BIOMETRICS ??, 1–**??**
? 2020

DOI: 10.1111/j.1541-0420.2005.00454.x

Modelling group movement with behavior switching in continuous time

Mu Niu[1,*], Fay Frost[2,], Jordan E. Milner[2,***] Anna Skarin[3,****] Paul G. Blackwell[2,*****],**

[1]School of Mathematics and Statistics, University of Glasgow, Glasgow GL 8QQ, U.K.

[2]School of Mathematics & Statistics, University of Sheffield, Sheffield S3 7RH, U.K.

[3]Department of Animal Nutrition and Management,

Swedish University of Agricultural Sciences, Uppsala, Sweden

**email:* mu.niu@glasgow.ac.uk

***email:* ffrost1@sheffield.ac.uk

****email:* jemilner1@sheffield.ac.uk

*****email:* anna.skarin@slu.se

******email:* p.blackwell@sheffield.ac.uk

SUMMARY: This article presents a new method for modelling collective movement in continuous time with behavioural switching, motivated by simultaneous tracking of wild or semi-domesticated animals. Each individual in the group is at times attracted to a unobserved leading point. However the behavioural state of each individual can switch between ‘following’ and ‘independent’. The ‘following’ movement is modelled through a linear stochastic differential equation, while the ‘independent’ movement is modelled as Brownian motion. The movement of the leading point is modelled either as an Ornstein Uhlenbeck process or as Brownian motion, which makes the whole system a higher-dimensional Ornstein Uhlenbeck process, possibly an intrinsic non-stationary version. An inhomogeneous Kalman filter Markov chain Monte Carlo algorithm is developed to estimate the diffusion and switching parameters and the behaviour states of each individual at a given time point. The method successfully recovers the true behavioural states in simulated datasets, and is also applied to model a group of simultaneously tracked reindeer (*Rangifer tarandus*).

KEY WORDS: Animal Movement; Bayesian inference; Kalman filter; Multivariate Ornstein Uhlenbeck process; Stochastic differential equation; Switching diffusion.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Understanding the collective movement of animals is an important challenge in ecology. Individual animals in many of the world's taxa do not move independently of each other. For most, conspecific or heterospecific interactions influence decision making, behavioural choices and movement (Schlägel et al., 2019; Couzin et al., 2005; Merkle et al., 2016). Many species of birds, fish, insects and ungulates demonstrate highly cohesive and coordinated movements whose social interactions are being increasingly investigated (Herbert-Read, 2016; Westley et al., 2018; Buhl et al., 2006; Croft et al., 2015). In the past, simulation models have provided useful insights into the movement and decision making of animal groups (Aoki, 1982; Huth and Wissel, 1992). By assuming underlying laws of interaction these predictive models help us to understand certain ecological phenomena such as information sharing (Couzin et al., 2005), the effect of group size in obstacle avoidance (Croft et al., 2013) and how variation among individuals impacts the overall cohesion of the group (del Mar Delgado et al., 2018). Now with a wealth of tracking technologies it is possible to analyse real data of aggregations without such complete reliance on simulation models.

However, recent studies using real data typically employ a metric based approach to quantify aspects of collective movement such as dependency in acceleration or proximity rather than explicitly providing a model of movement. What's more, most studies are restricted to dyadic interactions (Polansky and Wittemyer, 2011; Joo et al., 2018; Long et al., 2014) even when providing a model for movement for example by correlating an individual's acceleration or turning angle with that of other members of its group (Polansky and Wittemyer, 2011). Some approaches do offer a stochastic model for multi-individual movement and may even account for behavioural heterogeneity (Calabrese et al., 2018) but still operate with metric based analysis whose behavioural transitions are dictated by the sampling scheme and whose inference can produce different results on different time-scales. Haydon et al. (2008) uses social structures of large groups to infer population dynamics, mortality rates and fecundity. This is demonstrated with a unique dataset of elk where all individuals are tracked. Whilst movement in this approach is modelled explicitly, each animal's movement is modelled individually using a correlated random walk and the group structure is quantified by spatial-temporal proximity from other individuals. Nonetheless, the "socially informed" model gives enlightening results about the population growth rate in relation to fission-fusion processes. They found that solitary individuals (those outside of the proximity threshold) have a higher risk of mortality than their grouped counterparts. This stresses the importance of forming coherent models which capture the sophistication of collective motion and social interactions of gregarious animals.

Existing realistic models of movement, which typically combine continuous locations in space with a discrete representation of behaviour, are generally limited to modelling single individuals. Langrock et al. (2014) do model the movement of a group of animals explicitly, allowing both dependent and independent behaviours, but their model and inference method are limited to discrete time, and their 'centroid' mechanism to represent attraction is explicitly tied to the time-scale of observations. Niu et al. (2016) give a continuous-time collective movement model (see §2.1) which assumes consistent group movement at all times, without any variation in behaviour. Here, we develop novel methodology which allows exact Bayesian statistical analysis for a class of group movement models with behavioural switching in continuous time, without any need for time-discretization error. We represent the group movement as a multivariate Ornstein Uhlenbeck process and allow the individuals to switch behaviour, either following the group or moving independently as Brownian motion. The times of changes in behaviour are represented as a thinned Poisson process, allowing exact simulation and Markov chain Monte Carlo inference. The methodology can be applied to data that are regular or irregular in time, with or without missing or incomplete observations. As well as much greater flexibility in modelling, our approach gives improved computational efficiency by integrating out part of the group movement process using a Kalman filter. In a set of simulation experiments, motivated in part by our analysis of data from simultaneously tracked reindeer (*Rangifer tarandus*), we show that our approach can reconstruct unobserved behaviours from location data in a range of scenarios.

The structure of the remainder of the article is as follows. We first review the existing models and extend the leader-follower framework to the non-stationary case and to include behavioural switching in Section 2. The simulation of the trajectories of the group using uniformization and the state space form of the model is explained in Section 3. The inhomogeneous Kalman filter Markov chain Monte Carlo algorithm is developed in Section 4 to estimate the behaviour states and diffusion parameters. The method is applied to model a simulation dataset in Section 5 and real data from reindeer movement in Section 6.

2. Models

Continuous-time models for movement are usually taken to be diffusion processes, the simplest case of which is Brownian motion. More general diffusion models for movement can be defined as solutions to stochastic differential equations (Brillinger et al., 2002). Here, we build on the approach of Niu et al. (2016) of modelling the group movement with a leader-follower framework (originating with Langrock et al., 2014) using a multivariate Ornstein Uhlenbeck process, and we start by summarising that model.

2.1 *Existing model*

Niu et al. (2016) represent the interaction between animals as a shared attraction to a leader L, which may be another animal, or simply an abstraction. If the leader is an animal, it may of course be observed or unobserved; in the former case the model still applies but much of the calculation is greatly simplified. For ease of exposition, we assume here that this is not the case, and the leader is either an unobserved animal or a mathematical abstraction. The observed individuals are conditionally independent, given full information about L. Thus animals do not interact directly, but only through their interactions with L. This formulation means that the model is robust to incomplete observation of a group of animals, and to variation over time of the number or identity of the observed individuals. The interpretation of the parameters of the model does not depend on the numbers of observed or unobserved animals. This approach is therefore suitable for cases where there may be many unobserved animals, e.g. large herds of herbivores. See Langrock et al. (2014) and Niu et al. (2016) for further discussion.

The movement of the unobserved leader L is modelled as a stationary Ornstein Uhlenbeck process. Let the random variable L_t^y represent the location of the y coordinate of the location of the leading point at time t. A stochastic process $\{L_t^y : t \geqslant 0\}$ in which L_t^y is attracted to θ^y is given by the stochastic differential equation (Schach, 1971)

$$\mathrm{d}L_t^y = -\beta(L_t^y - \theta^y)\mathrm{d}t + \rho \mathrm{d}V_t^y \tag{1}$$

where β is the attraction rate to θ^y; θ^y is a fixed location; ρ is the coefficient for the noise; V_t^y is standard Brownian motion. By applying the rotational symmetry which is natural in practice (Blackwell, 1997), the model is identical for x coordinate L_t^x, with parameters β and ρ in common, and independent Brownian motions $\{V_t^x\}$ and $\{V_t^y\}$ used for L_t^x and L_t^y.

A similar stochastic differential equation can model the movement of each of n followers attracted at any instant to the current location of the leader. Let the random variable $F_t^{y,k}$ represent the y coordinate of the kth follower's location at time t. $\{F_t^{y,k} : t \geqslant 0\}$ is defined by the following stochastic differential equation with parameters α, σ, L_t^y and Brownian motion $\{W_t^{y,k}\}$, where $F_t^{y,k}$ is attracted to L_t^y:

$$\mathrm{d}F_t^{y,k} = -\alpha \left(F_t^{y,k} - L_t^y \right) \mathrm{d}t + \sigma \mathrm{d}W_t^{y,k}$$

with α the attraction rate to L_t^y; σ the coefficient for the noise. By rotational symmetry as before, $F_t^{x,k}$ and $F_t^{y,k}$ satisfy identical equations.

For the present work, we express the idea of attraction to a leader by restricting α to be positive. Taking α to be negative would imply repulsion from the moving point at L_t^y, which is not a useful model of collective behaviour of the form that we are interested in, although a related model involving repulsion of a single animal from a fixed centre is explored by Blackwell (1997) and Harris and Blackwell (2013).

Combining the equations for the leading point and followers gives a stochastic differential equation for the y coordinates of both leader and followers which defines a particular multivariate Ornstein Uhlenbeck process:

$$\mathrm{d}\boldsymbol{Y_t} = A\boldsymbol{Y_t}\mathrm{d}t + \Sigma \mathrm{d}\boldsymbol{B^y_t} \tag{2}$$

where

$$\boldsymbol{Y_t} = \begin{pmatrix} \theta^y \\ L^y_t \\ F^{y,1}_t \\ \vdots \\ F^{y,n}_t \end{pmatrix}, A = \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ \beta & -\beta & \ddots & & \vdots \\ 0 & \alpha & -\alpha & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \alpha & 0 & \cdots & -\alpha \end{pmatrix}, \Sigma = \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 \\ 0 & \rho & \ddots & & \vdots \\ \vdots & \ddots & \sigma & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \sigma \end{pmatrix},$$

$$\boldsymbol{B}^{y^T_t} = \begin{pmatrix} 0 & V^y_t & W^{y,1}_t & \cdots & W^{y,n}_t \end{pmatrix}.$$

$\boldsymbol{Y_t}$ is a vector representing the y coordinates of the attractor, the leader and the followers. The attractor θ^y is a constant in Niu et al. (2016), but in general it could be modelled by another diffusion process; we include it in the state vector for convenience in describing the inference algorithm later. Note that each $F^{y,k}_t$ is *indirectly* attracted to θ^y. The matrix A is the attraction rate matrix. We take the stochastic parts (Brownian motion) for the leader and the followers to be uncorrelated, therefore Σ is diagonal; each diagonal element of the Σ, except the initial zero, represents the coefficient of the individual variance. The solution of this multivariate stochastic differential equation can be written as a multivariate normal distribution:

$$\boldsymbol{Y}_t|\boldsymbol{Y}_0 \sim \mathrm{MVN}(\boldsymbol{\mu}, \Xi) \tag{3}$$

where

$$\boldsymbol{\mu}^T = \begin{pmatrix} \theta^y & \mu_{\mathrm{L}}(L^y_0, t) & \mu_{\mathrm{F}}(L^y_0, F^{y,1}_0, t) & \cdots & \mu_{\mathrm{F}}(L^y_0, F^{y,n}_0, t) \end{pmatrix}$$

with

$$\mu_{\mathrm{L}}(L_0^y, t) = \left(L_0^y - \theta^y\right) e^{-\beta t} + \theta^y, \tag{4}$$

$$\mu_{\mathrm{F}}(L_0^y, F_0^{y,k}, t) = \left(L_0^y - \theta^y\right) \frac{\alpha}{\alpha - \beta} \left(e^{-\beta t} - e^{-\alpha t}\right) + \left(F_0^{y,k} - \theta^y\right) e^{-\alpha t} + \theta^y, \tag{5}$$

and

$$\Xi = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \xi_{\mathrm{L}} & \xi_{\mathrm{LF}} & \cdots & \cdots & \xi_{\mathrm{LF}} \\ \vdots & \xi_{\mathrm{LF}} & \xi_{\mathrm{F}} & \xi_{\mathrm{FF}} & \cdots & \xi_{\mathrm{FF}} \\ \vdots & \vdots & \xi_{\mathrm{FF}} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \xi_{\mathrm{FF}} \\ 0 & \xi_{\mathrm{LF}} & \xi_{\mathrm{FF}} & \cdots & \xi_{\mathrm{FF}} & \xi_{\mathrm{F}} \end{pmatrix}$$

with

$$\xi_{\mathrm{L}}(t) = \frac{\rho^2}{2\beta}\left(1 - e^{-2\beta t}\right), \tag{6}$$

$$\xi_{\mathrm{LF}}(t) = \frac{\rho^2 \alpha}{2\beta\left(\alpha + \beta\right)} - \frac{\rho^2 \alpha}{2\beta\left(\alpha - \beta\right)} e^{-2\beta t} + \frac{\rho^2 \alpha}{\alpha^2 - \beta^2} e^{-(\beta+\alpha)t}, \tag{7}$$

$$\begin{aligned} \xi_{\mathrm{F}}(t) = & \left\{\frac{\sigma^2}{2\alpha} + \frac{\rho^2 \alpha}{2\beta\left(\alpha + \beta\right)}\right\}\left(1 - e^{-2\alpha t}\right) - \frac{\rho^2 \alpha^2}{2\beta\left(\alpha - \beta\right)^2}\left(e^{-\beta t} - e^{-\alpha t}\right)^2 \\ & - \frac{\rho^2 \alpha^2}{\beta\left(\alpha^2 - \beta^2\right)}\left\{e^{-(\alpha+\beta)t} - e^{-2\alpha t}\right\}, \end{aligned} \tag{8}$$

$$\begin{aligned} \xi_{\mathrm{FF}}(t) = & \frac{\rho^2 \alpha}{2\beta\left(\alpha + \beta\right)}\left(1 - e^{-2\alpha t}\right) - \frac{\rho^2 \alpha^2}{2\beta\left(\alpha - \beta\right)^2}\left(e^{-\beta t} - e^{-\alpha t}\right)^2 \\ & - \frac{\rho^2 \alpha^2}{\beta\left(\alpha^2 - \beta^2\right)}\left\{e^{-(\alpha+\beta)t} - e^{-2\alpha t}\right\}. \end{aligned} \tag{9}$$

For details of the derivation see Niu et al. (2016). The parameter α controls the strength of the attraction of the followers to the leader. One consequence of this is that higher values of α will lead to the followers typically being closer to the leader, although of course their distribution around it depends on the diffusion parameters ρ and σ too.

2.2 *Behavioural Switching*

For realism in modelling animal group movement, the individual animals may not always follow the leader; they can move independently from time to time. The behaviour of the followers can switch between following the leader and independent Brownian motion. The Brownian motion type of movement can be modelled as $F_t^{y,k}|F_0^{y,k} \sim \mathrm{N}(F_0^{y,k}, t\sigma^2_{\mathrm{BM}})$ where σ^2_{BM} is the diffusion rate of the Brownian motion of the non-following animals.

We need to combine the group diffusion model and independent movement model through a framework of behavioural switching, whereby animals switch between behavioural states with different movement characteristics (Blackwell, 1997, 2003). In mathematical terms, we can represent this as a Markov process in continuous time with both a diffusion component, location, and a discrete one, behaviour, as in Berman (1994). The more complex case where behaviour itself depends on location is discussed below.

For modelling a single animal, the idea of a switching diffusion process driven by a continuous-time Markov chain was proposed in Blackwell (1997) and formalised in Blackwell (2003). In group movement modelling, a discrete-time version was described by Langrock et al. (2014); here we develop a multivariate Ornstein Uhlenbeck process for a group of animals, driven by a continuous-time Markov chain on a space representing their joint behaviour. We let J_t^k denote the kth animal's behavioural state at time t, taking values in $\{1, 2\}$, where 1 represents the state of following the leader and 2 represents the state of independent Brownian motion. Then we write $\boldsymbol{J}_t$ for the behavioural state for the whole group of animals at time t, taking values in $\{1, 2\}^n$. We take each J_t^k independently to be a continuous-time Markov chain on $\{1, 2\}$ having transition rates $\lambda_{1,2}$ and $\lambda_{2,1}$, where $\lambda_{1,2}$ is the switching rate of an individual from following the leader to Brownian motion and $\lambda_{2,1}$ is the switching rate of an individual from Brownian motion to following the leader. The transition rates for $\boldsymbol{J}_t$ are then implied by that structure, although it would be straightforward to allow for additional structure.

The kth animal starts in some state $J_0^k = i$ and location $F_0^{x,k} = x_0, F_0^{y,k} = y_0$, then follows the ith movement model; that is, $F_t^{x,k}|F_0^{x,k}$ and $F_t^{y,k}|F_0^{y,k}$ are realizations of the ith diffusion process. If $i = 2$ the diffusion process is Brownian motion; if $i = 1$ the animal is following the leader, so that its movement jointly with that of the leader is multivariate Ornstein Uhlenbeck. This continues until the time of the first switch in behaviour, at time T_1, when the animal is at $F_{T_1}^{x,k}, F_{T_1}^{y,k}$. If the behaviour switches to $J_t^k = j$, the next part of the location trajectory is a realisation of the jth diffusion process, starting at $F_{T_1}^{x,k}, F_{T_1}^{y,k}$, and so on.

The behavioural switching allows a much wider range of observed movement patterns. For example, if switching between behaviours is relatively slow, following individuals will tend to be found closer to the leader—and therefore closer together—the higher the value of α, while non-following animals will tend to drift away. However, faster switching between behaviours can complicate this picture, depending on the absolute and relative switching rates. For example, short periods of non-following behaviour will lead to the animals moving independently in the short term while generally remaining close together.

2.3 *Non-stationary case*

In the model described in §2.1, the leader and followers jointly define a multivariate Ornstein Uhlenbeck process and therefore have a stationary joint distribution. However, in practice the leader may not have a point of attraction, or at least not one that is relevant on the time scale of available data. The most tractable way to allow for this is to simply allow the leader to undergo Brownian motion instead of an Ornstein Uhlenbeck process, by setting $\beta = 0$ in equation 1. The individual stochastic differential equations are then

$$\mathrm{d}L_t^y = \rho \mathrm{d}V_t^y, \qquad \mathrm{d}F_t^{y,k} = -\alpha \left(F_t^{y,k} - L_t^y \right) \mathrm{d}t + \sigma \mathrm{d}W_t^{y,k}.$$

Similarly to §2.1, we can combine the equations for the leading point and the followers to give a stochastic differential equation for the y coordinates of both leader and followers. This defines an intrinsic Ornstein Uhlenbeck process, in which the behaviour of the leader is no longer stationary, but the configuration of the followers around the leader *is* stationary. The solution of this stochastic differential equation can again be written as a multivariate normal distribution:

$$\boldsymbol{Y}_t|\boldsymbol{Y}_0 \sim \mathrm{MVN}(\boldsymbol{\mu}^*, \Xi^*), \tag{10}$$

where

$$\boldsymbol{\mu}^{*^T} = \begin{pmatrix} \theta^y & L_0^y & \mu_{\mathrm{F}}^*(L_0^y, F_0^{y,1}, t) & \cdots & \mu_{\mathrm{F}}^*(L_0^y, F_0^{y,n}, t) \end{pmatrix}, \tag{11}$$

with $\mu_{\mathrm{F}}^*(L_0^y, F_0^{y,k}, t) = L_0^y(1 - e^{-\alpha t}) + F_0^{y,k} e^{-\alpha t}$. We retain θ^y in equation 11 in order to be consistent with the existing model in equation 2. θ^y is fixed to be zero and not used in the inference. The variance matrix can be written as

$$\Xi^* = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \xi_{\mathrm{L}}^* & \xi_{\mathrm{LF}}^* & \cdots & \cdots & \xi_{\mathrm{LF}}^* \\ 0 & \xi_{\mathrm{LF}}^* & \xi_{\mathrm{F}}^* & \xi_{\mathrm{FF}}^* & \cdots & \xi_{\mathrm{FF}}^* \\ 0 & \vdots & \xi_{\mathrm{FF}}^* & \ddots & \ddots & \vdots \\ 0 & \vdots & \vdots & \ddots & \ddots & \xi_{\mathrm{FF}}^* \\ 0 & \xi_{\mathrm{LF}}^* & \xi_{\mathrm{FF}}^* & \cdots & \xi_{\mathrm{FF}}^* & \xi_{\mathrm{F}}^* \end{pmatrix}. \tag{12}$$

Obtaining the conditional distributions in this case requires additional work, as the derivation of the previous result in equation 3 relies on the stationarity. In this case the solution is as follows:

$$\begin{aligned} \xi_{\mathrm{L}}^*(t) =& \rho^2 t, \quad \xi_{\mathrm{LF}}^*(t) = \rho^2 t - \frac{\rho^2}{\alpha}\left(1 - e^{-\alpha t}\right), \\ \xi_{\mathrm{F}}^*(t) =& \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}) + \frac{\rho^2}{2\alpha}(2\alpha t - 3) + \frac{2e^{-\alpha t}\rho^2}{\alpha} - \frac{e^{-2\alpha t}\rho^2}{2\alpha}, \\ \xi_{\mathrm{FF}}^*(t) =& \frac{\rho^2}{2\alpha}(2\alpha t - 3) + \frac{2e^{-\alpha t}\rho^2}{\alpha} - \frac{e^{-2\alpha t}\rho^2}{2\alpha}. \end{aligned}$$

The full derivations of these equation are given in Web Appendix A.

3. Inference

3.1 *Exact Simulation*

The key to Bayesian inference for the above model is the simulation of trajectories augmented by switching times/locations, appropriately conditioned on the observed data. We introduce the central idea by giving an algorithm for simulation of these models. We carry out the simulation exactly, rather than the typical approach of making a discrete-time approximation (Langrock et al., 2014) and assuming that a switch can only happen at a discrete observation time point. Here we want to avoid this unnatural assumption and the poorly understood discretization error involved.

We take a uniformization approach, introduced in the animal movement context by Blackwell et al. (2016), where the times of switches of behaviour form a Poisson process with rate κ, which has been 'thinned', that is each potential switching time point either retained or deleted probabilistically (Guttorp and Minin, 2018), in a way that depends on the movement process. This device is not strictly necessary when the transition rates are spatially homogeneous, as in our examples here, but it gives a useful framework which readily allows spatial heterogeneity. The rate κ needs to be an upper bound for all the actual transition rates. Since the transition rates are all of the form $\mathcal{N}\lambda_{1,2} + (n - \mathcal{N})\lambda_{2,1}$, where n is number of members in the group and $\mathcal{N}$ is the number in state 1, we take $\kappa = n\max(\lambda_{1,2}, \lambda_{2,1})$. The waiting time from any instant until the next switch in behaviour is then bounded below, in a probabilistic sense, by the time that would apply if the rate of switching was always κ. Starting at some known vector of states $\boldsymbol{J}_{T_0}$ for all members of the group, we can simulate the process forward as follows. Let $T \sim \text{Exponential}(\kappa)$ be the time of the first event of a process with constant rate κ. This is the first potential time at which a change in behaviour might occur. We can then determine whether the potential switch at T is an actual switch, an event which has probability $\lambda(T)/\kappa$, where

$$\lambda(T) = \mathcal{N}^{1,T_0}\lambda_{1,2} + \mathcal{N}^{2,T_0}\lambda_{2,1}$$

is the actual transition rate at time T, $\mathcal{N}^{1,T_0}$ is the number of animals following the leader at time T_0, and $\mathcal{N}^{2,T_0}$ is the number of animals moving as Brownian motion at time T_0. If it is an actual switch, we switch the state of the kth animal with probability $\lambda^{k,T_0}/\lambda(T)$, where λ^{k,T_0} is the switching rate of the kth animal at time T_0. λ^{k,T_0} is $\lambda_{1,2}$ if the kth animal's previous state $J^k_{T_0}$ is 1 or $\lambda_{2,1}$ if the previous state $J^k_{T_0}$ is 2. If it is not an actual switch, nothing need be changed. Only one animal can switch at each actual switching time.

Knowing $\boldsymbol{J}_T$, we can iterate this procedure forwards. This leads to a natural way of extending the simulation over as long interval as we desire. If we denote the events of the Poisson (κ) process by $T_1, T_2, \ldots$, then for each T_j in turn, we generate location $\boldsymbol{Y}_{T_j}$ by forward simulation.

3.2 *The state space form of the model*

Given the behavioural states, we can transform the group dynamic model with behaviour switching into a linear state space model, which can be expressed in the following form:

$$\boldsymbol{Y}_{t_{i+1}} = e^{A_i(t_{i+1}-t_i)}\boldsymbol{Y}_{t_i} + q_i, \quad q_i \sim \mathrm{MVN}(0, \Xi_i) \tag{13}$$

$$\boldsymbol{Z}_{t_i} = H_i \boldsymbol{Y}_{t_i} + \epsilon \tag{14}$$

where $q_i \sim \mathrm{MVN}(0, \Xi_i)$ is the process noise, and A_i and Ξ_i can take different forms based on the behavioural states. The measurement model is constructed by defining H_i through which the model is observed at the discrete time step t_i. $\boldsymbol{Z}_{t_i}$ is the observation of the followers' location and the leader location L_{t_i} is unobserved; in the case where the leader *is* observed, this is straightforward to accommodate by modifying H_i and hence $\boldsymbol{Z}_{t_i}$. We assume there is no observation error, and therefore we can set ϵ to zero. The state space form of the model is the discrete-time version of the continuous Ornstein Uhlenbeck and Brownian motion behavioural switching model. Here this discretisation is not an approximation, but is the so-called mild solution to the stochastic differential equation (Da Prato and Zabczyk, 2014).

Given the behavioural states $\boldsymbol{J}_{t_i}$ of the whole group at time t_i, the covariance matrix Ξ_i and coefficient matrix A_i need to be changed by setting the corresponding row and column to the Ornstein Uhlenbeck or Brownian motion version of the coefficient. For example, if we have one leader and three followers, and at time t_i, the second follower is moving as Brownian motion while the rest following the leader, then $J_{t_i}^2 = 2$, $\boldsymbol{J}_{t_i} = (1\ 2\ 1)$. The corresponding A_i and Ξ_i become

$$A_i = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ \beta & -\beta & 0 & 0 & 0 \\ 0 & \alpha & -\alpha & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 & -\alpha \end{pmatrix} \quad \Xi_i = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \Xi_{11} & \Xi_{12} & 0 & \Xi_{14} \\ 0 & \Xi_{21} & \Xi_{22} & 0 & \Xi_{24} \\ 0 & 0 & 0 & \sigma_{\text{BM}}^2 & 0 \\ 0 & \Xi_{41} & \Xi_{42} & 0 & \Xi_{44} \end{pmatrix}.$$

The second follower being in the Brownian motion behavioural state is reflected by setting the fourth row and column of Ξ_{t_i} and A_{t_i} to zero, except for the diagonal element in Ξ_{t_i} which is set to σ_{BM}^2, the diffusion parameter of the Brownian motion. This animal's attraction rate to the leader is 0 and its movement is independent of the rest of group. Here the first and second row and column of A_i and Ξ_i correspond to the attractor θ and the leader L_{t_i}. We keep the row for θ to be consistent with the setting in Niu et al. (2016), but since we concentrate here on the non-stationary case, θ is fixed to be 0 and is not used in the inference.

4. Markov chain Monte Carlo and the inhomogeneous Kalman filter

4.1 *Sampling the trajectory*

Based on the simulation ideas above and the state space form of the model, we can produce an algorithm for Bayesian inference for these models combining Markov chain Monte Carlo with the inhomogeneous Kalman filter. Given data $\boldsymbol{Z}_0, \ldots, \boldsymbol{Z}_t$ we want to sample from the posterior distributions for the parameters of the diffusion process and of the switching rates. Our approach involves augmenting the data with the times of all changes of behavioural state, and associated locations. We actually sample times, locations and states for all potential changes, that is at all times of a Poisson(κ) process. Since the true transition rates $\lambda_{1,2}, \lambda_{2,1}$ are unknown, we take their priors to be bounded above by κ_1, κ_2 respectively, and set $\kappa = \mathcal{N} \max\{\kappa_1, \kappa_2\}$.

Let $\mathcal{T}_{observe} = \{t_0, \ldots, t_N\}$ be the set of the observation times, $\mathcal{T}_{potential} = \{T_{i,j}, i = 0, \ldots, N -$

$1, j = 1, \ldots, M_i\}$ be the set of all potential switching time points, where M_i is the number of potential switches with $t_i < T_{i,j} < t_{i+1}$, and $\mathcal{T}_{actual}$ be the actual switching time, with $\mathcal{T}_{actual} \subset \mathcal{T}_{potential}$. We may have zero, one or multiple switches between two consecutive observation time points. The state of our chain is the collection of all times $\mathcal{T} = \mathcal{T}_{observe} \cup \mathcal{T}_{potential}$, plus associated locations $\boldsymbol{Y}_t$ for the whole group at $t \in \mathcal{T}$, initial state $\boldsymbol{J}_{t_0}$, the states $\boldsymbol{J}_t$ at potential switching time points, and implied states at the times of observations $\boldsymbol{J}_{t_1}, \ldots, \boldsymbol{J}_{t_N}$.

The key Markov chain Monte Carlo step is to sample the trajectory, that is potential switches, locations and states, over some time interval (t_a, t_b) such that $t_0 \leqslant t_a < t_b \leqslant t_N$, conditional on the trajectory outside that interval, on the states J_{t_a}, J_{t_b} and on the movement and switching parameters. We define $\mathcal{T}^{ab\prime}_{potential} = \{T'_{i,j}, i = a, \ldots, b-1, j = 1, \ldots, m_i\}$ with $t_i < T'_{i,j} < t_{i+1}$ and m_i the number of potential switches between t_i and t_{i+1}, as the set of all proposed potential switching times in the interval (t_a, t_b), a realisation of a Poisson (κ) process on (t_a, t_b). Once we propose all potential switches time points in the time interval (t_a, t_b), we can also propose the actual switches and the behavioural states.

Starting with $\boldsymbol{J}_{t_a}$, the behavioural states for the whole group at t_a, the next actual switching time $T'_{a,1}$ and corresponding behavioural states $\boldsymbol{J}_{T'_{a,1}}$ can be proposed as in §3.1, iterating to obtain $\boldsymbol{J}_{T'_{a,j+1}}$ for $j = 1, \ldots, m_a - 1$. This proposal process is repeated on each subinterval (t_i, t_{i+1}) for $i = a, \ldots, b-1$. We require for consistency that the final simulated behavioural states $\boldsymbol{J}_{T'_{b,m_b-1}}$ match the existing augmentation $\boldsymbol{J}_{t_b}$; if not, rejection is automatic. Conditioning on the proposed behavioural states, we can sample the trajectory by simulating the diffusion process forward. The simulation and inference of the continuous dynamic models (stochastic differential equations, see Øksendal, 2003), which are turned into the state space form in equation 13, can be done using the Kalman filter.

4.2 *Inhomogeneous Kalman filter*

The Kalman filter (see e.g. Särkkä, 2013) is an algorithm for solving the state estimation problem, which refers to the inverse problem of estimating the state trajectory of the stochastic process $\boldsymbol{Y}_t$ based on the noisy observations $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k$. The Kalman filter can be used to compute the exact Bayesian posterior distributions of the state in the state space form of the group movement model with behaviour switching. The transition of the state of the Kalman filter depends on the behaviour states $\boldsymbol{J}_t$, and so the system dynamics of the Kalman filter are inhomogeneous over time.

Unlike the inference algorithm in Niu et al. (2016) which requires imputing the unobserved leader's location to compute the marginal likelihood, the Kalman filter allows us to integrate out the leader's location exactly, massively reducing the dimension of the space of unknown quantities which the algorithm must explore. Here, a two-step scheme is presented, which first calculates the marginal distribution of the next step using the known system dynamics, given the behavioural states. In the prediction step, the mean and covariance matrix can be derived as:

$$\begin{aligned} m_{t_i|t_{i-1}} &= e^{A_i(t_i - t_{i-1})} m_{t_i|t_{i-1}} \\ P_{t_i|t_{i-1}} &= e^{A_i(t_i - t_{i-1})} P_{t_i|t_{i-1}} (e^{A_i(t_i - t_{i-1})})^{\mathrm{T}} + \Xi_i \end{aligned}$$

Here the subscript $t_i|t_{i-1}$ represents the prediction at step t_i conditional on the state at t_{i-1}. The recursive iteration is initialised by presenting the prior information in the form $\boldsymbol{Y}_0 \sim \mathrm{MVN}(m_0, P_0)$. In the stationary case, m_0 and P_0 follow form the stationary distribution, but in the non-stationary case, some care is needed in the initialisation; details are given in Web Appendix B. The algorithm then uses each observation to update the distribution to match the new information obtained by the measurement in step t_i. This is the updating step.

$$\begin{aligned} k_{t_i} &= P_{t_i|t_{i-1}} H_i^{\mathrm{T}} (H_i P_{t_i|t_{i-1}} H_i^{\mathrm{T}})^{-1} \\ m_{t_i|t_i} &= m_{t_i|t_{i-1}} + k_{t_i}(\boldsymbol{Z}_{t_i} - H_i m_{t_i|t_{i-1}}) \\ P_{t_i|t_i} &= P_{t_i|t_{i-1}} - k_{t_i} H_i P_{t_i|t_{i-1}} H_i^{\mathrm{T}} k_{t_i}^{\mathrm{T}} \end{aligned}$$

where $(.)^{-1}$ denotes the matrix inverse and $(.)^{\mathrm{T}}$ the matrix transpose. As a result, the filtered forward-time posterior process in step t_i is given by $\boldsymbol{Y}_{t_i} \sim \mathrm{MVN}(m_{t_i|t_i} P_{t_i|t_i})$. In this iterative computation, A_i and Ξ_i will change according to the behavioural states $\boldsymbol{J}_{t_i}$. H_i will also change according to the availability of the observations at time step t_i. The updating step is only run when $t_i \in \mathcal{T}_{observe}$, whereas we need to run the prediction step at every potential switching time and observation time. Given the behavioural states $\boldsymbol{J}_{T'_{a,1}}, \ldots, \boldsymbol{J}_{T'_{b,m_b-1}}$ in the interval (t_a, t_b), the log likelihood of trajectories in the interval (t_a, t_b) is

$$-\sum_{i=a+1}^{b} \left\{\frac{1}{2} n \log 2\pi + \frac{1}{2} \log |H_i P_{t_i|t_{i-1}} H_i^{\mathrm{T}}| + \frac{1}{2}(Z_{t_i} - H_i m_{t_i|t_{i-1}})^{\mathrm{T}} (H_i P_{t_i|t_{i-1}} H_i^{\mathrm{T}})^{-1}\right\}. \qquad (15)$$

4.3 *Parameter inference*

The behavioural states, switching rates and the diffusion parameters can be estimated using Markov chain Monte Carlo with a standard Metropolis Hastings algorithm. We propose new switching rates $\boldsymbol{\lambda}'$ using the symmetric Gaussian proposal distribution centered on the previous values $\boldsymbol{\lambda}$. The acceptance probability for $\boldsymbol{\lambda}'$ depends only on $\boldsymbol{J}$, since movement is independent of the rates given the states, and since $\mathcal{T}$ depends only on κ. The new switching rates are accepted with probability $\min\{\mathrm{HR}, 1\}$ where HR is the Hastings ratio

$$\frac{p\left(\boldsymbol{\lambda}'|\boldsymbol{J}, \mathcal{T}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Theta}\right) q(\boldsymbol{\lambda}|\boldsymbol{\lambda}')}{p\left(\boldsymbol{\lambda}|\boldsymbol{J}, \mathcal{T}, \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{\Theta}\right) q(\boldsymbol{\lambda}'|\boldsymbol{\lambda})} = \frac{p(\boldsymbol{\lambda}')p(\boldsymbol{J}|\boldsymbol{\lambda}')}{p(\boldsymbol{\lambda})p(\boldsymbol{J}|\boldsymbol{\lambda})}$$

by conditional independence and symmetry.

Given the trajectory and states, we know exactly what type of the movement processes the group of animals were following, so the inference about the movement parameters is straightforward. From the Markov property, the trajectory log-likelihood is calculated by summing over terms of the form given in equation 15 for the whole time interval. All followers are considered jointly. We use uniform priors on $[0, +\infty)$ and standard random-walk Metropolis-Hastings updates for diffusion parameters. The only non-standard aspect is the calculation of the likelihood, and so other details are omitted. Similar, lower-dimensional updates for a model of a single animal are described in detail by Blackwell (2003).

5. Implementation with simulated data

We carried out several simulation experiments to demonstrate the model's ability to pick up on a wide range of behavioural patterns. The first uses parameter values similar to those obtained from the analysis of real data from reindeer tracking in §6; results are described in §5.1. In the other simulations, the diffusion parameters are modified from these values in contrasting ways. In §5.2, the data are simulated with relatively high values for both α and σ_{BM}. Here we can imagine the animals are tightly grouped when in the OU state and widely separated when in the BM state, perhaps representing individual exploring behaviour. In contrast, the data in §5.3 are simulated with a much smaller σ_{BM} This leads to movement behaviour where, when the animals are not grouped together, they forage locally, leading to rather stationary behaviour.

In each case, we simulated the location of five followers and one leading point in both x and y directions from the non-stationary intrinsic Ornstein Uhlenbeck model, for 50 steps forward by using equation 10 iteratively and taking each generated location as the origin for the next. We then applied the Markov chain Monte Carlo algorithm described above to reconstruct the trajectories and the parameters of the model.

5.1 *Reindeer-based simulation*

For simulated data based on the reindeer analysis, we ran the Markov chain Monte Carlo algorithm for 50,000 iterations after burn-in. The posterior mean and standard deviation of model parameters are shown in Table 1, along with the true values used in the simulation.

[Table 1 about here.]

Posterior density plots of the model parameters are given in the Supporting Information Web Appendix Figure C.1. All posterior distributions are consistent with the true values. The posterior means of the behavioural states (black crosses) for each follower at every time point are plotted against the true behavioural states (red circles) in Web Appendix Figure C.2. It is clear from Figure C.2 that most of the true states are captured by our estimates. However, some are more difficult to estimate, like the true states of animals 1 and 2 at around time 39. In these two cases, the animals only stay in the Brownian motion behavioural state for very short times, relative to the observation time intervals, and then switch back to the Ornstein Uhlenbeck state following the leader. Inevitably, this makes it harder for the inference algorithm to capture the switching. On the other hand, if the animals move in certain behavioural states for somewhat longer time periods like animals 2 and 3 in the time interval 1 to 10, the estimated behaviour states match the truth very well. All these results show that our fully Bayesian approach can reconstruct the states of followers and their diffusion trajectories.

5.2 *Simulation with high attraction and diffusion*

For simulated data with a high attraction parameter and diffusion coefficients, we ran the Markov chain Monte Carlo algorithm for 100,000 iterations after burn-in. The true values of the parameters used are given in Supporting Information Web Appendix C in Table C.1 along with point estimates and standard deviations of the posterior distributions for each parameter. The posterior densities are given in Figure C.4. Posterior means for the states of each animal are given in Figure C.5. The model performed well at retrieving the true values, even with widely dispersed initial values.

Visualisations of the movement trajectories are also given in in Web Appendix C. In Figure C.6, each animal's path is plotted in one dimension against time whilst simultaneously indicating the posterior state estimation at each time step. For completeness the trajectories in two dimensions are presented in Figure C.7. In this simulation study, we set the true value of $\sigma_{BM} = 5$, $\lambda_{1,2} = 0.1$ and $\lambda_{2,1} = 0.4$. Since $\lambda_{2,1} > \lambda_{1,2}$, each animal has a higher probability of being in the BM state than in the OU state. The state estimation also confirms that animals spent most of the time in OU states. The high value of σ_{BM} leads to large movement steps when animal is in BM states, as is clear from the trajectories plotted in Figures C.6 and C.7. Considering the individual trajectories, state estimation is difficult, compared with the results in §5.3, but carrying out the estimation jointly gives good results.

5.3 *Simulation with low diffusion*

For simulated data with a low diffusion coefficient in the non-following state, we ran the Markov chain Monte Carlo algorithm for 100,000 iterations after burn-in. The true values of the parameters used are given in Table C.2 along with the point estimates and standard deviations of the posterior distributions for each parameter. The posterior densities are given in Figure C.8, and posterior means for the states of each animal are given in Figure C.9. For comparison with the previous example, the movement trajectories in one and two dimensions are shown in Figures C.10 and C.11 respectively. In this simulation study, the true value of the BM diffusion parameter $\sigma_{BM} = 0.1$ is much smaller than in §5.2. The effect of this small BM diffusion parameter is clearly demonstrated in Figures C.10 and C.11, with movement in the BM state being much more localised than before. As expected, state estimation is generally good in this case; the parameter estimation also reflects the true values, and correctly captures the qualitative difference from the previous case.

6. Implementation with real data

We also illustrate this approach using the real movement data from 5 reindeer *(Rangifer tarandus)* from a study site in Njaarke reindeer herding community, Sweden. The data used are a subset of observations from 79 individual reindeer equipped with GPS collars, collected in 2009-2011 (Rivrud et al., 2018). In an effort to test the model's ability to capture behavioural heterogeneity, the specific subset was chosen through exploratory data analysis from which it appeared that at some times individuals switched from following the group to a Brownian motion behaviour. The data consist of up to 50 observations from each individual taken every two hours from 01/12/2009 until 5/12/2009. Whilst they are subject to some of the usual irregularities when dealing with real data, i.e. missing values and observation spacing inconsistencies, the observations are almost regular insofar as they occur up to only 2 minutes before/after the intended timing. Thus, for the simplicity of implementation, the time steps of the data were rounded to the nearest hour.

However, in principle the methodology accounts for irregular times between observations. The original GPS data in the geographical projection WGS84 or ESPG:4326 were transformed to the Swedish coordinate system SWEREF99 or ESPG:3006, using the `spTransform` function within the `sp` package in R (Pebesma and Bivand, 2005), and then further scaled down by a factor of 100 for numerical convenience, before analysis.

Table 2 shows posterior means and standard deviations for the parameters of the model, and density plots of the posterior distribution of the model parameters are shown in Web Appendix Figure C.3. The results here are based on 100,000 iterations of Markov chain Monte Carlo runs fitting the switching non-stationary model, with over-dispersed initial values, every second iteration being recorded after 10,000 iterations of burn-in. The corresponding posterior mean states are shown in Figure 1.

[Table 2 about here.]

[Figure 1 about here.]

In order to compare with the method in Niu et al. (2016), we also fitted the real data with the non-switching model as in equation 2. The posterior mean of the non-switching variance coefficients of the leader ρ is 11.3, compared with the much smaller parameter 4.58 in the switching case. Similarly, the attraction rate parameter α is 0.32 for the non-switching case and 1.33 for the switching case. The non-switching model treats the independent movement of followers as the part of the group movement. This leads to the larger estimated variance of the leader's location and smaller estimated attraction rate, while in the switching group movement model, we successfully distinguished the group movement and independent movement of the followers using the behavioural states.

A partial visualization of the data as a trajectory over time in one dimension is given in Figure 2. The points are classified as OU or BM if their point estimates are $\leqslant 1.5$ or > 1.5 respectively where 1 denotes the OU state and 2 denotes the BM state. Some experimentation was done with a less strict threshold to account for an uncertain category (say, between 1.4 and 1.6) but this had only a few points as most estimates are confidently assigned to one behaviour or the other.

[Figure 2 about here.]

Since the estimated switching rate $\lambda_{21} = 0.63$ is substantially higher than the reverse rate $\lambda_{12} = 0.16$, the proportion of time that the reindeer spend in the BM state is quite low, as can be seen in Figure 2. The BM diffusion parameter $\sigma_{BM} = 2.49$ is small compared with the actual movement in the OU state, due to attraction, even though the independent component of OU movement is even smaller ($\sigma_{OU} = 0.64$), and so the movement in the BM state is very localised, as is again clear from Figures 2. In the OU state, movement is generally faster, which is largely driven by the dependent component based on α. Furthermore, we can link the locations in two dimensions, categorised by estimated behaviour, to the actual terrain on the ground at that location; see Figure 3.

[Figure 3 about here.]

Comparing these plots with satellite data confirms that the forest areas being used by the reindeer contain lichen, on which the reindeer typically feed; so in this particular case, the grouping dynamics are likely to be driven by the costs and benefits of collective foraging for lichen in winter, as discussed in §7.

The true diffusion parameters and the behaviour states are unknown. However, we have generated the simulation data and behaviour states in §5.1 based on parameters similar to those estimated from the real data in this section. The results in §5 and §6 give us confidence about our approach and interpretation.

7. Discussion

We have described the formulation of a group movement model with behaviour switching in continuous time, building on some of the strengths of previous approaches, and an algorithm for fully Bayesian inference. We have shown that we can successfully estimate the behavioural states and diffusion parameters. Compared to Niu et al. (2016), we have introduced behaviour switching in the continuous movement model and also extended the model to the non-stationary case by defining the leader's movement process as Brownian motion.

Behaviour switching is important in real applications in realistic representation of movement c.f. Blackwell (1997, 2003), Gurarie et al. (2010), Haydon et al. (2008), Morales and Ellner (2002), Langrock et al. (2014). Simpler single-behaviour models fail to capture the heterogeneity of movement exhibited by animals as they respond to their environment. When considering multiple animals, these behaviours represent complex trade-offs between environmental and social factors. For example, although an individual reindeer may reduce its grazing competition by moving away from the herd, it then also stands a greater chance of being killed by predators or, in summer, being harassed by insects, and therefore the choice an individual reindeer makes about how and where to move is balanced between finding enough food for itself but also staying within the safety of the group (Mooring and Hart, 1992). In winter reindeer usually graze in groups digging for lichens underneath the snow. Staying with a group where several animals are digging could be beneficial for the individual reindeer as this saves time and energy from digging. However, this also means competition among the animals for the best lichens forage and individuals may be pushed away and thus need to search for new places to dig (Kojola, 1989). Our approach is unique in allowing this behavioural complexity for group movement while retaining the theoretical and practical benefits of formulation in continuous time.

Of course, if changes in behaviour are rapid compared to the time scale of the information from observations, for example if there are frequently multiple switches between observations, then it becomes impossible to reconstruct the sequence of behaviours, much less their precise timing, with any certainty. That is inevitable in any model of this kind; our approach does at least allow properly for the different underlying possibilities, and the associated uncertainty, rather than ignoring them as would be necessary in a discrete-time model. Our introduction of the Kalman filter also saves us from the need to impute the location of the leading point as in Niu et al. (2016). By massively reducing the dimension of the space to be explored by the Markov chain Monte Carlo algorithm, this makes the computation feasible even in this more complex model.

Even with the gains from the use of the Kalman filter, our exact approach to reconstruction of the animals' behaviour means that computational costs will limit the size of the dataset that can be analyzed in this way. For large datasets, it would be possible to carry out an approximate analysis, using time discretization. In such cases, we believe that it is preferable to formulate the model as we have done in continuous time, and then approximate, rather than attempting to formulate a discrete-time model that is unable to accommodate irregular data, missing values etc.

We have neglected observation error, as is common in movement modelling. However, the use of the Kalman filter means that it would be straightforward to allow for observation error, taking ϵ to be non-zero in equation 14. Similarly, the specific models discussed in detail and applied here have switching rates for each individual which are spatially and temporally homogeneous. However, the method is formulated and implemented within a uniformization approach which makes it possible to incorporate heterogeneity in switching rates, following Blackwell et al. (2016).

Our approach considers a herd represented by a single 'leader' and animals who follow the leader for part of the time. A model which allows switching between multiple separate leaders, suitable for species with more complex social structures, but which relies on a more complete tracking of individuals, is explored by Milner et al. (In press).

ACKNOWLEDGEMENTS

The authors are grateful to the Njaarke reindeer herding community for their support in the collection of the data.

FF is supported by the Centre for Advanced Biological Modelling at the University of Sheffield, funded by the Leverhulme Trust, award number DS-2014-081. JEM is supported by a studentship from the Engineering and Physical Sciences Research Council.

ORCID

Mu Niu https://orcid.org/0000-0002-3068-5501

Fay Frost https://orcid.org/0000-0003-2382-2990

Jordan E. Milner https://orcid.org/0000-0002-0863-3158

Anna Skarin https://orcid.org/0000-0003-3221-1024

Paul G. Blackwell https://orcid.org/0000-0002-3141-4914

DATA AVAILABILITY STATEMENT

The data analyzed in this paper are available on request from the authors.

REFERENCES

Aoki, I. (1982). A Simulation Study on the Schooling Mechanism in Fish. *Nippon Suisan Gakkaishi* **48,** 1081–1088.

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

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,** 613–627.

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

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 of the Brazilian Mathematical Society* **33,** 385–408.

Buhl, J., Sumpter, D. J., Couzin, I. D., Hale, J. J., Despland, E., Miller, E. R., and Simpson, S. J. (2006). From Disorder to Order in Marching Locusts. *Science* **312,** 1402–1406.

Calabrese, J. M., Fleming, C. H., Fagan, W. F., Rimmler, M., Kaczensky, P., Bewick, S., Leimgruber, P., and Mueller, T. (2018). Disentangling social interactions and environmental drivers in multi-individual wildlife tracking data. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373,**.

Couzin, I. D., Krause, J., Franks, N. R., and Levin, S. A. (2005). Effective leadership and decision-making in animal groups on the move. *Nature* **433,** 513–516.

Croft, S., Budgey, R., Pitchford, J. W., and Wood, A. J. (2013). The influence of group size and social interactions on collision risk with obstacles. *Ecological Complexity* **16,** 77–82.

Croft, S., Budgey, R., Pitchford, J. W., and Wood, A. J. (2015). Obstacle avoidance in social groups: New insights from asynchronous models. *Journal of the Royal Society Interface* **12,**.

Da Prato, G. and Zabczyk, J. (2014). *Stochastic equations in infinite dimensions*. Cambridge University Press, Cambridge, UK.

del Mar Delgado, M., Miranda, M., Alvarez, S. J., Gurarie, E., Fagan, W. F., Penteriani, V., di Virgilio, A., and Morales, J. M. (2018). The importance of individual variation in the dynamics of animal collective movements. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373,** 20170008.

Gurarie, E., Andrews, R. D., and Laidre, K. L. (2010). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters* **12,** 395–408.

Guttorp, P. and Minin, V. N. (2018). *Stochastic modeling of scientific data*. CRC Press.

Harris, K. J. and Blackwell, P. G. (2013). Flexible continuous-time modelling for heterogeneous animal movement. *Ecological Modelling* **255,** 29–37.

Haydon, D. T., Morales, J. M., Yott, A., Jenkins, D. A., Rosatte, R., and Fryxell, J. M. (2008). Socially informed random walks: incorporating group dynamics into models of population spread and growth. *Proceedings of the Royal Society B Biological Sciences* **275,** 1101–1109.

Herbert-Read, J. E. (2016). Understanding how animal groups achieve coordinated movement. *Journal of Experimental Biology* **219,** 2971–2983.

Huth, A. and Wissel, C. (1992). The simulation of the movement of fish schools. *Journal of Theoretical Biology* **156,** 365–385.

Joo, R., Etienne, M.-P., Bez, N., and Mahévas, S. (2018). Metrics for describing dyadic movement: a review.

Kojola, I. (1989). Mother's dominance status and differential investment in reindeer calves. *Animal Behaviour* **38,** 177–185.

Langrock, R., Hopcraft, J. G. C., Blackwell, P. G., Goodall, V., King, R., Niu, M., Patterson, T. A., Pedersen, M. W., Skarin, A., and Schick, R. S. (2014). Modelling group dynamic animal movement. *Methods in Ecology and Evolution* **5,** 190–199.

Long, J. A., Trisalyn, N. A., Webb, S. L., and Gee, K. L. (2014). A critical examination of indices of dynamic interaction for wildlife telemetry studies. *Journal of Animal Ecology.* **83,** 1216–1233.

Merkle, J. A., Monteith, K. L., Aikens, E. O., Hayes, M. M., Hersey, K. R., Middleton, A. D., Oates, B. A., Sawyer, H., Scurlock, B. M., and Kauffman, M. J. (2016). Large herbivores surf waves of green-up during spring. *Proceedings of the Royal Society B: Biological Sciences* **283,**.

Milner, J. E., Blackwell, P. G., and Niu, M. (In press.). Modelling and inference for the movement of interacting animals. In press, *Methods in Ecology and Evolution*.

Mooring, M. S. and Hart, B. L. (1992). Animal grouping for protection from parasites: Selfish herd and encounter-dilution effects. *Behaviour* **123,** 173–193.

Morales, J. M. and Ellner, S. P. (2002). Scaling up animal movements in heterogeneous landscapes: The importance of behavior. *Ecology* **83,** 2240–2247.

Niu, M., Blackwell, P. G., and Skarin, A. (2016). Modeling interdependent animal movement in continuous time. *Biometrics* **72,** 315–324.

Øksendal, B. (2003). *Stochastic differential equations: An introduction with applications*. Springer, Berlin.

Pebesma, E. J. and Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News* **5,** 9–13. https://CRAN.R-project.org/doc/Rnews/.

Polansky, L. and Wittemyer, G. (2011). A framework for understanding the architecture of collective movements using pairwise analyses of animal movement data. *Journal of the Royal Society Interface* **8,** 322–333.

Rivrud, I. M., Sivertsen, T. R., Mysterud, A., Åhman, B., Støen, O. G., and Skarin, A. (2018). Reindeer green-wave surfing constrained by predators. *Ecosphere* **9,**.

Särkkä, S. (2013). *Bayesian Filtering and Smoothing (Institute of Mathematical Statistics Textbooks)*. Cambridge University Press, Cambridge, UK.

Schach, S. (1971). Weak convergence results for a class of multivariate Markov processes. *The Annals of Mathematical Statistics* **42,** 451–465.

Schlägel, U. E., Signer, J., Herde, A., Eden, S., Jeltsch, F., Eccard, J. A., and Dammhahn, M. (2019). Estimating interactions between individuals from concurrent animal movements. *Methods in Ecology and Evolution* **10,** 1234–1245.

Westley, P. A. H., Berdahl, A. M., Torney, C. J., Biro, D., Westley, P. A. H., Torney, C. J., and Biro, D. (2018). Collective movement in ecology: from emerging technologies to conservation and management. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373,** 20170004.

Supporting information

Web Appendices A and B, containing derivations for the non-stationary case, referenced in Section 2.3 and 4.2, and Web Appendix C, containing additional tables and figures for simulated analyses, referenced in Section 5, are available with this paper at the Biometrics website on Wiley Online Library. Code for the simulations and analysis carried out in the paper are available online at the same location, and also at `https://github.com/mu2013/Group-Movement-Switching`.

Received November 2019. *Revised June* 2020. *Accepted ?* 2020.

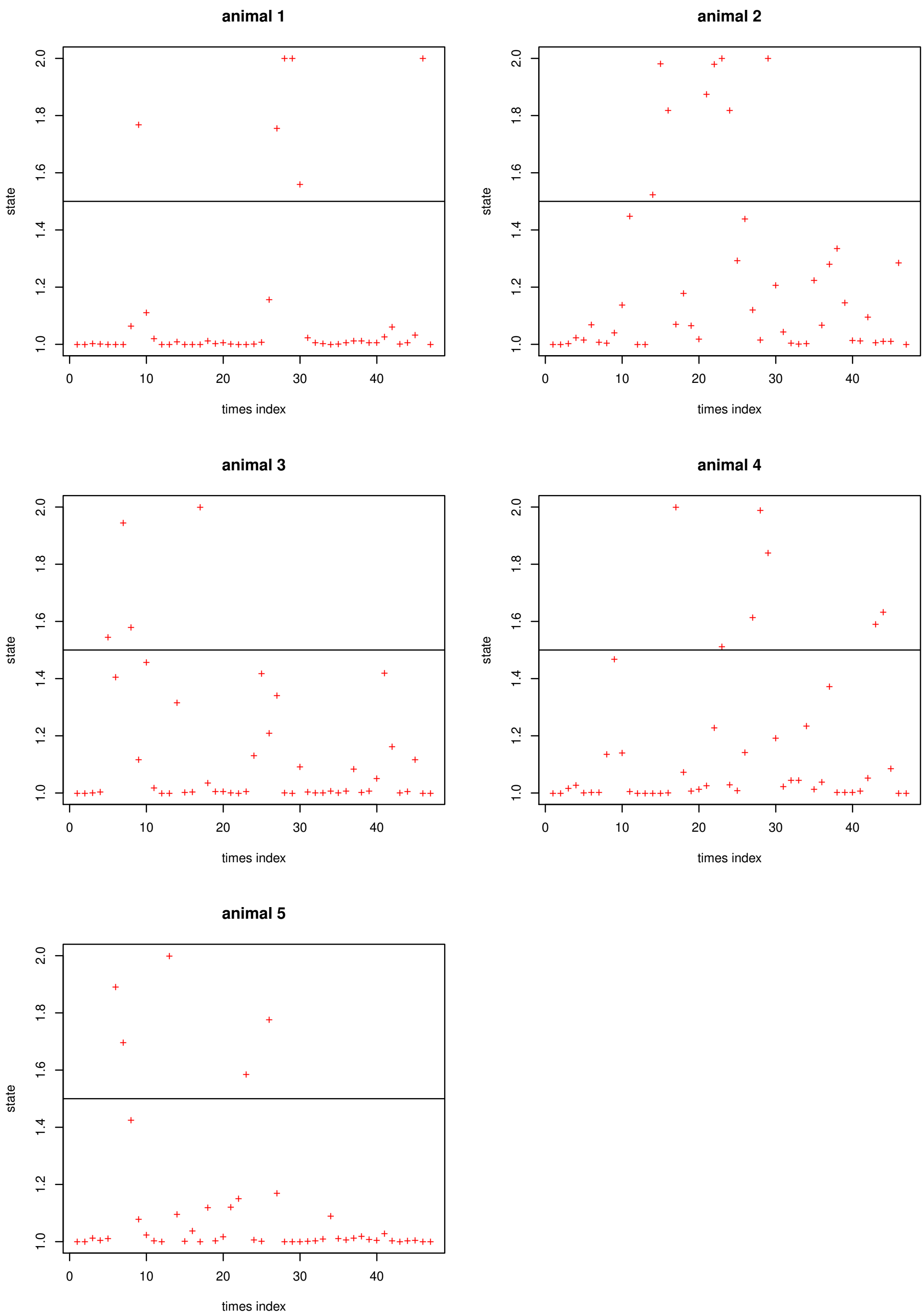

Figure 1: Posterior mean states of all followers for the real dataset . The circles (red) represent the true states of the follower. The vertical axis represents the states, 1 for Ornstein Uhlenbeck and 2 for Brownian motion. The crosses (black) represent the mean posterior of the estimated behaviour states. This figure appears in colour in the electronic version of this article, and the colours refer to that version.

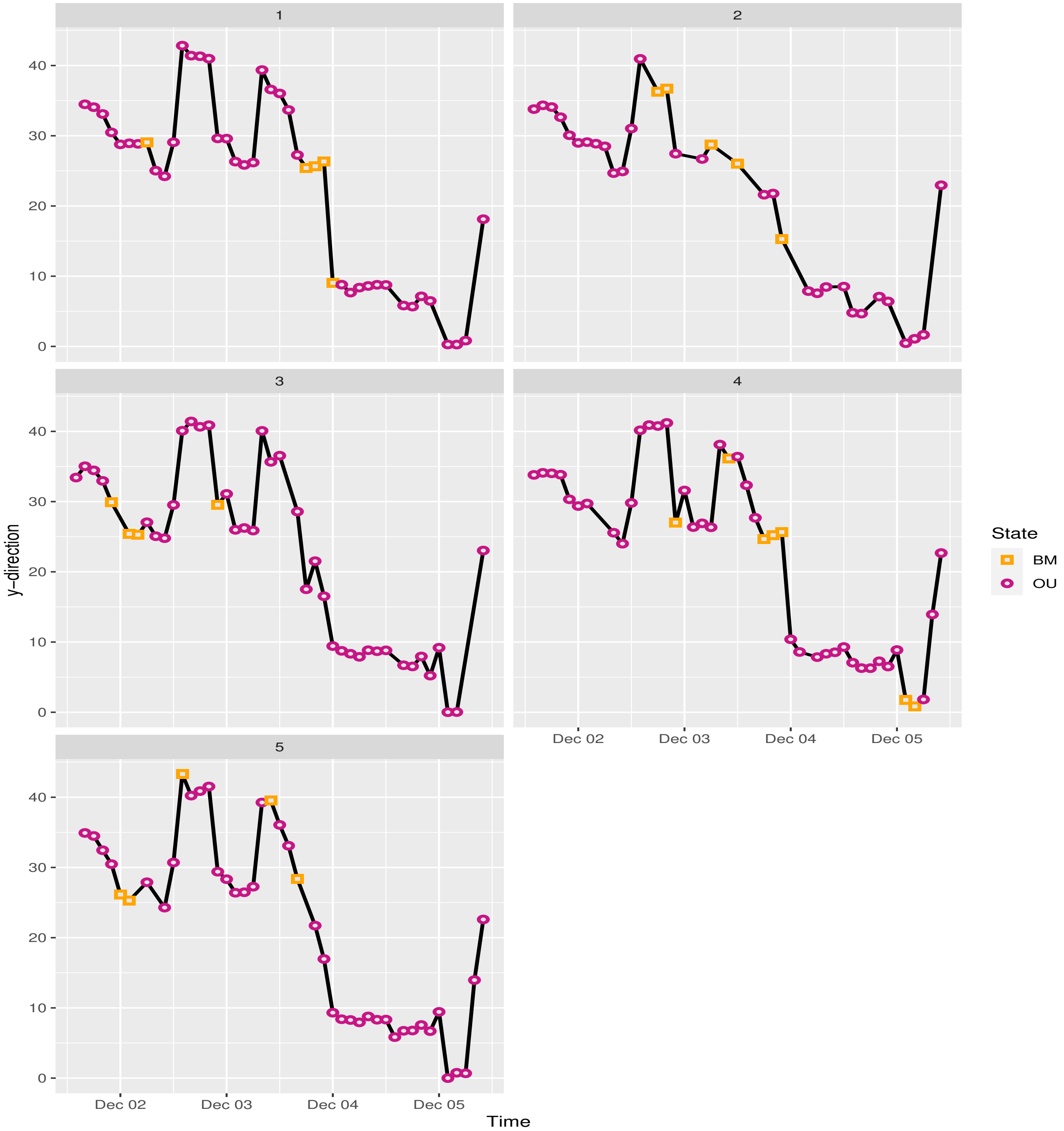

Figure 2: Plot of locations in the y-direction (i.e. scaled northings in SWEREF99; see main text) against time, for each animal in the real dataset. At each time step the points indicate whether the individual's posterior state is OU or BM. The orange square points indicate an BM state whilst the purple circular points indicate OU states. This figure appears in colour in the electronic version of this article, and the colours refer to that version.

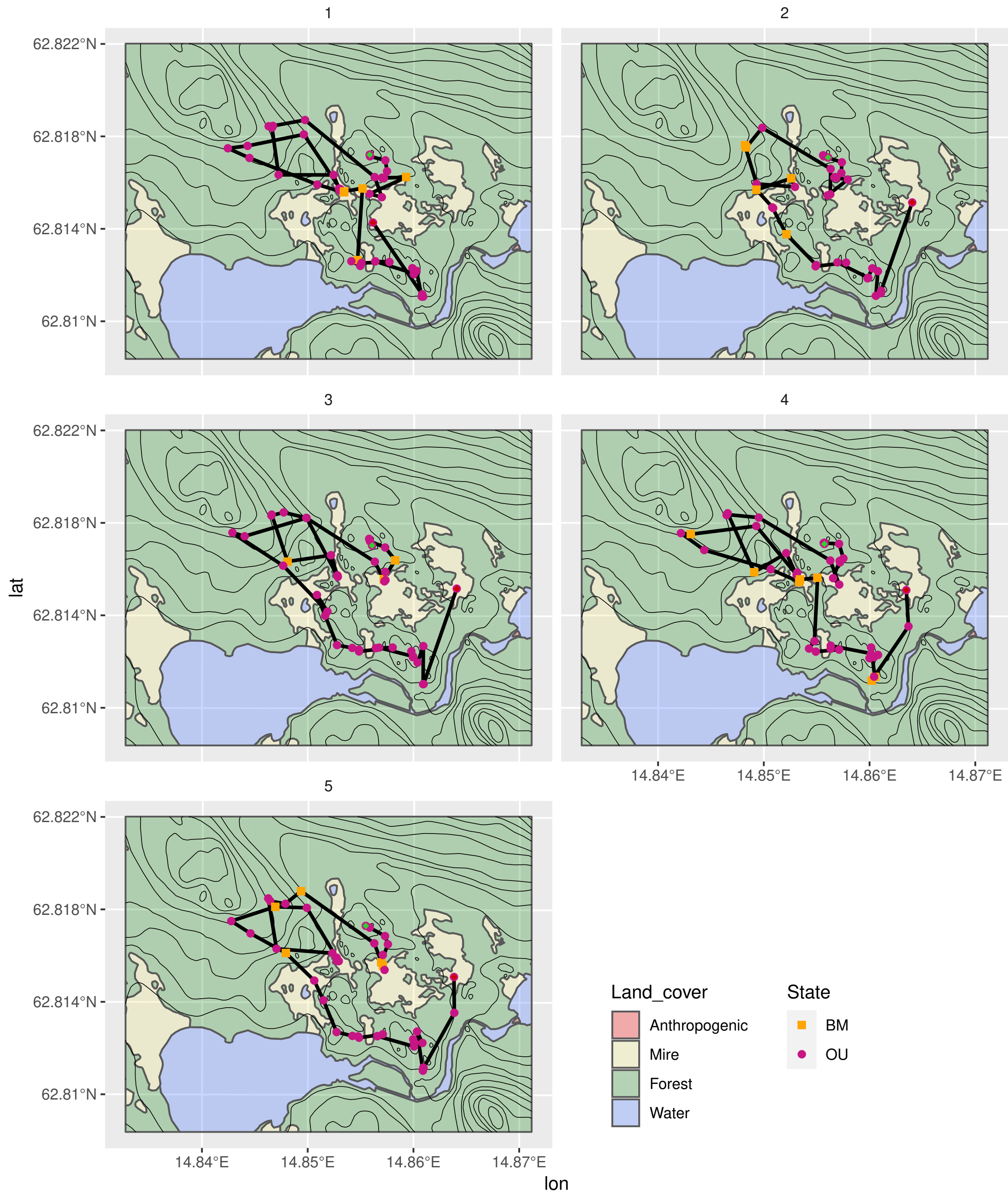

Figure 3: Plot of trajectories projected on to a terrain map for each animal in the real dataset. At each time step the points indicate whether the individual's posterior state is OU or BM. The orange square points indicate an BM state whilst the purple circular points indicate OU states. Start and end points are indicated by green and red diamonds respectively. The terrain is split into four categories: anthropogenic, water body, mire and forest given in red, blue, tan and green respectively. The latitude/longitude coordinates were used for visualization; see main text for relationship with data as analysed.

Table 1: Parameter estimates for the movement and switching model with reindeer-based simulation data

Parameter	Point estimate	Standard deviation	True value
α	1.23	0.06	1.2
ρ	5.02	0.34	5.0
σ	0.69	0.04	0.7
σ_{BM}	1.59	0.15	2.0
$\lambda_{1,2}$	0.14	0.02	0.1
$\lambda_{2,1}$	0.51	0.08	0.4

Table 2: Parameter estimates for reindeer movement and switching model with real dataset

Parameter	Point estimate	Standard deviation
α	1.33	0.24
ρ	4.58	0.41
σ	0.64	0.07
σ_{BM}	2.49	0.39
$\lambda_{1,2}$	0.16	0.03
$\lambda_{2,1}$	0.63	0.05