greater than 0.5 and 1 (for each implementation) minutes apart. Hence, in many cases, this resulted in a refined time scale that was less than 0.5 or 1 minutes, respectively. Interest here was in comparing the parameter estimation between the two different approximations.

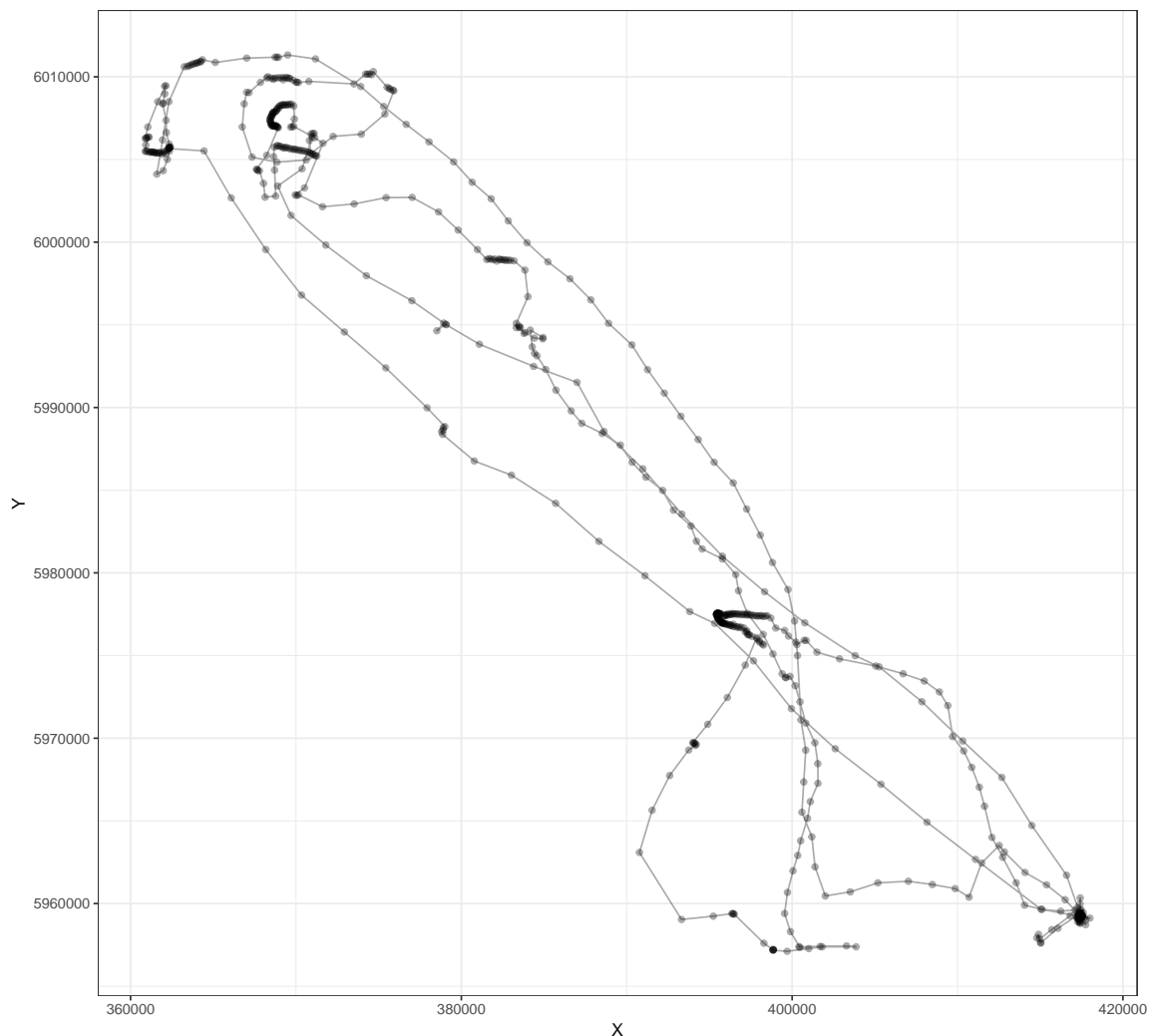


Fig. 5.10 Observed locations of gull-3 (**points**, linked chronologically with **lines**).

5.4.2 Implementing the inference algorithm

5.4.2.1 Initial values

The initial movement paths for both implementations were created using an interpolating cubic spline between observations, as in the previous example on reindeer data. The refined time scale in each of the two cases was chosen to not be greater than 0.5 and 1, and due to the sampling frequency of the observations, this led to initial refined paths with mean time scales of 0.44 and 0.78, respectively (with around 4 and 8 locations between each pair of observations). Initial parameters in each implementation were set as maximum likelihood estimates of the initial path configuration.

5.4.2.2 Prior information

As in the elk example, a prior distribution was applied to avoid state label switching by categorising state 1 as having a lower mean speed than that of state 2. The same prior on the speed correlation parameter was applied as that in the reindeer example of Sect. 5.3 to avoid drifting if the inferred value was at a level where no speed correlation would be present (the actual inferred range of this parameter was found to be so much smaller than this threshold that it was uninformative). The conjugate inverse gamma prior for the error variance was defined with a shape 5 and rate 300. A 90% credible interval for this was (32,152), with a median of 62. This was chosen to reflect the belief that some observation error would be present, but that the level of such is uncertain within the region of realistic GPS device errors. Due to the distances travelled between observations being large in this example (in comparison with the reindeer example of the previous section), observation error is not expected to affect the ability to estimate the movement parameters. The choice of prior for the error variance here is therefore not expected to be important. All remaining parameters had flat priors.

5.4.2.3 Implementation

Perturbation variances for the speed parameters were set based on pilot runs, and the inference algorithm was implemented for 5×10^6 iterations, with 50 path section updates for every parameter update (with path section updates ranging between 4–11 points). Samples were thinned (for memory storage purposes) by a factor of 2,000 and the first fifth treated as burn-in. Computational run-time for the two cases were under 4 days, with the implementation at time scale 0.5 taking longer than that at time scale 1.

5.4.3 Behaviour results

Posterior samples of the behavioural switching parameters for the two implementations are shown in Fig. 5.11, with sample trace and kernel density estimates given in the Appendix, Figs. B.17 and B.18. Posterior summary statistics are given in Table 5.2. The two implementations inferred similar values for the behavioural parameters, with the estimation being more uncertain in the case where the movement path was estimated at a coarser time scale. In both cases the inferred switching rate was higher for the travelling state over the foraging state, leading to 90% credible intervals for the expected residency time of (50,100) and

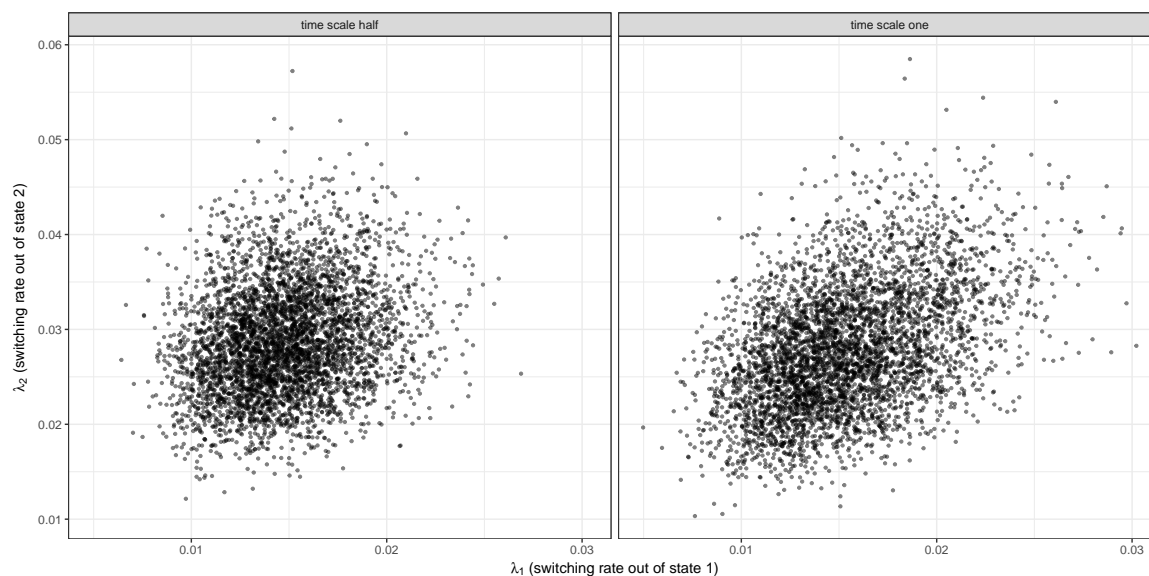


Fig. 5.11 Posterior samples of the behavioural parameters (on a log scale) for gull-3: λ_1 is the switching rate out of the ‘foraging’ state and λ_2 is the switching rate out of the ‘travelling’ state. Implementations with refined time scale of 0.5 **left panel** and 1 **right panel**.

(25,50) minutes in states 1 and 2, respectively. The similarity of the inferred behaviour parameters between the two implementations is unsurprising due to the behavioural reconstruction remaining exact in the algorithm for inference.

Table 5.2 Posterior credible intervals (5%, 50%, 95% quantiles) of the sampled behavioural parameters in the gull-3 example.

	δt	5%	50%	95%
λ_1	0.5	0.0103	0.0146	0.0200
	1	0.00990	0.0151	0.0220
λ_2	0.5	0.0202	0.0284	0.0392
	1	0.0181	0.0274	0.0399

Based on posterior samples of the reconstructed behavioural process, Fig. 5.12 shows the probability of being in each behaviour throughout the course of the sampling period for both implementations. The estimated behavioural process is similar in each case, which is to be expected due to the similarity in estimated behavioural parameters and because the reconstruction of the behavioural process is exact.

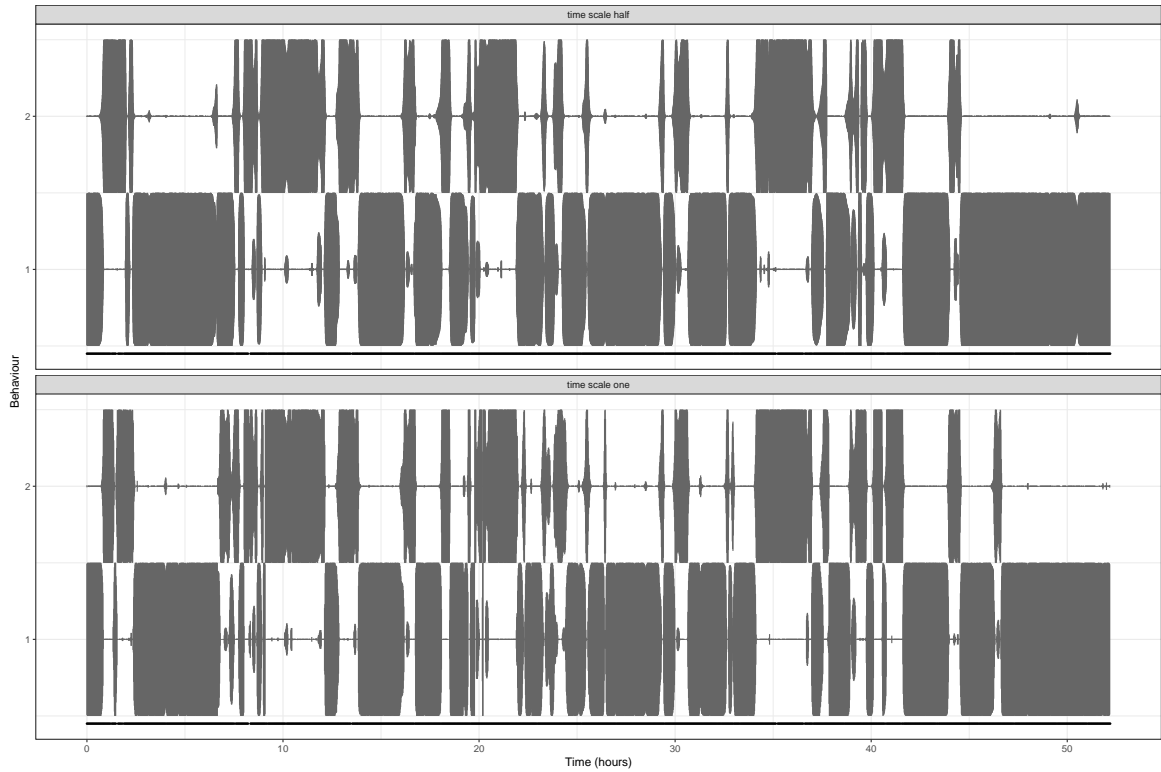


Fig. 5.12 Posterior probability of residing in each behavioural state over time, calculated from the sampled reconstructions of the behavioural process (**grey bars**). The taller a bar at a given time, the higher the proportion of sampled reconstructions classified as that behaviour (so each bar for state 1 and 2 sum to a constant value). The times/frequency of observations is highlighted by **points** along the horizontal axis. Implementations with refined time scale of 0.5 **top panel** and 1 **bottom panel**.

5.4.4 Movement results

Posterior samples of the movement parameters are shown in Fig. 5.13, with sample trace and kernel density estimates given in the Appendix, Figs. B.19 and B.20. Posterior summary statistics are given in Table 5.3. The estimates of the movement parameters for state 1 (foraging state) differ between the two implementations. The bearing parameter σ_{θ}^2 for this state is so high in each implementation as to produce uniform turning angles, rendering the difference in estimation of this parameter inconsequential. The estimate of the speed correlation, β , and long-term variance, σ_{ψ}^2 , in state 1 for the two implementations differ greatly between the two implementations (although the product of these quantities, determined by the slope in the scatter plot of Fig. 5.13, is comparable). In contrast, the estimates of the movement parameters for the travelling state (state 2) are similar for both implementations. The difficulty in estimating the speed correlation, β and variance, σ_{ψ}^2 at coarse time approximations

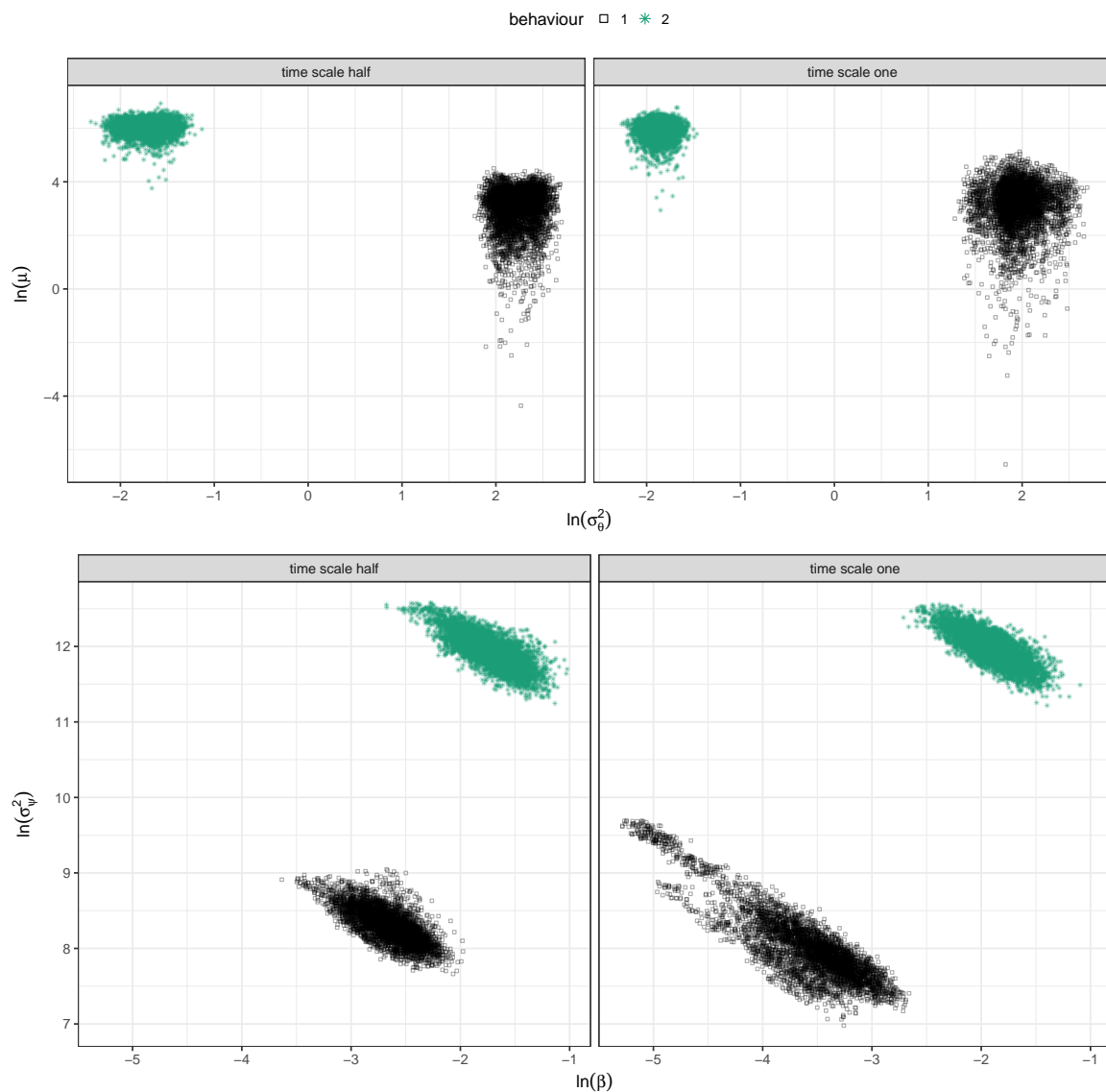


Fig. 5.13 Posterior samples of the movement parameters (on a log scale) for gull-3. Implementations with refined time scale of 0.5 **left panel** and 1 **right panel**.

was a property previously identified in the example of Sect. 3.5.2. That the uncertainty for these parameters is much larger for the implementation with coarser approximation to the movement process suggests that such an approximation is not fine enough to estimate these features of the movement process for this set of observations.

Table 5.3 Posterior credible intervals (5%, 50%, 95% quantiles) of the sampled movement parameters in the gull-3 example.

	δt	State 1			State 2		
		5%	50%	95%	5%	50%	95%
σ_θ^2	0.5	7.04	9.09	12.4	0.134	0.185	0.239
	1	4.76	6.98	10.9	0.129	0.154	0.185
μ	0.5	3.90	22.7	50.1	268	428	599
	1	3.89	25.4	69.6	216	389	562
β	0.5	0.0460	0.0715	0.101	0.117	0.178	0.256
	1	0.00909	0.0288	0.0506	0.100	0.153	0.218
σ_ψ^2	0.5	2830	3990	6060	108000	154000	225000
	1	1690	2810	8920	111000	154000	227000

Table 5.4 Posterior credible intervals (5%, 50%, 95% quantiles) of the sampled observation error variance in the gull-3 example.

	δt	5%	50%	95%
σ_ε^2	0.5	28.6	35.6	43.7
	1	35.6	51.7	68.8

5.4.5 Error results

Posterior kernel density estimates of the observation error variance are given in Fig. 5.14 and summary statistics in Table 5.4. The estimation of the level of observation error differed between the two implementations, estimating smaller levels of error with a better approximation to the movement process. The coarser approximation has a posterior closer to the prior distribution. It may be that when the approximation to the continuous-time movement process is too coarse, the inferred level of error must be large to allow for an inability to describe the movement process well enough at such an approximation. Alternatively, it may be that at this level of approximation there is not enough information to estimate the error variance, and so the posterior is dominated by the prior information. It is unknown whether the difference in estimated speed parameters is due to the refined path approximation directly, or whether the estimates differ indirectly because of the difference in inferred observation error. This lack of identifiability between the movement parameters and the observation error is explored further in the simulated example of Sect. 5.5, in which the error variance is fixed at a number of known quantities to investigate the effect this has on movement parameter estimation.

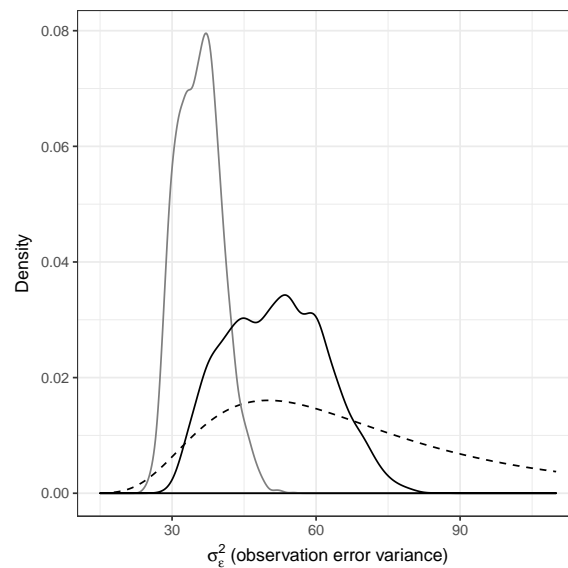


Fig. 5.14 Posterior kernel density estimate of the observation error variance for gull-3. Implementations with refined time scale of 0.5 **grey** and 1 **black**. The prior distribution for the error variance is also included (**dashed line**).

5.5 Simulated example

The following example applies a restricted version of the methods of this chapter to a simulated dataset of noisy observations. Given a set of observations simulated with a known error variance, the simplification applied implements the inference algorithm with the actual errors unknown, but assuming a known value for the error variance (i.e. the error parameter is fixed prior to implementation and Sect. 5.2.1 is not carried out). In the following this is applied for a number of fixed error parameters, from 0 (not considering error on the error-ridden observations) through a range of values including, and greater than, the true value used for the simulation. This example demonstrates the effect that the level of observation error has on the ability to estimate the parameters and reconstruct the unobserved path. The following results can be reproduced with the `single_simulation_error` example within the repository `CTStepTurn` available at <https://github.com/a-parton/CTStepTurn>.

5.5.1 Underlying movement and observations

Single state movement following Sect. 3.4 was simulated with parameters

$$\Phi = \{\sigma_\theta^2 = 0.7, \mu = 20, \beta = 0.2, \sigma_\psi^2 = 80\},$$

between the times 0–400 at an approximate time scale of 0.01.

The ‘true’ observations of the movement path were created by sub-sampling the simulated locations at intervals of 200, giving 200 observations at a time scale of 2. Noise was then added according to the independent model of Sect. 5.1, with error variance given by

$$\sigma_{\varepsilon}^2 = 25.$$

The simulated path, with corresponding true and noisy observations, is given in Fig. 5.15. Although the level of error applied here appears small at the scale of the full movement path, the inset, detailed section shows that the error applied alters the movement between successive observations noticeably and is expected to have a significant effect on the inferred movement.

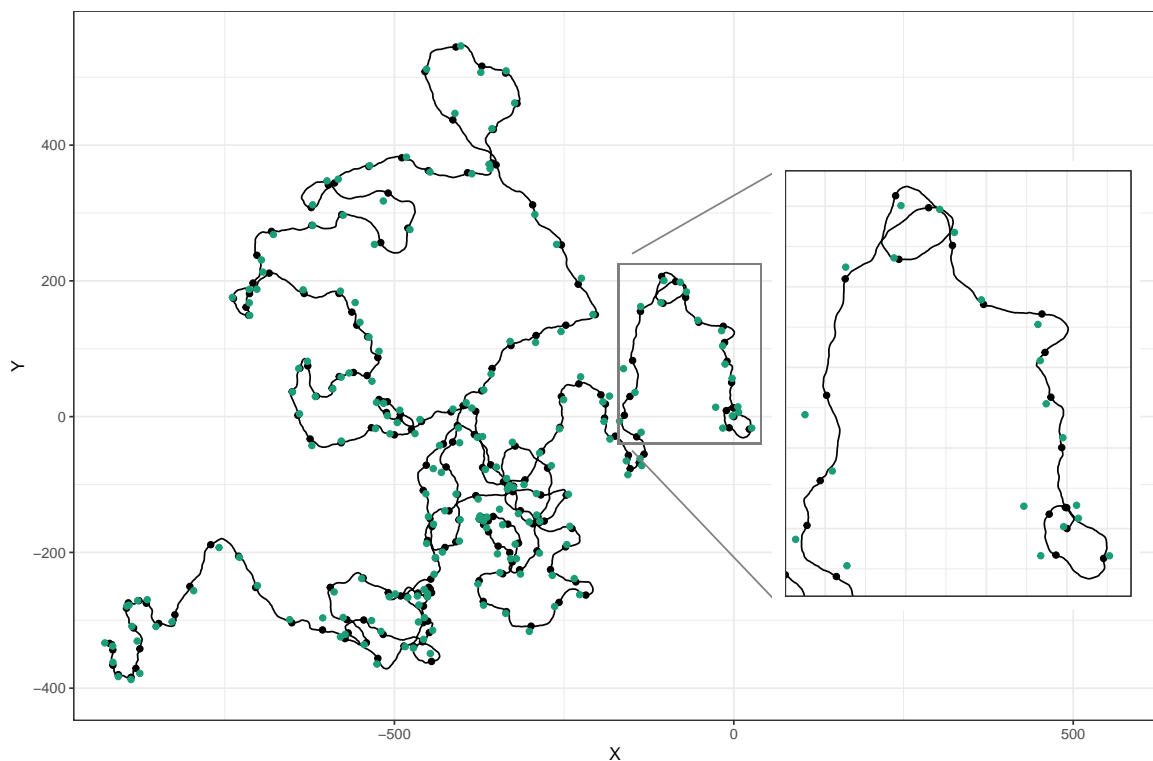


Fig. 5.15 Simulated single state movement path (**black line**) with true locations at the observation times (**black points**) and the locations observed with noise (**green points**). **Inset panel:** detailed section between times 0–40.

5.5.2 Implementing the inference algorithm

The ability of the algorithm at inferring the movement process in the presence of error, along with the effect the size of such an error has, is explored through a number of implementations. The observation error variance parameter was fixed in each run with values ranging at intervals of 5 between 0–50 (note that the run with variance of 0 is equivalent to ignoring error and the run with variance of 25 applied the actual variance used to simulated the observations). To ‘fix’ the error variance, the algorithm of Sect. 5.2 was applied ignoring the step to sample the error parameter in Sect. 5.2.1. The full inference algorithm, in which the observation error variance is assumed unknown and estimated, was also implemented for comparison.

5.5.2.1 Initial values

Each implementation with fixed error variance (with the exception of that with no error) used the same initial movement path. This was created by drawing a set of ‘true’ observations by perturbing the observed locations, according to the true underlying model for error. For the implementation with error variance of zero, the observed locations were used. The path was then created at a refined time scale of 0.25 by taking an interpolating cubic spline between the drawn, ‘true’ locations. This gave a reconstruction of seven unknown locations between each pair of observations. Initial movement parameters were set as estimates from this initial configuration.

When the observation error variance was assumed unknown, two implementations were carried out that differed only in the initial values. The initial movement path was created assuming that the error variance was 25 (the true value) and 50, with the initial value of the error variance also set at the respective value.

5.5.2.2 Prior information

A half normal prior was placed on the speed correlation parameter β with scale 25 (see the discussion on this prior in Sect. 5.3.2). Such a prior discourages large values of β (with 95% quantile of 9.82) and so for values of β that are within an estimable range, this is uninformative. The conjugate, inverse gamma prior for the bearing parameter σ_θ^2 was defined with both shape and scale of 0.001, representing weak information.

When the observation error variance was assumed unknown a prior was set representing no information about its value, being an inverse gamma with shape and rate both equal to 0.001.

5.5.2.3 Implementation

The inference algorithm was applied for 5×10^6 iterations, with each consisting of a single parameter update and 100 refined path updates on random sections of the path with lengths between 4–17 points. Perturbation standard deviations for the speed parameters were based on pilot runs. Samples were thinned (for memory storage purposes) by a factor of 1,000 with the first 500 samples treated as burn-in (10%), leaving 4,500 samples for estimation with ESS given in Table 5.5. Each implementation took less than 24 hours in computational time to complete.

5.5.3 Parameter results

The posterior samples of the parameters are given in Figs. 5.16–5.17, and the trace in the Appendix, Fig. B.21. Summaries of the posterior distributions are given as kernel density estimates in Fig. 5.18, boxplots in the Appendix, Fig. B.22, and in Table 5.5.

Fig. 5.16 shows the estimated bearing variance decreasing towards the simulation value as the error variance is increased. Fig. 5.18 shows that the uncertainty around the estimation of this parameter also decreases as the error variance is increased. The simulation value has not been captured well by any of the implementations with error variance less than the true value, with estimation improving as the error is increased. Both here, and in the reindeer example of Sect. 5.3. incorporating observation error allows for paths with lower bearing variance (and higher persistence) than would be the case when observation error is omitted, and although this is closer to the true value, the presence of this error makes estimation of the correct bearing parameter more difficult. For example, for the implementation with error variance set as the true value, the estimation of σ_θ^2 was too high, with posterior median of 0.867. For comparison, in the simulation example of Sect. 3.5.2, which did not include error but also had a bearing parameter of 0.7, the posterior median from the two implementations were 0.66 and 0.63 (the 97.5% quantile was 0.85 and 0.81, therefore still lower than the median in this example).

The mean speed was the easiest parameter to estimate; even with the complication of observation error. Similarly to the bearing parameter, the estimate of the mean speed reduces as the assumed error is increased (see Fig. 5.16). The uncertainty around such an estimate,

however, increases as the assumed error does also (see Fig. 5.18), with all implementations having essentially the same upper estimate of the mean speed but a lower median. This reflects the findings in the simulation example of Sect. 3.5.2 that did not include error, in which the uncertainty around the estimate of the mean speed remained high even when the true simulation at the 0.01 timescale was known. When no observation error is present, the mean speed is overestimated and not included in the 95% posterior credible interval in Table 5.5—this is to be expected when the bearing variance is also the highest of all the implementations. When there is observation error included in the implementation, the

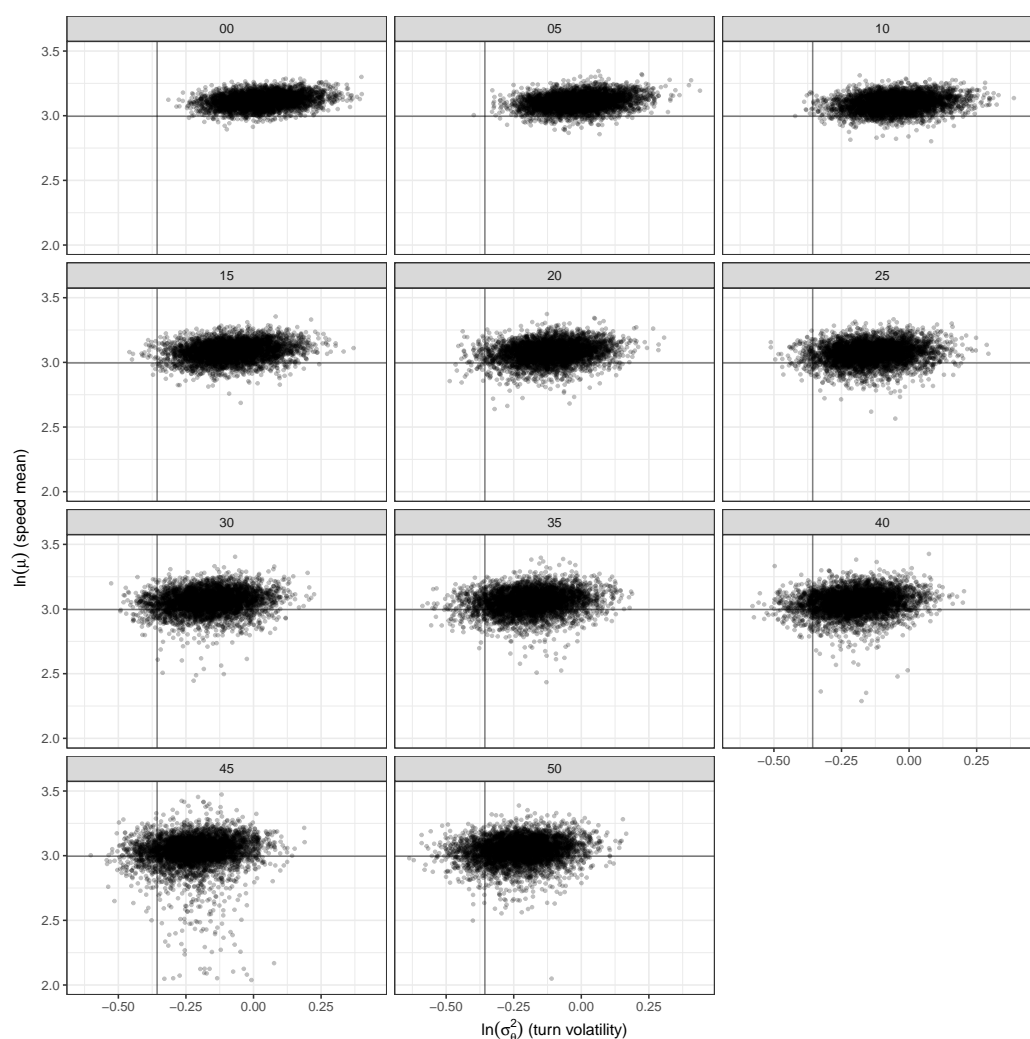


Fig. 5.16 Posterior samples of the bearing variance against mean speed (on log scale) for the observation error simulation example, faceted into panels for the level of fixed error variance. In each case, the true simulation value is given by **grey lines**.

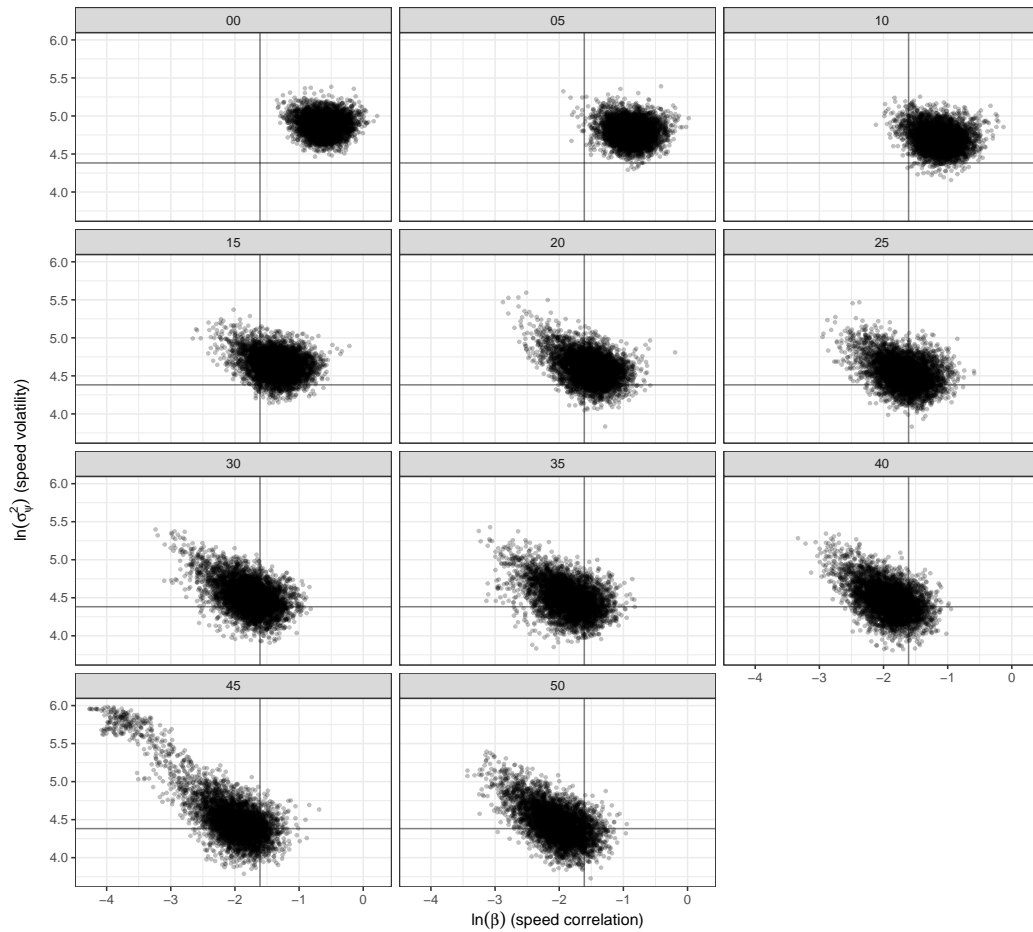


Fig. 5.17 Posterior samples of the speed correlation against variance (on log scale) for the observation error simulation example, faceted into panels for the level of fixed error variance. In each case, the true simulation value is given by **grey lines**.

estimate of the mean speed is not improved upon greatly as the error increases past the true simulation value.

When no, or little, observation error was allowed, the correlation in the speed process could not be estimated, with the sampled values leading to little correlation in the process. When the error variance was assumed higher than 20 (with the true value having been 25), the speed correlation was estimated well (see Fig. 5.17 and Table 5.5), however, once the observation error becomes large (values of 40 and over) β begins to be underestimated. As the assumed error increased, the uncertainty around the estimate for the speed correlation reduced, as shown in Fig. 5.18. The estimates of the long term speed variance are also similar; having been overestimated when no error was incorporated, with the true value only being included in the 95% posterior credible interval for error variances of 15 and greater. As shown in

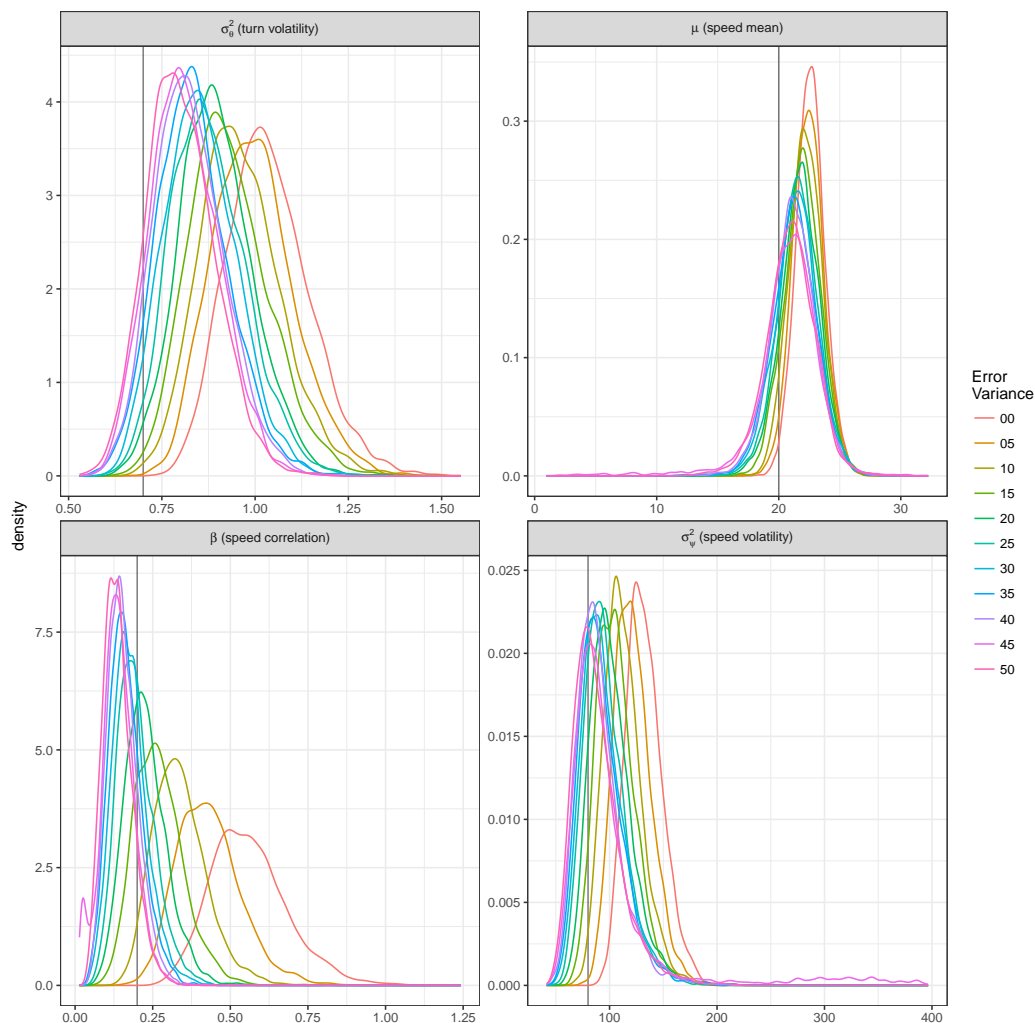


Fig. 5.18 Posterior kernel density estimates of the movement parameters for the observation error simulation example, shown for each implementation that used a different level of fixed error variance. In each case, the simulation value is given by **grey vertical line**.

Figs. 5.18 and B.22, for assumed errors at the true simulation value and greater, the estimate of the long term speed variance is relatively stable.

For comparison, Fig. 5.19 shows the posterior kernel density estimates of the movement parameters for the two implementations in which the error variance was assumed unknown and estimated, and Table 5.6 gives the posterior credible intervals. The initial configuration of the movement path and error variance parameter can be seen to have had little effect on the inferred parameter values. Estimated movement parameters are similar to those found when the error variance was fixed at the true value of 25 (see Table 5.5) and the error variance parameter has been estimated well in spite of weak prior information.

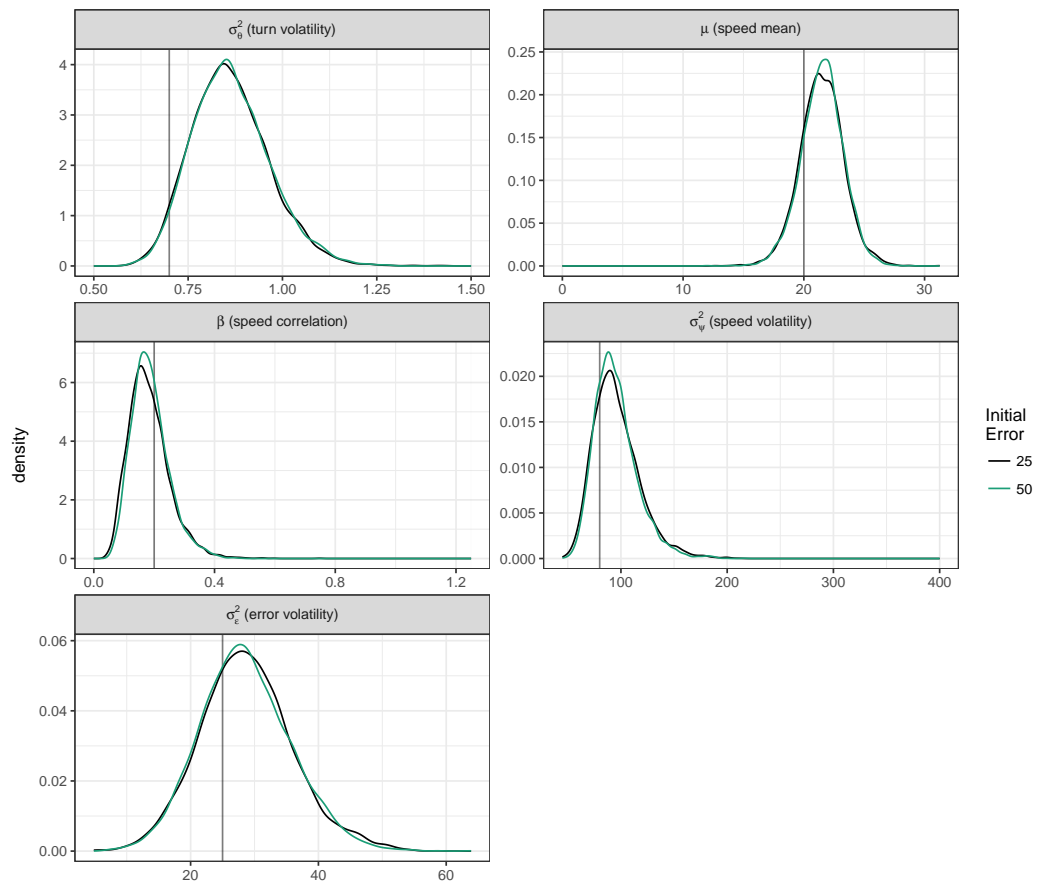


Fig. 5.19 Posterior kernel density estimates of the movement parameters for the observation error simulation example when the error parameter is unknown. The simulation value is given by **grey vertical line**.

Table 5.5 Posterior credible intervals (and ESS) for the movement parameters in the observation error simulation. The ESS is taken from the 4,500 samples thinned from 4.5×10^6 iterations. Note that the simulation values were $\{0.7, 20, 0.2, 80\}$ and the error variance was 25. Intervals that do not contain the true value are highlighted.

σ_{ϵ}^2		2.5	50	97.5	ESS		2.5	50	97.5	ESS
0		0.844	1.03	1.28	1719		20.3	22.6	25.0	3269
5		0.806	0.990	1.23	1561		19.9	22.4	25.0	4057
10		0.769	0.950	1.19	1654		19.5	22.2	24.8	4232
15		0.746	0.919	1.17	1550		18.9	21.9	24.9	4368
20		0.713	0.889	1.11	2231		18.4	21.8	24.9	4374
25	σ_{θ}^2	0.699	0.867	1.09	2031	μ	18.2	21.6	25.0	4414
30		0.681	0.849	1.06	2865		17.8	21.4	25.0	2374
35		0.660	0.829	1.04	1973		17.6	21.3	24.8	3346
40		0.660	0.819	1.02	1864		17.4	21.2	24.8	3052
45		0.649	0.806	1.01	1839		14.3	21.0	24.9	2993
50		0.638	0.795	0.997	2075		16.9	21.0	24.7	2582
0		0.350	0.549	0.836	726		103	131	170	1422
5		0.258	0.423	0.665	726		92.0	120	160	1891
10		0.197	0.329	0.542	723		82.9	110	151	1757
15		0.138	0.263	0.451	560		75.9	104	147	1812
20		0.112	0.221	0.379	585		70.3	98.6	149	1270
25	β	0.095	0.188	0.338	546	σ_{ψ}^2	65.7	92.9	143	1514
30		0.079	0.171	0.303	531		62.8	90.8	145	478
35		0.071	0.157	0.285	434		61.2	88.7	142	1049
40		0.073	0.148	0.264	641		59.1	86.5	140	1127
45		0.025	0.136	0.249	601		58.3	87.8	316	505
50		0.059	0.133	0.245	424		56.4	84.5	145	894

5.5.4 Path reconstruction results

The posterior distribution of the path reconstruction at the refined time scale of 0.25 minutes for four of the implementations (no error, error variance of 5, the true error variance of 25, and the largest error variance of 50) are shown in Fig. 5.20 between the times 0–40. Each **green line** represents a single path reconstruction, the **circular points** highlight the true locations at observation times, and the **square points** give the noisy observations.

For the implementation without observation error, the paths must move through each error-ridden observation, and so if the location process was to be estimated using the density of the path reconstructions, this implementation would not perform well, with a number of the true locations being missed by the reconstructed paths. Due to the constraint that the ‘observed’

Table 5.6 Posterior credible intervals (and ESS) for the movement parameters in the observation error simulation when the error variance was unknown and estimated. The ESS is taken from the 4,500 samples thinned from 4.5×10^6 iterations.

Initial σ_ε^2		2.5	50	97.5	ESS		2.5	50	97.5	ESS
25	σ_θ^2	0.686	0.855	1.081	1326	μ	17.9	21.5	25.1	2872
50		0.685	0.857	1.089	1307		17.9	21.6	24.8	3818
25	β	0.0800	0.174	0.340	325	σ_ψ^2	62.1	92.8	150	915
50		0.0870	0.179	0.334	440		64.1	92.4	143	1201
25	σ_ε^2	15.2	28.5	44.9	461					
50		15.4	28.2	43.3	429					

locations impose upon the implementation, there are fewer likely options for how the path could progress between these, and so the uncertainty of the estimated location process is low, with the reconstructions being similar to one another.

When including error, although the space use densities are more uncertain, they are able to capture the true locations within the estimated location process. The individual path reconstructions can be seen to reflect the trend in estimated bearing variance, being smoother and more persistent when observation error is incorporated in comparison to the ‘jagged’ reconstructions when error is suppressed. Comparing the implementations with error variances of 5, 25 (the true value) and 50, note the observation with high simulated error at the left of the plotted region (the probability of an error at least this large being simulated was 0.00012). When the error variance was 25, although this was the value used for the simulation, the reconstructed paths do not appear to capture this true location well, and pass symmetrically on either side of the observed location. When the error variance was much larger, the majority of reconstructed paths lie to the east of the observed location, towards the true simulated value. This highlights a limitation of the algorithm to identify the true location when the actual error lies in the extreme tail of the true error distribution.

To evaluate the performance at estimating the actual observation error of each of the 10 implementations in which error was incorporated, Fig. 5.21 shows posterior reconstructions of the true location at an individual observation time. In this first example, the time chosen had small actual error between the true simulation and the observation (the **black square** and **black point** are almost indistinguishable), with probability of an error at least this large under the simulation model of 0.88. The posterior reconstructions of the true location are given by **green points**, along with kernel density estimates highlighting the 50%, 75% and 95% quantiles as **black contours**. For comparison, the prior distribution of the true location

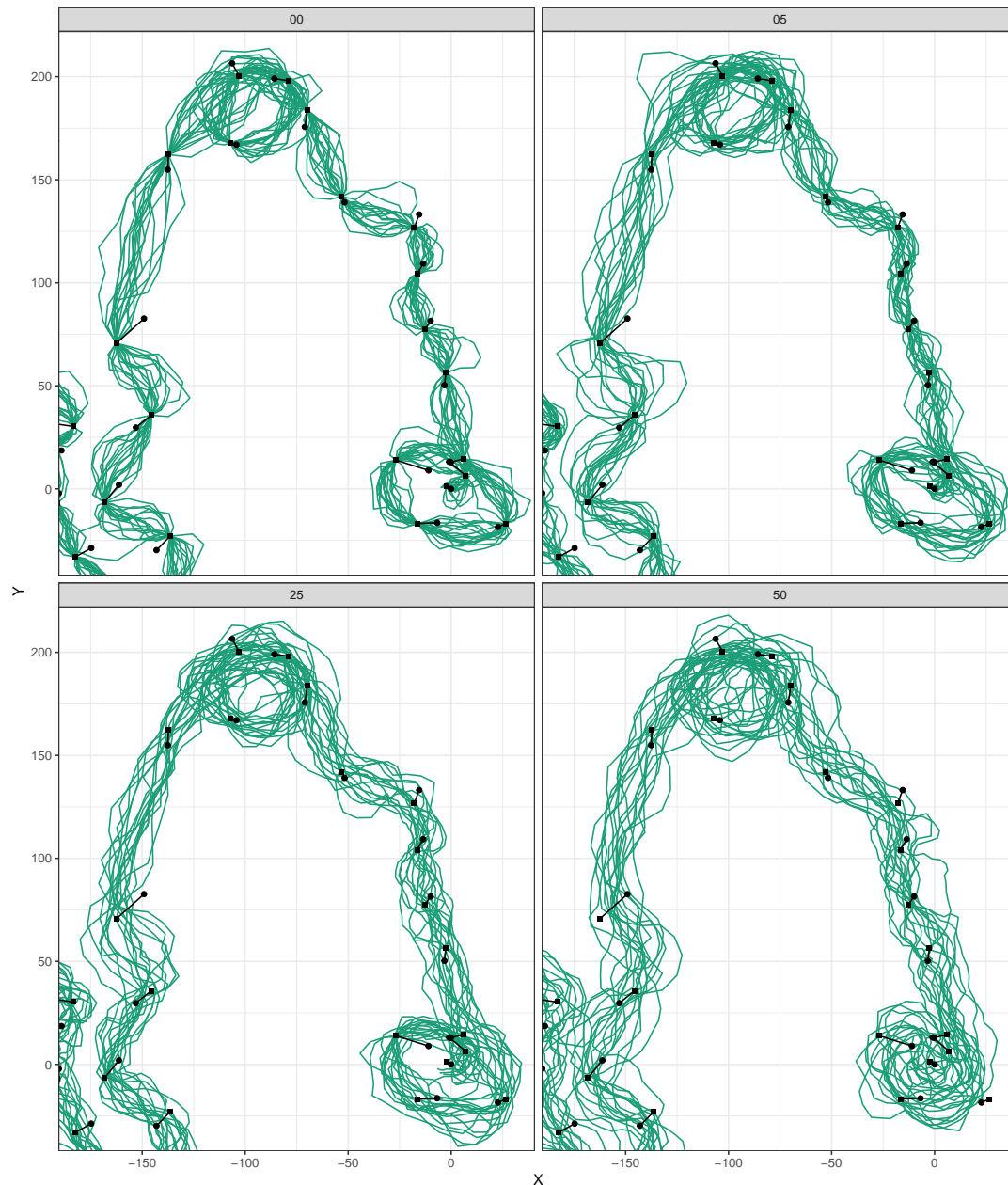


Fig. 5.20 Examples of the sampled refined path reconstructions between times 0–40 (**green lines**) in the observation error simulation example. **Panels:** reconstructions for four of the implementations: without observation error, with error variance of 5, with true error variance of 25 and with error variance of 50. The observations are shown (**square points**) along with the true locations at those observation times (**circular points**).

implied by the error model with fixed error variance in each implementation is shown by the **orange contours** at the same quantile levels. Each implementation in this case has a roughly circular/symmetric posterior distribution, closely centred around the observed location (the mean error not differing between the prior and posterior). For the cases with small error variance, the posterior distribution closely follows the prior model. For larger error variances, above 30, the posterior distribution of the true location is less dispersed than the prior model, having lower support for larger errors. Implementing a fixed error variance larger than the simulation value has made little difference to the inferred true location (for example, the 95% quantiles for the true location when $\sigma_{\epsilon}^2 = 25, 30, 35, 40$ are similar). This supports the findings above that the posterior distributions of the movement parameters were similar for these cases.

Fig. 5.22 gives the performance at estimating the actual observation error for a time point with larger error than the previous example (the probability of an error at least this large under the simulation model is 0.049). In this case, the true simulated location is not contained within the 95% quantile region for the prior model when the error variance was small (less than 20). In these implementations, the posterior distribution is centred towards the true location but is not captured well, remaining in a low density region. When the error variance was fixed at the true simulation value, the true location has been captured well, being in the highest 50% region of the posterior. This estimation differs little for the implementations with error variances larger than this. Unlike the previous example, all the implementations here show posterior estimates with similar levels of uncertainty to their model prior, but with the posterior mean differing from the prior and the correlation between the x - and y -directions being non-negligible (the quantiles are distinctly elliptical).

An example of extreme actual observation error is given in Fig. 5.23, in which the probability of simulating an error at least as large as this was 0.00012 (note that even for the implementation with the largest error variance, 50, the true location lies outside the model 95% quantile region). For reference, note that this is the observation with large error which appears towards the left of the location plots in Fig. 5.20. The implementations with error variance less than the simulation value have not been able to capture the true location, which lies outside the 95% quantile region in all cases. For the implementations with error variances equal to the simulation value (25), and higher, the true location has only been estimated well in the y -direction with large uncertainty in the x -direction (note the non-circular shape of the posterior distributions). Unlike the previous examples, for this observation time there is a noticeable difference in the posterior distributions when the error variance is larger than 25, with the estimate of the true location improving as the error variance is increased. As

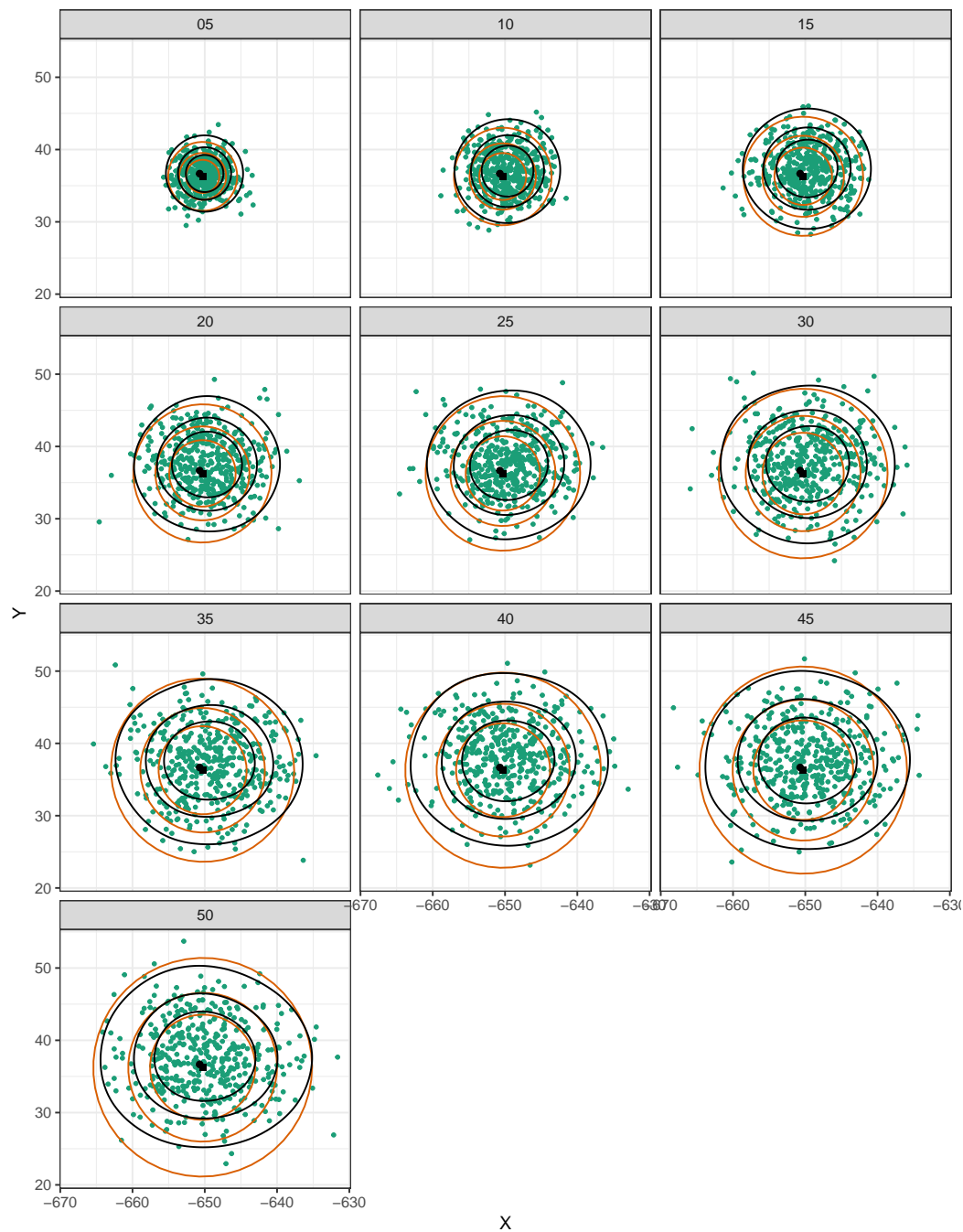


Fig. 5.21 Posterior reconstructions of the true location at an individual observation time (**green points** and kernel density estimate at 50%, 75%, 95% quantiles in **black contours**) in the observation error simulation example. **Panels:** reconstructions for the 10 implementations that included error, with the prior distribution of the true location highlighted in each case (**orange contours** also at 50%, 75%, 95% quantiles). The observation is shown (**square point**) along with the true location at that time (**circular point**). Note that the actual error at this time was so small that these points are almost indistinguishable.

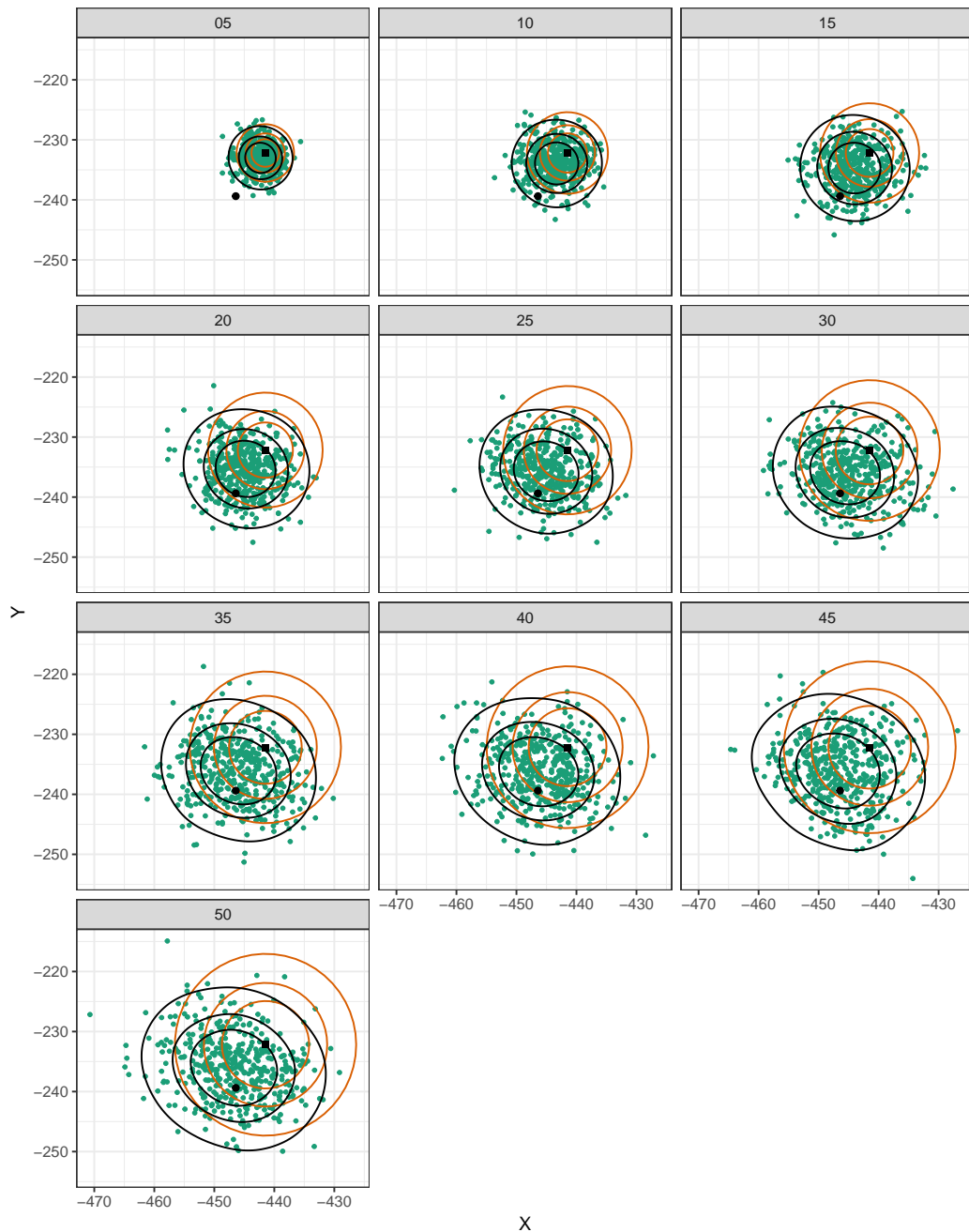


Fig. 5.22 Posterior reconstructions of the true location at an individual observation time (**green points** and kernel density estimate at 50%,75%,95% quantiles in **black contours**) in the observation error simulation example. **Panels:** reconstructions for the 10 implementations that included error, with the prior distribution of the true location highlighted in each case (**orange contours** also at 50%,75%,95% quantiles). The observation is shown (**square point**) along with the true location at that time (**circular point**). In this case the actual observation error is reasonably large.

identified earlier in the discussion of the reconstructions of Fig. 5.20, the posterior mean location in the x -direction when the error variance was 25 is close to the observed location (hence the reconstructions passed to the east and west of the observation). However, when the error variance was assumed to be larger, the posterior mean lies to the east of the observation—towards the true location.

5.6 Correlated error process

Observation errors from GPS transmitters have been found to be highly correlated when the sampling frequency is high (Breed and Severns, 2015). The model for observation error in Sect. 5.1 that assumes independence in the error process over time should ideally be applied only when the sampling interval is coarse enough that such an assumption is realistically valid (thought to be sampling intervals of 15–30 minutes and over (Breed and Severns, 2015)). The following presents an extension to this simple model that incorporates the realism of correlation in the observation error process.

The two-dimensional error process, $\boldsymbol{\epsilon}$, is assumed to follow a circular OU process with common mean 0. This process was introduced in Sect. 2.1.2, with the special circular case of the OU process assuming that the two dimensions are independent from one another (this assumption was also used in the error model of Sect. 5.1). The error at observation j , given the error at an earlier observation i , is given by

$$\boldsymbol{\epsilon}_j | \boldsymbol{\epsilon}_i \sim \text{N} \left(e^{-\kappa(t_j - t_i)} \boldsymbol{\epsilon}_i, \sigma_{\boldsymbol{\epsilon}}^2 \left(1 - e^{-\kappa(t_j - t_i)} \right) I_2 \right), \quad (5.5)$$

where I_2 is the two-dimensional identity matrix, $\kappa > 0$ is the correlation parameter and $\sigma_{\boldsymbol{\epsilon}}^2$ is the long-term variability. The observed locations are then given by Eq. 5.1, but with $\boldsymbol{\epsilon}$ now given by the correlated process described here.

5.6.1 Extending the method for fully Bayesian inference

Inference follows the same approach as that described in previous sections. The parameters are updated as in Chap. 4. The following describes the method for sampling the two error parameters from the full conditional distribution and reconstructing a section of the refined path.

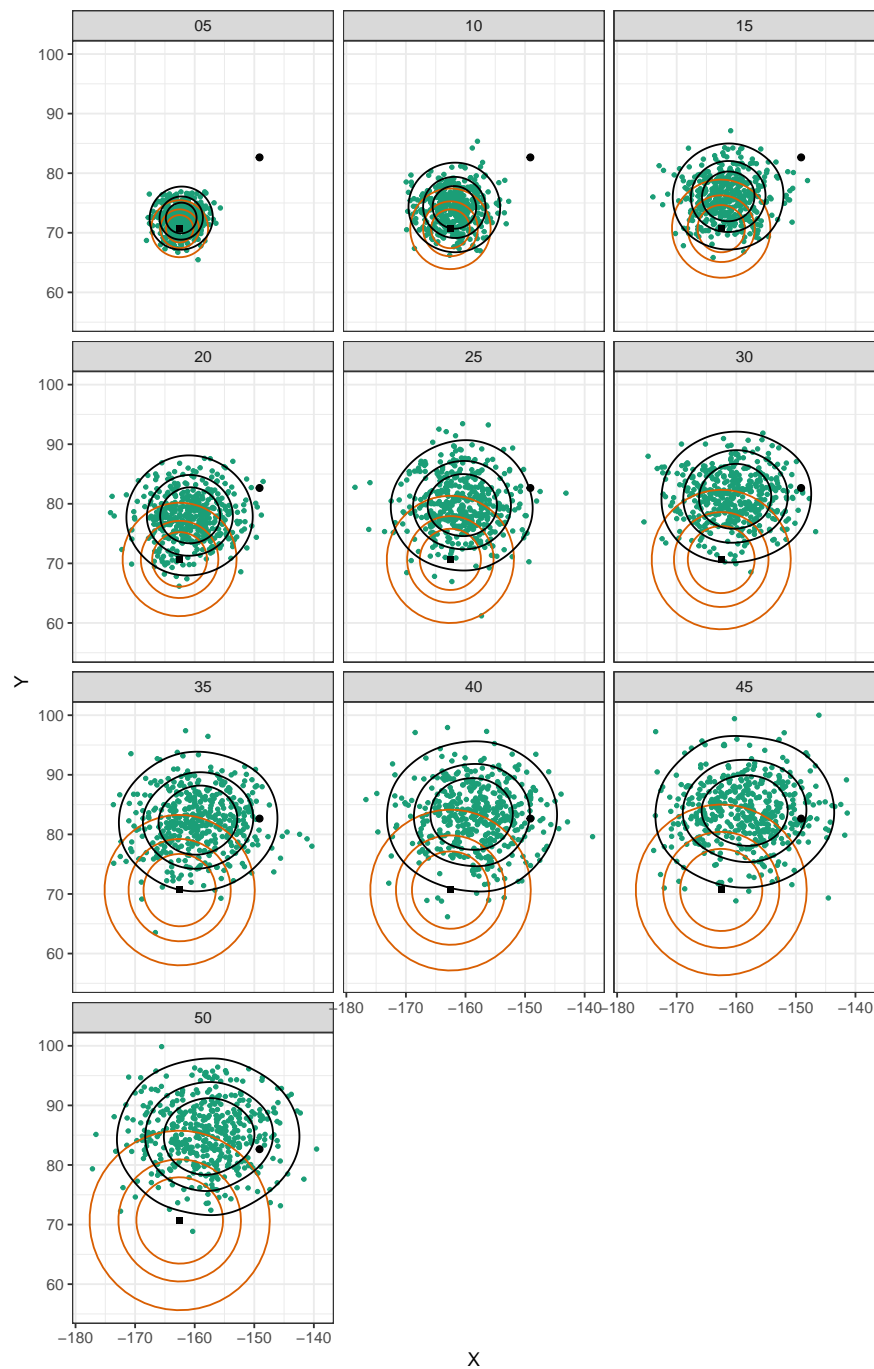


Fig. 5.23 Posterior reconstructions of the true location at an individual observation time (**green points** and kernel density estimate at 50%,75%,95% quantiles in **black contours**) in the observation error simulation example. **Panels:** reconstructions for the 10 implementations that included error, with the prior distribution of the true location highlighted in each case (**orange contours** also at 50%,75%,95% quantiles). The observation is shown (**square point**) along with the true location at that time (**circular point**). In this case the actual observation error is large.

5.6.1.1 Sampling the observation error parameters

The two parameters of the error process are updated simultaneously using a random walk MH step, with independent proposals for each parameter. As both parameters are constrained to be positive, univariate Gaussians truncated below at zero are used to generate the step in the random walk. The posterior conditional distribution, up to a constant, is

$$\begin{aligned} p\left(\kappa, \sigma_{\varepsilon}^2 \mid \Phi, \mathbf{B}, \boldsymbol{\theta}, \mathbf{v}, \mathbf{Z}^*\right) &= p\left(\kappa, \sigma_{\varepsilon}^2 \mid \boldsymbol{\theta}, \mathbf{v}, \mathbf{Z}^*\right) \\ &= p\left(\kappa, \sigma_{\varepsilon}^2 \mid \boldsymbol{\varepsilon}\right) \\ &\propto p\left(\kappa, \sigma_{\varepsilon}^2\right) p\left(\boldsymbol{\varepsilon}_1 \mid \kappa, \sigma_{\varepsilon}^2\right) \prod_{i=2}^M p\left(\boldsymbol{\varepsilon}_i \mid \boldsymbol{\varepsilon}_{i-1}, \kappa, \sigma_{\varepsilon}^2\right), \end{aligned}$$

where $p\left(\kappa, \sigma_{\varepsilon}^2\right)$ is given by an appropriate prior distribution, the errors $\boldsymbol{\varepsilon}$ are calculated as previously, the initial error is given by

$$\boldsymbol{\varepsilon}_1 \mid \kappa, \sigma_{\varepsilon}^2 \sim \mathbf{N}\left(\mathbf{0}, \sigma_{\varepsilon}^2 \mathbf{I}_2\right),$$

and $\boldsymbol{\varepsilon}_i \mid \boldsymbol{\varepsilon}_{i-1}, \kappa, \sigma_{\varepsilon}^2$ is distributed as in Eq. 5.5.

5.6.1.2 Reconstructing the unobserved refined path

In contrast to the independent case of Sect. 5.2.2, the actual observation error at the start and end of the refined path section must be taken into consideration because the process is no longer independent (this change is highlighted in Fig. 5.24, which should be compared with Fig. 5.3). To reflect this additional constraint, we now have

$$\mathcal{F} = \{\Phi, \sigma_{\varepsilon}^2, \mathbf{Z}(j), B(j), B(l), \boldsymbol{\theta}_0, \boldsymbol{\theta}_n, \mathbf{v}_0, \mathbf{v}_n, \boldsymbol{\varepsilon}_j, \boldsymbol{\varepsilon}_l\}.$$

The underlying movement process has not been changed and so the distribution of the true locations $(\mathbf{Z}(k), \mathbf{Z}(l))^{\text{T}}$ remains the same. It is only the model for the observation error that has changed, and so the relationship between the true and observed locations in Eq. 5.2 becomes

$$(\mathbf{Z}^*(k), \mathbf{Z}(l))^{\text{T}} \mid (\mathbf{Z}(k), \mathbf{Z}(l))^{\text{T}}, \mathbf{B}^*, \boldsymbol{\theta}^*, \mathcal{F} \sim \mathbf{N}\left((\mathbf{Z}(k), \mathbf{Z}(l))^{\text{T}} + \boldsymbol{\mu}_{\varepsilon}, \Sigma_{\varepsilon}\right),$$

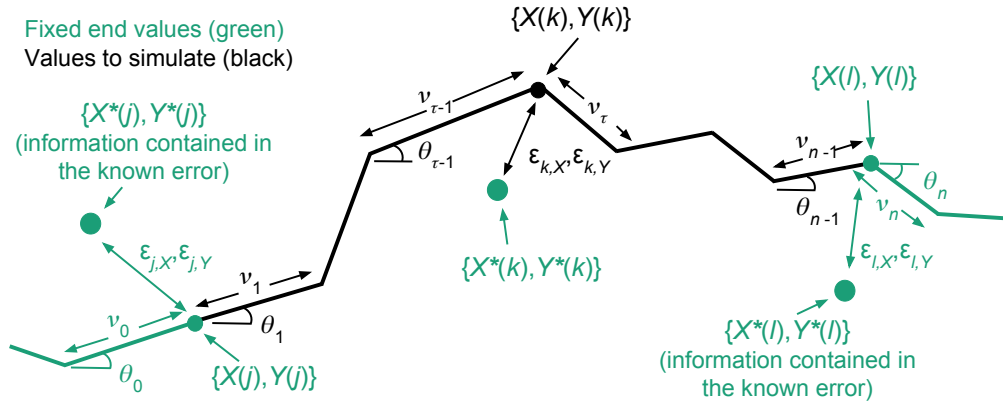


Fig. 5.24 Section of the full refined path to update over when correlated observation error is present. Fixed endpoint locations are given at the times j and l . Note that in the correlated case, both the true locations and the actual error at the endpoint times are needed (this information describes the noisy locations at these times also).

where $\boldsymbol{\mu}_\varepsilon, \Sigma_\varepsilon$ are constructed using the mean and covariance of an OU bridge between the known errors $\boldsymbol{\varepsilon}_j, \boldsymbol{\varepsilon}_l$. The distribution of the OU bridge can be constructed following the conditional Gaussian approach used earlier for the marginal step distribution, and is given by

$$\boldsymbol{\mu}_\varepsilon = \begin{pmatrix} \mathbb{E}(\varepsilon_{k,X}) \\ \mathbb{E}(\varepsilon_{k,Y}) \\ \varepsilon_{l,X} \\ \varepsilon_{l,Y} \end{pmatrix}, \quad \Sigma_\varepsilon = \begin{pmatrix} \text{Var}(\varepsilon_{k,X}) & 0 & 0 & 0 \\ 0 & \text{Var}(\varepsilon_{k,Y}) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where

$$\mathbb{E}(\varepsilon_{k,\cdot}) = \frac{\left(e^{-\kappa(k-j)} (1 - e^{-2\kappa(l-k)}) \boldsymbol{\varepsilon}_{j,\cdot} \right) + \left(e^{-\kappa(l-k)} (1 - e^{-2\kappa(k-j)}) \boldsymbol{\varepsilon}_{l,\cdot} \right)}{1 - e^{-2\kappa(l-j)}},$$

$$\text{Var}(\varepsilon_{k,\cdot}) = \sigma_\varepsilon^2 \left(1 - e^{-2\kappa(k-j)} - \frac{e^{-2\kappa(l-k)} (1 - e^{-2\kappa(k-j)})^2}{1 - e^{-2\kappa(l-j)}} \right).$$

Note that the covariance between $\boldsymbol{\varepsilon}_{k,X}, \boldsymbol{\varepsilon}_{k,Y}$ is zero because of the earlier assumption that the OU error process is circular. The proposal distribution for the steps in Eq. 5.3 then becomes

$$\begin{aligned} \mathbf{v} \mid \mathbf{B}^*, \boldsymbol{\theta}^*, \mathcal{F}, \mathbf{Z}^*(k), \mathbf{Z}(l) \\ \sim \text{N} \left(\mathbf{m}_1 + \Sigma_{1,2} (\Sigma_2 + \Sigma_\varepsilon)^{-1} \left((\mathbf{Z}^*(k), \mathbf{Z}(l))^\text{T} - \mathbf{m}_2 - \boldsymbol{\mu}_\varepsilon \right), \Sigma_1 - \Sigma_{1,2} (\Sigma_2 + \Sigma_\varepsilon)^{-1} \Sigma_{1,2}^\text{T} \right), \end{aligned} \quad (5.6)$$

and can be simulated (additional details on the validity of this are in Appendix A.2.3) as

$$\begin{aligned} \mathbf{x}^* &\sim \text{N}(\mathbf{m}_1, \Sigma_1), \\ \mathbf{y}^* &\sim \text{N} \left((\mathbf{Z}^*(k), \mathbf{Z}(l))^\text{T} - \boldsymbol{\mu}_\varepsilon, \Sigma_\varepsilon \right), \\ \mathbf{v}^* &= \mathbf{x}^* - \Sigma_{1,2} (\Sigma_2 + \Sigma_\varepsilon)^{-1} \left((\mathbf{Z}(j), \mathbf{Z}(j))^\text{T} + \mathbf{A}\mathbf{x}^* - \mathbf{y}^* \right). \end{aligned} \quad (5.7)$$

5.7 Discussion

Observation error is often lacking from current approaches for modelling movement; with the exception of SSMs (Patterson et al., 2008), the discrete-time step-and-turn models (including HMM's (Langrock et al., 2012)) do not allow for error unless the location space is discretised (Pedersen et al., 2011). In Pohle et al. (2017), it was shown that neglecting error leads to the selection of increased numbers of states, resulting in 'behavioural states' that are not ecologically relevant but merely a construct of the model having low support. Rather than accounting for this problem with a rigorous model for the error, more ad-hoc approaches are introduced to deal with the problem of increased state selection. Similarly, McClintock (2017) found that when observation error was present, models were less robust at estimating the behavioural process. This is a worrisome problem when recent reviews of approaches for identifying behavioural switches (see e.g. Gurarie et al. (2016)) do not include any approach to allow for error. Although it is highlighted as an issue (Edelhoff et al., 2016), there has been a lack of solutions provided other than simply searching for and removing outlier observations. In addition to problems with behavioural classification, Hurford (2009) showed that the presence of observation errors leads to discrete-time models identifying non-existent turning angles of around $\pm\pi$. With numerous discrete-time step-and-turn models identifying 'foraging' behavioural states with small steps lengths and mean turning angles of π , the lack of allowance for observation error should be prioritised.

This chapter has explored the introduction of errors on the observed locations, initially with a model of independent, circular errors and then an extension to this that allows for correlation in this error process. In both these cases, the errors were assumed to be Gaussian distributed for the simplicity this brings to the simulation technique used in the inference algorithm. Previous studies into observation error have argued that the errors produced by GPS devices are too dispersed for a normality assumption and have opted instead for t -distributions (Jonsen et al., 2005). It was found in the simulation example of Sect. 5.5 that large errors were difficult to estimate under the chosen model, and so could be a limitation of this assumption. However, although the model for error was simple, the posterior distribution of the true location given the errors is not necessarily a simple Gaussian. This was shown in both the reindeer example of Sect. 5.3 and the simulation in Sect. 5.5, in which the distribution of the inferred error was often non-circular and not centred on zero (i.e. the observed location). Therefore, although this presented a relatively simple model for error, this is not such a restrictive assumption and the introduction of correlated errors in Sect. 5.6 aims to introduce further realism into the model.

In the examples of Sects. 5.4 and 5.5 the problem of identifiability between the observation error variance and the movement parameters was discussed. When implementing different approximations to the underlying movement process, different levels of observation error variance, and hence speed parameters, were inferred for the gull example. In the simulation example, fixed levels of observation error variance led to different estimates of the movement parameters. Although this is the case, and the effects of this identifiability should be explored further, this should not necessarily be a hindrance to the modelling approach. The levels of observation error caused by specific GPS devices are well known, and informative prior knowledge of the error variance can be implemented and useful inferences made about the movement parameters.

Chapter 6

Discussion and further work

This thesis presents an approach for mechanistic modelling of animal movement that is defined in continuous time but aims to provide the interpretability found in discrete-time step-and-turn models. Movement models analogous to the popular models of Langrock et al. (2012); Morales et al. (2004) in which paths follow a CRW with speed that is independent over disjoint time intervals is provided, in addition to more realistic movement in which the speed process features correlation. Distinct behavioural modes are incorporated through a switching model in continuous time, with the movement process being dependent upon such behaviours. Gaussian error on observed locations is included, initially as a simple model, independent in both time and direction, which is then extended to include correlation in time.

Inference for movement, behaviour and error parameters is presented in the form of a hybrid Bayesian MCMC algorithm, augmenting the observed locations with the true trajectory (both movement and behaviour). Although the augmentation technique makes inference possible, this approach also provides comprehensible inference that aids in the interpretation of model estimates. These methods are demonstrated over a variety of simulated and real examples to showcase the ability to estimate parameter values, along with the possible ecological questions that this approach is able to address.

The following provides recommendations for improvement and future extensions and a discussion of how this work compares with current animal movement modelling.

6.1 Extending the model for behaviour

The behavioural process introduced in Chap. 4 is a CTMC on a known, finite number of states. This process is assumed to be homogeneous in time so that the transition rates between states are constant. As discussed in Sect. 4.6, methods from Blackwell et al. (2015) can be directly applied to the current methodology to allow for heterogeneous behavioural switching. This would allow the switching rate to depend on environmental covariates or, more generally, the location itself and also with time, such as periodic rates where a resting state is more likely between certain hours of the day. The current methods for inference are provided as an R package containing the necessary functions. Further work integrating the methods of this thesis with the location models of Blackwell et al. (2015) in a combined software suite would be desirable.

The HMM behaviour model of Langrock et al. (2012) discusses the use of semi-Markov models to describe a switching behavioural model in which the total time spent in a state affects the switching rate out of such state. Similarly, Fleming et al. (2014a) feature a stochastic behavioural process with slower decaying memory than an autoregressive process of order one. These models for behaviour appear more realistic than the current switching model, and so further work could be carried out regarding implementing semi-Markov behaviour in a continuous-time framework.

The methods discussed here have assumed that the number of behavioural states is known prior to analysis. Although it is thought that this number will be informed from ecological knowledge of the study species to provide interpretable inference, further work is required regarding how this number of states should be chosen. Likelihood criteria (e.g. Watanabe-Akaike (WAIC), see Gelman et al. (2013)) could be applied to compare the model fit at a number of known behavioural states, as is applied in the discrete-time literature (Langrock et al., 2012), but this has found in practise to be biased towards assigning a large number of behavioural states (Pohle et al., 2017). Alternatively, an unknown number of states could potentially be incorporated into the current approach through a reversible jump MCMC algorithm that can handle the varying numbers of parameters.

6.2 Identifiability in the presence of observation error

Chap. 5 introduced a Gaussian model for error in the observed locations, but in the examples of Sects. 5.4 and 5.5 potential problems with identifiability between movement and error

parameters were found. Although this problem is reduced by placing informative prior distributions upon the error parameter based on well-known error ranges of GPS devices, further work should be carried out into the relationship that exists between the parameters, in particular, focusing on the range of values for which these parameters can be effectively identified and estimated. For those combinations of parameters that cannot be estimated, it would be useful to identify the association that describes the lack of identifiability. Useful first steps in this direction will be to investigate known realistic scenarios for the error parameter (e.g. the error ranges identified in works such as Hurford (2009) and Breed and Severns (2015)) to identify how strong a prior distribution on this parameter must be to correctly identify the movement parameters.

In addition to investigating the link between the parameters, more research should be carried out into the effect that error has on the identification of the behavioural process. In discrete-time models, errors have been shown to make behavioural estimation difficult (McClintock, 2017; Pohle et al., 2017). In the gull example of Sect. 5.4, the difference in error variance estimation left the estimated behavioural process and parameters largely unchanged. This is likely due to the large distances that the gull was travelling between observations, with the median distance between observations being 91 metres and a 95% quantile interval being (11, 2240) metres, which resulted in the estimated mean speed in the two behavioural states being around 20 and 400 metres per minute. For the fast moving state, these distances are large in comparison with the level of observation error, estimated to be up to 11 metres. In studies where species are travelling smaller distances between observations, the observation error is likely to become more of an issue and have a larger effect on inference results. For example, the reindeer example in Sect. 5.3 had an estimated mean speed with median 26 metres per minute, and so observations taken at 2 minute intervals have a lower signal-to-noise ratio than in the gull example. If using the extended version of this dataset that displays clear heterogeneity in the movement process, it is likely that estimation of the behavioural process would be greatly affected by the level of inferred observation error. In particular, note the bursts of observations in Fig. 6.1, one of which is shown in detail in Fig. 6.2 in which the distance between successive locations is between 2–7 metres. Dependent upon the level of inferred observation error within the model, the behavioural process is likely to feature a stationary state in which the speed is zero and the distances observed are purely observation error.

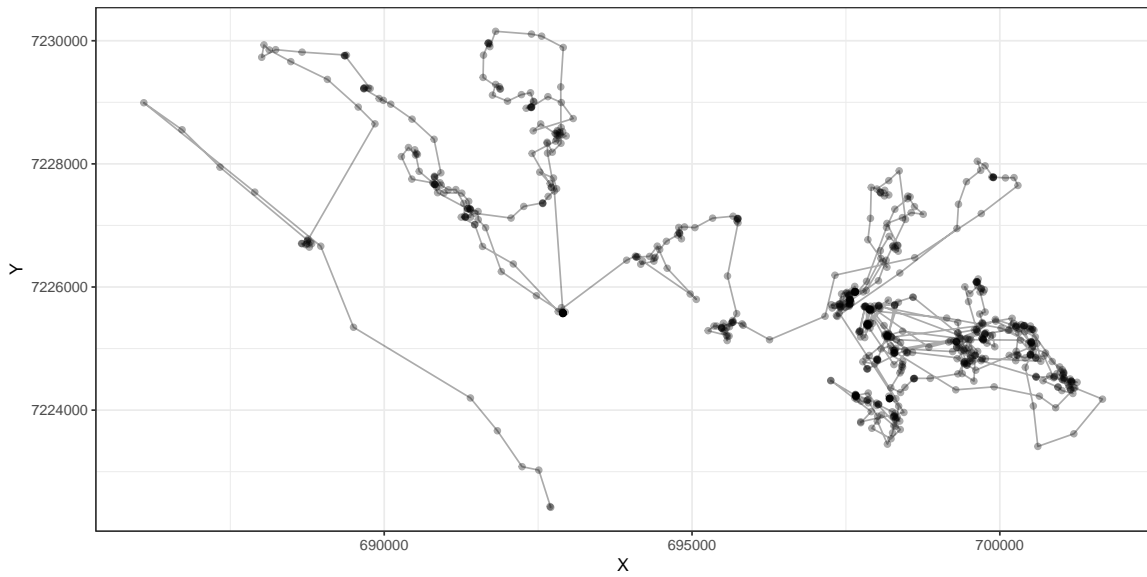


Fig. 6.1 Extended dataset of observed locations of reindeer b53.10, showing clear heterogeneity in the movement process over time, with bursts of similar successive observations.

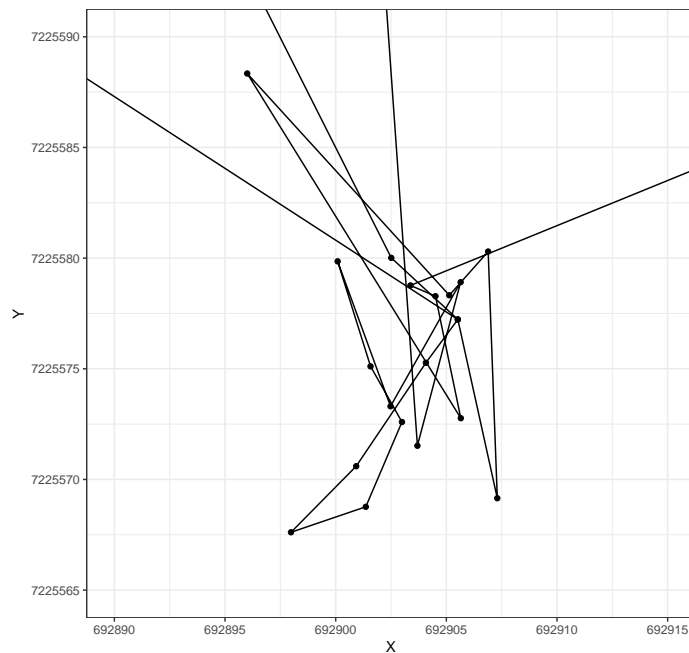


Fig. 6.2 Section of the reindeer observations of Fig. 6.1, showing a number of successive observations with distances between of 2–7 metres, where expected observation error is up to 10 m.

6.3 Efficiency of the inference algorithm

The MCMC algorithm for inference with augmentation gives interpolated location estimates at an arbitrarily fine scale. Not only does this aid in the interpretation of model parameters, but allows estimates (with robust quantification of uncertainty) of location density that can be combined with environmental covariates to learn about resource selection. The drawback of this scheme, however, is the slow speed/efficiency. Mixing of the Gibbs MCMC sampler is slow due to the (obviously) high correlation between the path reconstruction and the parameters that describe it. As the number of observations increase and the path becomes larger the high dimensionality of the reconstruction further reduces mixing and the sampler must be run for higher numbers of iterations to achieve similar ESS.

Solutions that can be explored to improve the efficiency of the current method for inference include implementing (at least parts of) the algorithm in other languages than R that have proved to be faster at specific tasks that are repeated often. Examples include inverting matrices, which is required to simulate a step proposal, that can be calculated quicker in the intel build of R or python. Other options include exploring parts of the algorithm that can be parallelised. For instance, path reconstruction proposals that are over disjoint sections of the path are independent from one another and could potentially be implemented in parallel if a saving is made despite the computational cost of ‘joining’ the path sections back together. Options to improve the mixing of the algorithm, and hence improve efficiency through fewer iterations, should also be explored. It was noted that in the two-state simulation and elk examples of Chap. 4, 20% and 18%, respectively, of the path proposals were instantly rejected at the behavioural stage within the rejection simulation. Implementing more sophisticated CTMC bridge simulations for the behavioural proposal, such as the newly released `ECctmc` (Fintzi, 2017) R package could improve path section acceptance, and hence mixing.

Rather than attempting to improve the efficiency of the current approach to inference, other methods could be explored. One option would be to formulate the model as an SSM, with the GPS locations being the observed process, and the joint bearing and speed being the underlying process. The observed process is non-linear, and so a Kalman filter is not possible, but a potential approach would be to use particle filtering MCMC (PMCMC, see Andrieu et al. (2010) for further details). Inference would still follow an augmentation approach, using Gibbs sampling to alternately update the parameters and the unobserved path, but with a new method for sampling the path than that used here. Sequential Monte Carlo methods are employed (a generalisation of the Kalman filter) to simulate path reconstructions

by passing over each observation in turn. This method for reconstruction would be more computationally demanding than the approach we have used, but would avoid the slow mixing brought about by only sampling small sections of the path at a time, and we would expect a higher acceptance rate.

Part of the approach to inference in the multistate case involves augmenting the unobserved, complete behavioural process. This allows the parameters corresponding to each state to be separated and easily sampled as part of the MCMC inference algorithm. However, this is a costly procedure that involves slow mixing and so methods could be implemented to avoid this by integrating out the behavioural process, similar to an HMM. Such an approach is currently being developed in regards the model of Blackwell et al. (2015) (Blackwell 2017, personal communications). In Blackwell et al. (2015) the unobserved behavioural process is reconstructed using a Poisson process of potential switches, that is then thinned to give the actual switch times. Taking the realisation of potential switches formulates the model as an HMM; there are fixed times where behavioural switches can occur and the probabilities of switching between pairs of states at each of these times can be calculated. Utilising this formulation, an overall likelihood for the observed locations can be calculated, rather than a combination of the separated behaviours. As with the PMCMC approach, the calculation of the likelihood will be more complex than the current approach but is likely to give an improvement in efficiency overall because there is no issue of mixing in the behavioural state sequence (or instant rejections through the proposed behavioural section not agreeing with the fixed end state).

6.4 Comparison to discrete-time step-and-turn

The movement model developed over Chaps. 3–5 is based in continuous time, but with parameters and movement patterns based on the discrete-time step-and-turn models such as Morales et al. (2004). The difference in time formulation allows behavioural switches to occur continuously in time rather than merely at observation times. When the sampling scheme is of no biological importance to the animal, as in the 3 hour observations of the bison example in Langrock et al. (2012), the relevance of the behavioural switching model is uncertain. In the elk example of Sect. 4.5, although there were general similarities between the behavioural processes estimated by the continuous- and discrete-time models, clear differences in residency times were found (see Fig. 4.21). A number of periods of time with less than 0.5 probability of being in behavioural state 1 in the discrete-time case were classified with high certainty (over 0.95) of being in state 1 in continuous-time. Although

both behavioural models are based on Markov chains, this demonstrates the difference in time formulation. Work quantifying the relationship between the two models would be useful for future comparisons of real data, including the link between estimated discrete-time parameters when applied to continuous-time movement paths with varying characteristics, including time-varying state-switching, as mentioned in Sect. 6.1, and how the estimated behavioural process is affected by these movement changes.

Many discrete-time examples use movement tracks from multiple animals, either by pooling this data and estimating a single, shared set of movement parameters between all individuals, or implementing a hierarchical model in which individual's movement characteristics may differ from one another. All of the examples that have been presented here include only a single individual's movement, however, the extension to multiple animals that has been applied in discrete time can be equally applied to the continuous-time setting. The limiting factor in this case would be the computational feasibility, however, because each individual's unknown movement path can be updated within the inference algorithm independently from one another (and therefore in parallel), with only the movement parameter information needing to be shared across individuals, it is expected that the computational increase would not make this extension impractical.

In a number of discrete-time step-and-turn applications, a behavioural state with mean turning angle of around $\pm\pi$ is determined. The significance of π turning angles is discussed in Sect. 3.6, highlighting the concern that this is not a feature of the movement process itself, but a construct of either the sampling scheme and/or observation error. For instance, for the elk data in Sect. 4.5 the HMM identifies a behavioural state with a mean turning angle of -3 (Michelot et al., 2016), and the Morales et al. (2004) model estimates a mean of 3 . In the continuous-time application to this data the corresponding behavioural state has a turning angle distribution that is uniform at the observation time scale. We believe that this is a more realistic interpretation of movement in an encamped behaviour. Similarly, in the reindeer example of Sect. 5.3 a discrete-time analysis would describe tortuous movement due to a number of observed turning angles of $\pm\pi$. In the continuous-time approach that made allowance for realistic levels of observation error, the inferred movement path had high persistence, indicating the ' π turns' as an erroneous consequence of the unaccounted error (as suggested in Hurford (2009)). Further work should be carried out to both demonstrate the occurrence of this effect and quantify the conditions under which it occurs. This would involve movement paths simulated in continuous-time with turning angles centred on zero, that when sub-sampled and with error applied lead to turning angles of π in a discrete-time analysis. Although similar analyses have been carried out that identified the bias in

estimation of step lengths (Rowcliffe et al., 2012) and correlation in turning angles (Nams, 2013) when continuous-time paths are discretised, we are not aware of an empirical study into the presence of these large turning angles.

The detailed section of reindeer observations in Fig. 6.2 suggests a behavioural bout that is stationary in location, with observation error accounting for the ‘observed’ movement. Stationary movement is not addressed in the discrete-time step-and-turn models of Langrock et al. (2012); Morales et al. (2004), with the distribution describing the step length only being defined for positive values. In the R package *MoveHMM* (Michelot et al., 2016), this issue is addressed by incorporating a zero-inflated step length distribution, however, this is only implemented if there are actual observed step lengths of zero as observation error is not incorporated in that model. In contrast, the continuous-time model developed here uses a speed distribution that allows for step lengths of zero. Future work modelling data of the type seen in Fig. 6.2 could include prior specification of a stationary behavioural state (the times at which the animal is in such a state are, of course, still unknown and estimated by the model) in which the speed process is fixed at zero, and the bearing distribution uniform. This essentially says that we learn nothing about the bearing of the animal when it is in the stationary state, other than its bearing once it switches into another state, which would be determined by the subsequent observations. This restricted case could be implemented without major modification to the current code for implementing inference.

The model described here assumes that movement follows a CRW, and so the change in bearing is centred at zero. BRW, such as that included in McClintock et al. (2012), and attracted walks, such as that in Blackwell et al. (2015), will be more appropriate assumptions in a variety of applications including migratory species. For instance, the detailed section of gull data (from Sect. 5.4) in Fig. 6.3 shows multiple trips to the same location, indicating that the animal has specifically navigated to such a location. Simulating movement under the continuous-time model that is biased towards a specified location is a simple extension, however, extending the inference method to account for this would require additional work. The mean of the bearing process would be dependent upon the combination of the attractive location and the current location, which would complicate the current simulation method of proposing bearings without dependence upon the observed locations. Development of an extension to the current inference algorithm that would allow this movement behaviour could be a focus for future work, with initial methods following the formulation for directed movement included in Michelot et al. (2017), in which restrictions are placed upon the possible sequences of behavioural states that can be performed and the attractive locations are known.

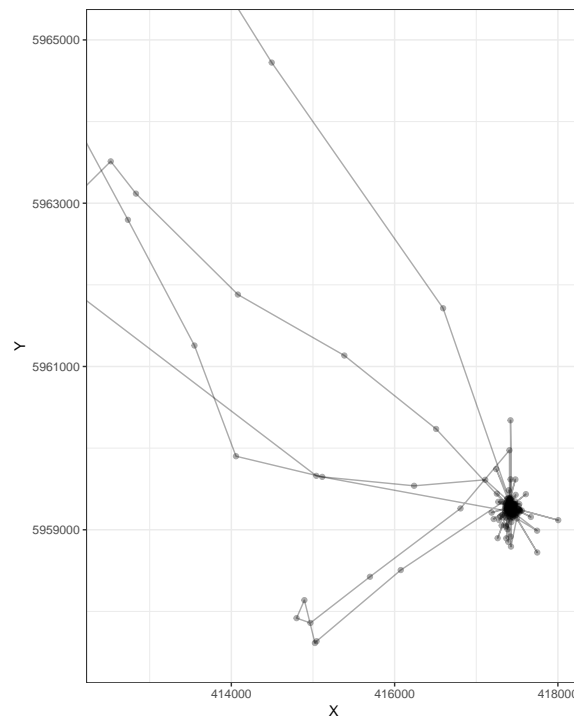


Fig. 6.3 Section of the gull observations of Fig. 5.10, showing a number of journeys returning to the same locations, suggesting directed/attractive movement may be a more appropriate assumption.

One of the main advantages of modelling in continuous time is the ability to handle irregular and missing observations with ease, unlike the restrictions placed upon discrete-time models. In the gull example of Sect. 5.4 the sampling period varied between 3.05–7.77 minutes. This presented no difficulty for the continuous-time model, in which the refined time scale chosen for reconstruction need not be regular, and the method to choose the refined times was easily automated. In a discrete-time analysis of this data, a linear interpolation of the locations on to a regular time grid would need to be implemented before calculating the observed turning angles and step lengths. Although the assumption this implies of constant speed between locations can be justified, the effect this has on the implied turning angle is not well understood and certainly not ecologically justified. Similarly, in the elk example of Sect. 4.5 all analyses treated the observation as daily despite the daily movement rate included alongside the observed locations suggesting that observations were mostly between 12–36 hours but included a gap spanning multiple days (our analysis also assumed observations were daily as these rates were not given with enough accuracy to calculate observation times effectively). Although an HMM could account for a gap of missing information if the process causing this was random, the information on the displacement

observed over the gappy data is discarded. This is a disadvantage, particularly in datasets with many missing observations, and does not occur in the continuous-time approach to modelling we have presented here.

References

- Albertsen, C. M., Whoriskey, K., Yurkowski, D., Nielsen, A. and Flemming, J. M. (2015). Fast fitting of non-Gaussian state-space models to animal movement data via Template Model Builder, *Ecology*, **96** (10): 2598–604.
- Anderson-Sprecher, R. C. and Ledolter, J. (1991). State-space analysis of wildlife telemetry data, *Journal of the American Statistical Association*, **86** (415): 596–602.
- Andrieu, C., Doucet, A. and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods, *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **72** (3): 269–342.
- Bagniewska, J. M., Hart, T., Harrington, L. A. and MacDonald, D. W. (2013). Hidden Markov analysis describes dive patterns in semiaquatic animals, *Behavioral Ecology*, **24** (3): 659–67.
- Barraquand, F. and Benhamou, S. (2008). Animal movements in heterogenous landscapes: Identifying profitable places and homogenous movement bouts, *Ecology*, **89** (12): 3336–48.
- Beyer, H. L., Morales, J. M., Murray, D. and Fortin, M. J. (2013). The effectiveness of Bayesian state-space models for estimating behavioural states from movement paths, *Methods in Ecology and Evolution*, **4** (5): 433–41.
- Blackwell, P. G. (1997). Random diffusion models for animal movement, *Ecological Modelling*, **100**: 87–102.
- Blackwell, P. G. (2003). Bayesian inference for Markov processes with diffusion and discrete components, *Biometrika*, **90** (3): 613–27.
- Blackwell, P. G., Niu, M., Lambert, M. S. and LaPoint, S. D. (2015). Exact Bayesian inference for animal movement in continuous time, *Methods in Ecology and Evolution*.
- Börger, L., Dalziel, B. D. and Fryxell, J. M. (2008). Are there general mechanisms of animal home range behaviour? A review and prospects for future research, *Ecology Letters*, **11**: 637–50.
- Bovet, P. and Benhamou, S. (1988). Spatial analysis of animals' movements using a correlated random walk model, *Journal of Theoretical Biology*, **131** (4): 419–33.

- Boyce, M. S., Pitt, J., Northrup, J. M., Morehouse, A. T., Knopff, K. H., Cristescu, B. and Stenhouse, G. B. (2010). Temporal autocorrelation functions for movement rates from global positioning system radiotelemetry data, *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **365** (1550): 2213–9.
- Breed, G. and Severns, P. (2015). Low relative error in consumer-grade GPS units make them ideal for measuring small-scale animal movement patterns, *PeerJ*, **3**: e1205.
- Breed, G. A., Costa, D. P., Jonsen, I. D., Robinson, P. W. and Mills-Flemming, J. (2012). State-space methods for more completely capturing behavioral dynamics from animal tracks, *Ecological Modelling*, **235**: 49–58.
- Brillinger, D. R., Preisler, H. K., Ager, A. A., Kie, J. G. and Stewart, B. S. (2002). Employing stochastic differential equations to model wildlife motion, *Bulletin Brazilian Mathematical Society*, **33** (3): 385–408.
- Brillinger, D. R. and Stewart, B. S. (1998). Elephant-seal movements: modelling migration, *The Canadian Journal of Statistics*, **26** (3): 431–43.
- Brost, B. M., Hooten, M. B., Hanks, E. M. and Small, R. J. (2015). Animal movement constraints improve resource selection inference in the presence of telemetry error, *Ecology*, **96** (10): 2590–7.
- Buderman, F. E., Hooten, M. B., Ivan, J. S. and Shenk, T. M. (2016). A functional model for characterizing long-distance movement behaviour, *Methods in Ecology and Evolution*, **7**: 264–73.
- Byrne, R. W., Noser, R., Bates, L. A. and Jupp, P. E. (2009). How did they get here from there? detecting changes of direction in terrestrial ranging, *Animal Behaviour*, **77**: 619–31.
- Cagnacci, F., Boitani, L., Powell, R. A. and Boyce, M. S. (2010). Animal ecology meets GPS-based radiotelemetry: A perfect storm of opportunities and challenges, *Philosophical Transactions of the Royal Society B-Biological Sciences*, **365** (1550): 2157–62.
- Calenge, C., Dray, S. and Royer-Carenzi, M. (2009). The concept of animals' trajectories from a data analysis perspective, *Ecological Informatics*, **4**: 34–41.
- Codling, E. A. and Hill, N. A. (2005). Sampling rate effects on measurements of correlated and biased random walks, *Journal of Theoretical Biology*, **233** (4): 573–88.
- Cooke, S. J., Hinch, S. G., Wikelski, M., Andrews, R. D., J, K. L., Wolcott, T. G. and Butler, P. J. (2004). Biotelemetry: a mechanistic approach to ecology, *Trends in Ecology and Evolution*, **19** (6): 334–43.
- Cooke, S. J., Midwood, J. D., Thiem, J. D., Klimley, P., Lucas, M. C., Thorstad, E. B., Eiler, J., Holbrook, C. and Ebner, B. C. (2013). Tracking animals in freshwater with electronic tags: past, present and future, *Animal Biotelemetry*, **1** (1): 1–5.
- Cushman, S. A. (2010). Animal movement data: GPS telemetry, autocorrelation and the need for path-level analysis, in *Spatial Complexity, Informatics, and Wildlife Conservation*, pp. 131–149, Tokyo, Japan: Springer.

- Dunn, J. E. and Gipson, P. S. (1977). Analysis of radio telemetry data in studies of home range, *Biometrics*, **33** (1): 85–101.
- Eaton, M. L. (2007). *Multivariate statistics: A vector space approach*, vol. 53 of *Lecture Notes—Monograph Series*, Beachwood, OH: Institute of Mathematical Statistics, second edn., ISBN 978094060069.
- Edelhoff, H., Signer, J. and Balkenhol, N. (2016). Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns, *Movement Ecology*, **4** (21): 1–21.
- Einicke, G. A. and White, L. B. (1999). Robust extended kalman filtering, *IEEE Transactions on Signal Processing*, **47** (9): 2596–9.
- Fauchald, P. and Tveraa, T. (2003). Using first-passage time in the analysis of area-restricted search and habitat selection, *Ecology*, **84** (2): 282–8.
- Fieberg, J., Matthiopoulos, J., Hebblewhite, M., Boyce, M. S. and Frair, J. L. (2010). Correlation and studies of habitat selection: problem, red herring or opportunity?, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365** (1550): 2233–44.
- Fintzi, J. (2017). *ECctmc: Simulation from endpoint-conditioned continuous-time Markov chains*, R package version 0.2.4, <https://CRAN.R-project.org/package=ECctmc>.
- Fleming, C. H., Calabrese, J. M., Mueller, T., Olson, K. A., Leimgruber, P. and Fagan, W. F. (2014a). From fine-scale foraging to home ranges: A semivariance approach to identifying movement modes across spatiotemporal scales, *The American Naturalist*, **183** (5): 154–67.
- Fleming, C. H., Calabrese, J. M., Mueller, T., Olson, K. A., Leimgruber, P. and Fagan, W. F. (2014b). Non-Markovian maximum likelihood estimation of autocorrelated movement processes, *Methods in Ecology and Evolution*, **5**: 462–72.
- Fleming, C. H., Fagan, W. F., Olson, K. A., Leimgruber, P. and Calabrese, J. M. (2016). Estimating where and how animals travel: an optimal framework for path reconstruction from autocorrelated tracking data, *Ecology*, **97** (3): 576–82.
- Fleming, C. H., Sheldon, D., Gurarie, E., Fagan, W. F., Lapoint, S. and Calabrese, J. M. (2017). Kálmán filters for continuous-time movement models, *Ecological Informatics*, **40**: 8–21.
- Forester, J. D., Ives, A. R., Turner, M. G., Anderson, D. P., Fortin, D., Beyer, H. L., Smith, D. W. and Boyce, M. S. (2007). State-space models link elk movement patterns to landscape characteristics in Yellowstone National Park, *Ecological Monographs*, **77** (2): 285–99.
- Frair, J. L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, J. and Pedrotti, L. (2010). Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365** (1550): 2187–200.

- Franke, A., Caelli, T. and Hudson, R. J. (2004). Analysis of movements and behavior of caribou (*Rangifer tarandus*) using hidden Markov models, *Ecological Modelling*, **173** (2-3): 259–70.
- Garthe, S., Schwemmer, P., Paiva, V. H., Corman, A., Fock, H. O., Voigt, C. C. and Adler, S. (2016a). Data from: Terrestrial and marine foraging strategies of an opportunistic seabird species breeding in the Wadden Sea, *Movebank Data Repository*.
- Garthe, S., Schwemmer, P., Paiva, V. H., Corman, A., Fock, H. O., Voigt, C. C. and Adler, S. (2016b). Terrestrial and marine foraging strategies of an opportunistic seabird species breeding in the Wadden Sea, *PLoS ONE*, **11** (8): e0159630.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian Data Analysis*, Texts in Statistical Science Series, Boca Raton, FL: CRC Press, third edn.
- Gurarie, E., Andrews, R. D. and Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data, *Ecology Letters*, **12** (5): 395–408.
- Gurarie, E., Bracis, C., Delgado, M., Meckley, T. D., Kojola, I. and Wagner, C. M. (2016). What is the animal doing? tools for exploring behavioural structure in animal movements, *Journal of Animal Ecology*, **85** (1): 69–84.
- Gurarie, E., Cagnacci, F., Peters, W., Fleming, C. H., Calabrese, J. M., Mueller, T. and Fagan, W. F. (2017). A framework for modelling range shifts and migrations: asking when, whither, whether and will it return, *Journal of Animal Ecology*, **86** (4): 943–59.
- Gurarie, E. and Ovaskainen, O. (2011). Characteristic spatial and temporal scales unify models of animal movement, *The American Naturalist*, **178** (1): 113–23.
- Guttorp, P. (1995). *Stochastic modeling of scientific data*, Stochastic Modeling Series, Boca Raton, FL: Chapman and Hall/CRC.
- Hanks, E. M., Hooten, M. B. and Alldredge, M. W. (2015). Continuous-time discrete-space models for animal movement, *The Annals of Applied Statistics*, **9** (1): 145–65.
- Hanks, E. M., Hooten, M. B., Johnson, D. S. and Sterling, J. T. (2011). Velocity-based movement modeling for individual and population level inference, *PLoS ONE*, **6** (8).
- Harris, K. J. and Blackwell, P. G. (2013). Flexible continuous-time modelling for heterogeneous animal movement, *Ecological Modelling*, **255**: 29–37.
- Harvey, A. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.
- Hebblewhite, M. and Haydon, D. T. (2010). Distinguishing technology from biology: A critical review of the use of GPS telemetry data in ecology, *Philosophical Transactions of the Royal Society B-Biological Sciences*, **365** (1550): 2303–12.
- Hefley, T. J., Broms, K. M., Brost, B. M., Buderman, F. E., Kay, S. L., Scharf, H. R., Tipton, J. R., Williams, P. J. and Hooten, M. B. (2017). The basis function approach for modeling autocorrelation in ecological data, *Ecology*, **98** (3): 632–46.

- Heidelberger, P. and Welch, P. D. (1983). Simulation run length control in the presence of an initial transient, *Operations Research*, **31** (6): 1109–44.
- Hobolth, A. and Stone, E. A. (2009). Simulation from endpoint-conditioned, continuous-time Markov chains on a finite state space, with applications to molecular evolution, *Annals of Applied Statistics*, **3** (3): 1204–31.
- Holyoak, M., Casagrandi, R., Nathan, R., Revilla, E. and Spiegel, O. (2008). Trends and missing parts in the study of movement ecology, *Proceedings of the National Academy of Sciences U S A*, **105** (49): 19060–5.
- Hooten, M. B. and Johnson, D. S. (2017). Basis function models for animal movement, *Journal of the American Statistical Association*.
- Horne, J. S., Garton, E. O., Krone, S. M. and Lewis, J. S. (2007). Analyzing animal movements using Brownian bridges, *Ecology*, **88** (9): 2354–63.
- Hurford, A. (2009). GPS measurement error gives rise to spurious 180 degrees turning angles and strong directional biases in animal movement data, *PLoS ONE*, **4** (5).
- Iacus, S. M. (2008). *Simulation and inference for stochastic differential equations*, Springer Series in Statistics, New York, NY: Springer.
- Jennrich, R. I. and Turner, F. B. (1969). Measurement of non-circular home range, *Journal of Theoretical Biology*, **22**: 227–37.
- Johnson, D. S., London, J. M. and Kuhn, C. E. (2011). Bayesian inference for animal space use and other movement metrics, *Journal of Agricultural, Biological and Environmental Statistics*, **16** (3): 357–70.
- Johnson, D. S., London, J. M., Lea, M. and Durban, J. W. (2008a). Continuous-time correlated random walk model for animal telemetry data, *Ecology*, **89** (5): 1208–15.
- Johnson, D. S., Thomas, D. L., Hoef, J. M. V. and Christ, A. (2008b). A general framework for the analysis of animal resource selection from telemetry data, *Biometrics*, **64** (3): 968–76.
- Jonsen, I. D., Basson, M., Bestley, S., Bravington, M. V., Patterson, T. A., Pedersen, M. W., Thomson, R., Thygesen, U. H. and Wotherspoon, S. J. (2013). State-space models for bio-loggers: A methodological road map, *Deep-Sea Research Part II*, **88–9**: 34–46.
- Jonsen, I. D., Flemming, J. A. and Myers, R. A. (2005). Robust state-space modeling of animal movement data, *Ecology*, **86** (11): 2874–80.
- Jonsen, I. D., Myers, R. A. and Flemming, J. A. (2003). Robust state-space modeling of animal movement data, *Ecology*, **84** (11): 3055–63.
- Julier, S. J. and Uhlmann, J. K. (1997). New extension of the kalman filter to nonlinear systems, *Proceedings of the SPIE*, **3068**: 182–93.
- Kareiva, P. M. and Shigesada, N. (1983). Analyzing insect movement as a correlated random walk, *Oecologia*, **56** (2): 234–8.

- Kranstauber, B., Kays, R., LaPoint, S. D., Wikelski, M. and Safi, K. (2012). A dynamic Brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement, *Journal of Animal Ecology*, **81**: 738–46.
- Kranstauber, B., Safi, K. and Bartumeus, F. (2014). Bivariate Gaussian bridges: directional factorization of diffusion in Brownian bridge models, *Movement Ecology*, **2** (5): 1–10.
- Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D. and Morales, J. M. (2012). Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions, *Ecology*, **93** (11): 2336–42.
- Lavielle, M. (2005). Using penalized contrasts for the change-point problem, *Signal Processing*, **85** (8): 1501–10.
- Leos-Barajas, V., Photopoulou, T., Langrock, R., Patterson, T. A., Watanabe, Y. Y., Murgatroyd, M. and Papastamatiou, Y. P. (2017). Analysis of animal accelerometer data using hidden Markov models, *Methods in Ecology and Evolution*, **8** (2): 161–73.
- Li, M. and Bolker, B. M. (2017). Incorporating periodic variability in hidden Markov models for animal movement, *Movement Ecology*, **5** (1).
- Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility, *Statistics and Computing*, **10**: 325–37.
- Madon, B. and Hingrat, Y. (2014). Deciphering behavioral changes in animal movement with a “multiple change point algorithm- classification tree” framework, *Frontiers in Ecology and Evolution*, **2** (30): 1–9.
- Marsh, L. M. and Jones, R. E. (1988). The form and consequences of random walk movement models, *Journal of Theoretical Biology*, **133** (1): 113–31.
- McClintock, B. T. (2017). Incorporating telemetry error into hidden Markov models of animal movement using multiple imputation, *Journal of Agricultural, Biological and Environmental Statistics*.
- McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M. and Morales, J. M. (2014). When to be discrete: The importance of time formulation in understanding animal movement, *Movement Ecology*, **2** (1): 1–21.
- McClintock, B. T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B. J. and Morales, J. M. (2012). A general discrete-time modeling framework for animal movement using multistate random walks, *Ecological Monographs*, **82** (3): 335–49.
- McClintock, B. T., London, J. M., Cameron, M. F. and Boveng, P. L. (2015). Modelling animal movement using the Argos satellite telemetry location error ellipse, *Methods in Ecology and Evolution*, **6**: 266–77.
- McKellar, A. E., Langrock, R., Walters, J. R. and Kesler, D. C. (2015). Using mixed hidden Markov models to examine behavioral states in a cooperatively breeding bird, *Behavioral Ecology*, **26** (1): 148–57.

- Michelot, T., Langrock, R., Bestley, S., Jonsen, I. D., Photopoulou, T. and Patterson, T. A. (2017). Estimation and simulation of foraging trips in land-based marine predators, *Ecology*, **98** (7): 1932–44.
- Michelot, T., Langrock, R. and Patterson, T. A. (2016). moveHMM: an R package for the statistical modelling of animal movement data using hidden Markov models, *Methods in Ecology and Evolution*, **7**: 1308–15.
- Morales, J. M. and Ellner, S. P. (2002). Scaling up animal movements in heterogeneous landscapes: The importance of behavior, *Ecology*, **83** (8): 2240–7.
- Morales, J. M., Haydon, D. T., Frair, J., Holsinger, K. E. and Fryxell, J. M. (2004). Extracting more out of relocation data: building movement models as mixtures of random walks, *Ecology*, **85** (9): 2436–45.
- Nakagawa, S. and Freckleton, R. P. (2008). Missing inaction: the dangers of ignoring missing data, *Trends in Ecology and Evolution*, **23** (11): 592–6.
- Nams, V. O. (2013). Sampling animal movement paths causes turn autocorrelation, *Acta Biotheoretica*, **61** (2): 269–84.
- Nams, V. O. (2014). Combining animal movements and behavioural data to detect behavioural states, *Ecology Letters*, **17** (10): 1228–1237.
- Nations, C. S. and Anderson-Sprecher, R. C. (2006). Estimation of animal location from radio telemetry data with temporal dependencies, *Journal of Agricultural, Biological, and Environmental Statistics*, **11** (1): 87–105.
- Ovaskainen, O. (2004). Habitat-specific movement parameters estimated using mark-recapture data and a diffusion model, *Ecology*, **85** (1): 242–257.
- Ovaskainen, O., Rekola, H., Meyke, E. and Arjas, E. (2008). Bayesian methods for analyzing movements in heterogeneous landscapes from mark-recapture data, *Ecology*, **89** (2): 542–54.
- Parton, A. and Blackwell, P. G. (2017). Bayesian inference for multistate ‘step and turn’ animal movement in continuous time, *Journal of Agricultural, Biological and Environmental Statistics*, **22** (3): 373–92.
- Parton, A., Blackwell, P. G. and Skarin, A. (2017). Bayesian inference for continuous-time animal movement based on steps and turns, in *Bayesian Statistics in Action*, edited by Argiento, R., Lanzarone, E., Villalobos, I. A. and Mattei, A., vol. 194 of *Springer Proceedings in Mathematics & Statistics*, Springer International Publishing, arXiv:1608.05583.
- Patterson, T. A., Basson, M., Bravington, M. V. and Gunn, J. S. (2009). Classifying movement behaviour in relation to environmental conditions using hidden Markov models, *The Journal of Animal Ecology*, **78** (6): 1113–23.
- Patterson, T. A., McConnell, B. J., Fedak, M. A., Bravington, M. V. and Hindell, M. A. (2010). Using GPS data to evaluate the accuracy of state-space methods for correction of Argos satellite telemetry error, *Ecology*, **91** (1): 273–85.

- Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L. and King, R. (2017). Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges, *Advances in Statistical Analysis*, **101** (4): 399–438.
- Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O. and Matthiopoulos, J. (2008). State-space models of individual animal movement, *Trends in Ecology and Evolution*, **23** (2): 87–94.
- Pedersen, M. W., Patterson, T. A., Thygesen, U. H. and Madsen, H. (2011). Estimating animal behavior and residency from movement data, *Oikos*, **120**: 1281–90.
- Plummer, M., Best, N., Cowles, K., Vines, K., Sarkar, D., Bates, D., Almond, R. and Magnusson, A. (2016). *coda: Output Analysis and Diagnostics for MCMC*, R package version 0.19-1, <https://CRAN.R-project.org/package=coda>.
- Pohle, J., Langrock, R., van Beest, F. M. and Schmidt, N. M. (2017). Selecting the number of states in hidden Markov models: Pragmatic solutions illustrated using animal movement, *Journal of Agricultural, Biological, and Environmental Statistics*.
- Postlethwaite, C. M., Brown, P. and Dennis, T. E. (2013). A new multi-scale measure for analysing animal movement data, *Journal of Theoretical Biology*, **317**: 175–85.
- Preisler, H. K., Ager, A. A., Johnson, B. K. and Kie, J. G. (2004). Modeling animal movements using stochastic differential equations, *Environmetrics*, **15** (7): 643–57.
- Preisler, H. K., Ager, A. A. and Wisdom, M. J. (2013). Analyzing animal movement patterns using potential functions, *Ecosphere*, **4** (3): 1–13.
- Pyke, G. H. (2015). Understanding movements of organisms: it's time to abandon the levy foraging hypothesis, *Methods in Ecology and Evolution*, **6**: 1–16.
- Rao, V. and Teh, Y. W. (2013). Fast MCMC sampling for Markov jump processes and extensions, *Journal of Machine Learning Research*, **14**: 3295–320.
- Roever, C. L., Beyer, H. L., Chase, M. J. and van Aarde, R. J. (2014). The pitfalls of ignoring behaviour when quantifying habitat selection, *Diversity and Distributions*, **20** (3): 322–33.
- Rowcliffe, J. M., Carbone, C., Kays, R., Kranstauber, B. and Jansen, P. A. (2012). Bias in estimating animal travel distance: The effect of sampling frequency, *Methods in Ecology and Evolution*, **3** (4): 653–62.
- Rue, H. and Held, L. (2005). *Gaussian Markov random fields Theory and applications*, Monographs on Statistics and Applied Probability 104, Boca Raton, FL: Chapman and Hall/CRC.
- Rutz, C. and Hays, G. C. (2009). New frontiers in biologging science, *Biology Letters*, **5** (3): 289–92.
- Scharf, H., Hooten, M. B. and Johnson, D. S. (2017). Imputation approaches for animal movement modeling, *Journal of Agricultural, Biological and Environmental Statistics*.
- Schliehe-Diecks, S., Kappeler, P. M. and Langrock, R. (2012). On the application of mixed hidden Markov models to multiple behavioural time series, *Interface focus*, **2** (2): 180–9.

- Siniff, D. B. and Jessen, C. R. (1969). A simulation model of animal movement patterns, *Advances in Ecological Research*, **6**: 185–217.
- Sur, M., Skidmore, A. K., Exo, K., Wang, T., Ens, B. J. and Toxopeus, A. G. (2014). Change detection in animal movement using discrete wavelet analysis, *Ecological Informatics*, **20**: 47–57.
- Thiebault, A. and Tremblay, Y. (2013). Splitting animal trajectories into fine-scale behaviorally consistent movement units: breaking points relate to external stimuli in a foraging seabird, *Behavioral Ecology and Sociobiology*, **67** (6): 1013–26.
- Towner, A. V., Leos-Barajas, V., Langrock, R., Schick, R. S., Smale, M. J., Kaschke, T., Jewell, O. J. and Papastamatiou, Y. P. (2016). Sex-specific and individual preferences for hunting strategies in white sharks, *Functional Ecology*, **30** (8): 1397–407.
- Tremblay, Y., Roberts, A. J. and Costa, D. P. (2007). Fractal landscape method: an alternative approach to measuring area-restricted searching behavior, *The Journal of Experimental Biology*, **210**: 935–45.
- Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of Brownian motion, *Physical Review Letters*, **36**: 823–41.
- Whitaker, G. A., Golightly, A., Boys, R. J. and Sherlock, C. (2017). Improved bridge constructs for stochastic differential equations, *Statistics and Computing*, **27**: 885–900.
- Wilmers, C. C., Nickel, B., Bryce, C. M., Smith, J. A., Wheat, R. E. and Yovovich, V. (2015). The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology, *Ecology*, **96** (7): 1741–53.
- Zhang, J., O'Reilly, K. M., Perry, G. L. W., Taylor, G. A. and Dennis, T. E. (2015). Extending the functionality of behavioural change-point analysis with k-means clustering: A case study with the little penguin (*Eudyptula minor*), *PLoS ONE*, **10** (4): 1–14.
- Zucchini, W., MacDonald, I. L. and Langrock, R. (2016). *Hidden Markov models for time series: An introduction using R*, CRC Press, Boca Raton FL, second edn.

Appendix A

Additional details

A.1 Derivation of Gibbs samplers

A.1.1 Sampling the bearing process parameter

In the following, all variables are as defined in Sect. 3.3.2.1. The full conditional posterior for the bearing process parameter is given by

$$\begin{aligned} p\left(\sigma_\theta^2 \mid \mu, \sigma_\varphi^2, \boldsymbol{\theta}, \mathbf{v}, \mathbf{Z}\right) &= p\left(\sigma_\theta^2 \mid \boldsymbol{\theta}\right) \\ &\propto p\left(\sigma_\theta^2\right) p\left(\boldsymbol{\theta} \mid \sigma_\theta^2\right) \\ &\propto \left[\left(\sigma_\theta^2\right)^{-a_\theta-1} \exp\left(-\frac{b_\theta}{\sigma_\theta^2}\right) \right] \left[\left(\sigma_\theta^2\right)^{-\frac{m-1}{2}} \exp\left(-\frac{1}{2\sigma_\theta^2} \sum_{i=1}^{m-1} \left(\frac{\delta\theta_i}{\delta t_i}\right)^2\right) \right] \\ &= \left(\sigma_\theta^2\right)^{-\left(a_\theta+\frac{m-1}{2}\right)-1} \exp\left(-\frac{1}{\sigma_\theta^2} \left(b_\theta + \sum_{i=1}^{m-1} \left(\frac{\delta\theta_i}{\delta t_i}\right)^2\right)\right) \\ &\propto \text{IG}\left(a_\theta + \frac{m-1}{2}, b_\theta + \sum_{i=1}^{m-1} \left(\frac{\delta\theta_i}{\delta t_i}\right)^2\right). \end{aligned}$$

A.1.2 Sampling the behavioural process parameters

In the following, all variables are as defined in Sect. 4.3.2. The full conditional posterior for the behavioural process parameters is given by

$$\begin{aligned}
& p(\Phi_B | \mathbf{B}, \boldsymbol{\theta}, \mathbf{v}, \mathbf{Z}, \Phi_M) \\
&= p(\Phi_B | \mathbf{B}) \\
&\propto p(\mathbf{B} | \Phi_B) p(\Phi_B) \\
&\propto \left[\exp\left(-\sum_i \lambda_i a_i\right) \prod_i \prod_{j \neq i} (q_{ij} \lambda_i)^{b_{ij}} \right] \left[\left(\prod_i \lambda_i^{c_i-1} \exp(-d_i \lambda_i) \right) \left(\prod_i \prod_{j \neq i} (q_{ij}^{f_{ij}-1}) \right) \right] \\
&= \left[\exp\left(-\sum_i (a_i + d_i) \lambda_i\right) \prod_i (\lambda_i^{c_i + \sum_{j \neq i} b_{ij}}) \right] \left[\prod_i \prod_{j \neq i} (q_{ij}^{b_{ij} + f_{ij} - 1}) \right] \\
&\propto \prod_i \text{Gam}\left(\lambda_i | c_i + \sum_i b_{ij}, d_i + a_i\right) \text{Dir}(\mathbf{q}_i | \mathbf{f}_i + \mathbf{b}_i),
\end{aligned}$$

where $i, j \in \{1, \dots, p\}$.

A.2 Conditioning by Kriging

A.2.1 In the absence of observation error

The following shows that the conditioning by Kriging method for sampling \mathbf{v}^* in Sect. 3.3.3.1 is equivalent to sampling from the target distribution in Eq. 3.8. With all variables as those defined in Sect. 3.3.3.1, the mean and covariance of Eq. 3.9 are given by

$$\begin{aligned}
\mathbf{E}(\mathbf{v}^*) &= \mathbf{E}\left(\mathbf{x}^* - \Sigma_{1,2} \Sigma_2^{-1} \left((\mathbf{Z}(j), \mathbf{Z}(j))^T + \mathbf{A} \mathbf{x}^* - \mathbf{Z}(k) \right)\right) \\
&= \mathbf{m}_1 - \Sigma_{1,2} \Sigma_2^{-1} \left((\mathbf{Z}(j), \mathbf{Z}(j))^T + \mathbf{A} \mathbf{m}_1 - \mathbf{Z}(k) \right) \\
&= \mathbf{m}_1 - \Sigma_{1,2} \Sigma_2^{-1} (\mathbf{m}_2 - \mathbf{Z}(k)) = \hat{\boldsymbol{\mu}},
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(\mathbf{v}^*) &= \text{Cov}\left(\mathbf{x}^* - \Sigma_{1,2}\Sigma_2^{-1}\left((\mathbf{Z}(j), \mathbf{Z}(j))^T + A\mathbf{x}^* - \mathbf{Z}(k)\right)\right) \\
&= \text{Cov}\left(\left(I - \Sigma_{1,2}\Sigma_2^{-1}A\right)\mathbf{x}^*\right) \\
&= \left(I - \Sigma_{1,2}\Sigma_2^{-1}A\right)\Sigma_1\left(I - A^T\Sigma_2^{-1}\Sigma_{1,2}^T\right) \\
&= \Sigma_1 - \Sigma_1A^T\Sigma_2^{-1}\Sigma_{1,2}^T - \Sigma_{1,2}\Sigma_2^{-1}A\Sigma_1 + \Sigma_{1,2}\Sigma_2^{-1}A\Sigma_1A^T\Sigma_2^{-1}\Sigma_{1,2}^T \\
&= \Sigma_1 - \Sigma_{1,2}\Sigma_2^{-1}\Sigma_{1,2}^T - \Sigma_{1,2}\Sigma_2^{-1}\Sigma_{1,2}^T + \Sigma_{1,2}\Sigma_2^{-1}\Sigma_2\Sigma_2^{-1}\Sigma_{1,2}^T \\
&= \Sigma_1 - \Sigma_{1,2}\Sigma_2^{-1}\Sigma_{1,2}^T = \hat{\Sigma}.
\end{aligned}$$

A.2.2 In the presence of independent observation errors

The following shows that the conditioning by Kriging method for sampling \mathbf{v}^* in Sect. 5.2.2.1 is equivalent to sampling from the target distribution in Eq. 5.3. With all variables as those defined in Sect.5.2.2.1, the mean and covariance of Eq. 5.4 are given by

$$\begin{aligned}
\mathbb{E}(\mathbf{v}^*) &= \mathbb{E}\left(\mathbf{x}^* - \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\left((\mathbf{Z}(j), \mathbf{Z}(j))^T + A\mathbf{x}^* - \mathbf{y}^*\right)\right) \\
&= \mathbf{m}_1 - \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\left((\mathbf{Z}(j), \mathbf{Z}(j))^T + A\mathbf{m}_1 - (\mathbf{Z}^*(k), \mathbf{Z}(l))^T\right) \\
&= \mathbf{m}_1 - \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\left(\mathbf{m}_2 - (\mathbf{Z}^*(k), \mathbf{Z}(l))^T\right) = \hat{\boldsymbol{\mu}},
\end{aligned}$$

$$\begin{aligned}
\text{Cov}(\mathbf{v}^*) &= \text{Cov}\left(\mathbf{x}^* - \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\left((\mathbf{Z}(j), \mathbf{Z}(j))^T + A\mathbf{x}^* - \mathbf{y}^*\right)\right) \\
&= \left(I - \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}A\right)\Sigma_1\left(I - \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}A\right)^T \\
&\quad + \left(\Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\right)\Sigma_\epsilon\left(\Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\right)^T \\
&= \Sigma_1 + \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}A\Sigma_1A^T(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_{1,2}^T - \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}A\Sigma_1 \\
&\quad - \Sigma_1A^T(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_{1,2}^T + \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_\epsilon(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_{1,2}^T \\
&= \Sigma_1 + \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_2(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_{1,2}^T - \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_{1,2}^T \\
&\quad - \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_{1,2}^T + \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_\epsilon(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_{1,2}^T \\
&= \Sigma_1 + \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}(\Sigma_2 + \Sigma_\epsilon)(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_{1,2}^T - 2\Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_{1,2}^T \\
&= \Sigma_1 + \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_{1,2}^T - 2\Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_{1,2}^T \\
&= \Sigma_1 - \Sigma_{1,2}(\Sigma_2 + \Sigma_\epsilon)^{-1}\Sigma_{1,2}^T = \hat{\Sigma}.
\end{aligned}$$

A.2.3 In the presence of correlated observation errors

The following shows that the conditioning by Kriging method for sampling \mathbf{v}^* in Sect. 5.6.1.2 is equivalent to sampling from the target distribution in Eq. 5.6. With all variables as those defined in Sect.5.6.1.2, the mean of Eq. 5.7 is given by

$$\begin{aligned} \mathbb{E}(\mathbf{v}^*) &= \mathbb{E}\left(\mathbf{x}^* - \Sigma_{1,2}(\Sigma_2 + \Sigma_\varepsilon)^{-1} \left((\mathbf{Z}(j), \mathbf{Z}(j))^T + \mathbf{A}\mathbf{x}^* - \mathbf{y}^* \right)\right) \\ &= \mathbf{m}_1 - \Sigma_{1,2}(\Sigma_2 + \Sigma_\varepsilon)^{-1} \left((\mathbf{Z}(j), \mathbf{Z}(j))^T + \mathbf{A}\mathbf{m}_1 - \left((\mathbf{Z}^*(k), \mathbf{Z}(l))^T - \boldsymbol{\mu}_\varepsilon \right) \right) \\ &= \mathbf{m}_1 - \Sigma_{1,2}(\Sigma_2 + \Sigma_\varepsilon)^{-1} \left(\mathbf{m}_2 + \boldsymbol{\mu}_\varepsilon - (\mathbf{Z}^*(k), \mathbf{Z}(l))^T \right). \end{aligned}$$

The covariance structure in this case is the same as that in Sect. A.2.2, however, note that the form of Σ_ε in those two cases differs.

Appendix B

Additional figures

B.1 Single state independent steps simulation

The following includes additional figures to those included in the main text for the simulation example of Sect. 3.5.1 that assumes a single state, independent step process (as described in Sect. 3.1).

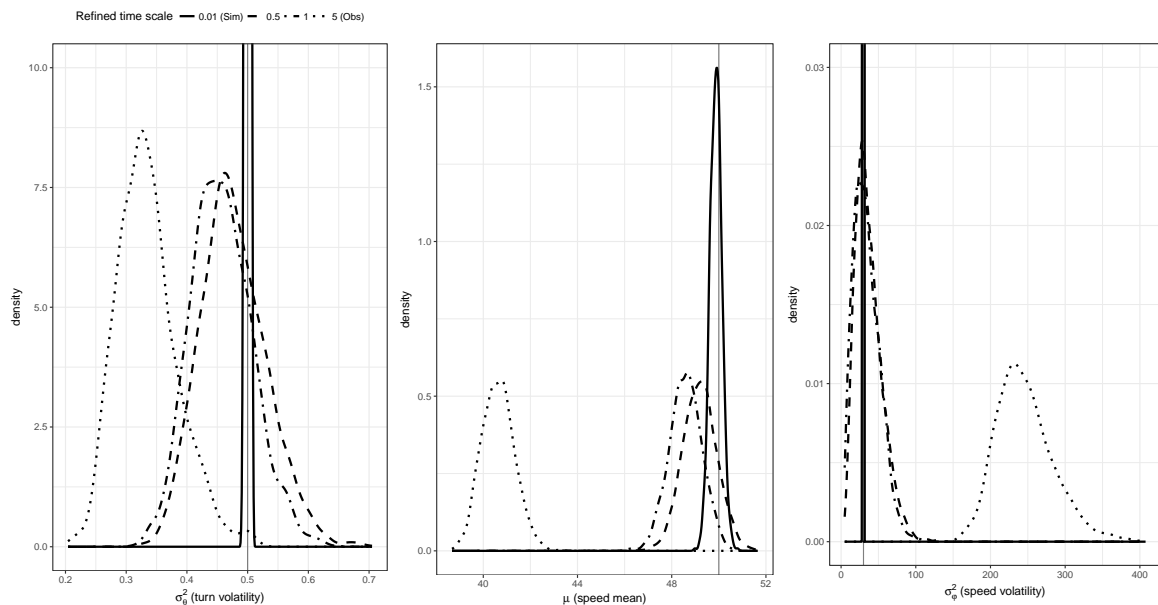


Fig. B.1 Posterior kernel density estimates of the movement parameters (using thinned samples and with burn-in time omitted) for the single state, independent step simulation, with true parameter values used for the simulation (**grey**). Note that the full height of the density estimates for the ‘base-line’ implementation is not included.

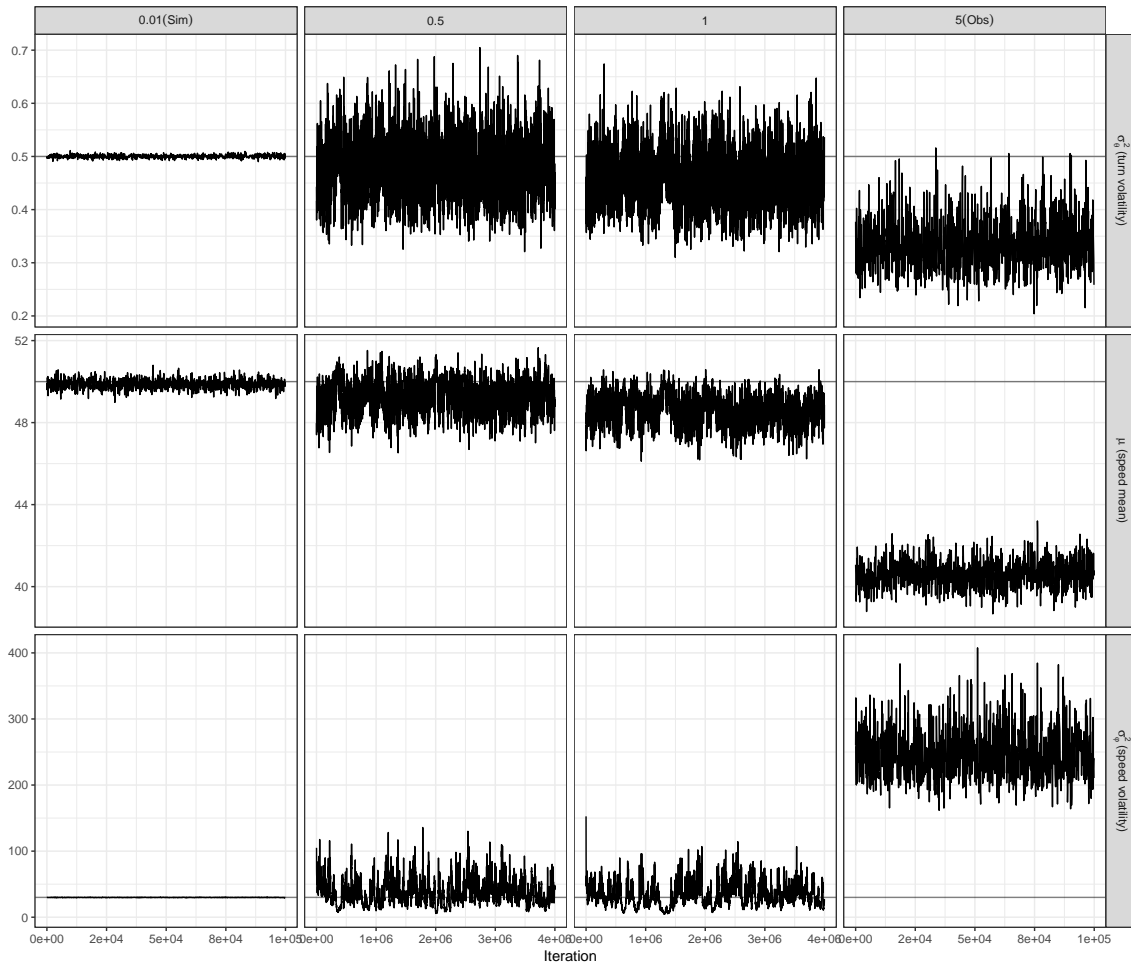


Fig. B.2 Movement parameter trace (thinned) in the single state, independent step simulation. For reference, true parameter values used for the simulation are highlighted (grey).

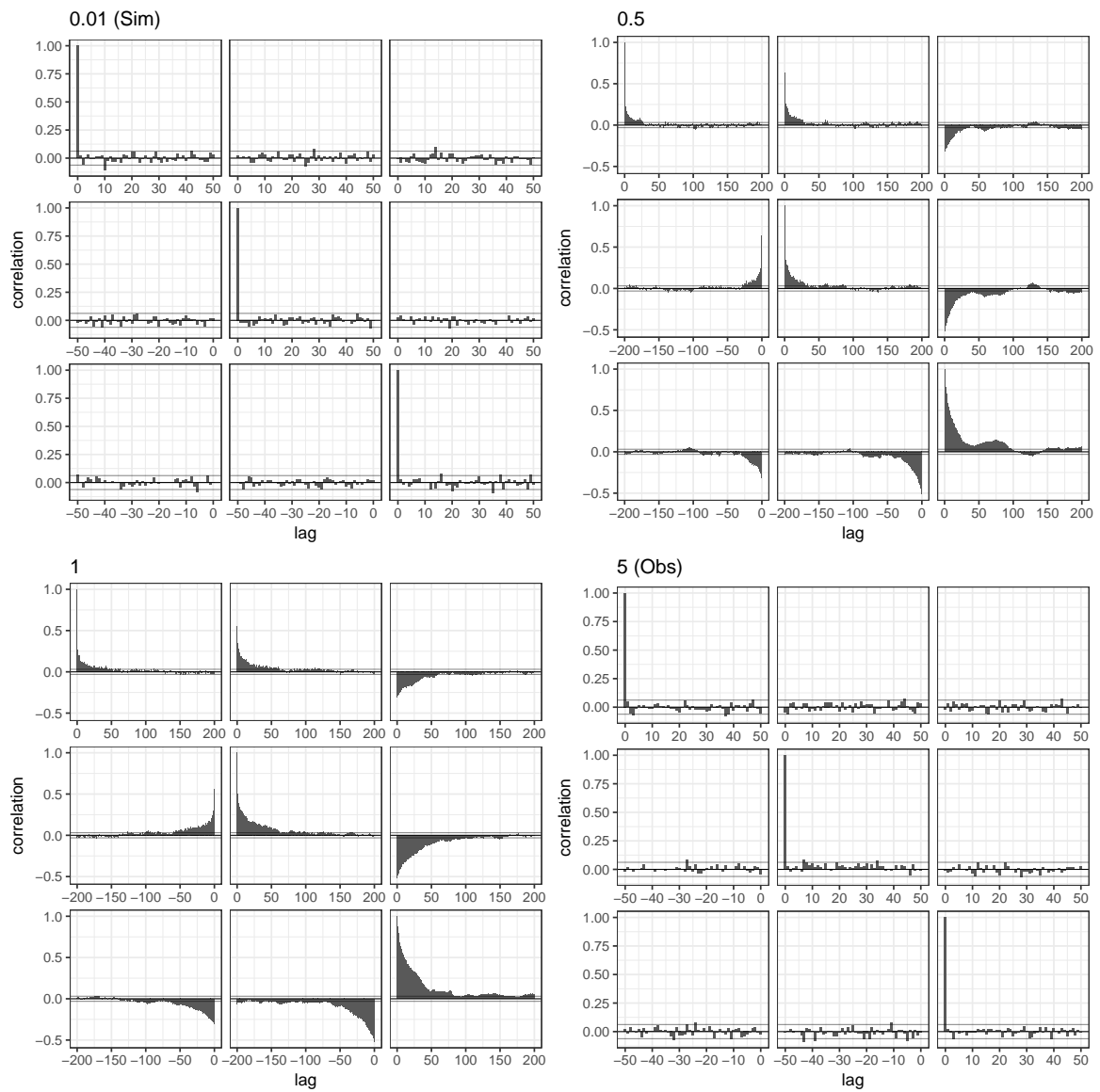


Fig. B.3 Autocorrelation in the sampled movement parameters, for the single state, independent step simulation. Maximum lag shown is given as 5% of the thinned sample size.

B.2 Single state correlated steps simulation

The following includes additional figures to those included in the main text for the simulation example of Sect. 3.5.2 that assumes a single state, correlated speed process (as described in Sect. 3.4).

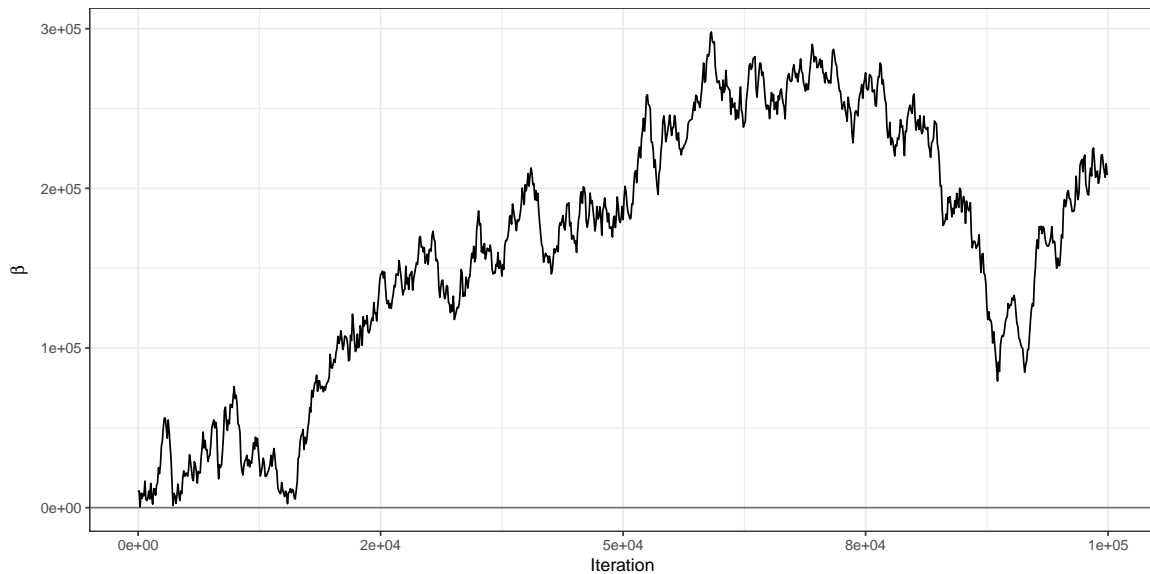


Fig. B.4 Speed correlation trace for the sampling time scale implementation, shown separately from Fig. B.5. See main text for discussion of the ‘drifting’ displayed here.

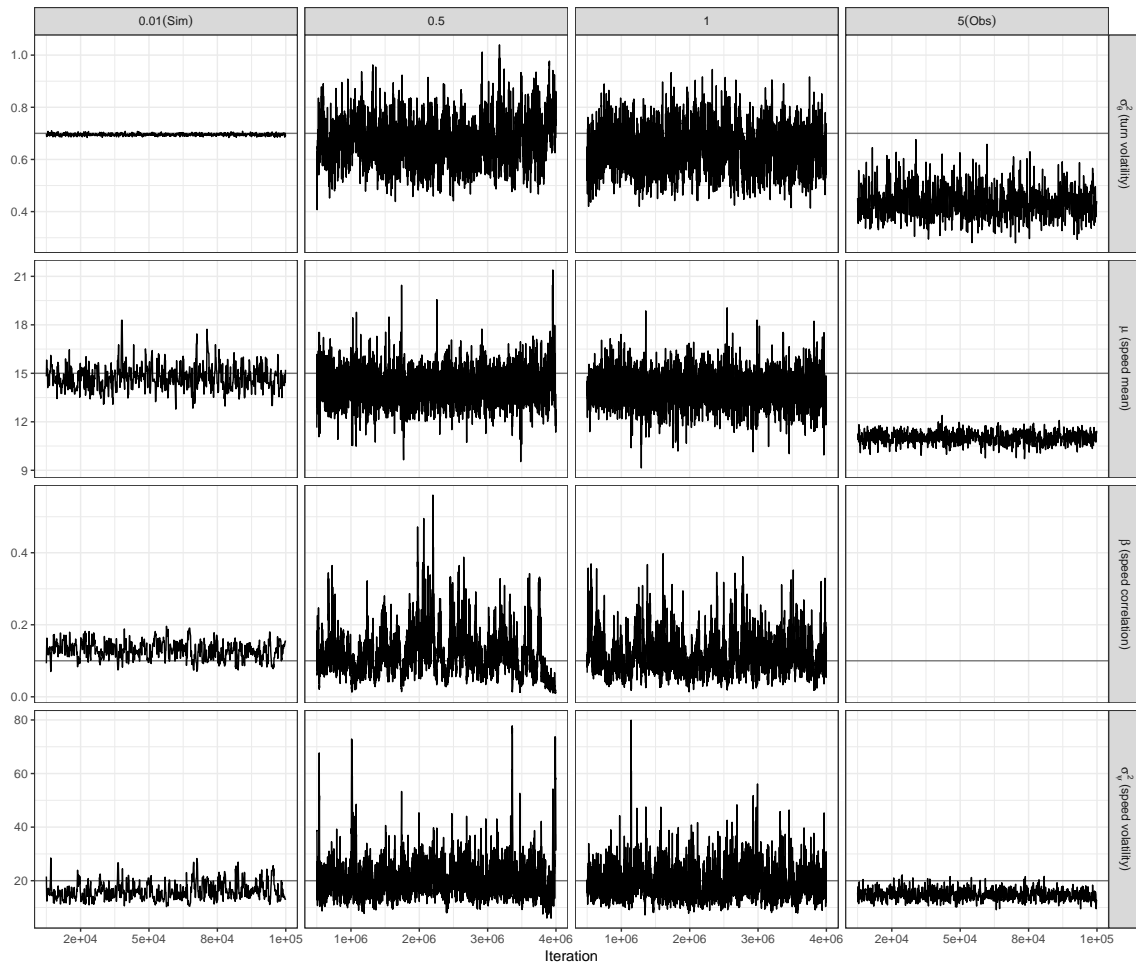


Fig. B.5 Movement parameter trace (thinned) in the single state, correlated step simulation. True parameter values used for the simulation are highlighted (grey) and the missing β trace is shown in Fig. B.4.

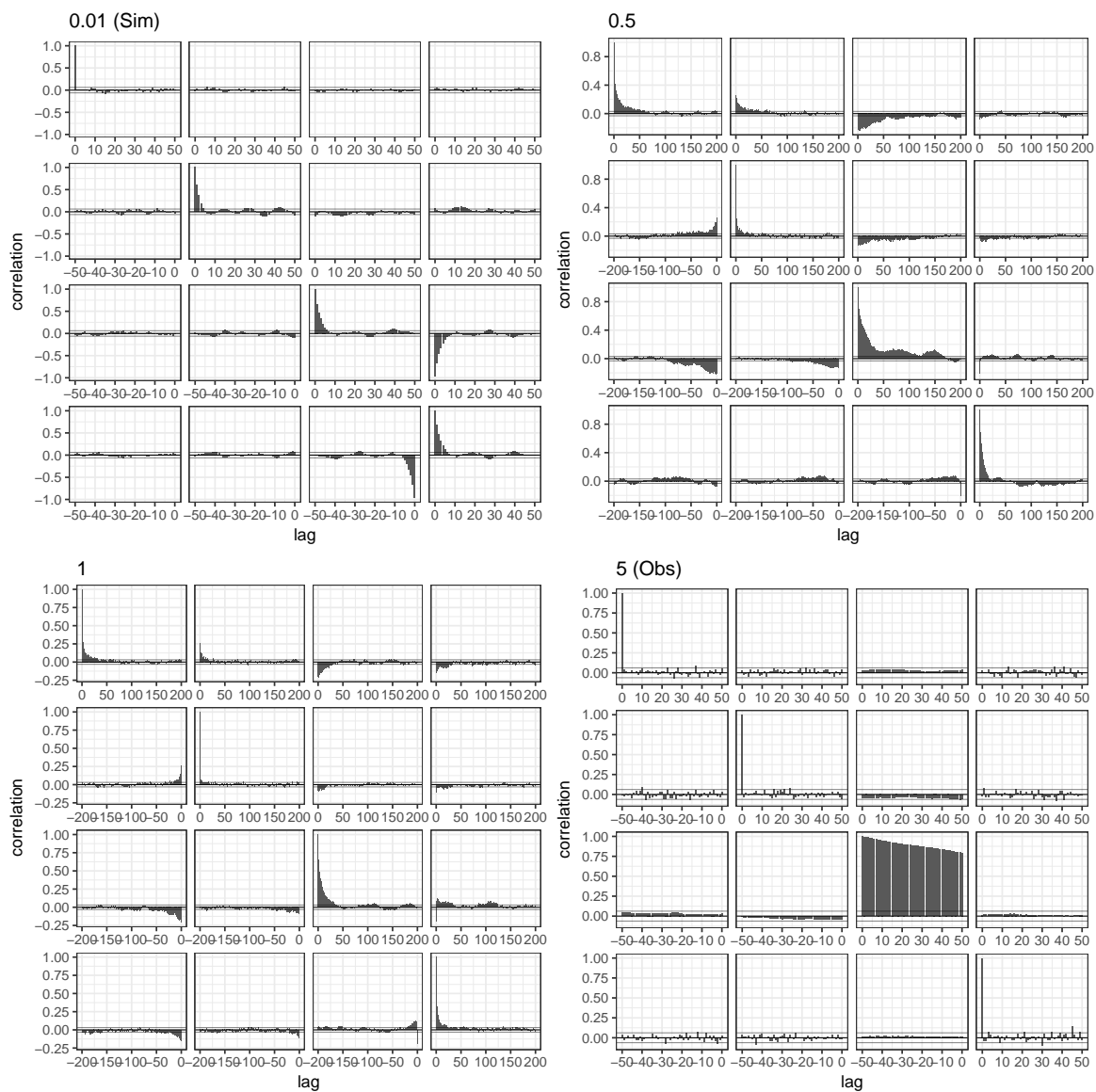


Fig. B.6 Autocorrelation in the sampled movement parameters, for the single state, correlated step simulation. Maximum lag shown is given as 5% of the thinned sample size.

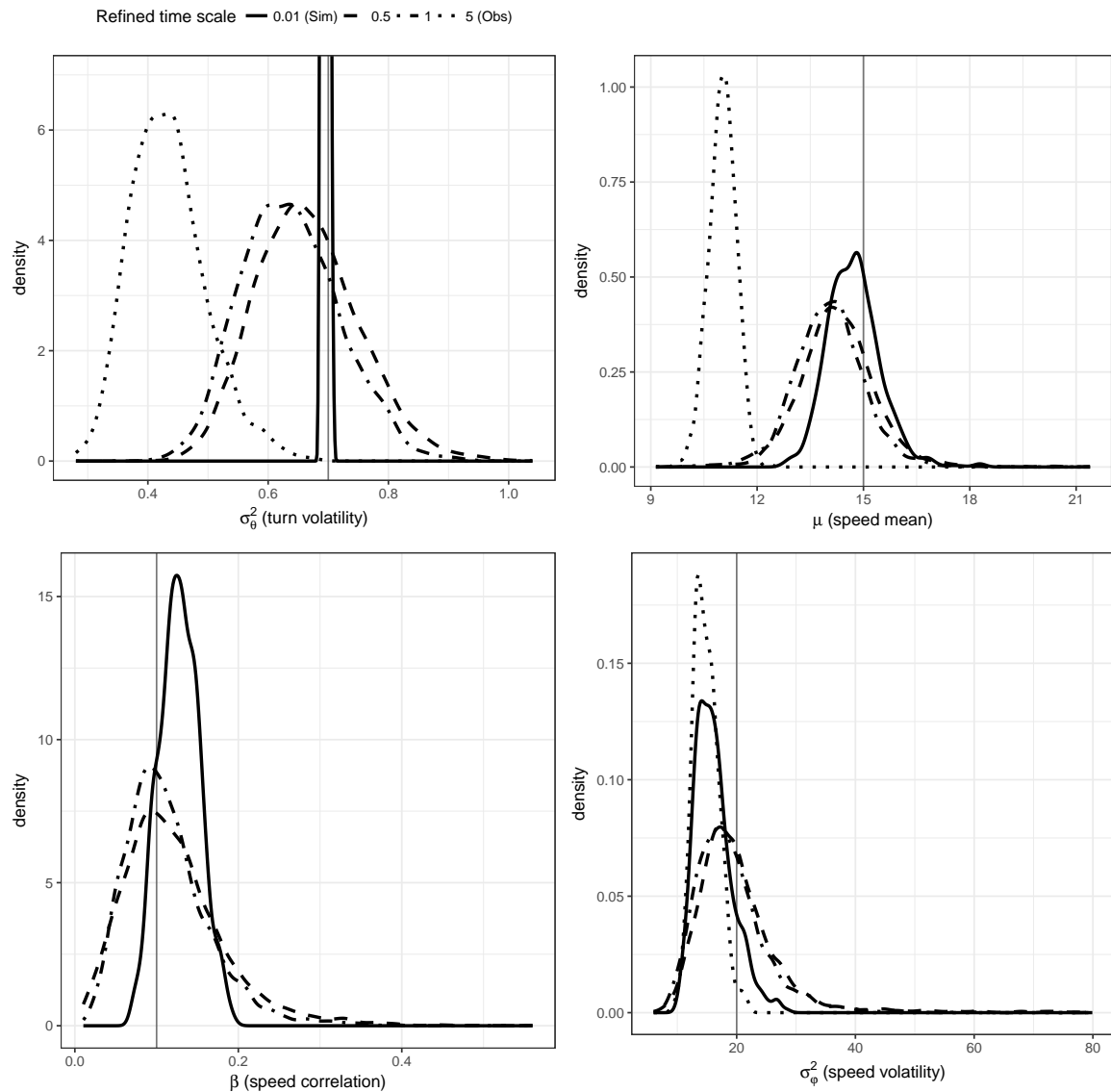


Fig. B.7 Posterior kernel density estimates of the movement parameters (using thinned samples and with burn-in time omitted) for the single state, correlated step simulation, with true parameter values used for the simulation (**grey**). Note that the full height of the density estimate of the bearing parameter for the ‘base-line’ implementation is not included, and the speed correlation parameter is not estimated for the implementation at the sampling time scale (see main text).

B.3 Two-state simulation

The following includes additional figures to those included in the main text for the simulation example of Sect. 4.4 that assumes a two-state, correlated step process.

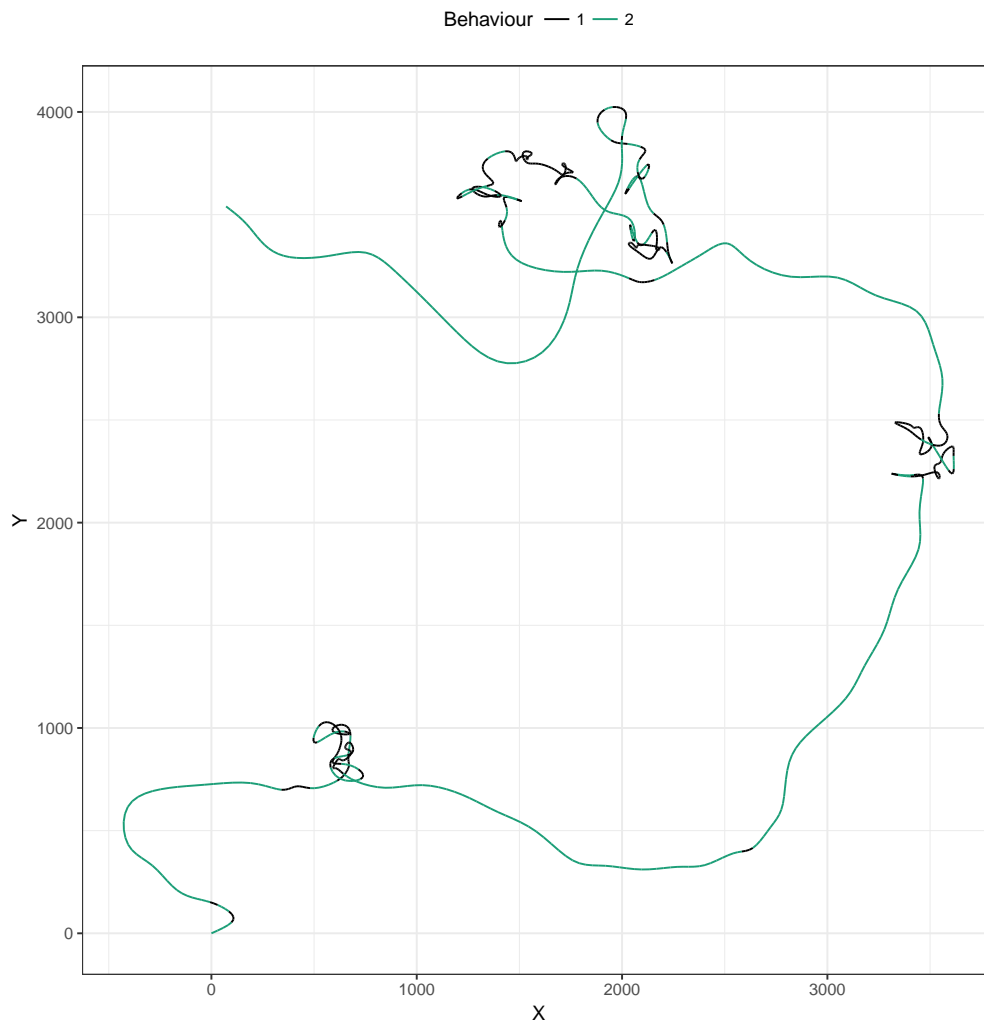


Fig. B.8 Initial two-state movement path, coloured by behaviour.

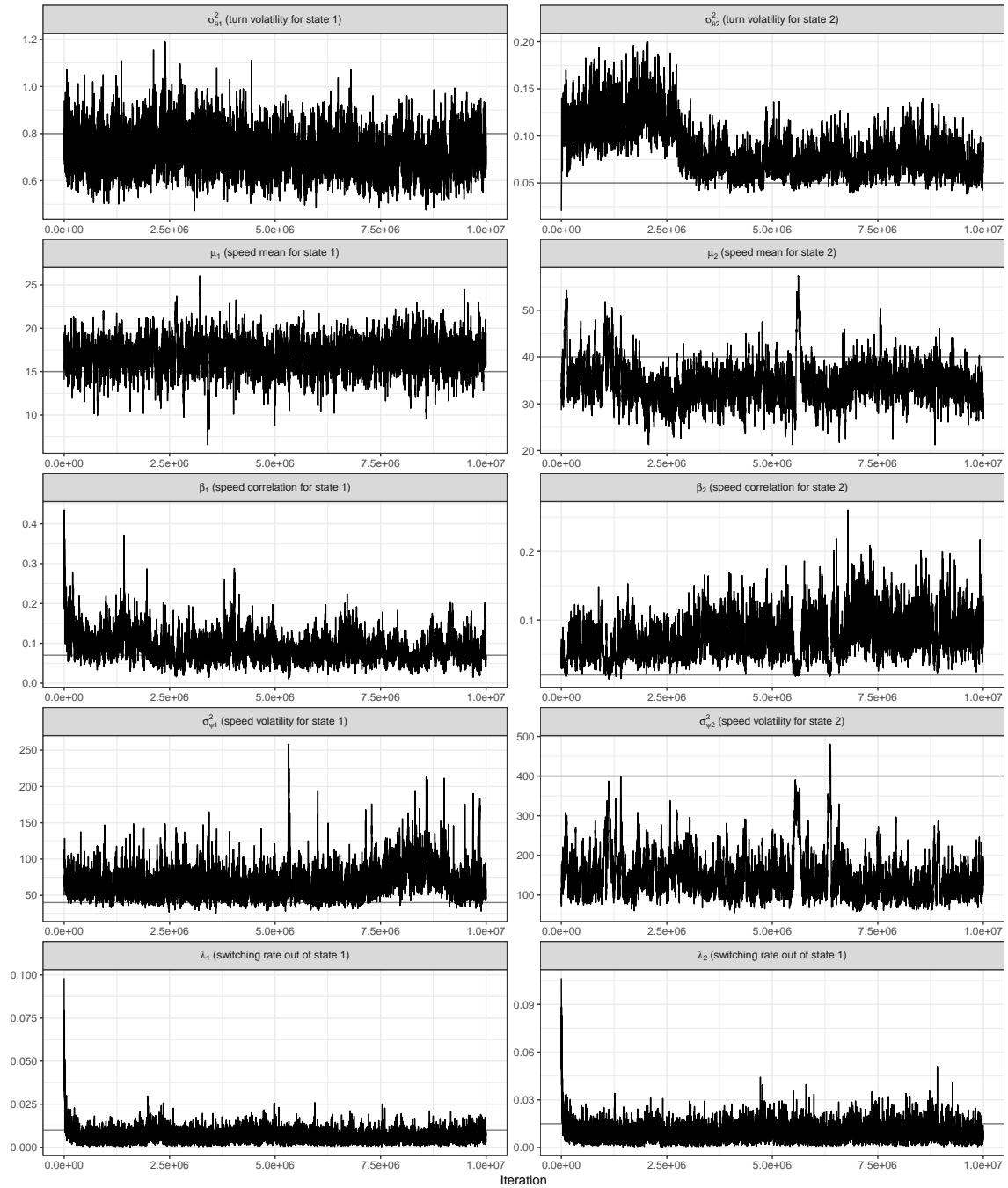


Fig. B.9 Sample trace of the parameters for the two-state simulation example (**black**). True values used for the simulation are highlighted (**grey**).

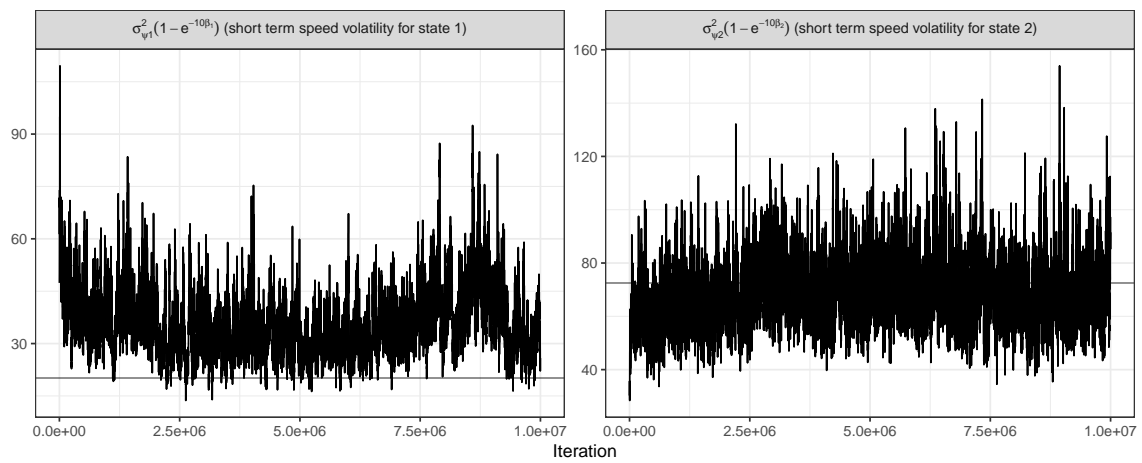


Fig. B.10 Sample trace of the short-term speed variance for the two-state simulation example (**black**). True values used for the simulation are highlighted (**grey**).

B.4 Two-state movement in elk

The following includes additional figures to those included in the main text for the two-state example of Sect. 4.5 that uses elk data.

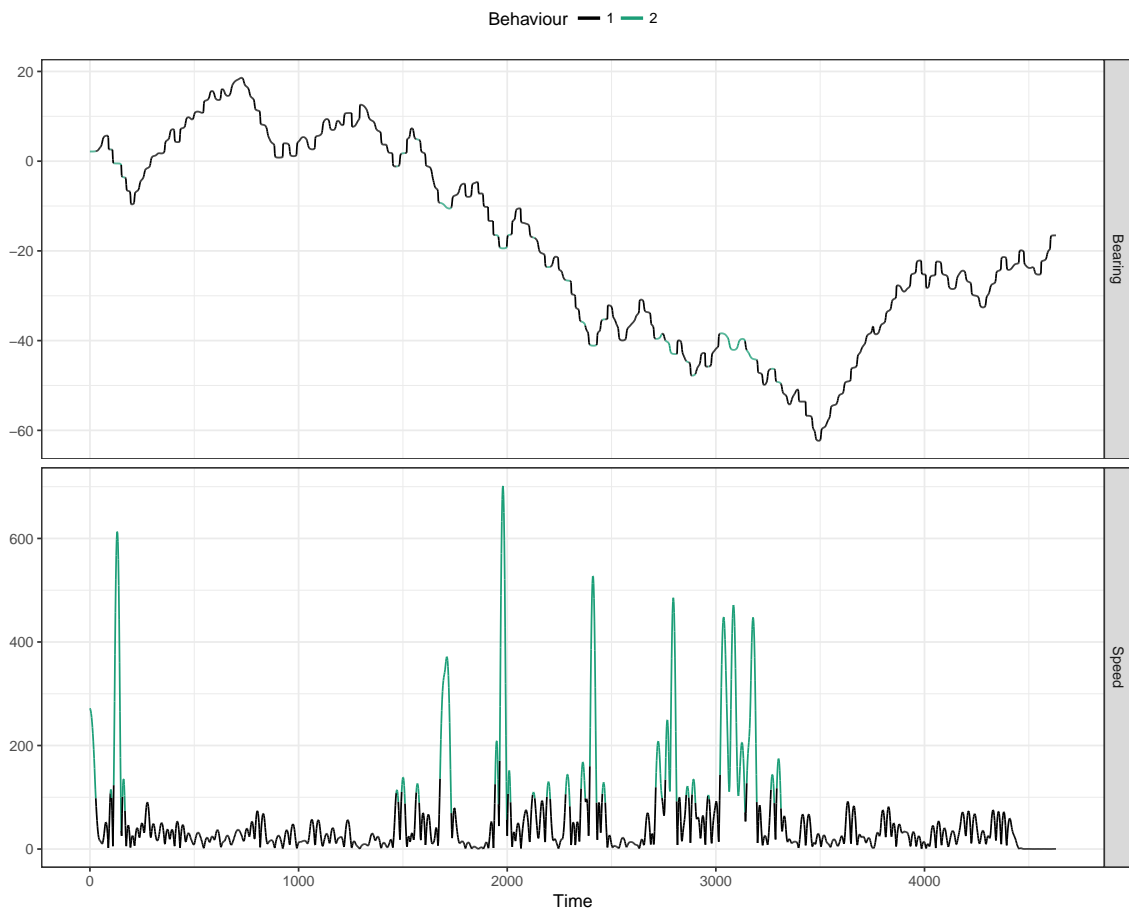


Fig. B.11 Initial path as bearings and speeds for the elk-115 example, coloured by the initial behavioural configuration.

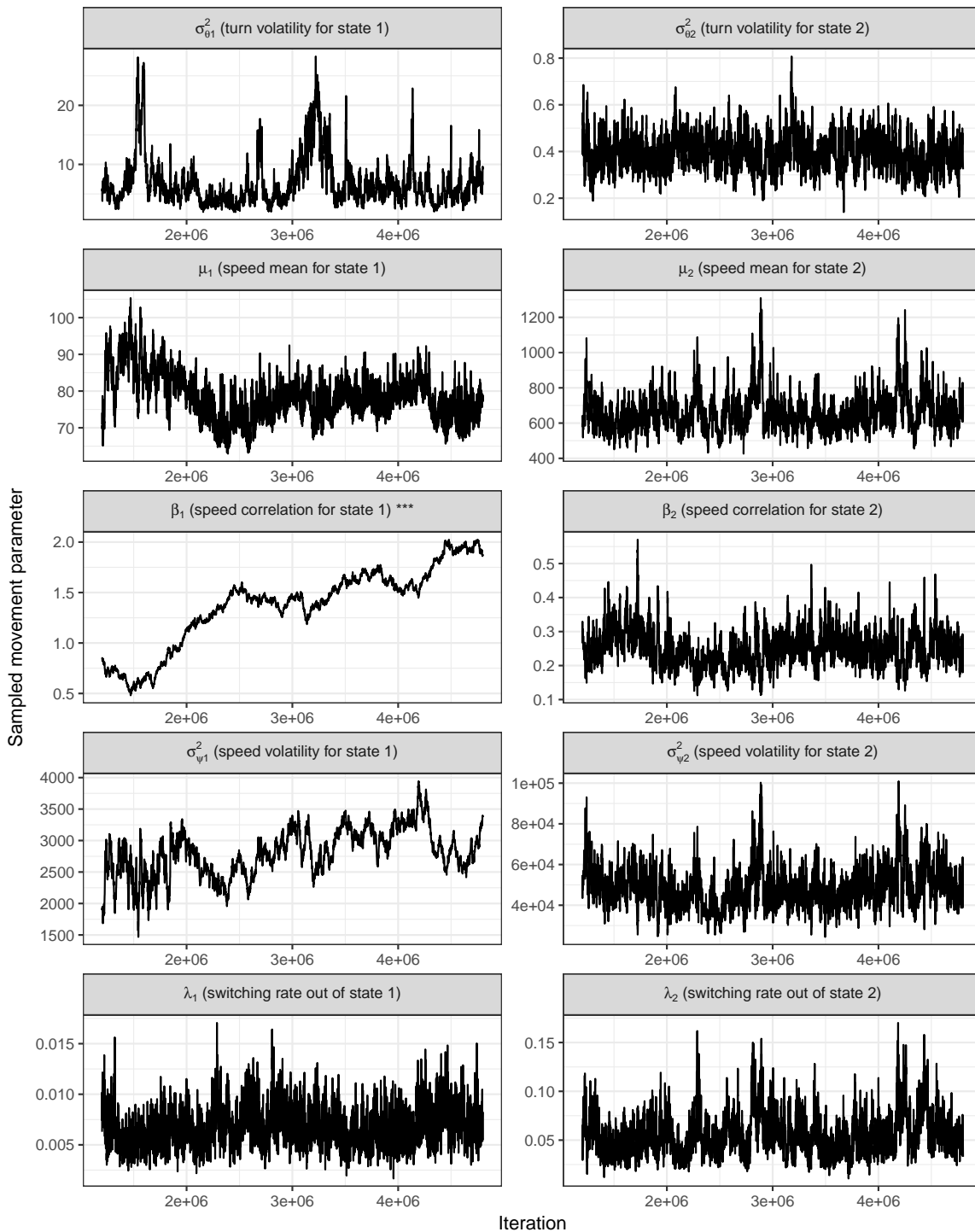


Fig. B.12 Trace of the posterior samples of the movement and behavioural parameters in the elk-115 example, split by behavioural state. Samples are thinned, and with burn-in period discarded. Note that the correlation parameter for state 2, β_2 , is discussed further in the main text.

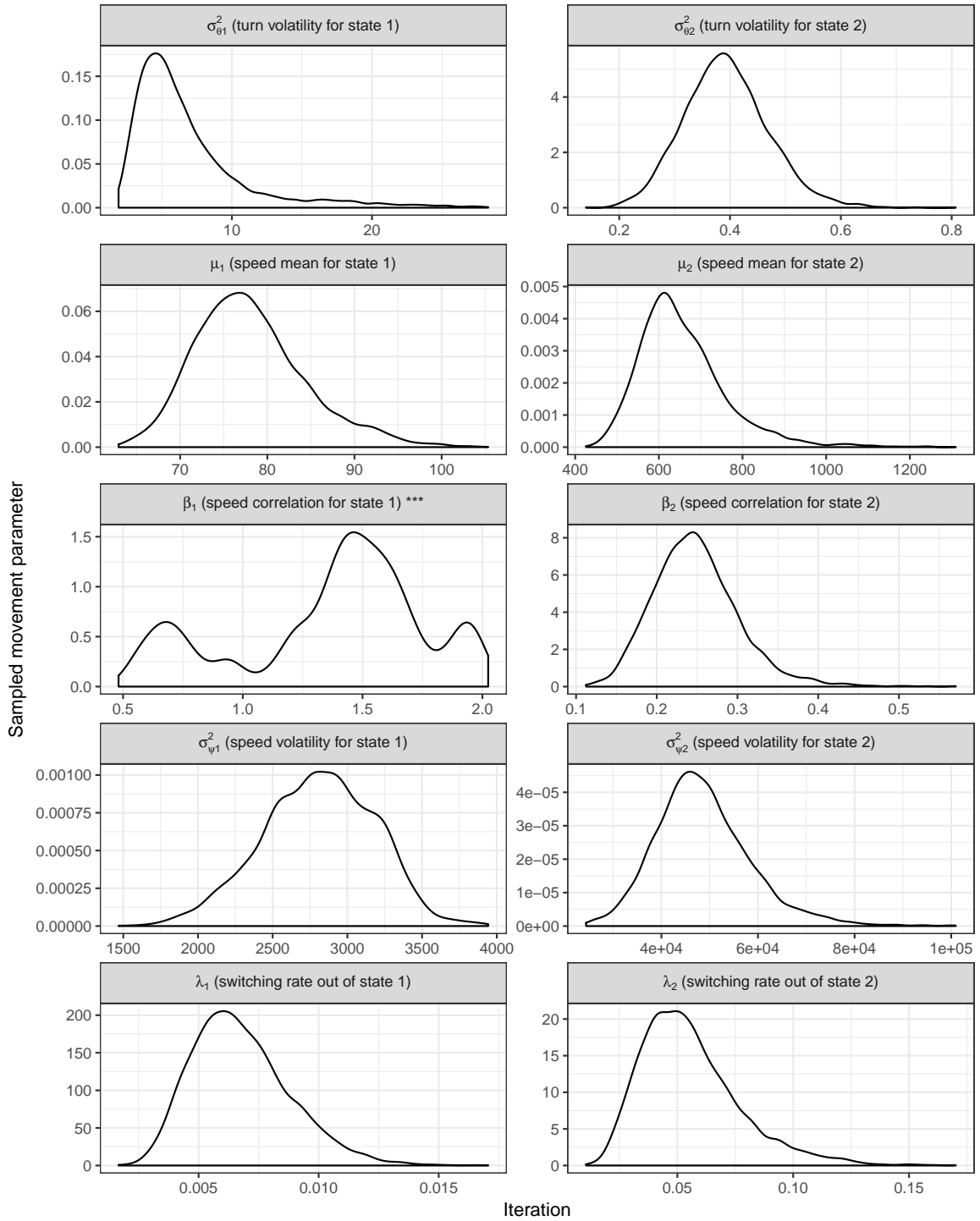


Fig. B.13 Kernel density estimates of the posterior samples of the movement and behavioural parameters in the elk-115 example, split by behavioural state. Estimates are based on thinned samples with burn-in period discarded. Note that the correlation parameter for state 2, β_2 , is discussed further in the main text.

B.5 Noisy observations of single state reindeer movement

The following includes additional figures to those included in the main text for the example of Sect. 5.3 that assumes single state movement observed with noise using reindeer data.

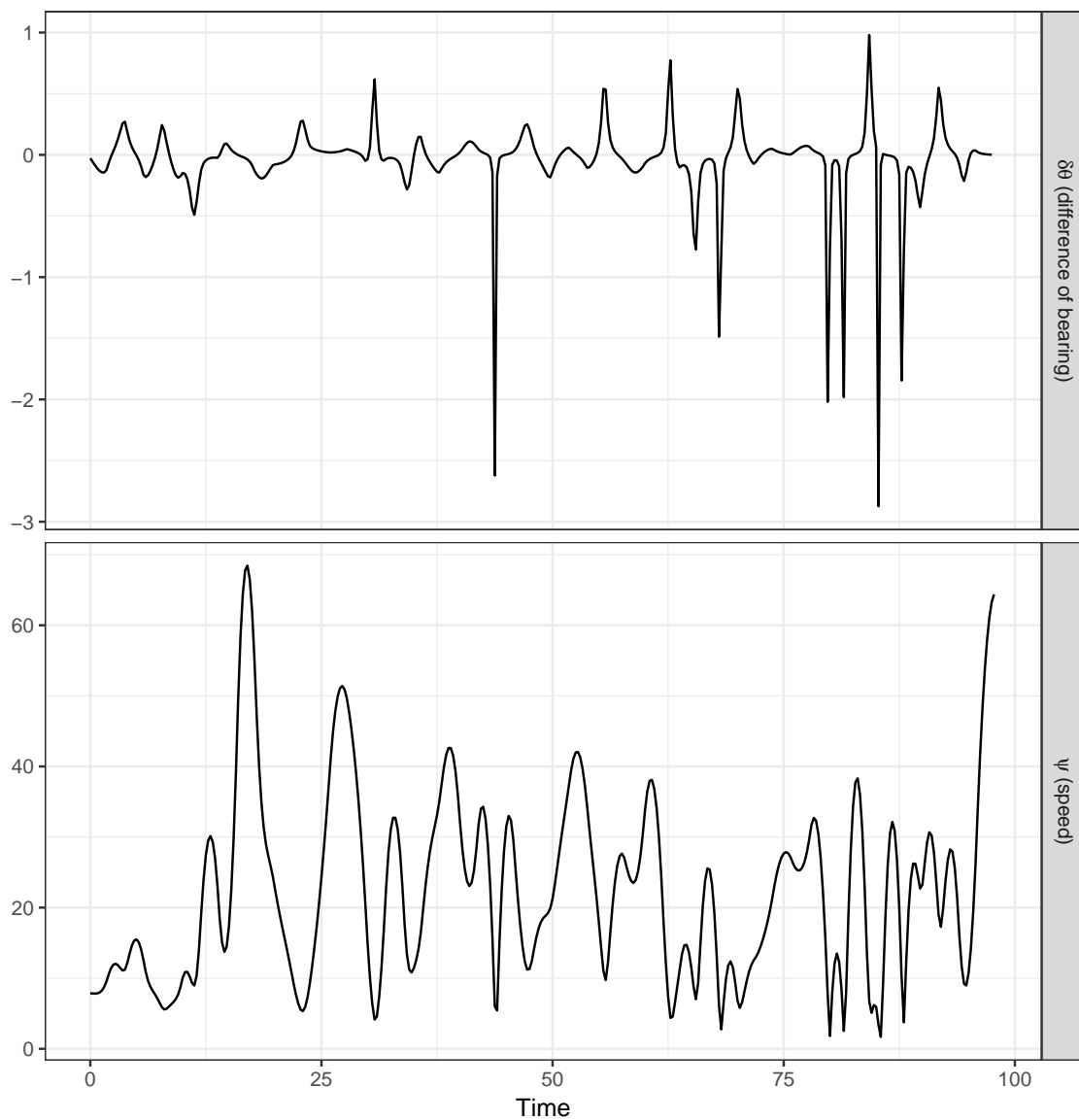


Fig. B.14 Initial movement path (as bearings and speeds) used within the example of single state reindeer movement with noisy observations. Initial paths was created used interpolating splines between perturbations of the observed locations to allow for observation error.

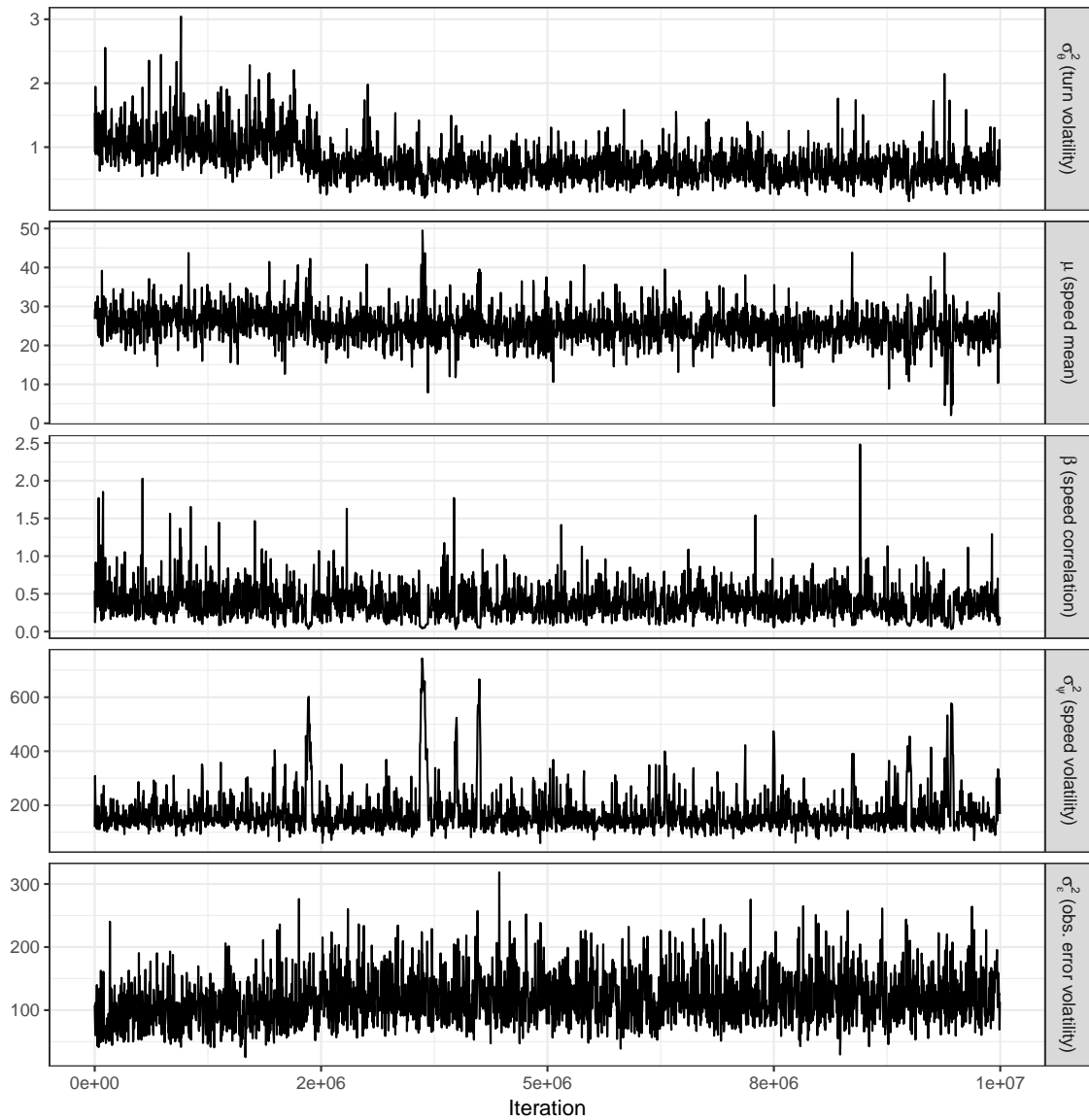


Fig. B.15 Trace of the posterior samples of the movement and observation error parameters in the reindeer b53.10 example. Trace is thinned, but includes burn-in period.

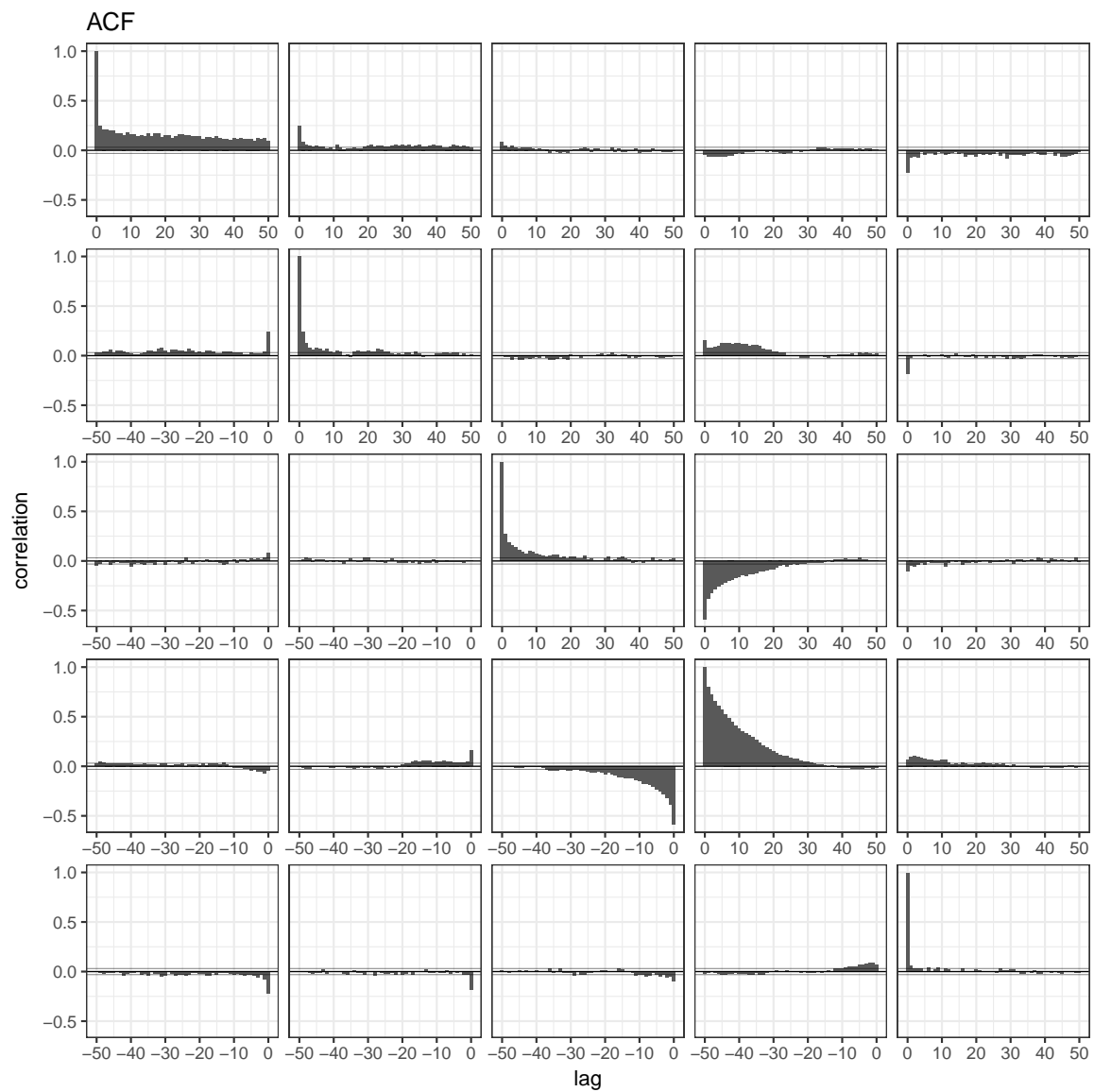


Fig. B.16 Autocorrelation in the MCMC movement and observation error parameter sampler, for the example of single state reindeer movement with noisy observations. Maximum lag shown up to 50, constituting 1.25% of the thinned sample size.

B.6 Noisy observations of two-state gull movement

The following includes additional figures to those included in the main text for the simulation example of Sect. 5.4 that assumes two-state movement of lesser black-backed gulls, observed with noise.

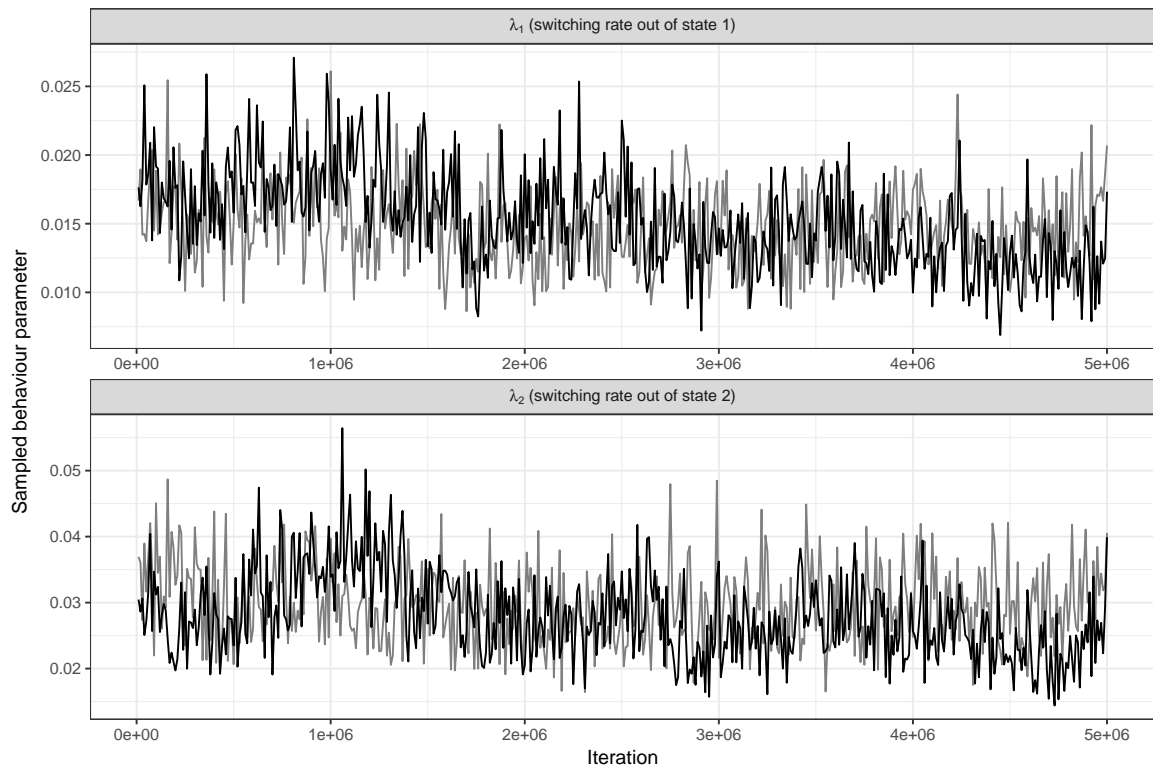


Fig. B.17 Trace of the posterior samples of the behavioural parameters in the gull-3 example, split by behavioural state. Samples are thinned, and with burn-in period discarded. The implementation with approximate refined time scale of 1 is displayed in **black** and that with time scale 0.5 in **grey**.

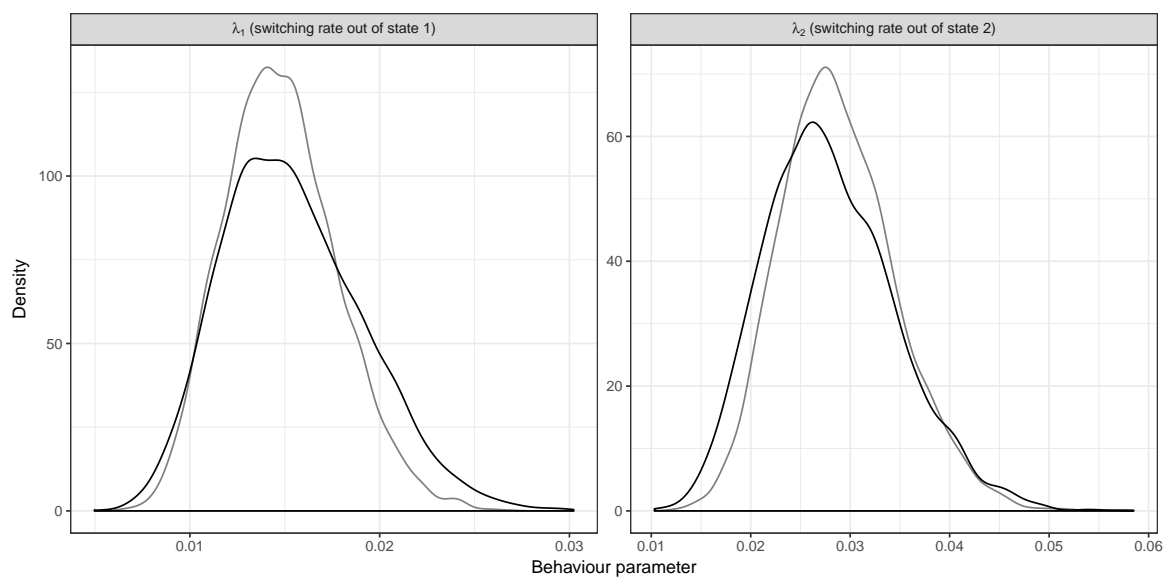


Fig. B.18 Kernel density estimates of the posterior samples of the behavioural parameters in the gull-3 example, split by behavioural state. Estimates are based on thinned samples with burn-in period discarded. The implementation with approximate refined time scale of 1 is displayed in **black** and that with time scale 0.5 in **grey**.

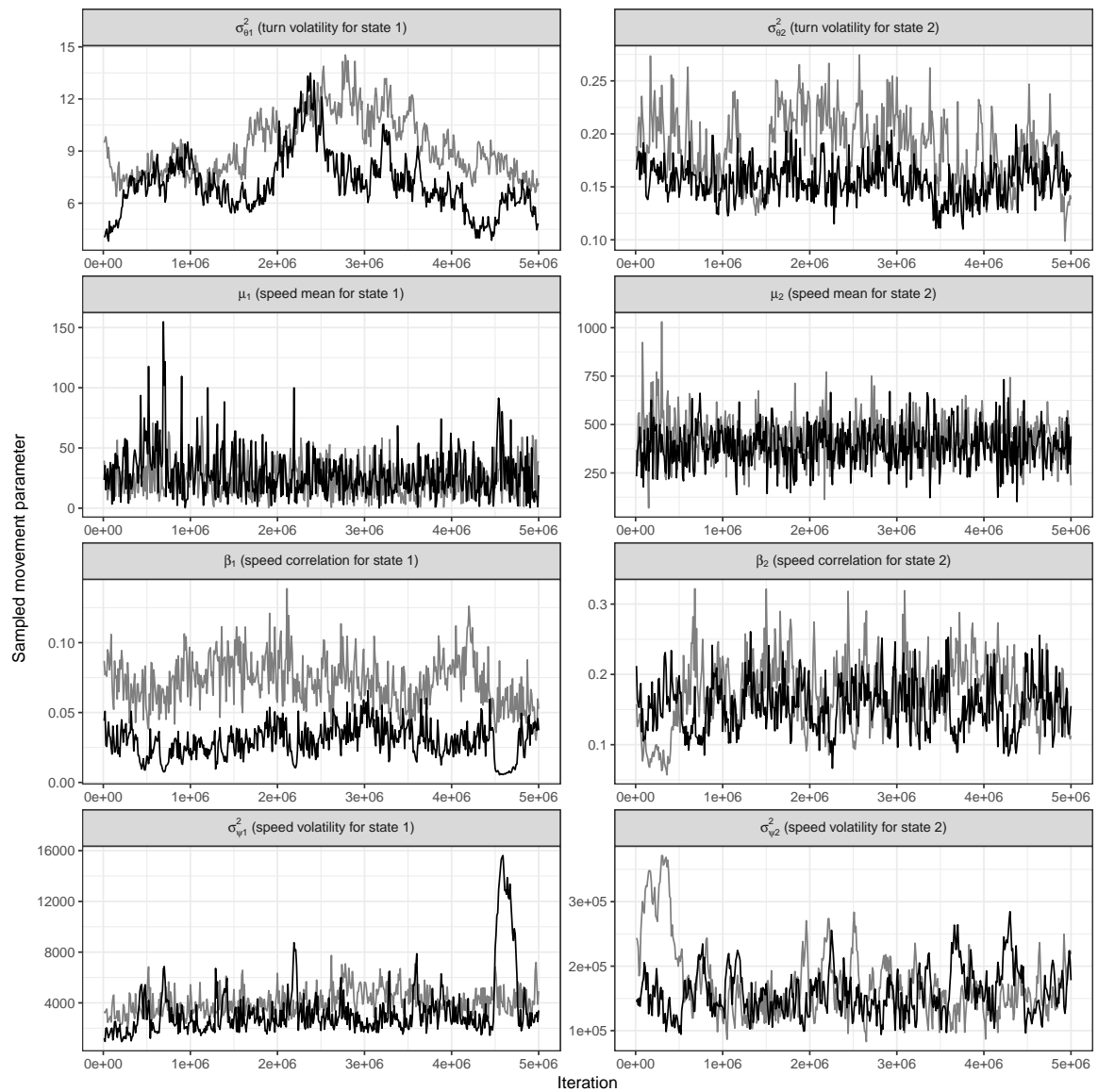


Fig. B.19 Trace of the posterior samples of the movement parameters in the gull-3 example, split by behavioural state. Samples are thinned, and with burn-in period discarded. The implementation with approximate refined time scale of 1 is displayed in **black** and that with time scale 0.5 in **grey**.

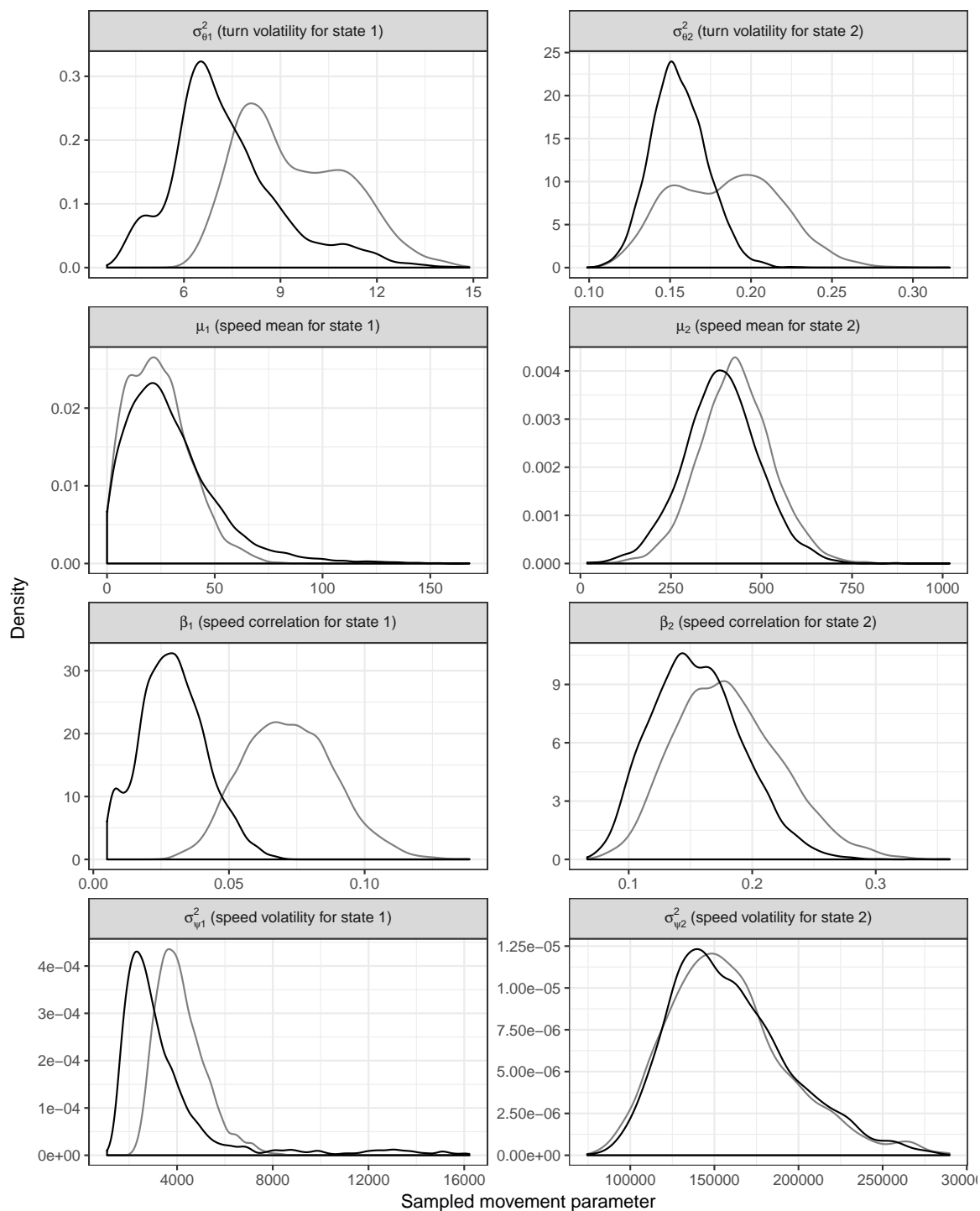


Fig. B.20 Kernel density estimates of the posterior samples of the movement parameters in the gull-3 example, split by behavioural state. Estimates are based on thinned samples with burn-in period discarded. The implementation with approximate refined time scale of 1 is displayed in **black** and that with time scale 0.5 in **grey**.

B.7 Observation error simulation example

The following includes additional figures to those included in the main text for the simulation example of Sect. 5.5 that assumes single state movement observed with noise.

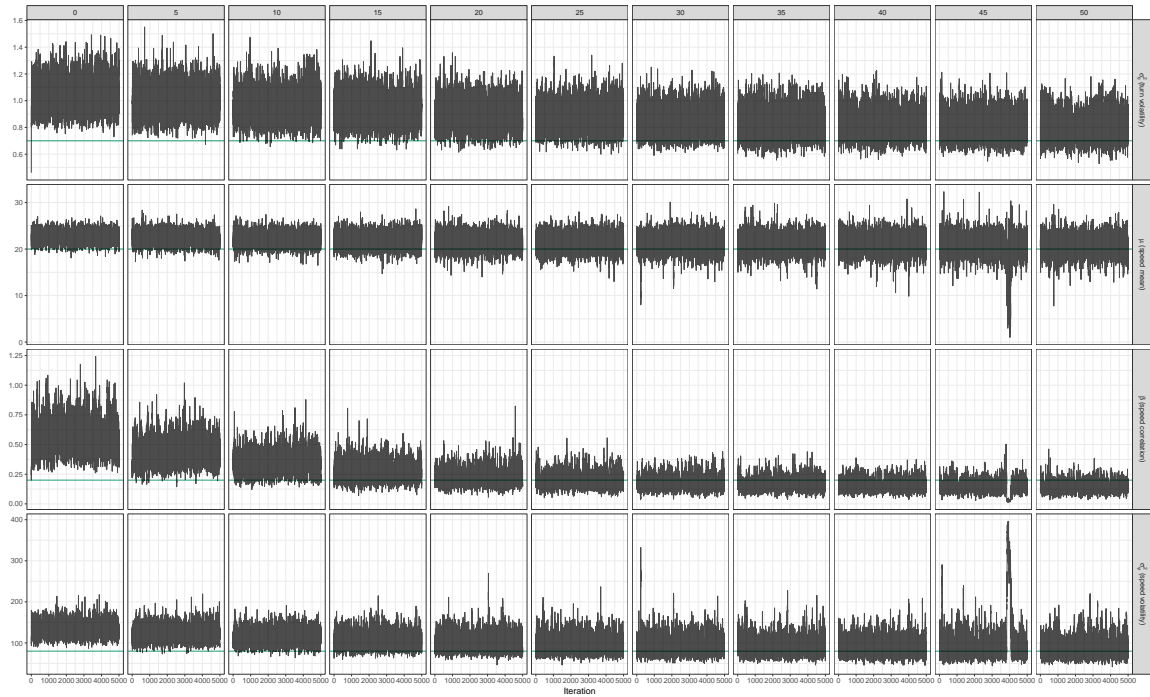


Fig. B.21 Sample trace of the posterior distribution of the movement parameters (**rows**) for the observation error simulation example, shown for each implementation (**columns**) that used a different level of fixed error variance. In each case, the true simulation value is given by the **horizontal line**.

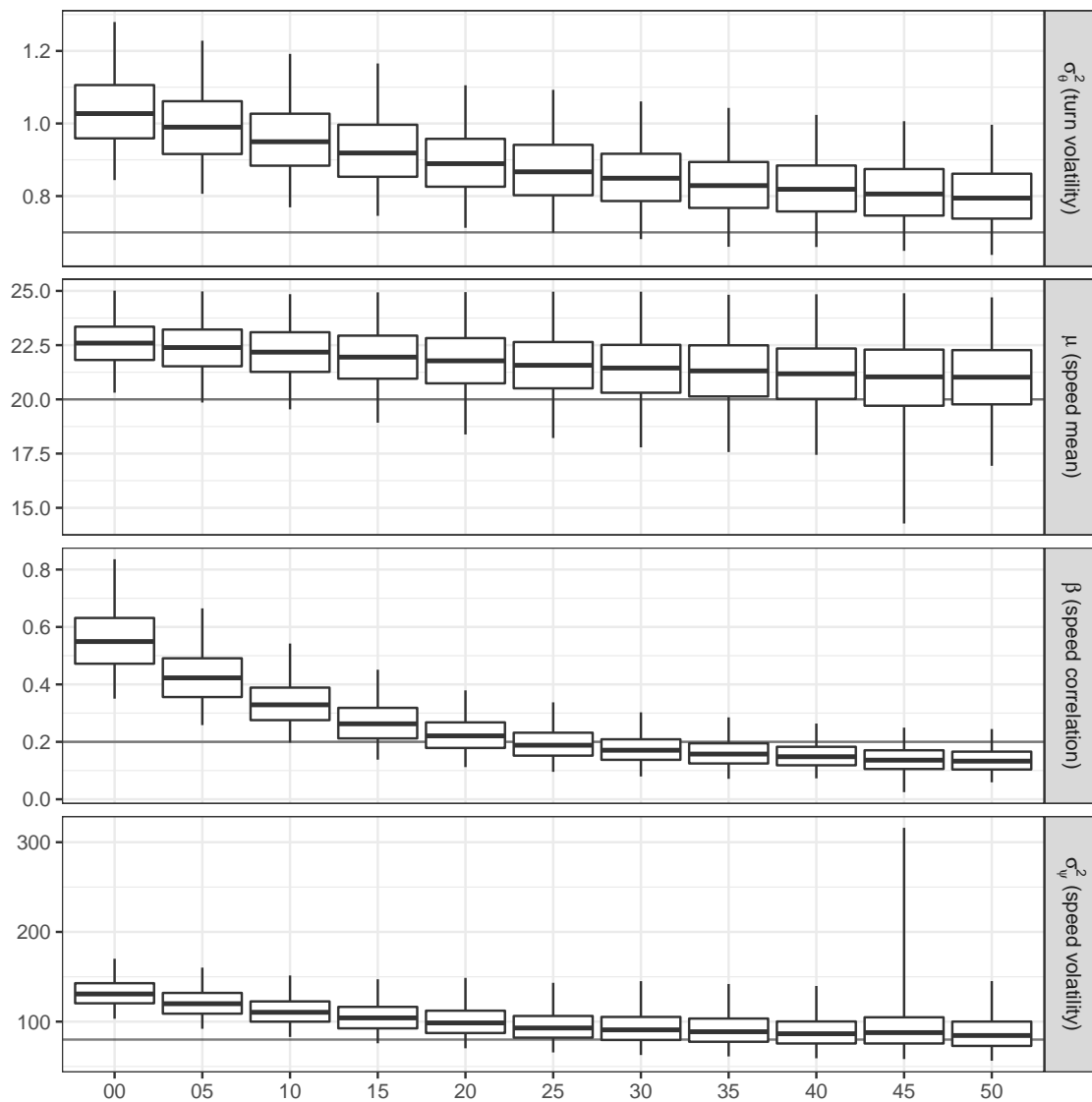


Fig. B.22 Boxplots showing the posterior distribution of the movement parameters for the observation error simulation example, shown for each implementation that used a different level of fixed error variance. In each case, the true simulation value is given by **grey horizontal line**.