# Modeling Interdependent Animal Movement in Continuous Time

**Mu Niu,**[1,*,†] **Paul G. Blackwell,**[1,**] **and Anna Skarin**[2,***]

[1]School of Mathematics & Statistics, University of Sheffield, Sheffield S3 7RH, U.K.
[2]Department of Animal Nutrition and Management, Swedish University of Agricultural Sciences,
Uppsala, Sweden
[†]Current address: School of Mathematics & Statistics, University of Glasgow
*email: mu.niu@glasgow.ac.uk
**email: p.blackwell@sheffield.ac.uk
***email: anna.skarin@slu.se

Summary. This article presents a new approach to modeling group animal movement in continuous time. The movement of a group of animals is modeled as a multivariate Ornstein Uhlenbeck diffusion process in a high-dimensional space. Each individual of the group is attracted to a leading point which is generally unobserved, and the movement of the leading point is also an Ornstein Uhlenbeck process attracted to an unknown attractor. The Ornstein Uhlenbeck bridge is applied to reconstruct the location of the leading point. All movement parameters are estimated using Markov chain Monte Carlo sampling, specifically a Metropolis Hastings algorithm. We apply the method to a small group of simultaneously tracked reindeer, *Rangifer tarandus tarandus*, showing that the method detects dependency in movement between individuals.

Key words: Animal movement; Bayesian inference; Multivariate Ornstein Uhlenbeck process; Ornstein Uhlenbeck bridge; Stochastic differential equation.

## 1. Introduction

In recent years, an important aspect of statistical ecology has been the improvement in the understanding of wildlife movement, making use of increasing amounts of data from radiotracking and the global positioning system (GPS) (see, e.g., Nathan et al., 2008). The understanding of movement patterns is in turn important in evaluating the environmental effects of land use, climate change, or anthropogenic activities (Bowler and Benton, 2005).

Many statistical approaches have been extensively discussed, including state–space models (e.g., Johnson et al., 2008; Patterson et al., 2008), stochastic differential equations (e.g., Brillinger et al., 2011), and combinations of both (Blackwell, 2003). However, simultaneous data on multiple individuals within a group are increasingly available, and the standard approaches typically ignore the potential interactions between individuals. Animals within a group generally do not move independently, and therefore it is inappropriate to model the simultaneous movement of group members without considering the dynamics of the group.

Among articles that do consider group dynamics, some concentrate on statistical descriptions of aspects of movement, rather than modeling the movement process; for example, del Mar Delgado et al. (2014) apply a linear model to derived measures of sociability, in a discrete time approach, while Strandburg-Peshkin et al. (2015) derive complex social structures from interaction events extracted from movement data. Haydon et al. (2008) incorporate interactions through behavioral switching rather than through the actual movement process. Potts et al. (2014) give a flexible framework allowing joint movement dependent on complex state variables,

in discrete time. In addition, there is a body of literature on theoretical aspects of group movement, such as Couzin et al. (2002) and Viscido et al. (2007), which explores how behavioral rules maintain and reorganize group structure in simulation experiments. In continuous time, Dunn and Gipson (1977) touched only briefly on the idea of interdependent movement. Brillinger et al. (2011) describe a flexible, continuous-time formulation, but their formal inference uses a simple discrete-time approximation, and their main example concentrates on just two elk. Recently, Langrock et al. (2014) proposed a discrete-time model of group movement, in which individual animals follow a biased random walk, attracted to a leading point which may represent either another individual with special status as a leader or simply a mathematical abstraction of the dependence between animals' movement. In their implementation, the position of the leading point is imputed as the one-step-ahead average location of the group.

The focus of the current article is to model group movement in continuous time, inspired by Dunn and Gipson (1977) and Langrock et al. (2014). Our work has similarities to Brillinger et al. (2011) but for our models we carry out inference exactly, without any time discretization or approximation. In addition, we have a much more parsimonious representation of interactions; instead of modeling the interaction between neighbors directly, as in the various approaches outlined above, we follow Langrock et al. (2014) and take all individuals in the group to be attracted to a latent moving point. Our continuous-time formulation gives a number of advantages, as discussed for single-animal models in Harris and Blackwell (2013) and McClintock et al. (2014). It gives a model which can be interpreted without reference to the timescale

of the data collection, which in general is arbitrary as far as the behavior of the animals is concerned, and it therefore adapts straightforwardly to irregular or missing data, and allows meaningful interpolation and comparison across different sampling regimes.

The structure of the article is as follows. In Section 2, group movement is modeled jointly using a multivariate Ornstein Uhlenbeck process. Section 3 describes fully Bayesian inference for this process, with the position of the unobserved leading point sampled using data augmentation. The method is applied and compared with the approach of Langrock et al. (2014) using simulation data in Section 4 and data from reindeer migration in Section 5.

## 2. Joint Movement Model

### 2.1. *Multivariate Ornstein Uhlenbeck Process*

We start by considering the movement of the leading point—or "leader," for brevity, although we retain the idea that it may be an abstraction rather than an actual animal. Its movement is modeled as a stationary, isotropic Ornstein Uhlenbeck (OU) process; stationarity allows us to model stable long-term behavior (though with Brownian motion possible as a limiting case). Rotational symmetry could be relaxed, but is rather natural in practice (Blackwell, 1997) and also implies that its $x$ and $y$ coordinates are two independent univariate OU processes. If the random variable $L_t^y$ represent the $y$ coordinate of the location of the leading point at time $t$, then a stochastic process $\{L_t^y : t \geq 0\}$ in which $L_t^y$ is attracted to $\theta^y$ is given by the stochastic differential equation (SDE; Schach, 1971; McNeil and Schach, 1973):

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

where $L_t^y$ is the location of the leader on the $y$ axis; $\beta$ is the attraction rate to $\theta^y$; $\rho$ is the coefficient for the noise; $V_t^y$ is standard Brownian motion. Here, we present only the movement model for the $y$ coordinate; the model is identical for the $x$ coordinate $L_t^x$, with parameters $\beta$ and $\rho$ in common, and independent Brownian motions $\{V_t^x\}$ and $\{V_t^y\}$ used for $L_t^x$ and $L_t^y$. Conditioning on the previous location of the leader $L_0^y$, the solution of this SDE can be written as

$$L_t^y = (L_0^y - \theta^y)\, e^{-\beta t} + \theta^y + \rho \int_0^t e^{-\beta(t-s)}\, \mathrm{d}V_s^y. \qquad (2)$$

Using Itō calculus (e.g., Lamberton and Lapeyre, 2007), given $L_0^y$ the expectation and variance of the current location $L_t^y$ are

$$\mathrm{E}\{L_t^y|L_0^y\} = e^{-\beta t}\,(L_0^y - \theta^y) + \theta^y, \qquad (3)$$

$$\mathrm{Var}\left\{L_t^y|L_0^y\right\} = \rho^2 \mathrm{E}\left\{\left(\int_0^t e^{-\beta(t-s)}\, \mathrm{d}V_s\right)^2\right\} = \frac{\rho^2}{2\beta}\left(1 - e^{-2\beta t}\right). \qquad (4)$$

A similar SDE can model the movement of a "follower" attracted at any instant to the current location of the leader.

Let random variables $F_t^x$ and $F_t^y$ represent the follower's location at time $t$. Under the same assumptions as before, $F_t^x$ and $F_t^y$ are independent, conditional on the movement process of the leader. $\{F_t^y : t \geq 0\}$ is defined by the following SDE with parameters $\alpha$, $\sigma$, $L_t^y$, and Brownian motion $\{W_t^y\}$, where $F_t^y$ is attracted to $L_t^y$:

$$\mathrm{d}F_t^y = -\alpha\,(F_t^y - L_t^y)\,\mathrm{d}t + \sigma \mathrm{d}W_t^y \qquad (5)$$

with $F_t^y$ the location of the follower and $L_t^y$ the location of the leader, on the $y$ axis; $\alpha$ the attraction rate to $L_t^y$; $\sigma$ the coefficient for the noise. A corresponding equation is used for the $x$ coordinate $F_t^x$, with common parameters and independent Brownian motion $\{W_t^x\}$.

Combining these equations with the model of the leading point gives an SDE for the $y$ coordinates of both leader and follower which defines a particular bivariate OU process:

$$\mathrm{d}\boldsymbol{Y_t} = A\,(\boldsymbol{Y_t} - \boldsymbol{\Theta^y})\,\mathrm{d}t + \Sigma \mathrm{d}\boldsymbol{B_t^y} \qquad (6)$$

where $\boldsymbol{Y_t} = \begin{pmatrix} L_t^y \\ F_t^y \end{pmatrix}$, $A = \begin{pmatrix} -\beta & 0 \\ \alpha & -\alpha \end{pmatrix}$, $\boldsymbol{\Theta^y} = \begin{pmatrix} \theta^y \\ \theta^y \end{pmatrix}$, $\Sigma = \begin{pmatrix} \rho & 0 \\ 0 & \sigma \end{pmatrix}$,

$\boldsymbol{B_t^y} = \begin{pmatrix} V_t^y \\ W_t^y \end{pmatrix}$. $\boldsymbol{Y_t}$ is a vector representing the $y$ coordinates of both leader and follower. The matrix $A$ is the attraction rate matrix. $\boldsymbol{\Theta^y}$ is a vector with both entries $\theta^y$; that the second element is $\theta^y$ follows from combining the equations, and indicates that their combined effect is that $F_t^y$ is indirectly attracted to $\theta^y$. We take the Brownian motions for the leader and the follower to be uncorrelated, therefore $\Sigma$ is a diagonal matrix; each diagonal element of the $\Sigma$ represents the coefficient of the individual variance. The solution of this multivariate SDE is given by

$$\boldsymbol{Y_t} = e^{At}\,(\boldsymbol{Y_0} - \boldsymbol{\Theta^y}) + \boldsymbol{\Theta^y} + \int_0^t \Sigma e^{A(t-s)}\, \mathrm{d}\boldsymbol{B_s^y} \qquad (7)$$

and can be shown to have an exact closed form as a multivariate normal distribution (Schach, 1971; McNeil and Schach, 1973):

$$\boldsymbol{Y_t}|\boldsymbol{Y_0} \sim \mathrm{MVN}\,(\boldsymbol{\mu}(L_0^y, F_0^y, t), \Xi(t))\,. \qquad (8)$$

The expectation of $\boldsymbol{Y_t}$ given $\boldsymbol{Y_0}$, denoted $\boldsymbol{\mu}(L_0^y, F_0^y, t)$, is

$$\mathrm{E}\,\{\boldsymbol{Y_t}|\boldsymbol{Y_0}\} = \boldsymbol{\mu}(L_0^y, F_0^y, t) = e^{At}\,(\boldsymbol{Y_0} - \boldsymbol{\Theta^y}) + \boldsymbol{\Theta^y}; \qquad (9)$$

$\boldsymbol{\mu}(L_0^y, F_0^y, t)$ has elements $\mu_{\mathrm{L}}(L_0^y, t)$ and $\mu_{\mathrm{F}}(L_0^y, F_0^y, t)$ given by

$$\mu_{\mathrm{L}}(L_0^y, t) = (L_0^y - \theta^y)\, e^{-\beta t} + \theta^y \qquad (10)$$

$$\mu_{\mathrm{F}}(L_0^y, F_0^y, t) = (L_0^y - \theta^y)\,\frac{\alpha}{\alpha - \beta}\left(e^{-\beta t} - e^{-\alpha t}\right)$$
$$+ (F_0^y - \theta^y)\, e^{-\alpha t} + \theta^y. \qquad (11)$$

The variance of $\boldsymbol{Y_t}$, denoted $\Xi(t)$, could be obtained as

$$
\begin{aligned}
\mathrm{Var}\{\boldsymbol{Y_t}|\boldsymbol{Y_0}\} &= \mathrm{E}\left\{\left(\int_0^t \Sigma e^{A(t-s)}\,\mathrm{d}\boldsymbol{B}_s^y\right)^2\right\} \\
&= e^{At}\left(\int_0^t e^{-As}\Sigma\Sigma^T e^{-A^T s}\mathrm{d}s\right)e^{A^T t} \quad (12)
\end{aligned}
$$

where $A^T$ denotes the transpose of $A$, etc. However, Schach (1971) indicated an alternative way to derive the variance $\mathrm{Var}\{\boldsymbol{Y_t}|\boldsymbol{Y_0}\}$ instead of solving the integral in equation (12) directly. The variance is determined by the Kolmogorov forward equations, and can be written as

$$
\mathrm{Var}\,\{\boldsymbol{Y_t}|\boldsymbol{Y_0}\} = \Xi(t) = \Delta - e^{At}\Delta e^{A^T t} \quad (13)
$$

where $\Delta$ is the stationary variance. It can be calculated from the local variance covariance

$$
\Sigma^T\Sigma = -\left(A\Delta + \Delta A^T\right) \quad (14)
$$

where the matrix $\Sigma$ is diagonal as above. This gives

$$
\Delta = \begin{pmatrix} \dfrac{\rho^2}{2\beta} & \dfrac{\rho^2\alpha}{2\beta(\alpha+\beta)} \\ \dfrac{\rho^2\alpha}{2\beta(\alpha+\beta)} & \dfrac{\sigma^2}{2\alpha}+\dfrac{\rho^2\alpha}{2\beta(\alpha+\beta)} \end{pmatrix} \quad (15)
$$

and hence from substitution into equation (13)

$$
\Xi(t) = \begin{pmatrix} \xi_{\mathrm{L}}(t) & \xi_{\mathrm{LF}}(t) \\ \xi_{\mathrm{LF}}(t) & \xi_{\mathrm{F}}(t) \end{pmatrix} \quad (16)
$$

with

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

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

- $\xi_{\mathrm{F}}(t) = \left\{\dfrac{\sigma^2}{2\alpha} + \dfrac{\rho^2\alpha}{2\beta\left(\alpha+\beta\right)}\right\}\left(1 - e^{-2\alpha t}\right) - \dfrac{\rho^2\alpha^2}{2\beta\left(\alpha-\beta\right)^2} \times$
$\left(e^{-\beta t} - e^{-\alpha t}\right)^2 - \dfrac{\rho^2\alpha^2}{\beta\left(\alpha^2-\beta^2\right)}\left\{e^{-(\alpha+\beta)t} - e^{-2\alpha t}\right\}.$

## 2.2. *Extension to More Followers*

The model in the previous section involves a single following animal, for illustration, but in reality our interest is almost always in cases with multiple followers. In the interests of parsimony, in contrast with the model proposed by Dunn and Gipson (1977) we assume that followers are conditionally independent of each other, given the location of the leader, though structured departures from that assumption could certainly be accommodated. For simplicity, we also assume that the followers have identical parameters; again, this assumption could be relaxed in principle, perhaps through a hierarchical model on individual parameters, provided there were sufficient data in a given case to make it feasible.

Modeling the interactions of animals through the leading point, rather than directly with each other, also has the advantage that the model is applicable in cases where some individuals are unobserved, as is unavoidable for many large herds; this applies to our main example in Section 5. In such a situation, observing only some of the animals obviously gives less information, but does not systematically change the results.

The formulation shown in equation (6), the form of its solution in equation (8), and the method of calculating the covariance still apply. To illustrate the general case, the key parameters are shown here for the case of two followers and one leading point; further extension is trivial. As before, $\alpha$ is the attraction rate of a follower to the leading point and $\sigma$ is the variance coefficient of followers; different followers are assumed to share the same $\alpha$ and $\sigma$. We have

$$
\boldsymbol{\mu}(L_0^y, F_0^y, t) = \begin{pmatrix} \mu_{\mathrm{L}}(L_0^y, t) \\ \mu_{\mathrm{F}}(L_0^y, F_0^y, t) \\ \mu_{\mathrm{F}}(L_0^y, F_0^y, t) \end{pmatrix}, \quad (17)
$$

$$
A = \begin{pmatrix} -\beta & 0 & 0 \\ \alpha & -\alpha & 0 \\ \alpha & 0 & -\alpha \end{pmatrix}, \quad (18)
$$

$$
\Sigma = \begin{pmatrix} \rho & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & \sigma \end{pmatrix} \quad (19)
$$

and the solution for $\Delta$ is

$$
\Delta = \begin{pmatrix} \dfrac{\rho^2}{2\beta} & \dfrac{\rho^2\alpha}{2\beta(\alpha+\beta)} & \dfrac{\rho^2\alpha}{2\beta(\alpha+\beta)} \\ \dfrac{\rho^2\alpha}{2\beta(\alpha+\beta)} & \dfrac{\sigma^2}{2\alpha}+\dfrac{\rho^2\alpha}{2\beta(\alpha+\beta)} & \dfrac{\rho^2\alpha}{2\beta(\alpha+\beta)} \\ \dfrac{\rho^2\alpha}{2\beta(\alpha+\beta)} & \dfrac{\rho^2\alpha}{2\beta(\alpha+\beta)} & \dfrac{\sigma^2}{2\alpha}+\dfrac{\rho^2\alpha}{2\beta(\alpha+\beta)} \end{pmatrix}. \quad (20)
$$

Note that all off-diagonal elements of $\Delta$ are identical and the diagonal elements are a function of individual variance terms plus the off-diagonal element. Finally, we have

$$
\Xi(t) = \begin{pmatrix} \xi_{\mathrm{L}}(t) & \xi_{\mathrm{LF}}(t) & \xi_{\mathrm{LF}}(t) \\ \xi_{\mathrm{LF}}(t) & \xi_{\mathrm{F}}(t) & \xi_{\mathrm{FF}}(t) \\ \xi_{\mathrm{LF}}(t) & \xi_{\mathrm{FF}}(t) & \xi_{\mathrm{F}}(t) \end{pmatrix} \quad (21)
$$

where $\xi_{\mathrm{L}}(t), \xi_{\mathrm{LF}}(t), \xi_{\mathrm{F}}(t)$ are as before, and

$$
\xi_{\mathrm{FF}}(t) = \dfrac{\rho^2\alpha}{2\beta\left(\alpha+\beta\right)}\left(1 - e^{-2\alpha t}\right) - \dfrac{\rho^2\alpha^2}{2\beta\left(\alpha-\beta\right)^2}\left(e^{-\beta t} - e^{-\alpha t}\right)^2
$$
$$
- \dfrac{\rho^2\alpha^2}{\beta\left(\alpha^2-\beta^2\right)}\left\{e^{-(\alpha+\beta)t} - e^{-2\alpha t}\right\}. \quad (22)
$$

For the purposes of Section 3, we will usually be dealing with updates over a fixed unit interval, in which case we simplify notation by suppressing the dependence on the interval $t$; it is also useful to write $\boldsymbol{\mu}(L_0^y, F_0^y)$ and $\Xi$ in block form as

$$\boldsymbol{\mu}(L_0^y, F_0^y) = \left( \frac{\mu_{\mathrm{L}}(L_0^y)}{\boldsymbol{\mu}_{\mathrm{F}}(L_0^y, F_0^y)} \right), \qquad (23)$$

$$\Xi = \left( \begin{array}{c|c} \xi_{\mathrm{L}} & \boldsymbol{\xi}_{\mathrm{LF}}{}^T \\ \hline \boldsymbol{\xi}_{\mathrm{LF}} & \Xi_{\mathrm{F}} \end{array} \right), \qquad (24)$$

where in the general case $\boldsymbol{\mu}(\cdot, \cdot)$ and $\boldsymbol{\xi}_{\mathrm{LF}}$ are vectors and $\Xi_{\mathrm{F}}$ a square matrix, of size equal to the number of followers.

## 3. Inference

Here, we describe a Markov Chain Monte Carlo approach to carrying out inference for this model, based on location data for the followers only, as is realistic for most of the applications we have in mind. For simplicity, we focus on data that are equally spaced in time, and define our time unit so that observations occur at times $0, 1, 2, \ldots$, though extension to irregular spacing, for example to accommodate missing data, is straightforward in principle.

Given the model in Section 2, the likelihood calculation in Section 3.1 can be carried out separately and identically for each coordinate; here we drop the superscript $x$ or $y$ to simplify notation.

### 3.1. Inference for the Location of the Leading Point

Since the leading point is assumed unobserved, we augment the data (Tanner and Wong, 1987) with its locations at the times of the observations; this augmentation gives us the complete state of the system at these instants, which combined with the Markov property gives a more tractable method for inference, since the likelihood for the parameters can then be written as a product over time. We sample a single value $L_t$ at a time, using a Metropolis–Hastings (MH) step. To do this, we need its conditional density

$$p(L_t | \boldsymbol{L}_{(t)}, \boldsymbol{F}) \propto p(L_t | \boldsymbol{L}_{(t)}) p(\boldsymbol{F} | L_t, \boldsymbol{L}_{(t)}) \qquad (25)$$

where $\boldsymbol{F} = (\boldsymbol{F_1} | \boldsymbol{F_2} | \cdots)$ represents the followers' locations at all time points, with each $\boldsymbol{F_t}$ a vector of length equal to the number of followers, $L_t$ is the leading point's location at time $t$ and $\boldsymbol{L}_{(t)}$ represents the leading point's locations at all other time points. The terms on the RHS of equation (25) are calculated in Sections 3.1.1 and 3.1.2; Section 3.1.3 gives details of the actual algorithm.

*3.1.1. Conditioning on the leader's movement process.* From the Markov property of our model for $L_t$ in each dimension, conditioning on $\boldsymbol{L}_{(t)}$ is equivalent to conditioning on $L_{t-1}$ and $L_{t+1}$ (the leading point locations at one time step before and after). Using standard results for the OU process, $L_t$ then has a normal distribution with mean $\mu_{L_t}$ and variance $\nu_{L_t}$,

$$L_t | \boldsymbol{L}_{(t)} \equiv L_t | L_{t-1}, L_{t+1} \sim N\left(\mu_{L_t}, \nu_{L_t}\right), \qquad (26)$$

where $\mu_{L_t}$ and $\nu_{L_t}$ can be calculated from the OU bridge (Levy, 2008):

$$\mu_{L_t} = \left\{ (L_{t-1} + L_{t+1} - 2\theta)\, e^{-\beta} + \theta \left(1 + e^{-2\beta}\right) \right\} \frac{\left(1 - e^{-2\beta}\right)}{\left(1 - e^{-4\beta}\right)}, \qquad (27)$$

$$\nu_{L_t} = \frac{\rho^2 \left(1 - e^{-2\beta}\right)^2}{2\beta \left(1 - e^{-4\beta}\right)}. \qquad (28)$$

*3.1.2. Conditioning on the data.* As a function of $L_t$, the second term on the RHS of equation (25) is

$$p\left(\boldsymbol{F} | L_t, \boldsymbol{L}_{(t)}\right) \propto p\left(\boldsymbol{F_t} | \boldsymbol{F_{t-1}}, L_{t-1}, L_t\right) p\left(\boldsymbol{F_{t+1}} | \boldsymbol{F_t}, L_t, L_{t+1}\right), \qquad (29)$$

again by the Markov property. For the first term on the RHS of equation (29), we have

$$\boldsymbol{F_t} | \boldsymbol{F_{t-1}}, L_{t-1}, L_t \sim MVN\left(\boldsymbol{\mu}_{F_t}, \Sigma_{F_t}\right). \qquad (30)$$

To get the parameters of this normal distribution, we take equation (8) describing the forward evolution of the whole system and condition on $L_t$. From the standard result for conditioning a multivariate normal on some of its components, and the partitioned form of the forward mean and covariance in equations (23) and (24), the conditioned mean of the followers $\boldsymbol{\mu}_{F_t}$ is

$$\boldsymbol{\mu}_{F_t} = \boldsymbol{\mu}_{\mathrm{F}}(L_{t-1}, \boldsymbol{F_{t-1}}) + \boldsymbol{\xi}_{\mathrm{LF}} \xi_{\mathrm{L}}^{-1} \left(L_t - \mu_{\mathrm{L}}(L_{t-1})\right) \qquad (31)$$

where $\mu_{\mathrm{L}}(L_{t-1})$ is the mean of the leading point at time $t$, given $L_{t-1}$, and similarly $\boldsymbol{\mu}_{\mathrm{F}}(L_{t-1}, \boldsymbol{F_{t-1}})$ is the mean of the followers at time $t$. $\boldsymbol{\xi}_{\mathrm{LF}}$ and $\xi_{\mathrm{L}}$ are as defined by equation (24). Similarly, $\Sigma_{F_t}$ is the followers' conditional variance

$$\Sigma_{F_t} = \Xi_{\mathrm{F}} + \boldsymbol{\xi}_{\mathrm{LF}} \xi_{\mathrm{L}}^{-1} \boldsymbol{\xi}_{\mathrm{LF}}. \qquad (32)$$

The second term on the RHS of equation (29) can be derived in a similar way, with

$$\boldsymbol{F_{t+1}} | \boldsymbol{F_t}, L_{t+1}, L_t \sim MVN\left(\boldsymbol{\mu}_{F_{t+1}}, \Sigma_{F_{t+1}}\right) \qquad (33)$$

$$\boldsymbol{\mu}_{F_{t+1}} = \boldsymbol{\mu}_{\mathrm{F}}(L_t, \boldsymbol{F_t}) + \boldsymbol{\xi}_{\mathrm{LF}} \xi_{\mathrm{L}}^{-1} \left(L_{t+1} - \mu_{\mathrm{L}}(L_t)\right) \qquad (34)$$

$$\Sigma_{F_{t+1}} = \Sigma_{F_t}. \qquad (35)$$

Note that it *is* necessary to consider the followers jointly in evaluating these likelihood terms; see Section 3.1.4 for a brief explanation.

*3.1.3. Sampling $L_t$.* From Sections 3.1.1 and 3.1.2, the conditional distribution in equation (25) is the product of

the densities of the normal distributions in equations (26), (30), and (33). For the sampling of $L_t$, we use a Metropolis–Hastings (MH) step. Given $L_{t-1}$ and $L_{t+1}$, we propose a new location $L'_t$ for the leader at time $t$ from the OU bridge defined in equation (26). The proposal density cancels with the OU bridge term in the likelihood; the Hastings ratio is therefore the product of the densities in equations (30) and (33) evaluated at $L'_t$ divided by the corresponding product evaluated at $L_t$.

*3.1.4. Conditional independence.* Because the augmentation with locations of the leader takes place only at the times of the observations, the followers' movements are not conditionally independent even given this augmentation; the "instantaneous" conditional independence involved in defining the model would carry over to finite intervals only given augmentation with the location of the leader at every instant, which is not feasible in practice. Equation (32) captures the dependence between followers, conditional on the location of the leader at the start and end of an interval, in terms of the variance and covariance terms in equation (21) et seq. We can gain some insight by considering the effect of varying the time interval (written here as $\tau$, to avoid confusion with the $t$ indexing the observations). As might be expected given that followers are conditionally independent given complete information about the position of the leader, $\xi_{\text{FF}}(\tau)$ is small compared with $\xi_{\text{L}}(\tau), \xi_{\text{F}}(\tau)$ and $\xi_{\text{LF}}(\tau)$; for small $\tau$, $\xi_{\text{FF}}(\tau)$ is of lower order in $\tau$ than the other terms. From a Taylor expansion around $\tau = 0$, purely for illustration (not used in the analysis, which is exact), $\xi_{\text{L}}(\tau) = \rho^2\tau + O(\tau^2)$, $\xi_{\text{F}}(\tau) = \sigma^2\tau + O(\tau^2)$, $\xi_{\text{LF}}(\tau) = \frac{1}{2}\rho^2\alpha\tau^2 + O(\tau^3)$, while $\xi_{\text{FF}}(\tau) = \frac{1}{3}\rho^2\alpha^2\tau^3 + O(\tau^4)$. This confirms that, for short intervals, the followers are very close to being conditionally independent given the initial location of the leader. Nevertheless, our exact analysis requires use of the full joint likelihood terms in equations (30) and (33).

### 3.2. Inference for Parameters

Given the augmentation of the data with the leader's locations, the parameters of the movement process can be sampled straightforwardly. From the Markov property, the likelihood is a product over time of terms of the form given in equation (30) or (33). Again, the followers need to be considered jointly, as explained in Section 3.1.4. We use a conjugate normal prior for the point of attraction $(\theta^x, \theta^y)$, and so a standard "normal–normal" Gibbs step can be used. For the other movement parameters, we use uniform priors on $[0, \infty)$ and standard random walk MH updates. In either case, 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).

### 4. Implementation with Simulated Data

Here, we look at inference for simulated data, where the position of the leader is known, but treated as unobserved. Langrock et al. (2014), in a similar model in discrete time, also used an unobserved leading point, but imputed its location approximately as the average of followers' locations from one step ahead. Here, our estimation of the leading point is compared visually against both the true location and the ap-

proximation of Langrock et al. (2014); a more formal statistical comparison, based on real data, is given in Section 5.

We simulated the location of six followers and one leading point from our model for 50 steps, using equation (8) iteratively and taking each generated location as the origin for the next. We then applied our MCMC algorithm to reconstruct the location of the leader and the model parameters, running it for 10,000 iterations after burn-in, from a starting point based on the same kind of one-step-ahead calculation that Langrock et al. (2014) used for their inference. Figure 1 shows selected data, point estimates, and posterior credible regions, with the four subfigures representing consecutive time steps as the leading point moves from top left to bottom right, taking it through the group of followers.

In Figure 1(a), the true leading point is outside the group. The OU estimate follows its movement and is located away from the group. In Figure 1(b), the leading point changes its direction and moves into the group. In this case, the one-step-ahead average is closer to the true point. In Figure 1(c), the OU estimate follows the movement of the true leading point and moves into the group. In Figure 1(d), the leading point moves away from the group and the OU estimate follows it well while the one-step-ahead average (red square) is still sitting within the group.

The posterior mean and standard deviation of model parameters are shown in Table 1, along with the actual values used in simulation. Posterior density plots for the parameters are given in Figure 2. All the posterior distributions are consistent with the true values; the key parameter $\alpha$, representing attraction of the followers to the leader, is estimated quite well, from this relatively small dataset, whereas the parameter $\beta$, representing the attraction of the (imputed) leader to its (estimated) center of attraction is estimated much less precisely. The MCMC trace plot is also provided in Web Appendix A.

These results show that our fully Bayesian approach can reconstruct the location of the unobserved leader in a case where the attraction mechanism is appropriate, and also appears from the plots to outperform the simpler intuitive one-step-ahead approach used by Langrock et al. (2014), giving some justification for our more elaborate modeling.

### 5. Implementation with Real Data

In this section, our group dynamic model is fitted to location data of reindeer from Skarin et al. (2008). The annual migration of reindeer follows a seasonal progression of snowmelt and fresh vegetative growth that broadly describes the general movement pattern of the population (Skarin et al., 2008, 2010). 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).

The data here are hourly GPS fixes on reindeer within a single herd, arising from a study of habitat use in the reindeer herding area in Sweden (Skarin et al., 2008). We consider a subset of the available reindeer data, with 58 observations
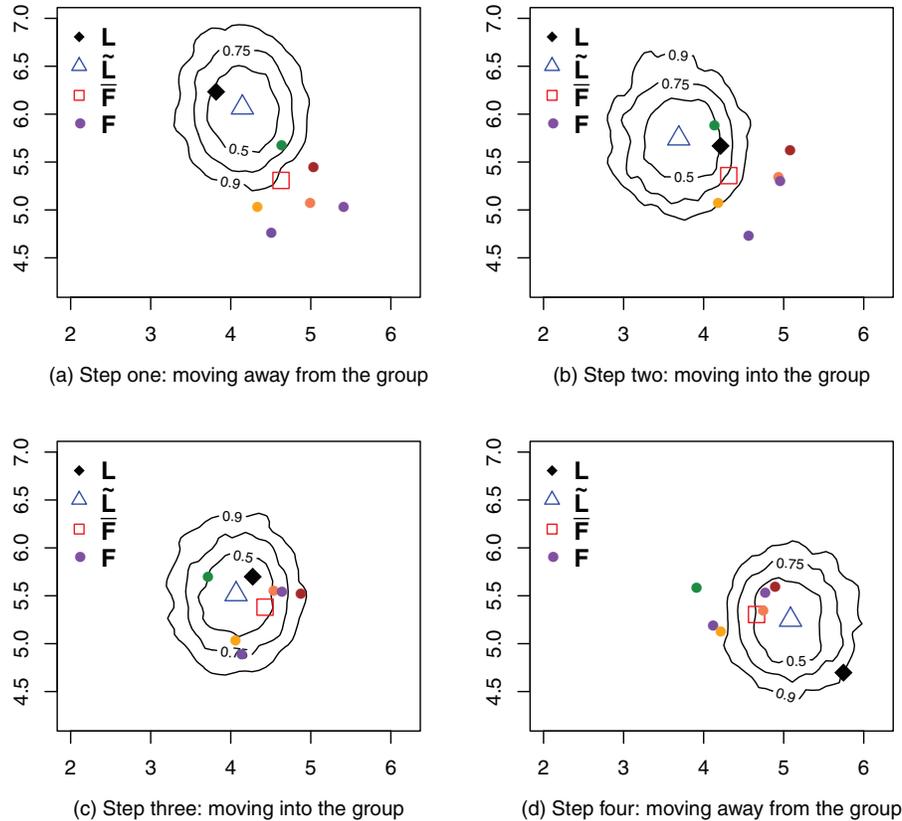
**Figure 1.** True leading point and estimates from simulated data, as the true leader moves from top left to bottom right. Each plot shows the true leading point L (diamond), its posterior mean location $\tilde{\mathsf{L}}$ (triangle), the individual followers F (circles), and their one-step-ahead mean $\bar{\mathsf{F}}$ (square). Thus, the point $\bar{\mathsf{F}}$ in (a) is the average of the followers' positions in (b), etc. Contours show credible regions, with posterior probabilities as indicated, for the estimated position of the leader. This figure appears in color in the electronic version of this article.

on each of 5 individual "followers." This subset is selected to give what appears to be reasonably homogeneous behavior, since incorporating switching between behaviors is beyond the scope of the current article—see Section 6—and to avoid missing data; while it is straightforward in principle to allow for incomplete observations, it entails quite a lot of additional coding.

The computational time for this example is a few hours on a standard PC, using code written wholly in R (R Core Team, 2014). The computational effort should scale reasonably well

**Table 1**
*Parameter estimates for the movement model with simulation data*

| Parameter | Point estimate | Standard deviation | True value |
|-----------|----------------|--------------------|------------|
| $\theta^x$ | 4.574 | 1.181 | 4.0 |
| $\theta^y$ | 4.569 | 1.183 | 4.0 |
| $\beta$ | 0.129 | 0.061 | 0.1 |
| $\sigma$ | 0.204 | 0.007 | 0.2 |
| $\alpha$ | 0.214 | 0.020 | 0.2 |
| $\rho$ | 0.710 | 0.102 | 0.86 |

with the number of followers; the most expensive element in the computation is the matrix exponential in equation (13) (using `expm`; Goulet et al., 2014) which needs to be carried out only when the parameters are updated, and not when the leader's location changes. Thus, we expect that this approach would be computationally feasible for most practical field experiments involving simultaneous tracking of tagged wildlife (perhaps less so for laboratory experiments involving observations of many individuals). Our code would undoubtedly benefit from recoding in a compiled language; we are also working on improvements to computational aspects of the algorithm. The current code is included as Web Appendix C.

All parameters of the model are unknown, but for simplicity we assume the parameters are the same for all followers. The initial location of the leader is the average of all followers' locations. The results here are based on 30,000 iterations of two MCMC runs with over-dispersed initial values, every second iteration being recorded after 5000 iterations of burn-in. The Gelman–Rubin diagnostic is used to check convergence. The potential scale reduction factors are close to 1 for all model parameters; see Web Appendix B.

Table 2 shows posterior means and standard deviations for the parameters of the model, indicating a number of useful
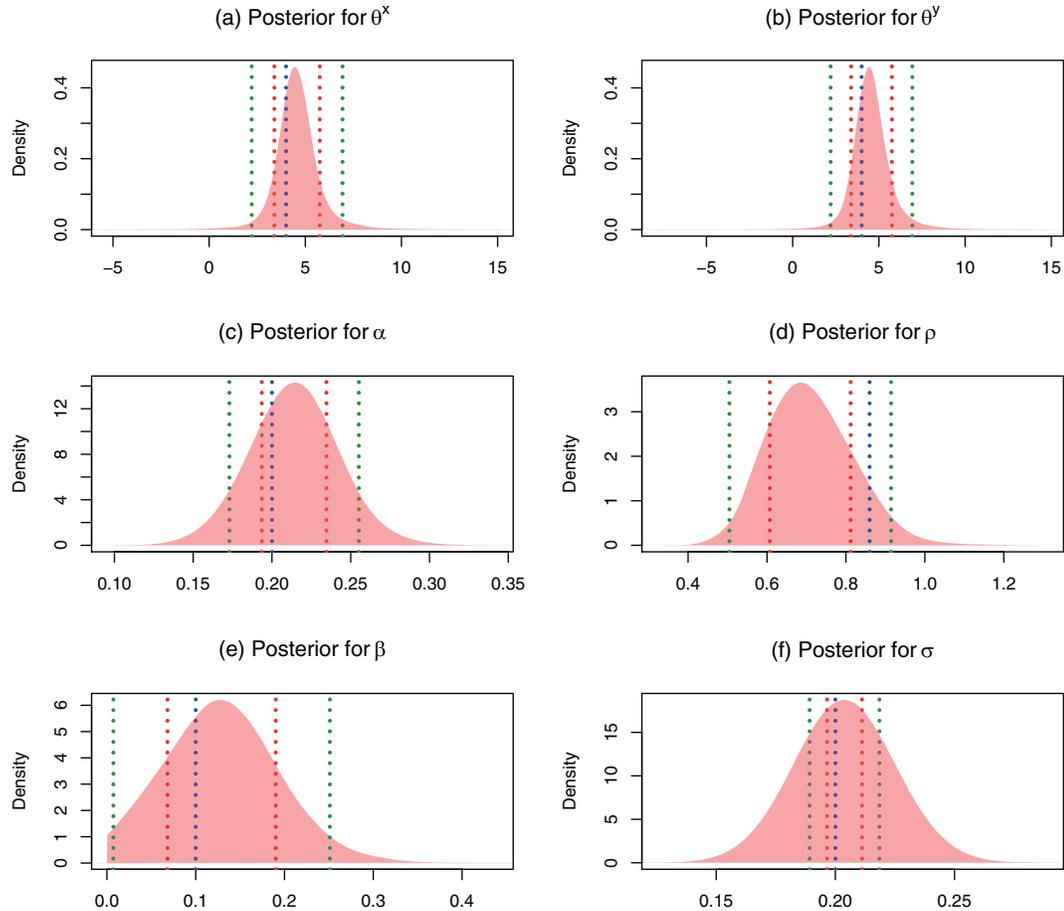
**Figure 2.** Posterior densities for model parameters with simulation data, based on the Markov Monte Carlo runs of 10,000 iterations. The solid line represents the true parameter; dotted lines represent the posterior mean $\pm 1$ and $\pm 2$ posterior standard deviations. (a) Posterior density for $\theta^x$, the fixed attractor on horizontal direction. (b) Posterior density for $\theta^y$, the fixed attractor on vertical direction. (c) Posterior density for $\alpha$, the attraction rate of the follower to the leading point. (d) Posterior density for $\sigma$, the individual variance coefficient of the follower. (e) Posterior density for $\beta$, the attraction rate of the leading point to the fixed attractor. (f) Posterior density for $\rho$, the individual variance coefficient of the leading point. This figure appears in color in the electronic version of this article.

conclusions. Most importantly, there is clear evidence from the posterior distribution for $\alpha$ that there *is* attraction of the individual animals to a leading point, i.e., clear dependence between their movements. This is pursued in Section 5.1. On the other hand, from the posterior distribution for $\beta$, there is no clear indication of attraction of this leading point to a fixed attractor; it is possible that, on the time-scale of the

data analyzed at least, a simpler model for the movement of the leading point may be adequate, e.g., Brownian motion. Finally, comparing the values of $\rho$ and $\sigma$ in Table 2 and Figure 3, we can see that the former is likely to be much larger than the latter, meaning that the movement of the leading point is much "noisier" than that of individual animals; much of the individual movement can be explained by the common movement toward the leading point. Locations of the leader and followers at selected times are plotted in Figure 4 to match the scenario of Figure 1. The plots correspond to $t = 29, 31, 32, 37$ within the dataset; thus Fig. 4(b) and (c) represent consecutive observations, with Fig. 4(a) and (d) further separated before and after them, to make the dynamics clearer.

### 5.1. Comparison with Other Models

We compare our model more formally with two alternatives. Firstly, we look at the case where $\alpha = 0$, so that the followers are not attracted to any common leader. The followers then undergo independent Brownian motions, and the result-
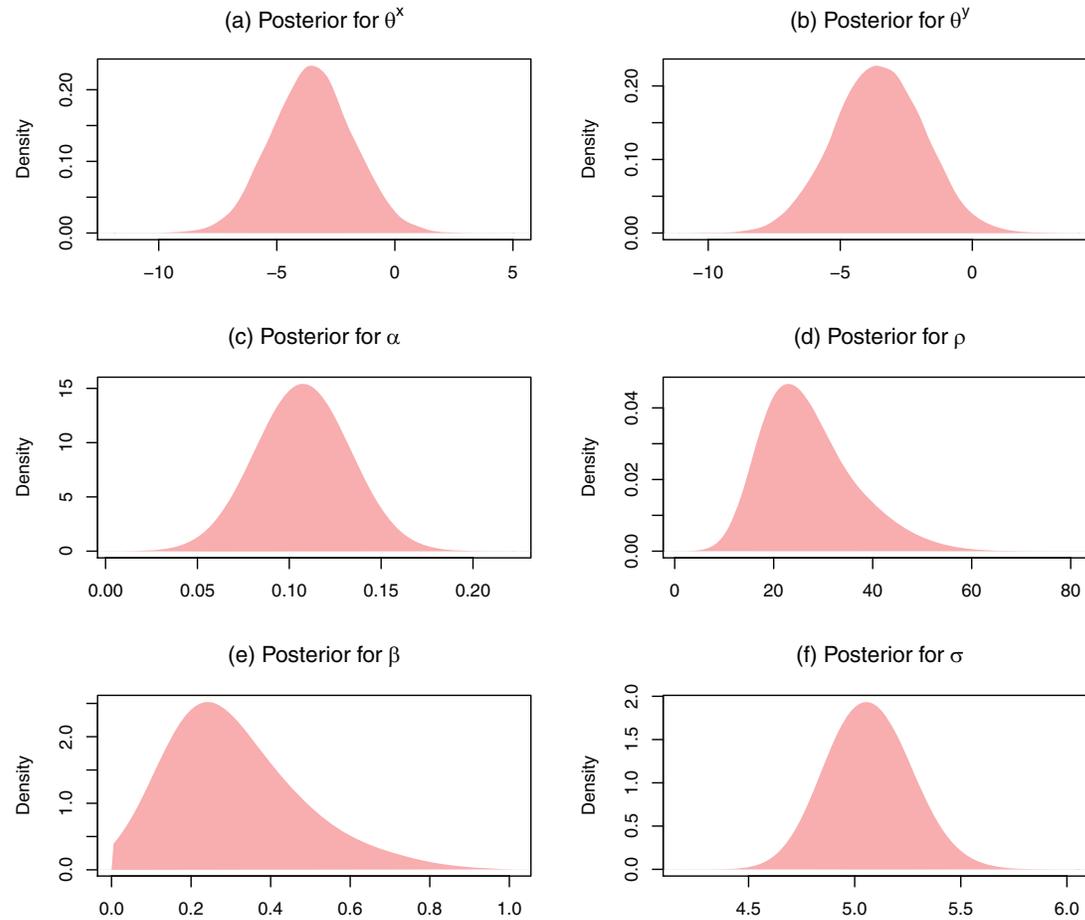
**Table 2**
*Parameter estimates for the reindeer movement model*

| Parameter | Point estimate | Standard deviation |
|-----------|----------------|--------------------|
| $\theta^x$ | $-3.508$ | 1.712 |
| $\theta^y$ | $-3.543$ | 1.699 |
| $\alpha$ | 0.107 | 0.016 |
| $\beta$ | 0.307 | 0.156 |
| $\sigma$ | 5.067 | 0.176 |
| $\rho$ | 27.05 | 8.78 |

**Figure 3.** Posterior densities for model parameters, based on the Markov Monte Carlo runs of 30,000 iterations. (a) Posterior density for $\theta^x$, the fixed attractor on horizontal direction. (b) Posterior density for $\theta^y$, the fixed attractor on vertical direction. (c) Posterior density for $\alpha$, the attraction rate of the follower to the leading point. (d) Posterior density for $\sigma$, the individual variance coefficient of the follower. (e) Posterior density for $\beta$, the attraction rate of the leading point to the fixed attractor. (f) Posterior density for $\rho$, the individual variance coefficient of the leading point. This figure appears in color in the electronic version of this article.

ing model has just one parameter, their shared diffusion parameter. Secondly, motivated by Langrock et al. (2014), we take the location of the leading point to be *known* and given by the one-step-ahead average of the locations of the followers. This simplifies the computation substantially. In practice, we can calculate these "known" locations in advance, and then use our algorithm omitting the updating of the leader. To compare the models, we use the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002). The values obtained are as follows: full model, 3426.5; model with known leader, 3526.7; independent Brownian motions, 3600.3. Thus, the full model, in which the trajectory of the leading point is reconstructed based on the model, with uncertainty, is substantially better than the simpler models in terms of DIC, despite its much larger effective number of parameters, around 40 for this example.

## 6. Conclusion

We have described the formulation of a group movement model in continuous time, building on some of the strengths of previous approaches, and an algorithm for fully Bayesian inference, which outperforms the simpler approach previously used in a related discrete-time analysis. We have shown that even with relatively small amounts of data, we can extract an estimated location for a notional leading point which gives a good representation of the common component of the movement of a group of real animals.

We have concentrated on the case where the position of the leader is unknown, either because the leader is an individual who is not tracked or because it is a modeling construct enabling a parsimonious representation of dependence between actual animals' movements. We could take a similar approach when the leader is a tracked animal; inference would be simpler and faster, since the leader's location would be known rather than sampled. If the leader was one of the tracked animals, but its identity was unknown, it would be straightforward to repeat this faster analysis with each animal in turn taken as leader, to determine the most likely candidate(s).

Our motivation and data, and most of our discussion here and in the Introduction, relate to dependent movement of
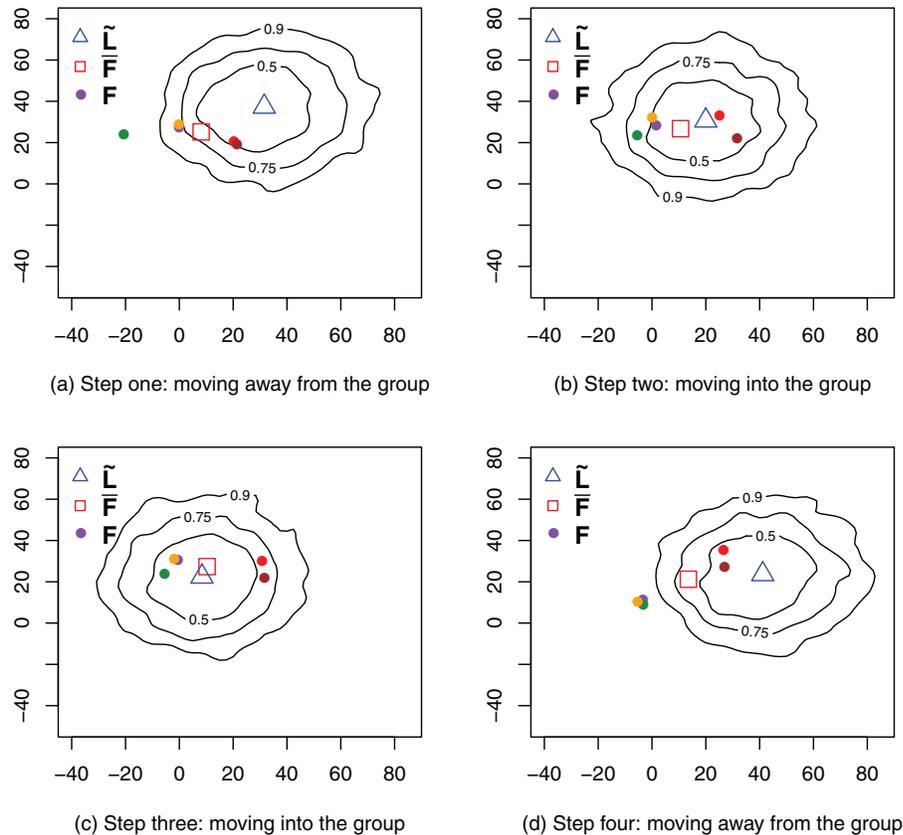
**Figure 4.** Estimation of the leading point from reindeer data. Co-ordinates are in hundreds of meters (with arbitrary origin). Each plot shows the posterior mean location of the leading point $\tilde{\mathsf{L}}$ (triangle), the individual followers $\mathsf{F}$ (circles), and their lagged mean $\bar{\mathsf{F}}$ (square). Contours show credible regions, with posterior probabilities as indicated, for the estimated position of the leader. The four cases are selected to illustrate different configurations that arise, similar to those in the simulations in Figure 1; they correspond to $t = 29, 31, 32, 37$ within the data-set. This figure appears in color in the electronic version of this article.

individual animals, particularly wildlife; but potentially the models and methods apply in other settings too. For example, there is considerable interest in modeling the collective movement of biological cells—see, for example, Binny et al. (2015)—but apparently much less work on inferential aspects. We hope that there may be scope for applying the current ideas in that context.

We have not considered here the possibility of animals switching between behavioral states, likely to be important in applying this work to a wider range of real situations c.f. Blackwell (1997, 2003); Gurarie et al. (2009); Haydon et al. (2008); Morales and Ellner (2002); Langrock et al. (2014). Work in progress allows followers to have multiple states, with switching between them modeled as a continuous time Markov chain, as in Blackwell (1997, 2003). Different states represent attraction to different leaders or independent movement such as Brownian motion. Conditional on the behavior, the overall system remains an Ornstein Uhlenbeck process and a similar approach to that developed here applies.

### 7. Supplementary Materials

Web Appendices A (MCMC trace plots), B (MCMC convergence diagnostics), and C (data and code) referenced in Sections 4 and 5 are available with the online version of this article at the *Biometrics* website on Wiley Online Library.

### References

Binny, R. N., Plank, M. J., and James, A. (2015). Spatial moment dynamics for collective cell movement incorporating a neighbour-dependent directional bias. *Journal of The Royal Society Interface* **12**, 20150228.

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.

Bowler, D. E. and Benton, T. G. (2005). Causes and consequences of animal dispersal strategies: Relating individual behaviour to spatial dynamics. *Biological Reviews* **80**, 205–225.

Brillinger, D. R., Preisler, H. K., and Wisdom, M. J. (2011). Modelling particles moving in a potential field with pairwise interactions and an application. *Brazilian Journal of Probability and Statistics* **25**, 421–436.

Couzin, I. D., Krause, J., James, R., Ruxton, G. D., and Franks, N. R. (2002). Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology* **218**, 1–11.

del Mar Delgado, M., Penteriani, V., Manuel Morales, J., Gurarie, E., and Ovaskainen, O. (2014). A statistical framework for inferring the influence of conspecifics on movement behaviour. *Methods in Ecology and Evolution* **5**, 183–189.

Dunn, J. E. and Gipson, P. S. (1977). Analysis of radio telemetry data in studies of home range. *Biometrics* **33**, 85–101.

Goulet, V., Dutang, C., Maechler, M., Firth, D., Shapira, M., Stadelmann, M., et al. (2014). *expm: Matrix exponential*. R package version 0.99-1.1.

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

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.

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**, 1208–1215.

Lamberton, D. and Lapeyre, B. (2007). *Introduction to Stochastic Calculus Applied to Finance*. Chapman and Hall/CRC; 2 edition; London.

Langrock, R., Hopcraft, J. G. C., Blackwell, P. G., Goodall, V., King, R., Niu, M., et al. (2014). Modelling group dynamic animal movement. *Methods in Ecology and Evolution* **5**, 190–199.

Levy, G. (2008). *Computational Finance Using C and C#*. New York: Academic Press.

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**, 21.

McNeil, D. R. and Schach, S. (1973). Central limit analogues for Markov population processes. *Journal of the Royal Statistical Society. Series B (Methodological)* **35**, 1–23.

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.

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

Patterson, T. A., Thomas, L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J. (2008). State–space models of individual animal movement. *Trends in Ecology & Evolution* **23**, 87–94.

Potts, J. R., Mokross, K., and Lewis, M. A. (2014). A unifying framework for quantifying the nature of animal interactions. *Journal of The Royal Society Interface* **11**, 20140333.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

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

Skarin, A., Danell, Ö., Bergström, R., and Moen, J. (2008). Summer habitat preferences of GPS-collared reindeer Rangifer tarandus tarandus. *Wildlife Biology* **14**, 1–15.

Skarin, A., Danell, Ö., Bergström, R., and Moen, J. (2010). Reindeer movement patterns in alpine summer ranges. *Polar Biology* **33**, 1263–1275.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 583–639.

Strandburg-Peshkin, A., Farine, D. R., Couzin, I. D., and Crofoot, M. C. (2015). Shared decision-making drives collective movement in wild baboons. *Science* **348**, 1358–1361.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**, 528–540.

Viscido, S. V., Parrish, J. K., and Grunbaum, D. (2007). Factors influencing the structure and maintenance of fish schools. *Ecological Modelling* **206**, 153–165.