

Modelling group dynamic animal movement[†]

Roland Langrock^{1*}, J. Grant C. Hopcraft^{1,2}, Paul G. Blackwell³, Victoria Goodall^{4,5}, Ruth King¹, Mu Niu³, Toby A. Patterson⁶, Martin W. Pedersen⁷, Anna Skarin⁸ and Robert S. Schick¹

¹Center for Research into Ecological and Environmental Modelling, School of Mathematics and Statistics, University of St Andrews, St Andrews, UK; ²Boyd Orr Centre for Population and Ecosystem Health, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow, UK; ³School of Mathematics and Statistics, University of Sheffield, Sheffield, UK; ⁴South African Environmental Observation Network, Fynbos Node, South Africa; ⁵School of Statistics and Actuarial Science, University of the Witwatersrand, Witwatersrand, South Africa; ⁶Commonwealth Scientific and Industrial Research Organisation, Wealth from Oceans Research Flagship, Hobart, Tas., Australia; ⁷National Institute of Aquatic Resources, Technical University of Denmark, Charlottenlund, Denmark; and ⁸Department of Animal Nutrition and Management, Swedish University of Agricultural Sciences, Uppsala, Sweden

Summary

1. Group dynamics are a fundamental aspect of many species' movements. The need to adequately model individuals' interactions with other group members has been recognized, particularly in order to differentiate the role of social forces in individual movement from environmental factors. However, to date, practical statistical methods, which can include group dynamics in animal movement models, have been lacking.
2. We consider a flexible modelling framework that distinguishes a group-level model, describing the movement of the group's centre, and an individual-level model, such that each individual makes its movement decisions relative to the group centroid. The basic idea is framed within the flexible class of hidden Markov models, extending previous work on modelling animal movement by means of multistate random walks.
3. While in simulation experiments parameter estimators exhibit some bias in non-ideal scenarios, we show that generally the estimation of models of this type is both feasible and ecologically informative.
4. We illustrate the approach using real movement data from 11 reindeer (*Rangifer tarandus*). Results indicate a directional bias towards a group centroid for reindeer in an encamped state. Though the attraction to the group centroid is relatively weak, our model successfully captures group-influenced movement dynamics. Specifically, as compared to a regular mixture of correlated random walks, the group dynamic model more accurately predicts the non-diffusive behaviour of a cohesive mobile group.
5. As technology continues to develop, it will become easier and less expensive to tag multiple individuals within a group in order to follow their movements. Our work provides a first inferential framework for understanding the relative influences of individual versus group-level movement decisions. This framework can be extended to include covariates corresponding to environmental influences or body condition. As such, this framework allows for a broader understanding of the many internal and external factors that can influence an individual's movement.

Key-words: behavioural state, hidden Markov model, maximum likelihood, random walk

Introduction

In ecology, there is growing interest in understanding and modelling animal movement (Nathan *et al.* 2008). For example, understanding movement patterns is critical when considering the potential consequences of land use, climate change or anthropogenic activities on range expansions, the spread of invasive species, or movement of hosts and their pathogens into new vulnerable areas (Bowler & Benton 2005). In addition, quantitative descriptions of animals' movements may also contribute to our understanding of the underlying

movement decisions (Schick *et al.* 2008). Many of these decisions are driven by social factors (Haydon *et al.* 2008), which highlights the importance of being able to adequately model group dynamic movement patterns.

Following advances in tracking technology, recent years have seen a fast growing body of literature concerned with the statistical modelling of animal movement. Different modelling approaches have been extensively discussed, comprising, *inter alia*, state-space models (Jonsen, Myers & James 2006; Patterson *et al.* 2008) and stochastic differential equations (Blackwell 2003; Harris & Blackwell 2013). Such approaches have mostly neglected potential interactions between different animals and, until now, most studies have assumed that the movement of one individual within a group is representative of the group's overall movement (Morales *et al.* 2010). However, animals often do not move

*Correspondence author. Roland Langrock, CREEM, The Observatory, Buchanan Gardens, University of St Andrews, St Andrews, KY16 9LZ, UK. E-mail: roland@mcs.st-and.ac.uk

[†]A previous version of this article is available at <http://arxiv.org/abs/1308.5850>.

independently of each other, and therefore, analysing the movement of individual animals without considering the dynamics of the group could be misleading. Reduced production costs and miniaturization of tracking technology mean that field researchers can monitor the movement of many more individuals than was previously possible, including simultaneously tracking several individuals within the same group. These new advances provide the opportunity for studying the inter- and intragroup dynamics within a population.

The movement of individuals in a group is typically investigated using self-propelled particle (SPP) models that capture the alignment between neighbours in self-organized swarms. For SPP models to maintain coordinated group movement, all individuals must adhere to basic mechanistic rules in which the forces of attraction (e.g. social interactions such as information sharing or vigilance) and repulsion (e.g. avoiding collisions with neighbours) are optimized within an interaction zone (Buhl *et al.* 2006; Mann 2011; Strombom 2011). In classic Lagrangian SPP models, individuals match their speed and alignment at discrete time intervals and are virtually homologous copies of one another, with the exception that some may be 'informed' while others 'naïve' (Couzin *et al.* 2005; Conradt *et al.* 2009). These models, while informed by biology, do not require any actual data input.

While in the existing literature on SPP models, the aim is to replicate movement patterns in forward simulations by defining certain rules of behaviour, we suggest a novel approach, which fundamentally differs from SPP models in that our model is fitted statistically to telemetry data. The aim here is to make inference on the influence of a group on the movement dynamics of social individuals. Our modelling approach allows individuals to switch between different behavioural states. Importantly, they can be gregarious members of a group under certain conditions, for example when they are exposed to risks or are uninformed. In turn, when conditions change and competition becomes intense, animals may break away from the group in order to capitalize on resources. We assume individuals to be periodically attracted to an abstract point, which we call the *group centroid*. This centroid could be the centre of mass of the group (a proxy for social networks, as described by Croft, James & Krause 2010) or a dominant 'informed' alpha animal that leads the group (as described by Nagy *et al.* 2012).

Here, we formulate the approach within an easily accessible framework, given by the class of hidden Markov models (HMMs) (Zucchini & MacDonald 2009). We show the feasibility of the HMM-based approach in a simulation before fitting the model to real data on the movement of 11 reindeer (*Rangifer tarandus*).

Materials and methods

GENERAL FORMULATION OF THE MODELLING APPROACH

We consider a modelling framework with a parent-child structure. First, a group-level model describes movement of some entity (the

'parent') that characterizes the group's centre of gravity and drives the movement behaviour of individuals. Depending on the system, this entity can have different interpretations: for example, it may be the group centre (represented by the mathematical centroid), or the location of the group leader. This entity is a tool that allows for modelling correlation between the individuals' movement paths in a flexible yet intuitive way. It is crucial that the entity adequately represents the point relative to which individual animals make their movement decisions. Hereafter, we will refer to this entity as the (*group*) *centroid*. Secondly, at the individual level, the animals (the 'children') make their movement decisions relative to the group centroid. Such decisions can involve attraction to the centroid, repulsion from the centroid or disregard of the centroid.

The suggested concept is immensely flexible and, in principle, can be implemented by means of different stochastic models. Here, we use a discrete-time HMM-based approach for observations that are regularly spaced in time (Patterson *et al.* 2009; Langrock *et al.* 2012; Langrock *et al.* in press). In these instances, it is convenient to model the bivariate time series of step lengths (between successive locations) and turning angles (between successive movement directions); see Morales *et al.* (2004).

THE BUILDING BLOCKS – CORRELATED AND BIASED RANDOM WALKS

Our model for group movement is composed of well-known movement models: correlated and biased random walks (CRWs and BRWs, respectively), and walks that are both correlated and biased (BCRWs) (Codling, Plank & Benhamou 2008; Langrock *et al.* 2012). CRWs involve positive (or negative) correlation in direction. In discrete time, they can be expressed by a turning angle distribution with mass centred around zero (or π). Biasedness of random walks can either refer to a general preference for some direction (e.g. East) or a bias towards a particular location. For example, in discrete time, a bias towards the location $(x^{(c)}, y^{(c)})$ is obtained by assuming that the expected movement direction at time $t + 1$ is the direction of the vector $(x^{(c)}, y^{(c)}) - (x_t, y_t)$, where (x_t, y_t) is the animal's location at time t . In our model, the location a given animal is attracted to will vary over time, denoted by $(x_t^{(c)}, y_t^{(c)})$, $t = 1, \dots, T$, so that $(x_t^{(c)}, y_t^{(c)})$ is the location of the centroid at time t . An animal may be attracted to the centroid (in which case, the expected movement direction at time $t + 1$ is the direction of the vector $(x_{t+1}^{(c)}, y_{t+1}^{(c)}) - (x_t, y_t)$), or it may be repulsed by the centroid (in which case, the expected movement direction is the direction of the vector $(x_t, y_t) - (x_{t+1}^{(c)}, y_{t+1}^{(c)})$), or it may move in disregard of the centroid (e.g. it may move according to a CRW, or according to a BRW with a fixed centre of attraction).

INDIVIDUAL-LEVEL AND CENTROID MOVEMENT MODELS

In multistate random walks – which are HMMs – the movement pattern of an animal is assumed to depend on the current underlying behavioural state of the animal (e.g. foraging, migrating and resting). The state sequence in these models is generated by a Markov chain, which leads to temporal autocorrelation in both the behavioural states and in the observed movement patterns. Notationally, we summarize the state transition probabilities of the (homogeneous) Markov chain in the $N \times N$ matrix $\Gamma = (\gamma_{ij})$, where γ_{ij} denotes the probability of the animal switching from state i (at any time t) to state j (at time $t + 1$).

In a multistate random walk, say with N states, each of the N states is associated with a distinct random walk pattern (CRW, BRW or

BCRW; cf. Langrock *et al.* 2012). In this manuscript, we focus on (individual-level) multistate random walks in which at least one of the N states involves either a BRW or a BCRW with bias relative to the centroid's location (either positive or negative). Crucially, if the movement of multiple individuals is considered, and the movement models all involve a BRW or BCRW relating to the same group centroid, then the collection of all individual-level models can capture various degrees of possible correlation of the multiple movement paths. For example, individuals may differ in their bias towards the centroid, but as long as they exhibit some tendency towards the centroid, the paths will be correlated.

In order to simulate and predict from such individual-level models, one also needs a model for the movement of the group centroid. The bivariate time series corresponding to the centroid's movement can, in principle, also be modelled using (dependent) mixtures of random walks, that is, HMMs. However, the location of the centroid often cannot be directly observed, and we address this issue below.

MODEL FITTING

The likelihood of an HMM can conveniently be calculated using an efficient recursive scheme called the *forward algorithm*, which effectively corresponds to a summation over all possible state sequences. Thus, the maximum-likelihood estimates (MLEs) of the parameters can usually easily be obtained by direct numerical likelihood maximization (e.g. using `nlm` in R). In the case where the parameters are constant over time, the likelihood is given by

$$\mathcal{L} = \delta^{(1)} \mathbf{P}(\mathbf{z}_1) \mathbf{\Gamma} \mathbf{P}(\mathbf{z}_2) \mathbf{\Gamma} \dots \mathbf{\Gamma} \mathbf{P}(\mathbf{z}_{T-1}) \mathbf{\Gamma} \mathbf{P}(\mathbf{z}_T) \mathbf{1}. \quad \text{eqn 1}$$

Here $\mathbf{P}(\mathbf{z}) = \text{diag}(f_1(\mathbf{z}), \dots, f_N(\mathbf{z}))$, with f_n denoting the state-dependent probability density function of the observations, given the animal is in state n , and $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_T$ is the bivariate series of observed step lengths and turning angles. The density f_n is determined by the type of random walk assumed in state n and the state-dependent distributions considered for step lengths and turning angles (cf. Langrock *et al.* 2012). Furthermore, $\mathbf{1}$ is a column vector of ones and $\delta^{(1)}$ denotes the initial distribution of the Markov chain.

If one uses multistate random walks to model both centroid and individual-level movement, then expression (1) applies to both the centroid and the individual-level movement models, in each case with \mathbf{P} , $\mathbf{\Gamma}$ and $\delta^{(1)}$ defined according to the model formulation. However, in both the centroid and individual-level cases, in the evaluation of (1) the locations of the group centroid at the observation times are required. The determination of these locations depends on the type of centroid considered and may be straightforward in some cases but intractable in others. For example, if the centroid corresponds to an animal leading the entire group, then it may be possible to identify and tag that animal, thus yielding the centroid's locations directly. We focus here on the case where the centroid roughly corresponds to the core of the group, that is, a virtual entity essentially corresponding to the mathematical centroid of the group. If multiple individuals within the same group are tagged, then, at any given time t , we simply compute the mean of all observed individuals' locations at that time and consider this mean as an estimate of the centroid's location. If the group of animals is such that individuals occasionally separate from the core group, then we expect that using an (outlier-)robust mean of the locations leads to a more accurate approximation of such a centroid's location. To do this, we used the minimum volume ellipsoid estimator provided in the function `cov.rob` from the R-package MASS.

EXPERIMENTS WITH SIMULATED DATA

We simulate data from the proposed group dynamic movement model, generating a movement path for the centroid comprising $T = 250$ locations, and, based on the simulated centroid locations, 20 individual movement paths, again each comprising $T = 250$ locations. The movement model for the centroid is a BCRW, with some bias towards the location $(0, 0)$, but mainly positive correlation in successive movement directions (for details see Table S1). Each individual-level movement path is generated by a two-state HMM, with state 1 a BRW (bias towards the moving centroid) and state 2 a CRW (positive correlation in successive movement directions). These states correspond to desire to stay within the group (state 1) and desire to forage independently of the group (state 2). We use gamma distributions to generate step lengths (letting μ_i and σ_i denote the mean and standard deviation, respectively, in state i), and von Mises distributions to generate turning angles (letting ν_i and κ_i denote the mean and concentration, respectively, in state i). We consider two different simulation scenarios, with the following parameter values considered for the individual-level movement models:

Scenario A:

State 1: $\mu_1 = 30$, $\sigma_1 = 15$, ν_1 determined by centroid location, $\kappa_1 = 2.5$

State 2: $\mu_2 = 50$, $\sigma_2 = 25$, $\nu_2 = 0$ (not estimated), $\kappa_2 = 5$

Scenario B:

State 1: $\mu_1 = 30$, $\sigma_1 = 15$, ν_1 determined by centroid location, $\kappa_1 = 2.5$

State 2: $\mu_2 = 30$, $\sigma_2 = 15$, $\nu_2 = 0$ (not estimated), $\kappa_2 = 5$

In both scenarios,

$$\mathbf{\Gamma} = \begin{pmatrix} 0.95 & 0.05 \\ 0.15 & 0.85 \end{pmatrix}.$$

This state process is such that individuals spend most of the time following the centroid (in state 1, occupied 75% of the time according to the stationary distribution), but occasionally split from the group and move solitarily (in state 2). Scenario A represents a setting in which the identification of the two hidden behavioural states based on the observations is expected to be accurate (since not only the turning angle distributions, but also the step length distributions differ across states), and Scenario B represents a setting in which the identification of the states is expected to be more challenging (since only the turning angle distributions differ across states, i.e. only the attraction to the centroid distinguishes state 1 from state 2 at the observation level).

By numerically maximizing the respective likelihood, we fit the model to the centroid's series of step lengths and turning angles (initially assuming that these are observed), and then simultaneously to all 20 individual-level series of step lengths and turning angles, assuming common parameters for all individuals. We also explore how well the estimation works in cases where the true location of the centroid is unknown and hence can only be approximated. To do this, we fit the centroid and the individual-level movement model using two different approximations of the centroid's locations: (i) the centroid calculated as the simple mean of all individuals' locations at each time and (ii) the centroid calculated as the robust mean of the individuals' locations at each time. This whole simulation and model fitting exercise was repeated 500 times, and performance is evaluated based on summary statistics of the parameter estimates. In each scenario, for each simulation run and three estimation methods (corresponding to different series of exact or approximate centroid locations), we use the Viterbi algorithm (see Chapter 5 in Zucchini & MacDonald 2009) to find the sequence of states s_1^*, \dots, s_T^* that under the fitted model is most likely to have given rise to the observed sequence,

$$(s_1^*, \dots, s_T^*) = \underset{(s_1, \dots, s_T) \in \{1,2\}^T}{\operatorname{argmax}} \Pr(s_1, \dots, s_T | z_1, \dots, z_T),$$

and establish the proportion of states correctly identified. The R code used in the simulation experiments is provided in the Supporting Information (Appendix S2), as is an example movie illustrating the movement pattern of the individuals and the centroid in Scenario B (Appendix S3).

APPLICATION TO REAL DATA

We fit our model to location data of reindeer. The annual migration of reindeer follows a seasonal progression of snow-melt and fresh vegetative growth that broadly describes the general movement pattern of the population (Skarin *et al.* 2008, 2010). However, herds often follow slightly different paths each year and individuals within these herds make independent decisions based on the environmental conditions they encounter *en route*. Although an individual reindeer may reduce its grazing competition by moving away from the herd, it also stands a greater chance of being killed by predators, and therefore, the choice an individual reindeer makes about how and where to move is balanced between finding enough food for itself but also by staying within the safety of the group.

We consider location data for 11 female reindeer, recorded in June and July 2003 (which is in the post-calving period) with GPS collars. For each of the 11 individuals, a maximum of 225 hourly fixes is considered (corresponding to 9.4 days of observation), with very few missing data (0.2%). From exploratory analysis, we deduce that these 11 individuals belong to the same herd. Although the total number of individuals in the herd varies over time, a core group tended to stay together for the period we considered.

Since individual reindeer occasionally separate from the group, we use a robust mean of the individual's locations observed at a given time to approximate the centroid location at that time. Modelling the movement of the (approximated) centroid requires only standard methods, but fitting ecologically complex and hence informative models at the centroid level is difficult in this case as the corresponding time series is short. Therefore, we focus exclusively on the individual-level model in this analysis. For simplicity and for the sake of parsimony in terms of the number of parameters, we assume the parameters of the individual-level model to be common to all individuals. We fit a two-state model, with state 1 involving a BRW with bias towards the (moving) centroid, and state 2 involving a CRW such that the turning angle distribution has mean zero, that is, with the reindeer tending to move straight ahead. We use gamma distributions for modelling the step lengths and von Mises distributions for modelling the turning angles.

From the suggested class of HMMs for movement at the individual level, this is the simplest example of a group dynamic movement model that still allows for the individuals to occasionally separate from the group. The purpose of the analysis is to illustrate that the suggested type of movement models can successfully be applied to real data and that this can lead to biologically interesting insights. As the model fitting is illustrative, we do not search for the optimal model from a suite of candidate models. We do however check the adequacy of the described model by computing forecast pseudoresiduals for both step lengths and turning angles, as described in Langrock *et al.* (2012), separately for each of the 11 individuals, resulting in 22 series of residuals. Such residuals are standard normally distributed if the fitted model is correct. We test for normality using Jarque-Bera tests.

COMPARISON OF GROUP DYNAMIC MODEL TO STANDARD MULTISTATE RANDOM WALK

We also explore the utility of our model in comparison with a standard multistate random walk. This is done to determine whether the additional layer of complexity regarding the centroid movement, and the response of the individuals to the centroid, leads to a more realistic description of observed movement patterns. In particular, we fit a simple 2-state HMM to the reindeer data, with each state involving a correlated random walk. Thus, in both states of the simple HMM, individuals move in complete disregard of the centroid. Parameters were again assumed to be common across individuals.

For each of the 11 individuals and each of the fitted models (standard HMM and group dynamic model), we then simulate 50 movement paths, each starting at the location where the corresponding individual was initially observed, and covering a time period of 200 days. For the group dynamic model, we simulate individual-level movement paths using an artificial series of centroid locations that start at the initial (approximate) centroid location, with subsequent movement straight northwards at the constant speed of 0.5 km h⁻¹. This is an essentially arbitrary choice, though it does illustrate the potential usefulness of the suggested approach regarding the modelling of migratory movement.

Results

SIMULATION EXPERIMENTS

All tables giving details on the simulation results are presented in Appendix S1. In the case where the exact locations of the centroid are known, there is no indication of bias of the parameter estimators. This holds true for both the centroid- and the individual-level movement models, and in both Scenarios A and B (cf. Tables S1 and S2). Fitting the suggested type of movement model thus is feasible if the centroid's locations are known.

In the case where the centroid locations can only be approximated, our fitting approach correctly identified the pattern of the individual-level movement, but for both centroid approximation methods, the parameter estimates exhibited some bias (cf. Tables S1 and S2). For the parameter estimates associated with the individual-level model, the bias was generally lower than for the parameter estimates associated with the group-level model. The bias was highest (in relative measure) for the concentration parameter of the directional distribution in the BRW component. At the individual level, the concentration parameter corresponds to the individual's bearing relative to the centroid, and this bias is not surprising, since the directional distribution relative to the centroid's location is blurred when using an approximation of the centroid's location. The estimates of the (individual-level) state transition probabilities were also found to be slightly biased, with the fitted models typically predicting more state switches than the true model that was used to simulate the data.

The results further show that when using approximate centroid's locations, the estimation seems to perform better in cases where the state-dependent distributions are fairly distinct (Scenario A) than in cases where they are not that easily distinguishable (Scenario B). In particular, the accuracy of the

Viterbi-based state decoding was higher in Scenario A (96.1% of the states correctly decoded, vs. 95.2% in Scenario B, for the case where the robust mean was applied). Use of the robust mean led to higher accuracy of the state decoding than use of the simple mean (96.1% vs. 94.7% in Scenario A, and 95.2% vs. 93.0% in Scenario B).

REINDEER DATA

Parameter estimates for the group dynamic model fitted to the reindeer data, together with 95% CIs obtained based on the Hessian of the log-likelihood, are listed in Table 1. The fitted state-dependent distributions for step lengths and directions are displayed in Fig. 1. In 22 Jarque–Bera tests for normality of the pseudoresiduals, the null hypothesis of normality is twice rejected at the 5% level, which is consistent with the model fitting well.

Table 1. Parameter estimates for the group dynamic individual-level model, fitted to the reindeer data, and 95% confidence intervals of the estimates (obtained based on the Hessian of the log-likelihood)

Parameter	Description	CI lower bound	Estimate	CI upper bound
State 1				
μ_1	gamma mean	399.5	432.9	469.1
σ_1	gamma standard deviation	524.8	571.2	621.8
κ_1	von Mises concentration	0.183	0.246	0.331
γ_{11}	persistence in state 1	0.830	0.866	0.895
State 2				
μ_2	gamma mean	805.0	896.6	998.6
σ_2	gamma standard deviation	743.8	829.7	925.5
κ_2	von Mises concentration	2.759	3.517	4.482
γ_{22}	persistence in state 2	0.510	0.584	0.655

In state 1, individuals on average make relatively short steps (gamma mean = 432.9) and are weakly attracted to the group's centroid (von Mises concentration = 0.246), while in state 2, individuals on average make longer steps (gamma mean = 896.6) and perform a CRW with high directional persistence (von Mises concentration = 3.517). Following Morales *et al.* (2004), we interpret state 1 as the behavioural state in which animals are 'encamped', here associated with a weak attraction to the group centroid, and state 2 as the behavioural state in which animals are 'exploring'. Such interpretations should not be taken literally, as they merely provide a rough classification of different movement patterns (Langrock *et al.* 2012). In particular, encamped does not necessarily mean with the group, rather it means the animal is making shorter steps and those steps are slightly biased towards the group centroid (Fig. 1).

According to the stationary distribution of the fitted Markov chain, approximately three quarters of the time is spent in state 1. Figure 2 displays the Viterbi-decoded state sequences, and Fig. 3 displays the estimated probabilities of being in the exploratory state (and thus moving in disregard of the group centroid), mapped onto the movement paths, for the 11 reindeer. Broadly speaking, the individual reindeer spend long stretches of time encamped (Fig. 2), and these periods are interspersed with brief exploratory forays (shown in yellow in Fig. 3).

GROUP DYNAMIC MODEL VERSUS STANDARD MULTISTATE RANDOM WALK

The centroid-driven individual-level model yields a much lower value for the Akaike Information Criterion (AIC) than the simple two-state random walk ($\Delta\text{AIC} = 43.6$). However, the AIC has to be interpreted very cautiously here: the additional centroid layer appears only in the former model, and the stated ΔAIC is based on regarding the centroid's positions as deterministic covariates within the individual-level model. We also consider the difference in predictive capacity between the two

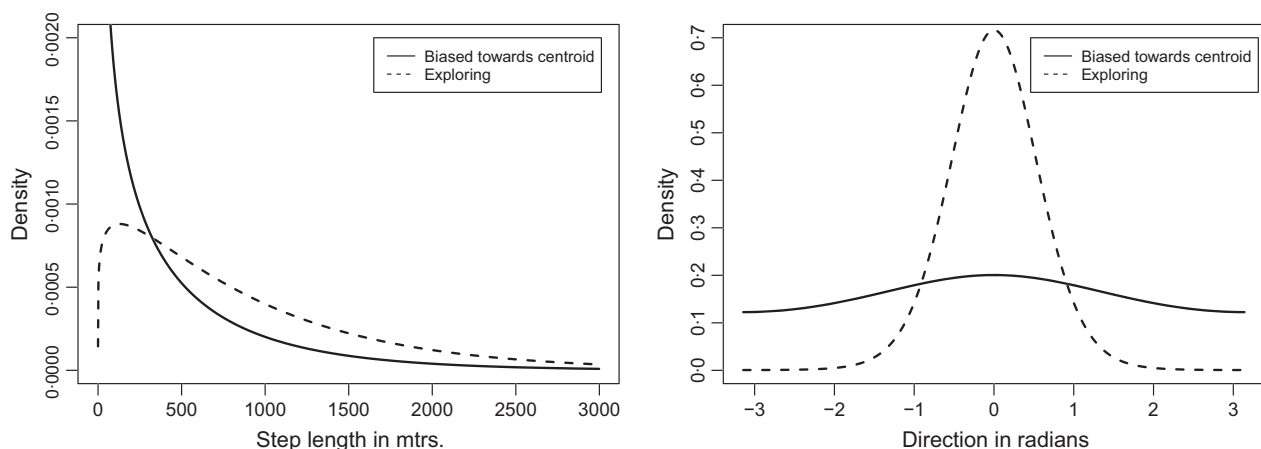


Fig. 1. Reindeer example, group dynamic model – fitted state-dependent step length and directional distributions. For the state involving a bias towards the centroid (state 1), radian 0 corresponds to the direction the centroid is in, relative to the individual. For the 'exploring' state (state 2), radian 0 corresponds to the previous movement direction of the individual, that is, while exploring the individuals tend to move straight ahead.

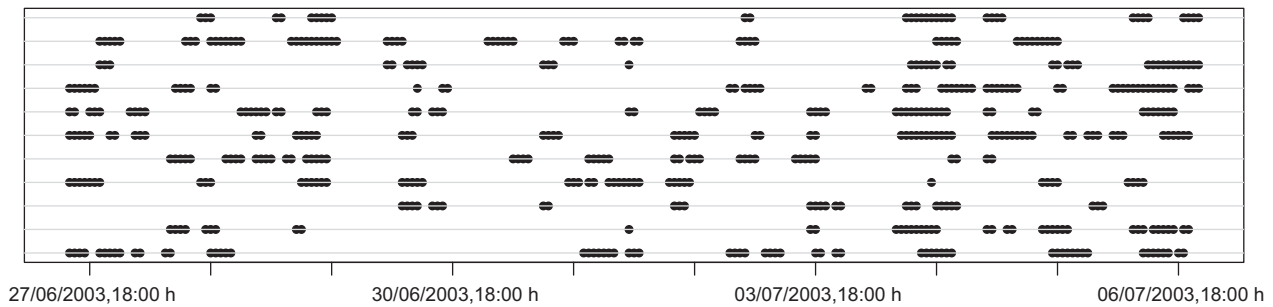


Fig. 2. Reindeer example – Viterbi-decoded state sequences for the 11 different individuals. Each grey horizontal line corresponds to one individual, with black dots indicating when the corresponding individual was estimated to be in state 2 of the fitted model.

models using simulation (Fig. 4). Although this is a simplified example, it highlights the effect of the group attraction mechanism, which ensures that individuals stick together and follow the group (Fig. 4, top row) rather than wander independently and randomly about (Fig. 4, bottom row), as is the case for the simple two-state CRW.

Discussion

GENERAL REMARKS ABOUT THE APPROACH

Through simulation and by fitting our model to real data, we have documented that reliable inference can be made on the influence of a group on the movement dynamics of social individuals. Our approach represents a novel step towards an inferential framework for understanding how social factors influence the movement decisions made within a group; this adds to the growing body of the literature (Polansky *et al.* 2010; Couzin *et al.* 2011; Dowd & Joy 2011; Polansky & Wittemyer 2011). The key advancement in this approach that distinguishes it from previous work is its inferential nature. The likelihood-based framework lends itself to including environmental or physiological parameters as potential explanatory covariates in the fission–fusion dynamics of communally living organisms, which makes itself immediately applicable to ecological research. In contrast to the common approaches in the SPP literature on movement of social groups (Couzin *et al.* 2005), here, we fit our model to actual movement data. In doing so, we are able to statistically compare the fit of competing models (model selection) and to assess the goodness-of-fit of our model (model checking).

The key component of our framework is the modelling of the individual's attraction to the centroid. Specifically, this is done by allowing individuals to switch between a 'group' state, where individuals are biased towards the group centre, and one or more 'exploratory' states where individual movement is independent of the group. In contrast to multistate CRWs, the fitted group dynamic movement model is non-diffusive. More precisely, even when the attraction to the centroid is weak, individuals will, in the long run, not move arbitrarily far away from either the centroid or from each other (cf. Fig. 4).

One pre-existing approach that is similar in spirit to ours is that of Dunn & Gipson (1977), who formulate all of their models in terms of multiple interacting animals. They think of

the position of m animals in d dimensions (taking $d = 2$ in practice, as we do here) as a single point in $k = m \times d$ dimensions, and model the movement of that point using a multivariate Ornstein–Uhlenbeck process, parameterized in terms of a k -vector and two $k \times k$ matrices. By contrast, our approach can be thought of as modelling the position of the group in terms of a point and having individuals interact directly with the centroid. Their models have very large numbers of parameters already for moderate values of m , and in practice, our representation is far more parsimonious and interpretable, certainly for $m > 2$.

Hierarchical models are an alternative way to capture similarities between individuals, by modelling the model parameters as stochastic variables with a common group or population mean. Thus, individuals' parameters are correlated, but conditional on those parameters, there is no interaction between individuals. We assumed here that all individuals shared movement parameters and focused on the interaction of their movement paths via the group centroid. In contrast to the traditional hierarchical approach, this models a fundamental mechanism of the herding behaviour and therefore has a more intuitive biological interpretation.

Using multiple yet relatively short real data time series, we were able to reveal an attraction to a centroid. However, much more analyses, including simulation experiments, need to be conducted in order to make general statements about how many individuals need to be tracked, or for how long, to make reliable inference on attraction to a centroid. It is also unclear as to how influential the proportion of tracked animals to untracked animals needs to be in terms of reliably estimating the centroid position, but it is intuitive that the model will work best for relatively cohesive groups, where the choice of particular individuals does not bias the analysis.

Discussion of results

We initially generated and analysed simulated data to assess the performance of our modelling approach. When the true location of the group centroid was known, the approach provided unbiased parameter estimates of both the centroid movement model and the individual-level movement model. In cases where the movement of the group is directed by an alpha individual, such as an elephant matriarch, the true centroid location is known when this animal is identified and tagged.

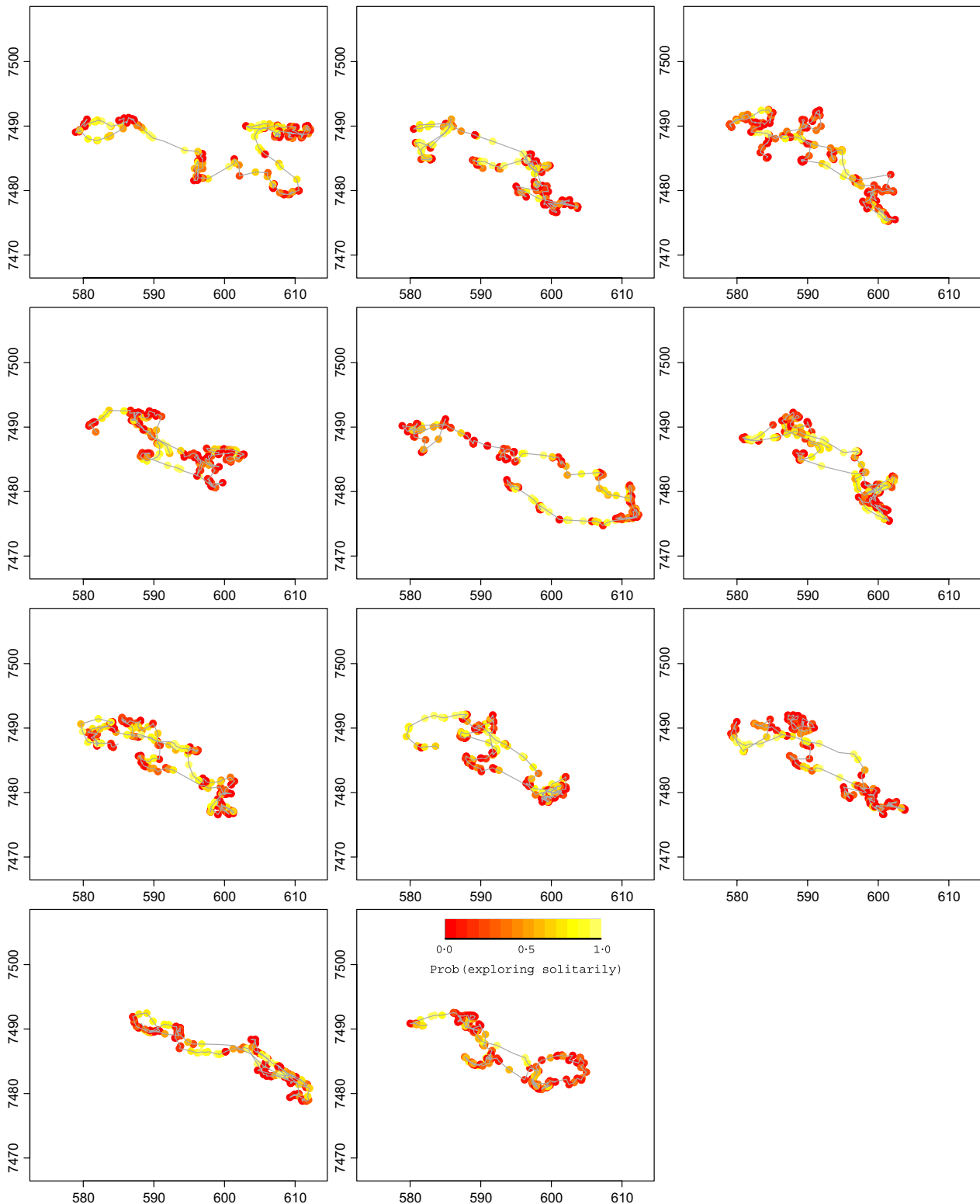


Fig. 3. Reindeer example – state categorizations mapped onto the 11 movement paths. The colour of the points mapped on the locations indicate the probability of the corresponding animal being in state 2 (exploring), that is, red indicates higher probability of being encamped and attracted to the group centroid. The horizontal and vertical axes give x and y , respectively, in kilometres using the Sweref99 projection from the National Land Survey of Sweden.

For scenarios in which the centroid location is unknown, we considered simple or robust means of the individuals' locations as approximations to the centroid's locations. In our simula-

tions, the estimation of the individual-level movement models worked reasonably well when the centroid's locations were approximated using the simple mean, and slightly better when

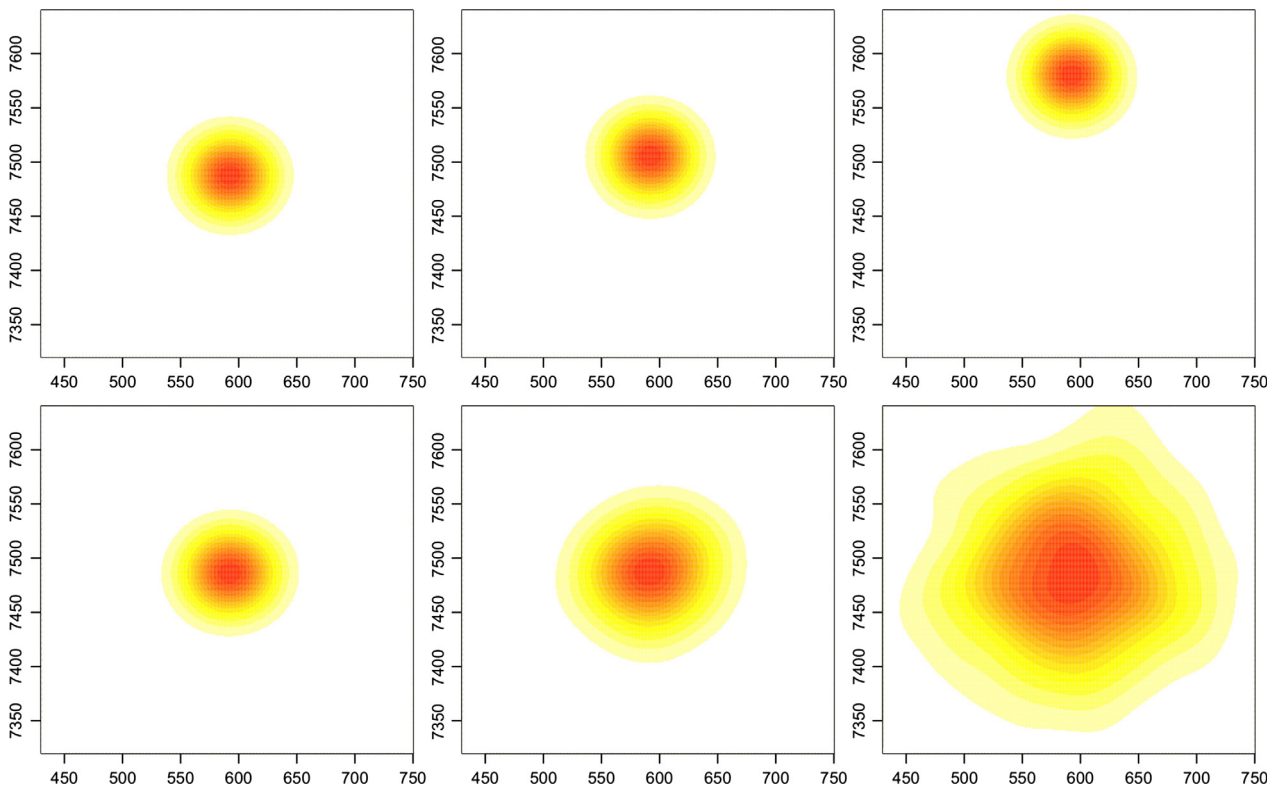


Fig. 4. Reindeer example – spatial densities of simulated locations of individuals after 9–4 days (left column), 50 days (middle column) and 200 days (right column), for the group dynamic model (top row; with artificial centroid movement straight northwards) and for the standard two-component mixture of correlated random walks (bottom row). The smoothed densities were obtained by applying a bivariate kernel density estimator to the generated locations. Within a short time span (left column), the two models provide similar predictions both in terms of mean and variance of the densities. For simulation over longer time periods (middle and right column), the attraction mechanism in the group dynamic model (top row) both bounds the extreme movements of the individuals (variance) and ensures that individuals follow the group centroid (mean), while the two-state model without attraction (bottom row) is purely diffusive.

using the robust mean. In addition, even in a scenario where the step length distribution was common across states, we were able to use the centroid attraction information to estimate behavioural states with high accuracy. Our approach thus represents a useful tool for drawing statistical and ecological inference on group movement dynamics, at least in particular situations.

We illustrated the modelling approach using movement data from 11 reindeer in Sweden (Skarin *et al.* 2008, 2010). We were able to infer, to some extent, when individuals are tracking group movement, and when they appear to be following their own movement drivers (cf. Fig. 1). The models we fitted to the reindeer data led to the classification that is characteristic of two-state random walks, with an encamped and an exploratory state and corresponding state-dependent step lengths and turning angle distributions (Morales *et al.* 2004; Langrock *et al.* 2012). In the models fitted here, the encamped state includes an additional biological feature of attraction to the group centroid. According to the fitted model and based on Viterbi-decoding of the underlying state sequences, some individuals spent periods of up to 3 days in the nominal encamped state during the time period covered by the observations.

The conducted analyses indicate the ecological inference that can be gleaned from our approach. First, we have quantified how the individual reindeer respond to the group.

This highlights how conspecific attraction can be included in movement models. Secondly, we have shown how behavioural states can be estimated with high accuracy, even in cases where important movement phenomena, that is, step lengths, are similar across states. Finally, we have outlined advantages of including the centroid information over simple multi-state random walks. Specifically, including the centroid layer in the model has led to a more accurate depiction of spatial density, highlighting the importance of incorporating social drivers.

EXTENSIONS

Several possible refinements and extension of our approach come to mind. First, in the simulation experiments, we found estimators to be biased when we directly approximated the centroid's locations using the individual's positions. As an alternative, the location of the centroid could be estimated within a state-space modelling framework, for example using a Bayesian data augmentation approach for the centroid location. Such an approach is appealing since the uncertainty about the centroid's location would be propagated through to the individual-level parameter estimates.

A simplifying assumption of our method was that all individuals shared movement parameters. A more realistic model

could include random effects for certain parameters such as the switching probabilities, thereby allowing for variations in behaviour between individuals (Ford, Bravington & Robbins 2012; Schliehe-Diecks, Kappeler & Langrock 2012), for example that some individuals are more adventurous than others. Random effects, however, describe individual differences as stochasticity (or noise) and hence do not explain the source of the variation. In the frequentist HMM framework, they also typically carry a high computational cost (Schliehe-Diecks, Kappeler & Langrock 2012). Instead, formulating individual-level parameters as functions of auxiliary information, such as gender, weight or age, would enable testing of hypotheses related to individual covariates and aid in uncovering reasons for differences between individuals within the group.

More generally, the model can be extended to include internal and external dynamic covariates. For example, non-homogeneous Markov chains could be considered for the behavioural state dynamics, with transition probabilities structured such that observations of the ambient environment or body condition mediate switches between movement phases. Such an approach may enable separation of the relative roles of group membership, individual body condition and habitat, and this represents an exciting avenue of analysis for the movement patterns of social animals. An important special case of such non-homogeneous Markov chains results from including time as a covariate, which can be useful in order to accommodate seasonality or the daily cycle in the model. Another potentially interesting covariate to consider is the individual-specific separation distance from the centroid, which could be informative about group membership. From the statistical and computational point of view, these extensions are fairly straightforward, since the likelihood structure of the HMM remains unaffected. However, it will often be the case that numerous possible model formulations seem plausible *a priori*, in which, case model selection can become a daunting task.

In this work, we chose to implement the group centroid concept in the (discrete-time) HMM framework. It is however also possible to extend the continuous-time models of Dunn & Gipson (1977) to incorporate some of the advantages of the current approach. Explicitly representing the centroid as part of the process means that the movement of the whole system is then a diffusion process in $(m + 1) \times d$ dimensions, but the modelling of that process is greatly simplified if the m individual animals are each interacting with the centroid, as here, rather than interacting directly with each other. One possibility currently being pursued is for each animal to follow a multivariate Ornstein–Uhlenbeck (MOU) process attracted to the centroid, while the centroid follows its own movement model. Under simplifying assumptions, the whole system can then be modelled as a partially observed $(m + 1) \times d$ -dimensional MOU process.

FINAL REMARKS

Our work was motivated by the necessity for an inferential statistical framework to investigate the effects of group

aggregations on individual choice patterns. The modelling approach could for example be used to investigate environmental cues for juvenile dispersal from natal colonies, the congregation of animals for security reasons, or the convergence of animals tracking an ephemeral yet moving resource (e.g. wildebeest migration in relation to fresh grazing). As it stands, the approach is suited to relatively simple situations where multiple individuals from a single group are tagged and environmental parameters are easily determined. As presently constructed, our model cannot address scenarios with multiple competing groups. Though it is conceptually straightforward in a Bayesian setting to impute multiple centroids, fitting such a model to actual data would be quite difficult. Future research should explore the flexibility of the considered approach with regard to such more complex fission–fusion dynamics. In general, the direction of a group's movement might correspond to a consensus by all members in response to the spatial pattern of resources (i.e. a common motivator detected by all members of the group), or it might result from some social structure within the group (i.e. informed versus naïve individuals). Future research should test our concept of a group centroid on real data to see whether meaningful biological responses can be inferred in different such scenarios.

Acknowledgements

Funding for RL was provided by the Engineering and Physical Sciences Research Council (EPSRC reference EP/F069766/1). Funding for RSS was provided by US Office of Naval Research Grant N00014-12-1-0286 to University of St Andrews. Funding for PGB and MN was provided by EPSRC/NERC Grant EP/1000917/1. Funding for JGCH was provided by the Lord Kelvin Adam Smith Fellowship. The ideas presented in this manuscript were developed during a workshop on animal movement modelling, held at the University of St Andrews in June 2012, which was funded via an EPSRC Strategic Partnership Fund. The authors would like to thank Tiago Marques, Bram Van Moorter, an anonymous referee and an associate editor for very helpful comments on previous versions of the manuscript.

References

- Blackwell, P.G. (2003) Bayesian inference for Markov processes with diffusion and discrete components. *Biometrika*, **90**, 613–627.
- Bowler, D.E. & Benton, T.G. (2005) Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. *Biological Reviews*, **80**, 205–225.
- Buhl, J., Sumpter, D.J.T., Couzin, I.D., Hale, J.J., Despland, E., Miller, E.R. & Simpson, S.J. (2006) From disorder to order in marching locusts. *Science*, **312**, 1402–1406.
- Codling, E.A., Plank, M.J. & Benhamou, S. (2008) Random walk models in biology. *Journal of the Royal Society Interface*, **5**, 813–834.
- Conradt, L., Krause, J., Couzin, I.D. & Roper, T.J. (2009) “Leading according to need” in self-organizing groups. *American Naturalist*, **173**, 304–312.
- Couzin, I.D., Krause, J., Franks, N.R. & Levin, S.A. (2005) Effective leadership and decision-making in animal groups on the move. *Nature*, **433**, 513–516.
- Couzin, I.D., Ioannou, C.C., Demirel, G., Gross, T., Torney, C.J., Hartnett, A., Conradt, L., Levin, S.A. & Leonard, N.E. (2011) Uninformed individuals promote democratic consensus in animal groups. *Science*, **334**, 1578–1580.
- Croft, D.P., James, R. & Krause, J. (2010) *Exploring Animal Social Networks*. Princeton University Press, Princeton, NJ, USA.
- Dowd, M. & Joy, R. (2011) Estimating behavioral parameters in animal movement models using a state-augmented particle filter. *Ecology*, **92**, 568–575.
- Dunn, J.E. & Gipson, P.S. (1977) Analysis of radio telemetry data in studies of home range. *Biometrics*, **33**, 85–101.
- Ford, J.H., Bravington, M.V. & Robbins, J. (2012) Incorporating individual variability into mark-recapture models. *Methods in Ecology and Evolution*, **3**, 1047–1054.

- Harris, K.J. & 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. & 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.
- Jonsen, I.D., Myers, R.A. & James, M.C. (2006) Robust hierarchical state-space models reveal diel variation in travel rates of migrating leatherback turtles. *Journal of Animal Ecology*, **75**, 1046–1057.
- Langrock, R., Marques, T.A., Baird, R.W. & Thomas, L. (in press) Modeling the diving behavior of whales: a latent-variable approach with feedback and semi-Markovian components. *Journal of Agricultural, Biological and Environmental Statistics*. DOI: 10.1007/s13253-013-0158-6.
- Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D. & Morales, J.M. (2012) Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*, **93**, 2336–2342.
- Mann, R.P. (2011) Bayesian inference for identifying interaction rules in moving animal groups. *PLoS ONE*, **6**, e22827.
- Morales, J.M., Haydon, D.T., Frair, J.L., Holsinger, K.E. & Fryxell, J.M. (2004) Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, **85**, 2436–2445.
- Morales, J.M., Moorcroft, P.R., Matthiopoulos, J., Frair, J.L., Kie, J.G., Powell, R.A., Merrill, E.H. & Haydon, D.T. (2010) Building the bridge between animal movement and population dynamics. *Philosophical Transactions of the Royal Society B*, **365**, 2289–2301.
- Nagy, M., Akos, Z., Biro, D. & Vicsek, T. (2012) Hierarchical group dynamics in pigeon flocks. *Nature*, **464**, 890–899.
- Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. & Smouse, P.E. (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. & Matthiopoulos, J. (2008) State-space models of individual animal movement. *Trends in Ecology and Evolution*, **23**, 87–94.
- Patterson, T.A., Basson, M., Bravington, M.V. & Gunn, J.S. (2009) Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *Journal of Animal Ecology*, **78**, 1113–1123.
- Polansky, L. & 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.
- Polansky, L., Wittemyer, G., Cross, P.C., Tambling, C.J. & Getz, W.M. (2010) From moonlight to movement and synchronized randomness: Fourier and wavelet analyses of animal location time series data. *Ecology*, **91**, 1506–1518.
- Schick, R.S., Loarie, S.R., Colchero, F., Best, B.D., Boustany, A., Conde, D.A. *et al.* (2008) Understanding movement data and movement processes: current and emerging directions. *Ecology Letters*, **11**, 1338–1350.
- Schliehe-Diecks, S., Kappeler, P.M. & Langrock, R. (2012) On the application of mixed hidden Markov models to multiple behavioural time series. *Interface Focus*, **2**, 180–189.
- Skarin, A., Danell, Ö., Bergström, R. & Moen, J. (2008) Summer habitat preferences of GPS-collared reindeer *Rangifer tarandus tarandus*. *Wildlife Biology*, **14**, 1–15.
- Skarin, A., Danell, Ö., Bergström, R. & Moen, J. (2010) Reindeer movement patterns in alpine summer ranges. *Polar Biology*, **33**, 1263–1275.
- Strombom, D. (2011) Collective motion from local attraction. *Journal of Theoretical Biology*, **283**, 145–151.
- Zucchini, W. & MacDonald, I.L. (2009) *Hidden Markov Models for Time Series: An Introduction Using R*. Chapman & Hall, London, UK.

Received 1 March 2013; accepted 11 December 2013

Handling Editor: Nigel Yoccoz

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Simulation results.

Appendix S2. R code used to conduct the simulation experiments.

Appendix S3. Animation of the movement paths simulated using the code given in Appendix S2 (black dot: true centroid locations; red dot: approximated centroid locations; grey dots: individuals' locations).