WILEY

Process convolution approaches for modeling interacting trajectories

Henry R. Scharf[1] | Mevin B. Hooten[1,2] | Devin S. Johnson[3] | John W. Durban[4]

[1]Department of Statistics, Colorado State University, Fort Collins, CO, U.S.A.

[2]U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO, U.S.A.

[3]Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, U.S.A.

[4]Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, La Jolla, CA, U.S.A.

Correspondence
Henry R. Scharf, Department of Statistics, Colorado State University, Fort Collins, CO, U.S.A.
Email: henry.scharf@colostate.edu

Gaussian processes are a fundamental statistical tool used in a wide range of applications. In the spatiotemporal setting, several families of covariance functions exist to accommodate a wide variety of dependence structures arising in different applications. These parametric families can be restrictive and are insufficient in some situations. In contrast, process convolutions represent a flexible, interpretable approach to defining the covariance of a Gaussian process and have modest requirements to ensure validity. We introduce a generalization of the process convolution approach that employs multiple convolutions sequentially to form a "process convolution chain". In our proposed multistage framework, complex dependencies that arise from a combination of different interacting mechanisms are decomposed into a series of interpretable kernel smoothers. We demonstrate an application of process convolution chains to model killer whale movement, in which the paths taken by multiple individuals are not independent but reflect dynamic social interactions within the population. Our proposed model for dependent movement provides inference for the latent dynamic social structure in the study population. Additionally, by leveraging the positive dependence among individual paths, we achieve a reduction in uncertainty for the estimated locations of the killer whales, compared with a model that treats paths as independent.

KEYWORDS

animal movement, kernel convolution, moving average, process convolution, social network

1 | INTRODUCTION

Models for continuous random processes using kernel convolutions are known in much of the spatial and spatiotemporal statistical literature as "process convolutions" (e.g., Bolin & Lindgren, 2013; Calder, 2007; Higdon, 2002). Also referred to as spatial moving averages (e.g., Cressie & Pavlicová, 2002), process convolutions first arose as a means for constructing valid covariance matrices for Gaussian processes (GPs) while relaxing the typical assumptions of stationarity and isotropy (e.g., Barry & Ver Hoef, 1996). One application area for process convolutions that has driven much of their recent theoretical development is for random spatial processes on stream networks (Ver Hoef, Peterson, & Theobald, 2006). Stream networks occupy submanifolds of physical space on which specialized measures of distance often complicate enforcement of nonnegative definiteness in covariance functions and have an inherent need for anisotropic dependence structures to accommodate the effects of direction in stream flow. Process convolutions offer an approach to modeling dependence that addresses both of these issues. We employ process convolutions to model the dependence observed in the trajectories of interacting animals.

A wealth of data arising from measurements of trajectories has spurred the development of many new statistical models for individual-based movement processes. These models have been used to study the individual movement patterns of a wide variety of animals, where scientific questions of interest are often focused on which exogenous environmental factors drive movement (Potts, Bastille-Rousseau, Murray, Schaefer, & Lewis, 2014). A fundamental commonality of research related to animal movement is a need for accurate and precise knowledge about the paths traversed by each individual, leading to the development of many models used for reconstructing the true underlying movement processes from telemetry data (e.g., Fleming et al., 2016). Several excellent models for telemetry data have been proposed in both discrete- and continuous-time formulations that have sought to address the significant challenges making the prediction of the true paths difficult (for a review, see Hooten & Johnson, 2017; McClintock, Johnson, Hooten, Ver Hoef, & Morales, 2014). Telemetry devices are often subject to environmental conditions and technological limitations that restrict their ability to deliver precise measurements of location at dense temporal resolutions. Thus, many analyses of animal movement must account for both measurement error and irregular temporal observations when estimating the true path traversed by an individual. In some cases, it may be reasonable to assume that the data include independent Gaussian noise; however, for many telemetry devices, it is necessary to account for complex, non-Gaussian forms of measurement error (Brost, Hooten, Hanks, & Small, 2015; Buderman, Hooten, Ivan, & Shenk, 2016). Irregularity in the frequency of telemetry observations can complicate the implementation of discrete-time models (although for one solution using a multiple-imputation procedure, see Scharf, Hooten, & Johnson, 2017) and potentially lead to large uncertainty during time intervals with few or no observations.

Ecologists are also interested in the extent to which individuals in a population move in direct response to each other. Particularly in the case of large mammals, complex and dynamic social connections within a population can play a critical role in the movement behavior of individuals (e.g., Goldenberg, de Silva, Rasmussen, Douglas- Hamilton, & Wittemyer, 2014; Scharf et al., 2016; Williams & Lusseau, 2006). For example, two individuals with a strong social connection may exhibit similar movement patterns and visit the same locations, resulting in positive dependence between their respective paths. Inference about the social ties within a population, and how those ties change over time, can provide valuable information about the behavioral ecology of a population. However, there are few existing methods available that explicitly account for interactions among individuals (e.g., Codling & Bode, 2014; Haydon et al., 2008; Langrock et al., 2014; Russell, Hanks, & Haran, 2016; Scharf et al., 2016). Especially when direct observation of a species is infeasible, careful analysis of data gathered from telemetry devices can reveal useful information about a population's social structure (Scharf et al., 2016).

We propose a joint model for the movement of multiple individuals that expands the suite of statistical methodology available for studying animal movement. Our approach allows researchers to study dependence among the paths of a population of individuals that arises as the result of social interactions. Also, by taking into account the dependence among individuals in a population, our approach has the potential to reduce both bias and uncertainty in reconstructing trajectories in the study population compared with models that treat individuals independently. We obtain improvements in path reconstruction by modeling the true movement processes of the population conditioned on an unobserved, dynamic social network. The inferred network provides a description of the social ties within a population and how those ties change in time.

In what follows, we present a generalized approach to process convolutions in which a random process is decomposed into a sequence of one or more smoothing kernels that are convolved with a white noise process. Multiple stages of smoothing allow us to model dependence in time and between individuals separately, resulting in a multivariate GP that captures the combination of effects. It is important to note that, while the kernels responsible for inducing temporal and pathwise dependence are constructed separately, the resulting covariance function is not separable in the geostatistical sense.

We provide a detailed introduction to our generalized approach to process convolutions in Sections 2.1 and 2.2 and construct our joint model for movement in Sections 2.3 and 2.4. After discussing the implementation details in Section 3, we demonstrate path reconstruction in a simulation study in Section 4. We apply our model to the study of telemetry data arising from the movement of killer whales in Section 5 and close with a summary of our findings and future directions.

2 | METHODS

2.1 | A multiple-kernel convolution framework

We first outline a new flexible framework for the development of GP models based on a synthesis of ideas from geostatistics, multivariate time series, and trajectory modeling. Our hierarchical framework relies on the kernel convolution

approach for modeling random processes, known in the spatial and spatiotemporal statistical literature as "process convolutions" (e.g., Bolin & Lindgren, 2013; Calder, 2007; Higdon, 2002). A mean-zero random process $\mu(\cdot)$ is called a process convolution if it is constructed by convolving a continuous random process, $dB(\cdot)$, with a kernel function, h, over a domain, $\mathcal{T}$, so that

$$\mu(t) = \int_{\mathcal{T}} h(t, \tau) dB(\tau). \tag{1}$$

If the random process $dB(\cdot)$ is Gaussian, the resulting process $\mu(\cdot)$ will also be Gaussian with covariance function, as follows:

$$\text{Cov}(\mu(t), \mu(t^*)) = \int_{\mathcal{T}} \int_{\mathcal{T}} h(t, \tau) h(t^*, \tau^*) \text{Cov}(dB(\tau), dB(\tau^*)). \tag{2}$$

It is common to define the process $dB(\cdot)$ to be Gaussian white noise, which yields the simplified covariance function, as follows:

$$\text{Cov}(\mu(t), \mu(t^*)) = \int_{\mathcal{T}} h(t, \tau) h(t^*, \tau) d\tau, \tag{3}$$

(although for an example of a spatial process convolution where $dB(\cdot)$ has covariance specified through a Gaussian Markov random field, see Nychka, Bandyopadhyay, Hammerling, Lindgren, & Sain, 2015). A process convolution with kernel function $h(\cdot, \cdot)$, as in (1), represents a smoothing of the process $dB(\cdot)$, and the kernel is therefore often referred to as a smoother. In what follows, we refer to h as both a kernel function and a smoother, interchangeably.

The function defined in (3) is guaranteed to be nonnegative definite and therefore a valid covariance function, if $\int_{\mathcal{T}} h(t, \tau) d\tau < \infty$ and $\int_{\mathcal{T}} h^2(t, \tau) d\tau < \infty$ for all t (Higdon, 2002). In many applications, it may be easier to specify the proper form of dependence in a GP through the form of a kernel smoother rather than that of a covariance function. For example, by constructing process convolutions with asymmetrical kernels, Ver Hoef et al. (2006) accounted for directional dependence in spatial processes arising on stream networks, where directional flow plays a crucial role. For certain choices of h, it is possible to express the covariance function analytically (e.g., Higdon, 2002; Paciorek, 2003; Ver Hoef & Peterson, 2010), although this is not necessary for nonnegative definiteness. In the cases where an analytic solution to (3) is not available, one can evaluate the integral numerically. The class of covariance functions constructed using kernel convolutions is general, containing many of the parametric families commonly used in geostatistical settings (e.g., exponential, Gaussian, and spherical), as well as more flexible dependence structures. For example, Higdon (1998) and Paciorek (2003) showed that process convolutions allow researchers to model both anisotropic and nonstationary dependence by letting $h(t, \tau)$ vary with both $|t - \tau|$ and t. In principle, any covariance function evaluated over a discretized domain has a process convolution representation on the same discretized grid, where the kernel function's values over the grid may be defined using a decomposition of the covariance matrix (e.g., Cholesky).

Hooten and Johnson (2017) used process convolutions to develop new models for trajectories by convolving a Wiener process, rather than white noise, with different kernel functions. Using Wiener process convolutions, they specified realistic models for animal movement without the characteristic "roughness" found in Brownian motion and defined a framework that incorporated other existing models for movement, such as that in Johnson, London, Lea, and Durban (2008). Brownian motion itself can be thought of as a process convolution where the kernel is a step function given by $h^{(bm)}(t, \tau) \equiv \mathbf{1}_{\{\tau \leq t\}}$ and the process is Gaussian white noise. Thus, the Hooten and Johnson (2017) framework can be viewed as a nested smoothing procedure, where the position at time t is as follows:

$$\mu(t) = \int_{\mathcal{T}} h(t, \tau) \left(\int_{\mathcal{T}} h^{(bm)}(\tau, \tilde{\tau}) dB(\tilde{\tau}) \right) d\tau. \tag{4}$$

To account for Brownian motion with an initial position far from the origin, one may define the process $dB(\cdot)$ to be a continuous white noise process for all $t > 0$, with arbitrary initial variance so that $dB(0) \sim \mathcal{N}(0, \sigma_0^2)$. Hooten and Johnson (2017) provided several examples of possible kernel functions and demonstrated a computationally efficient procedure to incorporate temporal nonstationarity into the kernel, allowing them to fit highly flexible models to telemetry data arising from animal movement trajectories.

It is possible to write the two-stage process convolution in (4) in terms of a single effective smoothing kernel $\tilde{h}(t, \tilde{\tau}) = \int_{\mathcal{T}} h(t, \tau) h^{(bm)}(\tau, \tilde{\tau}) d\tau$ using Fubini's theorem to change the order of integration. Moreover, the equivalent, single-stage representation for $\mu(\cdot)$ is not limited to two-stage smoothing processes. One may specify any number of kernel functions, and the resulting multistage kernel convolution can be written in terms of a single effective kernel. This nested structure

motivates a novel model-building approach in which an arbitrarily large collection of convolution kernels is used to specify a valid GP. We refer to a stochastic process resulting from the iterative convolution of a chain of kernels $h^{(1)}, \ldots, h^{(L)}$ as a "process convolution chain" (PCC) and write it in its expanded form as follows:

$$\mu^{(1)}(\tau_1) = \int_{\mathcal{T}} h^{(1)}(\tau_1, \tau_0) dB(\tau_0)$$

$$\mu^{(2)}(\tau_2) = \int_{\mathcal{T}} h^{(2)}(\tau_2, \tau_1) \mu^{(1)}(\tau_1) d\tau_1$$

$$\vdots \tag{5}$$

$$\underline{\mu^{(L)}}(\tau_l) = \int_{\mathcal{T}} h^{(L)}(\tau_L, \tau_{L-1}) \underline{\mu^{(L-1)}}(\tau_{L-1}) d\tau_{L-1}.$$

The underline in stages $2 \leq l \leq L$ (i.e., $\underline{\mu^{(l)}}(\tau_l)$) distinguishes the PCC composed of all kernels $\{h^{(k)} : k \leq l\}$ from that of convolving $h^{(l)}$ with the process $dB(\cdot)$ directly (written $\mu^{(l)}(\tau_l)$). We use matching superscripts to denote the covariance function for a given process (e.g., $\text{Cov}(\underline{\mu^{(l)}}(t), \underline{\mu^{(l)}}(t^*)) = \underline{C^{(l)}}(t, t^*)$). A collapsed version of a PCC using a single effective kernel and the white noise process $dB(\cdot)$ can be written as follows:

$$\underline{\mu^{(l)}}(\tau_l) = \int_{\mathcal{T}} \underline{h^{(l)}}(\tau_l, \tau_0) dB(\tau_0) \tag{6}$$

if we define the effective kernel as follows:

$$\underline{h^{(l)}}(\tau_l, \tau_0) \equiv \int_{\mathcal{T}} \cdots \int_{\mathcal{T}} h^{(l)}(\tau_l, \tau_{l-1}) \ldots h^{(1)}(\tau_1, \tau_0) d\tau_1 \ldots d\tau_{l-1}. \tag{7}$$

To simplify notation, we write $\tilde{h} = \underline{h^{(L)}}$ for the effective kernel constructed from the entire chain, $\tilde{\mu}$ for the full process $\underline{\mu^{(L)}}$, and $\widetilde{C}$ for the corresponding covariance function. In general, the kernels $h^{(k)}$ are not commutative, in the sense that

$$\int_{\mathcal{T}} \int_{\mathcal{T}} h^{(l_1)}(t, \tau) h^{(l_2)}(\tau, \tau_0) dB(\tau_0) d\tau \neq \int_{\mathcal{T}} \int_{\mathcal{T}} h^{(l_2)}(t, \tau) h^{(l_1)}(\tau, \tau_0) dB(\tau_0) d\tau. \tag{8}$$

Thus, both the forms of the kernels and their order are important when specifying a PCC.

Specifying a GP through an ordered chain of smoothers allows for considerable flexibility. For example, if we specify a chain of size $L = 1$ with kernel function $h^{(1)}(t, \tau_1) \equiv \mathbf{1}_{\{\tau_1 < t\}}$, we recover simple Brownian motion. If we increase the length of the chain by including a second kernel, $h^{(2)}(t, \tau_2) = (1 + \frac{\tau_2 - t}{\phi}) \mathbf{1}_{\{-\phi < \tau_2 - t \leq 0\}}$, we recover a structure used by Ver Hoef and Peterson (2010) for modeling dependence in stream networks (also mentioned in Hooten & Johnson, 2017; Figure 2, row 3). Another specific case of PCCs are the so-called $(k-1)$-fold integrated Wiener processes (Rue & Held, 2005; Shepp, 1966; Wecker & Ansley, 1983) that use $L = k$ kernels of the form $h^{(l)}(t, \tau) = \mathbf{1}_{\{\tau < t\}}$ and allow one to model GPs with exactly $k-1$ continuous derivatives. The multistage decomposition is similar in spirit to the way random processes with multiple dependence scales are decomposed additively in multiresolution processes (Higdon, 2002; Katzfuss, 2017; Nychka et al., 2015); the difference is that the convolution chain approach combines model components using convolution rather than addition.

In what follows, we construct convolution kernels that allow us to specify multivariate GPs; these provide a powerful tool for modeling multiple trajectories arising from interacting individuals. We return to this specific application of PCCs in Section 2.3.

2.2 | Finite representation

The joint distribution of $\tilde{\boldsymbol{\mu}} \equiv (\tilde{\mu}(t_1), \ldots, \tilde{\mu}(t_n))'$ for $\mathbf{t} \equiv (t_1, \ldots, t_n)' \subseteq \mathcal{T}$ is mean-zero Gaussian with covariance function $\widetilde{C}(t, t^*) = \int_{\mathcal{T}} \tilde{h}(t, \tau) \tilde{h}(t^*, \tau) d\tau$ for any $t, t^* \in \mathbf{t}$ (if the integrals of $\tilde{h}$ and $\tilde{h}^2$ exist). When this integral cannot be computed analytically, one can perform the integration numerically by selecting a fine grid of values $\boldsymbol{\tau} \equiv (\tau_0, \ldots, \tau_m)'$ from the domain $\mathcal{T}$ and by evaluating the discrete sum $\sum_{k=1}^{m} \tilde{h}(t, \tau_k) \tilde{h}(t^*, \tau_k) \Delta \tau_k$, where $\Delta \tau_k = \tau_k - \tau_{k-1}$. In the discretized setting, the process convolution may be written as a matrix product $\tilde{\boldsymbol{\mu}} = \widetilde{\mathbf{H}} \boldsymbol{\varepsilon}_0$, where the jth row of $\widetilde{\mathbf{H}}_{n \times m}$ is the function $h(t_j, \tau)$ evaluated at all $\tau \in \boldsymbol{\tau}$, and $\boldsymbol{\varepsilon}_0 \equiv (\varepsilon_0(\tau_1), \ldots, \varepsilon_0(\tau_m))'$ with each $\varepsilon_0(\tau_k) \sim \mathcal{N}(0, \Delta \tau_k)$. Assuming the grid times $\boldsymbol{\tau}$ are equally

spaced with intervals of size $\Delta\tau$, the joint distribution of $\tilde{\boldsymbol{\mu}}$ can be expressed using matrix notation as follows:

$$\tilde{\boldsymbol{\mu}} \sim \mathcal{N}\left(\mathbf{0}, \Delta\tau\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}'\right). \tag{9}$$

As the density of the grid grows to infinity ($m \to \infty$), the outer product in (9) approaches the covariance defined by the continuous integral over $\mathcal{T}$. The granularity of the grid (i.e., the size of m) required to adequately approximate the integral will generally depend on the characteristics of $\tilde{h}$. Alternatively, one can follow the approach of Higdon (2002), who used finite process convolutions (also called discrete process convolutions in Calder, 2008) as an approximation, and choose $m < n$ to yield a fixed-rank model for the continuous process. Fixed-rank models can offer considerable computational efficiency and provide an adjustable level of implicit regularization (Wikle, 2010).

2.3 | A PCC for dependent movement

2.3.1 | Social smoothing

We construct a novel model for animal movement using the PCC approach within a Bayesian hierarchical modeling framework. To describe the basic procedure, we first consider paths in one dimension. The model we describe can be readily extended to movement in two or more dimensions, and we demonstrate that in the example that follows. We use a three-stage ($L = 3$) PCC that includes a specialized kernel function constructed to induce dependence among the paths of p different individuals. The interpath dependence that arises is based on a weighted, undirected, and latent social network (Goldenberg, Zheng, Fienberg, & Airoldi, 2010).

Let the random variable $\mu_i(t)$ represent the position of individual i at time t, and let $w_{ij}(\tau) \in [0, 1]$ denote the connection weight between individuals i and j at time τ, with $w_{ii}(\tau) \equiv 1$ and $i, j \in \{1, \ldots, p\}$. We propose a "social" smoothing kernel of the following form:

$$h_{ij}^{(soc)}(\tau_{soc}, \tau) \equiv \mathbf{1}_{\{\tau=\tau_{soc}\}} \frac{w_{ij}(\tau_{soc})}{|w_{i\cdot}(\tau_{soc})|}, \tag{10}$$

$$|w_{i\cdot}(\tau_{soc})| \equiv \sum_{j=1}^{p} w_{ij}(\tau_{soc}). \tag{11}$$

We write the kernel in (10) with explicit dependence on arbitrary time τ for completeness but will hereafter suppress the second argument for brevity. To provide intuition for the effect of convolving with the social smoothing kernel, we first consider the case of a multivariate process with two individuals. We assume that, for the process $\boldsymbol{\mu}(\cdot) \equiv (\mu_1(\cdot), \mu_2(\cdot))'$, the variables $\mu_1(\tau)$ and $\mu_2(\tau^*)$ are independent *a priori* for all τ and τ^*. After social smoothing, the resulting process for individual 1 is defined by the following:

$$\mu_1^{(soc)}(\tau_{soc}) \equiv \sum_{j=1}^{2} \int_{\mathcal{T}} h_{1j}^{(soc)}(\tau_{soc}, \tau)\mu_j(\tau)d\tau \tag{12}$$

$$= h_{11}(\tau_{soc})\mu_1(\tau_{soc}) + h_{12}(\tau_{soc})\mu_2(\tau_{soc}) \tag{13}$$

$$= \frac{\mu_1(\tau_{soc})}{1 + |w_{12}(\tau_{soc})|} + \frac{w_{12}(\tau_{soc})\mu_2(\tau_{soc})}{1 + |w_{12}(\tau_{soc})|}. \tag{14}$$

From (13) and (14), it is clear that $\mu_1^{(soc)}(\tau_{soc})$ is a weighted average of the independent variables $\mu_1(\tau_{soc})$ and $\mu_2(\tau_{soc})$, where the weights are equal to $h_{11}(\tau_{soc})$ and $h_{12}(\tau_{soc})$, respectively. When the network connection is strong (i.e., $w_{12}(\tau_{soc}) \approx 1$), the two weights $h_{11}(\tau_{soc})$ and $h_{12}(\tau_{soc})$ will be approximately equal, and the effect of the kernel is to "squeeze" the previously independent paths toward their mutual mean. When the network connection is weak ($w_{12}(\tau_{soc}) \approx 0$), the weights will be close to 1 and 0, respectively, and the processes will be unaffected by the kernel, retaining the *a priori* independence between individuals. These effects are visible in the simulated process shown in the top plot of Figure 1.

The social smoothing kernel we propose is motivated by an interest in describing a mechanistic, interpretable driver of dependent movement. Specifically, it is useful for modeling the paths of individuals who have a tendency to move toward, and alongside, other individuals with whom they share connections. This type of dependence is ubiquitous in animal movement but is by no means the only meaningful type of interaction. As an alternative, consider the movement of two highly territorial animals with overlapping territories in which observed paths would appear to "avoid" each other. To model mutual avoidance, we require a social kernel that "spreads apart" two otherwise independent paths when a strong social connection is present. In what follows, we focus on the particular kernel defined in (16), but we emphasize that PCCs, with suitably constructed kernels, offer a flexible way to specify models for a wide range of possible behavioral
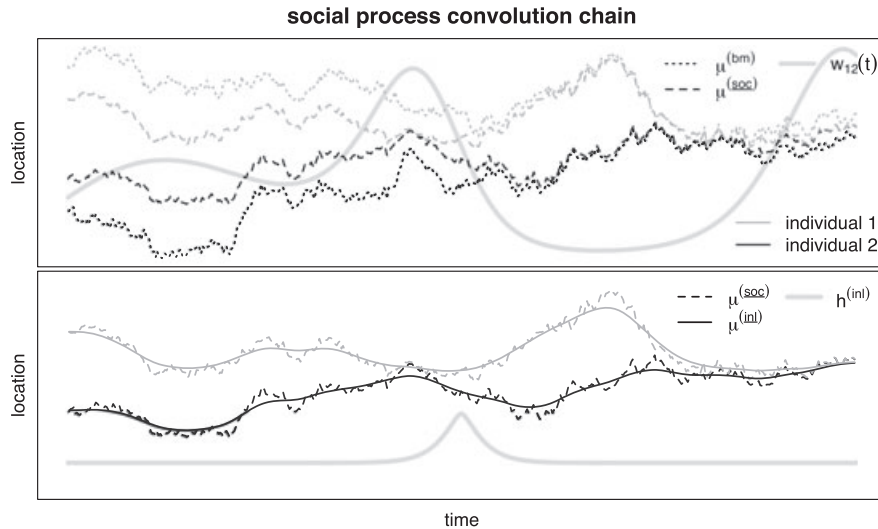
social process convolution chain

FIGURE 1 An example of a one-dimensional GP arising from a three-stage social PCC for two individuals (light and dark). All three intermediate stages of smoothing are shown, beginning with Brownian motion (top, dotted), followed by "social" smoothing (top and bottom, dashed), and finally, "inertial" smoothing (bottom, solid). The thick gray line in the top plot is proportional to the strength of the social tie $w_{12}(t)$ at all times t, where the extent of the y-axis corresponds to a range of $[0, 1]$. The thick gray line in the bottom plot is proportional to the inertial smoothing kernel $h^{(inl)}(t)$ for a fixed value of τ

mechanisms. Additionally, our choice of social smoothing kernel is most appropriate when social connections among individuals vary slowly, relative to the movement processes. When relationships among individuals are allowed to vary too rapidly, it may become difficult for the model to distinguish between brief encounters that arise due to social effects, and those that arise as part of the stochasticity inherent in Brownian motion.

Equipped with the social kernel given in (16), we outline the first two stages of our full three-stage PCC model for the dependent movement of p individuals. We begin with Gaussian white noise and smooth at the first stage by convolving $h_{ij}^{(bm)}(t, \tau) \equiv \mathbf{1}_{\{\tau \leq t\}} \mathbf{1}_{\{i=j\}}$ with $dB_i(\cdot)$, for each $i = (1, \dots, p)$, to generate p independent instances of Brownian motion, denoted $\mu_i^{(bm)}(\cdot)$, each with its own initial position. At the second stage, we apply the social smoothing kernel defined in (10) to the collection of all p processes, $\boldsymbol{\mu}^{(bm)}(t) \equiv (\mu_1(t), \dots, \mu_p(t))'$. Smoothing with the social kernel returns weighted averages of the Brownian processes, where the weights are proportional to the latent social weights $w_{ij}(t)$. Thus, two individuals i and j for which $w_{ij}(t)$ is close to 1 will tend to have smoothed locations $\mu_i^{(soc)}(t)$ and $\mu_j^{(soc)}(t)$ that are close together in space. A third and final stage in the PCC, introduced in the following section, ensures that the random process has the proper temporal smoothness required for modeling animal movement.

2.3.2 | Inertial smoothing

Marginally, the individual paths $\underline{\mu_i^{soc}}$ generated from the two-stage PCC constructed from $h^{(bm)}$ and $h^{(soc)}$ are each an instance of Brownian motion. However, Brownain motion is unsuitable for the direct modeling of animal movement because the instantaneous velocity of a particle traversing a Brownian path is discontinuous, and therefore, the acceleration is not well defined. Discontinuities in the first derivative of the path processes imply that individuals are capable of instantaneous changes in their velocities and are inconsistent with the physical laws governing the mechanics of massive bodies (e.g., Feynman, Leighton, & Sands, 1963).

Several process convolutions have been explored that impart specific smoothness properties on a GP. For example, Rue and Held (2005), Shepp (1966), and Wecker and Ansley (1983) discussed applications of $(k − 1)$-fold integrated Wiener processes, which ensure the existence of $k − 1$ continuous derivatives. Johnson et al. (2008) constructed a model for the movement of individual harbor and northern fur seals by specifying an Ornstein–Uhlenbeck process for the velocity, rather than the position. Hooten and Johnson (2017) modeled the true locations of an individual animal using a two-stage PCC in which the second kernel function is Gaussian, and Buderman et al. (2016) used a similar model with cubic splines as their kernels. In all of these examples, smoothness is imparted to a process through convolution with a kernel, either implicitly or explicitly. We introduce a final stage of smoothing to yield paths that are guaranteed to be continuously differentiable.

Enforcing the existence of a continuous derivative for the movement processes ensures that individuals have an interpretable acceleration at every time t. Therefore, we refer to the final kernel function as an "inertial" smoother because its purpose is to generate paths with the physical properties required to obey the laws of classical mechanics. We specify a kernel from the Matérn family of correlation functions (Cressie, 1991) with an unknown range parameter, ϕ_{inl}, and smoothness $\nu = 1$ as the inertial smoother. The effect of this final kernel, $h_{ij}^{(inl)}(t, \tau) \equiv \frac{|t-\tau|}{\phi_{inl}} K_1(|t - \tau|/\phi_{inl}) \mathbf{1}_{\{i=j\}}$, where $\mathbf{K}_1$ is a modified Bessel function, is visible in the bottom plot of Figure 1. As we did for the social smoothing kernel, in what follows, we suppress the superfluous indices and write $h^{(inl)}(\cdot, \cdot)$.

2.3.3 | The full three-stage PCC

The expanded specification of the full three-stage PCC is given by the following:

$$\mu_i^{(bm)}(\tau_{bm}) = \int_{\mathcal{T}} h^{(bm)}(\tau_{bm}, \tau_0) dB_i(\tau_0) \tag{15}$$

$$\mu_i^{(soc)}(\tau_{soc}) = \sum_{j=1}^{p} h_{ij}^{(soc)}(\tau_{soc}) \mu_j^{(bm)}(\tau_{soc}) \tag{16}$$

$$\tilde{\mu}_i(\tau_{inl}) = \mu_i^{(inl)}(\tau_{inl}) = \int_{\mathcal{T}} h^{(inl)}(\tau_{inl}, \tau_{soc}) \mu_i^{(soc)}(\tau_{soc}) d\tau_{soc}. \tag{17}$$

One of the primary benefits of using a PCC framework is that it allows one to decompose complex dependencies into an iterative sequence of relatively simple mechanisms. Positions of multiple individuals can, in general, exhibit dependencies in both time and among individuals, and the characteristics of these dependencies may be dynamic themselves. We construct a PCC using kernels that compartmentalize the mechanisms generating the complex dependencies in the collective movement of interacting individuals. The first stage of smoothing is in time, and not across individuals, so that $\mu_i^{(bm)}(\tau_{bm})$ and $\mu_j^{(bm)}(\tau_{bm})$ are independent for $i \neq j$. The second stage of smoothing is across individuals, and not in time, in the sense that $h_{ij}^{(soc)}(t, t^*) = 0$ for $t \neq t^*$. The final stage of smoothing is in time only.

Restricting individual smoothing components to operate in only one "dimension" at a time (in this case, either temporal or social) is not necessary. However, a primary motivation for constructing GP models in the PCC framework is to decompose a complex mechanism into components that are easier to understand separately. Therefore, it is natural to form PCC models with kernel functions that operate in a limited number of dimensions simultaneously.

Let $s_i(t)$ denote the observed position of individual $i \in \{1, \ldots, p\}$ at time $t \in \mathcal{T}$. We model the true, unobserved position of individual i at time t proportional to $\tilde{\mu}_i(t)$ and add a measurement error process, $\varepsilon_i(t)$, to yield the data model as follows:

$$s_i(t) \equiv \sigma_\mu \tilde{\mu}_i(t) + \varepsilon_i(t). \tag{18}$$

The parameter σ_μ scales the random process $\tilde{\mu}(\cdot)$ to account for the overall spatial domain of the locations. We model the measurement error $\varepsilon_i(t)$ as independent, identically distributed (i.i.d.) mean-zero Gaussian random variables with variance σ_s^2, although other measurement error processes could also be used (e.g., Brost et al., 2015; Buderman et al., 2016).

2.4 | Process prior for the dynamic social network

The general model for dependent movement is highly parameterized. To evaluate the covariance $\Delta \tau \widetilde{\mathbf{H}} \widetilde{\mathbf{H}}'$ for a grid of m times, one must estimate $\binom{p}{2} m$ parameters associated with the social network that drives the second stage of the PCC. As we describe in Section 3, we fit the proposed model to data within a Markov chain Monte Carlo paradigm, which requires repeated inversion of a large, dense covariance matrix. Without further structure for $w_{ij}(\cdot)$, estimating the underlying network would be computationally infeasible. We increase computational efficiency by leveraging reasonable assumptions about the underlying network to constrain the space of possible latent social networks, thereby reducing the effective number of parameters in the model. We impose constraints on the network's marginal complexity at each fixed time and the smoothness of its temporal evolution.

Our approach makes use of recently developed methods for modeling dynamic stochastic networks. To reduce the dimensionality of the latent network at a fixed time, we follow the approaches of Durante, Scarpa, and Dunson (2014); Hoff (2008); and Hoff, Raftery, and Handcock (2002), who modeled social connections using a latent space. Latent space models for stochastic networks proceed by defining network connections through functions of the locations of particles in a latent d-dimensional "social" space. As an example, consider the following latent space model for the edges in a weighted,

undirected, and dynamic social network. Suppose that the latent variables $\boldsymbol{\mu}_i^{(w)}(t)$, $\boldsymbol{\mu}_j^{(w)}(t)$, and $\boldsymbol{\mu}_k^{(w)}(t)$ are positions in $\mathbb{R}^2$, and edge weights are defined by $w_{ij}(t) = e^{-d_{ij}(t)}$, where $d_{ij}(t)$ is the Euclidean distance between $\boldsymbol{\mu}_i^{(w)}(t)$ and $\boldsymbol{\mu}_j^{(w)}(t)$. Defining edge weights based on the distance between particles in a latent space induces positive dependence among the connections $w_{ij}(t)$, because the latent distances are constrained by the characteristics of the latent space and the measure of distance. That is, if $d_{ij}(t)$ and $d_{jk}(t)$ are small, resulting in edge weights $w_{ij}(t)$ and $w_{jk}(t)$ that are close to 1, the triangle inequality implies that $d_{ik}(t)$ must also be small, and $w_{ik}(t)$ will also be close to 1. Therefore, latent space approaches for modeling social networks can be understood as a way to induce a tendency to complete triangles, or transitivity, in a network. This phenomenon is often observed in human social networks (e.g., friends of friends tend to themselves be friends; Hansell, 1984) and is also reasonable for many applications to animal social networks. A latent space approach to constraining the underlying social network has the computational advantage of reducing the number of parameters to estimate by a factor of p, because we only need to estimate p latent paths instead of $\binom{p}{2}$ pairwise relationships.

We define the latent social connection between individuals i and j at time τ_{soc} using the latent positions and an appropriately chosen functional g by the following:

$$w_{ij}(\tau_{soc}) \equiv g\left(\boldsymbol{\mu}_i^{(w)}(\tau_{soc}), \boldsymbol{\mu}_j^{(w)}(\tau_{soc})\right). \tag{19}$$

In general, the type of network desired for the application (e.g., binary and weighted) may inform the particular choice for g. As discussed in Section 2.3, we construct the kernel $h_{ij}^{(soc)}$ under the assumption that the weights in the social network have support $[0, 1]$. The compact support of $w_{ij}(\cdot)$ motivates our choice of g, defined by $g(\mathbf{x}, \mathbf{y}) \equiv e^{-\|\mathbf{x}-\mathbf{y}\|_2^2}$. The functional g maps two vectors in the latent space to the unit interval $[0, 1]$ and follows the methods used for latent space network modeling by Hoff (2008), who showed that this construction induces positive dependence in the edges $w_{ij}(\cdot)$.

$$\boldsymbol{\mu}_i^{(w)}(\tau_w) \equiv \sigma_w \int_{\mathcal{T}} h^{(w)}(\tau_w, \tau) dB_i^{(w)}(\tau), \tag{20}$$

$$h^{(w)}(\tau_w, \tau) = e^{\frac{-(\tau-\tau_w)^2}{\phi_w^2}}, \tag{21}$$

and $dB_i^{(w)}(\cdot)$ are instances of two-dimensional Brownian motion, now used to define the positions in the latent social space. The parameter σ_w controls the dispersion of the latent paths $\boldsymbol{\mu}_i^{(w)}$ and is therefore related to the overall density of the social network, with smaller σ_w corresponding to higher connectivity. The parameter ϕ_w^2 controls the tortuosity of the trajectories through the latent social space and is related to the temporal stability of the network over time, with larger values of ϕ_w^2 corresponding to a more stable network. As described in Section 2.2, we approximate the continuous stochastic processes $\boldsymbol{\mu}_i^{(w)}(\cdot)$ using a set of independent normal random variables anchored at a finite number of knots yielding the following:

$$\boldsymbol{\mu}_i^{(w)} \sim \mathcal{N}\left(\mathbf{0}, \sigma_w^2 \Delta t \widetilde{\mathbf{H}}_w' \widetilde{\mathbf{H}}_w\right). \tag{22}$$

The complete specification of the hierarchical model is provided in Appendix A.

3 | MODEL IMPLEMENTATION

We obtain realizations from the posterior distribution of the model parameters using a Markov chain Monte Carlo algorithm. In this section, we briefly discuss the most relevant features of model implementation. A more detailed description, including our choice of priors and hyperparameters, is available in Appendix B.

We specify independent normal priors for the initial-location parameters and a conjugate hyperprior on the population-level variance so that $d\mathbf{B}_i(0) \sim \mathcal{N}(\mathbf{0}, \sigma_0^2 \mathbf{I}_2)$ and $\sigma_0^2 \sim \mathrm{IG}(a_0, b_0)$. Additionally, we integrate the true location process $\tilde{\boldsymbol{\mu}}$ out of the likelihood, which allows us to avoid sampling the continuous movement process directly. This is a common technique used in spatial statistics (e.g., Finley, Banerjee, & Gelfand, 2013; Gelfand, Kim, Sirmans, & Banerjee, 2003) to improve mixing for the remaining parameters in the model. The underlying movement process $\tilde{\boldsymbol{\mu}}$ can be recovered post hoc using composition sampling (e.g. Finley et al., 2013). The resulting integrated model formulation is as follows:

$$\mathbf{s}_i(t)|\varepsilon_w, \phi_w, \sigma_w^2, \phi_{inl}, \sigma_s^2, \sigma_\mu^2, \sigma_0^2 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{23}$$

$$\boldsymbol{\Sigma} \equiv \sigma_s^2 \mathbf{I} + \sigma_\mu^2 \Delta\tau \widetilde{\mathbf{H}}' \widetilde{\mathbf{H}}, \tag{24}$$

where the parameters $\varepsilon_w, \phi_w, \sigma_w^2, \phi_{inl},$ and σ_0^2 enter the density through the definition of $\widetilde{\mathbf{H}}$ in (15)–(17). To update each latent space path, $\boldsymbol{\mu}_i^{(w)}$, we employ a Metropolis–Hastings algorithm.

4 | SIMULATION

As the dependence among individuals in a population weakens (i.e., $w_{ij}(t) \to 0$ for all t, $i \neq j$), our proposed model simplifies to one that considers each path independently but shares information about all nonsocial parameters across individuals. The PCC model constructed from a fixed, empty social network presents a natural baseline for comparison with our full model for dependent movement. Estimating the latent social network that gives rise to the dependence among paths represents the majority of the computational effort required to fit our proposed model to data and grows rapidly with the number of individuals under study. Thus, there is a natural incentive to model movement under the assumption of path independence unless it can be shown to be deficient. We compare the performances of the full model for dependent movement (IP-DEP) with the special case of the model under the assumption of interpath independence (IP-IND). In many cases, we find that accounting for the dependence among individuals results in a significantly improved reconstruction of the true underlying paths.

Telemetry devices are subject to a wide range of environmental conditions that frequently result in large intervals of time during which no observed locations are recorded. Estimates of the true locations, $\tilde{\boldsymbol{\mu}}$, during long gaps between observations are often accompanied by large amounts of uncertainty that present a challenge for researchers studying animal behavior. In simulation, we show that the presence of moderate to strong dependence among observed paths provides an opportunity for improvement in path reconstruction when we take into account the joint distribution of all observed individuals. To illustrate the potential gains, we consider a simple setting of two individuals in which the connection status, w_{12}, is constant over the period of observation. We assume regular, uninterrupted observations for individual 1, but a gap in time, $\mathcal{T}_g$, exists in the observed sequence of telemetry locations for individual 2.

We evaluate the quality of a given model for path reconstruction in terms of both accuracy and precision. To assess the accuracy of a path reconstruction, we define a loss function termed the "squared path error" (SPE) that quantifies the agreement between a specific path reconstruction $\hat{\boldsymbol{\mu}}$ and the true path $\tilde{\boldsymbol{\mu}}_{true}$ as follows:

$$\text{SPE}(\hat{\boldsymbol{\mu}}; \mathcal{T}_g, \tilde{\boldsymbol{\mu}}_{true}) = \frac{1}{|\mathcal{T}_g|} \int_{\mathcal{T}_g} \|\tilde{\boldsymbol{\mu}}_{true}(t_g) - \hat{\boldsymbol{\mu}}(t_g)\|_2^2 dt_g. \qquad (25)$$

We define the reconstruction $\hat{\boldsymbol{\mu}}$ to be the posterior mean for $\tilde{\boldsymbol{\mu}}$ so that (25) represents a measure of how accurately the center of the posterior distribution matches the true underlying path. To assess the precision associated with a model, we compute the radii for circular 95% credible regions surrounding each point in the path $\hat{\boldsymbol{\mu}}(t_g)$ and average across the entire temporal gap in observations, $\mathcal{T}_g$, to yield an overall summary of precision that we term the average circular 95% credible region radius (ACCRR$_{0.95}$).

To provide a general sense of when accounting for path dependence results in the greatest gains in path reconstruction, we vary the strength of dependence, w_{12}; the proportion of the study interval made up by the observation gap, $|\mathcal{T}_g|/|\mathcal{T}|$; and the tortuosity of the paths and use SPE in tandem with ACCRR$_{0.95}$ to compare the performances of IP-DEP and IP-IND. We vary tortuosity through the range parameter ϕ_{inl} that appears in the inertial smoother, with smaller values corresponding to more tortuous paths on average. For each combination of w_{12}, $\mathcal{T}_g$, and tortuosity, we used the following five-step procedure:

1. Simulate a realization of $\tilde{\boldsymbol{\mu}}_{true}$ and **s** from our proposed model (see Figure 2).
2. Fit both the dependent and independent path models to **s**.
3. Sample 1,000 paths from the posterior distribution of $\tilde{\boldsymbol{\mu}}$ using composition sampling.
4. Compute the SPE defined in (25) for the posterior mean of $\tilde{\boldsymbol{\mu}}_2$.
5. Use the same 1,000 draws from the posterior to compute the ACCRR$_{0.95}$ for $\tilde{\boldsymbol{\mu}}_2$.

We repeat the procedure 20 times for each combination of parameters to obtain an estimate of the variability among realizations of the simulated paths. Values for the parameters used to simulate each path and all prior distributions and hyper parameters are provided in Table 1, and an example simulation is shown in Figure 2.

To facilitate direct comparison between the two models under consideration, we examine the ratios of SPE and ACCRR$_{0.95}$ for IP-DEP and IP-IND, defining ratios such that the relevant value for IP-DEP appears in the denominator and that values greater than 1 show support for IP-DEP. Figure 3 displays the ratios of SPE (top row) and ACCRR$_{0.95}$ (bottom row) under all combinations of values for the parameters w_{12}, $|\mathcal{T}_g|/|\mathcal{T}|$, and ϕ_{inl} (tortuosity). Individual plots show the median ratio across all simulations for high (solid) and low (dashed) tortuosities, with an associated polygon
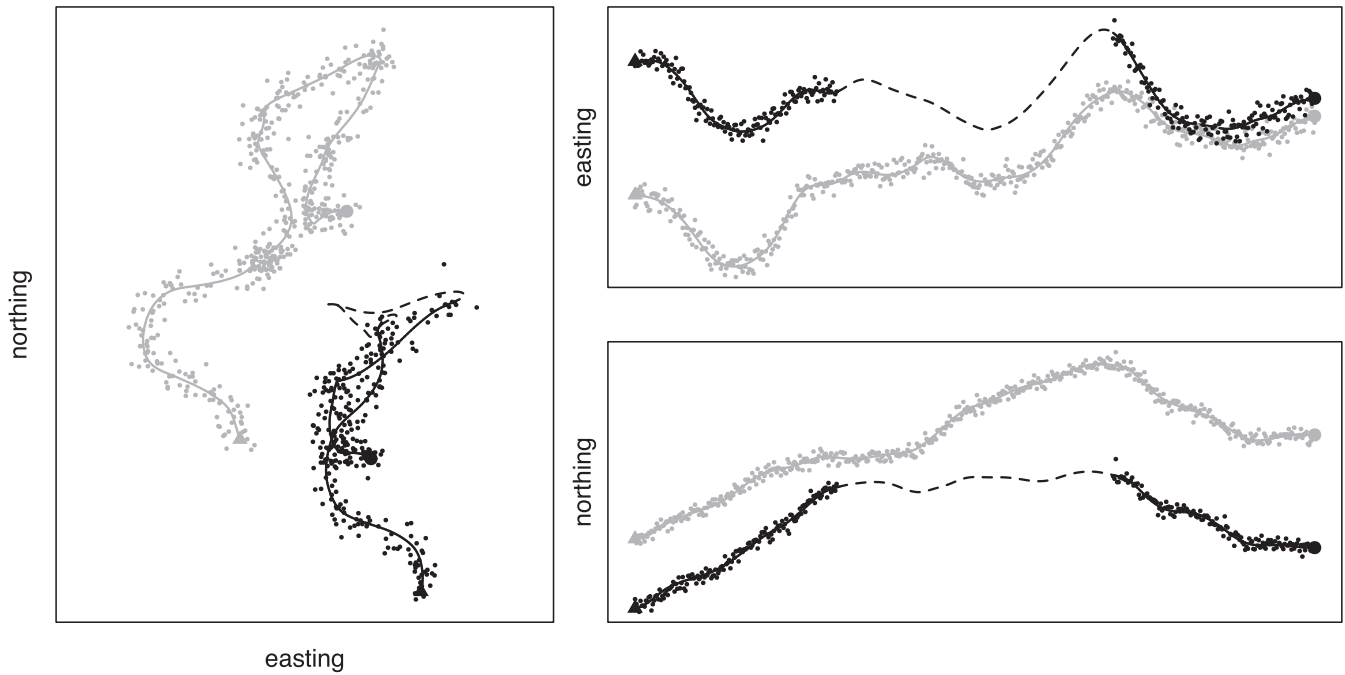
FIGURE 2 Example of a realization of two simulated true paths, $\tilde{\mu}_1$ and $\tilde{\mu}_2$ (gray and black lines, respectively), and observed locations, $\mathbf{s}_1$ and $\mathbf{s}_2$ (points) used in the simulation study. The dashed line represents the portion of the path taken by individual 2 where a gap in observation occurs. The figure corresponds to the case of moderate strength of tie ($w_{12} = 0.5$), large gap ($|\mathcal{T}_g|/|\mathcal{T}| = 0.4|$), and low tortuosity (see also Figure 3)

TABLE 1 True values and prior distibutions used in the simulation study

parameter	true	prior density
ϕ_{inl} ("low tortuosity")	0.04	Gamma(2, 100)
ϕ_{inl} ("high tortuosity")	0.04/3	Gamma(2, 100)
σ_0^2	1	IG($10^{-3}, 10^{-3}$)
$\sigma_{\mu/s}^2$	800	IG($10^{-3}, 10^{-3}$)
σ_μ^2	10	
σ_s^2	0.0125	IG($10^{-3}, 10^{-3}$)

delineating the 25% and 75% quantiles across simulations. Columns organize the plots by gap size, and the strength of the social connection w_{12} increases along the *x*-axis within each individual plot.

Several general observations can be made about the circumstances under which fitting the IP-DEP offers the greatest improvements in path reconstruction. First, both precision and accuracy improve near monotonically with increasing w_{12}. Second, the greatest gains come when the gap size is moderate to large (columns 3 and 4). When the gap is brief (column 1), the difference between the reconstructed paths for each model is modest. Finally, the gains in performance for the full model are greater for the case of high tortuosity. In our simulation study, more tortuous paths are characterized by a shorter range of dependence in time; thus, increasing tortuosity is similar to increasing the size of the observation gap.

5 | KILLER WHALES

We analyzed telemetry data for four killer whales near the Antarctic Peninsula (see Figure 4) over the course of 5 days in February 2014. Geographic positions were measured using Argos telemetry tags (for a complete description of the tags and study area, see Andrews, Pitman, & Ballance, 2008; Durban & Pitman, 2012). Although multiple types of killer whales have been described in this area, all four tags were deployed on individuals from the same population of the most common Type B2 killer whales (Durban, Fearnbach, Burrows, Ylitalo, & Pitman, 2016).
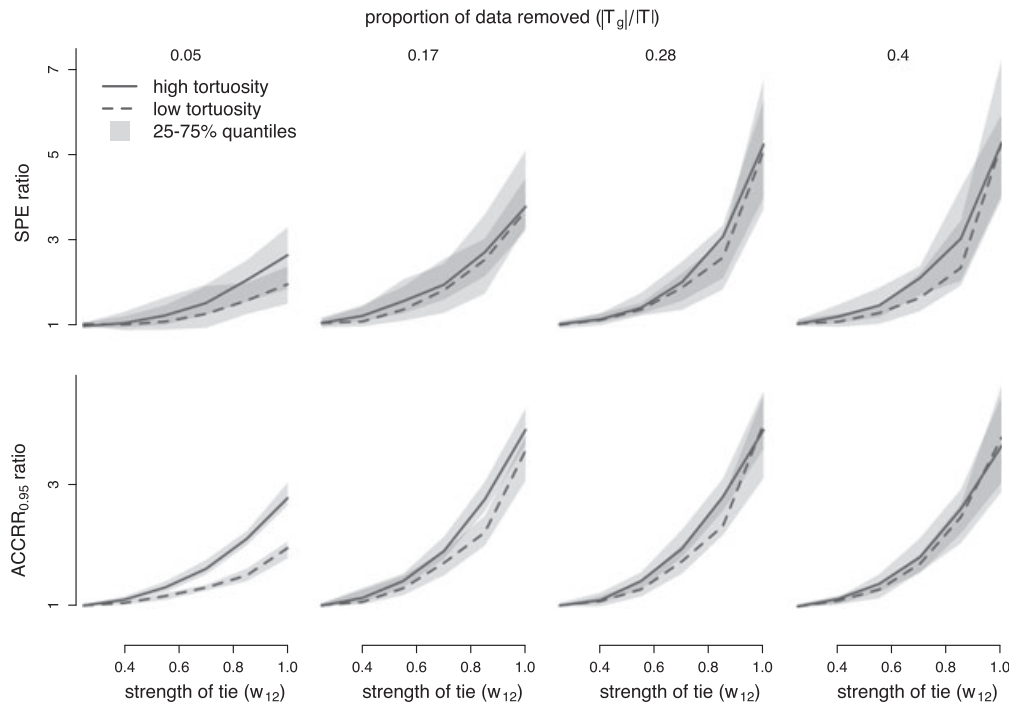
FIGURE 3 Path errors. The top row of plots shows the ratios of squared path error (SPE) for the independent to full models (thus, larger ratios correspond to the full model outperforming the alternative). The bottom row shows the analogous ratios of average circular 95% credible region radius (ACCRR$_{0.95}$). Each column represents a fixed value for the proportion of the study interval made up by the gap in observations ($|\mathcal{T}_g|/|\mathcal{T}|$). Within each plot, the strength of the social connection increases along the *x*-axis. The lines represent the median value of all ratios across the 20 simulations, and the associated polygons represent the 25% and 75% quantile boundaries. Finally, the solid lines correspond to simulations with high tortuosity in the true paths, whereas dashed lines correspond to simulations with low tortuosity

It is immediately apparent from the data that individuals 1 and 2 (Figure 4, bottom left) show potential evidence of a close connection. In addition, there is some ambiguity about the relationship between individuals 3 and 4 (Figure 4 top right), because they occupy approximately the same spatial region during the study period. In contrast, there is little reason to suspect dependence between the two pairs of individuals. We analyzed the movement of all four individuals jointly, which allowed for the pooling of information about measurement error across the entire group, and provided an opportunity for basic model validation. By fitting a model for the joint movement of all four individuals with a fully flexible latent network structure, we were able to check for the presence of potentially spurious network connections, because existing knowledge about the system suggests that the underlying network should exclude connections between the two subgroups.

The top plot of Figure 5 shows the observation times for the killer whales in our study, with darker regions corresponding to a denser rate of telemetry measurements in time. Two different daylong gaps in observation occur for individual 1 (Figure 5), a consequence of the original study design used in the deployment of the telemetry tags that sought to balance the need for temporally dense observations with limitations in the battery life of the tags by collecting measurement only every other day for select individuals. The observation times for the other three individuals cover these gaps, suggesting that modeling the four paths jointly may allow for more precise and/or accurate estimates of the true path taken by individual 1 if moderate to strong dependence exists between it and any of the other three whales.

As in the simulation study (Section 4), we investigated path reconstructions generated by our proposed IP-DEP, as well as the analogous model under an assumption of IP-IND. We found that the full model generates path reconstructions in which uncertainty about the position of individual 1 is dramatically reduced, compared with the reconstructions generated under the assumption of IP-IND as measured by 95% circular credible region radii (bottom plot of Figure 5). Circular credible regions for the true position of the individual during the two large gaps in observation occuring on February 12th and 14th for the IP-IND model have large radii, sometimes exceeding 70 km. In contrast, circular credible regions for the true position based on the IP-DEP model are less than a third this large and similar in magnitude to the uncertainties for individuals for which we have dense observations. Despite the uncertainty about the latent social network, the IP-DEP model offers a substantial reduction in uncertainty about the true path taken by individual 1.
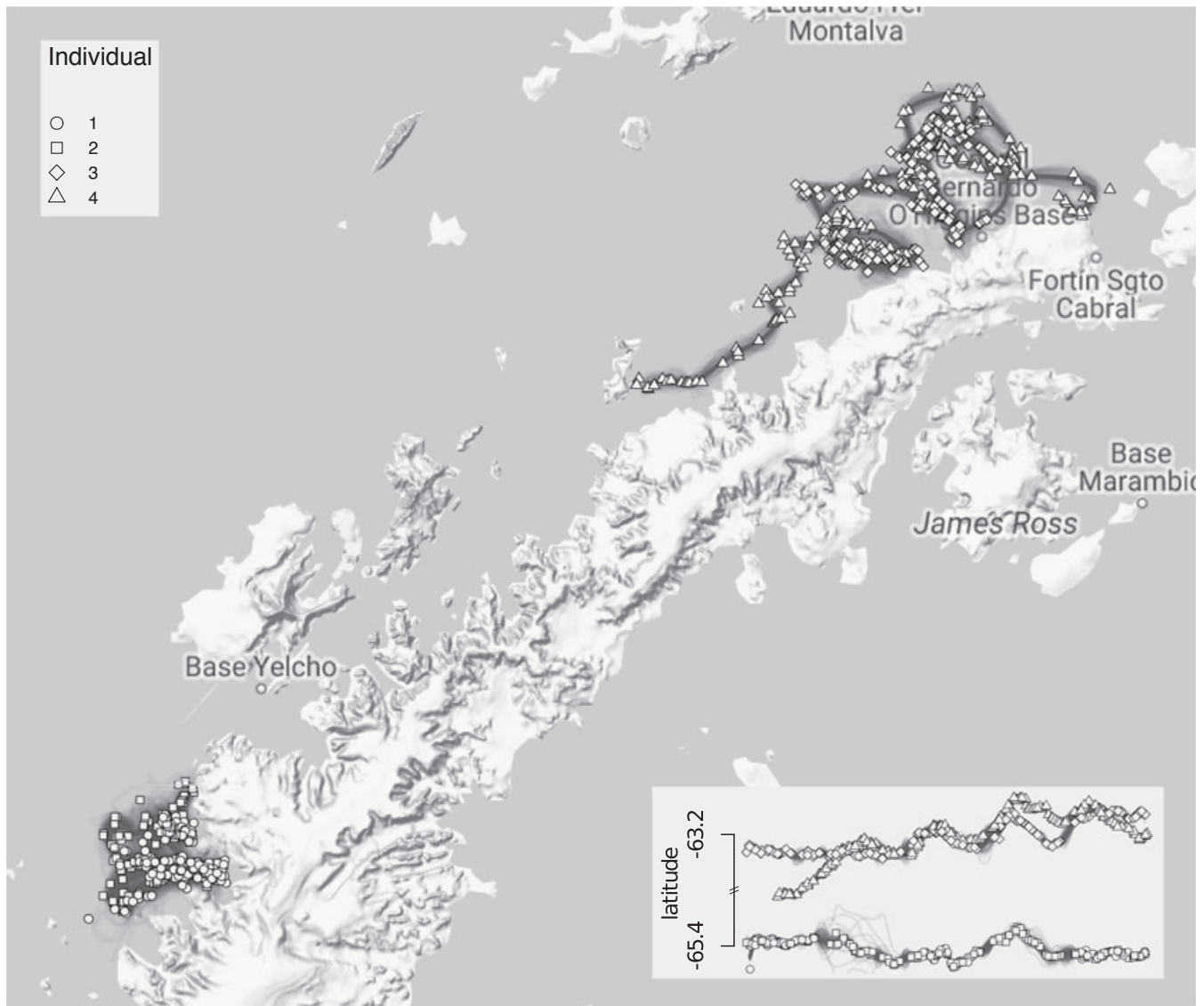
FIGURE 4 Observed telemetry data and joint posterior distribution of the true paths of all individuals ($\tilde{\mu}$). The solid lines represent the posterior mean, and the semitransparent lines are draws from the posterior distribution to illustrate uncertainty in the path reconstruction. The points are the observed locations from the Argos satellite system. The shapes correspond to the four individuals in the study and match those used in Figure 5. The subplot shows the latitude of each individual over time. Map created with Kahle and Wickham (2013). Map data ©2017 Google

The posterior distributions for the network relationships are shown in Figure 6. We can see strong evidence for a significant relationship between individuals 1 and 2; however, there is weak evidence of a meaningful social connection between individuals 3 and 4. Similarly, there is no evidence of connections existing between any other pairs of individuals. Credible intervals for other model parameters, as well as the specified prior distributions and hyperparameter values, are provided in Table 2.

6 | DISCUSSION

Appropriate models for dependent, multivariate data are application dependent. GPs offer a flexible, parsimonious tool for analyzing complex data, such as those that arise from measuring the movement of animals, if one can specify a valid form for the covariance. The need for covariances that are mathematically sound, realistic, and interpretable has motivated decades of research with a vast array of applications because satisfying all three of these goals simultaneously is
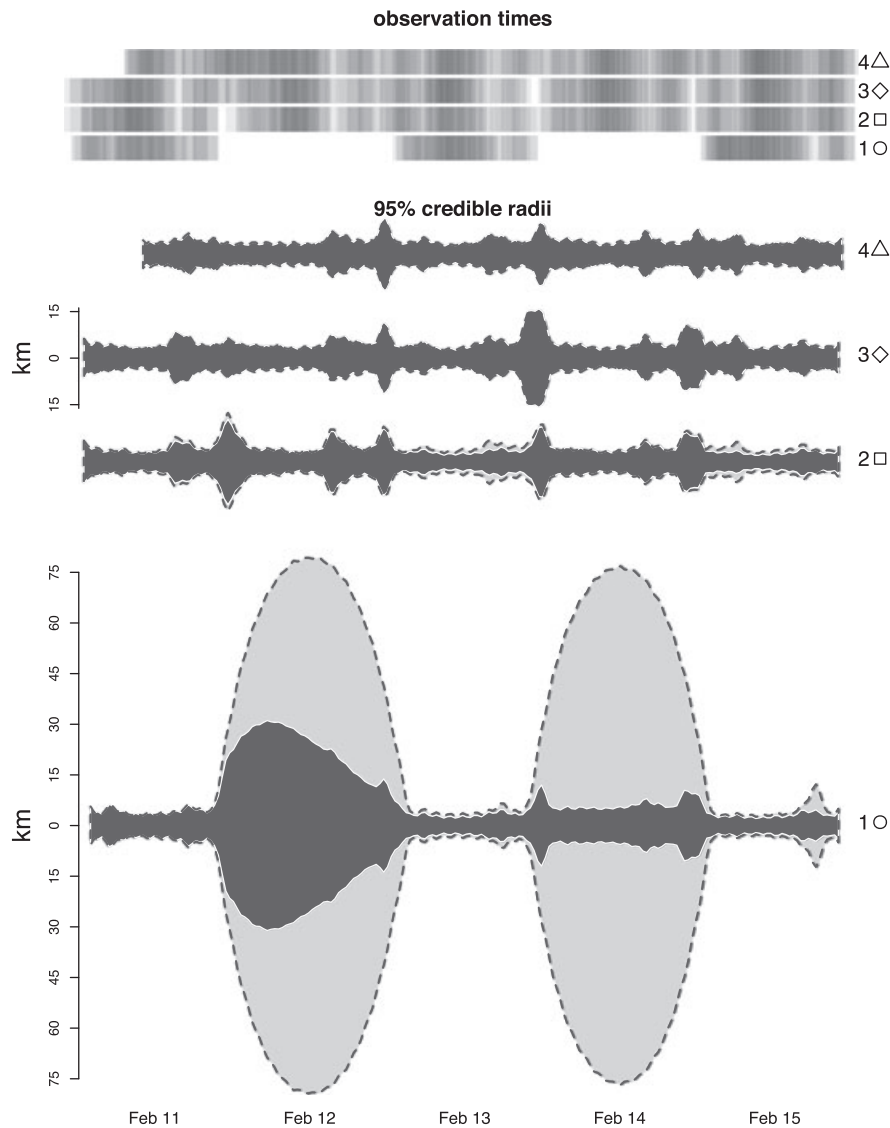
FIGURE 5 *Top*: Observation times for the three killer whales. Darker regions correspond to denser observation times or, equivalently, shorter gaps between observations. *Bottom*: 95% circular credible region radii at each time point for all four individuals. The dashed outer line shows the radii for the independent model, and the solid polygons show the radii for the full model

challenging. Our proposed PCC framework represents a novel perspective in the construction of covariance functions. Through an application involving the movement of killer whales, we demonstrated that the PCC framework can be used to create sophisticated covariance functions that account for several important mechanisms, without resorting to unrealistic assumptions such as separability.

The joint movement of interacting individuals can be viewed as a multivariate temporal process. Methods for interpolating multivariate spatial and spatiotemporal processes, also called "cokriging," have been studied for several decades (e.g., Cressie, 1991; Myers, 1982). The primary challenge has been to develop models for multivariate processes that accurately capture the dynamics within each process, as well as across processes, while ensuring that the resulting cross-covariance structure is valid (symmetric and nonnegative definite). While not explicitly framed as a PCC, Ver Hoef and Barry (1998) proposed a model for cokriging that shares important connections with our proposed model for the joint movement of interacting animals. Ver Hoef and Barry (1998) also approached the problem of specifying a valid, mechanistically motivated cross-covariance structure for multivariate processes through the use of kernel convolutions, relaxing unrealistic assumptions about the covariance structure. Similar to the way dependence among individuals arises through the application of a social smoothing kernel (Section 2.3.1), Ver Hoef and Barry (1998) induced dependence in a multivariate process by smoothing across variables. The cross-covariance described by Ver Hoef and Barry (1998) also accommodates spatiotemporal lags in the dependence among variables. This generalization may also be a useful
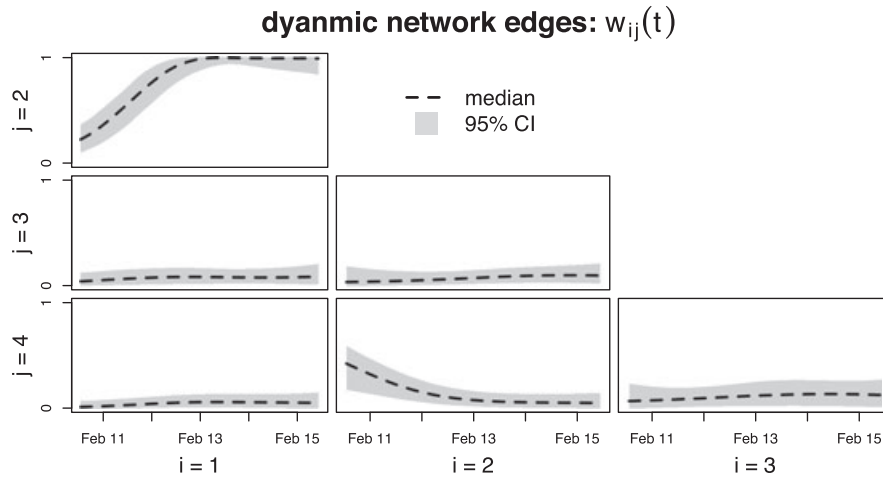
FIGURE 6 Posterior distribution of each edge in the dynamic, weighted social network that defines the social smoother used in stage two of the proposed PCC model (IP-DEP). Each plot shows the time evolution for the edge, $w_{ij}(t)$, relating a specific pair of killer whales indexed by row (i) and column (j) for all times t

TABLE 2 Posterior credible intervals and prior distributions for the IP-DEP model in killer whale application

	posterior (IP-DEP)		posterior (IP-IND)		prior
parameter	median	(2.5%, 97.5%)	median	(2.5%, 97.5%)	density
ϕ_{inl}	0.00848	(0.00756, 0.00954)	0.00850	(0.00762, 0.00951)	Gamma(2, 100)
σ_0^2	705	(292, 2,040)	1,440	(627, 4,510)	IG(1, 10)
$\sigma_{\mu/s}^2$	3,710	(2,600, 5,490)	1,700	(1,380, 2,400)	IG($10^{-3}, 10^{-3}$)
σ_μ^2	4.94	(3.52, 7.21)	2.39	(1.88, 3.14)	NA
σ_s^2	0.00133	(0.00122, 0.00145)	0.00133	(0.00123, 0.00145)	IG($10^{-3}, 10^{-3}$)
σ_w^2	0.491	(0.370, 0.660)	NA	NA	IG(52, 10)

Note. IP-DEP = the full model for dependent movement; IP-IND = interpath independence.

feature in future models for animal movement, where a temporal lag may allow researchers to capture the effect of animals following one another.

The particular model we constructed for animal movement cannot be used to understand all forms of dependence that may exist among interacting individuals. Rather, it has been tailored to the case in which interactions among animals manifest themselves as movement along proximally close trajectories. Particles may also exhibit other forms of dependence, such as a tendency to repel in the case of strong territorial behavior, which would not be well described by the same covariance function. The PCC framework allows researchers to define and order kernels as necessary to appropriately model the mechanisms under study. Under a different ordering, the same three kernels we employed result in another useful covariance function for the study of particle movement. We briefly discuss this alternative to highlight the flexibility offered by the PCC framework and provide some intuition for the certain role kernel functions play in the characteristics of the resulting random process.

As mentioned in Section 2.1, choosing the ordering of the kernels in a PCC represents a meaningful modeling decision when kernel convolutions do not commute. In our proposed model for dependent movement, the second and third stages of smoothing commute, whereas the first and second do not. We can interpret the commutativity between the social and inertial smoothers mechanistically, by noting that it does not matter whether the effect of the latent social connections operates on the raw, Brownian paths, or the temporally smoothed paths. In contrast, smoothing white noise first with the social kernel, and then with the step function kernel results in a very different form of dependence. This alternative ordering ($h^{(soc)} \rightarrow h^{(bm)} \rightarrow h^{(inl)}$) presents another plausible mechanism through which dependence might arise in paths taken by multiple particles.

When applied to the velocity of a particle, smoothing with the step function kernel $h^{(bm)}$ returns the associated position process (after taking into consideration the appropriate initial location of the particle). Therefore, Brownian motion can be thought of as a random position process in which the velocity of the particle during each infinitesimal span of time is a realization from a Gaussian white noise process. In our proposed PCC model for dependent movement, the social smoother

has the effect of "squeezing" the positions of strongly connected particles toward each other. If we instead employed the social smoother before the step function kernel, the result would be to "squeeze" the velocities of the particles together, rather than their positions. Thus, we would be inducing a tendency for particles to move in similar directions, although not necessarily a tendency to be in similar locations. Such an effect would be visible in the particles as movement in parallel, perhaps with a considerable distance between connected particles.

A PCC approach to constructing covariance functions for GPs allows for broad flexibility in model development, offering researchers a highly customizable framework that can be used in a wide variety of applications. We demonstrated the value of this approach with an application to animal movement; however, PCCs can be used to model a broad range of random processes. Rather than relying on parametric families of covariance functions, PCCs encourage the use of interpretable, intuitive, and problem-specific convolution kernels, allowing for direct incorporation of scientific knowledge.

ORCID

Henry R. Scharf http://orcid.org/0000-0002-9266-4191

REFERENCES

Andrews, R. D., Pitman, R. L., & Ballance, L. T. (2008). Satellite tracking reveals distinct movement patterns for Type B and Type C killer whales in the southern Ross Sea, Antarctica. *Polar Biology*, *31*(12), 1461–1468.

Barry, R. P., & Ver Hoef, J. M. (1996). Blackbox kriging: Spatial prediction without specifying variogram models. *Journal of Agricultural, Biological, and Environmental Statistics*, *1*(3), 297–322.

Bolin, D., & Lindgren, F. (2013). A comparison between Markov approximations and other methods for large spatial data sets. *Computational Statistics & Data Analysis*, *61*, 7–21.

Brost, B. M., Hooten, M. B., Hanks, E. M., & Small, R. J. (2015). Animal movement constraints improve resource selection inference in the presence of telemetry error. *Ecology*, *96*(10), 2590–2597.

Buderman, F. E., Hooten, M. B., Ivan, J. S., & Shenk, T. M. (2016). A functional model for characterizing long-distance movement behaviour. *Methods in Ecology and Evolution*, *7*(3), 264–273.

Calder, C. A. (2007). Dynamic factor process convolution models for multivariate space-time data with application to air quality assessment. *Environmental and Ecological Statistics*, *14*(3), 229–247.

Calder, C. A. (2008). A dynamic process convolution approach to modeling ambient particulate matter concentrations. *Environmetrics*, *19*(1), 39–48.

Codling, E. A., & Bode, N. W. (2014). Copycat dynamics in leaderless animal group navigation. *Movement Ecology*, *2*, 11.

Cressie, N. (1991). *Statistics for spatial data* (1st ed). New York, NY: John Wiley & Sons.

Cressie, N., & Pavlicová, M. (2002). Calibrated spatial moving average simulations. *Statistical Modelling*, *2*(4), 267–279.

Durante, D., Scarpa, B., & Dunson, D. B. (2014). Locally adaptive factor processes for multivariate time series. *Journal of Machine Learning Research*, *15*, 1493–1522.

Durban, J. W., Fearnbach, H., Burrows, D. G., Ylitalo, G. M., & Pitman, R. L. (2016). Morphological and ecological evidence for two sympatric forms of Type B killer whale around the Antarctic Peninsula. *Polar Biology*, 1–6.

Durban, J. W., & Pitman, R. L. (2012). Antarctic killer whales make rapid, round-trip movements to subtropical waters: Evidence for physiological maintenance migrations? *Biology Letters*, *8*(2), 274–277.

Feynman, R. P., Leighton, R. B., & Sands, M. (1963). *The Feynman lectures on physics*. Reading, MA: Addison-Wesley.

Finley, A. O., Banerjee, S., & Gelfand, A. E. (2013). spBayes for large univariate and multivariate point-referenced spatio-temporal data models. arXiv preprint arXiv:1310.8192.

Fleming, C. H., Fagan, W. F., Mueller, T., Olson, K. A., Leimgruber, P., & Calabrese, J. M. (2016). Estimating where and how animals travel: An optimal framework for path reconstruction from autocorrelated tracking data. *Ecology*, *97*(3), 576–582.

Gelfand, A. E., Kim, H.-J., Sirmans, C. F., & Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, *98*(462), 387–396.

Goldenberg, A., Zheng, A. X., Fienberg, S. E., & Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning*, *2*(2), 129–233.

Goldenberg, S. Z., de Silva, S., Rasmussen, H. B., Douglas-Hamilton, I., & Wittemyer, G. (2014). Controlling for behavioural state reveals social dynamics among male African elephants, Loxodonta africana. *Animal Behaviour*, *95*, 111–119.

Hansell, S. (1984). Cooperative groups, weak ties, and the integration of peer friendships. *Social Psychology Quarterly*, *47*(4), 316–328.

Haydon, D. T., Morales, J. M., Yott, A., Jenkins, D. A., Rosatte, R., & Fryxell, J. M. (2008). Socially informed random walks: Incorporating group dynamics into models of population spread and growth. *Proceedings. Biological sciences / The Royal Society*, *275*(1638), 1101–1109.

Higdon, D. M. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics*, *5*(2), 173–190.

Higdon, D. M. (2002). Space and space-time modeling using process convolutions. In *Quantitative methods for current environmental issues* (pp. 37–56). London: Springer. https://link.springer.com/chapter/10.1007/978-1-4471-0657-9_2#citeas

Hoff, P. D. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in neural information processing systems* (pp. 657–664). http://papers.nips.cc/paper/3294-modeling-homophily-and-stochastic-equivalence-in-symmetric-relational-data

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, *97*(460), 1090–1098.

Hooten, M. B., & Johnson, D. S. (2017). Basis function models for animal movement. *Journal of the American Statistical Association*. *112*(518), 578–589.

Johnson, D. S., London, J. M., Lea, M.-A., & Durban, J. W. (2008). Continuous-time correlated random walk model for animal telemetry data. *Ecology*, *89*(5), 1208–1215.

Kahle, D., & Wickham, H. (2013). ggmap: Spatial visualization with ggplot2. *The R Journal*, *5*(1), 144–166.

Katzfuss, M. (2017). A multi-resolution approximation for massive spatial datasets. *Journal of the American Statistical Association*. *112*(517), 201–214.

Langrock, R., Hopcraft, J. G. C., Blackwell, P. G., Goodall, V., King, R., Niu, M., Patterson, T. A., … Schick, R. S. (2014) Modelling group dynamic animal movement. *Methods in Ecology and Evolution*, *5*(2), 190–199.

McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M., & Morales, J. M. (2014). When to be discrete: The importance of time formulation in understanding animal movement. *Movement Ecology*, *2*(1), 21.

Myers, D. E. (1982). Matrix formulation of co-kriging. *Journal of the International Association for Mathematical Geology*, *143*, 249–257.

Nychka, D. W., Bandyopadhyay, S., Hammerling, D., Lindgren, F., & Sain, S. (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, *24*(2), 579–599.

Paciorek, C. J. (2003). Nonstationary Gaussian processes for regression and spatial modelling [Doctoral dissertation, PhD thesis]. Carnegie Mellon University: Pittsburgh, Pennsylvania.

Potts, J. R., Bastille-Rousseau, G., Murray, D. L., Schaefer, J. A., & Lewis, M. A. (2014). Predicting local and non-local effects of resources on animal space use using a mechanistic step selection model. *Methods in Ecology and Evolution*, *5*(3), 253–262.

Rue, H., & Held, L. (2005). Continuous time random walks. In *Gaussian Markov random fields: Theory and application* (pp. 123–130). Boca Raton, FL: Chapman and Hall/CRC.

Russell, J. C., Hanks, E. M., & Haran, M. (2016). Dynamic models of animal movement with spatial point process interactions. *Journal of Agricultural, Biological, and Environmental Statistics*, *21*(1), 22–40.

Scharf, H. R., Hooten, M. B., Fosdick, B. K., Johnson, D. S., London, J. M., & Durban, J. W. (2016). Dynamic social networks based on movement. *Annals of Applied Statistics*, *10*(4), 2182–2202.

Scharf, H. R., Hooten, M. B., & Johnson, D. S. (2017). Imputation approaches for animal movement modeling. *Journal of Agricultural, Biological, and Environmental Statistics*. https://doi.org/10.1007/s13253-017-0294-5

Shepp, L. A. (1966). Radon-Nikodym derivatives of Gaussian measures. *Annals of Mathematical Statistics*, *37*(2), 321–354.

Ver Hoef, J. M., & Barry, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, *69*(907), 275–294.

Ver Hoef, J. M., Peterson, E., & Theobald, D. (2006). Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics*, *13*(4), 449–464.

Ver Hoef, J. M., & Peterson, E. E. (2010). A moving average approach for spatial statistical models of stream networks. *Journal of the American Statistical Association*, *105*(489), 6–18.

Wecker, W. P., & Ansley, C. F. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association*, *78*(381), 81–89.

Wikle, C. K. (2010). Low-rank representations for spatial processes. In *Handbook of spatial statistics* (pp. 107–118). Boca Raton, FL: CRC Press Taylor & Francis Group.

Williams, R., & Lusseau, D. (2006). A killer whale social network is vulnerable to targeted removals. *Biology Letters*, *2*(4), 497–500.

APPENDIX A: MODEL SPECIFICATION

A.1 | Data model

Each orthogonal spatial direction (longitude and latitude) is modeled independently using the same specification.

$$\mathbf{s} \sim \mathcal{N}\left(\mathbf{0},\ \sigma_s^2\left(\mathbf{I} + \sigma_{\mu/s}^2 \Delta\tau \widetilde{\mathbf{H}}\boldsymbol{\Sigma}_{dB}\widetilde{\mathbf{H}}'\right)\right)$$

$$\widetilde{\mathbf{H}} = \mathbf{H}^{(inl)}(\phi_{inl})\mathbf{H}^{(soc)}(d\mathbf{B}_w, \phi_w, \sigma_w^2)\mathbf{H}^{(bm)}$$

$$\boldsymbol{\Sigma}_{dB}(t, \tau, i, j) = \mathbf{1}_{\{t=\tau\}}\left(\sigma_0^2 \mathbf{1}_{\{t=0\}} + \mathbf{1}_{\{t>0\}}\right)\mathbf{1}_{\{i=j\}}$$

$$\mathbf{H}^{(bm)}(t, \tau, i, j) = \mathbf{1}_{\{\tau<t\}}\mathbf{1}_{\{i=j\}}$$

$$\mathbf{H}^{(soc)}(t, \tau, i, j) = \mathbf{1}_{\{\tau=t\}}\frac{w_{ij}(t)}{|w_{i\cdot}(t)|}$$

$$\mathbf{H}^{(inl)}(t, \tau, i, j) \equiv \frac{|t-\tau|}{\phi_{inl}}K_1\left(|t-\tau|/\phi_{inl}\right)\mathbf{1}_{\{i=j\}}$$

A.2 | Process model

Movement (integrated out for model fitting)

$$\boldsymbol{\tau} \equiv (\tau_1 = 0, \tau_2, \ldots, \tau_{m-1}, \tau_m = 1)'$$

$$\Delta\tau_k \equiv \tau_k - \tau_{k-1}, \quad 1 < k \le m$$

$$d\boldsymbol{B}_i(\tau_k) \sim \begin{cases} \mathcal{N}\left(0, \sigma_0^2\right); & k = 1 \\ \mathcal{N}\left(0, \Delta\tau_k\right); & k > 1 \end{cases}$$

$$d\mathbf{B}_i = (d\boldsymbol{B}_i(\tau_1), \ldots, d\boldsymbol{B}_i(\tau_m))'$$

$$d\mathbf{B} = \left(d\mathbf{B}_1', \ldots, d\mathbf{B}_p'\right)'$$

$$\tilde{\boldsymbol{\mu}} = \mathbf{H}^{(inl)}\mathbf{H}^{(soc)}\mathbf{H}^{(bm)}d\mathbf{B} = \widetilde{\mathbf{H}}d\mathbf{B}$$

Dynamic network

$$\boldsymbol{\tau}^{(w)} \equiv (\tau_1^{(w)} = 0, \tau_2^{(w)}, \ldots, \tau_{m_w-1}^{(w)}, \tau_{m_w}^{(w)} = 1)'$$

$$\Delta\tau_k^{(w)} \equiv \tau_k^{(w)} - \tau_{k-1}^{(w)}, \quad 1 < k \le m_w$$

$$d\boldsymbol{B}_i^{(w)}(\tau_k) \sim \begin{cases} \mathcal{N}(0, 1); & k = 1 \\ \mathcal{N}\left(0, \Delta\tau_k^{(w)}\right); & k > 1 \end{cases}$$

$$\mathbf{H}^{(w)}(t, \tau) = e^{\frac{-(t-\tau)^2}{\phi_w^2}}$$

$$\boldsymbol{\mu}_i^{(w)} = \sigma_w \mathbf{H}^{(w)}d\mathbf{B}_i$$

$$w_{ij}(\tau_{soc}) \equiv \exp\left\{-\left\|\boldsymbol{\mu}_i^{(w)}(\tau_{soc}) - \boldsymbol{\mu}_j^{(w)}(\tau_{soc})\right\|_2^2\right\}$$

A.3 | Prior model

$$\phi_{inl} \sim \text{Gamma}(\alpha_s, \beta_s) \qquad \sigma_0^2 \sim \text{IG}(a_0, b_0)$$

$$\sigma_{\mu/s} \sim \text{IG}(a_{\mu/s}, b_{\mu/s}) \qquad \sigma_s^2 \sim \text{IG}(a_s, b_s)$$

Application only:

$$\sigma_w^2 \sim \text{IG}(a_w, b_w)$$

A.4 | Fixed parameters

Simulation study:

$$\phi_w \quad \text{and} \quad \sigma_w^2$$

Application:

$$\phi_w$$

APPENDIX B: MODEL FITTING DETAILS

B.1 | Simulation Study

We simulated approximately continuous true paths using a grid of 500 equally spaced time points on the unit interval. We simulated 100 observation times drawn uniformly from the same unit interval. For estimation, we fixed the hyperparameters associated with the network to $\phi_w = 4/15$ and $\sigma_w^2 = 10$ and used the priors given in Table 1. For each fit, we obtained 4,000 iterations and discard the first 3,000 as burn-in. The entire simulation study was parallelized across six 3 GHz cores and required approximately one week of computation time.

B.2 | Killer whales

We acquired 100,000 iterations on a single computing node and discarded the first 50,000 as burn-in to yield a sample size of 50,000. We employ diffuse priors for all parameters. We set the hyperparameters associated with the network to $\phi_w = 0.3$ and $\sigma_w^2 = 10$. Model fitting was performed using a processor speed of 3 GHz and required approximately 100 hrs of computing time.