State-switching continuous-time correlated random walks

Théo Michelot[1][*], Paul G. Blackwell[1]

[1]*University of Sheffield, UK*

Abstract

1. Continuous-time models have been developed to capture features of animal movement across temporal scales. In particular, one popular model is the continuous-time correlated random walk, in which the velocity of an animal is formulated as an Ornstein-Uhlenbeck process, to capture the autocorrelation in the speed and direction of its movement. In telemetry analyses, discrete-time state-switching models (such as hidden Markov models) have been increasingly popular to identify behavioural phases from animal tracking data.

2. We propose a multistate formulation of the continuous-time correlated random walk, with an underlying Markov process used as a proxy for the animal's behavioural state process. We present a Markov chain Monte Carlo algorithm to carry out Bayesian inference for this multistate continuous-time model.

3. Posterior samples of the hidden state sequence, of the state transition rates, and of the state-dependent movement parameters can be obtained. We investigate the performance of the method in a simulation study, and we illustrate its use in a case study of grey seal (Halichoerus grypus) tracking data.

4. The method we present makes use of the state-space model formulation of the continuous-time correlated random walk, and can accommodate irregular sampling frequency and measurement error. It will facilitate the use of continuous-time models to estimate movement characteristics and infer behavioural states from animal telemetry data.

Keywords: animal movement, continuous time, Ornstein-Uhlenbeck process, state-space model, multistate model, random walk

1   Introduction

The collection of large high-resolution animal tracking data sets has motivated the development of a wide range of statistical methods (Patterson et al., 2017; Hooten et al., 2017). For the analysis of animal movement, an important conceptual modelling choice is the time formulation, i.e. the choice between discrete-time and continuous-time models (McClintock et al., 2014). Although animals move in continuous time, their location may only be observed at discrete intervals (e.g. every minute or every hour). Dicrete-time approaches are based on the assumption that the underlying movement process can be appropriately modelled at the time scale of the observations. Most often, movement is described by the "step lengths" (distances between successive locations) and "turning angles" (angles between successive directions), derived from the location data (Morales et al., 2004; Jonsen et al., 2005). However, the distributions of these metrics of movement strongly depend on the sampling rate, such that the resulting inference is tied to a specific temporal scale (Codling and Hill, 2005; Schlägel and Lewis, 2016). One of the consequences is that discrete-time methods require locations to be collected at regular time intervals through the period of the study. Many telemetry data sets are collected at irregular time intervals, for example with marine mammals which may only be observed when they surface. To use a discrete-time model in such cases, it is then necessary to interpolate the data points on a regular time grid, introducing approximation uncertainty.

---

[*]Corresponding author: tmichelot1@sheffield.ac.uk

On the other hand, continuous-time models consider that telemetry observations arise from a continuous movement process. As such, they can naturally accommodate different temporal scales, and irregular sampling rates (Patterson et al., 2017). Various approaches have been used to model animal movement in continuous time, most of them based on diffusion processes. These include Ornstein-Uhlenbeck processes (Dunn and Gipson, 1977; Blackwell, 1997, 2003), Brownian bridges (Horne et al., 2007), and potential functions (Preisler et al., 2013). A very popular model is the continuous-time correlated random walk (CTCRW) introduced by Johnson et al. (2008), in which the velocity of the animal is formulated as an Ornstein-Uhlenbeck process. Through the velocity, this model incorporates autocorrelation into both the speed and the direction of the movement, similarly to discrete-time correlated random walks based on step lengths and turning angles. Johnson et al. (2008) formulated the CTCRW as a state-space model, making fast inference possible through the Kalman filter, and made it available in the R package crawl (Johnson, 2017). Fleming et al. (2017) extended this implementation to a wider family of diffusion processes, including their "OUF" model of correlated movement around a centre of attraction. Gurarie et al. (2017) recently reviewed the use of the CTCRW model for the analysis of animal tracking data.

Random walks have been used as "building blocks" for more complex, multistate, models. These state-switching models describe animal movements as the outcome of several distinct behaviours, e.g. "foraging", "resting", "exploring", based on the notion that the behavioural states of the animal differ noticeably in terms of some metrics of the movement, e.g. speed or sinuosity (Morales et al., 2004). Although this idea has received a lot of attention for discrete-time models, with the growing popularity of hidden Markov models (Patterson et al., 2009; Langrock et al., 2012), it has been underutilised in continuous-time approaches. Blackwell (1997) introduced a multistate movement model, where the location of the animal is modelled with an Ornstein-Uhlenbeck process. That model does not directly capture the movement persistence in speed and direction, which makes its application limited for high-frequency tracking data. More recently, Parton and Blackwell (2017) described a multistate approach in which the speed and the bearing of the animal are modelled with diffusion processes, analogously to discrete-time models based on step lengths and turning angles. However, their method requires computationally-costly numerical approximation to reconstruct the movement path at a fine time scale, a disadvantage in dealing with large tracking data sets. McClintock et al. (2014) presented a multistate analysis based on the CTCRW, but they constrained the state process to be constant over each time interval between two observations. Therefore, they do not carry out exact inference from the continuous-time model.

Alternatively, to circumvent the limitations of discrete-time models, McClintock (2017) suggested a two-stage approach based on multiple imputation methods. A one-state continuous-time model (such as the CTCRW) is fitted to the data, and a large number m of possible realisations of the movement process are simulated from the model on a regular time grid. Then, a hidden Markov model is fitted to each realisation, to investigate the state-switching dynamics. The m sets of estimates are pooled, such that the resulting model takes into account the uncertainty in the locations. Note that, since the realisations are generated without taking into account the possible behaviours, this is not fully equivalent to fitting a multistate CTCRW model.

Here, we extend the framework of Johnson et al. (2008) to incorporate behavioural states directly into the CTCRW framework, with an underlying continuous-time Markov process. We present the model formulation and describe a Bayesian estimation method to infer hidden states and movement parameters in this framework. We investigate the performance of the method in a simulation study, and then use it to analyse a trajectory of grey seal (*Halichoerus grypus*).

2 Model formulation

2.1 Continuous-time correlated random walk

The continuous-time correlated random walk (CTCRW) was introduced as a model of animal movement by Johnson et al. (2008). The underlying stochastic process was originally developed by Uhlenbeck and Ornstein (1930) to describe the movement of a particle under friction.

Denote $\boldsymbol{z}_t = (x_t, y_t)^\top$ the location of the animal at time t, and $\boldsymbol{v}_t = (v_t^x, v_t^y)^\top$ its velocity, linked by the equation

$$d\boldsymbol{z}_t = \boldsymbol{v}_t dt. \tag{1}$$

In the CTCRW model, the velocity of the animal is modelled by an Ornstein-Uhlenbeck process, defined as the solution of the stochastic differential equation

$$dv_t = \beta(\gamma - z_t)dt + \sigma dw_t, \tag{2}$$

where w_t denotes a Wiener process, $\gamma \in \mathbb{R}^2$ is the mean velocity, $\beta > 0$ measures the reversion of the velocity to its mean, and $\sigma > 0$ measures the spread of the velocity around its mean. In practice, the mean velocity parameter γ is often taken to be zero, corresponding to the case where there is no systematic drift in the animal's movement (although see Johnson et al., 2008, for an example of analysis with drift). Here, for simplicity, we consider that β and σ are scalar parameters, corresponding to the isotropic case, but they could be taken as matrices for a more general formulation (Blackwell, 2003; Gurarie et al., 2017). We will sometimes refer to the location process z_t as an integrated Ornstein-Uhlenbeck process, to indicate that its derivative (with respect to time) is an Ornstein-Uhlenbeck process.

This formulation is very convenient because it is possible to derive the transition densities of the velocity process v_t and of the position process x_t analytically. In the following, we assume $\gamma = 0$. Using Itô calculus, it can be shown that, under Equations 1 and 2, we have

$$v_{t+\delta} = e^{-\beta\delta}v_t + \zeta(\delta),$$

and

$$z_{t+\delta} = z_t + \left(\frac{1 - e^{-\beta\delta}}{\beta}\right)v_t + \xi(\delta),$$

for any time interval $\delta > 0$ and, for either dimension $c \in \{x, y\}$,

$$\zeta_c(\delta) \sim N\left[0, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta\delta})\right], \tag{3}$$

$$\xi_c(\delta) \sim N\left[0, \left(\frac{\sigma}{\beta}\right)^2\left(\delta + \frac{1 - e^{-2\beta\delta}}{2\beta} - \frac{2(1 - e^{-\beta\delta})}{\beta}\right)\right]. \tag{4}$$

The steps of the calculation are detailed in Appendix S1.

2.2 Multistate model

In this paper, we use the CTCRW as a building block for more complex and realistic movement models. Multistate models of animal movement have been developed to account for behavioural heterogeneity. In the most common formulation, a (discrete- or continuous-time) Markov process models switches between discrete "behavioural" states, on which depend the parameters of the movement process (Blackwell, 1997; Morales et al., 2004). Following this idea, we introduce a N-state continuous-time Markov process $(s_t)_{t\geq 0}$, characterised by its infinitesimal generator matrix,

$$\Lambda = \begin{pmatrix} -\lambda_1 & \lambda_{12} & \cdots & \lambda_{1N} \\ \lambda_{21} & -\lambda_2 & \cdots & \lambda_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{N1} & \lambda_{N2} & \cdots & -\lambda_N \end{pmatrix}, \tag{5}$$

where $\forall i \in 1, \ldots, N$, $\lambda_i = \sum_{j\neq i}\lambda_{ij}$. At any time $t \geq 0$, the discrete state s_t takes one of N values $\{1, \ldots, N\}$, typically used as proxies for the behavioural states of the animal (e.g. "foraging", "exploring"). The generator matrix is the continuous-time analogue of the transition probability matrix used in hidden Markov models, and its entries determine the state-switching dynamics. In particular, as a consequence of the Markov property, the dwell times in state i follow an exponential distribution with rate λ_i.

We now consider that the parameters of the CTCRW model (β and σ in Equation 2) are state-dependent, so that each behavioural state can be associated with a different type of movement. Using the notation introduced in Section 2.1, the multistate CTCRW model is defined by

$$\begin{cases} dz_t = v_t dt, \\ dv_t = -\beta_{s_t}v_t dt + \sigma_{s_t}dw_t. \end{cases}$$

3

This can be viewed as a higher-dimension continuous-time Markov process composed of a continuous component (the location and velocity processes) and a discrete component (the discrete state process), as described e.g. by Berman (1994). In the following, we develop a method of Bayesian inference for the multistate CTCRW model.

3 Bayesian inference

3.1 Likelihood evaluation with Kalman filter

We present a method to evaluate the likelihood of the multistate CTCRW model, given a reconstruction of the underlying state process. Johnson et al. (2008) implemented a Kalman filter to obtain the likelihood of movement trajectories in the single-state CTCRW model. With only minor changes, it can be extended to evaluate the likelihood of the multistate model, conditionally on the state sequence. The values of the state process are typically not known, and will thus need to be imputed. One method of reconstructing the state sequence is presented in Section 3.2.

We consider a data set of observed locations, augmented with the times of the reconstructed state transitions. The locations associated with the transitions are generally not observed, and they are thus treated as missing data. We denote by $\{\tilde{z}_1, \ldots, \tilde{z}_n\}$ the augmented sequence of locations, possibly observed with measurement error, and $\{t_1, \ldots, t_n\}$ the associated times. We denote by s_i the (imputed) behavioural state between t_i and t_{i+1}.

Following Johnson et al. (2008), the model can be written as a state-space model, where the observation process is the animal's location (possibly observed with noise), and the continuous state process is composed of both the true location and the velocity. We denote by $\boldsymbol{\omega}_i = (x_i, v_i^x, y_i, v_i^y)^\top$ the continuous state vector at time t_i, $\Delta_i = t_{i+1} - t_i$ the time intervals, and $\boldsymbol{\xi}_i = \boldsymbol{\xi}(\Delta_i)$ and $\boldsymbol{\zeta}_i = \boldsymbol{\zeta}(\Delta_i)$ the stochastic terms of the transition densities for the location and the velocity, respectively (Equations 3 and 4, substituting β_{s_i} for β and σ_{s_i} for σ).

Both s_i and $\boldsymbol{\omega}_i$ are referred to as "state", in state-switching models and state-space models, respectively. Here, we combine both and, to avoid confusion, we will refer to $\boldsymbol{\omega}_i$ as the "continuous" state of the model (as opposed to the "discrete" behavioural state s_i).

The observation equation of the CTCRW is

$$\tilde{z}_i = \boldsymbol{Z}\boldsymbol{\omega}_i + \boldsymbol{\varepsilon}_i, \qquad \boldsymbol{\varepsilon}_i \sim N(\boldsymbol{0}, \boldsymbol{H}_i),$$

where $\boldsymbol{H}_i$ is the 2×2 measurement error covariance matrix, and

$$\boldsymbol{Z} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

That is, the observed location $\tilde{z}_i$ is the sum of the true location $\boldsymbol{z}_i = (x_i, y_i)$ and an error term $\boldsymbol{\varepsilon}_i$. Using block matrix notation, the continuous state equation is

$$\boldsymbol{\omega}_{i+1} = \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix} \boldsymbol{\omega}_i + \boldsymbol{\eta}_i, \qquad \eta_i \sim N\left[\boldsymbol{0}, \begin{pmatrix} \boldsymbol{Q}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_i \end{pmatrix}\right]$$

where

$$\boldsymbol{T}_i = \begin{pmatrix} 1 & (1 - e^{-\beta_{s_i}\Delta_i})/\beta_{s_i} \\ 0 & e^{-\beta_{s_i}\Delta_i} \end{pmatrix}, \quad \boldsymbol{Q}_i = \begin{pmatrix} \text{Var}(\xi_{ci}) & \text{Cov}(\xi_{ci}, \zeta_{ci}) \\ \text{Cov}(\xi_{ci}, \zeta_{ci}) & \text{Var}(\zeta_{ci}) \end{pmatrix}.$$

The variances are given in Equation 3 and 4. The derivation of the covariance is given in Appendix S1, and yields

$$\text{Cov}(\xi_{ci}, \zeta_{ci}) = \frac{\sigma_{s_i}^2}{2\beta_{s_i}^2} \left(1 - 2e^{-\beta_{s_i}\Delta_i} + e^{-2\beta_{s_i}\Delta_i}\right).$$

Under this state-space model formulation, the Kalman filter can be used to obtain the log-likelihood of observed locations $\{\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_n\}$. In addition, the Kalman filter and smoother provide an estimate $\hat{\boldsymbol{\omega}}_i$ of the continuous state at each time step, as well as the covariance matrix $\hat{\boldsymbol{\Sigma}}_i$ of the estimate, conditional on all observations (Johnson et al., 2008). Appendix S2 gives the Kalman filter and smoother equations for the model, and the expression of the log-likelihood. For more details on state-space models and the Kalman filter, see e.g. Durbin and Koopman (2012).

3.2   MCMC algorithm

We use a Markov chain Monte Carlo (MCMC) algorithm to carry out inference on the parameters of the multistate CTCRW model introduced in Section 2.2, following the Metropolis-within-Gibbs approach introduced by Blackwell (2003). It relies on the successive updates of the three components of the model: the reconstructed behavioural state process, the movement parameters, and the transition rates. We denote $p(\tilde{z}|\boldsymbol{\theta}, \mathcal{S})$ the likelihood of a sequence of observed locations $\tilde{z}$, given the movement parameters $\boldsymbol{\theta}$ and the reconstructed state sequence $\mathcal{S}$, as obtained from the Kalman filter presented in Section 3.1.

We initialise the state sequence to $\mathcal{S}^{(0)}$, and the movement parameters to $\boldsymbol{\theta}^{(0)}$. Then, for K iterations $(k = 1, \ldots, K)$, we run the three following steps alternately.

Update of the behavioural state process   The evaluation of the likelihood of the model, described in Section 3.1, is conditional on the sequence of underlying states. In practice, the states are generally not observed, such that we need to impute them. The sequence of states $\mathcal{S}$ is composed of the times of the state transitions, and the values of the states. At each iteration, an updated state sequence $\mathcal{S}^*$ is proposed as follows. We choose an interval $[t_a, t_b]$, where a and b are two integers such that $t_1 \le t_a < t_b \le t_n$. We simulate the state process s_t from t_a to t_b, conditional on s_{t_a} and s_{t_b}, e.g. using the endpoint-conditioned continuous-time Markov process simulation methods from Hobolth and Stone (2009). The proposed sequence of states $\mathcal{S}^*$ remains identical to $\mathcal{S}^{(k-1)}$ outside $[t_a, t_b]$. The acceptance ratio for $\mathcal{S}^*$ is

$$r = \frac{p(\tilde{z}|\boldsymbol{\theta}^{(k-1)}, \mathcal{S}^*)}{p(\tilde{z}|\boldsymbol{\theta}^{(k-1)}, \mathcal{S}^{(k-1)})}.$$

The proposed state process reconstruction is accepted with probability $\min(1, r)$. The length of the interval $[t_a, t_b]$ over which the state sequence is updated is a tuning parameter of the sampler, and updates over longer intervals are generally less likely to be accepted.

Update of the movement parameters   Denote $\boldsymbol{\theta}$ the vector of parameters of the movement process, i.e. $\boldsymbol{\theta} = (\beta_1, \ldots, \beta_N, \sigma_1, \ldots, \sigma_N)$ for a N-state model. We use a Metropolis-Hastings step to update the movement parameters. At iteration k, we propose new movement parameters $\boldsymbol{\theta}^*$, from a proposal density $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(k-1)})$, and the acceptance ratio is

$$r = \frac{p(\tilde{z}|\boldsymbol{\theta}^*, \mathcal{S}^{(k)})q(\boldsymbol{\theta}^{(k-1)}|\boldsymbol{\theta}^*)}{p(\tilde{z}|\boldsymbol{\theta}, \mathcal{S}^{(k)})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(k-1)})}.$$

The parameters are updated to $\boldsymbol{\theta}^*$ with probability $\min(1, r)$. Note that, if the proposal distribution is symmetric such that $\forall \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, q(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) = q(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1)$, then r simplifies to

$$r = \frac{p(\tilde{z}|\boldsymbol{\theta}^*, \mathcal{S}^{(k)})}{p(\tilde{z}|\boldsymbol{\theta}^{(k-1)}, \mathcal{S}^{(k)})}.$$

In practice, a standard choice is to use a multivariate normal proposal distribution on the working scale of the parameters (in this case, on the log scale). Its variance can be tuned to obtain different acceptance rates, and covariance structure can be added to explore the parameter space more efficiently.

Update of the transition rates   Following Blackwell (2003), using conjugate priors, the transition rates can be directly sampled from their posterior distribution, which is known conditionally on the reconstructed state sequence $\mathcal{S}$. We find it convenient to parametrise the generator matrix as

$$\boldsymbol{\Lambda} = \begin{pmatrix} -\lambda_1 & \lambda_1 p_{12} & \cdots & \lambda_1 p_{1N} \\ \lambda_2 p_{21} & -\lambda_2 & \cdots & \lambda_2 p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_N p_{N1} & \lambda_N p_{N2} & \cdots & -\lambda_N \end{pmatrix}$$

where $\lambda_i > 0$ is the rate of transition out of state i, and the $p_{ij} \in [0, 1]$ are the transition probabilities out of state i. For each state i, they satisfy $\sum_j p_{ij} = 1$. The transition rates and the transition probabilities can be sampled separately.

For each $i \in \{1, \ldots, N\}$, we denote n_i the number of time intervals spent in state i, and $(\tau_i^{(1)}, \tau_i^{(2)}, \ldots, \tau_i^{(n_i)})$ their lengths. These dwell times are exponentially distributed with rate λ_i. The conjugate prior of the exponential distribution is the gamma distribution such that, with the prior

$$\lambda_i \sim \text{gamma}(\alpha_1, \alpha_2),$$

the transition rates are sampled from the posterior distribution

$$\lambda_i | \mathcal{S} \sim \text{gamma}\left(\alpha_1 + n_i, \ \alpha_2 + \sum_{j=1}^{n_i} \tau_i^{(j)}\right).$$

For $i \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, N\}$ such that $i \neq j$, we denote n_{ij} the number of transitions from state i to state j, and $n_i = \sum_j n_{ij}$ the number of transitions out of state i. Then,

$$n_{i1}, n_{i2}, \ldots, n_{iN} \sim \text{multinom}(n_i, p_{i1}, p_{i2}, \ldots, p_{iN}).$$

The conjugate prior of the multinomial distribution is the Dirichlet distribution such that, with the prior

$$p_{i1}, p_{i2}, \ldots, p_{iN} \sim \text{Dir}(\kappa_{i1}, \kappa_{i2}, \ldots, \kappa_{iN}),$$

the posterior distribution of the transition probabilities is

$$p_{i1}, p_{i2}, \ldots, p_{iN} | \mathcal{S} \sim \text{Dir}(\kappa_{i1} + n_{i1}, \kappa_{i2} + n_{i2}, \ldots, \kappa_{iN} + n_{iN}).$$

4 Simulation study

We simulated a trajectory from the multistate CTCRW model described in Section 2.2, and used the MCMC algorithm described in Section 3 to infer the hidden state sequence and the movement parameters from the simulated data.

4.1 Simulated data

We simulated 10,000 locations from a 2-state model at a fine time scale (every 0.1 time unit), and thinned them by keeping 10% of the points at random (i.e. 1000 irregularly-spaced locations), to emulate real movement data. The time intervals in the resulting data set ranged between 0.1 and 8 time units.

The movement parameters and switching rates of the simulated process were chosen as

$$(\beta_1, \beta_2) = (1, 0.3), \quad (\sigma_1, \sigma_2) = (1, 3), \quad \mathbf{\Lambda} = \begin{pmatrix} -0.03 & 0.03 \\ 0.03 & -0.03 \end{pmatrix}.$$

In state 1, the variance was smaller and the reversion to the mean larger, which resulted in slower and more sinuous movement (perhaps analogous to "area-restricted search" behaviour). State 2 corresponded to faster and more directed movement. The transition rates were chosen such that the process would on average stay about 30 time units in a state before switching to the other state. The simulated track (after thinning) is shown in Figure 1.

4.2 Estimation

We initialised the reconstructed state sequence by classifying each observation randomly as being in state 1 or state 2, with probability 0.5 each. At each iteration, the state process was updated over a randomly-selected interval of five time steps. We used independent normal proposal distributions (on the working log scale) to update the movement parameters. The proposal variances were tuned based on initial test runs, to obtain near-optimal acceptance rates. We chose normal prior distributions on the working scale for the movement parameters, centred on the true values of the parameters, and with large variances.

We ran 10^5 MCMC iterations, which took around 20 min on an 2GHz i5 CPU, and discarded the first 5×10^4 as burn-in. Figure 2 shows the posterior probabilities of being in state 2 at the times of the
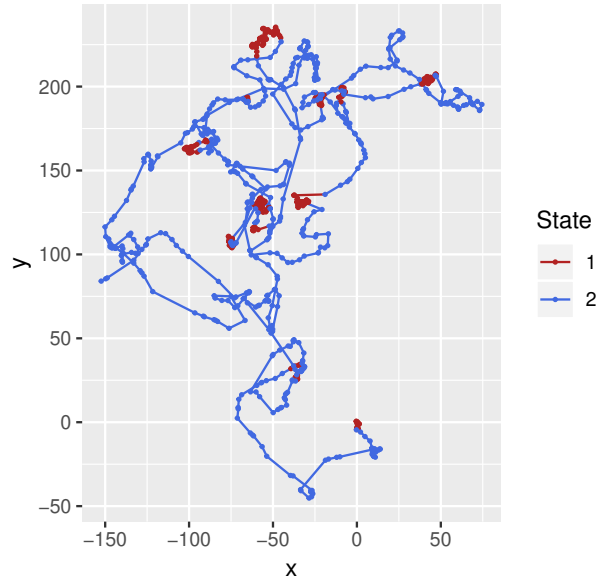
Figure 1: *Track simulated from a 2-state CTCRW model.*

observations, to compare with the "true" simulated state sequence. For each $i = 1, \ldots, n$, we calculated the posterior probability as the proportion of reconstructions of the state process in which $\boldsymbol{x}_i$ was classified in state 2. (It would have been equivalent to consider the posterior probability of being in state 1; the two probabilities sum to 1.) We considered that a step was misclassified if the posterior probability of being in the true state was less than 0.5. The states were correctly estimated in the vast majority of time steps, with only 2% of misclassified steps.
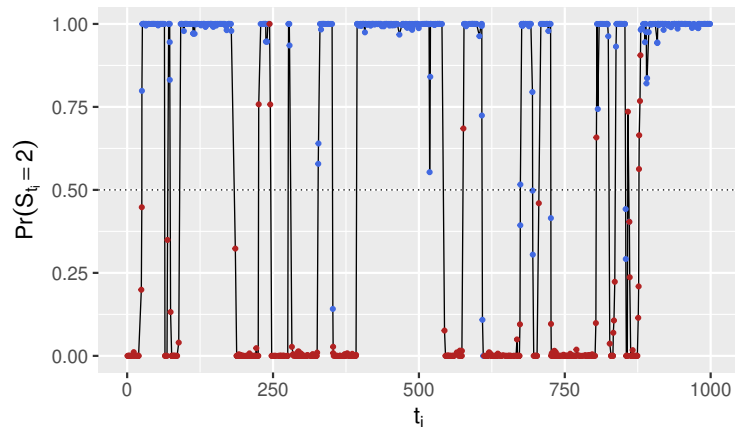

Figure 2: *Posterior probabilities of being in state 2 at the times of the observations. The true (simulated) states are shown by the colours (red: state 1, blue: state 2).*

Figure 3 displays posterior samples for the state-dependent movement parameters, β_1, β_2, σ_1, and σ_2, as well as the true parameter values used in the simulation. The posterior distributions seem to appropriately estimate all movement parameters.

We were able to recover the values of the state process and of the state-dependent movement parameters from a simulated track thinned to irregular intervals. This demonstrates the ability of the method to work across temporal scales, and to cope with irregular sampling.
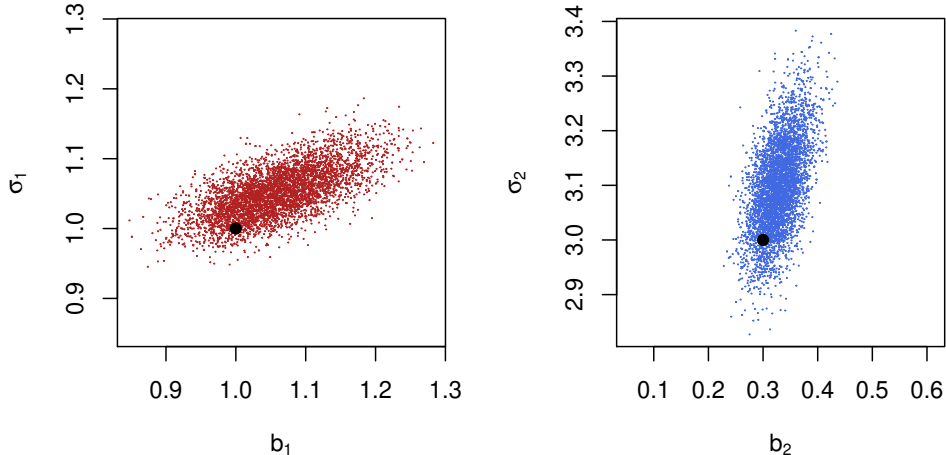
7

Figure 3: *Posterior samples of the movement parameters in state 1 (left) and in state 2 (right), in the simulation study. The black dots are the true values of the parameters, used in the simulation. The samples are thinned to every tenth value, for visualisation purposes.*

5 Grey seal case study

We illustrate the use of the method described in Section 3 for the analysis of a grey seal (*Halichoerus grypus*) movement track. We considered a trajectory of 1875 observations, collected in the North Sea between April and August 2008, and previously described by Russell et al. (2015). The base sampling frequency was of one location every 30 minutes, but many fixes were missed, and the resulting time grid was highly irregular ($P_{0.025} = 27$ min, $P_{0.975} = 8$ hours). Note that the CTCRW model describes movement on a plane, and thus requires that the longitude-latitude locations be projected to UTM coordinates for the analysis.

We considered a 2-state CTCRW model, with four movement parameters to estimate: β_1, β_2, σ_1, σ_2. Similarly to the simulation study, we initialised the state reconstruction to a random sequence of 1s and 2s (with probability 0.5 each). We used independent normal proposal and prior distributions on the working scale of the movement parameters. We selected the proposal variances based on test runs, and used weakly informative prior distributions. We ran 2 million MCMC iterations, discarding the first half as burn-in. We only saved every 100th reconstructed state sequence, because of memory limitations.

Figure 4 shows a map of the track, coloured by posterior state probabilities, and Figure 5 shows posterior samples for the four movement parameters $(\beta_1, \beta_2, \sigma_1, \sigma_2)$. State 2 captured very directed movements, corresponding to periods of transit between areas of interest, and state 1 captured more tortuous phases of the track. This can be seen in Figure 5: the posterior distribution of β_1 covers much larger values than that of β_2 (posterior means of 1.73 and 0.06, respectively), indicating stronger reversion to the mean in state 1, and thus less movement persistence. There were no signs of label switching in the posterior samples; if there were, a straightforward solution would be to constrain (β_1, β_2) and (σ_1, σ_2) to be ordered (McClintock et al., 2014).

The Kalman filter and smoother recursions given in Appendix S2 can be used to compute estimated velocities at the times of the observations. The velocities obtained with the mean posterior movement parameters are displayed in Figure 6, and coloured by posterior state probabilities. The strong movement autocorrelation in state 2 can be seen in the outer rim of the plot, where the velocity sometimes persists with little variation over many time steps. The Kalman algorithm can also provide estimates of the locations (and possibly velocities) of the animal – and associated standard errors – on any time grid, e.g. on a finer time grid than that of the observations.

An alternative parametrisation of the CTCRW is defined with $\tau = 1/\beta$ and $\nu = \sqrt{\pi}\sigma/(2\sqrt{\beta})$ (Gurarie et al., 2017). The parameter τ is the time interval over which the autocorrelation function of the velocity process decreases by a factor e, such that larger values indicate longer-term movement persistence. The parameter ν is the mean speed of movement of the animal. We transformed the posterior samples of
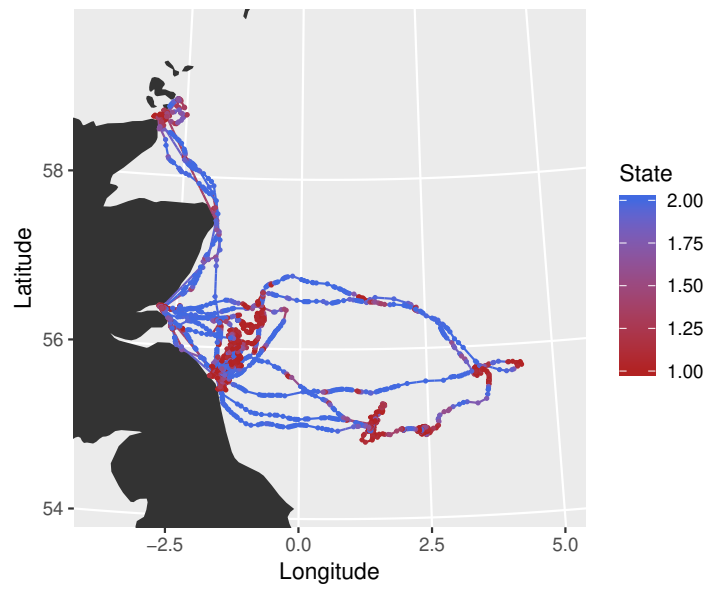
Figure 4: *Grey seal track, off the East coast of Great Britain, coloured by posterior state probabilities.*
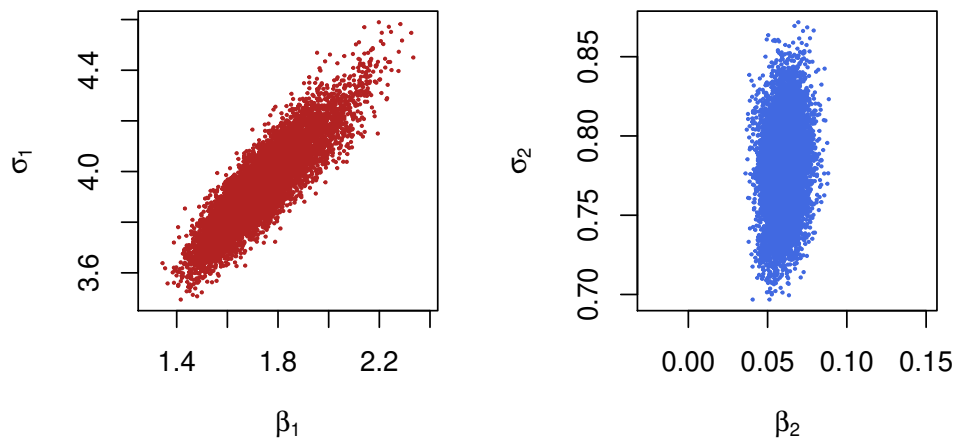

Figure 5: *Posterior samples of the movement parameters in state 1 (left) and in state 2 (right), in the grey seal case study. The samples are thinned to every 100th value, for visualisation purposes.*
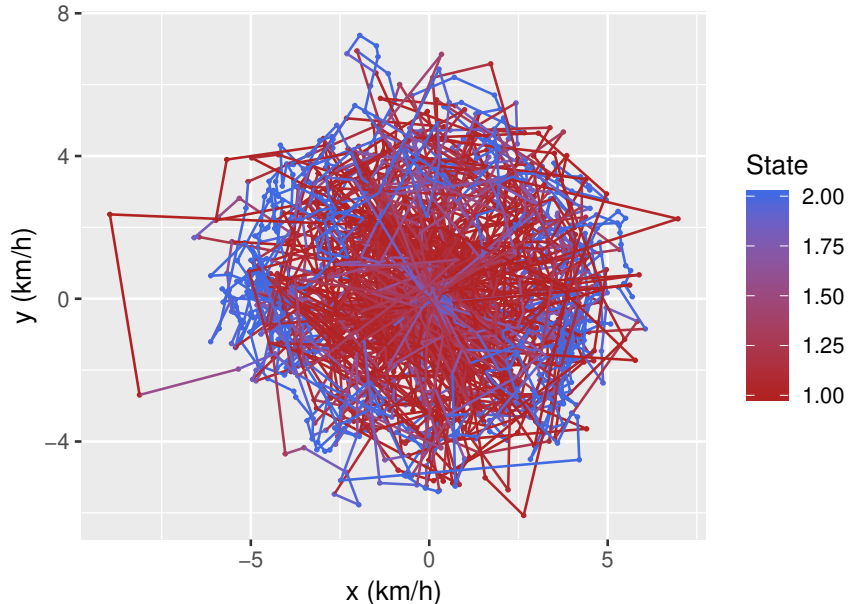
Figure 6: *Predicted velocities from the grey seal example, obtained with the mean posterior movement parameter estimates. The colour reflects the mean (over the sampled state sequences) of the states to which each observation was allocated.*

movement parameters to obtain estimates of τ and ν in both states. In the following, we report posterior mean estimates, and histograms of the posterior samples for the transformed parameters can be found in Appendix S3. The posterior means were $\hat{\tau}_1 = 0.55$h, $\hat{\tau}_2 = 17.21$h, $\hat{\nu}_1 = 2.63$km.h^{-1}, and $\hat{\nu}_2 = 2.90$km.h^{-1}. This indicates that the two states are very similar in terms of the speed of movement, but that the autocorrelation function of the velocity drops much faster in state 1 than in state 2.

The posterior samples of the transition rates can be used to derive mean dwell times in each state, and long-term activity budgets. The dwell times in state i follow an exponential distribution with rate λ_i (the rate of transition out of state i). The mean dwell time can thus be derived as $d_i = 1/\lambda_i$. In the grey seal analysis, the posterior means for the mean dwell time were $\hat{d}_1 = 7.5$h in state 1, and $\hat{d}_2 = 7.8$h in state 2, indicating similar dwell times in both states. Activity budgets refer to the proportion of time spent by an animal in each of its behavioural states (Russell et al., 2015). In a time-homogeneous state-switching model, an estimate of the long-term activity budget can be calculated as the stationary distribution of the underlying Markov process. The stationary distribution of a N-state Markov process is the vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)$ which satisfies $\boldsymbol{\pi\Lambda} = \mathbf{0}$, subject to the constraint $\sum_{i=1}^{N} \pi_i = 1$, where $\boldsymbol{\Lambda}$ is the generator matrix defined in Equation 5. In the 2-state case, solving the equation yields $\pi_1 = \lambda_2/(\lambda_1 + \lambda_2)$ and $\pi_2 = \lambda_1/(\lambda_1 + \lambda_2)$. The posterior mean estimate for the stationary distribution was $(\hat{\pi}_1, \hat{\pi}_2) = (0.48, 0.52)$, i.e. the seal will tend to spend roughly the same proportion of time in both states, in the long term. Histograms of posterior draws for the dwell times and stationary distribution are displayed in Appendix S3.

6   Discussion

We presented a Bayesian framework to infer discrete behavioural states and movement parameters from a multistate continuous-time model of animal movement. The continuous-time formulation can accommodate irregular time intervals, and is consistent across temporal scales. The conditional likelihood of the model, used in the MCMC algorithm, is implemented using the Kalman filter, making it relatively fast and allowing for the inclusion of measurement error.

The inferential approach introduced in this paper could in principle be used to implement a state-switching version of the OUF model described by Fleming et al. (2014). The OUF process is a generalisation of the CTCRW used in this paper, and of the Ornstein-Uhlenbeck location process used e.g. by Blackwell et al. (2016). It features both persistence in velocity and long-term attraction towards a point in space, making it a very flexible model of animal movement. Like the CTCRW, it can be written as a state-space model, and the Kalman filter can be used to derive the likelihood of the model (Fleming et al., 2017). The MCMC algorithm described in Section 3.2 could then be used to fit a multistate OUF model to animal movement. However, the OUF process has five parameters (against 2 only for the CTCRW), which could make estimation more challenging. More generally, this methodology could be applied to a model switching between processes with different formulations (e.g. a 2-state model switching between a CTCRW and OUF process).

Analyses of animal movement and behaviour often combine telemetry and environmental data. In state-switching models, the effect of environmental (or other) covariates on the transition probabilities is of particular interest, and is used to uncover the drivers of animal behavioural and movement decisions (Patterson et al., 2009; Bestley et al., 2012; Blackwell et al., 2016). The MCMC algorithm of Section 3.2 could be extended, following Blackwell et al. (2016), to allow for the inclusion of covariates in the state-switching dynamics.

Acknowledgements

References

Berman, S. M. (1994). A bivariate markov process with diffusion and discrete components. *Stochastic Models*, 10(2):271–308.

Bestley, S., Jonsen, I. D., Hindell, M. A., Guinet, C., and Charrassin, J.-B. (2012). Integrative modelling of animal movement: incorporating in situ habitat and behavioural information for a migratory marine predator. *Proceedings of the Royal Society of London B: Biological Sciences*, page rspb20122262.

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

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

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

Codling, E. and Hill, N. (2005). Sampling rate effects on measurements of correlated and biased random walks. *Journal of Theoretical Biology*, 233(4):573–588.

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

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*, volume 38. OUP Oxford.

Fleming, C. H., Calabrese, J. M., Mueller, T., Olson, K. A., Leimgruber, P., and Fagan, W. F. (2014). From fine-scale foraging to home ranges: a semivariance approach to identifying movement modes across spatiotemporal scales. *The American Naturalist*, 183(5):E154–E167.

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.

Gurarie, E., Fleming, C. H., Fagan, W. F., Laidre, K. L., Hernández-Pliego, J., and Ovaskainen, O. (2017). Correlated velocity models as a fundamental unit of animal movement: synthesis and applications. *Movement Ecology*, 5(1):13.

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. *The annals of applied statistics*, 3(3):1204.

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

Horne, J. S., Garton, E. O., Krone, S. M., and Lewis, J. S. (2007). Analyzing animal movements using brownian bridges. *Ecology*, 88(9):2354–2363.

Johnson, D. S. (2017). *crawl: Fit Continuous-Time Correlated Random Walk Models to Animal Movement Data*. R package version 2.1.1.

Johnson, D. S., London, J. M., Lea, M.-A., and Durban, J. W. (2008). Continuous-time correlated random walk model for animal telemetry data. *Ecology*, 89(5):1208–1215.

Jonsen, I. D., Flemming, J. M., and Myers, R. A. (2005). Robust state–space modeling of animal movement data. *Ecology*, 86(11):2874–2880.

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):2336–2342.

McClintock, B. T. (2017). Incorporating telemetry error into hidden markov models of animal movement using multiple imputation. *Journal of Agricultural, Biological and Environmental Statistics*, 22(3):249–269.

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):21.

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

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):373–392.

Patterson, T. A., Basson, M., Bravington, M. V., and Gunn, J. S. (2009). Classifying movement behaviour in relation to environmental conditions using hidden markov models. *Journal of Animal Ecology*, 78(6):1113–1123.

Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., and King, R. (2017). Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis*, 101(4):399–438.

Preisler, H. K., Ager, A. A., and Wisdom, M. J. (2013). Analyzing animal movement patterns using potential functions. *Ecosphere*, 4(3):1–13.

Russell, D. J., McClintock, B. T., Matthiopoulos, J., Thompson, P. M., Thompson, D., Hammond, P. S., Jones, E. L., MacKenzie, M. L., Moss, S., and McConnell, B. J. (2015). Intrinsic and extrinsic drivers of activity budgets in sympatric grey and harbour seals. *Oikos*, 124(11):1462–1472.

Schlägel, U. E. and Lewis, M. A. (2016). Robustness of movement models: can models bridge the gap between temporal scales of data sets and behavioural processes? *Journal of mathematical biology*, 73(6-7):1691–1726.

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of the brownian motion. *Physical review*, 36(5):823.

Appendix S1. CTCRW transition density

The CTCRW, or integrated Ornstein-Uhlenbeck process $(\boldsymbol{z}_t)_{t>0}$, is defined as the solution to

$$\begin{cases} d\boldsymbol{z}_t = \boldsymbol{v}_t dt, \\ d\boldsymbol{v}_t = -\beta \boldsymbol{v}_t dt + \sigma d\boldsymbol{w}_t. \end{cases}$$

We consider here the isotropic case only, so it is sufficient to solve this system of equations in either dimension. In the following, we consider the univariate processes z_t, v_t, and w_t.

Solution to the Ornstein-Uhlenbeck equation

We first derive the solution for the equation of the velocity v_t,

$$dv_t = -\beta v_t dt + \sigma dw_t \tag{1}$$

Note that

$$d(e^{\beta t} v_t) = \beta e^{\beta t} v_t + e^{\beta t} dv_t \tag{2}$$

Then, from (1) and (2),

$$d(e^{\beta t} v_t) = e^{\beta t} \sigma dw_t$$

Integrating both sides between t and $t + \delta$,

$$e^{\beta(t+\delta)} v_{t+\delta} - e^{\beta t} v_t = \sigma \int_{s=t}^{t+\delta} e^{\beta s} dw_s,$$

and we obtain the solution

$$v_{t+\delta} = e^{-\beta\delta} v_t + \sigma \int_{s=t}^{t+\delta} e^{-\beta(t+\delta-s)} dw_s. \tag{3}$$

Denote $\zeta(\delta) = \sigma \int_{s=t}^{t+\delta} e^{-\beta(t+\delta-s)} dw_s$. By property of the Itô integral, $\zeta(\delta)$ is a Gaussian random variable with mean 0 and variance

$$\text{Var}(\zeta(\delta)) = E\left[\zeta(\delta)^2\right] = \sigma^2 E\left[\left(\int_{s=t}^{t+\delta} e^{-\beta(t+\delta-s)} dw_s\right)^2\right]$$

Using the Itô isometry,

$$\text{Var}(\zeta(\delta)) = \sigma^2 \int_{s=t}^{t+\delta} e^{-2\beta(t+\delta-s)} ds$$

$$= \sigma^2 \left[\frac{e^{-2\beta(t+\delta-s)}}{2\beta}\right]_{s=t}^{t+\delta}$$

$$= \frac{\sigma^2}{2\beta}(1 - e^{-2\beta\delta})$$

Eventually,

$$\zeta(\delta) \sim N\left(0, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta\delta})\right)$$

Solution to the integrated Ornstein-Uhlenbeck equation

We can now solve for the location process z_t. Integrating both sides of $dz_t = v_t dt$ between t and $t + \delta$, and using the solution found in Equation 3, we have

$$z_{t+\delta} - z_t = \int_{s=t}^{t+\delta} v_s ds = \int_{s=t}^{t+\delta} \left(e^{-\beta(s-t)} v_t + \sigma \int_{u=t}^{s} e^{-\beta(s-u)} dw_u \right) ds$$

Thus,

$$
\begin{aligned}
z_{t+\delta} &= z_t + v_t \int_{s=t}^{t+\delta} e^{-\beta(s-t)} ds + \sigma \int_{s=t}^{t+\delta} \int_{u=t}^{s} e^{-\beta(s-u)} dw_u ds \\
&= z_t + v_t \left[-\frac{e^{-\beta(s-t)}}{\beta} \right]_{s=t}^{t+\delta} + \sigma \int_{u=t}^{t+\delta} \int_{s=u}^{t+\delta} e^{-\beta(s-u)} ds dw_s \\
&= z_t + \left(\frac{1 - e^{-\beta\delta}}{\beta} \right) v_t + \sigma \int_{u=t}^{t+\delta} \left[-\frac{e^{-\beta(s-u)}}{\beta} \right]_{s=u}^{t+\delta} dw_u \\
&= z_t + \left(\frac{1 - e^{-\beta\delta}}{\beta} \right) v_t + \frac{\sigma}{\beta} \int_{u=t}^{t+\delta} (1 - e^{-\beta(t+\delta-u)}) dw_u
\end{aligned}
$$

The location process z_t also has a normal transition density, with mean

$$E(z_{t+\delta} | z_t, v_t) = z_t + \left(\frac{1 - e^{-\beta\delta}}{\beta} \right) v_t.$$

Denote $\xi(\delta)$ the Gaussian error term

$$\xi(\delta) = \frac{\sigma}{\beta} \int_{u=t}^{t+\delta} (1 + e^{-\beta(t+\delta-u)}) dw_u.$$

The variance of the transition density of the location process is given by $\mathrm{Var}(\delta)$,

$$\mathrm{Var}(\xi(\delta)) = \left(\frac{\sigma}{\beta} \right)^2 E \left[\left(\int_{u=t}^{t+\delta} (1 - e^{-\beta(t+\delta-u)}) dw_u \right)^2 \right].$$

Using the Itô isometry,

$$
\begin{aligned}
\mathrm{Var}(\xi(\delta)) &= \left(\frac{\sigma}{\beta} \right)^2 \int_{u=t}^{t+\delta} (1 - e^{-\beta(t+\delta-u)})^2 du \\
&= \left(\frac{\sigma}{\beta} \right)^2 \int_{u=t}^{t+\delta} (1 + e^{-2\beta(t+\delta-u)} - 2e^{-\beta(t+\delta-u)}) du \\
&= \left(\frac{\sigma}{\beta} \right)^2 \left[u + \frac{e^{-2\beta(t+\delta-u)}}{2\beta} - \frac{2e^{-\beta(t+\delta-u)}}{\beta} \right]_{u=t}^{t+\delta} \\
&= \left(\frac{\sigma}{\beta} \right)^2 \left(\delta + \frac{1 - e^{-2\beta\delta}}{2\beta} - \frac{2(1 - e^{-\beta\delta})}{\beta} \right)
\end{aligned}
$$

Finally, we obtain

$$\xi(\delta) \sim N \left(0, \left(\frac{\sigma}{\beta} \right)^2 \left(\delta + \frac{1 - e^{-2\beta\delta}}{2\beta} - \frac{2(1 - e^{-\beta\delta})}{\beta} \right) \right)$$

Covariance of the location process and velocity process

To formulate the CTCRW as a state-space model, we also need the covariance of the location process and the velocity process, i.e. $\text{Cov}(\xi(\delta), \zeta(\delta))$.

$$\text{Cov}(\zeta(\delta), \xi(\delta)) = E(\zeta(\delta)\xi(\delta))$$

$$= E\left[\left(\sigma \int_{s=t}^{t+\delta} e^{-\beta(t+\delta-s)} dw_s\right)\left(\frac{\sigma}{\beta} \int_{s=t}^{t+\delta} (1 - e^{-\beta(t+\delta-s)}) dw_s\right)\right].$$

Using the Itô isometry,

$$\text{Cov}(\zeta(\delta), \xi(\delta)) = \frac{\sigma^2}{\beta} \int_{s=t}^{t+\delta} e^{-\beta(t+\delta-s)}(1 - e^{-\beta(t+\delta-s)}) ds$$

$$= \frac{\sigma^2}{\beta} \int_{s=t}^{t+\delta} (e^{-\beta(t+\delta-s)} - e^{-2\beta(t+\delta-s)}) ds$$

$$= \frac{\sigma^2}{\beta} \left[\frac{e^{-\beta(t+\delta-s)}}{\beta} - \frac{e^{-2\beta(t+\delta-s)}}{2\beta}\right]_{s=t}^{t+\delta}$$

$$= \frac{\sigma^2}{\beta} \left(\frac{1 - e^{-2\beta\delta}}{\beta} - \frac{1 - e^{-2\beta\delta}}{2\beta}\right)$$

$$= \frac{\sigma^2}{2\beta^2} \left(1 - 2e^{-\beta\delta} + e^{-2\beta\delta}\right)$$

Appendix S2. Kalman filtering and smoothing

Here we give the equations of the Kalman filter and smoother, for the multistate CTCRW model. These equations can be found for the single-state CTCRW in Appendix B of Johnson et al. (2008), and in the case of a general state-space model in Durbin and Koopman (2012).

Notation   We use the following notation:

- $\{\tilde{\boldsymbol{z}}_1, \ldots, \tilde{\boldsymbol{z}}_n\}$ are the observed locations, and $\{t_1, \ldots, t_n\}$ the corresponding times. Note that $\tilde{\boldsymbol{z}}_i$ is treated as missing data if t_i corresponds to a behavioural transition (see below for processing of missing data).

- $\tilde{\boldsymbol{z}}_i = (\tilde{x}_i, \tilde{y}_i)^\top$.

- s_i is the behavioural state between t_i and t_{i+1}.

Filtering   A standard choice to initialise the state estimate is $\hat{\boldsymbol{a}}_1 = (\tilde{x}_1, 0, \tilde{y}_1, 0)^\top$, i.e. the mean initial location is the first observation $\tilde{\boldsymbol{z}}_1$, and the mean initial velocity is $\boldsymbol{0}$. The estimate covariance matrix $\hat{\boldsymbol{P}}_1$ is typically taken to be diagonal, and its elements measure the uncertainty on the initial state estimate $\hat{\boldsymbol{a}}_1$.

Then, for $i = 1, 2, \ldots, n-1$, the Kalman filter equations are

- $\boldsymbol{u}_i = \tilde{\boldsymbol{z}}_i - \boldsymbol{Z}\hat{\boldsymbol{a}}_i$,

- $\boldsymbol{F}_i = \boldsymbol{Z}\hat{\boldsymbol{P}}_i\boldsymbol{Z}^\top + \boldsymbol{H}_i$,

- $\boldsymbol{K}_i = \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix}\hat{\boldsymbol{P}}_i\boldsymbol{Z}^\top\boldsymbol{F}_i^{-1}$,

- $\hat{\boldsymbol{a}}_{i+1} = \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix}\hat{\boldsymbol{a}}_i + \boldsymbol{K}_i\boldsymbol{u}_i$,

- $\hat{\boldsymbol{P}}_{i+1} = \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix}\hat{\boldsymbol{P}}_i\left[\begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix} - \boldsymbol{K}_i\boldsymbol{Z}^\top\right]^\top + \begin{pmatrix} \boldsymbol{Q}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_i \end{pmatrix}$,

with the model matrices

$$\boldsymbol{Z} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad \boldsymbol{T}_i = \begin{pmatrix} 1 & (1 - e^{-\beta_{s_i}\Delta_i})/\beta_{s_i} \\ 0 & e^{-\beta_{s_i}\Delta_i} \end{pmatrix}, \quad \boldsymbol{Q}_i = \begin{pmatrix} \mathrm{Var}(\xi_{ci}) & \mathrm{Cov}(\xi_{ci}, \zeta_{ci}) \\ \mathrm{Cov}(\xi_{ci}, \zeta_{ci}) & \mathrm{Var}(\zeta_{ci}) \end{pmatrix}.$$

Log-likelihood   Due to the assumptions of normality, the log-likelihood can be calculated as a by-product of the Kalman filter,

$$l(\tilde{\boldsymbol{z}}_1, \ldots, \tilde{\boldsymbol{z}}_n) = -n\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\left(\log|\boldsymbol{F}_i| + \boldsymbol{u}_i^\top\boldsymbol{F}_i^{-1}\boldsymbol{u}_i\right),$$

where $|\boldsymbol{F}_i|$ denotes the determinant of $\boldsymbol{F}_i$.

Smoothing   We initialise $\boldsymbol{r}_n = \boldsymbol{0}$ and $\boldsymbol{N}_n = \boldsymbol{0}$. Then, for $i = n, n-1, \ldots, 1$, the Kalman smoother equations are

- $\boldsymbol{L}_i = \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix} - \boldsymbol{K}_i\boldsymbol{Z}$,

- $\boldsymbol{r}_{i-1} = \boldsymbol{Z}^\top\boldsymbol{F}_i^{-1}\boldsymbol{u}_i + \boldsymbol{L}_i^\top\boldsymbol{r}_i$,

- $\boldsymbol{N}_{i-1} = \boldsymbol{Z}^\top\boldsymbol{F}_i^{-1}\boldsymbol{Z} + \boldsymbol{L}_i^\top\boldsymbol{N}_i\boldsymbol{L}_i$,

- $\hat{\boldsymbol{\omega}}_i = \hat{\boldsymbol{a}}_i + \hat{\boldsymbol{P}}_i\boldsymbol{r}_{i-1}$,

- $\hat{\boldsymbol{\Sigma}}_i = \hat{\boldsymbol{P}}_i - \hat{\boldsymbol{P}}_i\boldsymbol{N}_{i-1}\hat{\boldsymbol{P}}_i$.

Then, $\hat{\boldsymbol{\omega}}_i$ is the smoothed state vector at time t_i, and $\hat{\boldsymbol{\Sigma}}_i$ the smoothed state variance matrix.

Missing data The locations of the state transitions are generally not observed, and they must be treated as missing data. If the index i corresponds to a missing value, the filtering equations become

- $\hat{\boldsymbol{a}}_{i+1} = \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix} \hat{\boldsymbol{a}}_i,$

- $\hat{\boldsymbol{P}}_{i+1} = \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix} \hat{\boldsymbol{P}}_i \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix}^\top + \begin{pmatrix} \boldsymbol{Q}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_i \end{pmatrix},$

and the smoothing equations,

- $\boldsymbol{r}_{i-1} = \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix}^\top \boldsymbol{r}_i,$

- $\boldsymbol{N}_{i-1} = \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix}^\top \boldsymbol{N}_i \begin{pmatrix} \boldsymbol{T}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{T}_i \end{pmatrix}.$

The other equations are unchanged. Note also that missing observations do not contribute to the log-likelihood.

Appendix S3. Posterior samples for the grey seal analysis

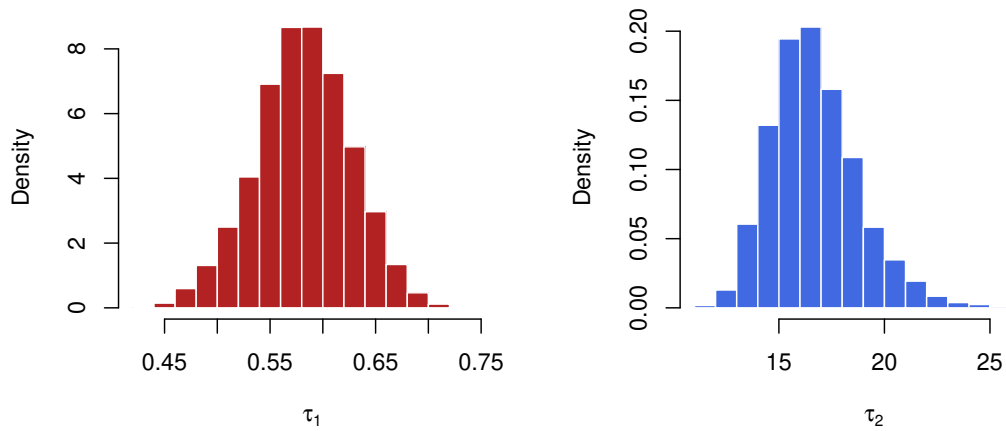Below are histograms of the posterior samples obtained for different model parameters in the grey seal case study.

Figure S1: *Histograms of posterior samples for the time scales of autocorrelation τ_1 and τ_2 in the grey seal case study.*
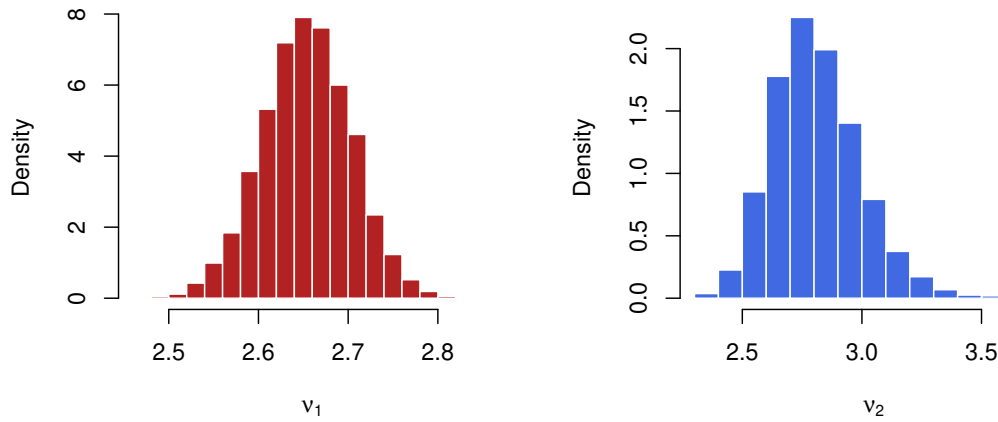

Figure S2: *Histograms of posterior samples for the mean speed parameters ν_1 and ν_2 in the grey seal case study.*
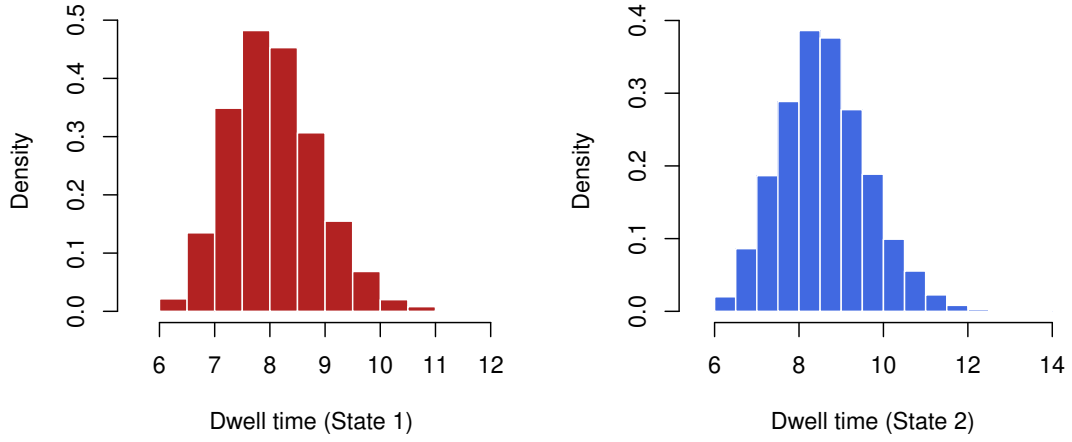
Figure S3: *Histograms of posterior samples for the mean dwell times in states 1 and 2, in the grey seal case study.*
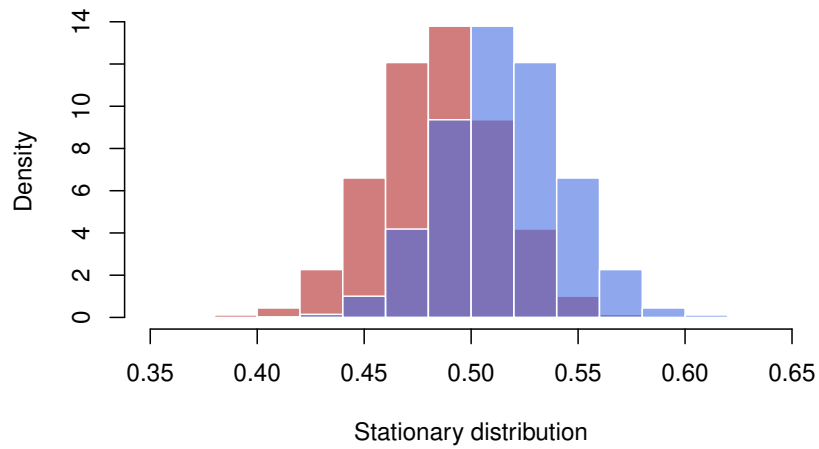

Figure S4: *Histograms of posterior samples for the stationary distribution of the Markov state process, in the grey seal case study. The red histogram shows the posterior sample for π_1 (the stationary probability of being in state 1), and the blue histogram shows the posterior sample for $\pi_2 = 1 - \pi_1$.*