PHILOSOPHICAL TRANSACTIONS B

rstb.royalsocietypublishing.org

Research

Cite this article: Calabrese JM, Fleming CH, Fagan WF, Rimmler M, Kaczensky P, Bewick S, Leimgruber P, Mueller T. 2018 Disentangling social interactions and environmental drivers in multi-individual wildlife tracking data. *Phil. Trans. R. Soc. B* **373**: 20170007. http://dx.doi.org/10.1098/rstb.2017.0007

Accepted: 31 October 2017

One contribution of 16 to a theme issue 'Collective movement ecology'.

Subject Areas:
behaviour, ecology

Keywords:
caribou, correlated diffusion, khulan, movement correlation indices, shared drift, wildlife tracking data

Author for correspondence:
Justin M. Calabrese
e-mail: calabresej@si.edu

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3999492.

THE ROYAL SOCIETY PUBLISHING

Disentangling social interactions and environmental drivers in multi-individual wildlife tracking data

Justin M. Calabrese[1], Christen H. Fleming[1,2], William F. Fagan[2], Martin Rimmler[3], Petra Kaczensky[4], Sharon Bewick[2], Peter Leimgruber[1] and Thomas Mueller[5,6]

[1]Conservation Ecology Center, Smithsonian Conservation Biology Institute, Front Royal, VA, USA
[2]Department of Biology, University of Maryland, College Park, MD, USA
[3]Department of Biology, University of Stuttgart, Stuttgart, Germany
[4]Research Institute of Wildlife Ecology, Vienna, Austria
[5]Senckenberg Biodiversity and Climate Research Centre, Frankfurt, Germany
[6]Department of Biological Sciences, University Frankfurt, Frankfurt, Germany

JMC, 0000-0003-0575-6408

While many animal species exhibit strong conspecific interactions, movement analyses of wildlife tracking datasets still largely focus on single individuals. Multi-individual wildlife tracking studies provide new opportunities to explore how individuals move relative to one another, but such datasets are frequently too sparse for the detailed, acceleration-based analytical methods typically employed in collective motion studies. Here, we address the methodological gap between wildlife tracking data and collective motion by developing a general method for quantifying movement correlation from sparsely sampled data. Unlike most existing techniques for studying the non-independence of individual movements with wildlife tracking data, our approach is derived from an analytically tractable stochastic model of correlated movement. Our approach partitions correlation into a deterministic tendency to move in the same direction termed 'drift correlation' and a stochastic component called 'diffusive correlation'. These components suggest the mechanisms that coordinate movements, with drift correlation indicating external influences, and diffusive correlation pointing to social interactions. We use two case studies to highlight the ability of our approach both to quantify correlated movements in tracking data and to suggest the mechanisms that generate the correlation. First, we use an abrupt change in movement correlation to pinpoint the onset of spring migration in barren-ground caribou. Second, we show how spatial proximity mediates intermittently correlated movements among khulans in the Gobi desert. We conclude by discussing the linkages of our approach to the theory of collective motion.

This article is part of the theme issue 'Collective movement ecology'.

1. Introduction

Understanding the causes and consequences of animal movement remains one of ecology's greatest challenges [1]. Movement ecology has rapidly progressed in recent years, fuelled by advances in animal tracking technology coupled with increasingly sophisticated movement models [2,3]. Many studies now track multiple animals simultaneously, and these datasets open new possibilities for going beyond traditional individual-level analyses to explore how animals move relative to one another. Identifying when, where and to what extent animals exhibit correlated movements is important because correlation affects the way individual movements scale up to population-level distribution patterns. Consequently, ecologists increasingly recognize the need to understand non-independent movements among individuals [4–12]. Despite this growing

awareness, general methods for rigorously quantifying movement correlation from wildlife tracking data are lacking.

Correlated movements among individuals may arise from diverse sources, but two broad classes of non-independence can be identified [4,7,13]. First, individuals may respond to the same external influences, such as an environmental gradient or current, or may share a navigation goal. Second, individuals can respond directly to each other's movements, and for social species, may directly coordinate their movements. Ecologically, both forms of non-independence are important because they both affect the way that animals use space. Additionally, both kinds of non-independence may act in concert, particularly for species that live in fission–fusion groups. For example, a shared navigation goal or gradient following behaviour at key times of the year (e.g. breeding season) might bring individuals who had been moving independently into close enough proximity to begin directly coordinating their movements. Teasing apart these different sources of correlation and identifying the conditions under which they occur can elucidate the factors governing movement decisions.

Quantifying correlated movements in wildlife tracking studies is conceptually related to studies of collective animal movement, which focus on the fine-scale mechanisms allowing individuals in flocks, schools or swarms to coordinate their movements [5,7]. However, empirical collective movement studies typically rely on extremely high density datasets (e.g. multiple location fixes per second) usually obtained in controlled condition, short duration, replicated experiments (e.g. [5,7]). By contrast, technical constraints in tracking wildlife across long time frames and broad spatial scales frequently result in datasets that are coarse, include relatively few individuals, and are not replicated. Analytical methods designed for collective movement data are therefore often not applicable to wildlife tracking studies.

While more finely sampled wildlife tracking studies are beginning to appear in the literature (e.g. [14,15]), there are, and will remain, many multi-individual datasets that are too sparse for typical collective motion approaches. Even if tracking devices are capable of collecting very-high-frequency data, study goals may dictate a sparser collection strategy. For example, if the primary goal of a study is quantifying home-range areas, then a sparser, longer-term data collection schedule would be more appropriate than collecting a short burst of high-frequency data [16]. Methods that allow information on movement correlation to be extracted from such datasets can thus be viewed as complementary to those that work on datasets purpose-built for studying collective motion.

Metrics designed to estimate some aspect of non-independence [6,8,10,11] have been the main approaches to quantifying correlated movements in wildlife tracking studies, but none of these techniques are derived directly from a stochastic model of correlated movements. This disconnect between metrics and underlying movement models can make these approaches difficult to interpret. Recently, approaches derived from stochastic process models of non-independent movement have been developed [4,9,12,13]. However, these methods are either limited to pairwise analyses [4,13], or are restricted to situations where all tracked individuals belong to a group with a well-defined centroid [9,12].

Here, we develop a general framework for quantifying movement correlation from multi-individual tracking data. Unlike most existing approaches, we derive our movement correlation indices (MCIs) directly from a stochastic process model of correlated movements among individuals. Specifically, we model movement as a family of Brownian motion processes, potentially with shared drift and correlated diffusion among individuals. We then derive closed-form, minimum-variance unbiased (MVU) estimators and associated confidence intervals for the parameters of this family of models. We also derive exact, small sample size-corrected Akaike information criterion (AIC_C) formulae for our models to facilitate model selection. We then develop a partitioning algorithm, similar in principle to behavioural change point analysis (BCPA) [17], that facilitates quantification of non-stationarity in both the nature and degree of movement correlation among individuals over time. Finally, we provide an easy-to-use R package, `corrMove` (see the electronic supplementary material), that implements these calculations and facilitates visualization of results.

Our MCIs decompose total correlation into a shared tendency to move in a particular direction, termed 'drift correlation', and a shared tendency to move in random directions, termed 'diffusive correlation'. By partitioning correlated movements into drift and diffusive components, our approach suggests where and when different classes of mechanisms may underpin observed non-independence. We study two empirical examples to illustrate the utility of our MCIs: barren-ground caribou (*Rangifer tarandus granti*) in Alaska and Canada initiating their spring migration, and khulans (*Equus hemionus*) in the Gobi desert coming together in a large herd after an unusually severe winter displaced them from their normal range. In both cases, our approach yields novel insights into correlated movement.

2. Methods

(a) Movement modelling framework

We consider a family of movement models based on a continuous-time Brownian motion process that may contain correlated drift and/or correlated diffusion among N simultaneously tracked individuals that may or may not belong to the same species (e.g. [18]). The Langevin equation for our multivariate Brownian process is given by

$$\dot{\mathbf{r}}_n(t) = \underbrace{\boldsymbol{\xi}_n(t)}_{\text{stochastic}} + \underbrace{\boldsymbol{\nu}_n}_{\text{deterministic}}, \tag{2.1}$$

where $\mathbf{r}_n(t) = (x_n(t), y_n(t))$ is the location vector of the nth of N total walkers, $\dot{x} \equiv \mathrm{d}x/\mathrm{d}t$ denotes differentiation with respect to time, $\boldsymbol{\nu}_n$ is the drift vector and (t) is a white noise process. Solutions to this equation are straightforwardly given by integration to be

$$\Delta\mathbf{r}_n(t_i) = \mathbf{w}_n(t_i) + \boldsymbol{\nu}_n \Delta t_i, \tag{2.2}$$

where $\Delta\mathbf{r}(t)$ denotes the joint step vector, $\mathbf{w}(t)$ is a Brownian process initially at the origin and Δt denotes the time step:

$$\Delta\mathbf{r}(t_i) \equiv \mathbf{r}(t_i) - \mathbf{r}(t_{i-1}),\ \mathbf{w}(t_i) \equiv \int_{t_{i-1}}^{t_i} \mathrm{d}t\ \boldsymbol{\xi}(t),\ \Delta t_i \equiv t_i - t_{i-1}. \tag{2.3}$$

The locations $\mathbf{r}(t)$ constitute a Markov process, and the steps $\Delta\mathbf{r}(t)$ constitute an independent and identically distributed (IID) process with first (mean) and second (covariance) cumulants

$$\langle \Delta\mathbf{r}_n(t) \rangle = \boldsymbol{\nu}_n \Delta t, \quad \mathrm{COV}[\Delta\mathbf{r}_n(t), \Delta\mathbf{r}_m(t)] = D_{nm} \Delta t \mathbf{I}_2, \tag{2.4}$$

where $\mathbf{D}$ is the diffusion matrix (see below), and for simplicity, we take the spatial correlations to be the same in all directions

with the 2×2 identity matrix $\mathbf{I}_2$. The average amount of movement per individual (and second moment of the step process) is given by

$$\frac{1}{N}\sum_{i=1}^{N}\langle \Delta \mathbf{r}_n(t)^2\rangle = \underbrace{2\frac{\Delta t}{N}\mathrm{tr}[\mathbf{D}]}_{\text{stochastic}} + \underbrace{\frac{\Delta t^2}{N}\sum_{i=1}^{N}\boldsymbol{\nu}_n^2}_{\text{deterministic}}. \tag{2.5}$$

The average amount of movement per individual, per unit time is then given by

$$M = \frac{2}{N}\mathrm{tr}[\mathbf{D}] + \frac{\Delta t}{N}\sum_{i=1}^{N}\boldsymbol{\nu}_n^2. \tag{2.6}$$

(b) Movement correlation indices

We now wish to address, on average, how much of the drift and diffusion is correlated. For the drift, we consider the average correlation in drift vectors

$$\omega = \frac{1}{N(N-1)}\sum_{n\neq m}^{N}\boldsymbol{\nu}_n\cdot\boldsymbol{\nu}_m, \quad |\omega| \leq \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\nu}_n^2. \tag{2.7}$$

Similarly, for the diffusion we consider the average correlation

$$\rho = \frac{1}{N(N-1)}\sum_{n\neq m} D_{nm}, \quad |\rho| \leq \frac{1}{N}\mathrm{tr}[\mathbf{D}]. \tag{2.8}$$

In both cases, the correlations are zero if the movement is uncorrelated and equal to the total amount of movement if the movement is perfectly correlated. Finally, we construct our movement correlation indices (MCIs) accordingly as

$$\eta_{\mathrm{dif}} \equiv \frac{2\rho}{M}, \quad \eta_{\mathrm{dft}} \equiv \frac{\Delta t\omega}{M} \quad \text{and} \quad \eta_{\mathrm{tot}} \equiv \frac{2\rho + \Delta t\omega}{M}. \tag{2.9}$$

Note that the MCIs are necessarily scale dependent, as the amount of movement attributable to diffusion scales linearly with time lag, while the amount of movement attributable to drift scales quadratically with time lag. This is a limitation of coarsely sampled data, and with finer data the instantaneous diffusion will scale quadratically with time lag, as in [19–21].

The basic form of MCI we propose is only scale independent in the limit of $\Delta t \to 0$, with the interpretation of comparing infinitesimal displacements rather than gross displacements, and thus being more closely related to distance travelled and speed. To minimize scale dependence, we therefore base our MCIs on the smallest regular time interval Δt in the data where the Brownian motion model is still valid [22].

(c) Specific movement models

The underlying Brownian motion with drift model is parameterized by the uniform drift vector, $\boldsymbol{\nu} = \{\nu_x, \nu_y\}$, the individual diffusion rate, σ, and the cross-diffusion rate, ρ. Our MCIs then summarize diffusion by the average diagonal element ($\propto$ variance) $\sigma \equiv \mathrm{tr}[\mathbf{D}]/N$ and average off-diagonal element ($\propto$ covariance) ρ, while summarizing drift by the mean square velocity $\boldsymbol{\nu}^2/N$ and average velocity correlation ω. To model correlated random movement (diffusion), we consider the family of models given by the matrix

$$D_{nm} = \begin{cases} \sigma, & n = m \\ \rho, & n \neq m, \end{cases} \tag{2.10}$$

where the parameter ρ scales η_{dif}. We also consider the special case of no correlated diffusion ($\rho = 0$). To model correlated deterministic movement, we consider the model of uniform correlated drift $\boldsymbol{\nu}_n = \boldsymbol{\nu}$, where $\omega = \boldsymbol{\nu}^2$ scales η_{dft}. We also consider the special case of no drift ($\boldsymbol{\nu} = \mathbf{0}$). This results in a family of four models: (1) uncorrelated drift and uncorrelated diffusion (UU), (2) correlated drift and uncorrelated diffusion (CU), (3) uncorrelated drift and correlated diffusion (UC) and (4) correlated drift and correlated diffusion (CC). Note that in these model labels, the symbols C and U refer to correlated and uncorrelated, respectively, while the first position of each label is for the drift, and the second position is for the diffusion.

For each model, we derive ideal minimum-variance unbiased (MVU) parameter estimators (electronic supplementary material, appendices S1 and S2), which are necessary because the standard maximum likelihood estimators (MLEs) for variance and covariance parameters in our models can be substantially biased for small numbers of individuals N and/or a small number of time steps S. Standard model selection via small sample size-corrected (Akaike information criterion ($\mathrm{AIC_C}$)) is based on the maximized likelihood function [23]. However, when MVU estimators differ from the standard MLEs, the MVUs do not maximize the likelihood function, and thus it is no longer valid to calculate $\mathrm{AIC_C}$ values in the usual way. We therefore derive exact $\mathrm{AIC_C}$ formulae based on our MVU estimators (electronic supplementary material, appendix S3). Finally, while our model parameter estimates are unbiased, the MCIs are nonlinear functions of these estimates and thus can be slightly biased. We therefore derive a first-order correction for our MCIs to remove most of this bias (electronic supplementary material, appendix S4). Additionally, we provide accurate 95% CIs on the MCI estimates to facilitate inference and comparison. All of our estimators can be expressed in closed form and can be efficiently calculated even for very large datasets without resorting to complicated numerical methods. Finally, all of our estimators are designed to handle missing data. Specifically, if any observation for a given timestamp is missing, all observations for that timestamp are treated as missing, and the estimators explicitly account for the time elapsed in between observations. Electronic supplementary material, appendix S5 collects the relevant equations for calculating our estimators and their confidence intervals, and the `corrMove` R package included in the supplementary material automates these calculations.

(d) Analyses and interpretation

When movement behaviour is stationary in the statistical sense, meaning that the among-individual correlation structure in the tracking data does not change in time, an MCI analysis can be performed on an entire dataset. When non-stationarity is of interest, for example when focal individuals only occasionally display correlated movements, then either a moving window analysis or a behavioural change point analysis (BCPA, *sensu* [17]) may be appropriate. While a moving window analysis is straightforward to implement, it can produce erratic MCI estimates for windows that straddle different behaviours (results not shown). We therefore focus on a partitioning algorithm for our non-stationary analyses, which allows the MCIs to be calculated over more behaviourally homogeneous partitions of the data.

Our algorithm proceeds by testing each observed time in the data as a potential transition point, selecting the best models within each of the resulting partitions, and returning the $\mathrm{AIC_C}$ values of each partition, which are then summed to obtain the $\mathrm{AIC_C}$ of the entire dataset after partitioning. The partition point that results in the lowest combined $\mathrm{AIC_C}$ is then accepted if

(i) it is lower than the best $\mathrm{AIC_C}$ of the data before new partition was introduced, and
(ii) it does not result in any partitions that are smaller than the user-defined minimum partition length W (number of observations).

The algorithm repeats, introducing one new partition point in each step until either of the above conditions is violated. The minimum partition length, W, is the only tunable parameter in the algorithm and is necessary to combat the natural tendency of many unsupervised classification algorithms, including ours, to over-fit the data. We chose $W = 25$ days in both our simulated

and empirical examples to strike a balance between retaining enough data within each partition to obtain reliable parameter estimates, and allowing the algorithm to reveal transitions in movement correlation in the data. In general, values of W that are not grossly different from the smallest true partition in the data will give reasonable results, while very small and very large W values will lead to over- and under-partitioning of the data, respectively.

A shared tendency to move in the same direction will manifest in η_{dft}. This may occur when individuals follow the same environmental gradient or current (e.g. [24]). This can be viewed as a passive or external form of movement correlation that does not require direct coordination of movement behaviour among individuals. In contrast, a tendency to turn in a correlated way will show up in η_{dif}. This happens when the movement behaviour of one individual has a measurable effect on the movement behaviour of other individuals, as would be the case when individuals directly coordinate their movements. This type of movement coordination may be particularly apparent when multiple individuals simultaneously deviate in the same direction from following an environmental gradient [11]. The diffusive correlation index, η_{dif}, can thus be interpreted as quantifying social coordination when individuals are in close enough proximity for interaction to be possible, either through direct pairwise interactions among individuals, or through interaction chains where each individual influences its neighbour(s). Our MCIs thus allow researchers to separate passive movement correlation (e.g. gradient following) from active coordination of movement behaviour, even in datasets where both forms of correlation occur simultaneously. The `corrMove` R package, which implements these methods, as well as an R script demonstrating its use, are provided in the electronic supplementary material.

(e) Simulation study

We conducted a simulation study to test the performance of our MCI estimators and partitioning algorithm. First, we examined the bias of the MCI estimators and the realized coverage of their confidence intervals, where bias is defined as the average difference between the true value of the index and its estimate under repeated sampling, and coverage is the proportion of confidence interval estimates that contained the true index value under repeated sampling. We defined a baseline set of parameters ($\nu_x = \nu_y = 2$, $\sigma = 2$, $\rho = 0.5$), and then varied sample size via both number of individuals, N, and number of movement steps, S. For each combination of N and S, we also considered intact data, and data with 25% of the observations randomly deleted. We ran 10 000 simulations for each $N \times S \times$ *missingness* combination.

We defined two parameter scenarios to test the partitioning algorithm, and within each we varied W across four values (15, 25, 35 and 45), N across five values (5, 10, 15, 20 and 25) and the percentage of missing observations over two values (0% and 25%). Scenario 1 mixed the baseline parameters above with a null parameter set ($\nu_x = \nu_y = 0$, $\rho = 0$) which resulted in different partitions in the data being defined by transitions among movement models. Specifically, our scenario 1 simulations consisted of a block of 50 steps from the UU model ($\nu_x = \nu_y = 0$, $\rho = 0$), followed by 50 steps from the CC model ($\nu_x = \nu_y = 2$, $\rho = 0.5$), 100 steps from the CU model ($\nu_x = \nu_y = 2$, $\rho = 0$), 100 steps from the UU model and finally 50 steps from the UC model ($\nu_x = \nu_y = 0$, $\rho = 0.5$). This set-up is shown in figure 1*a*. For scenario 2, we mixed the baseline set with a different non-null parameter set ($\nu_x = 5$, $\nu_y = -1$, $\sigma = 4$, $\rho = 0.25$), which resulted in different partitions in the data being defined by transitions in parameter values, with the CC model underpinning all partitions. This arrangement is displayed in figure 1*b*. We conducted 500 replicate simulations for each *scenario* $\times W \times N \times$ *missingness* combination and then tallied the proportion of simulations for which the simulated data were correctly partitioned. We considered a set of partitions to be correctly identified when both the total number of estimated partitions matched the true number of partitions, and each estimated partition point was within three observations of the corresponding true partition point.

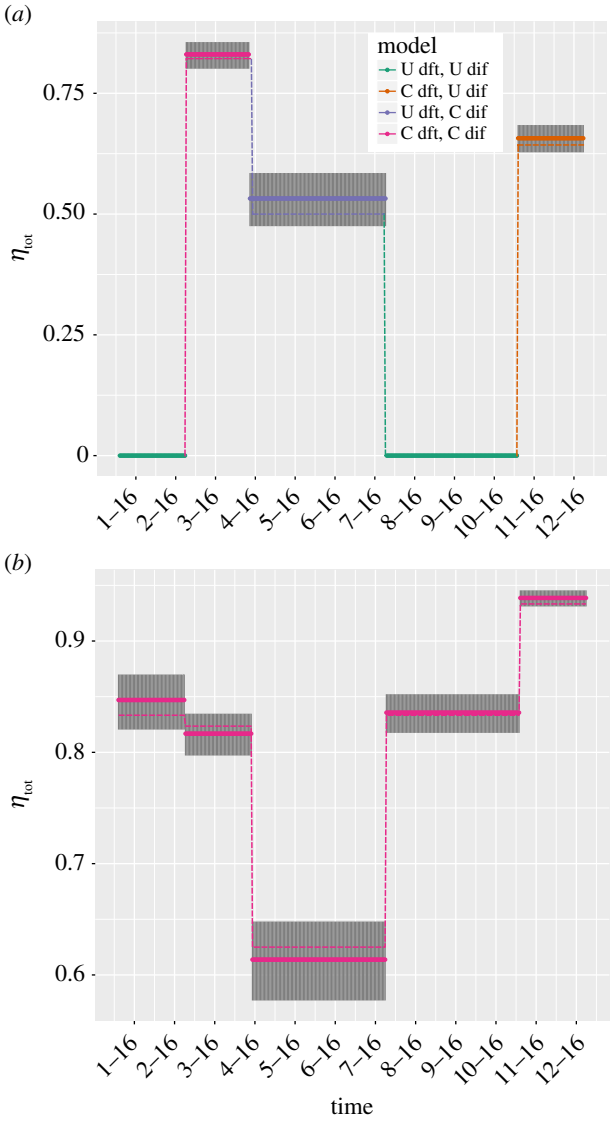

Figure 1. (*a*) A realization of a simulated partitioning experiment for parameter scenario 1 displaying the total correlation index, η_{tot}, with the dashed line denoting the true η_{tot} values and true model (via colour). Estimated η_{tot} values are displayed by points with the colour indicating the selected model for a given partition. Error bars on the η_{tot} estimates are 95% CIs. (*b*) A realization of a simulated partitioning experiment from parameter scenario 2. Note that in scenario 1, transitions between partitions correspond with transitions between models, whereas in scenario 2, transitions between partitions are driven by variation in parameter values, with the CC model being constant across all partitions.

(f) Empirical examples

Caribou. The barren-ground caribou (*R. tarandus granti*) of the Porcupine herd range across the Arctic National Wildlife Refuge in Alaska (USA) and the northern Yukon (Canada). These caribou are long-distance migrants; they use the northern coastal tundra plains as calving grounds and overwinter farther inland where mountainous alpine and taiga vegetation dominates [25,26]. We used VHF relocation data on five animals spanning a time period from January to May 1988, with daily resolution. This time span encompasses both the wintering period and the spring migration. Objectively defining the start of a migration is an active area of research in movement ecology [27,28]. We therefore focused on

Table 1. Simulation results for movement correlation index (MCI) estimation bias and confidence interval coverage. All results are based on 10 000 replicate simulations from the correlated drift, correlated diffusion (CC) model. For the coverage experiments, the nominal coverage in all cases was 95%. The parameters varied include N, S and % missing observations, with their values indicated in the table. The other parameters were fixed at $\mu_x = \mu_y = 2$, $\sigma = 2$ and $\rho = 0.5$.

		η_{dif}		η_{dft}		η_{tot}	
scenario	% missing	bias	coverage	bias	coverage	bias	coverage
$N = 2$, $S = 50$	0	−0.0004	94.2	0.0006	94.6	−0.0002	94.7
	25	−0.0003	94.3	0.0008	94.4	<−0.0001	94.7
$N = 5$, $S = 100$	0	<0.0001	94.1	<0.0001	94.6	<−0.0001	95.0
	25	<0.0001	94.7	0.0002	95.0	<0.0001	95.2
$N = 10$, $S = 200$	0	<0.0001	94.5	−0.0001	94.9	−0.0002	94.7
	25	<−0.0001	94.8	<−0.0001	95.1	−0.0002	94.9

searching for a pronounced transition in the degree of movement correlation between these two distinct behavioural periods as an indication of migration onset. To do this, we employed the partitioning algorithm with the *a priori* hypothesis that there would be a sudden increase in drift correlation at the onset of migration. The caribou dataset had a substantial proportion of missing observations (16.6%), and thus would be difficult to analyse with methods that require perfectly regular data. The caribou data are provided in the electronic supplementary material.

Khulan. Asiatic wild asses (*E. hemionus*), or khulans, live in fission–fusion groups in the Mongolian Gobi desert, and in the southwest Gobi, have non-exclusive home ranges of 4449–6835 km^2, and show little preference for any particular plant community [29]. We used data on five animals collected by store-on-board GPS collars over a six-month period (February–July 2010) with daily resolution. The fission–fusion nature of khulan groups suggests that movement correlation should vary over time. Furthermore, an extreme winter during the study period forced an unusual westward displacement and subsequent eastward spring return movement [30]. We thus used the partitioning algorithm to identify periods of the year when khulan movements were significantly correlated, and then explored how both median pairwise distance and median daily movement distance changed relative to diffusive (potentially social) correlation. The khulan data were intact over the observation period, with no missing observations. The khulan dataset is provided in the electronic supplementary material.

3. Results

(a) Simulation study

Estimates for all three MCIs showed negligible bias across our simulation study, and were largely unaffected when 25% of observations were randomly deleted (table 1). Specifically, the magnitude of the bias was always less than 0.001 across all of the scenarios we considered (table 1). Similarly, CI coverage for all three MCIs was always within 1% of the nominal 95% rate, and was also mostly unaffected by missing data (table 1). Across all of our simulation experiments, the classification accuracy of our partitioning algorithm was consistently over 70%, with the exception of simulations based on scenario 1, $W = 15$ and $N = 5$, which returned partitioning accuracies of approximately 58% for both 0 and 25% missingness (figure 2). Classification accuracy was over 99% across much of the parameter space that we tested, with accuracy tending to improve substantially as N increased and as W became more similar to the minimum true partition length in the data (50 observations; figure 2). While partitioning success did depend on W to some extent, it was not overly sensitive to this tuning parameter, and even in most cases where W was more than three times smaller than the true minimum partition length, the classification accuracy was still more than 70%. The algorithm was also quite robust to missing data, with little decrease in classification success between the intact and 25% missing observations scenarios, all else equal.

The caribou oscillated between bouts of no correlation (UU model) and correlated drift (CU model; figure 3*a*). In particular, a pronounced jump in drift correlation occurred on 14 April 1988 (figure 3*a*). At this time, four of the five tracked individuals, who had been moving independently, all simultaneously began moving northwest in the direction of the spring migration, and the fifth individual followed suite shortly thereafter. All individuals subsequently maintained their northwest heading, causing prolonged period of drift correlation (figure 3*a*). The movement tracks for three of these individuals with the estimated overwintering and migration periods colour coded are shown in figure 3*b*. When the migration began, tracked individuals were on average more than 100 km apart, suggesting that a commonly perceived environmental trigger, rather than direct coordination of movement behaviour, initiated the migration. While the exact partitioning shown in figure 3*a* was stable between $W = 18$ and $W = 25$, the 14 April 1988 partition point associated with migration onset was stable from $W = 5$ (the lowest value tested) to $W = 46$ (equal to the length of the migratory period), and was always the first and most important partition chosen, as evidenced by an AIC_C decrease of 462 relative to the original (unpartitioned) data.

The khulan analysis revealed intermittent correlation with transitions between uncorrelated movement, drift correlation and diffusive correlation. In particular, a large peak in η_{tot} occurred in April 2010 (figure 4*c*). This peak was initiated by shared drift (figure 4*a*), as all individuals had a persistent tendency to move eastward at this time. This initial bout of drift correlation was subsequently bolstered by diffusive correlation, which peaked roughly three weeks later (figure 4*b*), and then declined somewhat thereafter. By this time, all five individuals had come together as part of a large, diffuse herd and were close enough (median pairwise distance approximately 20 km, with median daily movement distance approximately 7 km d^{-1}) to coordinate their movements with other herd members (figure 4*c*, solid line). In mid-June, the period of diffusive correlation abruptly ends despite individuals being relatively close together. This transition is likely the onset of the breeding season, which typically occurs between June and August,

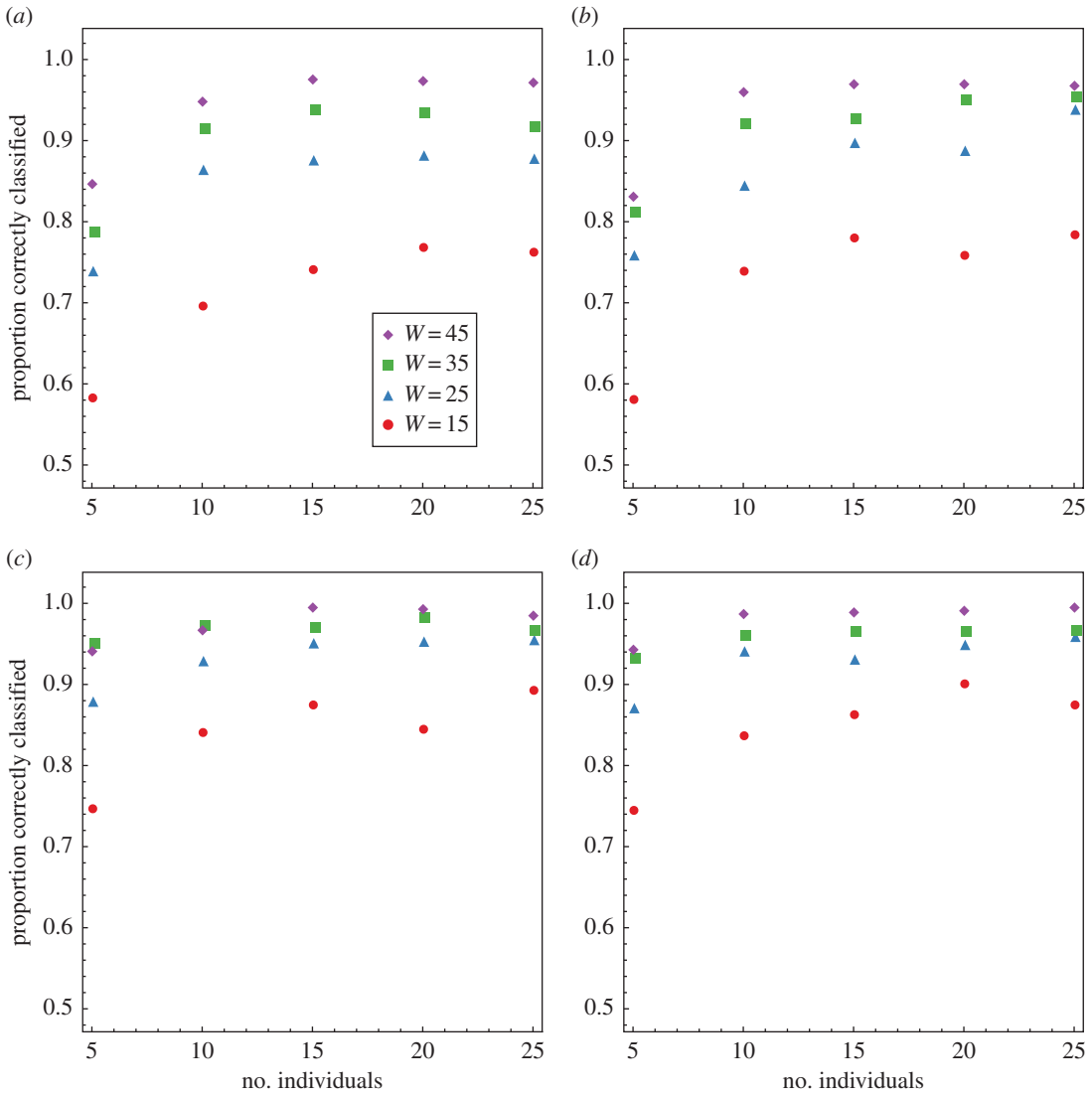

Figure 2. Classification accuracy of the partitioning algorithm on simulated data as a function of number of individuals, *N*, across parameter scenarios 1 (*a*,*b*) and 2 (*c*,*d*), with intact data (*a*,*c*) and with 25% of observations randomly missing (*b*,*d*). Curves within each panel represent different values of *W*.

and is accompanied by an abrupt reduction in daily movement distance (figure 4*c*, dashed line) as individuals become temporarily range resident. The exact partitioning shown in figure 4 was stable between $W = 14$ and $W = 28$.

4. Discussion

Understanding movement correlation provides insights into the relationship between individual movements and population-level redistribution patterns [6], and as we show in this study, it can also help disentangle environmental from social drivers of movement. Rapidly advancing tracking technology and decreasing tag costs will make multi-individual relocation datasets increasingly common, and multi-individual movement analyses increasingly relevant. While many methods aimed at understanding non-independent movement exist in the field of collective motion, most of those methods do not apply to the broad-scale, long-term data common in wildlife tracking studies. Currently, only a few approaches to quantifying movement correlation specifically from wildlife tracking data exist, and all are limited in scope or applicability.

Our MCIs combine several important properties. First, our metrics are derived from a stochastic movement model, yielding a clear interpretation of the results in terms of the shared diffusive (random) and drift (deterministic) components of movement. Second, our approach works with the relatively coarse datasets that typify wildlife ecology studies, and does not require additional information such as knowledge of a target location [7] or detailed, fine-scale estimates of velocities and accelerations of individuals relative to each other (e.g. [5]). Third, our MCIs are based on closed-form MVU estimators for model parameters, and so have good statistical properties including minimal bias and accurate CIs. Closed-form estimators allow our MCIs to be efficiently calculated on even very large datasets, whereas techniques employing complex numerical methods (e.g. [9,12] can be prohibitively slow on large datasets. Our framework easily accommodates any number of individuals, while approaches such as wavelet coherence analyses [4,13] are difficult to extend beyond pairwise analyses (but see [31]). Finally, unlike all other model-based approaches of which we are aware, our approach accommodates non-stationarity in the degree and nature of correlated movements, which is likely ubiquitous in the long-term data typical of wildlife tracking studies.

Of recent efforts to quantify correlated movements among individuals, Dalziel *et al.* [11] is perhaps the most similar, in terms of intent, to the method we introduce here. Both of these approaches focus on separating correlated diffusion from correlated drift for an arbitrary number of individuals

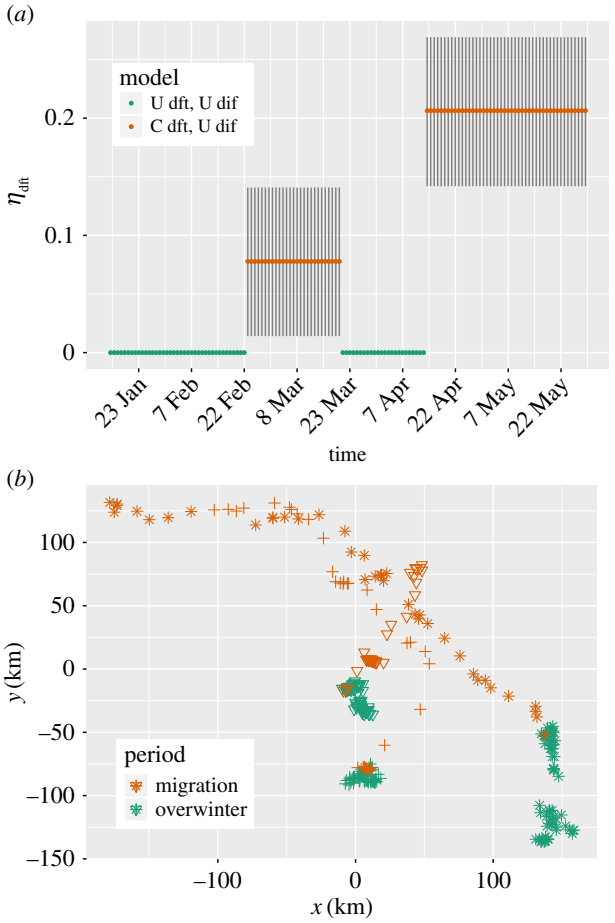

Figure 3. (*a*) Partitioned MCI analysis with a minimum partition length of $W = 25$ days for caribou focusing on η_{dft}. Colours indicate the selected model for each partition of the data, while error bars on the MCI estimates are 95% confidence intervals. Drift correlation jumps abruptly on 14 April 1988 and then remains high thereafter, which is consistent with this date being the onset of spring migration. (*b*) Caribou tracks with individuals denoted by different symbols and the overwintering period (green) and spring migration (orange) denoted by colour. To eliminate visual clutter, only three of the five individuals are shown, and the period of drift correlation apparent in (*a*) from late February to late March is not shown in (*b*).

when the correlations in the movement may be non-stationary in time. The approach of Dalziel *et al*. [11] has the advantage of not requiring assumptions on the particular form of the underlying movement model and has a very intuitive interpretation. However, their method relies on five manually tuned parameters (spatial and temporal bandwidths for the drift, spatial and temporal bandwidths for conspecific interactions, and the window over which interquartile ranges are calculated), while our approach requires only a single tuning parameter, *W*, to which we have demonstrated that both our simulated and empirical results are not overly sensitive. Additionally, their approach uses only interquartile ranges as a basis for inference, while our method is grounded in statistical theory and produces point estimates of known quality accompanied by accurate confidence intervals. Finally, their approach lacks a software implementation, while we provide an R package that implements our methods and facilitates visualization of the results.

For caribou, our MCIs identified the onset of migration via an abrupt transition from uncorrelated to highly correlated movement. Importantly, we identified this transition without using any outside information on when the spring migration begins, and the result is consistent with expert knowledge on migration start date [26]. Identifying migration onset from relocation data is currently an area of active research [27,28], and using the interrelationships among individuals to estimate the start date of a mass migration is a novel solution to this problem. That the caribou began moving in a correlated way despite being more than 100 km apart on average suggests that a coarse-grained environmental cue triggers migration. Alternatively, it might be possible that some kind of social signal that propagated via interaction chains across the very large Porcupine herd could have triggered migration onset. For example, Torney *et al*. [32] have identified fine-scale movement coordination among migrating caribou based on the analysis of aerial video footage. However, linking such a fine-grained signal to broad-scale migration initiation would require additional information such as the density of unobserved individuals in between tracked individuals. In either case, we note that an approach that imposed an *a priori* upper limit on the interaction distance over which correlated movements are thought to be possible (see examples in [10,11]) would have missed this long-distance signal entirely.

We showed that khulans feature intermittent movement correlation. This is a novel finding as it was previously unknown to what extent and at which times khulans display correlated movements. Decomposing khulan movement into correlated diffusive and drift components suggested that the April peak in correlation was initiated by a shared tendency to move from west to east. This coherent movement represented a return to their normal range after an extreme winter forced the usually nomadic khulans to 'migrate' to the west [30]. As individuals moved to the east, they came together in a large herd and began directly coordinating their movements, likely through chains of interactions. Coordinated movement within the herd continued until mid-June, when daily move distances suddenly plummeted, which likely corresponded to the onset of the breeding season. This result shows how drift correlation can set the stage for diffusive correlation by bringing individuals in closer proximity to one another. More generally, these results suggest that an MCI analysis could detect the influence of changing environmental conditions on movement correlation.

Despite its advantages, there are several areas where our approach could be improved. First, because of the assumptions made in the Brownian motion model, our MCIs cannot be directly compared across different sampling schedules. This kind of sampling dependence is also an issue for most current path-based movement analyses, such as discrete-time correlated random walks and their extensions [3]. Our model also assumes that movement steps (not positions) are uncorrelated in time, which will not be true for finely sampled data, though with the exceptions of Polansky *et al*. [4] and Polansky & Wittemyer [13], other model-based methods for quantifying movement correlation also ignore fine-scale autocorrelation (e.g. [12]). Future efforts that develop MCIs from more sophisticated movement models that feature correlated velocities (e.g. [3,19,21,33]) could overcome both of these limitations and would allow our approach to be applied to the high-density, controlled condition datasets that typify collective motion studies. However, basing MCIs on continuous-velocity movement processes will be a major undertaking, and will likely break some of the analytical tractability of our approach. Finally, one biologically relevant scenario that presents difficulties for our method occurs when individuals approach a point (e.g. a watering hole) from opposite directions.

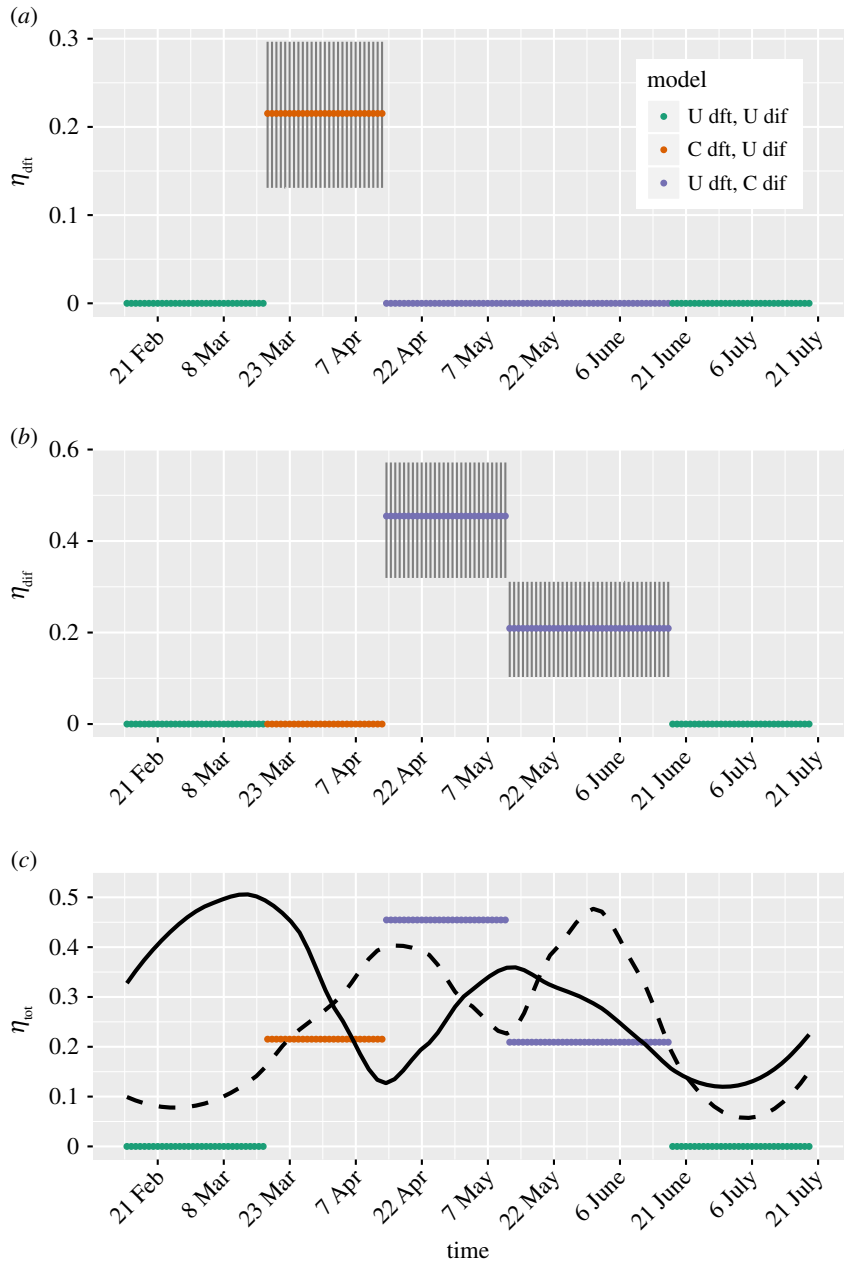

Figure 4. Partitioned MCI analysis with a minimum partition length of $W = 25$ days for the khulans. (*a*) Shows the drift correlation index, (*b*) presents the diffusive correlation index and (*c*) shows the total correlation index. Colours indicate the selected model for each partition, while error bars on the MCI estimates are 95% confidence intervals. The solid black curve in (*c*) is the Loess fit to the median pairwise distance between individuals, while the dashed lines is a Loess fit to median daily move distance. Both distance curves are scaled so they can be displayed on the MCI plot. Note that the CIs on the η_{tot} in (*c*) are suppressed to prevent visual clutter, but are equivalent to those on each of the components in (*a*) (η_{dft}) and (*b*) (η_{dif}).

Generalizing the underlying model on which our MCIs are based to allow for individual-specific drift components would alleviate this limitation, but will come at the cost of additional mathematical complexity.

Regarding our partitioning approach, unsupervised classification algorithms, which typically focus on identifying clusters, partitions or segments in a dataset, are used across the sciences and have a notorious tendency to overfit the data [34,35]. To the best of our knowledge, there is no generally applicable formal solution to this problem. The task of identifying behaviourally homogeneous subsets of a movement path belongs to this class of problems and is an active area of research in movement ecology (e.g. [17,36]). Indeed, across the collection of such methods reviewed by Gurarie *et al.* [36], no single approach worked well across all scenarios considered. Similar to other unsupervised classification approaches, our path segmentation algorithm requires a user-defined tuning parameter (W), but as long as W is not grossly different from the smallest true partition size in the data, our approach will return reasonable results. Developing more objective algorithms that do not require such parameters is an obvious avenue for further research.

Collective motion studies often focus on behaviours that allow information transmission within groups. One example is group navigation, where one or a few individuals experience an environmental stimulus (such as a predator, shelter or forage), and these 'informed' individuals transmit behavioural information to others, allowing efficient navigation [37–39]. Another core focus is flocking or schooling behaviours, where individuals average movement parameters to arrive at a consensus direction that, for larger groups, allows efficient group movements (many wrong principles [39–41]). Like our approach, these studies exist at the interface of environmental and social interactions, but they differ from the approach taken here by focusing on fine-scale mechanisms of navigation.

In contrast, we have taken a phenomenological approach that aims to quantity two broad categories of correlations (drift versus diffusive) in tracking data, and thus to suggest the class of mechanisms (environmental or social) that generated them. Our method is therefore built on a statistical model of correlated movements rather than on explicit behavioural mechanisms of interacting organisms. The detailed information required in many collective movement studies for examining behavioural mechanisms is often missing in wildlife tracking studies, yet such species will often feature highly non-independent movements. Our approach can, therefore, serve as a bridge between fine-scale, mechanistic collective motion studies and traditional analyses of wildlife tracking data that treat individuals independently.

Ethics. The caribou data were collected by the Porcupine Caribou Technical Committee, which consists of the authority-granting agencies governing monitoring of the Porcupine Caribou Herd. The khulan data were collected under the authority of a Memorandum of Understanding between the International Takhi Group, the Mongolian Ministry of the Environment and the Committee for the Protection of Rare Species.

Data accessibility. The datasets supporting this article have been uploaded as part of the electronic supplementary material. An archival version of the `corrMove` R package is included in the electronic supplementary material, and the current version of the package can be obtained at https://github.com/jmcalabrese/corrMove. An R script demonstrating use of the `corrMove` package has also been uploaded as part of the electronic supplementary material.

Authors' contributions. J.M.C., T.M., C.H.F. and W.F.F. conceived the study. C.H.F., J.M.C. and S.B. developed the models and statistical methods. T.M. and M.R. obtained and processed the data. P.K. provided the khulan data and provided guidance on khulan biology. J.M.C. wrote the R package and performed the analyses. J.M.C., T.M., C.H.F. and W.F.F. wrote the first draft of the paper. All authors contributed to subsequent revisions of the manuscript and approved the final version to be published.

Competing interests. We declare we have no competing interests.

Funding. NSF ABI 1062411 (to T.M.) and ABI 1458748 (to J.M.C.) supported this research, C.H.F. was supported by a Smithsonian Postdoctoral Fellowship, and T.M. was supported by the Bosch Foundation. The khulan field work was funded by the Austrian Science Foundation (FWF) project P 18624.

Acknowledgments. We thank the Porcupine Caribou Technical Committee for their assistance and for providing the caribou data.

References

1. Nathan R *et al.* 2008 A movement ecology paradigm for unifying organismal movement research. *Proc. Natl Acad. Sci. USA* **105**, 19 052–19 059. (doi:10.1073/pnas.0800375105)
2. Schick RS *et al.* 2008 Understanding movement data and movement processes: current and emerging directions. *Ecol. Lett.* **11**, 1338–1350. (doi:10.1111/j.1461-0248.2008.01249.x)
3. Fleming CH, Calabrese JM, Mueller T, Olson KA, Leimgruber P, Fagan WF. 2014 From fine-scale foraging to home ranges: a semi-variance approach to identifying movement modes across spatiotemporal scales. *Am. Nat.* **183**, E154–E167. (doi:10.1086/675504)
4. Polansky L, Wittemyer G, Cross PC, Tambling CJ, Getz WM. 2010 From moonlight to movement and synchronized randomness: Fourier and wavelet analyses of animal location time series data. *Ecology* **91**, 1506–1518. (doi:10.1890/08-2159.1)
5. Katz Y, Tunstrøm K, Ioannou CC, Huepe C, Couzin ID. 2011 Inferring the structure and dynamics of interactions in schooling fish. *Proc. Natl Acad. Sci. USA* **108**, 18 720–18 725. (doi:10.1073/pnas.1107583108)
6. Mueller T *et al.* 2011 How landscape dynamics link individual to population-level movement patterns: a multispecies comparison of ungulate relocation data. *Glob. Ecol. Biogeogr.* **20**, 683–694. (doi:10.1111/j.1466-8238.2010.00638.x)
7. Bode NWF, Franks DW, Wood AJ, Piercy JJB, Croft DP, Codling EA. 2012 Distinguishing social from nonsocial navigation in moving animal groups. *Am. Nat.* **179**, 621–632. (doi:10.1086/665005)
8. Delgado MdM, Penteriani V, Morales JM, Gurarie E, Ovaskainen O. 2014 A statistical framework for inferring the influence of conspecifics on movement behaviour. *Methods Ecol. Evol.* **5**, 183–189. (doi:10.1111/2041-210X.12154)
9. Langrock R *et al.* 2014 Modelling group dynamic animal movement. *Methods Ecol. Evol.* **5**, 190–199. (doi:10.1111/2041-210X.12155)
10. Long JA, Nelson TA, Webb SL, Gee KL. 2014 A critical examination of indices of dynamic interaction for wildlife telemetry studies. *J. Anim. Ecol.* **83**, 1216–1233. (doi:10.1111/1365-2656.12198)
11. Dalziel BD, Corre ML, Côté SD, Ellner SP. 2016 Detecting collective behaviour in animal relocation data, with application to migrating caribou. *Methods Ecol. Evol.* **7**, 30–41. (doi:10.1111/2041-210X.12437)
12. Niu M, Blackwell PG, Skarin A. 2016 Modeling interdependent animal movement in continuous time. *Biometrics* **72**, 315–324. (doi:10.1111/biom.12454)
13. Polansky L, Wittemyer G. 2011 A framework for understanding the architecture of collective movements using pairwise analyses of animal movement data. *J. R. Soc. Interface* **8**, 322–333. (doi:10.1098/rsif.2010.0389)
14. Strandburg-Peshkin A, Farine DR, Couzin ID, Crofoot MC. 2015 Shared decision-making drives collective movement in wild baboons. *Science* **348**, 1358–1361. (doi:10.1126/science.aaa5099)
15. Hughey LF, Hein AM, Strandburg-Peshkin A, Jensen FH. 2018 Challenges and solutions for studying collective animal behaviour in the wild. *Phil. Trans. R. Soc. B* **373**, 20170005. (doi:10.1098/rstb.2017.0005)
16. Fleming CH, Calabrese JM. 2017 A new kernel density estimator for accurate home-range and species-range area estimation. *Methods Ecol. Evol.* **8**, 571–579. (doi:10.1111/2041-210X.12673)
17. Gurarie E, Andrews RD, Laidre KL. 2009 A novel method for identifying behavioural changes in animal movement data. *Ecol. Lett.* **12**, 395–408. (doi:10.1111/j.1461-0248.2009.01293.x)
18. Sridhar H, Guttal V. 2018 Friendship across species borders: factors that facilitate and constrain heterospecific sociality. *Phil. Trans. R. Soc. B* **373**, 20170014. (doi:10.1098/rstb.2017.0014)
19. Johnson DS, London JM, Lea MA, Durban JW. 2008 Continuous-time correlated random walk model for animal telemetry data. *Ecology* **89**, 1208–1215. (doi:10.1890/07-1032.1)
20. Gurarie E, Ovaskainen O. 2011 Characteristic spatial and temporal scales unify models of animal movement. *Am. Nat.* **178**, 113–123. (doi:10.1086/660285)
21. Fleming CH, Calabrese JM, Mueller T, Olson KA, Leimgruber P, Fagan WF. 2014 Non-Markovian maximum likelihood estimation of autocorrelated movement processes. *Methods Ecol. Evol.* **5**, 462–472. (doi:10.1111/2041-210X.12176)
22. Fleming CH, Fagan WF, Mueller T, Olson KA, Leimgruber P, Calabrese JM. 2016 Estimating where and how animals travel: an optimal framework for path reconstruction from autocorrelated tracking data. *Ecology* **97**, 576–583. (doi:10.1890/15-1607)
23. Burnham KP, Anderson DR. 2002 *Model selection and multimodel inference: a practical information-theoretic approach*. Berlin, Germany: Springer Science & Business Media.
24. Nagy M, Couzin ID, Fiedler W, Wikelski M, Flack A. 2018 Synchronization, coordination and collective sensing during thermalling flight of freely migrating white storks. *Phil. Trans. R. Soc. B* **373**, 20170011. (doi:10.1098/rstb.2017.0011)
25. Fancy SG, Pank LF, Whitten KR, Regelin WL. 1989 Seasonal movements of caribou in arctic Alaska as determined by satellite. *Can. J. Zool.* **67**, 644–650. (doi:10.1139/z89-093)

26. Griffith B *et al.* 2002 Section 3: the porcupine caribou herd. *US Geological Survey, Biological Resources Division, Biological Science Report* no. USGS/BRD/BSR-2002-0001, pp. 8–37. Reston, VA: USGS. See https://alaska.usgs.gov/products/pubs/2002/2002-USGS-BRD-BSR-2002-0001.pdf.
27. Bunnefeld N, Börger L, van Moorter B, Rolandsen CM, Dettki H, Solberg EJ, Ericsson G. 2011 A model-driven approach to quantify migration patterns: individual, regional and yearly differences. *J. Anim. Ecol.* **80**, 466–476. (doi:10.1111/j.1365-2656.2010.01776.x)
28. Gurarie E, Cagnacci F, Peters W, Fleming CH, Calabrese JM, Mueller T, Fagan WF. 2017 A framework for modelling range shifts and migrations: asking when, whither, whether and will it return. *J. Anim. Ecol.* **86**, 943–959. (doi:10.1111/1365-2656.12674)
29. Kaczensky P, Ganbaatar O, Von Wehrden H, Walzer C. 2008 Resource selection by sympatric wild equids in the Mongolian Gobi. *J. Appl. Ecol.* **45**, 1762–1769. (doi:10.1111/j.1365-2664.2008.01565.x)
30. Kaczensky P, Ganbataar O, Altansukh N, Enkhsaikhan N, Stauffer C, Walzer C. 2011 The danger of having all your eggs in one basket–winter crash of the re-introduced Przewalski's horses in the Mongolian Gobi. *PLoS ONE* **6**, e28057. (doi:10.1371/journal.pone.0028057)
31. Rouyer T, Fromentin JM, Stenseth NC, Cazelles B. 2008 Analysing multiple time series and extending significance testing in wavelet analysis. *Mar. Ecol. Prog. Ser.* **359**, 11–23. (doi:10.3354/meps07330)
32. Torney CJ, Lamont M, Debell L, Angohiatok RJ, Leclerc L-M, Berdahl AM. 2018 Inferring the rules of social interaction in migrating caribou. *Phil. Trans. R. Soc. B* **373**, 20170385. (doi:10.1098/rstb.2017.0385)
33. Fleming CH, Sheldon D, Gurarie E, Fagan WF, LaPoint S, Calabrese JM. 2017 Kálmán filters for continuous-time movement models. *Ecol. Inform.* **40**, 8–21. (doi:10.1016/j.ecoinf.2017.04.008)
34. Hastie T, Tibshirani R, Friedman J. 2009 *The elements of statistical learning*. Berlin, Germany: Springer.
35. Murphy K. 2012 *Machine learning*. Cambridge, MA: MIT Press.
36. Gurarie E, Bracis C, Delgado M, Meckley TD, Kojola I, Wagner CM. 2016 What is the animal doing? Tools for exploring behavioural structure in animal movements. *J. Anim. Ecol.* **85**, 69–84. (doi:10.1111/1365-2656.12379)
37. Couzin I. 2007 Collective minds. *Nature* **445**, 715. (doi:10.1038/445715a)
38. Berdahl A, Torney CJ, Ioannou CC, Faria JJ, Couzin ID. 2013 Emergent sensing of complex environments by mobile animal groups. *Science* **339**, 574–576. (doi:10.1126/science.1225883)
39. Berdahl AM, Kao AB, Flack A, Westley PAH, Codling EA, Couzin ID, Dell AI, Biro D. 2018 Collective animal navigation and migratory culture: from theoretical models to empirical evidence. *Phil. Trans. R. Soc. B* **373**, 20170009. (doi:10.1098/rstb.2017.0009)
40. Simons AM. 2004 Many wrongs: the advantage of group navigation. *Trends Ecol. Evol.* **19**, 453–455. (doi:10.1016/j.tree.2004.07.001)
41. Codling E, Pitchford J, Simpson S. 2007 Group navigation and the 'many-wrongs' principle in models of animal movement. *Ecology* **88**, 1864–1870. (doi:10.1890/06-0854.1)