CrossMark

Hierarchical Nonlinear Spatio-temporal Agent-Based Models for Collective Animal Movement

Patrick L. McDermott, Christopher K. Wikle, and
Joshua Millspaugh

Modeling complex collective animal movement presents distinct challenges. In particular, modeling the interactions between animals and the nonlinear behaviors associated with these interactions, while accounting for uncertainty in data, model, and parameters, requires a flexible modeling framework. To address these challenges, we propose a general hierarchical framework for modeling collective movement behavior with multiple stages. Each of these stages can be thought of as processes that are flexible enough to model a variety of complex behaviors. For example, self-propelled particle (SPP) models (e.g., Vicsek et al. in Phys Rev Lett 75:1226–1229, 1995) represent collective behavior and are often applied in the physics and biology literature. To date, the study and application of these models has almost exclusively focused on simulation studies, with less attention given to rigorously quantifying the uncertainty. Here, we demonstrate our general framework with a hierarchical version of the SPP model applied to collective animal movement. This structure allows us to make inference on potential covariates (e.g., habitat) that describe the behavior of agents and rigorously quantify uncertainty. Further, this framework allows for the discrete time prediction of animal locations in the presence of missing observations. Due to the computational challenges associated with the proposed model, we develop an approximate Bayesian computation algorithm for estimation. We illustrate the hierarchical SPP methodology with a simulation study and by modeling the movement of guppies.

Supplementary materials accompanying this paper appear online.

Key Words: Agent-based models; Approximate Bayesian computation; Collective movement; Hierarchical model; Nonlinear modeling; Self-propelled particle model.

1. INTRODUCTION

Due to the rapid increase in available radio-tracking technology (Millspaugh et al. 2012), researchers have been tasked increasingly with modeling complex animal behavior from

P. L. McDermott (✉) and C. K. Wikle Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211, USA (E-mail: *plmyt7@mail.missouri.edu*). J. Millspaugh Wildlife Biology Program, University of Montana, Missoula, MT 59812, USA.

large datasets. Although data availability has improved the breadth of questions that can be addressed, modeling species interactions is complicated by the nonlinear nature of the interactions, complexity of data structures, uncertainty in data, and interactions with environmental covariates. Modeling animal movement in a formal statistical framework has been studied extensively in the literature (e.g., for a recent overview see Hooten et al. 2017), but fewer advancements have been made in modeling collective animal movements and interactions using fine-scale data now offered by recent technology. Understanding of complex species interactions, such as social hierarchies and predator–prey relationships, could benefit from the development of a flexible modeling framework that evaluates collective movement behavior.

Although modeling individual animal movement has a long history in the ecological and statistics literature (e.g., see the overview in Hooten et al. 2017), until more recently there has been less focus on collective animal movement. Recent work in this area has considered a wide range of modeling frameworks (e.g., Perna et al. 2014; Bonnell et al. 2016). In addition, there is increasingly more focus on analyzing collective animal movement within a formal statistical framework that can account for uncertainty data, process, and parameters (e.g., Langrock et al. 2014; Scharf et al. 2017; Dalziel et al. 2016; Russell et al. 2016). Yet, the inherent nonlinear behavior present in collective movement processes is difficult to capture with many traditional statistical models. Therefore, the challenges associated with modeling collective animal movement requires researchers to employ complex models, and many of the motivating mechanistic models can come from other subject matter disciplines. For instance, collective animal movement is modeled with a spatial point process interaction function in Russell et al. (2016) and with network modeling in Scharf et al. (2017).

Modeling the collective behavior of animals across many temporal and spatial scales fits naturally within the broader class of dynamic spatio-temporal models (DSTMs) (e.g., Cressie and Wikle 2011). However, in the statistical literature these models have tended to focus on global process behavior, rather than local decision-based behavior. Exceptions include agent-based models (ABMs), which have shown promise as a DSTM for animal movement (Hooten and Wikle 2010). ABMs take a "bottom-up" approach by modeling the behavior of individual agents in terms of a set of relatively simple "rules." As detailed in Wolfram (1983), these simple rules can lead to complex behavior and thus lead to complex collective behavior for the group of agents.

The highly cited self-propelled particle (SPP) model from Vicsek et al. (1995) is an example of a widely applied ABM for modeling the collective behavior of objects, both inanimate and animate. SPP models are popular due to their simplicity and demonstrated ability to describe collective motion of objects. For example, SPP models have been used to simulate the collective movement of animals that flock or school such as birds or fish (e.g., Yates et al. 2010). Typically, SPP models are utilized as simulation models to investigate various types of collective behavior (e.g., Yates et al. 2010; Hubbard et al. 2004; Mann 2011) as a function of the few parameters that control behavior (namely, particles move in a direction similar to their neighbors). Indeed, the appeal of these models is that they can capture quite complex and nonlinear spatio-temporal behavior with relatively few parameters. However, these models are rarely used in a rigorous statistical estimation and prediction framework to model animal behavior.

The methodology proposed here relies on a hierarchical model to characterize the uncertainty in the agent (animal) observations (such as measurement error) as well as that associated with the processes that describe an agent's movement. In particular, latent agent locations are updated at each discrete time period with separate speed and directional angle sub-models (see Eqs. (3) and (4), respectively) that determine the agent's velocity vector. Discussed in detail in Hooten et al. (2017), the so-called velocity models provide a discrete time approach for modeling movement dynamics. By decomposing an agent's velocity into separate processes, we can incorporate more flexible sub-models that capture complex behavior at multiple stages. Placing the proposed methodology within a hierarchical structure also allows us to model uncertainty at multiple levels, as well as environmental covariates at various stages in the model (e.g., location, time, habitat features). Within this framework, inference on the environmental covariates that determine an agent's movement can be conducted, and one can also predict the latent process for any discrete time in the presence of missing observations. The parameters that control these processes can also be thought of as processes (e.g., spatial, temporal) and further modeled with covariates or other dependence models (e.g., Cressie and Wikle 2011). Although the specific implementation considered here for illustration utilizes the SPP model, other interaction models (e.g., such as the model used in Langrock et al. 2014) could easily be incorporated.

The proposed hierarchical model is implemented within a Bayesian estimation paradigm. By using a Bayesian hierarchical model (BHM), we quantify uncertainty at each stage of the model, and with a separate data model, we can make predictions when data are missing. Due to the unique computational challenges associated with the proposed modeling framework, inference is carried out with approximate Bayesian computation (ABC) in conjunction with Markov chain Monte Carlo (MCMC). ABC algorithms are often used for likelihoods that are either intractable or difficult to compute (e.g., Beaumont et al. 2010; Marin et al. 2012), such as the SPP model. In general, the literature on implementing ABC methods with multi-stage hierarchical models (i.e., more than two levels) is limited. Thus, we propose a hybrid ABC–MCMC algorithm to implement the proposed hierarchical collective movement model.

2. HIERARCHICAL COLLECTIVE MOVEMENT MODEL

We begin by introducing a general hierarchical framework for modeling collective animal movement. We follow the commonly used hierarchical modeling paradigm of specifying data, process, and parameter distributions as outlined in Berliner (1996).

Suppose we have discrete referenced time periods $t = 0, \ldots, T$ and denote the agents in the model by $i = 1, \ldots, N$. For each time period, the observed locations of m_t of N possible agents is defined as $\underline{\mathbf{z}}_t \equiv \{\mathbf{z}'_{i,t}\}_{i \in \{1, \ldots, m_t\}}$, where $\mathbf{z}_{i,t} \equiv (x_{i,t}, y_{i,t})'$, and $x_{i,t}$ and $y_{i,t}$ are the coordinates (in $\mathbb{R}^2$) of the location of the ith agent at time t. Likewise, let the latent location process be given by $\underline{\mathbf{s}}_t \equiv (\mathbf{s}'_{1,t}, \ldots, \mathbf{s}'_{N,t})'$, which denotes the latent locations of the agents at time period t, where $\mathbf{s}_{i,t} \equiv (\tilde{x}_{i,t}, \tilde{y}_{i,t})'$, and $\tilde{x}_{i,t}$ and $\tilde{y}_{i,t}$ are the latent location coordinates (in $\mathbb{R}^2$) of the ith agent at time t. At each time period t, we then have the following data model:

$$\underline{\mathbf{z}}_t = \mathbf{H}_t \, \underline{\mathbf{s}}_t + \underline{\boldsymbol{\epsilon}}_t, \qquad \underline{\boldsymbol{\epsilon}}_t \sim \text{Gau}(\mathbf{0}, \mathbf{R}_{z,t}), \qquad (1)$$

where $\mathbf{H}_t$ is an $2m_t \times 2N$ incidence matrix of zeros and ones used to account for missing locations (e.g., see Chapter 7 of Cressie and Wikle 2011, for further details) and $\mathbf{R}_{z,t}$ is an $2m_t \times 2m_t$ error variance–covariance matrix for the $2m_t$-dimensional error process, $\boldsymbol{\epsilon}_t$. In the example to follow, we make a simplifying assumption that $\mathbf{R}_{z,t} \equiv \sigma_\epsilon^2 \mathbf{I}$, which assumes independent and homogeneous measurement error. However, more complex data models that account for dependence, change-of-support, dimension reduction, or nonlinear transformations could also be used here as is common in spatio-temporal statistics (e.g., see chapter 7 of Cressie and Wikle 2011, for other examples), or one can incorporate the unique measurement error structures associated with telemetry data (e.g., see the discussion in Hooten et al. 2017).

Next, the process model for each agent at time t is defined generally as:

$$\mathbf{s}_{i,t} = f(\mathbf{s}_{i,t-1} + u_{i,t}\boldsymbol{\delta}_{i,t}; \Theta_f) + \eta_{i,t}, \quad \eta_{i,t} \overset{iid}{\sim} [\eta], \tag{2}$$

where $\eta_{i,t} = (\eta_{x,i,t}, \eta_{y,i,t})'$ and $[\eta]$ is any valid probability density function (pdf). For (2) and the process equations to follow, the notation iid implies that the noise terms are independently and identically distributed in both space and time. The iid assumption for the noise terms can easily be altered. Regarding the process model, $u_{i,t}$ is a scalar representing speed, while $\boldsymbol{\delta}_{i,t} = (\delta_{x,i,t}, \delta_{y,i,t})'$ is a unit vector denoting the directional component of an agent's velocity. Thus, the term $u_{i,t}\boldsymbol{\delta}_{i,t}$ defines a vector corresponding to an agent's velocity. In addition, the parameter-dependent function $f(\cdot)$ is included for the sake of generality. For instance, in the application to follow we use $f(\cdot)$ to enforce boundary conditions that are modeled with specific parameters (i.e., Θ_f).

In this framework, the movement of an agent is largely driven by the processes used to model the speed and directional angle of an agent. The directional component of an agent's velocity (i.e., the compass heading of the agent) is determined by the angle parameter $\theta_{i,t}$ where $(\delta_{x,i,t}, \delta_{y,i,t})' = (\cos(\theta_{i,t}), \sin(\theta_{i,t}))'$. Generally, we define the directional angle model for agent i at time period t as:

$$\theta_{i,t} = g(\mathbf{s}_{i,t-1}; \Theta_\delta) + \gamma_{i,t}, \quad \gamma_{i,t} \overset{iid}{\sim} [\gamma], \tag{3}$$

where $[\gamma]$ is a pdf used to model the so-called angular noise, $\gamma_{i,t}$ (e.g., Yates et al. 2010) of the agent. Because an agent often has an imperfect ability to sense the location of nearby agents (Yates et al. 2010), angular noise is often incorporated into the classic SPP model. In this case, Θ_δ could include parameters associated with terms for environmental covariates and the interactions between agent locations. Moreover, for collective movement applications, it is likely that $g(\cdot)$ will be a nonlinear function. For example, in collective animal behavior models it is common to include terms to model avoidance, alignment, and attraction behaviors among animals (e.g., Gautrais et al. 2008; Scharf et al. 2017). Although all of these behaviors are not explicitly modeled in the application to follow, $g(\cdot)$ could account for them (with each behavior having specific parameters modeled as processes). We adopt the common practice in the SPP literature of modeling the angular noise of an agent with a uniform distribution (e.g., Yates et al. 2010). Alternatively, the angular noise could also be modeled with a von Misses distribution (i.e., circular Normal distribution) or with a wrapped Gaussian distribution (e.g., Mann 2011).

To complete the general hierarchical movement model framework, we specify the speed model as follows:

$$u_{i,t} = h(\mathbf{s}_{i,t-1}; \Theta_u) + \phi_{i,t}, \quad \phi_{i,t} \overset{iid}{\sim} [\phi], \tag{4}$$

where $[\phi]$ is a pdf that accounts for the speed variation in an agent's movement. For many applications, it is likely that environmental covariates could explain some of the variation in an agent's speed and thus, these would be included in $h(\cdot)$. Further, interaction models can also be included in $h(\cdot)$. For example, Hubbard et al. (2004) used SPP interaction models in both direction and speed through the use of a common parameter. While some collective movement models in the literature specify the speed of an agent deterministically (e.g., Mann 2011), by modeling the speed probabilistically, the model has the potential to capture more realistic behavior. To complete the specification of the model, we define all of the parameters as $\Theta \equiv (\Theta_f, \Theta_\delta, \Theta_u)'$ and let $[\Theta]$ represent the prior distribution on these parameters (where the notation $[\cdot]$ denotes a pdf). These prior distributions are application specific (e.g., see "Appendix B" for the specific priors corresponding to the model for our application).

3. APPROXIMATE COMPUTATION METHODOLOGY

The general hierarchical movement model presented above presents distinct computational challenges for conducting inference on the model parameters and state variables. One solution that has recently become popular is approximate Bayesian computation (ABC). Often referred to as a "likelihood-free" method, ABC is a Bayesian computational method for likelihoods that are either completely intractable or difficult to evaluate. First proposed for solving genetic problems (e.g., Tavaré et al. 1997), ABC has been applied to a wide range of applications (e.g., Beaumont et al. 2010; Marin et al. 2012).

3.1. BASIC ABC ALGORITHM

In the following, we briefly review the basic mechanics of applying ABC. Suppose the likelihood $[\mathbf{y} \mid \theta]$ generates the process of interest, where $\mathbf{y}$ denotes the observed data and θ represents a vector of parameters with prior distribution $[\theta]$. Assuming we want to sample from $[\mathbf{y} \mid \theta]$, the following simple accept–reject algorithm can be implemented to approximate the posterior distribution $[\theta \mid \mathbf{y}]$ (Marin et al. 2012):

Algorithm 1 Basic ABC algorithm

Select a summary statistic(s) $m(\cdot)$, distance metric $d(\cdot)$, and tolerance $\epsilon \geq 0$.
for $j = 1$ to J **do**
   repeat
    1. Sample θ^* from prior distribution $[\theta]$.
    2. Simulate a dataset $\widetilde{\mathbf{y}}$ from the model $[\mathbf{y} \mid \theta^*]$.
   until $d(m(\mathbf{y}), m(\widetilde{\mathbf{y}})) \leq \epsilon$
   set $\theta^{(j)} = \theta^*$
end for

Throughout the ABC literature, the problem of how to select an appropriate summary statistic (i.e., $m(\cdot)$ in the above algorithm) has been thoroughly studied. In general, summary statistics are used as a way to efficiently capture information about the observed data. Ideally, the summary statistics for the observed data will be similar to the summary statistics for the simulated data. It can be shown that when this summary statistic is a sufficient statistic for θ and the tolerance is zero (i.e., $\epsilon = 0$) ABC produces samples from the exact posterior. Unfortunately, in complicated models (such as hierarchical DSTMs) sufficient statistics are rarely available and one must select summary statistics to reduce the approximation as much as possible (i.e., to make ϵ as small as possible). Multiple methods have been proposed for selecting summary statistics (e.g., Fearnhead and Prangle 2012; Aeschbacher et al. 2012), although some of these methods do not appear feasible (for computational reasons) for complicated hierarchical models with minimal *a priori* information. From our experience, the choice of summary statistic is application dependent (as has been demonstrated throughout the ABC literature). In the application to follow, we apply two different sets of summary statistics for the proposed methodology.

There are a variety of other computational techniques besides ABC that could be employed for nonlinear dynamical collective movement models. For example, Fasiolo et al. (2016) made a comprehensive comparison of ABC methods and state-space methods (i.e., particle Markov chain Monte Carlo and iterated filtering) with regard to nonlinear dynamical state-space models. In addition, Nott et al. (2012) showed the equivalence between certain types of ABC algorithms and the so-called ensemble Kalman filter state-space methods. Overall, ABC is a very appealing option for complex models that can be represented as generative or simulation models (e.g., see Lagarrigues et al. 2015, for an ecological example). Although Fasiolo et al. (2016) described the potential loss of precision in inference for information reduction methods such as ABC, for this particular problem any loss in precision seems to be minimal (since we are able to recover many of the true parameter values in the simulation study below).

3.2. HYBRID MCMC–ABC ALGORITHM

Although the original ABC algorithm presented above is relatively simple, many more complicated ABC algorithms have been proposed. The methodology developed here, in part, utilizes the ABC–MCMC method first introduced in Marjoram et al. (2003). Specifically, the proposed algorithm uses the ABC–MCMC method to sample multiple levels of state variables and parameters within a traditional MCMC algorithm. We call this approach the "hybrid MCMC–ABC algorithm." To our knowledge, none of the existing ABC algorithms provide a framework in which the parameters, state variables, and latent locations from the presented model could all be estimated. For example, Pereyra et al. (2013) used a hybrid framework in which parameters are estimated with ABC and state variables are separately estimated with traditional Gibbs and/or Metropolis–Hastings (MH) steps. Alternatively, Picchini (2014) and Sisson and Fan (2011), along with others, provide an ABC algorithm in which both parameters and state variables can jointly be sampled. Our hybrid algorithm essentially combines both of these algorithms, while also estimating multiple sets of state variables within the ABC algorithm (i.e., speed and direction variables). That is, previous

hybrid algorithms have focused on using ABC for much simpler processes (involving only a few parameters and no state variables) than the algorithm proposed here.

We describe the details of our hybrid MCMC–ABC algorithm in the context of the presented animal movement model. First, the latent locations in (2) are sampled with one-variable-at-a-time MH steps (see supplementary materials for further details), and the σ_ϵ^2 error variance parameter is sampled via a Gibbs step. Importantly, all of the remaining parameters in the model (denoted $\Theta \equiv (\Theta_f, \Theta_\delta, \Theta_u)'$) and state variables (denoted $\mathbf{A}_{1:T} \equiv (\underline{u}_{1:T}, \underline{\delta}_{1:T})'$) are sampled using the ABC algorithm to follow. We denote the simulated locations in the ABC algorithm as $\tilde{\underline{\mathbf{s}}}_{1:T}$, where $\tilde{\underline{\mathbf{s}}}_{1:T} \equiv (\tilde{\underline{\mathbf{s}}}_1, \ldots, \tilde{\underline{\mathbf{s}}}_T)'$. The known initial locations and state variables are denoted as $\tilde{\underline{\mathbf{s}}}_0$ and $\mathbf{A}_0$, respectively. Using the previously defined notation, the hybrid MCMC–ABC algorithm is as follows:

Algorithm 2 Hybrid MCMC-ABC algorithm

At the $(j + 1)$st iteration of the MCMC-ABC algorithm:

1. Sample $\mathbf{s}_{i,t}^{(j+1)}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$ using one-variable-at-a-time MH steps (see supplementary materials).

2. Sample $\sigma_\epsilon^{2(j+1)}$ using a Gibbs step (see supplementary materials).

3. Generate proposal $\Theta^* \sim q[\Theta^* \mid \Theta^{(j)}]$ where $q[\cdot]$ is an arbitrary proposal distribution.

4. Generate $(\tilde{\underline{\mathbf{s}}}_{1:T}^*, \mathbf{A}_{1:T}^*) \sim \pi[\tilde{\underline{\mathbf{s}}}_0, \mathbf{A}_0; \Theta^*]$ where, $(\tilde{\underline{\mathbf{s}}}_0^*, \mathbf{A}_0^*) = (\tilde{\underline{\mathbf{s}}}_0, \mathbf{A}_0)$ and

$$\pi[\tilde{\underline{\mathbf{s}}}_0, \mathbf{A}_0; \Theta^*] = \prod_{t=0}^{T-1} \prod_{i=1}^{N} [\ \tilde{\mathbf{s}}_{i,t+1}^* \mid \tilde{\mathbf{s}}_{i,t}^*, \mathbf{A}_{i,t+1}^*, \Theta^*\ ][\ \mathbf{A}_{i,t+1}^* \mid \tilde{\mathbf{s}}_{i,t}^*, \Theta^*\ ]$$ (i.e., we are essentially simulating the movement path for all of the agents in this step).

5. Set $(\Theta^{(j+1)}, \tilde{\underline{\mathbf{s}}}_{1:T}^{(j+1)}, \mathbf{A}_{1:T}^{(j+1)}) = (\Theta^*, \tilde{\underline{\mathbf{s}}}_{1:T}^*, \mathbf{A}_{1:T}^*)$ with probability:

$$\alpha = \min\left(1, \frac{\pi_\epsilon[\ \underline{\mathbf{s}}_{1:T}^{(j+1)} \mid \tilde{\underline{\mathbf{s}}}_{1:T}^*, \mathbf{A}_{1:T}^*, \Theta^*\ ][\Theta^*]\ q[\Theta^{(j)} \mid \Theta^*]}{\pi_\epsilon[\ \underline{\mathbf{s}}_{1:T}^{(j+1)} \mid \tilde{\underline{\mathbf{s}}}_{1:T}^{(j)}, \mathbf{A}_{1:T}^{(j)}, \Theta^{(j)}\ ][\Theta^{(j)}]\ q[\Theta^* \mid \Theta^{(j)}]}\right).$$

Otherwise, $(\Theta^{(j+1)}, \tilde{\underline{\mathbf{s}}}_{1:T}^{(j+1)}, \mathbf{A}_{1:T}^{(j+1)}) = (\Theta^{(j)}, \tilde{\underline{\mathbf{s}}}_{1:T}^{(j)}, \mathbf{A}_{1:T}^{(j)})$.

In the ABC algorithm above, $\pi_\epsilon[\ \underline{\mathbf{s}}_{1:T}^{(j+1)} \mid \tilde{\underline{\mathbf{s}}}_{1:T}^*, \mathbf{A}_{1:T}^*, \Theta^{(j)}\ ]$ is the so-called uniform kernel defined as:

$$\pi_\epsilon\left[\ \underline{\mathbf{s}}_{1:T}^{(j+1)} \mid \tilde{\underline{\mathbf{s}}}_{1:T}^*, \mathbf{A}_{1:T}^*, \Theta^{(j)}\ \right] = \begin{cases} 1, & \text{if } d\left(m\left(\underline{\mathbf{s}}_{1:T}^{(j+1)}\right), m(\tilde{\underline{\mathbf{s}}}_{1:T}^*)\right) \leq \epsilon \\ 0, & \text{otherwise}, \end{cases}$$

where $\underline{\mathbf{s}}_{1:T}^{(j+1)}$ represents the current sampled latent locations. It should be noted that the MH ratio (i.e., α) above is considerably simplified since the state variables $\tilde{\underline{\mathbf{s}}}_{1:T}$ and $\mathbf{A}_{1:T}$ have proposal distributions that are equal to their respective prior distributions. Perhaps it is most useful to think of $\tilde{\underline{\mathbf{s}}}_{1:T}$ and $\mathbf{A}_{1:T}$ as auxiliary variables as described in Sisson and Fan (2011). Furthermore, despite the relatively high dimensionality of $\tilde{\underline{\mathbf{s}}}_{1:T}$ and $\mathbf{A}_{1:T}$, we have found that with appropriately chosen summary statistics (see Sect. 4.3 for a discussion of the

particular summary statistics used here) one can still achieve reasonable acceptance rates. For ABC, reasonable acceptance rates can range from 1% (e.g., Fearnhead and Prangle 2012) to 8.5% (e.g., Picchini 2014) depending on the application. Importantly, the ABC algorithm presented above is able to jointly sample the state variable and parameters, thereby reducing some of the computational cost associated with hierarchical ABMs.

4. HIERARCHICAL SPATIO-TEMPORAL SPP MODEL

In this section, we provide details based on the general movement model described in Sect. 2 (using similar notation as Sect. 2), along with a description of the guppy application and details of our implementation. The model presented in this section will be referred to as the hierarchical SPP (HiSPP) model.

4.1. GUPPY APPLICATION

Before detailing the proposed model, we start by describing the motivating application of modeling the collective movement of guppies (*Poecilia reticulate*). In particular, the HiSPP model is applied to a subset of the experimental guppy movement data from Bode et al. (2012). Detailed in Bode et al. (2012), to provide shelter for the experiment, one corner of the guppy tank was covered with gravel and cardboard. Shown in Fig. 1a, the guppies were
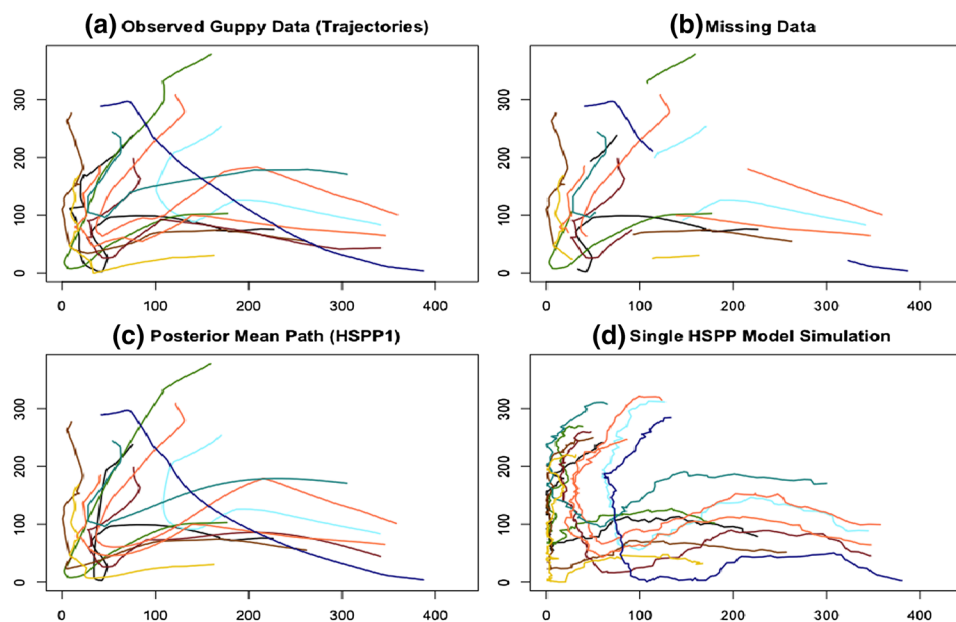

Figure 1. Summary of the data and posterior results for the guppy data with summary statistic HiSPP1. **a** Observed guppy movement paths, where each color represents a separate guppy. Note, the *top left corner* was covered with gravel and cardboard to provide shelter. **b** Plot of the guppy data with approximately 35% of the data removed and considered missing. **c** Posterior mean path calculated using HiSPP1. **d** Example of a single simulation from the HiSPP model using a set of parameters sampled from the posterior.

released from the opposite corner (i.e., the bottom right corner) of the sheltered area (top left corner). The movement of the guppies was captured with a standard definition camera recording at 10 frames per second. The observed guppy paths displayed in Fig. 1a provide realistic movement with clear collective behavior (Bode et al. 2012).

4.2. Hierarchical Spatio-Temporal SPP Model

The data model for the HiSPP model is the same as the data model in Sect. 2 (i.e., equation (1)), with the further assumption that $\mathbf{R}_{z,t} = \sigma_\epsilon^2 \mathbf{I}$. Now, suppose that the latent agent locations, $\mathbf{s}_{i,t}$, are confined to a bounded region such that $\tilde{x}_{i,t} \in [0, x_m]$ and $\tilde{y}_{i,t} \in [0, y_m]$. The process model is specified by:

$$\mathbf{s}_{i,t} \sim \text{TN}_{[0,x_m] \times [0,y_m]}(\, f(\mathbf{s}_{i,t-1}; \Theta_f), \sigma_\eta^2 \mathbf{I}\,). \tag{5}$$

A truncated normal (i.e., TN) distribution is utilized to accommodate the bounded locations. Further, using the notation from Sect. 2 for the speed and directional components of an agent's velocity vector, we have the following process model:

$$f(\mathbf{s}_{i,t-1}; \Theta_f) = (\tilde{x}_{i,t-1} + u_{i,t}\delta_{x,i,t}, \ \tilde{y}_{i,t-1} + u_{i,t}\delta_{y,i,t})', \tag{6}$$

if $\tilde{x}_{i,t-1} + u_{i,t}\delta_{x,i,t} \in [0, x_m]$ and $\tilde{y}_{i,t-1} + u_{i,t}\delta_{y,i,t} \in [0, y_m]$, implying that the updated locations are within the bounded region. For the cases where the updated locations are located outside of the boundary, we give the specific update equations in "Appendix A." We note that when the updated locations stray outside the bounded region, the parameters, Θ_f, are employed to enforce reflective conditions (i.e., the agents are reflected off the boundary back into the region of interest). Other types of boundary conditions may be more useful for other applications; see Hubbard et al. (2004) for further examples. For the HiSPP model $\Theta_f = (x_B, y_B)'$, where x_B and y_B define the angle of inflection for the x and y axis, respectively (see "Appendix A" for further details).

Selecting an appropriate model for the angle, and thus the direction of an agent, is instrumental in modeling an agent's movement. With regard to the HiSPP model, we use the following model to update a given agent's angle:

$$\theta_{i,t} = \arctan\left(\frac{\frac{1}{|\mathcal{N}_{i,t-1}^r|} \sum_{j \in \mathcal{N}_{i,t-1}^r} \sin(\theta_{j,t-1}) + \lambda_1 \tilde{x}_{i,t-1} + \lambda_2 \tilde{x}_{i,t-1}^2}{\frac{1}{|\mathcal{N}_{i,t-1}^r|} \sum_{j \in \mathcal{N}_{i,t-1}^r} \cos(\theta_{j,t-1})}\right) + \gamma_{i,t}, \ \gamma_{i,t} \sim \text{U}(-\tau/2, \tau/2),$$
$$\tag{7}$$

where $\mathcal{N}_{i,t-1}^r$ denotes the neighborhood of influence for an agent defined as $\mathcal{N}_{i,t-1}^r = \{j \in \{1, \dots, N\} : ||\mathbf{s}_{i,t-1} - \mathbf{s}_{j,t-1}|| < r\}$. The parameter r denotes the radius in which an agent's angle can be influenced, thus defining an agent's neighborhood of influence. While the radius is assumed to be the same for all agents here, with additional information one could allow the radius to vary by modeling it hierarchically (with covariates or other models). In essence, an agent's angle is updated as the average angle of its neighbors, while also considering relevant environmental covariates (i.e., $\tilde{x}_{i,t-1}$ and $\tilde{x}_{i,t-1}^2$). The covariates in (7)

are necessary to model the portion of an agent's movement that is not accounted for by their interaction with their neighbors (i.e., see Fig. 1a above). Other specifications of (7), such as including the direction of the refuge (with respect to an agent's current location) to bias the guppies toward the refuge, could also be useful here. The angular noise of an agent is assumed to be uniformly distributed over a bounded region, as is typical in the SPP literature.

The speed of an agent in the HiSPP model is specified using covariates as follows:

$$u_{i,t} = \max(\beta_0 + \beta_1 \tilde{x}_{i,t-1}, 0) + \phi_{i,t}, \quad \phi_{i,t} \sim \mathrm{U}(0, \rho). \tag{8}$$

Although for this particular application, we found that much of the variation in an agent's speed is explained simply by its x-coordinate, for other applications alternative covariates are likely necessary. Note, (8) could also take on alternative forms by using a log-link or truncated normal distribution. In the application to follow, we found that accounting for the nonnegative support of an agent's speed with a uniform distribution resulted in strong mixing with the hybrid MCMC–ABC algorithm. For the sake of brevity, we specify the prior distributions for all of the hyperparameters in the HiSPP model in "Appendix B."

4.3. MODEL SETUP

Here we briefly discuss details for the implementation of the hybrid MCMC–ABC algorithm for the HiSPP model. As previously mentioned, we implement the HiSPP model with two different sets of summary statistics. We will refer to the first set of summary statistics as HiSPP1 and the second as HiSPP2. The identity function (i.e., $m(\mathbf{s}_{i,t}) = \mathbf{s}_{i,t}$) is employed for the first set of summary statistics. For summary statistic HiSPP2, the temporal dimension of the data is reduced by thinning the data and only retaining every kth time period for each agent. More importantly, at every kth period the following four statistics are calculated for each agent: speed, directional angle, x-coordinate, and y-coordinate. Hence, for every kth period, $4N$ summary statistics are calculated (for both applications $k = 9$). The HiSPP2 summary statistic is very general while also capturing important information for describing an agent's movement, thus making it applicable for a variety of animal movement applications.

With regard to the distance metric necessary for the ABC algorithm described above, we use the following metric for both sets of summary statistics:

$$d(\cdot) = \Big(\sum_{i=1}^{N} \sum_{\ell \in \mathcal{L}} \frac{|m(\mathbf{s}_{i,\ell}) - m(\tilde{\underline{\mathbf{s}}}_{i,\ell})|}{\sigma_{m_{i,\ell}}} \Big)^{\frac{1}{2}}, \tag{9}$$

where $\mathcal{L}$ denotes the set of time indexes (which differ for HiSPP1 and HiSPP2) and $\sigma_{m_{i,\ell}}$ denotes a known scale parameter to ensure the statistics are weighted approximately equally (i.e., see Picchini and Forman 2016). Specifically, for HiSPP1, each $\sigma_{m_{i,\ell}} = 1$ and for HiSPP2, $\sigma_{m_{i,\ell}} = 7.75$ and $\sigma_{m_{i,\ell}} = 6$ for the speed/direction statistics, respectively, while $\sigma_{m_{i,\ell}} = 120$ and $\sigma_{m_{i,\ell}} = 150$ for the x/y-coordinates, respectively. To our knowledge, and as discussed in Fasiolo et al. (2016), there is currently no formal way of selecting these

weights for complicated hierarchical models. Hence, we use a series of simulations from the HiSPP model, to chose these particular values such that the respective summary statistics get approximately equal weight. Due to the small number of agents in both applications, we found that using a distance metric based on absolute deviations recovers the movement paths more accurately (based on mean-squared error) than a metric based on squared deviations. Finally, the tolerance parameter for the ABC algorithm is set so that the two sets of summary statistics have approximately equal acceptance rates (between 6 and 6.5% for both applications). Within a range of 5–8% (acceptance rates), the model was not overly sensitive to the tolerance level. These acceptance rates are similar to those found using a hierarchical model in Picchini (2014), which contains a more thorough discussion of acceptance rates for ABC.

For each implementation, the hybrid MCMC–ABC algorithm is run for 350,000 iterations with the first 50,000 iterations treated as burn-in. As is typical in the ABC literature, convergence of all chains was assessed through visual inspection of the posterior trace plots and no evidence of non-convergence was detected. The same priors (see "Appendix B") were applied for both the simulation study and the guppy application. A detailed discussion of the priors used for both applications can be found in supplementary materials (S.5).

5. RESULTS

5.1. SIMULATION STUDY RESULTS

To evaluate the performance of the HiSPP model in terms of both parameter estimation and prediction, we simulated movement paths for 10 agents from the HiSPP model using the parameter values listed in the second column of Table 1. We choose these particular parameters values to simulate trajectory paths that were similar to the overall movement pattern in the guppy application. For the sake of comparison, the number of agents (i.e., $N = 10$) and number of time periods (i.e., $T = 105$) were set equal to the corresponding values from the guppy application. Figure 2a shows the simulated paths for the simulation study (i.e., the "true" paths). By design, there is clearly a level of interaction between the agents in the simulated paths. As displayed in Fig. 2b, approximately 35% of the data were removed and considered missing. Frair et al. (2004) noted that 30% represents the upper bound for the amount of missing data in the telemetry data literature. Therefore, we removed approximately 35% of the data to represent a worst-case scenario for a realistic telemetry dataset. Additionally, for each agent, the same size block of data in time (i.e., multiple consecutive periods equaling 35% of an agent's complete path) was randomly removed (i.e., opposed to randomly removing individual observations). By removing observations, we are also able to investigate the effect of unobserved individuals on the observed agents (for certain time periods). Both summary statistics described in Sect. 4.3 were applied to the simulated dataset using the settings specified in Sect. 4.3.

Despite missing approximately 35% of the data, all of the true parameter values are contained within the 95% credible intervals (C.I.s) for both methods (with the exception of σ_ϵ^2 for HiSPP2). Overall, the two summary statistics provide similar posterior estimates for most of the parameters, with the radius parameter being the clear exception. It is likely that,

Table 1. Posterior summaries for the parameters in the simulation study with $N = 10$ agents and $T = 105$ time periods.

Parameter	Truth	HiSPP1			HiSPP2		
		Mean (median)	2.5%	97.5%	Mean (median)	2.5%	97.5%
σ_ϵ^2	0.75	1.582 (1.504)	0.6814	2.964	1.582 (1.509)	0.7591	2.769
σ_η^2	5	4.951 (4.523)	2.307	9.897	4.342 (4.127)	2.116	7.786
r	90	199.9 (180.9)	77.45	367.7	142.4 (114.4)	62.62	330.2
β_0	3.5	3.274 (3.293)	1.122	5.322	3.056 (3.118)	1.11	4.785
β_1	0.01	0.009308 (0.009139)	0.001893	0.01752	0.009537 (0.009354)	0.003004	0.01699
λ_1	−0.0013	−0.001594 (−0.001531)	−0.003731	4.721e−05	−0.001143 (−0.001109)	−0.002277	−0.0002437
λ_2	4.6e−06	5.765e−06 (5.501e−06)	−5.803e−07	1.414e−05	4.084e−06 (3.948e−06)	6.979e−07	8.257e−06
x_B	1.15	1.137 (1.206)	0.3416	1.556	1.123 (1.177)	0.4159	1.556
y_B	2.4	2.368 (2.381)	1.617	3.094	2.354 (2.346)	1.605	3.105
τ	0.15	0.2567 (0.2197)	0.009837	0.6769	0.2097 (0.1713)	0.007795	0.6096
ρ	1.5	2.775 (2.677)	0.1502	5.737	2.58 (2.37)	0.1341	5.741

Both summary statistics HiSPP1 and HiSPP2 were evaluated on the same simulated dataset. The mean (median) column denotes the posterior mean (median), while the 2.5 and 97.5% columns represent the lower and upper bound, respectively, for the 95% posterior credible intervals.
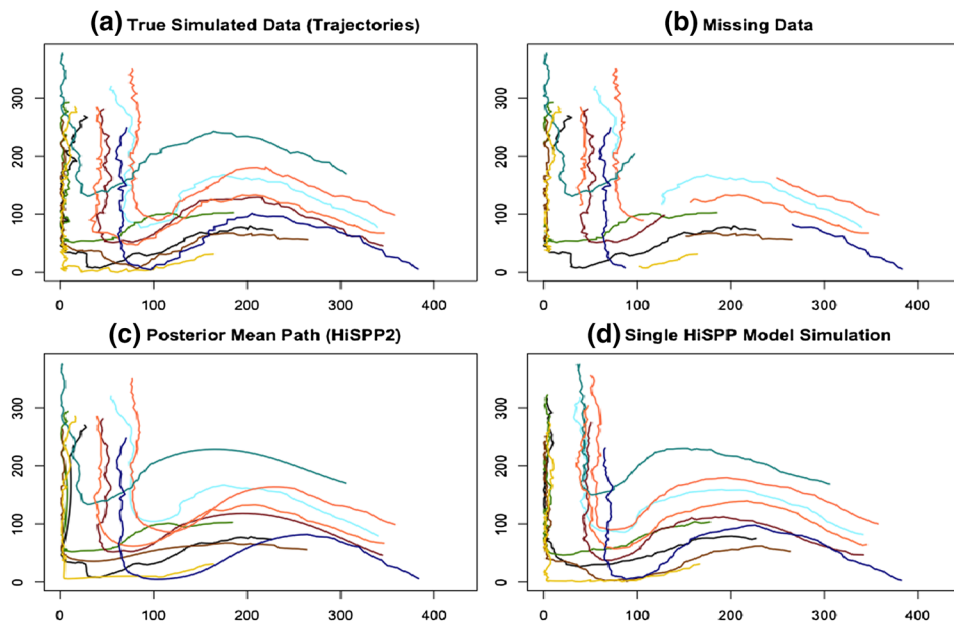
Figure 2. Summary of the data and posterior results for the simulated data with summary statistic HiSPP2. **a** True simulated data produced from the HiSPP model with the parameter values from Table 1. **b** Plot of the simulated data with approximately 35% of the data removed and considered missing. **c** Posterior mean path calculated using HiSPP2. **d** Example of a single simulation from the HiSPP model using a set of parameters sampled from the posterior.

with only 10 agents, it was difficult to estimate the magnitude of interaction among agents. With more data, the model would be able to provide more accurate estimates of the radius parameter. We should also note there appears to be less learning (i.e., wide C.I.s relative to the priors) for some of the parameters such as the reflection parameters (i.e., x_B and y_B). This may be attributed to the few number of agents that approached the boundary (as shown in Fig. 2a). Evaluation of the posterior mean paths was carried out using mean-squared error (MSE), defined here as the average squared distance between the observed paths and the posterior mean paths for all agent locations at each time period. Table 2 shows how HiSPP2 out-performed HiSPP1 in terms of filling in the missing data with HiSPP2 having a lower MSE for the missing observations. Further, the posterior mean path shown in Fig. 2c visually illustrates the predictive power of HiSPP2, with the predicted paths for the missing samples appearing similar to the true paths. Results for two additional simulation studies can be found in supplementary materials (S.4). For one of the additional simulation studies, we set both λ_1 and λ_2 in (7) to zero. The purpose of this simulation was to investigate possible confounding effects between the covariates in (7) and the radius parameter. The posterior mean of both λ_1 and λ_2 are centered around zero (along with the C.I.'s), suggesting the model was able to separate the effects of the radius and the covariates. For the other additional simulation study, we replaced the covariates in (7) and (8) with covariates based on an agent's distance from a refuge. The results for the second additional simulation study suggested that we were once again able to recovery many of the parameters and much of the movement paths.

Table 2. Mean-squared error (MSE) for both the simulated data and guppy data with summary statistic HiSPP1 and HiSPP2.

Data	HiSPP1		HiSPP2		Relative MSE	
	Overall MSE	Missing Obs. MSE	Overall MSE	Missing Obs. MSE	Overall	Missing Obs.
Simulation study	80.312	240.103	50.04	149.191	1.605	1.609
Guppy data	60.483	181.233	61.125	183.108	0.989	0.990

In order to evaluate the quality of the posterior paths for the missing data, the MSE was calculated using all of the data (i.e., Overall MSE) and with only the missing observations (i.e., Missing Obs. MSE). Relative MSE is defined as the MSE for summary statistic HiSPP1 divided by the MSE for summary statistic HiSPP2.

5.2. GUPPY APPLICATION RESULTS

To demonstrate the ability of the model to predict the latent collective movement process, the guppy application was implemented by removing approximately 35% of the data (in the same manner as the simulation study) and treating the removed observations as missing (i.e., see Fig. 1b). For our implementation, we utilized a subset of the data analyzed in Bode et al. (2012) consisting of 10 guppies and 105 time periods. Posterior inference for both summary statistics (see supplementary materials) suggested an interaction among the guppies (since the posterior mean for the radius parameter is over half the height of the guppy tank). Similar results regarding the interaction between the guppies in these data were found in Russell et al. (2016). Furthermore, both directional angle parameters (i.e., λ_1 and λ_2) have nonzero values, implying a common nonlinear direction pattern for the observed guppies. Overall, the two summary statistics provided similar results in terms of inference for all of the parameters, with a slight disagreement over the magnitude of the radius parameter. For the sake of brevity, the posterior results not shown here are given in supplementary materials (S.3).

With regard to recovering the missing guppy data, Table 2 shows how HiSPP1 and HiSPP2 performed similarly in terms of prediction accuracy. Moreover, the posterior mean paths in Fig. 1c appear similar to the observed paths, with visible differences only apparent for a few of the missing observations. Along with filling in the missing data, we are also interested in quantifying the uncertainty of the posterior paths, with particular interest in the uncertainty for the missing data. The posterior 95% C.I.s and mean paths for each of the 10 guppies are displayed in Fig. 3. As is common in spatial prediction, the level of uncertainty is much higher for the missing observations compared to the known observations. The majority of the true values for the missing observations are contained within the 95% intervals. In general, the missing observations with wider intervals correspond to observations with either unique individual movements or sharp turns, while the missing observations that correspond to more collective smooth movements have more narrow intervals.

6. DISCUSSION

The proposed hierarchical animal movement framework and the HiSPP model were very effective at recovering the movement paths for both the simulated collective movement data
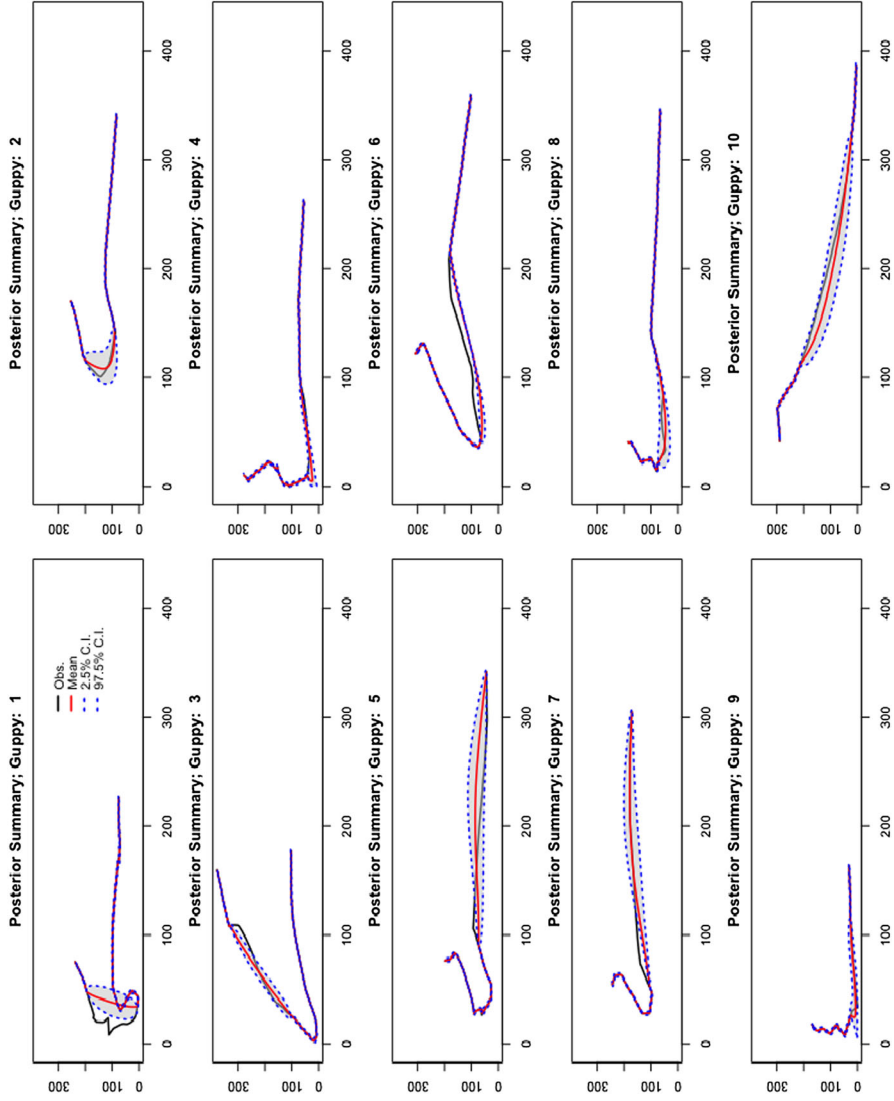
Figure 3. Individual posterior summaries for each of the 10 guppies in the guppy application using summary statistic HiSPP1. The *solid red line* in each *plot* denotes the posterior mean, while the *dashed blue lines* represent the 95% posterior credible intervals for each guppy. The observed paths are denoted by the *solid black lines* (Color figure online).

and the guppy data. Although a substantial amount of data were missing in the simulation study, the model was able to recover the parameters and much of the missing movement paths correctly. Through the use of Bayesian hierarchical modeling, we rigorously quantified the uncertainty associated with missing data from a realistic movement application of guppies. The model's overall ability to recover the guppy trajectories demonstrates the potential for utilizing ABC, even within a hierarchical framework.

Further, the developed ABC algorithm, along with the hybrid MCMC–ABC algorithm, provide an alternative to existing state-space estimation algorithms. From a scientific perspective the computational mechanisms that drive ABC is very intuitive, especially for simulation models such as complex collective animal movement models. The literature on approximate likelihood methods is currently growing at a rapid pace, ABC is just one of many available approximate methods (see Drovandi et al. 2015, for an overview). Along with the overall increase in approximation methods, there has also been an increase in approximate likelihood methods designed specifically for state-space models (e.g., Ehrlich et al. 2015; Fasiolo et al. 2016; Martin et al. 2016, to name a few). The unique challenges associated with modeling collective movement will likely require researchers to consider such approximate methods in the future.

While the presented model succeeded in estimating the few covariates included in the model, there is a great potential to include a variety of environmental covariates in future applications. For example, if considering terrestrial collective animal movement, information involving spatially referenced habitat features or weather patterns could easily be incorporated into the presented framework. Inclusion of such covariates would improve our understanding of interactions between collective animal movements and important covariates which might accentuate or mitigate species interactions. Such insights could help address potential impacts from current conservation challenges, such as climate change and shifting species distributions (Van der Putten et al. 2010). We also note that the generality of the model implies that a variety of other animals could be modeled with our methodology.

Although the presented model is formulated in discrete time, as described in McClintock et al. (2014), there are important consequences with regard to using continuous or discrete time when modeling animal movements. Indeed, Russell et al. (2016) described some advantages of using a continuous time collective movement model with the guppy data from Bode et al. (2012). While SPP models are usually formulated in discrete time, Yates et al. (2010) described various alternative SPP models formulated in continuous time. Thus, modifying the proposed hierarchical agent-based framework for continuous time would be an obvious extension of the model.

Many realistic collective behaviors may be difficult or impossible to model with standard statistical models, due to the nonlinear nature of these processes. Going forward, borrowing elements from mechanistic collective movement models and associated ideas from seemingly disparate subject matter sciences such as physics, chemistry, and engineering will be important when solving complex collective animal behavior problems. We note that many collective movement models, such as the SPP model, still have several *a priori* assumptions regarding collective behavior. Future applications may need to include mechanisms for learning complex behaviors that are not specified *a priori*.

APPENDIX A: FULL HISPP PROCESS MODEL EQUATION

The process model (i.e., Eq. 5) from Sect. 4.2 contains various update rules based on the values of the updated locations. In particular, if one or both of the updated x and y locations are outside the boundary of interest, then separate update conditions are invoked. These separate update conditions ensure that the updated locations are within the boundary of interest. We use reflective conditions to push the agents back toward the region of interest. The direction in which the agents reflect back into the region of interest is estimated for the cases where an agent's updated location is beyond the x or y boundary. If an agent's updated location is outside both the x and y boundary, then the reflective angle is set to a fixed value (this scenario was not common for our application). Recall that the latent locations, $\mathbf{s}_{i,t}$, are confined to a bounded region such that $\tilde{x}_{i,t} \in [0, x_m]$ and $\tilde{y}_{i,t} \in [0, y_m]$. The full process model is then given by the following:

$$f(\mathbf{s}_{i,t-1}; \Theta) = \begin{cases} (\tilde{x}_{i,t-1} + u_{i,t}\delta_{x,i,t}, \ \tilde{y}_{i,t-1} + u_{i,t}\delta_{y,i,t})', & \text{if } \tilde{x}_{i,t-1} + u_{i,t}\delta_{x,i,t} \in [0, x_m], \\ & \tilde{y}_{i,t-1} + u_{i,t}\delta_{y,i,t} \in [0, y_m] \\ (\tilde{x}_{i,t-1} + x_B\delta_{x,i,t}, \ \tilde{y}_{i,t-1} + u_{i,t}\delta_{y,i,t})', & \text{if } \tilde{x}_{i,t-1} + u_{i,t}\delta_{x,i,t} < 0 \\ (\tilde{x}_{i,t-1} + u_{i,t}\delta_{x,i,t}, \ \tilde{y}_{i,t-1} + u_{i,t}y_B)', & \text{if } \tilde{y}_{i,t-1} + u_{i,t}\delta_{y,i,t} < 0 \\ \left(\tilde{x}_{i,t-1} + u_{i,t}\left(\frac{\pi}{2} + x_B\right), \ \tilde{y}_{i,t-1} + u_{i,t}\delta_{y,i,t}\right)', & \text{if } \tilde{x}_{i,t-1} + u_{i,t}\delta_{x,i,t} > x_m \\ \left(\tilde{x}_{i,t-1} + u_{i,t}\delta_{x,i,t}, \ \tilde{y}_{i,t-1} + u_{i,t}\left(\frac{\pi}{2} + y_B\right)\right)', & \text{if } \tilde{y}_{i,t-1} + u_{i,t}\delta_{y,i,t} > y_m \\ \left(\tilde{x}_{i,t-1} + u_{i,t}\frac{5\pi}{4}, \ \tilde{y}_{i,t-1} + u_{i,t}\frac{5\pi}{4}\right)', & \text{if } \tilde{x}_{i,t-1} + u_{i,t}\delta_{x,i,t} > x_m, \\ & \tilde{y}_{i,t-1} + u_{i,t}\delta_{y,i,t} > y_m \\ \left(\tilde{x}_{i,t-1} + u_{i,t}\frac{3\pi}{4}, \ \tilde{y}_{i,t-1} + u_{i,t}\frac{3\pi}{4}\right)', & \text{if } \tilde{x}_{i,t-1} + u_{i,t}\delta_{x,i,t} > x_m, \\ & \tilde{y}_{i,t-1} + u_{i,t}\delta_{y,i,t} < 0 \\ \left(\tilde{x}_{i,t-1} + u_{i,t}\frac{7\pi}{4}, \ \tilde{y}_{i,t-1} + u_{i,t}\frac{7\pi}{4}\right)', & \text{if } \tilde{x}_{i,t-1} + u_{i,t}\delta_{x,i,t} < 0, \\ & \tilde{y}_{i,t-1} + u_{i,t}\delta_{y,i,t} > y_m \\ \left(\tilde{x}_{i,t-1} + u_{i,t}\frac{\pi}{4}, \ \tilde{y}_{i,t-1} + u_{i,t}\frac{\pi}{4}\right)', & \text{if } \tilde{x}_{i,t-1} + u_{i,t}\delta_{x,i,t} < 0, \\ & \tilde{y}_{i,t-1} + u_{i,t}\delta_{y,i,t} < 0. \end{cases}$$

APPENDIX B: SPECIFICATION OF PRIORS

Listed below are all of the prior distributions and corresponding hyperparameter values for the HiSPP model from Sect. 4.2.

$\sigma_\epsilon^2 \sim \text{IG}(\alpha_\epsilon, \beta_\epsilon)$, where $\alpha_\epsilon = .001$ and $\beta_\epsilon = .001$.

$\sigma_\eta^2 \sim \text{IG}(\alpha_\eta, \beta_\eta)$, where $\alpha_\eta = 4.25$ and $\beta_\eta = 24.375$.

$r \sim \text{TN}_{[0,r_{\max}]}(\mu_r, \sigma_r^2)$, where $\mu_r = 0$, $\sigma_r^2 = 100000$, and $r_{\max} = 378.4$.

$\beta_0 \sim \text{Gau}(\mu_\beta, \sigma_\beta^2)$ and $\beta_1 \sim \text{Gau}(\mu_\beta, \sigma_\beta^2)$, where $\mu_\beta = 0$ and $\sigma_\beta^2 = 1000$.

$\lambda_1 \sim \text{Gau}(\mu_\lambda, \sigma_\lambda^2)$ and $\lambda_2 \sim \text{Gau}(\mu_\lambda, \sigma_\lambda^2)$, where $\mu_\lambda = 0$ and $\sigma_\lambda^2 = 1000$.

$x_B \sim \text{U}(a_x, b_x)$, where $a_x = 0$ and $b_x = \frac{\pi}{2}$.

$y_B \sim \text{U}(a_y, b_y)$, where $a_y = \frac{\pi}{2}$ and $b_y = \pi$.

$\tau \sim \text{U}(a_\tau, b_\tau)$, where $a_\tau = 0$ and $b_\tau = .75$.

$\rho \sim \text{U}(a_\rho, b_\rho)$, where $a_\rho = 0$ and $b_\rho = 6$.

Since many of the hyperparameters above are utilized at the bottom of the hierarchical model, they are given relatively uninformative prior distributions. The hyperparameters for the boundary conditions (i.e., x_B and y_B) were selected such that agents would always be reflected back into the region of interest. See supplementary materials (S.5) for a further discussion on the priors used with the HiSPP model.

REFERENCES

Aeschbacher, S., Beaumont, M. A., and Futschik, A. (2012), "A novel approach for choosing summary statistics in approximate Bayesian computation," *Genetics*, 192, 1027–1047.

Beaumont, M. A. et al. (2010), "Approximate Bayesian computation in evolution and ecology," *Annual Review of Ecology, Evolution and Systematics*, 41.

Berliner, L. M. (1996), "Hierarchical Bayesian time series models," in *Maximum entropy and Bayesian methods*, Springer, vol. 79, pp. 15–22.

Bode, N. W., Franks, D. W., Wood, A. J., Piercy, J. J., Croft, D. P., and Codling, E. A. (2012), "Distinguishing social from nonsocial navigation in moving animal groups," *The American Naturalist*, 179, 621–632.

Bonnell, T. R., Henzi, S. P., and Barrett, L. (2016), "Direction matching for sparse movement data sets: determining interaction rules in social groups," *Behavioral Ecology*.

Cressie, N. and Wikle, C. (2011), *Statistics for Spatio-Temporal Data*, New York: John Wiley & Sons.

Dalziel, B. D., Corre, M. L., Côté, S. D., and Ellner, S. P. (2016), "Detecting collective behaviour in animal relocation data, with application to migrating caribou," *Methods in Ecology and Evolution*, 7, 30–41.

Drovandi, C. C., Pettitt, A. N., and Lee, A. (2015), "Bayesian indirect inference using a parametric auxiliary model," *Statistical Science*, 30, 72–95.

Ehrlich, E., Jasra, A., and Kantas, N. (2015), "Gradient free parameter estimation for hidden Markov models with intractable likelihoods," *Methodology and Computing in Applied Probability*, 17, 315–349.

Fasiolo, M., Pya, N., Wood, S. N., et al. (2016), "A Comparison of Inferential Methods for Highly Nonlinear State Space Models in Ecology and Epidemiology," *Statistical Science*, 31, 96–118.

Fearnhead, P. and Prangle, D. (2012), "Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74, 419–474.

Frair, J. L., Nielsen, S. E., Merrill, E. H., Lele, S. R., Boyce, M. S., Munro, R. H., Stenhouse, G. B., and Beyer, H. L. (2004), "Removing GPS collar bias in habitat selection studies," *Journal of Applied Ecology*, 41, 201–212.

Gautrais, J., Jost, C., and Theraulaz, G. (2008), "Key behavioural factors in a self-organised fish school model," in *Annales Zoologici Fennici*, pp. 415–428.

Hooten, M., Johnson, D., McClintock, B., and Morales, J. (2017), *Animal Movement: Statistical Models for Telemetry Data*, Chapman & Hall/CRC Boca Raton, Florida, USA.

Hooten, M. B. and Wikle, C. K. (2010), "Statistical agent-based models for discrete spatio-temporal systems," *Journal of the American Statistical Association*, 105, 236–248.

Hubbard, S., Babak, P., Sigurdsson, S. T., and Magnússon, K. G. (2004), "A model of the formation of fish schools and migrations of fish," *Ecological Modelling*, 174, 359–374.

Lagarrigues, G., Jabot, F., Lafond, V., and Courbaud, B. (2015), "Approximate Bayesian computation to recalibrate individual-based models with population data: illustration with a forest simulation model," *Ecological Modelling*, 306, 278–286.

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

Mann, R. P. (2011), "Bayesian inference for identifying interaction rules in moving animal groups," *PLoS One*, 6.

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012), "Approximate Bayesian computational methods," *Statistics and Computing*, 22, 1167–1180.

Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003), "Markov chain Monte Carlo without likelihoods," *Proceedings of the National Academy of Sciences*, 100, 15324–15328.

Martin, G. M., McCabe, B. P., Frazier, D. T., Maneesoonthorn, W., and Robert, C. P. (2016), "Auxiliary Likelihood-Based Approximate Bayesian Computation in State Space Models," arXiv preprint arXiv:1604.07949.

McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M., and Morales, J. M. (2014), "When to be discrete: the importance of time formulation in understanding animal movement," *Movement Ecology*, 2.

Millspaugh, J., Kesler, D., Kays, R., Gitzen, R., Schulz, J., Rota, C., Bodinof, C., Belant, J., and Keller, B. (2012), "Wildlife radiotelemetry and remote monitoring," *The Wildlife Techniques Manual*, 1, 258–283.

Nott, D. J., Marshall, L., and Ngoc, T. M. (2012), "The ensemble Kalman filter is an ABC algorithm," *Statistics and Computing*, 22, 1273–1276.

Pereyra, M., Dobigeon, N., Batatia, H., and Tourneret, J.-Y. (2013), "Estimating the granularity coefficient of a Potts-Markov random field within a Markov chain Monte Carlo algorithm," *IEEE Transactions on Image Processing*, 22, 2385–2397.

Perna, A., Grégoire, G., and Mann, R. P. (2014), "On the duality between interaction responses and mutual positions in flocking and schooling," *Movement Ecology*, 2.

Picchini, U. (2014), "Inference for SDE models via approximate Bayesian computation," *Journal of Computational and Graphical Statistics*, 23, 1080–1100.

Picchini, U. and Forman, J. L. (2016), "Accelerating inference for diffusions observed with measurement error and large sample sizes using approximate Bayesian computation," *Journal of Statistical Computation and Simulation*, 86, 195–213.

Russell, J. C., Hanks, E. M., and Haran, M. (2016), "Dynamic Models of Animal Movement with Spatial Point Process Interactions," *Journal of Agricultural, Biological, and Environmental Statistics*, 21, 22–40.

Scharf, H. R., Hooten, M. B., Fosdick, B. K., Johnson, D. S., London, J. M., and Durban, J. W. (2017), "Dynamic social networks based on movement," *The Annals of Applied Statistics*, 10, 2182–2202.

Sisson, S. and Fan, Y. (2011), "Likelihood-free MCMC," *Handbook of Markov Chain Monte Carlo*, 313–335.

Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. (1997), "Inferring coalescence times from DNA sequence data," *Genetics*, 145, 505–518.

Van der Putten, W. H., Macel, M., and Visser, M. E. (2010), "Predicting species distribution and abundance responses to climate change: why it is essential to include biotic interactions across trophic levels," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 2025–2034.

Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., and Shochet, O. (1995), "Novel type of phase transition in a system of self-driven particles," *Physical review letters*, 75:1226–1229

Wolfram, S. (1983), "Statistical mechanics of cellular automata," *Reviews of modern physics*, 55, 601.

Yates, C. A., Baker, R. E., Erban, R., and Maini, P. K. (2010), "Refining self-propelled particle models for collective behaviour," *Canadian Applied Mathematics Quarterly*, 18, 299–350.