

DISSERTATION

STATISTICAL MODELS FOR DEPENDENT TRAJECTORIES WITH APPLICATION TO
ANIMAL MOVEMENT

Submitted by

Henry R. Scharf

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2017

Doctoral Committee:

Advisor: Mevin B. Hooten

Daniel S. Cooley

Bailey K. Fosdick

N. Thompson Hobbs

Copyright by Henry R. Scharf 2017

All Rights Reserved

ABSTRACT

STATISTICAL MODELS FOR DEPENDENT TRAJECTORIES WITH APPLICATION TO ANIMAL MOVEMENT

In this dissertation, I present novel methodology to study the way animals interact with each other and the landscape they inhabit. I propose two statistical models for dependent trajectories in which dependencies among paths arise from pairwise relationships defined using latent dynamic networks. The first model for dependent trajectories is formulated in a discrete-time framework. The model allows researchers to make inference on a latent social network that describes pairwise connections among actors in the population, as well as parameters that govern the type of behavior induced by the social network. The second model for dependent trajectories is formulated in a continuous-time framework and is motivated primarily by reducing uncertainty in interpolations of the continuous trajectories by leveraging positive dependence among individuals. Both models are used in applications to killer whales. In addition to the two models for multiple trajectories, I introduce a new model for the movement of an individual showing a preference for areas in a landscape near a complex-shaped, dynamic feature. To facilitate estimation, I propose an approximation technique that exploits of locally linear structure in the feature of interest. I demonstrate the model for the movement of an individual responding to a dynamic feature, as well as the approximation technique, in an application to polar bears for which the changing boundary of Arctic sea ice represents the relevant dynamic feature.

ACKNOWLEDGEMENTS

I have had the help of many talented, passionate, considerate, and level-headed friends and colleagues over the course of my PhD, to whom I will be forever grateful.

First, thanks to my family, most of all my loving wife Stephanie. Her support during my PhD has been unconditional and bottomless.

Thanks to the Hootenannies, that evolving group of research associates, graduate students, and post-docs who work alongside Mevin Hooten at the exciting interface of statistics and ecology. Among them, Franny Buderman, Perry Williams, Brian Brost, John Tipton, Trevor Hefley, and Christopher Peck were fantastic colleagues and friends, and were always available to listen and to help me overcome the challenges and frustrations that come with statistical research. A special thank you to Franny, whose dry humor and empathy gave me clarity and perspective whenever I needed it, regardless of whether I realized I did.

Thanks to all the supportive colleagues I've worked with over the course of several fascinating research projects, especially Devin Johnson, John Durban, and Josh London at the National Oceanic and Atmospheric Administration, and Ryan Wilson at the U.S. Fish and Wildlife Service.

Thanks to Ryan Elmore, who helped me begin my PhD career in statistics. Working with Ryan during his time at the National Renewable Energy Lab gave me my first taste of what statistical research was like, and he has continued to be a mentor for me well beyond the time frame of that initial project.

Thanks to Geof Givens who taught the first linear regression course I ever took, in which my mind was forced to grow at an alarming rate, and which continues to be my favorite statistics course ever.

Thanks to Alison Ketz, a wonderful friend and compatriot whose wisdom, confidence, and determination have been a guide for me during the entirety of my degree.

Thanks to my fellow graduate students in the statistics department at Colorado State University. To our collective benefit, the department has continually been a positive and supportive place

to work and study. In particular, Cristian Oliva, Soo Young Kim, Zach Weller, Joshua Hewitt, Miranda Fix, and Ben Prytherch gave me their time and encouragement more times than I can remember.

Thanks to Kristin Stephens, Melissa Frances, Katy Jackson, Karena Alons, and Zube for all their assistance and patience.

Thanks to Dan Cooley and Tom Hobbs for their time and effort as committee members and their thoughtful professional advice.

Thanks to Bailey Fosdick, who has been generous enough to be a committee member, a research collaborator, a teacher, a mentor, and a friend.

Finally, thanks to my advisor Mevin Hooten. Mevin's dedication to all his students is profound and immediate, and I consider myself immensely fortunate to have had the privilege of his advice and mentorship.

DEDICATION

For my loving wife Stephanie, whose patience with me during my PhD amounts to nothing compared with her patience with me during my (repeated) tellings of bad jokes.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
DEDICATION	v
LIST OF TABLES	viii
LIST OF FIGURES	ix
Chapter 1 Introduction	1
1.1 Motivation	1
1.2 Distributions for dependent variables	3
1.2.1 Gaussian processes: continuous index	6
1.2.2 Gaussian processes: discrete index	10
1.2.3 Models for resource selection	12
1.3 Overview	14
Chapter 2 Dynamic Social Networks Based on Movement	16
2.1 Introduction	16
2.2 Methods	18
2.2.1 Position process	18
2.2.2 Dynamic social network	23
2.2.3 Measurement error and time alignment	25
2.2.4 Priors	27
2.3 Simulation	27
2.4 Killer whales	29
2.5 Discussion	32
2.6 Declaration of support	34
Chapter 3 Process Convolution Approaches for Modeling Interacting Trajectories	37
3.1 Introduction	37
3.2 Methods	39
3.2.1 A multiple-kernel convolution framework	39
3.2.2 Finite representation	43
3.2.3 A process convolution chain for dependent movement	44
3.2.4 Process prior for the dynamic social network	50
3.3 Model implementation	52
3.4 Simulation	53
3.5 Killer whales	55
3.6 Discussion	60
3.7 Declaration of support	63

Chapter 4	Accounting for Phenology in the Analysis of Animal Movement	64
4.1	Introduction	64
4.2	Model Development	65
4.2.1	Feature preference	65
4.2.2	Linearization approximation	68
4.3	Application	70
4.3.1	Goals and previous work	70
4.3.2	Movement process model	72
4.3.3	Activity centers and sub-population membership	75
4.3.4	Measurement error	75
4.3.5	Estimation	77
4.3.6	Results	79
4.4	Simulation Study	82
4.5	Discussion	84
Chapter 5	Conclusion	85
5.1	Models for dependent trajectories	85
5.1.1	Randomized matrix decompositions	85
5.1.2	Modeling the precision matrix	86
5.2	Resource Selection Functions	87
5.3	Conclusion	88
Appendix A	Supplemental Material for Chapter 2	105
A.1	Priors	105
A.2	Full conditionals	106
A.3	Hyper-/tuning-parameters	110
A.4	Dynamic Social Network between killer whale types	110
Appendix B	Supplemental Material for Chapter 3	113
B.1	Model specification	113
B.2	Model fitting details	114
Appendix C	Supplemental Material for Chapter 4	116
C.1	MCMC Implementation details	116

LIST OF TABLES

2.1	Marginal posterior medians and 95% credible intervals for model parameters. True values for the simulation were chosen to yield plausible movement paths. The right column describes the prior distributions used.	29
2.2	Marginal posterior medians and 95% credible intervals for model parameters when fit to the killer whale tagging data. The values reflect a strong alignment effect (α), weak attraction effect (β), and a stable (ϕ), sparse (p_1) social network. The right column describes the prior distributions used.	30
4.1	Posterior medians and equal-tailed credible intervals for all model parameters, as well as prior distributions and hyper-parameters.	79
A.1	Simulation	110
A.2	Killer whales	110
B.1	True values and prior distributions used in simulation study.	115
B.2	Posterior credible intervals and prior distributions for the IP-DEP model in killer whale application.	115

LIST OF FIGURES

1.1	Realizations of stochastic processes in two dimensions corresponding to <i>left</i> : Brownian motion, <i>middle</i> : integrated Brownian motion of Johnson et al. (2008a), and <i>right</i> : smoothed Brownian motion of Hooten and Johnson (2017)	7
1.2	Graphs representing neighborhood structure in discretely-indexed GP with index representing <i>left</i> : time and <i>right</i> : individual.	11
2.1	This schematic illustrates the two channels through which the dynamic social network influences movement. The dashed lines represent where the ego-network of individual i would be expected to be at time t under attraction alone, and the parallel dotted lines represent the alignment between individual i and the average of the differences $\boldsymbol{\mu}_j(t) - (\boldsymbol{\mu}_j(t-1) + \beta \tilde{\boldsymbol{\mu}}_j(t-1))$	23
2.2	A subset of the complete estimated dynamic network for the simulated data on six individuals. The titles correspond to the i^{th} and j^{th} individuals in $w_{ij}(t)$. The dashed line is the true network, the solid line is the posterior mean from the proposed Bayesian model, and the gray region represents one standard deviation above and below the posterior mean. The dotted line shows the network defined by \mathbf{W}^R , where individuals are deemed connected whenever they are separated by a distance less than R (see Section 2.3). (Note: The lines are offset slightly near 0 and 1 for visual clarity.)	35
2.3	A selection of the $\binom{7}{2} = 21$ possible pairs of individuals in the killer whale study sample. The left column is all pairs of killer whales of type B2 (labeled 2, 3, 4), and the right column is all pairs of killer whales of type A (labeled 5, 6, 7). The solid line in each plot shows the posterior mean for w_{ij} and the gray region represents one standard deviation above and below the posterior mean. The dotted line shows the network defined by \mathbf{W}^R , where individuals are deemed connected whenever they are separated by a distance less than R . (Note: The lines are offset slightly near 0 and 1 for visual clarity.)	36
3.1	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	48

3.2	Path errors. The top row of plots shows the ratios of SPE for the independent (IP-IND) to full (IP-DEP) models (thus, larger ratios correspond to the full model outperforming the alternative). The bottom row shows the analogous ratios of $ACRR_{0.95}$. 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} $). Within each plot, the strength of the social connection increases along the x-axis. The lines represent the median value of all ratios across the 20 simulations, and the associated polygons represent the 25% and 75% quantile boundaries. Finally, the solid lines correspond to simulations with high tortuosity in the true paths, while dashed lines correspond to simulations with low tortuosity.	56
3.3	Observed telemetry data and joint posterior distribution of the true paths of all individuals ($\tilde{\mu}$). The solid lines represent the posterior mean, and the semi-transparent 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 colors correspond to the four individuals in the study, and match those used in Figure 3.4. The subplot shows the latitude of each individual over time. Map created with Kahle and Wickham (2013). Map data ©2017 Google.	58
3.4	<i>TOP</i> : Observation times for the three killer whales. Darker regions correspond to denser observation times, or equivalently, shorter gaps between observations. <i>BOTTOM</i> : 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.	59
3.5	Posterior distribution of the dynamic 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 a specific pair of killer whales indexed by row and column.	60
4.1	The product of the two densities in the left plot (one of which is improper) result in the density shown in the right plot.	70
4.2	Sub-population boundaries of polar bears (Obbard et al., 2010; U.S. Fish and Wildlife, 2016). Sub-population abbreviations are: Southern Beaufort Sea (SB), Northern Beaufort (NB), Kane Basin (KB), Norwegian Bay (NW), Lancaster Sound (LS), Gulf of Boothia (GB), M’Clintock Channel (MC), Viscount Melville Sound (VM), and Western Hudson Bay (WH).	71
4.3	Posterior means for each z_i , organized by the agency responsible for tagging the individual. Darker points reflect posterior means closer to 0 (SB, orange) and 1 (CS, purple).	80
4.4	The orange and purple lines represent trajectories drawn from the process imputation distributions (Section 4.3.5) of all polar bears from 2008–2016. Orange lines correspond to polar bears associated with the Southern Beaufort Sea sub-population, and purple lines correspond to bears associated with the Chukchi Sea sub-population. The darkness of the color corresponds to the probability we assign to a given sub-population. The black line shows the posterior mean of the boundary where, marginally, the probability of sub-population membership is balanced (Section 4.3.6). The dashed lines show pointwise equal-tailed 95% credible intervals perpendicular to the mean boundary. The large polygons with thin black borders show the current sub-population delineations (CS, SB, and NB from left to right; see Figure 4.2).	81

4.5	Distributions of three summary statistics for paths simulated from the exact (solid) and approximate (dotted) movement models. Thin, faint lines show the distribution of each statistic for each of 16 simulated individuals over 366 days. Thick, dark lines show the distribution of each statistic across all individuals and times.	83
A.1	A selection of the $\binom{7}{2} = 21$ possible pairs of individuals in the killer whale study sample. The plots displayed are for all inter-type pairs of killer whales of type B2 (labeled 2, 3, 4) and A (labeled 5, 6, 7). The solid line in each plot shows the posterior mean for w_{ij} and the gray region represents one standard deviation above and below the posterior mean. The dotted line shows the network defined by \mathbf{W}^R , where individuals are deemed connected whenever they are separated by a distance less than R . No posterior means above 0.5 were predicted for inter-type connections. (Note: The lines are offset slightly near 0 and 1 for visual clarity.)	111
A.2	A selection of the $\binom{7}{2} = 21$ possible pairs of individuals in the killer whale study sample. The plots displayed are for all inter-type pairs of killer whales between the sole individual of type B1 (labeled 1) and those of type B2 (labeled 2, 3, 4) and A (labeled 5, 6, 7). The solid line in each plot shows the posterior mean for w_{ij} and the gray region represents one standard deviation above and below the posterior mean. The dotted line shows the network defined by \mathbf{W}^R , where individuals are deemed connected whenever they are separated by a distance less than R . No posterior means above 0.5 were predicted for inter-type connections. (Note: The lines are offset slightly near 0 and 1 for visual clarity.)	112
C.1	Realizations from the process imputation distribution for two polar bears in 2008. The polar bear to the right has fewer observations with greater measurement uncertainty, leading to greater variability among the sampled paths. The polar bear on the right has dense observations from both GPS and Argos devices. Sampled paths are much less variable.	117
C.2	Posterior medians for each p_i , organized by the agency responsible for tagging the individual. Darker points reflect posterior medians closer to 0 (SB, orange) and 1 (CS, purple). Vertical lines represent posterior 1st and 3rd quartiles marginally for each p_i	118

Chapter 1

Introduction

1.1 Motivation

In this dissertation, the scientific questions I answer involve the way animals move and make use of the space around them. I make the assumption that the actions of individuals are based primarily on information they have about their environment and other individuals of the same species (conspecifics), and I endeavor to learn how individuals make use of that information. Especially with regard to environmental factors, questions about space use have been a subject of research for decades, and a considerable body of literature has been produced containing many valuable discoveries (e.g., Manly et al. 2002; Hooten et al. 2017a). By comparison, attempts to understand the types of interactions that take place among multiple individuals are in their infancy, largely due to the high demand on data quality and computational resources associated with such analyses. Data that contain information about interactions between individuals in a population must, of course, be gathered for multiple individuals, typically on a fine time scale, and with a reasonably high degree of precision. Neither the volume of data, nor the methodological tools, nor the computational resources required to analyze them have been available until recently. However, improvements in technology have led to the development of telemetry devices that can gather extremely fine-scale data, fundamentally changing both the types of questions posed by researchers, and the analytic tools developed to handle such dense data sets (e.g., Prange et al. 2006; Rutz and Hays 2009; Cagnacci et al. 2010; Davis et al. 2012).

Accurately modeling the dependence among multiple interacting trajectories is of value whether one views the interactions as processes of fundamental interest, or random effects capturing unexplained dependence in the data, perhaps arising due to unobserved environmental factors. In Chapter 2, I introduce a model for dependent movement intended to capture effects directly related to dynamic social connections in killer whales near the Antarctic peninsula. The model allows

researchers to learn about a dynamic social network that describes pairwise relationships among the killer whales, as well as the way those connections are manifested in behavior. In contrast, the model described in Chapter 3 is motivated by an interest in accounting for dependence among trajectories to improve predictions of the true locations of individuals during long intervals between measurements. In both models, I make use of a latent, unobserved social network to characterize pairwise dependence between individuals, but in Chapter 3 the latent social network is not of direct scientific interest. Rather, it is a flexible component in the model used to capture dependence for the purposes of interpolation.

The data used to learn about animal movement typically take the form of remotely gathered, or telemetered, observations of the location of an individual. Observations can be made in a wide variety of ways, however, most standard techniques require a transmitter attached to an individual that can be remotely sensed by a secondary device such as a hand-held receiver or satellite (White and Garrott 1990; Millspaugh and Marzluff 2001; Davis et al. 2012). The options available to wildlife biologists depend on factors including the size of the species of interest, its habitat, the desired observation frequency, and the battery life of the transmitter. Myriad techniques and devices have been developed, each with benefits and drawbacks, and each involving a usually non-negligible amount of measurement error (e.g., Hays et al., 2001). While it would often be convenient from a modeling perspective to have observations made at regular time intervals, generally the conditions under which observations are made do not permit such designs. Rather, measurements are made whenever environmental and technological constraints allow.

The contributions made by work presented here arose from a need to further the methodological tools available to scientists studying animal movement. As a general rule, I sought to develop models and inferential procedures that incorporate scientifically and mechanistically motivated features and could be operationalized on modest, widely available computing hardware.

1.2 Distributions for dependent variables

The path an individual takes through a landscape can be modeled as a, possibly infinite, collection of random variables indexed in time. Part of the challenge to modeling animal movement comes from the fact that these random variables are not independent. Rather, two locations that are close together in time will covary because an individual at time t is limited in how far it can be from its location at time $t + \Delta t$ by the amount of time that has passed in the interim. Thus, it is reasonable to expect the strength of dependence between two random variables corresponding to the location of an individual at times t and $t + \Delta t$ to be a function of how far the individual is conceivably able to move during an interval of length Δt . This temporal dependence, and how it may change in time or in response to different environmental factors is the basis of much ongoing research. When the locations of multiple interacting individuals are considered, the possibility for even more complex dependence structures arises. Two of the tools commonly used by statisticians to account for the dependence structure present in animal movement data are Gaussian processes (GPs) and resource selection functions (RSFs).

While I have written this introductory chapter so that it is broadly accessible and qualitative in nature, I nevertheless use mathematical notation to help make some concepts precise. Let $s(t)$ denote the random variable giving the location of an individual in space at time t , where t a member of some temporal indexing set \mathcal{T} . For example, one might use $\mathcal{T} = \{1, \dots, 365\}$ to index days in a year, and define $s(t)$ to be a two-dimensional vector representing the location of an individual in latitude and longitude on the t^{th} day of the year. A collection of random variables, particularly those with an ordered index such as time, are often referred to as a stochastic process.

Among the many distributions that have been used to describe stochastic processes, by far the most widely used is the multivariate Gaussian distribution. Much of its popularity is attributable to two convenient characteristics. First, the multivariate Gaussian is one of very few parametric families of multivariate distributions whose density function can be evaluated exactly using a closed-form expression. This makes it possible to fit multivariate Gaussian distributions to observed data and make inference about parameters of interest using a variety of standard estimation

procedures. Second, the multivariate Gaussian distribution characterizes dependence among all the random variables $s(t), t \in \mathcal{T}$ using one parameter for each pair of quantities $s(t)$ and $s(t^*), t \in \mathcal{T}$. This is a strikingly simple parameterization of a multivariate distribution when one considers that the multivariate Bernoulli distribution, which describes dependent random variables that take on only values of 0 or 1, requires a parameter for every possible pair of random variables, every possible triple, every possible quadruple, and so on, up to the total number of values in the indexing set \mathcal{T} to uniquely specify a member of the parametric family. Such a dense parameterization is daunting both from a computational perspective, as well as an interpretive one. It is usually much more intuitive for one to think about dependence between a pair of random quantities than among groups of arbitrary size. For instance, in the case of animal movement, it is natural to suppose that the dependence between two locations $s(t^*)$ and $s(t)$ is related to the amount of time that passes between t^* and t . It is less clear what is meant by a parameter describing the dependence among random variables corresponding to three or more times.

The convenience of the multivariate Gaussian distribution as a tool for modeling dependent random variables, such as consecutive locations of an animal in space, comes with some limitations. One of the most fundamental weaknesses of the multivariate Gaussian distribution is that it only describes random variables with real support. That is, for each variable, non-zero probability is assigned to every value on the real line. For random variables that are, for example, constrained to be positive or integer-valued, the multivariate Gaussian is fundamentally inadequate. In many cases, a “link” function that maps the real line to the proper support of the random variables can effectively address this mismatch. However, for the case of an animal constrained to move in a subset of the real plane, perhaps by a fence or large body of water, a convenient link function is not typically available. Indeed, although it is not directly addressed in this dissertation, models for the movement of individuals in bounded landscapes is a subject of challenging, ongoing research (e.g., Brillinger 2003; Brost et al. 2015; Hanks et al. 2017). Another weakness of the multivariate Gaussian distribution is its limited ability to model extremely unusual events (e.g., de Haan and Ferreira, 2006). Gaussian distributions imply a specific form for the probability of outcomes

far from the center of the distribution that do not always comply with the observed behavior in the data. Finally, while multivariate Gaussian distributions do have the benefit of a closed-form probability density function, evaluating the function often requires inverting a dense, potentially massive covariance matrix. Inverting an $n \times n$ matrix is an operation for which the number of individual calculations grows proportional to n^3 . For values of n greater than a few thousand, the limitations of modern computing hardware result in a significant obstacle to model fitting. Yet even with these significant shortcomings, multivariate Gaussian distributions represent a flexible family of probability measures, and I use them throughout this dissertation.

If the collection of random variables $\mathbf{s} = (s_1, \dots, s_n)'$ arises from a multivariate Gaussian distribution with mean $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ and covariances $\text{cov}(s_i, s_j) = \Sigma_{ij}$, then the probability density function is given by

$$[\mathbf{s}|\boldsymbol{\mu}, \Sigma] = |2\pi\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2}(\mathbf{s} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{s} - \boldsymbol{\mu}) \right\}, \quad (1.1)$$

where Σ is a matrix whose $(i, j)^{th}$ element is Σ_{ij} , and the square brackets denote a probability density. The particular form of the Gaussian density function is helpful to see because it highlights the fact that the entire dependence structure of the n random variables can be completely described by the covariance matrix Σ , yet also reveals one of the frailties of the multivariate Gaussian distribution. Namely, to evaluate the density, one must compute the matrix inverse Σ^{-1} . We discuss a few of the most common approaches to mitigating this computational challenge in Sections 1.2.1 and 1.2.2.

Depending on the type of random phenomenon one is modeling, the index on the random variables may be discrete or continuous. Discrete indices may be unordered, such as labels for each individual in a population of animals, or ordered, such as times corresponding to regular (e.g., hourly or daily) intervals between locations. Continuous indices almost always come with a natural ordering. The most common continuous indices correspond to time and/or space (e.g., the temperature of water in a lake might be indexed both in time, and three continuous spatial indices). A collection of random variables may have multiple indices of differing types. For example, one

may denote by $s_i(t)$ the location of a discretely-indexed individual i at continuously-indexed time t .

In the case of animal movement, indices typically correspond to either times or individuals, and in the case of the former, may be discrete or continuous. While it is often more natural to think of the movement of an individual as a continuous random process, there are advantages to modeling it using a discrete temporal index, and sometimes such models can adequately describe the most salient features of continuous movement.

When at least one index of a multivariate Gaussian distribution is continuous, the collection is often called a Gaussian process (GP) (e.g., Cressie, 1991). Some authors reserve this term exclusively for continuously-indexed processes, but I use it for ordered, discretely-indexed Gaussian distributions as well. In the next two sections, I provide a short background of discretely and continuously-indexed GPs in the context of animal movement.

1.2.1 Gaussian processes: continuous index

When specifying mechanistically-motivated models for animal movement in two dimensions, it is common to model the true path taken by an individual using two processes continuously-indexed in time corresponding to two orthogonal spatial coordinates such as latitude and longitude. Typically, it may be assumed that the probability of a particular path should be the same regardless of our choice of coordinate system; thus, the two processes are taken to be independent and identically distributed. It is also common to simplify mathematical notation by considering a single dimension of movement when describing a model, which I do throughout this dissertation, because generalizing the model to two dimensions is straightforward.

Let $s(t)$ denote the location of an individual in a single spatial dimension at time $t \in \mathcal{T}$, and assume for simplicity that \mathcal{T} is an uninterrupted interval of the real line, $[0, T]$, representing time. One might begin constructing a model for $s(\cdot)$ by incorporating the fact that movement processes must be continuous. This means that as $\Delta t \rightarrow 0$, it must be that $s(t + \Delta t) \rightarrow s(t)$. One way to define a process with this characteristic is by using a random walk. Suppose an individual begins

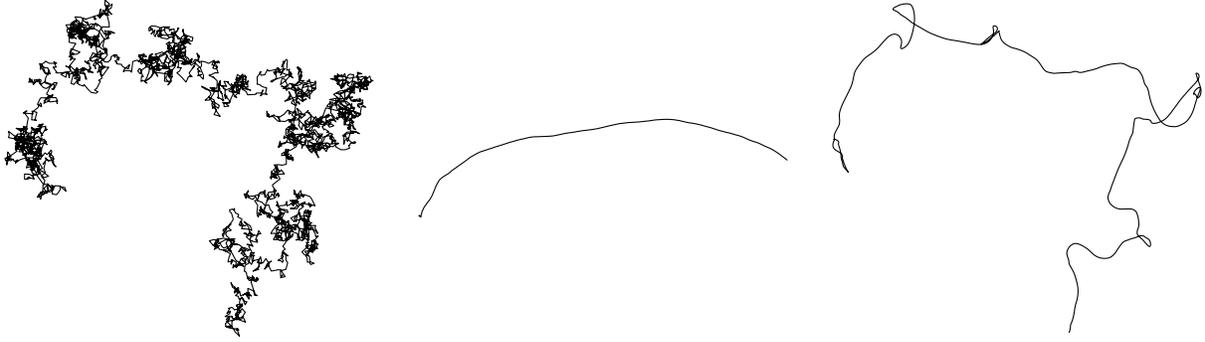


Figure 1.1: Realizations of stochastic processes in two dimensions corresponding to *left*: Brownian motion, *middle*: integrated Brownian motion of Johnson et al. (2008a), and *right*: smoothed Brownian motion of Hooten and Johnson (2017)

at location $s(t)$ and over the course of a short interval of time Δt , takes a random step to arrive at new location $s(t + \Delta t)$. If there is nothing on the landscape to cause the individual to prefer one direction over the other, then the step might occur in either direction with equal probability. The length of the step may also be random in length, however the probability of a very large step should be very low. Additionally, the average step length should be a function of the length of time Δt , as an individual will be able to travel farther during longer intervals. One possible model that incorporates all these features is defined by

$$s(t + \Delta t) = s(t) + \varepsilon(t), \quad \varepsilon(t) \sim \text{N}(0, \Delta t \sigma^2). \quad (1.2)$$

This model for movement assumes that, at each time t , an individual takes a random step coming from a Gaussian distribution centered on zero with variance $\Delta t \sigma^2$. Centering $\varepsilon(t)$ on zero corresponds to the case when the animal shows no directional preference, because the Gaussian distribution is symmetric. The variance of the distribution is proportional to Δt , where the proportionality is determined by a parameter σ^2 that controls how large the expected step length is for a given Δt . If we take the limit $\Delta t \rightarrow 0$, the distribution for $s(\cdot)$ is known as Brownian motion, sometimes also called a Wiener process.

Brownian motion represents the foundation on which many movement models in the literature are based (e.g., Dunn and Gipson 1977; Blackwell 1997; Brillinger and Stewart 1998; Jonsen et al. 2005; Johnson et al. 2008a; Hooten and Johnson 2017), and is an example of a GP with $\text{cov}(s(t^*), s(t)) = \sigma^2 \min(t^*, t)$. Brownian motion satisfies many of our fundamental requirements for models of animal movement, however, it has limitations. One weakness has to do with accurately modeling the physics associated with movement. It can be shown that realizations of paths defined by Brownian motion do not have well-defined first derivatives at any point in time, which means that one cannot define the instantaneous velocity of a physical body following Brownian motion. Visually, this characteristic of the process appears as a “roughness” in the paths (left plot in Figure 1.1). One way to define continuous GPs that do not have this roughness is to let Brownian motion define the velocity, rather than the position process of an individual (Johnson et al., 2008a). Let $v(t)$ denote the velocity of an individual at time t . For a given starting position $s(0)$, the position at time t can be computed by integrating the velocity process (i.e., $s(t) = s(0) + \int_0^t v(\tau)d\tau$). If $v(\cdot)$ is defined as Brownian motion, it can be shown that $\mu(\cdot)$ is a Gaussian process, now with $\text{cov}(s(t^*), s(t)) = t^*t \min(t^*, t)$ and well-defined velocity for all t (middle plot in Figure 1.1). Other modified Brownian processes can be obtained by smoothing Brownian motion with various kernel functions (right plot in Figure 1.1) (Hooten and Johnson, 2017).

In applications other than animal movement, GPs are frequently defined by direct specification of the covariance function. Flexible families of covariance functions, such as the Matérn class (e.g., Cressie, 1991), can capture a wide variety of the features observed in spatio-temporal data, although, the parameters for these families are not always directly interpretable. Interpretable covariance parameters may or may not be of high priority depending on the research goals of an analysis. Fitting the model to data generally requires inverting a matrix defined by evaluations of the covariance function at a finite number of indices, regardless of how a parametric family of covariance functions is defined. As mentioned above, this calculation can be quite demanding, and the size of the calculation grows rapidly with the size of the indexing set.

Several approaches have been devised to address the computational demands of fitting continuously-indexed GPs to data. In general, the motivation for each technique can be described as an attempt to exploit the properties of certain classes of covariance matrices that facilitate matrix inversion. For instance, matrices in which many elements are assumed to be zero (Furrer et al., 2006), and matrices that can be decomposed into a product of matrices of low dimension (or low “rank”) (e.g., Vecchia, 1988; Banerjee et al., 2008; Cressie and Johannesson, 2008; Rue et al., 2009; Datta et al., 2016) are easier to invert compared to the general case. By constraining attention only to GPs whose covariance functions conform to certain structural requirements, it is possible to significantly improve the computational efficiency of model fitting. Such constraints are usually motivated either by claiming that one has good reason to make strong assumptions about covariance structure *a priori*, or by proposing to approximate the true covariance function with one from the sub-class. Often, one can show that the error introduced by a given approximation is modest (e.g., Banerjee et al., 2013a). Regardless of the argument employed, the practical implication is that inference proceeds by limiting attention to a sub-class of GPs with convenient computational characteristics.

In Chapter 3, I use continuously-indexed GPs to model the movement of four interacting killer whales near the Antarctic peninsula over a period of five days in February 2014. The decision to use a continuous index for time was motivated by several factors. First, movement is most naturally thought of as a continuous-time stochastic process. No discrete set of times can ever completely describe the movement of an individual, because between two consecutive time points there is always another one for which the individual must have occupied some intermediary location. Mathematically, the index of the true movement process is “dense,” a feature captured by an interval of the real line, but not by any discrete index. Second, and related to the first point, the telemetered observations used in the application were made at irregular intervals. In some cases, when observations of an individual are made at fixed, equally-spaced time points, and when there is no scientific interest in estimating the location of an individual between those time points, a

discrete-time model is a natural consideration. Neither of these conditions were satisfied for the research presented in Chapter 3.

Even for the case of just four individuals studied over five days, the computational demands of model-fitting were substantial. It would not be possible to implement the same analysis for a group of even a few dozen individuals over a comparable time scale without significant modifications. We discuss potential approaches for addressing computational challenges to scaling inference up to larger data sets in Chapter 5.

1.2.2 Gaussian processes: discrete index

Thus far, the use of GPs to model animal movement has been motivated by the ability of GPs to concisely summarize the dependence structure in a random process through the collection of pairwise relationships contained in the covariance matrix, Σ . When fitting a GP model to data, however, it is the inverse of the covariance matrix, commonly called the precision matrix, that represents the more practically relevant object. If it were possible to specify a GP through the precision matrix directly, the computational expense of inverting the covariance matrix could be saved, and model fitting would be a far more efficient procedure. In the case of continuous indices, the relationship between covariance and precision matrices is understood only for special cases (Lindgren et al., 2011), which do not apply to the models discussed in this dissertation. In contrast, elements of precision matrices corresponding to discretely-indexed GPs have a clear interpretation, and it was shown by Besag (1974) that one can uniquely specify a GP through careful parameterization of the precision matrix.

Let $\mathbf{s} \equiv (s_1, \dots, s_n)'$ be a random vector arising from a GP with precision matrix $\mathbf{Q} \equiv \Sigma^{-1}$ and mean $\boldsymbol{\mu} \equiv (\mu_1, \dots, \mu_n)'$. Further, let $[s_i | \mathbf{s}_{-i}]$ denote the probability density of the variable s_i conditioned on the values of all other elements in the random process, \mathbf{s}_{-i} , where $\mathbf{s}_{-i} \equiv (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_n)'$. The so-called “full conditional” distribution is Gaussian with mean and variance given by

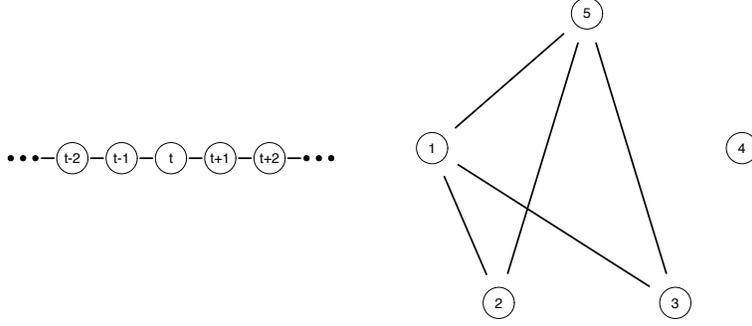


Figure 1.2: Graphs representing neighborhood structure in discretely-indexed GP with index representing *left: time* and *right: individual*.

$$\mathbb{E}(s_i | \mathbf{s}_{-i}) = \mu_i - Q_{ii}^{-1} \sum_{j \neq i} Q_{ij} (s_j - \mu_j) \quad (1.3)$$

$$\text{var}(s_i | \mathbf{s}_{-i}) = Q_{ii}^{-1} \quad (1.4)$$

(Rue and Held, 2005b). Equations (1.3) and (1.4) reveal an important feature of the precision matrix, which is that the element Q_{ij} is 0 if and only if $s_i | \mathbf{s}_{-i}$ is conditionally independent of s_j . The non-zero off-diagonal elements of a precision matrix correspond to edges in a network or graph that describes the conditional dependencies among the elements of \mathbf{s} . The network interpretation offers a natural way to model GPs when the notion of a “neighborhood” exists for each element s_i . When discretely-indexed GPs are specified based on the conditional dependence structure, they are often referred to as Gaussian Markov random fields (GMRFs).

For the case of a discrete temporal index, the neighborhood of s_i might consist of the elements s_{i-1} and s_{i+1} , or some other small group of temporally proximate variables. For the case of an unordered discrete index corresponding to individuals in a population, conditional dependencies may reflect ties within the population corresponding to, for example, social relationships. The plots in Figure 1.2 show graphical representations of conditional dependencies in which nodes or vertices represent indices, and line segments or edges connect variables that are conditionally dependent.

Discrete-time processes represent an attractive modeling approach to researchers because they allow one to construct mechanistically-driven models that can capture a wide variety of behav-

iors. Often, discrete-time model specifications admit efficient estimation procedures, although in the case of animal movement, the price for such computational benefits is usually an assumption that telemetry observations are made at regular intervals and without measurement error. These assumptions are frequently unrealistic in practice, although there are special situations where the strict requirements are satisfied (Langrock et al., 2012). In other cases, as I discuss in Chapters 2 and 4, two-stage estimation procedures can bridge the gap between discrete-time process models and irregularly observed data contaminated with non-ignorable measurement errors (Hanks et al. 2015a; Scharf et al. 2017).

In Chapter 2, I use a discrete-time model to analyze the joint movement of seven killer whales (different from those analyzed in Chapter 3). The goal of the research was to learn about a social network that explained dependence among the paths, as well as the way in which those connections influenced movement. I used a discrete-time stochastic process to model the collection of all paths, because the discrete-time framework admitted a simple parameterization that accommodated a wide variety of behaviors, including “attractive” movement, “repulsive” movement (such as might be observed among highly territorial individuals), and “aligning” movement (i.e., movement along parallel trajectories). In Chapter 4, I model the movements of polar bears in the Chukchi and Beaufort Seas over the course of 9 years using a discrete time index that aligns with auxiliary data describing the shape and extent of seasonally shifting sea ice.

1.2.3 Models for resource selection

Some of the most important scientific questions that can be answered through the analysis of animal movement concern the ways in which individuals use available resources in the landscape (Manly et al., 2002). Answers to these questions can help researchers determine which types of habitat are critical to sustaining healthy populations of sensitive species (Wilson et al., 2016), where corridors may exist that facilitate movement between desirable habitats, and where barriers may be inhibiting movement (Chetkiewicz and Boyce, 2009).

In the resource selection framework, the probability density function for the location $\mathbf{s}(t)$ is modeled as proportional to the product of two functions f and g such that

$$[\mathbf{s}(t)|\boldsymbol{\theta}(t), \boldsymbol{\phi}(t)] = f(\mathbf{s}(t)|\boldsymbol{\theta}(t))g(\mathbf{s}(t)|\boldsymbol{\phi}(t))K^{-1}(t), \quad (1.5)$$

where

$$K(t) \equiv \int_{\mathcal{D}} f(\mathbf{s}^*(t)|\boldsymbol{\theta}(t))g(\mathbf{s}^*(t)|\boldsymbol{\phi}(t))d\mathbf{s}^*(t) \quad (1.6)$$

(e.g., Manly et al. 2002; Lele and Keim 2006; Hooten et al. 2017a). The proportionality constant $K^{-1}(t)$ is defined so that the probability of observing the individual anywhere in the entire domain \mathcal{D} is 1, ensuring that $[\mathbf{s}(t)|\boldsymbol{\theta}(t), \boldsymbol{\phi}(t)]$ is a proper density function.

The function f defines the availability of the location $\mathbf{s}(t)$ and its inclusion represents an understanding that not all locations in the landscape are equally accessible at time t . Often, the availability function is used to account for the fact that an individual can only travel a finite distance between consecutive observations, in which case $f(\mathbf{s}(t)|\boldsymbol{\theta}(t))$ will be largest when $\mathbf{s}(t)$ is close to the recently visited location, $\mathbf{s}(t - \Delta t)$. The availability of a location is generally allowed to depend on some time-varying parameters $\boldsymbol{\theta}(t)$, which are often unknown and must be estimated from the data (e.g., Arthur et al., 1996). Availability is determined by constraints to movement or other individual-based features, such as an animal's home range, and taken to be independent of the environmental features near $\mathbf{s}(t)$.

The function g is called the resource selection function (RSF) and defines the inherent preference an individual has for the point $\mathbf{s}(t)$ independent of its availability. The RSF is therefore largest for locations that correspond to the most desirable habitat. It is important to note that the preference an individual has for a location is not necessarily the same as the value of that point, in the sense that increased use of preferential parts of the landscape do not always correspond to increased fitness or health (e.g., Hobbs and Hanley, 1990). RSFs may depend on another collection of time-varying parameters, $\boldsymbol{\phi}(t)$, which may also be unknown. Modeling space use as a product of

individual-based factors and location-based factors provides researchers with an intuitive approach to learn about the way individuals select particular regions of the landscape (Manly et al., 2002).

The proportionality constant $K(t)$ in (1.5) must be evaluated as part of any estimation procedure. In general, closed form expressions for the $K(t)$ are not available, and one must compute it numerically by evaluating (1.5) over a fine grid of values for $s^*(t)$, at every time t . Moreover, many inferential procedures require repeated evaluation of the constant of proportionality for a larger number of parameter combinations, which can make the time required to fit the movement model grow to lengths that effectively render inference inaccessible.

Many approximations have been proposed that trade exact inference for computational tractability. One approach involves discretizing the spatial domain \mathcal{D} into a finite number of cells, and modeling the number of times each cell is visited as either a Bernoulli or Poisson random variable. In this framework, one can define the RSF as a transformation of a linear combination of spatially-references covariates, in which case the RSF model can be recast as a generalized linear model (GLM; McCullagh and Nelder 1983) and standard software can be used to estimate the parameters in g (e.g., Warton and Shepherd 2010; Johnson et al. 2013). Another approach is to use an equally-spaced quadrature to directly approximate the integral in (1.6) (e.g., Warton and Shepherd 2010; Aarts et al. 2012; Brost et al. 2015). In Chapter 4, I introduce a model that builds on the existing literature by including dynamic environmental features with complex shapes in the RSF, g . In addition, I introduce a novel approximation technique used to evaluate the necessary proportionality constant that is suited to our particular class of RSFs.

1.3 Overview

In Chapter 2, I introduce the first of two models for a collection of dependent trajectories. The methods developed in Chapter 2 represent one of the first attempts made in the statistical literature to model unknown dependencies among trajectories. The model is hierarchical, and dependence is specified using a latent network that can be interpreted as social relationships among the observed individuals.

In Chapter 3, I introduce an alternative model for dependent trajectories in a continuous, rather than discrete-time framework. The model is developed using a formulation referred to in the spatial statistics literature as a “process convolution.” Process convolutions are stochastic processes with sophisticated dependence structures constructed using weighted averages of simpler, underlying processes, and have been used to model temporal, spatial, and spatio-temporal processes for over seventy years (Doob, 1944). I generalize the process convolution framework by allowing for multiple iterative stages of convolutions, which can then be used to derive processes for populations of animals with dependence both in time and across individuals. The new model construction framework, termed “process convolution chaining,” allows researchers to specify mechanistically-motivated models for dependence through a compartmentalized sequence of weighted averages.

In Chapter 4, focus shifts away from interactions among conspecifics to resource selection. I introduce a model for resource selection that depends explicitly on seasonal, or phenological, features in the landscape, and develop an approximation technique that is used to obtain approximate inference about parameters of interest. I apply the method to the study of polar bears, whose patterns of movement reflect a preference for habitat near the shifting boundary between sea ice and ocean.

Chapter 2

Dynamic Social Networks Based on Movement

2.1 Introduction

Dynamic social networks are an important topic of study among ecologists for a variety of species and ecological processes (Pinter-Wollman et al. 2014; Krause et al. 2007; Croft et al. 2008; Wey et al. 2008; Sih et al. 2009). Social networks can help explain a myriad of behavioral activities in a population, including the characteristics of animal movement. Therefore, it is common to define social networks based on directly observable behavior such as the duration of time animals spend in close proximity to one another (e.g., African elephants, *Loxodonta africana*, Goldenberg et al. 2014), discrete counts of interactions (e.g., yellow (*Papio cynocephalus*) and anubis baboons (*Papio anubis*) Franz et al. 2015), or discrete counts of close encounters (e.g., barn swallows (*Hirundo rustica erythrogaster*) Levin et al. 2015). Challenges for researchers interested in studying animal social networks include expensive data collection procedures, and potential biases due to opportunistic observation.

Killer whales (*Orcinus orca*), like many marine mammals, are complex and highly social creatures (Baird and Whitehead 2000; Williams and Lusseau 2006; Parsons et al. 2009; Pitman and Durban 2012). To better understand the behavior of killer whales, we seek to characterize their social relationships. Unfortunately, direct observation of killer whale interactions is challenging; it is not uncommon for individuals to travel 50km a day and to range over thousands of kilometers in a season (Andrews et al. 2008; Durban and Pitman 2012). Furthermore, observation of killer whales at close proximity has been found to significantly influence their movement behavior (Williams et al. 2002), which could directly affect measurements of social connectivity. In contrast, satellite tracking tags have been used to gather movement data for killer whales over several months (Andrews et al. 2008; Durban and Pitman 2012), and there is little evidence to suggest that tags alter behavior. Thus, a potential alternative to costly personal observations are telemetry data, which

contain rich movement information at the individual level, and can be collected in remote areas at a much lower cost.

The suite of models for animal telemetry data is vast and rapidly changing, including both continuous- and discrete-time approaches (see McClintock et al. 2014 for a review). Yet there are only a few models that explicitly account for interactions among individuals in the population (e.g., Morales et al. 2010; Codling and Bode 2014; Langrock et al. 2014; Russell et al. 2016). Moreover, methods are lacking that attempt to characterize pairwise connections between all members of the population. We propose a model for movement that incorporates plausible mechanistic effects on movement due to an underlying social network. Our model allows us to infer the specific characteristics of interaction in a given population and the underlying dynamic social network itself.

In our proposed discrete-time continuous-space model, we assume there exists an underlying (latent) dynamic social network among the individuals in the population. Conditional on the network characteristics and the positions of animals in the previous time step, the expected positions of individuals at the next time point are modeled jointly using a Gaussian Markov random field (GMRF) (Besag 1974; Besag and Kooperberg 1995; Rue and Held 2005b). The model is temporally Markovian for both the animal positions and the social network. In our model, the underlying social structure influences movement through two channels: an attractive effect and an alignment effect. These channels of interaction allow us to model a wide variety of behaviors, and they have a precedent for use in the context of interaction behavior (Lemasson et al. 2013). The connection between the underlying social network and position is an example of a hidden Markov model (HMM). HMMs represent a flexible class of hierarchical models popular in analyses of wildlife data (see, for example, Langrock et al. 2012) in which an observable process (in our case, position) is driven by an unobserved Markovian process (the underlying social network).

We introduce the details of our proposed method in Section 2.2. We demonstrate and assess inference from the model with simulated data in Section 2.3. In Section 2.4, we analyze data for seven killer whales tagged concurrently near the coast of the Antarctic Peninsula. Within the

tagged sample, there are three genetically distinct types of killer whale (Pitman and Ensor 2003; Morin et al. 2015) characterized by their size, coloration, and diet. The spatial distributions for each type overlap, and while strong social interaction is typical within each type, there have been no observed social associations among animals of different types. We demonstrate that inferences from our method are consistent with this history of observation. Furthermore, we find strong evidence for dynamic social connections forming and dissolving within each type, but no indication of connections between types. Finally, in Section 2.5, we discuss potential extensions for the model, including the incorporation of environmental covariates and approaches for mediating the large computational demands for the model when the study sample is large.

2.2 Methods

We propose new methodology based on a general hierarchical modeling framework that accommodates measurement, process, and parameter uncertainty (Berliner 1996). We introduce the GMRF that describes animal movement in Section 2.2.1 and describe our method for modeling the dynamic social network in Section 2.2.2. Then in Section 2.2.3, we detail how we account for the fact that telemetry data are typically measured at individual-specific, irregularly spaced times with error.

2.2.1 Position process

A GMRF is a description of a Gaussian random vector where conditional dependence between elements is specified based on a neighborhood structure (Rue and Held 2005b). For example, data occurring at regular intervals in time, or on a lattice in space, are often modeled with GMRFs because natural neighborhoods exist for each datum (e.g., the preceding measurement in time, or the four closest spatial locations). Thus, GMRFs present a natural mathematical structure for modeling trajectories of connected individuals, as they provide a way to model dependence between connected or “neighboring” individuals.

We expect that social structure among individuals will influence their movement with respect to one another. Let $\boldsymbol{\mu}_i(t)$ denote the position of individual i at time t . Assuming we know the population social structure (i.e., which individuals are socially affiliated with which other individuals), we model the movements of all individuals simultaneously using a GMRF involving two social behavioral mechanisms: one related to attraction toward the mean position of connected individuals, and the other related to alignment, or movement parallel to the paths taken by connected individuals. Although our model is flexible enough to capture attraction or repulsion, as well as alignment or anti-alignment, in most cases, we expect to infer assortative relations whereby individuals that are socially connected move “together.” For this reason, we discuss movement of connected individuals as aligned and attractive.

Attraction and alignment mechanisms are critical features of the mean positions of each individual at regular synchronous time steps. Models for locations on regular intervals have been developed by several others, including Brillinger and Stewart (1998), Jonsen et al. (2005), and Forester et al. (2007). We define the social relations in terms of a dynamic binary network $\mathbf{W}(t)$ indexed at times $t = 1, \dots, T$, where entry $w_{ij}(t) = 1$ indicates a connection between individuals i and j at time t and $w_{ij}(t) = 0$ indicates a lack thereof.

We specify a GMRF conditionally, from the perspective of a single individual at a given time. The mean position of each individual i at time t conditioned on all other individuals’ positions at time t , denoted $\boldsymbol{\mu}_{-i}(t)$, and all positions at the previous time, $\boldsymbol{\mu}(t-1)$, follows a normal distribution with mean

$$\mathbb{E}(\boldsymbol{\mu}_i(t) | \boldsymbol{\mu}_{-i}(t), \boldsymbol{\mu}(t-1), \mathbf{W}(t), \mathbf{W}(t-1), \alpha, \beta, \sigma^2, c) = \underbrace{\boldsymbol{\mu}_i(t-1) + \beta \tilde{\boldsymbol{\mu}}_i(t-1)}_{\text{attraction}} + \underbrace{\sum_{j \neq i} \alpha \frac{w_{ij}(t)}{w_{i+}^c(t)} (\boldsymbol{\mu}_j(t) - (\boldsymbol{\mu}_j(t-1) + \beta \tilde{\boldsymbol{\mu}}_j(t-1)))}_{\text{alignment}} \quad (2.1)$$

and precision

$$\text{Prec}(\boldsymbol{\mu}_i(t) | \mathbf{W}(t), \sigma^2, c) = \sigma^{-2} w_{i+}^c(t) \mathbf{I}_2. \quad (2.2)$$

Focusing on (2.1), we model the expected location of individual i as the sum of three terms: the individual's location in the previous time period, $\boldsymbol{\mu}_i(t-1)$; an attraction term capturing the tendency for the individual to move toward other individuals it is socially connected to; and an alignment term accounting for groups of interconnected individuals moving in the same general direction.

The term $\tilde{\boldsymbol{\mu}}_i(t)$, in the attraction component of (2.1), is a unit vector pointing from individual i 's position $\boldsymbol{\mu}_i(t)$ to the mean position $\bar{\boldsymbol{\mu}}_i(t)$ of all the individuals it is connected to in $\mathbf{W}(t)$ (i.e., its ego-network):

$$\bar{\boldsymbol{\mu}}_i(t) \equiv \sum_{j \neq i}^n \frac{w_{ij}(t)}{w_{i+}^c(t)} \boldsymbol{\mu}_j(t) \quad (2.3)$$

$$\tilde{\boldsymbol{\mu}}_i(t) \equiv \begin{cases} \frac{\bar{\boldsymbol{\mu}}_i(t) - \boldsymbol{\mu}_i(t)}{\|\bar{\boldsymbol{\mu}}_i(t) - \boldsymbol{\mu}_i(t)\|_2}, & \sum_{j \neq i} w_{ij}(t) > 0 \\ 0, & \sum_{j \neq i} w_{ij}(t) = 0. \end{cases} \quad (2.4)$$

The parameter β controls the strength of the attractive effect of a social connection. On average, individual i moves a distance β in the direction $\tilde{\boldsymbol{\mu}}_i(t)$ during each time step.

In the above expression, $w_{i+}^c(t)$ is the size of individual i 's ego-network at time t if the individual has at least one connection (i.e., $w_{i+}^c(t) = \sum_{j \neq i} w_{ij}(t)$), and equal to a constant $w_{i+}^c(t) = c > 0$ otherwise. We require c to be strictly positive so the precision in (2.2) is non-zero for unconnected individuals.

The alignment term in (2.1) quantifies the mean displacement in position from $t-1$ to t for only those individuals that individual i is socially connected to, and after accounting for attraction. Although the sum is over all individuals j , the social network indicators $w_{ij}(t)$ eliminate the effects of an individual's direction if it is not connected to individual i . The parameter α controls the strength of the aligning effect, with 0 corresponding to no alignment, and $\alpha \rightarrow 1$ corresponding to perfect alignment. The case $\alpha = 1$ corresponds to an intrinsic conditional autoregressive model

with an improper covariance matrix. However, we limit our consideration to $\alpha < 1$, precluding this special case.

Finally, the expression for the precision in (2.2) has the property that individuals who are more socially connected (i.e., have larger ego-networks $w_{i+}^c(t)$), have larger precision. The proportional relationship between precision and $w_{i+}^c(t)$ is required for a valid GMRF, and aligns with our intuition that, conditioned on the position of all other individuals, the movement of an individual with few or no social connections is more difficult to predict than one that experiences strong attraction and alignment toward a large group of individuals. The parameter c can be thought of as the effective size of the ego-network for an unconnected individual with regard to precision.

The specification of the model in (2.1) and (2.2) properly defines a GMRF where the elements of the precision matrix at time t are

$$Q_{ij}(t) = \begin{cases} -\alpha w_{ij}(t)\sigma^{-2}\mathbf{I}_2, & j \neq i \\ w_{i+}^c(t)\sigma^{-2}\mathbf{I}_2, & j = i. \end{cases} \quad (2.5)$$

Therefore, we can write the multivariate version of the model for $t = 2, \dots, T$ as

$$[\boldsymbol{\mu}(t)|\boldsymbol{\mu}(t-1), \boldsymbol{\theta}] = \mathbf{N}(\boldsymbol{\mu}(t-1) + \beta\tilde{\boldsymbol{\mu}}(t-1), \mathbf{Q}(t)), \quad (2.6)$$

where we have concatenated the model parameters $(\alpha, \beta, p_1, \phi, \sigma^2, c, \mathbf{W})$ into a single vector $\boldsymbol{\theta}$ (note: p_1 and ϕ are parameters associated with the dynamic network \mathbf{W} and are introduced in Section 2.2.2).

Notice that, for the joint distribution in (2.6), the attraction effect remains in the mean structure because the attraction force for an individual is toward the previous location of the individuals in the ego-network. However, the alignment effects are accounted for in the precision matrix because alignment is characterized by simultaneous movement of grouped individuals in the same direction. Figure 2.1 shows the alignment and attraction effects graphically.

The model for movement based on the normalized vector $\tilde{\boldsymbol{\mu}}_i(t)$, instead of $\bar{\boldsymbol{\mu}}_i(t) - \boldsymbol{\mu}_i(t)$, reflects a mechanistic understanding that attractive movement is often restricted by the distance an animal can reasonably travel in a given time step. We assume the maximum distance an individual is capable of moving during one time step to be approximately constant. Thus, when the gap between an individual and the center of its ego-network is large compared to its step size, an animal feeling an attractive pull will appear to take several steps of similar length in that direction.

If we had used the difference $\bar{\boldsymbol{\mu}}_i(t) - \boldsymbol{\mu}_i(t)$ instead of $\tilde{\boldsymbol{\mu}}_i(t)$ in the attraction component of (2.1), the attractive pull an individual experienced when its ego-network was far away could be far greater than the distance it was able to travel in a single time step. To see this, note that the interpretation of β in (2.6) would change to reflect the average proportion of the gap between an individual and the center of its ego-network covered during each time step. A value of $\beta = 0.5$ would imply that an animal closes half the distance between itself and the center of its ego-network, regardless of the size of that gap. In some cases, the proportional gap coverage model may be more appropriate. In our application with killer whales, it is reasonable for connections between animals to form across relatively large gaps in space relative to the distance an animal might be able to cover in a single time step. Thus, the former interpretation is the most appropriate for our application.

In (2.1) and (2.3) we define the vector $\bar{\boldsymbol{\mu}}_i(t - 1)$ using the status of the social network at time $t - 1$. Another possibility is to define $\bar{\boldsymbol{\mu}}_i(t - 1)$ using the social network at the current time t as

$$\bar{\boldsymbol{\mu}}_i(t - 1) \equiv \sum_{j \neq i}^n \frac{w_{ij}(t)}{w_{i+}^c(t)} \boldsymbol{\mu}_j(t - 1). \quad (2.7)$$

In practice, the differences that arise in the estimated social network depending on this modeling decision will only be noticeable near times when a connection status changes (i.e., whenever $\mathbf{w}(t) \neq \mathbf{w}(t - 1)$). Hence, when the estimated social network is slowly varying, like the one we observe in our application, we expect that these two definitions will result in essentially identical inference. However, for applications when the frequency of changes in social connections is high

relative to the scale at which telemetry observations are made, the impact of the decision of how to define $\bar{\mu}$ may be more significant.

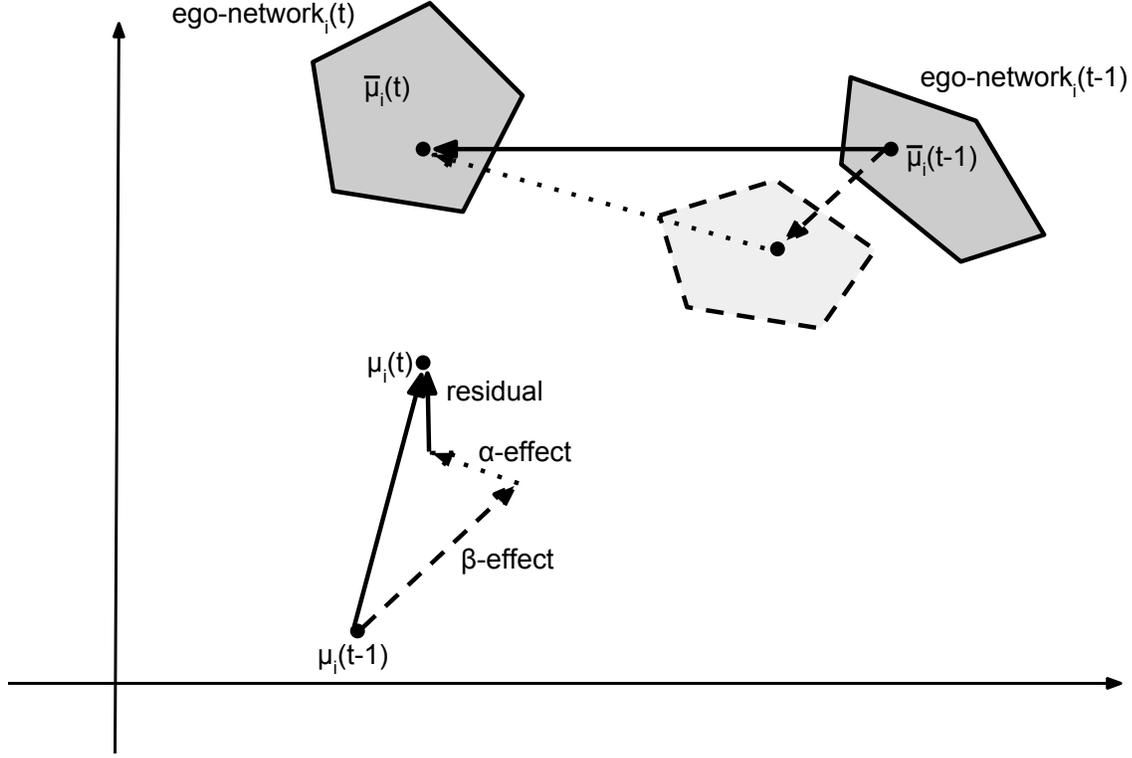


Figure 2.1: This schematic illustrates the two channels through which the dynamic social network influences movement. The dashed lines represent where the ego-network of individual i would be expected to be at time t under attraction alone, and the parallel dotted lines represent the alignment between individual i and the average of the differences $\mu_j(t) - (\mu_j(t-1) + \beta\tilde{\mu}_j(t-1))$.

2.2.2 Dynamic social network

We model the dynamic process that gives rise to $\mathbf{W}(t)$ as a collection of pairwise independent Bernoulli random variables with a Markovian dependence in time, where

$$w_{ij}(1) \sim \text{Bern}(p_1) \quad (2.8)$$

$$w_{ij}(t)|w_{ij}(t-1) \sim \begin{cases} \text{Bern}(p_{1|0}), & w_{ij}(t-1) = 0 \\ \text{Bern}(p_{1|1}), & w_{ij}(t-1) = 1 \end{cases} \quad t = 2, \dots, T. \quad (2.9)$$

The parameter p_1 is the probability of a social connection between any two individuals at time $t = 1$, $p_{1|0}$ is the conditional probability that a pair of individuals who are not connected at $t - 1$ become connected at time t , and $p_{1|1}$ is the conditional probability that a pair connected at time $t - 1$ remain connected at time t . Thus, our model for μ can be thought of as a HMM, where the latent social network \mathbf{W} takes on the role of the hidden Markovian process. Though the model for the dynamic social network could be used exactly as specified in (2.8) and (2.9), we make two refinements that reduce the number of parameters we are required to estimate, and facilitate solicitation of priors.

First, we take advantage of the fact that, in many cases, it is reasonable to assume the mean density of a study population's dynamic social network remains constant in time. This is equivalent to requiring that the stationary distribution of the Markov process governing the overall network density match the expected density at time $t = 1$. Recall that we model the conditional distributions of the edges, $\mathbf{w}(t)|\mathbf{w}(t-1)$ as independent Bernoulli random variables. Thus, the expected density of the network at time t is equal to the marginal probability of an edge between any two vertices, $\Pr(w_{ij}(t) = 1)$. The Markov process controlling network density is therefore the same as the process for the sequence of social connections \mathbf{w}_{ij} . Requiring that the initial density, p_1 , match the stationary distribution of the Markov process is equivalent to the condition

$$p_1 = \frac{p_{1|0}}{p_{1|0} + 1 - p_{1|1}}. \quad (2.10)$$

Condition (2.10) implicitly reduces the number of parameters to be estimated from three to two.

The second refinement we make is a reparameterization that allows for more intuitive interpretation of model parameters, and hence, facilitates the solicitation of priors. We define a new variable, ϕ , that controls the temporal stability of the dynamic network via

$$p_{1|0} \equiv (1 - \phi)p_1, \quad (2.11)$$

which implies we can write $p_{1|1} = 1 - (1 - \phi)(1 - p_1)$. As ϕ varies from 0 to 1, the social network transitions smoothly from complete temporal independence, to complete temporal dependence (i.e., a static network where no edges form or dissolve in time). This can be expressed mathematically as $\lim_{\phi \rightarrow 0} p_{1|0} = \lim_{\phi \rightarrow 0} p_{1|1} = p_1$ and $\lim_{\phi \rightarrow 1} (1 - p_{1|0}) = \lim_{\phi \rightarrow 1} p_{1|1} = 1$. Thus, ϕ can be thought of as a measure of the temporal range of dependence in the network. Under the parameterization using p_1 and ϕ , researchers can construct priors for the network density and stability independently of one another.

2.2.3 Measurement error and time alignment

Our model can be used to make inference about the posterior distribution of the model parameters θ conditioned on the mean position process μ (denoted $[\theta|\mu]$). However, in practice, we are rarely able to observe μ directly. Rather, we observe noisy measurements of position at asynchronous, irregularly occurring times, which we denote s , and the inference we wish to make is for the posterior distribution conditioned on observed data, not μ . Let $s_i(\tau_i)$ denote the observed position of individual i at time τ_i , and $[s|\mu]$ the joint density of all observed locations conditioned on the unobserved processes μ . The top level of our hierarchical model provides a connection between the locations $\mu_i(t)$, which occur at regular synchronized times, and the observations $s_i(t)$.

We could obtain the desired posterior distribution by evaluating the integral

$$[\theta|s] = \int [\theta|\mu, s] [\mu|s] d\mu, \quad (2.12)$$

using Markov chain Monte Carlo (MCMC), provided we could sample from the distribution $[\mu|s]$. Unfortunately, because of the inherent complexities in the irregular, asynchronous observation times and the high dimensionality of the vector μ , sampling from this distribution becomes computationally infeasible when a study population contains more than a few individuals and a few dozen observation times per individual. We address this issue by making use of a multiple imputation procedure employed by Hooten et al. (2010) and Hanks et al. (2015b, 2011), paired with a continuous-time correlated random walk model from Johnson et al. (2008a). Multiple imputation

offers a computationally efficient way to account for asynchronous, noisy position measurements while still permitting us to use a discrete-time, step-aligned structure for movement informed by a dynamic social network. We outline the procedure briefly below, and refer the reader to Hooten et al. (2010) and Hanks et al. (2011) for further details.

The premise of the multiple imputation strategy assumes the existence of a distribution that is very similar to $[\boldsymbol{\mu}|\mathbf{s}]$ from which we can sample paths easily. If we can define such a distribution, which we call $[\boldsymbol{\mu}^*|\mathbf{s}]$, then we can closely approximate the integral in (2.12) by

$$[\boldsymbol{\theta}|\mathbf{s}] \approx \int [\boldsymbol{\theta}|\boldsymbol{\mu} = \boldsymbol{\mu}^*] [\boldsymbol{\mu}^*|\mathbf{s}] d\boldsymbol{\mu}^*. \quad (2.13)$$

We can evaluate the integral in (2.13) up to a constant of proportionality by drawing a realization from $[\boldsymbol{\mu}^*|\mathbf{s}]$ at every iteration of our MCMC algorithm, and updating model parameters $\boldsymbol{\theta}$ conditioned on the realization.

Johnson et al. (2008a) introduced a continuous-time correlated random walk model for movement with measurement error that relies on an Ornstein-Uhlenbeck process for velocity, and treats the observed paths for each individual as conditionally independent (i.e., $[\mathbf{s}_i|\boldsymbol{\mu}^*] = [\mathbf{s}_i|\boldsymbol{\mu}_i^*]$). Continuing with the same model, Johnson et al. (2011) provided an approach for sampling from the posterior predictive path, $[\boldsymbol{\mu}_i^*|\mathbf{s}_i]$, which we use to evaluate the integral in (2.13).

We approximate the desired posterior using the following two-step procedure:

1. Draw K different realizations from $[\boldsymbol{\mu}^*|\mathbf{s}]$ using the R package `crawl` (Johnson et al. 2008a).
2. At each iteration of the MCMC sampler, draw one of the K samples and condition on $\boldsymbol{\mu}^*$ for parameter updates.

Choosing too small a value for K will result in inference for the social network that does not properly account for the uncertainty in $\boldsymbol{\mu}$ arising due to measurement error and temporal asynchronicity, and can potentially be biased depending on the particular draws from $[\boldsymbol{\mu}^*|\mathbf{s}]$. In practice, we found a sufficiently large K in our application to be on the order of 50, as parameter estimates were

essentially unchanged for larger K . By making use of the two-stage sequential procedure, we are fitting a close approximation to the full Bayesian hierarchical model.

2.2.4 Priors

To demonstrate the value of our model when little is known *a priori* about the social ties in a study population, we specify diffuse priors for most parameters in both the simulation and application. We select conjugate parametric families whenever possible. The priors used in our simulation and application are shown in the right columns of Tables 2.1 and 2.2. While more informative priors could be used when expert knowledge is available, we found most parameters to be insensitive to the choice of hyperparameters. The one exception is the network stability parameter ϕ (see Section 2.2.2). The stability of the network determines the range of temporal dependence in the dynamic social network. Similar to analogous range parameters in the geostatistical setting (see, for example, Chapter 6 of Gelfand et al. 2010), ϕ can prove difficult to estimate from the data. In our application (Section 2.4), we used an informative prior that implies a strongly stable network because we expected the social network to change slowly relative to the time scale at which the telemetry data were gathered.

2.3 Simulation

The primary parameters of scientific interest are in the network \mathbf{W} . Thus, we evaluate the quality of our model by assessing its ability to recover the network. A baseline model for comparison is one using only proximity as a criterion for social connectivity. We consider the proximity-based network defined by

$$W_{ij}^R(t) = I_{\|\mu_i - \mu_j\|_2 < R}. \quad (2.14)$$

Though it does not explicitly incorporate the behaviors of attraction and alignment, defining the network using (2.14) is computationally cheap and closely mirrors the way some data are collected in the field (Levin et al. 2015; Goldenberg et al. 2014). The proximity-based approach therefore

represents a viable alternative against which we can compare our model. However, failing to consider attraction and alignment effects, as well as temporal stability in a dynamic social network can lead to spurious associations that arise when two unconnected individuals happen to pass each other by chance. Our simulation shows that our model is able to avoid such pitfalls.

In the following simulation, we generate directly from the proposed process model and fit the model using paths μ . We use parameter values (shown in Table 2.1) that generate paths closely resembling the data in our application for killer whales. Details of the methods we used to fit the model, and the R code used to produce this simulation study is provided in Appendix A. We used the posterior mean of \mathbf{W} as a summary of the network, and investigated a variety of radii R with the proximity-based network, \mathbf{W}^R , to define a suite of alternatives. Because we know the true mean density of the network, p_1 , we select the proximity-based network for which the radius yields a mean density as close as possible to the true value. Choosing a radius that recovers the true mean density would not generally be possible, thus, we compared our model to a particularly favorable proximity-based alternative. However, we found that proximity alone provides a poor estimate of the true network relative to our proposed dynamic network model.

Figure 2.2 shows estimates of \mathbf{W} for a random selection of pairs. Included on each plot are the true network (dashed), the posterior mean from the model fit (solid), and the proximity-based estimate (dotted). Although the posterior mode of $w_{ij}(t)$ would be a natural choice for a prediction of the true dynamic social network, we plot the posterior mean because it provides a visual description about uncertainty in our predicted network. For example, posterior means of $w_{ij}(t)$ near 0.5 indicate larger uncertainty about the true connection status of individuals i and j at time t than posterior means near 0 or 1. The pairs 1-5 (top left) and 1-6 (bottom right), show how the proximity-based network can both find spurious connections, and fail to identify connected behavior when it takes place over too large a distance. Table 2.1 shows 95% credible intervals for all parameters in the model except \mathbf{W} . All credible intervals capture the true parameter values, except those for ϕ . We observed moderate systematic bias in the posterior distribution of ϕ toward zero, however posterior inferences for \mathbf{W} were robust despite the bias in ϕ . In most applications

we expect that the primary questions of scientific interest concern the network \mathbf{W} , and ϕ can be treated as a nuisance parameter.

Table 2.1: Marginal posterior medians and 95% credible intervals for model parameters. True values for the simulation were chosen to yield plausible movement paths. The right column describes the prior distributions used.

parameter	true	posterior		prior
		median	(2.5%, 97.5%)	density
α	0.9	0.92	(0.77, 0.96)	Unif(-1, 1)
β	0.5	0.46	(0.37, 0.55)	N(0, 10 ³)
p_1	0.2	0.15	(0.0096, 0.21)	Unif(0, 1)
ϕ	0.95	0.84	(0.78, 0.90)	Beta(17.2, 1.5)
c	0.33	0.30	(0.24, 2.36)	IG(1.5, 3.5)
σ^2	1	0.92	(0.74, 7.91)	IG(10 ⁻¹ , 10 ⁻³)

Any study of a social network is ultimately based on a definition for connection specific to the population of interest. Thus, it is incorrect to say that the proximity-based network fails to capture the true network. Rather, the proximity-based network simply does a poorer job describing the connections that influence movement than the network based on our proposed model. It is impossible to perfectly define a given social network, but if there is reason to believe that a study population might exhibit the commonly observed behaviors of attraction and alignment, then our model offers a way to study it. We have shown that ignoring these mechanisms can result in misleading inference.

2.4 Killer whales

We analyzed observed data for seven individuals near the Antarctic Peninsula over the course of a week in February 2013 (for a description of the tags and study area see Durban and Pitman 2012; Andrews et al. 2008). Geographic positions were measured using Argos transmitter tags. Within the study area, three genetically distinct types of killer whales (termed A, B1, and B2) are known to exist (Durban et al. 2017; Morin et al. 2015; Pitman and Ensor 2003) and are characterized primarily by their size, coloration, and diet. Type A killer whales are the largest and feed primarily

on Antarctic minke whales (*Balaenoptera bonaerensis*) (Pitman and Ensor 2003). Of the two type B killer whales, B1 is larger and is distinguished by a diet consisting primarily of ice seals (Durban and Pitman 2012). Finally, type B2 killer whales are distinguished by an observed diet of penguins and likely also fish during deep dives (personal communication J. W. Durban 2015; Pitman and Durban 2010). Although all types of killer whales have been observed exhibiting social behavior within type, association between types has not been observed. The study sample of seven tagged whales consisted of three whales of Type A, one of type B1 and three of type B2.

Credible intervals for all parameters except the network \mathbf{W} are shown in Table 2.2. When we examine the mean step size across all individuals and times, we found it to be several times larger than the contribution of attraction, suggesting only a moderate attractive effect. The fit also suggests a strong alignment effect evidenced by the posterior median for α near 1. Therefore, we conclude that connectivity in this population of killer whales manifests itself predominantly as movement in parallel, with some additional tendency for connected individuals to move toward one another.

Table 2.2: Marginal posterior medians and 95% credible intervals for model parameters when fit to the killer whale tagging data. The values reflect a strong alignment effect (α), weak attraction effect (β), and a stable (ϕ), sparse (p_1) social network. The right column describes the prior distributions used.

parameter	posterior		prior
	median	(2.5%, 97.5%)	density
α	0.88	(0.40, 0.94)	Unif(-1, 1)
β	0.022	(0.012, 0.030)	N(0, 10^3)
p_1	0.11	(0.005, 0.20)	Unif(0, 1)
ϕ	0.95	(0.90, 0.98)	Beta(100, $\frac{100}{9}$)
c	0.35	(0.24, 2.87)	IG(1.5, 3.5)
σ	0.0033	(0.0026, 0.025)	IG(10^{-1} , 10^{-3})

The credible intervals for p_1 and ϕ suggest that the network is very stable, but also fairly sparse. Enduring connections are directly visible in Figure 2.3. The left column shows all pairwise dynamics between the three individuals of type B2, and the right column shows all pairwise dynamics between the three individuals of type A. All three individuals of type B2 show strong connection

through the study period and, in fact, all three of these individuals moved as a group during this time. The only social interaction involving individuals in type A occurred during the first few days of the study period between individuals 5 and 6. There was strong evidence for complete independence between all individuals not in the same type (see Figures A.1 and A.2 in the appendix), consistent with expert knowledge. Of the 15 inter-type connections in \mathbf{W} , there were no posterior means above 0.5 at any time in the study period. A visualization of the movement and estimated social connections between these individuals can be found in Appendix A.

As in the simulation (Section 2.3), we investigated an alternative definition for the social network, based purely on proximity, given by (2.14). To account for the uncertainty in μ , we constructed the proximity-based network defined by a particular choice of R for each of the K draws from $[\mu^*|s]$ used for multiple imputation (see Section 2.2.3), and averaged across these networks. The primary means of communication at a distance between killer whales is acoustic signaling. Therefore, we selected values for R based on the typical distances across which killer whales are known to communicate acoustically. Miller (2006) observed killer whales in the Pacific Northwest and estimated signals between individuals were detectable at distances of 5-15km. This range is consistent with expert knowledge about the killer whales in our study region. We inspected the corresponding dynamic social networks for radii between 5-15km and found little variation in the resulting networks. Figures 2.3, A.1, and A.2 show the proximity-based network for $R = 10\text{km}$.

While we observed some similarities in the proximity-based and model-based networks, there are several notable discrepancies. For instance, all proximity-based networks for radii between 5-15km included numerous connections between individuals of different types (Figures A.1 and A.2). The presence of inter-type connections conflicts with expert knowledge that killer whales of differing types do not form social bonds, suggesting that the proximity-based network may be defining spurious social ties. Moreover, because the proximity-based network does not account for temporal stability in social connections, we observe instances of implausibly rapid oscillation in connection status (Figure A.2). The proximity-based networks and our model-based network provide similar inference for within-type ties (Figure 2.3), but our model-based approach also

provides rigorous uncertainty estimates. A researcher might arguably make an *ad hoc* adjustment to the network and simply discard all inter-type connections on the basis of prior knowledge, thereby arriving at the same conclusion regardless of which rule was used to define the social network. However, the feasibility of such an approach is unique to this study for two reasons. First, supplementary individual-level information, such as killer whale type, is often unavailable. Second, in many populations, the relationship among covariates and social connections is largely unknown, prohibiting covariate-based pruning of the proximity network.

2.5 Discussion

Existing methods for measuring and studying dynamic social networks in animal populations typically involve *ad hoc* definitions for connectivity based on direct observation of study populations. Our model offers a flexible, but interpretable, hierarchical framework that allows researchers to rigorously study dynamic social networks informed by relatively inexpensive telemetry data. Moreover, our proposed model can easily be coupled with existing analyses on dynamic networks. Fundamentally, the study of dynamic social networks often begins with descriptive statistics such as network density, node degree, transitivity, and others (Pinter-Wollman et al. 2014). All of these common summaries can be obtained as derived quantities in our Bayesian framework with estimates of uncertainty. More sophisticated models for dynamic networks (e.g., Sarkar and Moore 2006; Durante and Dunson 2014; Sewell and Chen 2015) can take the posterior mode of \mathbf{W} as input, or be incorporated as part of a larger hierarchical modeling structure.

We have shown, through simulation, that our proposed model is able to capture information about a population’s social structure in a way that a simplistic proximity-based measure cannot, both by avoiding spurious connections and detecting interactions that occur over large distances. Through an application on killer whale movement, we showed that the model captures connections consistent with expert knowledge based on non-quantitative observation, and can therefore be relied upon to deliver credible and practical inference.

When auxiliary covariates are available on the individuals, the proposed model can be extended to include such data. A potential generalization is to allow the spatial covariates to influence the mean position process of each individual, $\boldsymbol{\mu}_i(t)$, linearly. If we denote the matrix containing spatial covariates $\mathbf{X}_C(t)$, we arrive at a familiar additive form

$$[\boldsymbol{\mu}(t) | \boldsymbol{\mu}(t-1), \boldsymbol{\theta}, \boldsymbol{\gamma}] = \mathbf{N} \left(\underbrace{\mathbf{X}_C(t-1)\boldsymbol{\gamma}}_{\text{covariate effect}} + \underbrace{\mathbf{X}_W(t-1)\boldsymbol{\beta}}_{\text{attraction}}, \underbrace{\mathbf{Q}(t)}_{\text{alignment}} \right) \quad (2.15)$$

where $\mathbf{X}_W(t-1)$ is defined as the matrix with columns $\boldsymbol{\mu}(t-1)$ and $\tilde{\boldsymbol{\mu}}(t-1)$, and $\boldsymbol{\beta} \equiv (1, \beta)'$.

One limitation of our model is that it is computationally intensive for large study samples. The number of parameters in our model grows at a rate of $\binom{n}{2}T$ as the number of individuals, n , and number of time points, T , increase. The most dominant factor in computation time is typically n , and when the number of individuals grows beyond a few dozen, fitting the model on a laptop computer using MCMC becomes infeasible. One way to decrease the computational cost of fitting the model is to introduce additional structure on \mathbf{W} . We suggest two possible approaches.

The first way to introduce structure to \mathbf{W} is to define a maximum radius of interaction, R_{\max} , beyond which the probability of a social connection is zero. For example, the radius might be chosen to be the maximum distance at which two individuals are able to detect one another. After modifying the conditional distribution of $w_{ij}(t)$ based on R_{\max} , it is no longer necessary to update all $w_{ij}(t)$ in each step of the MCMC algorithm, only those for which $\|\boldsymbol{\mu}_i(t) - \boldsymbol{\mu}_j(t)\| < R_{\max}$. If R_{\max} is small relative to the spatial extent of the trajectories, this proximity-based modification offers a substantial reduction in the computational cost of fitting the model. This idea is somewhat related to covariance tapering for spatially referenced Gaussian random variables. Furrer et al. (2006) decrease the computational burden of interpolating, or kriging, by deliberately introducing zeros into the covariance matrix. In our setting, we would introduce zeros into the precision matrix.

Another way to alleviate the computational burden is to enforce structure directly on \mathbf{W} to reduce number of parameters in the model. For instance, it may be reasonable to assume that the social connections in a given population form as complete subgroups or cliques. In this case, the

network describes a clustering process with only nT parameters. Though motivated by straightforward mechanisms, both of these approaches to reducing the computational burden are non-trivial to implement. In the first case, setting a maximum radius of interaction complicates the enforcement of stability in the density of the network (introduced in Section 2.2.2) and offers modest or no gains when R_{\max} is large relative to the spatial extent of the individual paths. In the second case, updating the clustering process \mathbf{W} requires the exploration of a very large space (of cardinality equal to the Bell number B_n) for every t .

Although further developments are required before data for large populations of individuals can be accommodated, our framework provides a strong foundation for modeling relationships between movement and social networks.

2.6 Declaration of support

Killer whale tagging was conducted under permit #14097 from the National Marine Fisheries Service and Antarctic Conservation Act permit #2009-013. Shipboard tagging operations were supported by Lindblad Expeditions and the National Geographic Society, and by an NSF rapid grant to Ari Friedlaender. Robert Pitman helped with tag deployments and identification of killer whale types in the field. This material is based in part upon research supported by NOAA ACK 188000 and NSF SES-1461495 and DMS-1614392.

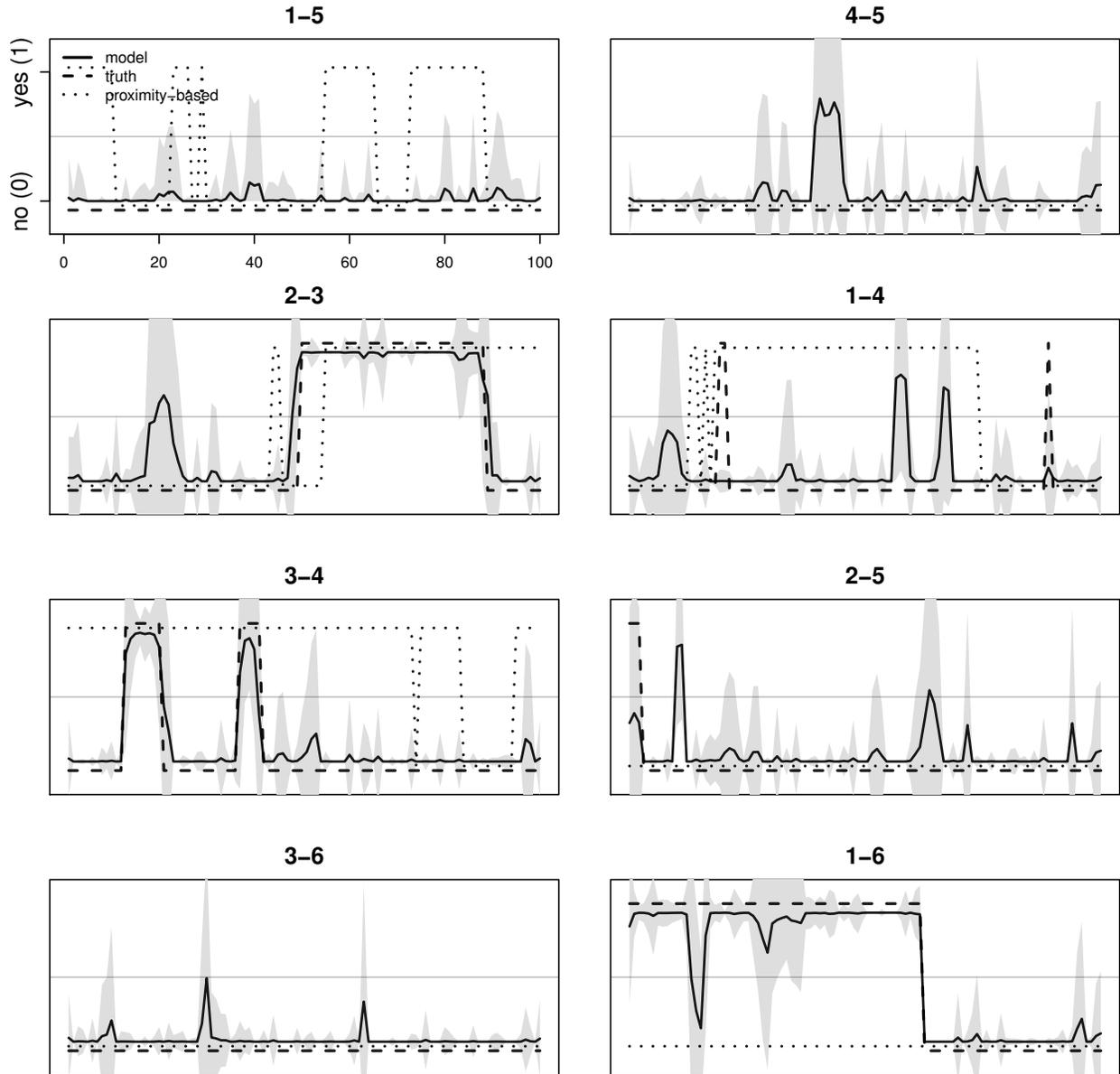


Figure 2.2: A subset of the complete estimated dynamic network for the simulated data on six individuals. The titles correspond to the i^{th} and j^{th} individuals in $w_{ij}(t)$. The dashed line is the true network, the solid line is the posterior mean from the proposed Bayesian model, and the gray region represents one standard deviation above and below the posterior mean. The dotted line shows the network defined by \mathbf{W}^R , where individuals are deemed connected whenever they are separated by a distance less than R (see Section 2.3). (Note: The lines are offset slightly near 0 and 1 for visual clarity.)

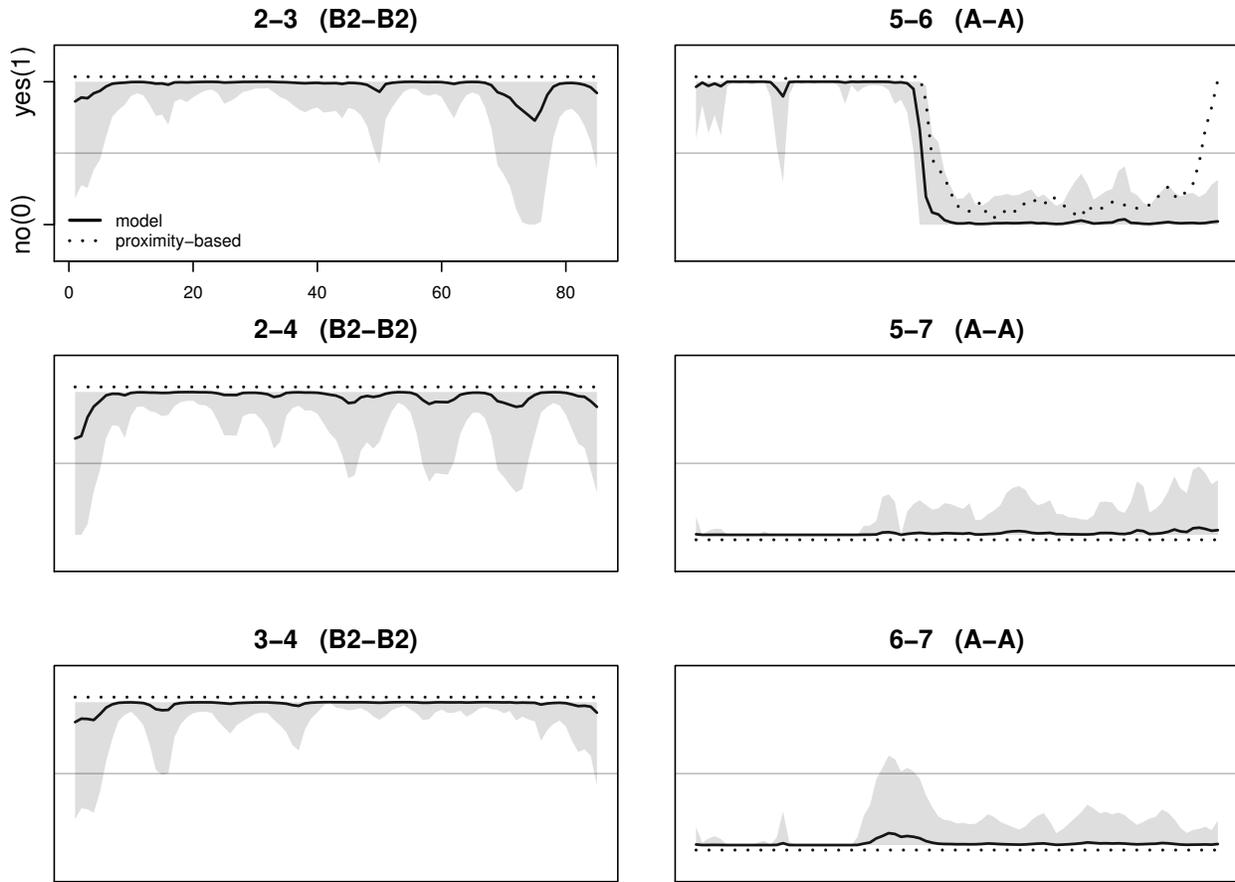


Figure 2.3: A selection of the $\binom{7}{2} = 21$ possible pairs of individuals in the killer whale study sample. The left column is all pairs of killer whales of type B2 (labeled 2, 3, 4), and the right column is all pairs of killer whales of type A (labeled 5, 6, 7). The solid line in each plot shows the posterior mean for w_{ij} and the gray region represents one standard deviation above and below the posterior mean. The dotted line shows the network defined by \mathbf{W}^R , where individuals are deemed connected whenever they are separated by a distance less than R . (Note: The lines are offset slightly near 0 and 1 for visual clarity.)

Chapter 3

Process Convolution Approaches for Modeling

Interacting Trajectories

3.1 Introduction

Models for continuous random processes using kernel convolutions are known in much of the spatial and spatio-temporal statistical literature as “process convolutions” (e.g., Higdon, 2002; Calder, 2007; Bolin and Lindgren, 2013). Also referred to as spatial moving averages (e.g., Cressie and 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 and Ver Hoef, 1996). One application area for process convolutions that has driven much of their recent theoretical development is for random processes taking place on stream networks (Ver Hoef et al., 2006). Stream networks occupy sub-manifolds of physical space on which specialized measures of distance often complicate enforcement of non-negative 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, and scientific questions of interest are often focused on which exogenous environmental factors drive movement (Potts et al., 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 prediction of the true paths difficult (for a review, see McClintock et al., 2014; Hooten et al., 2017a). 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 et al., 2015; Buderman et al., 2016). Irregularity in the frequency of telemetry observations can complicate the implementation of discrete-time models (although see Scharf et al., 2017, for one solution using a multiple-imputation procedure) 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., Williams and Lusseau, 2006; Goldenberg et al., 2014; Scharf et al., 2016). 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., Haydon et al., 2008; Codling and Bode, 2014; Langrock et al., 2014; Russell et al., 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 to 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 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 Gaussian process that captures the combination of effects. It is important to note that, while the kernels responsible for inducing temporal and path-wise 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 3.2.1 and 3.2.2 and construct our joint model for movement in Sections 3.2.3 and 3.2.4. After discussing the implementation details in Section 3.3, we demonstrate path reconstruction in a simulation study in Section 3.4. We apply our model to the study of telemetry data arising from the movement of killer whales in Section 3.5 and close with a summary of our findings and future directions.

3.2 Methods

3.2.1 A multiple-kernel convolution framework

We first outline a new flexible framework for the development of Gaussian process 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 spatio-temporal statistical literature, as “process convolutions”

(e.g., Higdon, 2002; Calder, 2007; Bolin and Lindgren, 2013). 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

$$\mu(t) = \int_{\mathcal{T}} h(t, \tau) dB(\tau). \quad (3.1)$$

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

$$\text{Cov}(\mu(t), \mu(t^*)) = \int_{\mathcal{T}} \int_{\mathcal{T}} h(t, \tau) h(t^*, \tau^*) \text{Cov}(dB(\tau), dB(\tau^*)). \quad (3.2)$$

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

$$\text{Cov}(\mu(t), \mu(t^*)) = \int_{\mathcal{T}} h(t, \tau) h(t^*, \tau) d\tau \quad (3.3)$$

(although see Nychka et al. (2015) for an example of a spatial process convolution where $dB(\cdot)$ has covariance specified through a Gaussian Markov random field). A process convolution with kernel function $h(\cdot, \cdot)$, as in (3.1), 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 (3.3) is guaranteed to be non-negative 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 and Peterson, 2010), although this is not necessary for non-negative definiteness. In the cases where an analytic solution to (3.3) 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, 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 non-stationary 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 et al. (2008a). 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

$$\mu(t) = \int_{\mathcal{T}} h(t, \tau) \left(\int_{\mathcal{T}} h^{(bm)}(\tau, \tilde{\tau}) dB(\tilde{\tau}) \right) d\tau. \quad (3.4)$$

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 $dB(0) \sim N(0, \sigma_0^2)$. Hooten and Johnson (2017) provided several examples of possible kernel functions and demonstrated a computationally efficient procedure to incorporate temporal non-stationarity 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 (3.4) 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 multi-stage 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 $h^{(1)}, \dots, h^{(L)}$ as a “process convolution chain” (PCC) and write it in its expanded form as

$$\begin{aligned}
\mu^{(1)}(\tau_1) &= \int_{\mathcal{T}} h^{(1)}(\tau_1, \tau_0) dB(\tau_0) \\
\mu^{(2)}(\tau_2) &= \int_{\mathcal{T}} h^{(2)}(\tau_2, \tau_1) d\mu^{(1)}(\tau_1) \\
&\vdots \\
\mu^{(L)}(\tau_L) &= \int_{\mathcal{T}} h^{(L)}(\tau_L, \tau_{L-1}) d\mu^{(L-1)}(\tau_{L-1}).
\end{aligned} \tag{3.5}$$

The underline in stages $2 \leq l \leq L$ (i.e., $\mu^{(l)}(\tau_l)$) 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 analogous superscripts for the covariance function of a given process (e.g., $\text{Cov}(\mu^{(l)}(t), \mu^{(l)}(t^*)) = C^{(l)}(t, t^*)$). A collapsed version of a PCC using a single effective kernel and the white noise process $dB(\cdot)$ can be written as

$$\mu^{(l)}(\tau_l) = \int_{\mathcal{T}} h^{(l)}(\tau_l, \tau_0) dB(\tau_0) \tag{3.6}$$

if we define the effective kernel as

$$h^{(l)}(\tau_l, \tau_0) \equiv \int_{\mathcal{T}} \dots \int_{\mathcal{T}} h^{(l)}(\tau_l, \tau_{l-1}) \dots h^{(1)}(\tau_1, \tau_0) d\tau_1 \dots d\tau_{l-1}. \tag{3.7}$$

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

$$\int_{\mathcal{T}} \int_{\mathcal{T}} h^{(l_1)}(t, \tau) h^{(l_2)}(\tau, \tau_0) dB(\tau_0) \neq \int_{\mathcal{T}} \int_{\mathcal{T}} h^{(l_2)}(t, \tau) h^{(l_1)}(\tau, \tau_0) dB(\tau_0). \quad (3.8)$$

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 and Johnson, 2017, Figure 2, row 3). Another specific case of PCCs are so-called $(k - 1)$ -fold integrated Wiener processes (Shepp, 1966; Wecker and Ansley, 1983; Rue and Held, 2005a), which use $L = k - 1$ kernels of the form $h^{(l)}(t, \tau) = (t - \tau)/(l - 1)$ and allow one to model GPs with exactly m continuous derivatives. The multi-stage decomposition is similar in spirit to the way random processes with multiple dependence scales are decomposed additively in multi-resolution processes (Higdon, 2002; Nychka et al., 2015; Katzfuss, 2017). The difference is the convolution chain approach combines model components using convolutions, rather than additively.

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 3.2.3.

3.2.2 Finite representation

The joint distribution of $\tilde{\boldsymbol{\mu}} \equiv (\tilde{\mu}(t_1), \dots, \tilde{\mu}(t_n))'$ for $\mathbf{t} \equiv (t_1, \dots, t_n)' \subseteq \mathcal{T}$ is mean-zero Gaussian with covariance function $\tilde{C}(t, t^*) = \int_{\mathcal{T}} \tilde{h}(t, \tau) \tilde{h}(t^*, \tau) d\tau$ 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 $\boldsymbol{\tau} \equiv (\tau_0, \dots, \tau_m)'$ from the domain \mathcal{T} ,

and evaluating the discrete sum $\sum_{i=2}^m \tilde{h}(t, \tau_i) \tilde{h}(t^*, \tau_i) \Delta\tau_i$, where $\Delta\tau_i = \tau_i - \tau_{i-1}$. In the discretized setting, the process convolution may be written as a matrix product $\tilde{\boldsymbol{\mu}} = \tilde{\mathbf{H}}\boldsymbol{\varepsilon}_0$, where the j^{th} row of $\tilde{\mathbf{H}}_{n \times m}$ is the function $h(t_j, \tau)$ evaluated at all $\tau \in \boldsymbol{\tau}$, and $\boldsymbol{\varepsilon}_0 \equiv (\varepsilon_0(\tau_1), \dots, \varepsilon_0(\tau_m))'$ with each $\varepsilon_0(\tau_i) \sim \text{N}(0, \Delta\tau_i)$. 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

$$\tilde{\boldsymbol{\mu}} \sim \text{N}\left(\mathbf{0}, \Delta\tau \tilde{\mathbf{H}}\tilde{\mathbf{H}}'\right). \quad (3.9)$$

As the density of the grid grows to infinity ($m \rightarrow \infty$), the outer product in (3.9) 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 computational efficiency and provide an adjustable level of implicit regularization (Wikle, 2010).

3.2.3 A process convolution chain for dependent movement

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 inter-path dependence that arises is based on a weighted, undirected, latent social network (e.g., Goldenberg et al., 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 $i, j \in \{1, \dots, p\}$. We propose a ‘‘social’’ smoothing kernel of the form

$$h_{ij}^{(soc)}(\tau_{soc}, \tau) \equiv \mathbf{1}_{\{\tau=\tau_{soc}\}} \frac{w_{ij}(\tau_{soc})}{|w_{i\cdot}(\tau_{soc})|}, \quad (3.10)$$

$$|w_{i\cdot}(\tau_{soc})| \equiv \sum_{j=1}^p w_{ij}(\tau_{soc}). \quad (3.11)$$

We write the kernel in (3.10) 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

$$\mu_1^{(soc)}(\tau_{soc}) \equiv \sum_{j=1}^2 \int_{\mathcal{T}} h_{1j}^{(soc)}(\tau_{soc}, \tau) \mu_j(\tau) d\tau \quad (3.12)$$

$$= h_{11}(\tau_{soc}) \mu_1(\tau_{soc}) + h_{12}(\tau_{soc}) \mu_2(\tau_{soc}) \quad (3.13)$$

$$= \frac{\mu_1(\tau_{soc})}{1 + |w_{12}(\tau_{soc})|} + \frac{w_{12}(\tau_{soc}) \mu_2(\tau_{soc})}{1 + |w_{12}(\tau_{soc})|}. \quad (3.14)$$

From (3.13) and (3.14), 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 a scaled version of 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 3.1.

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 (3.16), but we emphasize that PCCs, with suitably constructed kernels, offer a flexible way to specify models for a wide range of possible behavioral 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 (3.16), 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 convolving $h_{ij}^{(bm)}(t, \tau) \equiv \mathbf{1}_{\{\tau \leq t\}} \mathbf{1}_{\{i=j\}}$ with $dB(\cdot)$ 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 (3.10) 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 whom $w_{ij}(t)$ 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.

Inertial smoothing

Marginally, the individual paths μ_i^{soc} generated from the two-stage PCC constructed from $h^{(bm)}$ and $h^{(soc)}$ are each an instance of Brownian motion. However, Brownian motion is unsuitable for 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 is inconsistent with the physical laws governing the mechanics of massive bodies (e.g., Feynman et al., 1963).

Several process convolutions have been explored that impart specific smoothness properties on a GP. For example, Shepp (1966), Wecker and Ansley (1983), and Rue and Held (2005a) discuss applications of $(k - 1)$ -fold integrated Wiener processes, which ensure the existence of $k - 1$ continuous derivatives. Johnson et al. (2008a) 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, $h_{ij}^{(inl)}(t, \tau) \equiv \frac{|\tau_{inl} - \tau_{soc}|}{\phi_{inl}} K_1(|\tau_{inl} - \tau_{soc}| / \phi_{inl}) \mathbf{1}_{\{i=j\}}$, where K_1 is a modified Bessel function, is visible in the bottom plot of Figure 3.1. As we did for the social smoothing kernel, in what follows we suppress the superfluous index and write $h^{(inl)}(\cdot, \cdot)$.

The full three-stage process convolution chain

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

$$\mu_i^{(bm)}(\tau_{bm}) = \int_{\mathcal{T}} h^{(bm)}(\tau_{bm}, \tau_0) dB_i(\tau_0) \quad (3.15)$$

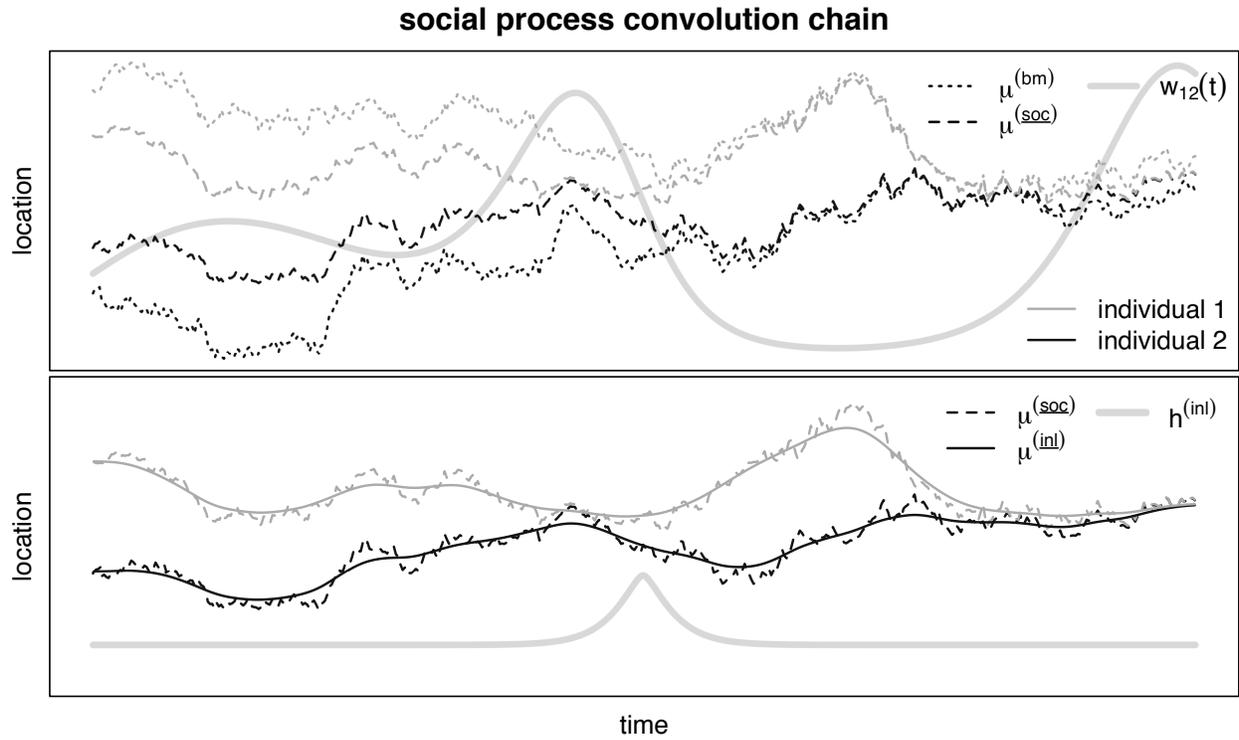


Figure 3.1: An example of a one dimensional GP arising from a three-stage social PCC for two individuals (light and dark). All three intermediate stages of smoothing are shown, beginning with Brownian motion (top, dotted), followed by “social” smoothing (top and bottom, dashed), and finally “inertial” 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 .

$$\mu_i^{(soc)}(\tau_{soc}) = \sum_{j=1}^p h_{ij}^{(soc)}(\tau_{soc}) \mu_j^{(bm)}(\tau_{soc}) \quad (3.16)$$

$$\tilde{\mu}_i(\tau_{inl}) = \mu_i^{(inl)}(\tau_{inl}) = \int_{\mathcal{T}} h^{(inl)}(\tau_{inl}, \tau_{soc}) \mu_i^{(soc)}(\tau_{soc}) d\tau_{soc}. \quad (3.17)$$

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 themselves be dynamic. 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 $i \in \{1, \dots, p\}$ 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

$$s_i(t) \equiv \sigma_\mu \tilde{\mu}_i(t) + \varepsilon_i(t). \quad (3.18)$$

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 i.i.d. mean-zero Gaussian random variables with variance σ_s^2 , although more complex measurement error processes could also be used (e.g., Brost et al., 2015; Buderman et al., 2016).

3.2.4 Process prior for the dynamic social network

The general model for dependent movement is highly parameterized. To evaluate the covariance $\Delta\tau\tilde{\mathbf{H}}\tilde{\mathbf{H}}'$ 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 4.3.5, we fit the proposed model to data within a Markov chain Monte Carlo (MCMC) 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 as well as 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 Hoff et al. (2002), Hoff (2008), and Durante et al. (2014), 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, dynamic social network. Suppose 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 $d_{ik}(t)$ 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 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 reason-

able 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

$$w_{ij}(\tau_{soc}) \equiv g\left(\boldsymbol{\mu}_i^{(w)}(\tau_{soc}), \boldsymbol{\mu}_j^{(w)}(\tau_{soc})\right). \quad (3.19)$$

In general, the type of network desired for the application (e.g., binary, weighted) may inform the particular choice for g . As discussed in Section 3.2.3, 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 $w_{ij}(\cdot)$.

Defining the weights in the social smoother through the relative positions of actors in a latent space reduces the dimensionality of the dynamic social network marginally at each time. In addition, we constrain the evolution of the weights in time to further reduce dimensionality and reflect an assumption that social connections among individuals should be stable in time. One approach for constraining the temporal smoothness in $w_{ij}(\tau_{soc})$ is to enforce temporal smoothness in the locations of the latent actors (e.g., Sarkar and Moore, 2006; Sewell and Chen, 2015). We model the location of the actors in the latent social space using another process convolution with kernel $h^{(w)}(\cdot, \cdot)$ as

$$\boldsymbol{\mu}_i^{(w)}(\tau_w) \equiv \int_{\mathcal{T}} h^{(w)}(\tau_w, \tau) d\mathbf{B}_i^{(w)}(\tau), \quad (3.20)$$

$$h^{(w)}(\tau_w, \tau) = \sigma_w e^{-\frac{(\tau-\tau_w)^2}{\phi_w^2}}, \quad (3.21)$$

where $d\mathbf{B}_i^{(w)}(\cdot)$ 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 3.2.2, 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

$$\boldsymbol{\mu}_i^{(w)} \sim \mathbf{N}\left(\mathbf{0}, \sigma_w^2 \Delta t \tilde{\mathbf{H}}_w' \tilde{\mathbf{H}}_w\right). \quad (3.22)$$

3.3 Model implementation

We obtain realizations from the posterior distribution of the model parameters using a Markov chain Monte Carlo (MCMC) 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 $\mu_{0i} \sim \mathbf{N}(0, \sigma_0^2)$ and $\sigma_0^2 \sim \text{IG}(a_0, b_0)$. 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., Gelfand et al., 2003; Finley et al., 2015) 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 (Finley et al., 2015). The resulting integrated model formulation is

$$\mathbf{s}_i(t) | \boldsymbol{\varepsilon}_w, \phi_{inl}, \sigma_s^2, \sigma_\mu^2, \sigma_0^2 \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}) \quad (3.23)$$

$$\boldsymbol{\Sigma} \equiv \sigma_s^2 \mathbf{I} + \sigma_\mu^2 \Delta \tau \tilde{\mathbf{H}}' \tilde{\mathbf{H}}, \quad (3.24)$$

where the parameters ϕ_{inl} , σ_0^2 , and ε_w enter the density through the definition of $\tilde{\mathbf{H}}$ in (3.15) - (3.17). To update each latent-space path, $\boldsymbol{\mu}_i^{(w)}$, we employ a Metropolis-Hastings algorithm.

3.4 Simulation

As the dependence between individuals in a population weakens (i.e., $w_{ij}(t) \rightarrow 0$ for all t , $i \neq j$), our proposed model simplifies to one that considers each path independently, but shares information about all non-social 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 to fit our proposed model, 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 inter-path independence (IP-IND). In many cases, we find that accounting for the dependence among individuals results in 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

$$\text{SPE}(\hat{\boldsymbol{\mu}}; \mathcal{T}_g, \tilde{\boldsymbol{\mu}}_{true}) = \frac{1}{|\mathcal{T}_g|} \int_{\mathcal{T}_g} \|\tilde{\boldsymbol{\mu}}_{true}(t_g) - \hat{\boldsymbol{\mu}}(t_g)\|_2^2 dt_g. \quad (3.25)$$

We define the reconstruction $\hat{\boldsymbol{\mu}}$ to be the posterior mean for $\tilde{\boldsymbol{\mu}}$ so that (3.25) 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 we term the average circular 95% credible region radius ($\text{ACRR}_{0.95}$).

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 $\text{ACRR}_{0.95}$ to compare the performance 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 five step procedure:

- (1) Simulate a realization of $\tilde{\boldsymbol{\mu}}_{true}$ and \mathbf{s} from our proposed model.
- (2) Fit both the dependent and independent-paths models to \mathbf{s} .
- (3) Sample 1000 paths from each of the two posterior distributions using composition sampling.
- (4) Compute the path error defined in (3.25) for the posterior mean.
- (5) Use the same 1000 draws from the posterior to compute the average radius of a circular 95% credible region surrounding a point $\mu_2(t_g)$.

We repeat the procedure 20 times for each combination of parameters to obtain an estimate of the variability between realizations of the simulated paths. Values for the parameters used to simulate each path as well as all prior distributions and hyper parameters are provided in Table B.1.

To facilitate direct comparison between the two models under consideration, we examined the ratios of SPE and $\text{ACRR}_{0.95}$ for IP-DEP and IP-IND, defining ratios such that the relevant value for IP-DEP appears in the denominator and values greater than 1 show support for IP-DEP. Figure 3.2 displays the ratios of SPE (top row) and $\text{ACRR}_{0.95}$ (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, red) and low (dashed, blue) tortuosities, with an associated polygon 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 full model for dependent movement 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 slightly 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.

3.5 Killer whales

We analyzed telemetry data for four killer whales near the Antarctic Peninsula (see Figure 3.3) over the course of five days in February 2014. Geographic positions were measured using Argos telemetry tags (for a complete description of the tags and study area see Andrews et al., 2008; Durban and Pitman, 2012). Although multiple types of killer whales have been described in this

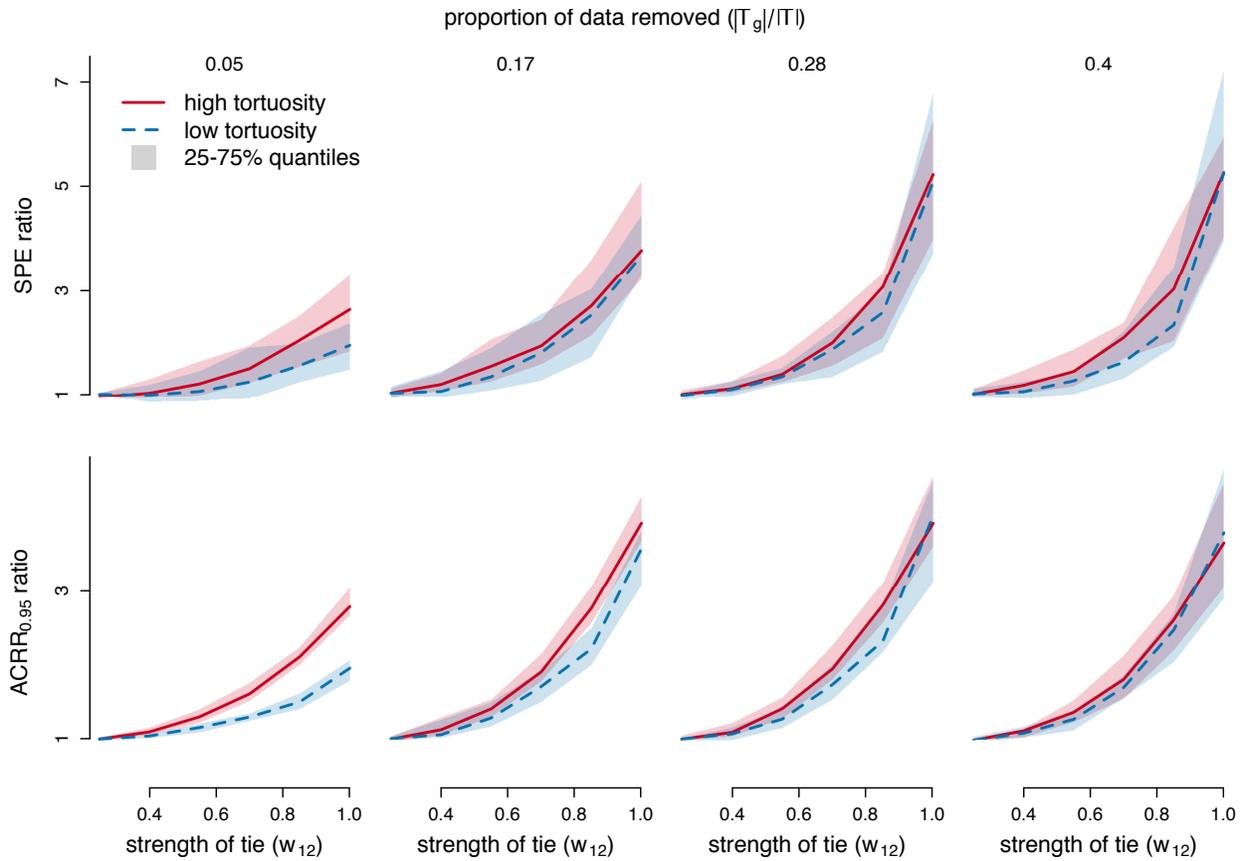


Figure 3.2: Path errors. The top row of plots shows the ratios of SPE for the independent (IP-IND) to full (IP-DEP) models (thus, larger ratios correspond to the full model outperforming the alternative). The bottom row shows the analogous ratios of $ACRR_{0.95}$. Each column represents a fixed value 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, while dashed lines correspond to simulations with low tortuosity.

area, all four tags were deployed on individuals from the same population of the most common type of “B2” killer whales (Durban et al., 2017).

It is immediately apparent from the data that individuals 1 and 2 (Figure 3.3, bottom left) show potential evidence of a close connection. In addition, there is some ambiguity about the relationship between individuals 3 and 4 (Figure 3.3 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 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 we have a strong, *a priori*, belief that the underlying network should exclude connections between the two subgroups.

The top plot of Figure 3.4 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 day-long gaps in observation occur for individual 1 (Figure 3.4), a consequence of the original study design used in the deployment of the telemetry tags which sought to balance the need for temporally dense observations with limitations on 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 could 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 3.4), we investigated path reconstructions generated by our proposed full model for dependent movement (IP-DEP), as well as the analogous model under an assumption of inter-path independence (IP-IND). We found that the full model generates path reconstructions in which the uncertainty about the position of individual 1 is dramatically reduced, compared to the reconstructions generated under the assumption of inter-path independence as measured by 95% circular credible interval radii (bottom plot of Figure 3.4). Circular credible

regions for the true position of the whale during the two large gaps in observation occurring on February 12th and 14th for the IP-IND model have large radii, sometimes exceeding 75km. In contrast, circular credible regions for the true position based on the IP-DEP model are on par with 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.

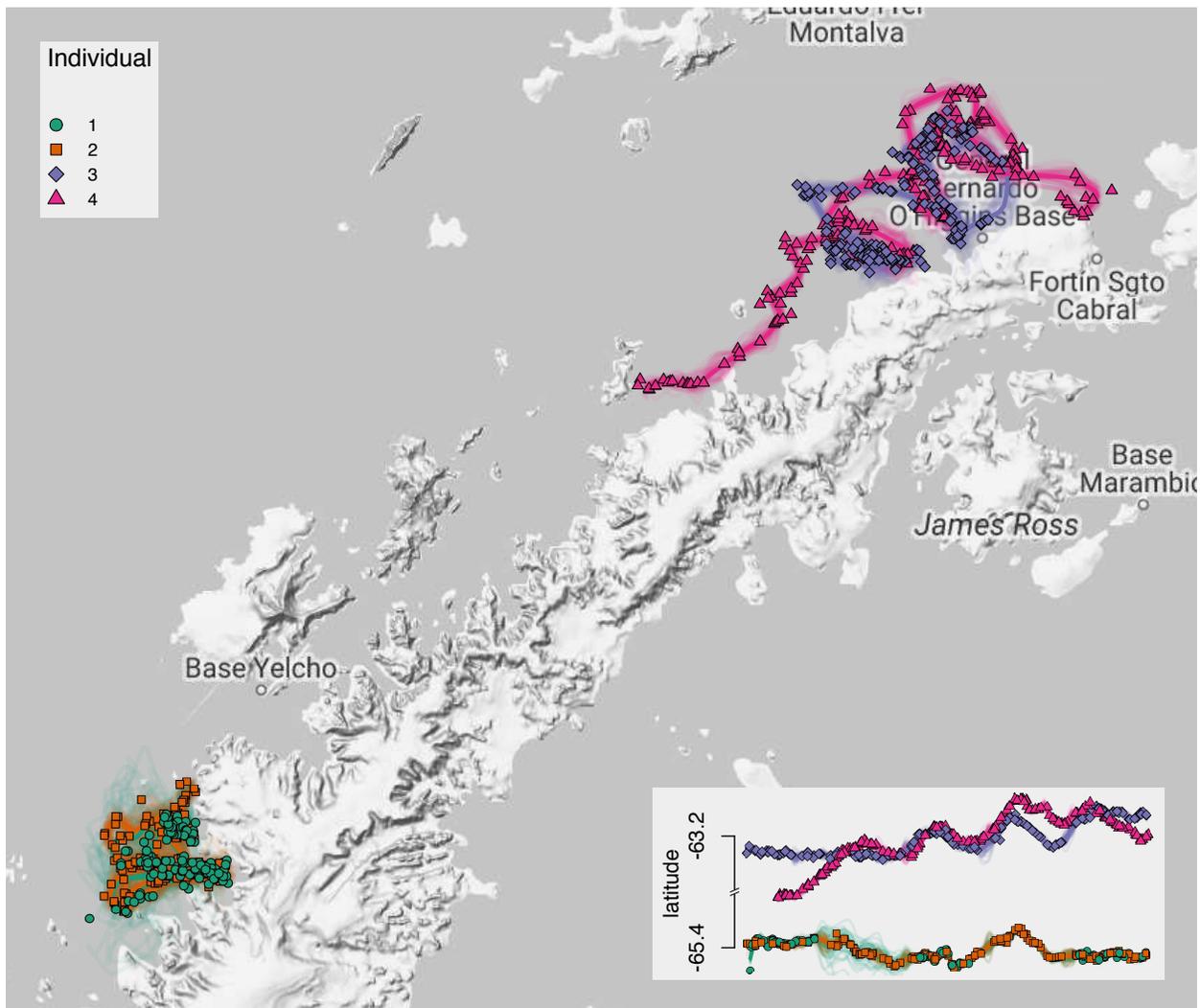


Figure 3.3: Observed telemetry data and joint posterior distribution of the true paths of all individuals ($\tilde{\mu}$). The solid lines represent the posterior mean, and the semi-transparent lines are draws from the posterior distribution to illustrate uncertainty in the path reconstruction. The points are the observed locations from the Argos satellite system. The colors correspond to the four individuals in the study, and match those used in Figure 3.4. 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 3.5. 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 B.2.

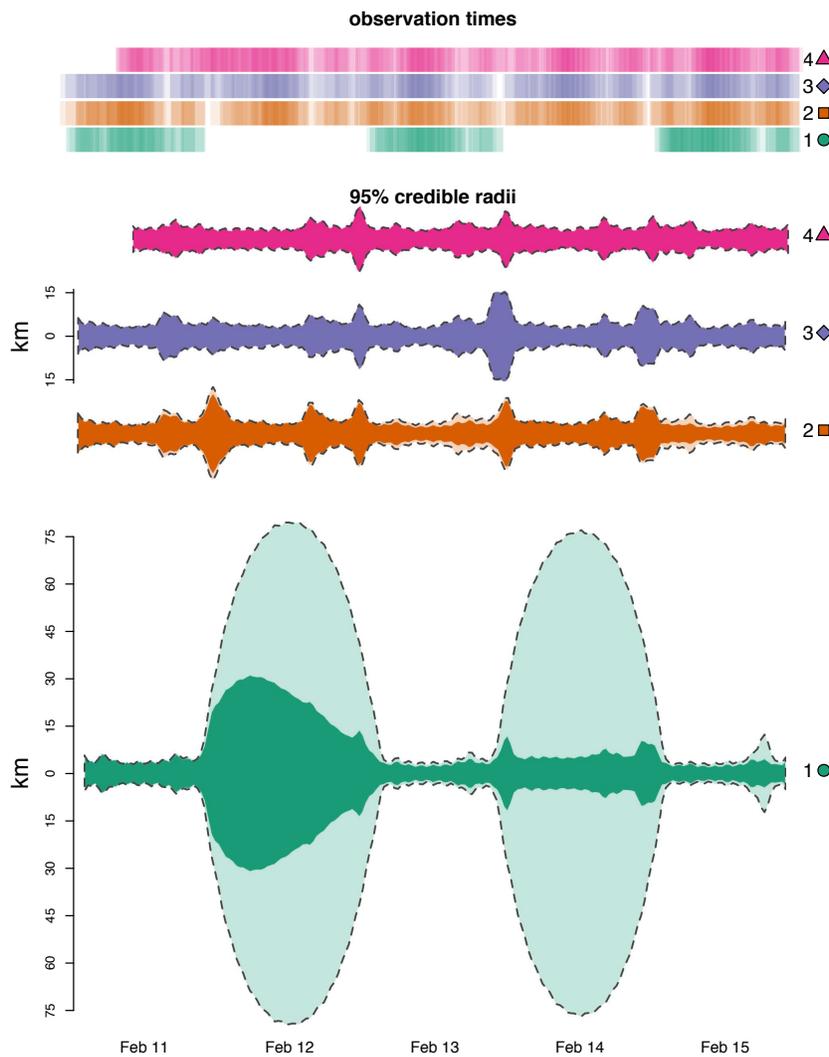


Figure 3.4: *TOP:* Observation times for the three killer whales. Darker regions correspond to denser observation times, or equivalently, shorter gaps between observations. *BOTTOM:* 95% circular credible region radii at each time point for all four individuals. The dashed outer line shows the radii for the independent model, and the solid polygons show the radii for the full model.

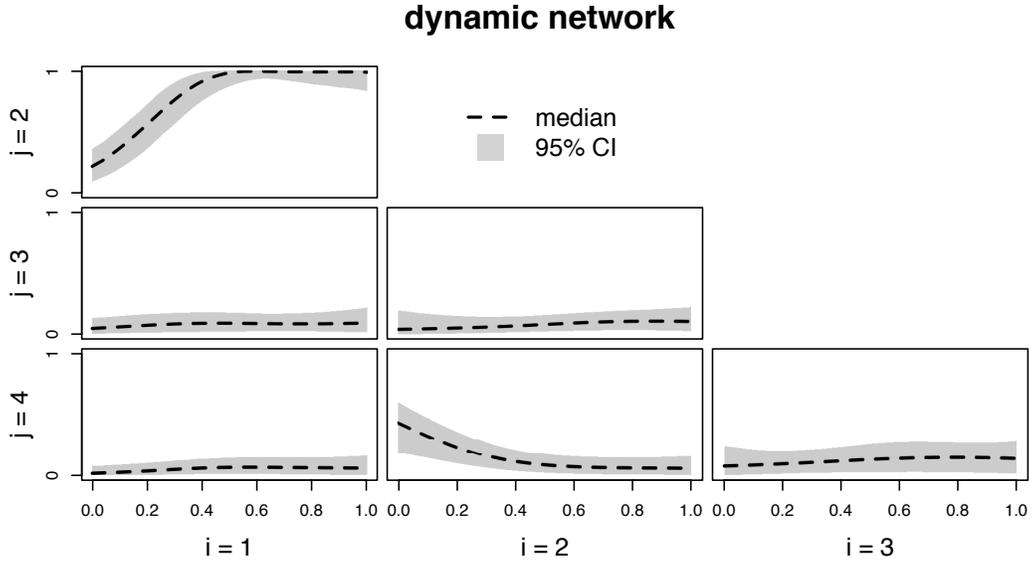


Figure 3.5: Posterior distribution of the dynamic social network that defines the social smoother used in stage two of the proposed PCC model (IP-DEP). Each plot shows the time-evolution for a specific pair of killer whales indexed by row and column.

3.6 Discussion

Appropriate models for dependent, multivariate data are application-dependent. Gaussian processes 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 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 spatio-temporal processes, also called “cokriging,” have been studied for several decades (e.g., Myers, 1982; Cressie, 1991). The pri-

mary challenge has been to develop models for multivariate processes that accurately capture the complex forms of data within each process, as well as across processes, while ensuring the resulting cross-covariance structure is valid (symmetric and non-negative 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 approach the problem of specifying a valid, mechanistically motivated cross-covariance structure for multivariate processes through the use of kernel convolutions, allowing the authors to relax unrealistic assumptions about the covariance structure. Similar to the way dependence among individuals arises through the application of a social smoothing kernel (Section 3.2.3), Ver Hoef and Barry (1998) induced dependence in a multivariate process by smoothing across variables. The cross-covariance in Ver Hoef and Barry (1998) also accommodates spatio-temporal lags in the dependence among variables. This generalization may also be a useful 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. Animals may also exhibit other forms of dependence, such as a tendency to repel in the case of strong territorial behavior, that 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 role certain kernel functions play in the characteristics of the resulting random process.

As mentioned in Section 3.2.1, 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, while 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, if we socially smoothed white noise and then used the kernel associated with Brownian motion, a very different form of dependence results. This alternative ordering ($h^{(soc)} \rightarrow h^{(bm)} \rightarrow h^{(intl)}$) 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 “shrinking” 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 “shrink” the velocities of the particles together, rather than their positions. Thus, we would be inducing a tendency for particles to move in similar directions, though 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.

3.7 Declaration of support

Killer whale tagging was conducted under permit #14097 from the National Marine Fisheries Service and Antarctic Conservation Act permit #2009-013. Shipboard tagging operations were supported by Lindblad Expeditions and the National Geographic Society, and by an NSF rapid grant to Ari Friedlaender. Robert Pitman helped with tag deployments and identification of killer whale types in the field. This material is based in part upon research supported by NSF DMS-1614392.

Chapter 4

Accounting for Phenology in the Analysis of Animal Movement

4.1 Introduction

For decades, analysis of the movement of animals has been an important source of scientific understanding and discovery in ecology (Hooten et al., 2017a). Observations of animal trajectories using telemetry devices, such as radio collars, have provided researchers with information about the way animals interact with their environment (e.g., Manly et al. 2002; Johnson et al. 2008a) and each other (e.g., Niu et al., 2016; Scharf et al., 2016). In the case of the former, one of the most common approaches used by ecologists is to analyze the frequencies with which animals use certain types of habitats relative to the distribution of habitat types available to them. Such analyses typically make use of a so-called “use-availability” framework in which the probability of an individual using a particular location is modeled as a weighted combination of all available locations (e.g., Northrup et al., 2013). The weights are then referred to as the resource selection function (RSF), and provide insight into which portions of a landscape are most desirable to the study species. What constitutes an available location depends on the characteristics of the particular species under study and the rate at which telemetry observations are gathered. Typically, however, the degree to which a location is available is a function of how far an animal can reasonably be expected to move between observation times, and/or the size of the home range of the individual (e.g., Christ et al. 2008; Brost et al. 2015).

For many species, specific geographical features in the landscape can have a strong effect on where individuals choose to move. Such features are sometimes well-summarized by a single point (e.g., dens or kill sites), but may also correspond to higher-dimensional subspaces (e.g., rivers or lakes). Their locations may be relatively static in time (e.g., coastlines or home-range centers), or

may be dynamic (e.g., sea ice extent or areas of high-quality forage for herbivores). While not always framed in the context of resource selection, these landscape features can nevertheless be thought of as resources, and the behavior of animals may demonstrate selection for (or against) points near the feature. We introduce a novel contribution to the suite of models available for the analysis of animal movement that incorporates active selection for features in a landscape that may have complex and dynamic shapes.

Our modeling framework is motivated by the study of polar bear (*Ursus maritimus*) movement. Polar bears spend much of their time on sea ice over shallow, biologically productive water where they depredate seals. During the sea ice melt season, the part of the year when sea ice first breaks up and contracts toward the pole, then freezes and expands southward again, polar bears spend much of their time at the interface between sea ice and the ocean (Durner et al. 2009; Rode et al. 2015; Atwood et al. 2016). The changing distribution and characteristics of sea ice throughout the late spring through early fall means that the location of desirable ice-edge habitat is constantly shifting. As climate change alters the rate at which sea ice thaws and freezes, as well as the size of its minimum and maximum extents, there is increasing concern about how polar bears are responding to these dramatic shifts in their environment (Rode et al., 2014).

Our goal was to develop a model for the movement of polar bears that explicitly incorporates the effect of the changing sea ice and can be used in a variety of hierarchical models to better understand polar bear ecology. In Section 4.3, we use our proposed model to answer a particular question posed by wildlife managers who seek to summarize the spatial boundaries of two sub-populations of polar bears from the Beaufort and Chukchi seas.

4.2 Model Development

4.2.1 Feature preference

To account for an individual's preference for areas in a landscape near (or far from) a particular feature of interest, we model the true, unobserved locations of each individual using a spatio-temporal point process approach (Hooten et al., 2017a). We define the probability density for

each unobserved location as proportional to a product of two parametric components describing the availability and desirability of every point on the landscape. The desirability of a particular point is defined based on its Euclidean distance to the feature of interest. Estimates of the relevant parameters provide a summary of how strong an effect the feature has on the behavior of the observed individuals. We model availability similar to Hjermann (2000), Christ et al. (2008), Johnson et al. (2008b), and Brost et al. (2015) who used radial distributions centered on the most recently-observed location to define the continuously-valued availability at each point in time.

Let $\boldsymbol{\mu}(t)$ be the location of an individual at time $t \in \{1, \dots, T\}$, where we will assume for now that observation times are equally spaced with no missing values. Define the conditional probability density for $\boldsymbol{\mu}(t)$ as

$$[\boldsymbol{\mu}(t) | \boldsymbol{\mu}(t-1), \sigma_\mu^2, \mathcal{M}(t), \tau^2] \propto \begin{cases} \mathbf{N}(\boldsymbol{\mu}(t); \boldsymbol{\mu}(t-1), \sigma_\mu^2 \mathbf{I}_2) g(\boldsymbol{\mu}(t); \mathcal{M}(t), \tau^2), & t > 1 \\ g(\boldsymbol{\mu}(1); \mathcal{M}(1), \tau^2), & t = 1 \end{cases} \quad (4.1)$$

where we use square brackets to denote a general probability density function, and $\mathbf{N}(\cdot; \cdot, \cdot)$ to denote the Gaussian density function in particular. For $t > 1$, the conditional distribution is proportional to the product of two components, the first of which is the density of a bivariate Gaussian distribution centered on the previous location of the individual, and defines the availability of each point on the landscape as in Christ et al. (2008). The availability component induces positive auto-correlation in the joint process $\boldsymbol{\mu} \equiv (\boldsymbol{\mu}(1), \dots, \boldsymbol{\mu}(T))'$, with larger values of σ_μ^2 resulting in processes with greater distances between consecutive locations and faster, more erratic movement.

The function g is a RSF that controls the effect of a particular feature in the landscape on an individual's movement. Let $\mathcal{M}(t)$ denote the set of points that make up the feature of interest (e.g., the interface between sea ice and ocean). We define the function g as

$$g(\boldsymbol{\mu}(t); \mathcal{M}(t), \tau^2) = \exp \left\{ - \min_{\mathbf{x} \in \mathcal{M}(t)} \|\boldsymbol{\mu}(t) - \mathbf{x}\|_2^2 / 2\tau^2 \right\}, \quad (4.2)$$

(where $\|\cdot\|_2$ is the ℓ_2 or Euclidean norm) so that the value of g is highest along $\mathcal{M}(t)$, and reduces to zero as $\boldsymbol{\mu}(t)$ moves away from $\mathcal{M}(t)$. The value of τ^2 controls the range at which g effectively reduces to zero. We show in Section 4.2.2 why this particular parametric form for the RSF leads to computational efficiencies in parameter estimation.

We defined g such that it achieves its largest values at locations $\boldsymbol{\mu}(t)$ near $\mathcal{M}(t)$ so that the conditional density given in (4.1) has probability mass concentrated near $\mathcal{M}(t)$. Specifying g in this way provides a method for modeling movement that exhibits preference for the region of the landscape near the feature of interest. Alternative specifications could also be used to model the movement of individuals displaying preference for portions of the landscape far from the feature $\mathcal{M}(t)$.

The model for the discrete-time process $\boldsymbol{\mu}$ provides a useful tool for modeling the movement of an individual responding to a one-dimensional feature on a landscape. In Section 4.3, we apply the model to the movement of polar bears with the ultimate goal of clustering individuals into disjoint sub-populations based on space use. By including availability and resource selection as part of a larger hierarchical structure, we were able to account for polar bears preference for habitats that facilitate the depredation of seals, which, if ignored, might result in biased inference about sub-population membership.

The conditional density in (4.1) is only defined up to a constant of proportionality that must be computed as part of any conventional estimation procedure. In Section 4.3, we employ a Bayesian hierarchical methodology and fit our model for polar bear movement using Markov chain Monte Carlo (MCMC), which requires computation of the normalizing constant several times at each iteration of the algorithm. For a general feature, $\mathcal{M}(t)$, the normalization constant is not analytically tractable. Thus, numerical integration is required to fit the model to data, the computational cost of which precludes such an approach in our application. In the RSF literature, the normalization constant has typically been approximated using either a coarse spatial discretization (e.g., Warton and Shepherd 2010; Brost et al. 2015), or a randomized scheme based on an “availability sample” (e.g., Northrup et al., 2013). In the next section, we introduce a novel approach for computing the nec-

essary normalization constant in which we approximate the RSF in a way that induces conjugacy in the distributional form for $\boldsymbol{\mu}(t)$ and greatly reduces the computational cost of model fitting.

4.2.2 Linearization approximation

We implement a novel approximation technique that assumes locally linear structure in the shape of $\mathcal{M}(t)$, allowing for efficient approximation of the true conditional density of $\boldsymbol{\mu}(t)$. To motivate our approximation, note that, for the special case when $\mathcal{M}(t)$ is a straight line, the RSF as defined in (4.2) can be written in a form similar to that of a bivariate Gaussian density function with a rank-deficient covariance matrix.

Resource selection function for straight lines

First, consider the case of a vertical line, $\widetilde{\mathcal{M}}(t)$, in the real plane so that $\widetilde{\mathcal{M}}(t) \equiv \{(x, y) \in \mathbb{R}^2 : x = h\}$. For this case, we have

$$g(\boldsymbol{\mu}(t); \widetilde{\mathcal{M}}(t), \tau^2) = \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu}(t) - (h, y)')' \mathbf{Q}(\tau^2) (\boldsymbol{\mu}(t) - (h, y)') \right\},$$

$$\mathbf{Q}(\tau^2) \equiv \begin{pmatrix} \tau^{-2} & 0 \\ 0 & 0 \end{pmatrix}$$

for all real-valued y .

To allow for $\widetilde{\mathcal{M}}(t)$ that are straight, but not necessarily vertically oriented, we rotate the coordinate system through an angle θ . Let $\mathbf{R}(\theta)$ be the rotation matrix defined as

$$\mathbf{R}(\theta) \equiv \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix},$$

and let θ be defined such that $\mathbf{R}'(\theta)\widetilde{\mathcal{M}}(t) = \{(x, y) \in \mathcal{R}^2 : x = h\}$ for some real-valued h . Note that the inverse of a rotation matrix, $\mathbf{R}^{-1}(\theta) = \mathbf{R}(-\theta)$, is also equal to its transpose, $\mathbf{R}'(\theta)$. The RSF defined in (4.2) is invariant under rigid transformations such as rotations, therefore

$$\begin{aligned}
g(\boldsymbol{\mu}(t); \widetilde{\mathcal{M}}(t), \tau^2) &= g(\mathbf{R}'(\theta)\boldsymbol{\mu}(t); \mathbf{R}'(\theta)\widetilde{\mathcal{M}}(t), \tau^2) \\
&= \exp \left\{ -\frac{1}{2} (\mathbf{R}'(\theta)\boldsymbol{\mu}(t) - \mathbf{R}'(\theta)(h, y)')' \mathbf{Q}(\tau^2) (\mathbf{R}'(\theta)\boldsymbol{\mu}(t) - \mathbf{R}'(\theta)(h, y)') \right\} \\
&= \exp \left\{ -\frac{1}{2} (\boldsymbol{\mu}(t) - (h, y)')' \mathbf{R}(\theta)\mathbf{Q}(\tau^2)\mathbf{R}'(\theta) (\boldsymbol{\mu}(t) - (h, y)') \right\}.
\end{aligned}$$

Linearizing complex landscape features

The resulting form for $[\boldsymbol{\mu}(t)|\boldsymbol{\mu}(t-1), \widetilde{\mathcal{M}}(t)]$ is proportional to the product of two Gaussian distributions, one of which is improper. Provided $\tau^2 < 0$, the product in (4.1) is a proper bivariate Gaussian distribution with mean, $\boldsymbol{\mu}^*$, and covariance, $\boldsymbol{\Sigma}^*$, given by

$$\begin{aligned}
\boldsymbol{\Sigma}^* &\equiv (\sigma_\mu^{-2}\mathbf{I}_2 + \mathbf{R}(\theta)\mathbf{Q}(\tau^2)\mathbf{R}'(\theta))^{-1} \\
\boldsymbol{\mu}^* &\equiv \boldsymbol{\Sigma}^* (\sigma_\mu^{-2}\boldsymbol{\mu}(t-1) + \mathbf{R}(\theta)\mathbf{Q}(\tau^2)\mathbf{R}'(\theta)\widetilde{\mathbf{m}}),
\end{aligned}$$

where $\widetilde{\mathbf{m}}$ is any point in $\widetilde{\mathcal{M}}(t)$. The distributional form of $\boldsymbol{\mu}(t)|\boldsymbol{\mu}(t-1), \widetilde{\mathcal{M}}(t)$ implicitly defines the appropriate normalization constant in (4.1). Thus, if there exists some straight line $\widetilde{\mathcal{M}}(t)$ that represents a close approximation to $\mathcal{M}(t)$ near $\boldsymbol{\mu}(t-1)$, $g(\boldsymbol{\mu}(t); \widetilde{\mathcal{M}}(t), \tau^2)$ may provide a reasonable approximation for $g(\boldsymbol{\mu}(t); \mathcal{M}(t), \tau^2)$ that alleviates the computational burden of repeatedly calculating the necessary normalization constant. A natural candidate for $\widetilde{\mathcal{M}}(t)$ is the line that is tangent to $\mathcal{M}(t)$ at the point on $\mathcal{M}(t)$ closest to $\boldsymbol{\mu}(t-1)$, because this is the portion of the feature that contributes most to the conditional distribution of $\boldsymbol{\mu}(t)$.

Let $\mathbf{m}(t)$ be the point in $\mathcal{M}(t)$ nearest to $\boldsymbol{\mu}(t-1)$, and let $\widetilde{\mathcal{M}}(t)$ be the set of points that lie on the line tangent to $\mathcal{M}(t)$ at $\mathbf{m}(t)$. Figure 4.1 shows a schematic illustrating how the product of the Gaussian components in (4.1) result in another Gaussian distribution. The dashed line represents an example of the edge of a one-dimensional feature of interest, and the solid line represents the linearized edge.

For $\widetilde{\mathcal{M}}(t)$ to result in an adequate approximation of the RSF, the linearized feature need only resemble the true feature in the vicinity of $\boldsymbol{\mu}(t-1)$. Outside of the immediate neighborhood of $\boldsymbol{\mu}(t-1)$, the availability distribution will be essentially zero, reducing the impact of errors in the

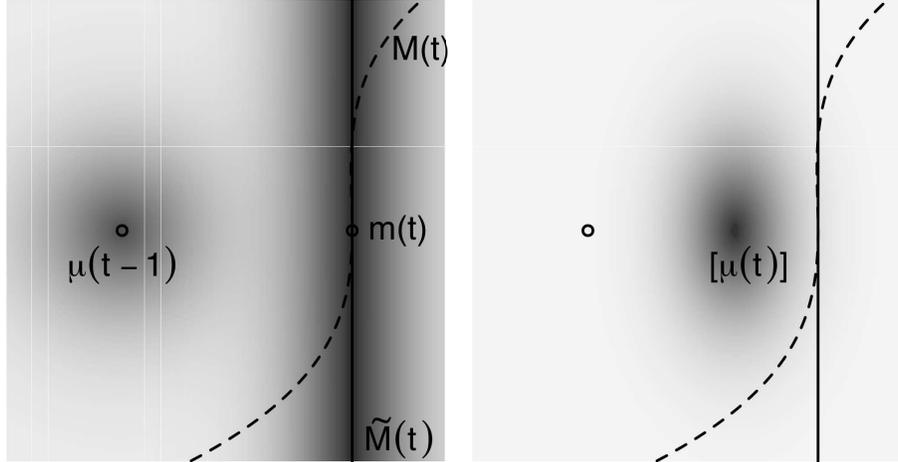


Figure 4.1: The product of the two densities in the left plot (one of which is improper) result in the density shown in the right plot.

RSF approximation. Therefore, the linearization approximation is justifiable if it is reasonable to assume that, for the scale at which locations are measured, the edge of the feature of interest can be considered approximately linear.

4.3 Application

4.3.1 Goals and previous work

There are a total of 19 recognized polar bear sub-populations in the circumpolar Arctic (Figure 4.2, Obbard et al., 2010). However, the boundaries that delineate the sub-populations are challenging to precisely define because there are few barriers to movement for polar bears, and the changing extent and drift of the sea ice leads to periods of the year when individuals from different sub-populations may use overlapping portions of the landscape. Nevertheless, there are important reasons to determine a clear delineation of the sub-population boundaries. For example, wildlife management agencies such as the U.S. Fish and Wildlife Service (USFWS) use sub-population boundaries to help guide management decisions for polar bears, which are currently listed as ‘threatened’ under the Endangered Species Act (U.S. Fish and Wildlife, 2016). There is also evidence that polar bears from different sub-populations are responding to climate change with differing degrees of success (Rode et al., 2014; Ware et al., 2017). In what follows, we focus

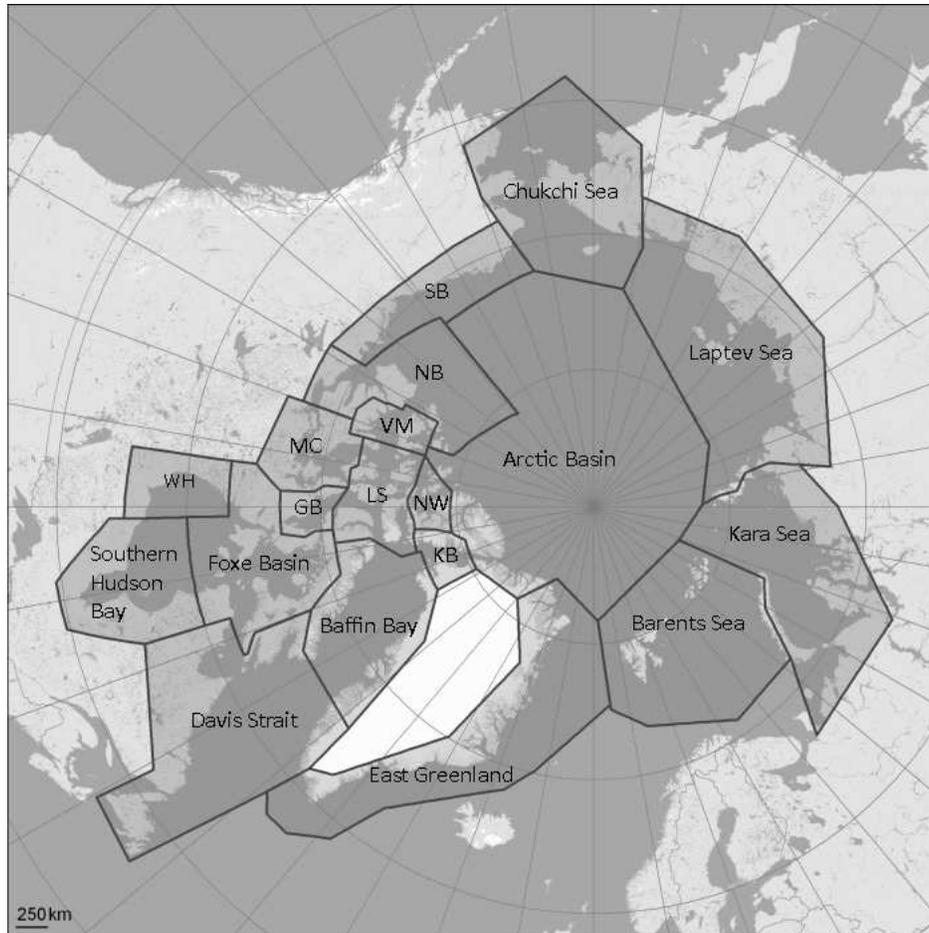


Figure 4.2: Sub-population boundaries of polar bears (Obbard et al., 2010; U.S. Fish and Wildlife, 2016). Sub-population abbreviations are: Southern Beaufort Sea (SB), Northern Beaufort (NB), Kane Basin (KB), Norwegian Bay (NW), Lancaster Sound (LS), Gulf of Boothia (GB), M'Clintock Channel (MC), Viscount Melville Sound (VM), and Western Hudson Bay (WH).

on estimating a sub-population boundary between the Chukchi Sea (CS) and Southern Beaufort Sea (SB) sub-populations.

Previous efforts to determine sub-population boundaries provided important initial estimates; however, the substantial physical changes taking place in the Arctic ecosystem mean that boundary estimates have the potential to become outdated quickly. Amstrup et al. (2005) used a multi-stage non-parametric clustering procedure to estimate boundaries based on utilization distributions estimated from telemetry observations. While the methods employed provide a probabilistic basis for inference, there are three primary opportunities for continued development.

First, Amstrup et al. (2005) did not make use of all observations. They discarded observations with measurement error classes deemed insufficiently precise, and thinned observations so that only one location is used per bear per 6-day duty cycle. Thinning the data represents an effort to diminish the dependence between consecutive observations so that locations can be treated as independent in the analysis. In our analysis, we directly incorporated the dependence between consecutive observations. We also modeled the wide variety of measurement error structures in the data allowing us to include all observations.

Second, the Amstrup et al. (2005) analysis did not provide probabilistic estimates of sub-population membership for each observed individual. Our analysis allowed us to compute the probability of sub-population membership for each polar bear, with estimates of uncertainty. Moreover, our methodology allowed us to estimate the sub-population membership probabilities for a newly observed individual, conditioned on the data used to fit the model.

Finally, Amstrup et al. (2005) did not directly incorporate the effect changing sea ice had on the observed locations of polar bears. Estimates of space use for each sub-population were summarized across all seasons, potentially confounding the roles of sea ice and sub-population boundaries. In contrast, we directly incorporated the effect of sea ice extent on movement with a RSF. Our novel linearization approximation allowed us to obtain inference from the resulting mechanistic model using an efficient sampling algorithm.

We analyzed telemetry observations of $N = 279$ polar bears made by the USFWS and the U.S. Geological Survey (USGS) over the course of 2008–2016 with the goal of estimating the boundary between the CS and SB sub-populations.

4.3.2 Movement process model

With the exception of mothers denning on land, most polar bears remain on the ice in the summer, following it north as it retreats away from continental land masses. Polar bears are specialist carnivores that use areas near the interface of sea ice and ocean to hunt seals during the sea ice melt season (Durner et al., 2009). However, polar bears do occasionally remain on land through

the warmest part of the year and, in fact, Atwood et al. (2016) concluded that there is evidence the number of individuals exhibiting this behavior is increasing.

Two of the primary characteristics of polar bear movement are a tendency for most individuals to prefer portions of the landscape near the edge of the sea ice, and a tendency for individuals to occupy a general spatial region corresponding to the particular unobserved sub-population to which an individual polar bear is a member. To address the effect of the sea ice, we specified a model for movement that incorporates preference for areas near the sea ice boundary as it changes in time. We accounted for polar bears that remain on land during the summer by also including coastline as a feature on the landscape associated with increased rates of use, because individuals that spend the summer on land tend to remain on islands and/or near the coastline for most of the season (Rode et al., 2015). By accounting for habitat variability, we minimized the possibility of clustering bears into sub-populations that confound true sub-population spatial regions with movement responding to changing sea ice.

We used a Bayesian hierarchical model, and specified models for both measurement error and the true unobserved movement process in a single coherent framework. We used a discrete-time approach to specify a model for polar bear movement, in which the conditional probability density for the location of an individual is proportional to a product of components corresponding to resource availability and selection. The discrete-time approach was motivated by the scale at which we have measurements of Arctic sea ice, which features as a direct effect in our model for movement. We used estimates of the extent of sea ice provided by the National Snow and Ice Data Center (Fetterer et al., 2010) that are available at daily intervals; thus, we modeled movement as a discrete-time stochastic process on a daily scale.

Let $\mu_i(t)$ denote the location of individual $i \in \{1, \dots, N\}$ at time $t \in \mathcal{T}_i$, where \mathcal{T}_i is a consecutive set of times from the reference set $\mathcal{T} = \{1, \dots, T\}$, and let z_i be a binary random variable that is 1 if individual i belongs to the CS sub-population, and 0 if it belongs to the SB sub-population. Additionally, denote by $\mathcal{M}(t)$ the set of points in the plane defined by the union of coast line, and the edge of the sea ice. We modeled the conditional distribution of $\mu_i(t)$ as

$$\begin{aligned}
& [\boldsymbol{\mu}_i(t) | \boldsymbol{\mu}_i(t-1), \sigma_\mu^2, \boldsymbol{\mu}_{CS}, \boldsymbol{\mu}_{SB}, \sigma_{ac}^2, z_i, \tau^2] \propto \\
& \begin{cases} \text{N}(\boldsymbol{\mu}_i(t); \boldsymbol{\mu}_i(t-1), \sigma_\mu^2) \text{N}(\boldsymbol{\mu}_i(t); \boldsymbol{\mu}_{CS}^{z_i} + \boldsymbol{\mu}_{SB}^{1-z_i}, \sigma_{ac}^2 \mathbf{I}_2) \\ \quad \times g(\boldsymbol{\mu}_i(t); \mathcal{M}(t), \tau^2), & t > \min(\mathcal{T}_i) \\ \text{N}(\boldsymbol{\mu}_i(t); \boldsymbol{\mu}_{CS}^{z_i} + \boldsymbol{\mu}_{SB}^{1-z_i}, \sigma_{ac}^2 \mathbf{I}_2), & t = \min(\mathcal{T}_i). \end{cases} \quad (4.3)
\end{aligned}$$

Each component in (4.3), $t > \min(\mathcal{T}_i)$ captures a different feature of the movement process. Namely, these are (1) the temporal dependence between locations on consecutive days, (2) the association of each individual bear with a sub-population-level central place, and (3) a RSF that appropriately weights locations near a coastline or the edge of the sea ice. The first two terms can be thought of as a two-component availability function that incorporates movement constraints and a sub-population activity center, similar in many respects to the modeling specification of Christ et al. (2008) and Johnson et al. (2008b). The third term is a RSF that models the preference polar bears exhibit for habitat near either a coastline, or the sea ice boundary.

The model for movement specified in (4.3) also captures a documented secondary effect that changing sea ice has on polar bears (Durner et al., 2017). Especially in the summer months, polar bears that remain on the sea ice expend a substantial amount of energy merely keeping up with the retreating ice sheet. Increased rates of sea ice retreat associated with global climate change have been estimated to impose an energetic cost on polar bears of 1–3 additional seals per year compared to historic norms (Durner et al., 2017). Sea ice does not retreat at a uniform rate at all points along the boundary. Rather, the change in shape is highly variable with some regions retreating more slowly than others (Steele et al., 2015). Our proposed RSF-based movement model has the effect of placing higher probability mass on paths that track with regions of the ice that have the slowest daily rates of change. Our model therefore captures an inherent incentive for polar bears to travel along those portions of the ice that minimize their expected energy expenditure associated with keeping pace with the changing ice.

4.3.3 Activity centers and sub-population membership

The second component in (4.3), for $t > \min(\mathcal{T}_i)$, and first component for $t = \min(\mathcal{T}_i)$ is a bivariate Gaussian distribution centered on one of two central places, $\boldsymbol{\mu}_{SB}$ and $\boldsymbol{\mu}_{CS}$, corresponding to the centers of the SB and CS sub-populations, respectively. The variance parameter, σ_{ac}^2 , controls the strength of the effect the sub-population center has on the movement of each individual. As σ_{ac}^2 increases, individuals are allowed to range farther from their central place.

The binary random variables, z_i , indicate the particular sub-population with which individual i associates. We specified prior distributions for each z_i such that

$$z_i \sim \text{Bern}(p_i), \quad p_i \sim \text{Beta}(\alpha, \beta).$$

The global Beta-distributed hyperprior we specify for each p_i allows for the inclusion of informative priors that shrink the probabilities of sub-population membership toward 0 and 1, encouraging separation between the two sub-populations.

4.3.4 Measurement error

Telemetry observations of polar bears made by the USFWS and the USGS employed a variety of different tracking device types, yielding data contaminated with measurement errors of varying severities. In each case, we modeled the observed locations as centered on the true, unobserved location of the individual with an additive measurement error process as

$$\mathbf{s}_i(t^*) = \boldsymbol{\mu}_i(t^*) + \boldsymbol{\varepsilon}_i(t^*), \quad t^* \in (1, T),$$

where the distribution for $\boldsymbol{\varepsilon}_i(t^*)$ depends on the particular device used to make the observation $\mathbf{s}_i(t^*)$, and we allowed measurements to be made at any point on the continuous interval $(1, T)$. We return to the misalignment between this continuous-time scale and the discrete scale used to model the movement process in Section 4.3.5.

The simplest of the three measurement error structures is that associated with GPS devices, which generate small measurement errors that are well-modeled by a circular Gaussian distribution as

$$\varepsilon_{GPS}(t^*) \sim \mathbf{N}(\mathbf{0}, \sigma_{GPS}^2 \mathbf{I}_2), \quad \sigma_{GPS} \sim \text{Unif}(0, 10^8).$$

We specified the upper bound on the uniform prior distribution based on the spatial extent of the area used by all polar bears. The upper bound corresponds to a measurement error distribution that is nearly uniform over the size of the study area.

The second and third types of measurement errors are those associated with Argos-type transmitters (Service Argos, 2016). Argos devices use polar orbiting satellites to record the location of the device and have the advantage of providing values to researchers in real time. However, the satellite-based location estimates are often made with a substantial amount of error, and the structure of the errors can be highly non-circular (Costa et al., 2010). The severity of Argos measurement errors are related to the direction of satellite travel, leading to elliptical error distributions with an orientation that varies with each observation. Service Argos provide categorical error-class labels (3, 2, 1, 0, A, B, or Z) that correspond to measurements made with increasing severity levels. Recently, Argos has included estimates of the maximum and minimum axis lengths, and angle of rotation for the error ellipse associated with each observation (e.g., McClintock et al., 2015).

For Argos observations that lack auxiliary ellipse information, we use the provided classes to model the measurement error process as

$$\varepsilon_{Argos,I}(t^*) \sim \mathbf{N}(\mathbf{0}, \sigma_{Argos,k}^2 \mathbf{I}_2), \quad \sigma_{Argos,k} \sim \text{Unif}(0, 10^8)$$

where $k = \{3, 2, 1, 0, A, B, Z\}$ indexes unique variance parameters for each Argos error class. For an alternative treatment of Argos error structure in the absence of auxiliary ellipse information, see Brost et al. (2016) and Buderman et al. (2016).

For Argos observations with auxiliary ellipse information, we model the measurement error process as

$$\boldsymbol{\varepsilon}_{Argos,II}(t^*) \sim \mathbf{N}(\mathbf{0}, \sigma_{Argos,II}^2 \boldsymbol{\Sigma}(t^*)), \quad \sigma_{Argos,II} \sim \text{Unif}(0, 10^8)$$

where $\boldsymbol{\Sigma}(t^*)$ is a scaled covariance matrix corresponding to a bivariate Gaussian distribution with elliptical contours that match the ratio of axis lengths and angle of rotation provided by Argos for each observation time.

4.3.5 Estimation

Process imputation: Stage 1

Our proposed model for the movement and measurement error process are based on discrete- and continuous-time scales, respectively. To reconcile this difference, we employed a two-stage estimation procedure called “process imputation” (e.g., Hooten et al., 2010; Hanks et al., 2011; Scharf et al., 2016, 2017). The procedure is based on the well-established method of multiple imputation for missing values (Rubin, 1996), and fits within a larger MCMC algorithm. In essence, it consists of first fitting a flexible, continuous-time model to the telemetry data and making a finite number of draws from the posterior distribution of the true continuous process at times $\{1, \dots, T\}$. The realizations from the first stage are selected uniformly at random and treated as the true discrete-time paths in the second stage of the process imputation procedure. Sampling uniformly from the collection of first-stage realizations propagates the measurement error incorporated in the first stage through to the second stage.

Using the process imputation procedure allows us to account for misaligned time scales in the measurement and movement processes, account for measurement error, and ease the computational demand of model fitting in exchange for a small amount of approximation error in the overall estimation procedure. We refer the reader to Scharf et al. (2017) for a more detailed treatment of the procedure, and justification of its use.

We used the Ornstein-Uhlenbeck (OU) movement model of Johnson et al. (2008b), implemented in the R package `crawl` (Johnson, 2016), as the first stage of a process imputation estimation procedure. Johnson et al. (2008b) model the velocity of an individual using Brownian motion, integrating the velocity process over time to yield the movement process. The OU model has been shown to perform well as a first-stage model for process imputation (Scharf et al., 2017), and accommodates a broad range of measurement error models. In stage one, we fit each individual separately, using the measurement error structures defined in Section 4.3.4 as appropriate for each individual’s device type, and drew $K = 24$ realizations of the continuous path evaluated on the daily discrete-time scale for all days falling between the first and last observation times. Additional details associated with model fitting are provided in the Supplementary Materials.

Process imputation: Stage 2

Conditioned on the true discrete-time locations, $\boldsymbol{\mu} \equiv (\boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_N)'$, we estimated the posterior density, $[\boldsymbol{\theta}|\boldsymbol{\mu}]$, for model parameters $\boldsymbol{\theta} \equiv (\sigma^2_\mu, \sigma^2_{ac}, \tau^2, \mathbf{z}, \mathbf{p}, \boldsymbol{\mu}_{SB}, \boldsymbol{\mu}_{CS})'$ with a MCMC algorithm, using standard Metropolis within Gibbs samplers for each parameter. We used visual examination of trace plots, as well as Gelman-Rubin statistics and estimates of effective sample size to assess convergence. A single chain of length 10^4 was generated, with the first half discarded as burnin. Full conditional distributions and additional model fitting details may be found in the Supplementary Material.

Pre-computation

In addition to providing a method for reconciling the time domains at which observations are made and the movement process is modeled, the process imputation procedure provides opportunities for improvements in computational efficiency as well. In the second stage of our fitting procedure, all parameter updates are conditioned on one of K known draws from the distribution of true paths generated in stage 1. Hence, there are only K possible values of $\mathbf{m}_i(t)$, the location on $\mathcal{M}(t)$ closest to $\boldsymbol{\mu}_i(t - 1)$, which we can compute and store before running the MCMC algorithm. Computing $\mathbf{m}_i(t)$ requires evaluating the distance between $\boldsymbol{\mu}_i(t)$ and every point in $\mathcal{M}(t)$,

Table 4.1: Posterior medians and equal-tailed credible intervals for all model parameters, as well as prior distributions and hyper-parameters.

parameter	posterior summary		prior
	median	(2.5%, 97.5%)	density
σ_μ^2	1780	(1680, 1880)	Unif(0, 10^8)
σ_{ac}^2	251000	(211000, 307000)	Unif(0, 10^8)
τ^2	205000	(113000, 368000)	Unif(0, 10^8)
z_i	see Figure 4.3		Bern(p_i)
p_i	see Figure C.2		Beta(0.1, 0.1)

and must be done for all $N = 279$ individuals and all $\mathcal{T}_i, i = 1, \dots, N$ time points. Pre-computing all possible values of $\mathbf{m}_i(t)$ reduces the number of times we must perform the complete search from 10^4 (the number of iterations used in the MCMC algorithm) to 24 (the number of imputed paths) resulting in a substantial decrease in the time required for estimation.

4.3.6 Results

Using the two-stage approach, we fit the model to all observations made by USFWS and USGS during 2008–2016. Table 4.1 provides the posterior medians and equal-tailed 95% credible intervals for each variance parameter, as well as relevant prior distributions and hyperpriors. Figure 4.3 shows the posterior mean of each z_i grouped by the agency that tagged the individuals. Posterior means of the class indicator variables can be interpreted as the posterior probability that individual i is a member of the Chukchi sea sub-population.

To meet our stated goal of producing a meaningful spatial delineation of the two sub-populations from which our study animals were drawn, we use a derived quantity related to the inferred locations of the sub-population activity centers. If we consider the observation of a single new location, $\boldsymbol{\mu}_{J+1}(t)$, and integrate across all arbitrary features, effectively removing the effect of the RSF, it can be shown that the posterior probability that $z_{J+1} = 1$ is given by

$$\Pr(z_{J+1} = 1 | \boldsymbol{\mu}_{J+1}(t), \sigma_{ac}^2, \boldsymbol{\mu}_{CS}, \boldsymbol{\mu}_{SB}) = \frac{p_{J+1} \mathbf{N}(\boldsymbol{\mu}_{CS}, \sigma_{ac}^2 \mathbf{I}_2)}{p_{J+1} \mathbf{N}(\boldsymbol{\mu}_{CS}, \sigma_{ac}^2 \mathbf{I}_2) + (1 - p_{J+1}) \mathbf{N}(\boldsymbol{\mu}_{SB}, \sigma_{ac}^2 \mathbf{I}_2)}.$$

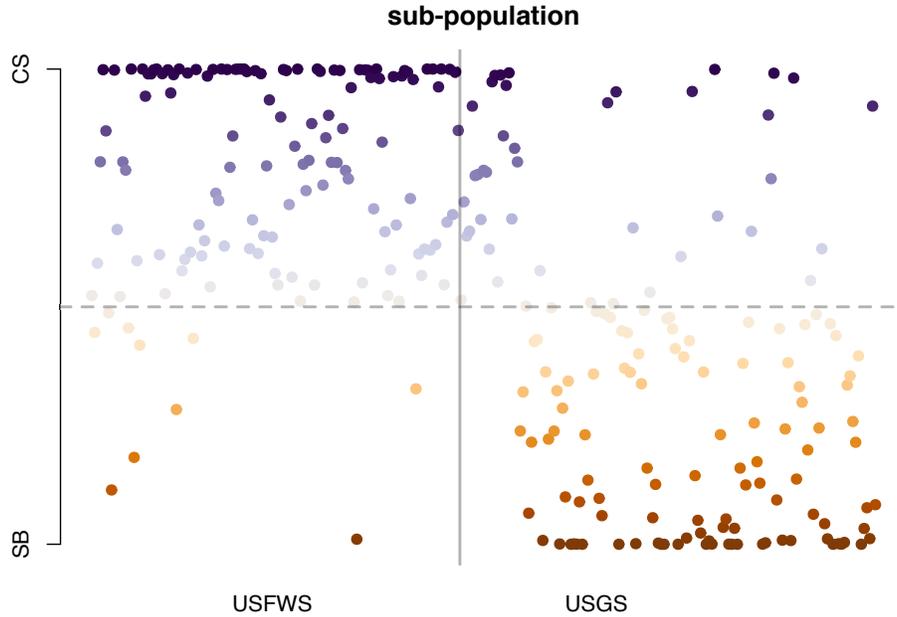


Figure 4.3: Posterior means for each z_i , organized by the agency responsible for tagging the individual. Darker points reflect posterior means closer to 0 (SB, orange) and 1 (CS, purple).

In the case where $p_{J+1} = 0.5$, the contour corresponding to $\Pr(z_{J+1} = 1) = 0.5$ is defined by the points in the plane where the two normal densities are equal. Because the two densities share a common covariance, $\sigma_{ac}^2 \mathbf{I}_2$, it is straightforward to show that this contour coincides with the set of points that are equidistant from the two activity center means, or the perpendicular bisector of the points μ_{CS} and μ_{SB} . By computing the perpendicular bisector at each iteration in the MCMC algorithm, we can infer its posterior distribution, a summary of which provides wildlife managers with a way to delineate the boundary between the two sub-populations.

Figure 4.4 shows a map of the region encompassing the Chukchi and southern Beaufort seas. In the background, weekly measurements of sea ice extent for February–September 2008 are shown as light blue polygons, with the darkest polygon corresponding to open ocean, the second darkest polygon to February 1st, and the lightest polygon to September 30th. Orange and purple lines show the paths drawn in the first stage of the process imputation framework. The colors correspond to the posterior means of z_i , with dark orange hues corresponding to values close to 0, white to values close to 0.5, and dark purple to values close to 1. The solid black line corresponds to the posterior

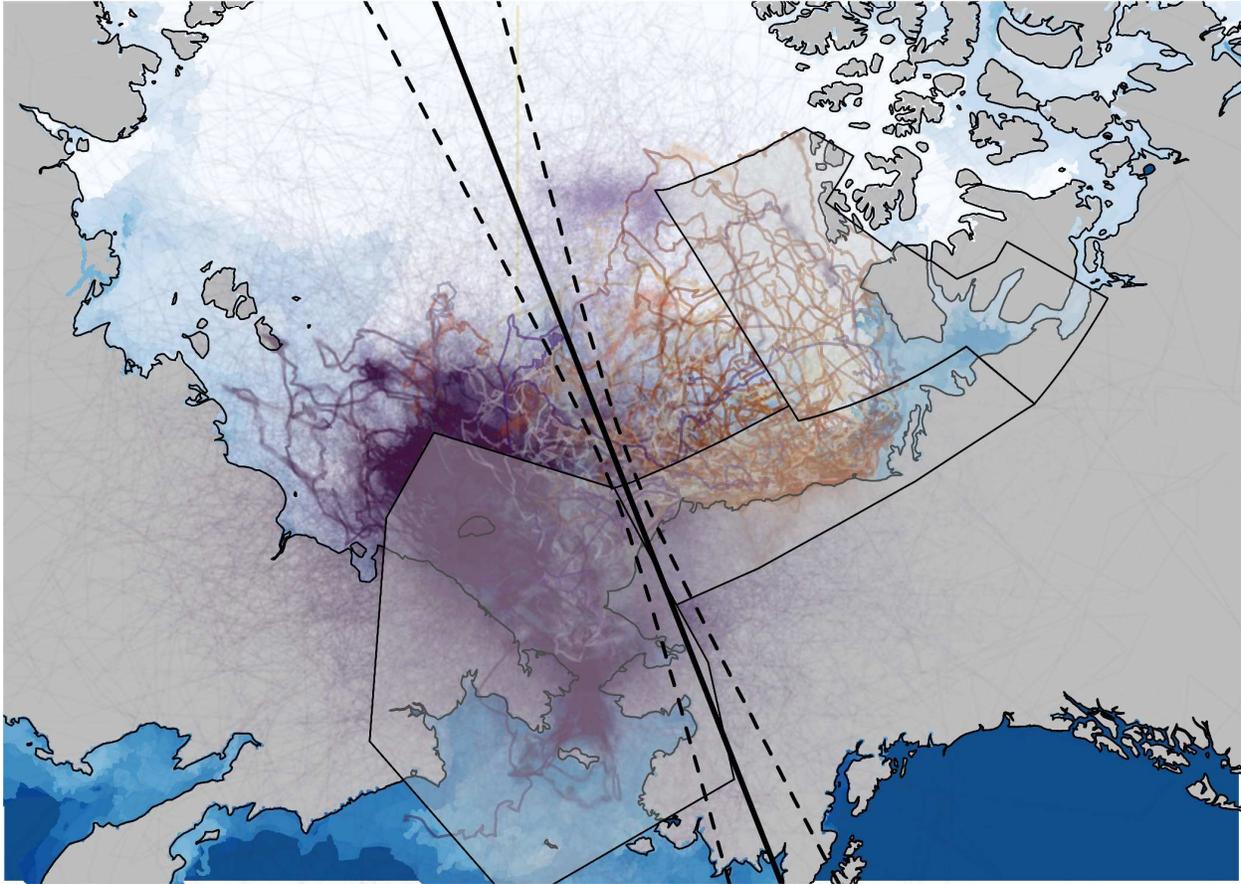


Figure 4.4: The orange and purple lines represent trajectories drawn from the process imputation distributions (Section 4.3.5) of all polar bears from 2008–2016. Orange lines correspond to polar bears associated with the Southern Beaufort Sea sub-population, and purple lines correspond to bears associated with the Chukchi Sea sub-population. The darkness of the color corresponds to the probability we assign to a given sub-population. The black line shows the posterior mean of the boundary where, marginally, the probability of sub-population membership is balanced (Section 4.3.6). The dashed lines show pointwise equal-tailed 95% credible intervals perpendicular to the mean boundary. The large polygons with thin black borders show the current sub-population delineations (CS, SB, and NB from left to right; see Figure 4.2).

mean of the slopes and intercepts of the perpendicular bisector computed at each iteration of the MCMC algorithm. The dashed lines show equal-tailed 95% pointwise credible intervals computed orthogonal to the solid line. The inferred boundary suggests only a slight shift from the currently accepted sub-population delineation, denoted by the large polygons with thin black lines (also shown in Figure 4.2), although our methods for computing the boundary are quite different from those used in previous analyses (Amstrup et al., 2005) and based on newly acquired data.

4.4 Simulation Study

The linearization approximation implemented in the estimation procedure (Section 4.2.2) introduces approximation error into the likelihood calculation unless the shape of the linear boundary is exactly a straight line. In many applications, the shape of a particular feature may be quite complex at the scale of the entire landscape, but well-approximated by a linear shape at the scale of movement increments. If the boundary of the feature is straight enough in the relevant vicinity of the individual reacting to it, then it is reasonable to assume that the approximation error will not adversely affect estimation.

To check the assumption of local linearity, we investigated differences between data simulated from the exact and approximated movement models. We used the same observed union of sea ice boundary and coastline from 2008 that we used in our analysis of the movement of polar bears (Section 4.3), and the posterior medians of model parameters σ_μ^2 and τ^2 to simulate 16 paths over all 366 days from both the exact and approximate movement process models. For the purposes of the simulation study, we disregarded the effect of sub-population activity centers, effectively taking the limit of the movement model in Section 4.3 as $\sigma_{ac}^2 \rightarrow \infty$. We then compared the distributions of three summary statistics: (1) the distance between each simulated location and the nearest point on $\mathcal{M}(t)$, (2) the distance between consecutive locations, and (3) the turning angle between consecutive locations. The first statistic was chosen to reveal important discrepancies in the way the exact and linearized models account for the effect of the RSF. The second and third statistics are commonly investigated quantities in animal movement studies.

Figure 4.5 shows kernel density estimates of the distributions for each statistic. Each of the narrow, faint lines represents the distribution of a statistic for a single individual. The thick, dark lines represent distributions taken across all simulated individuals. The strong similarity between the solid and dashed lines suggest that our linearization method is providing an adequate approximation to the true likelihood.

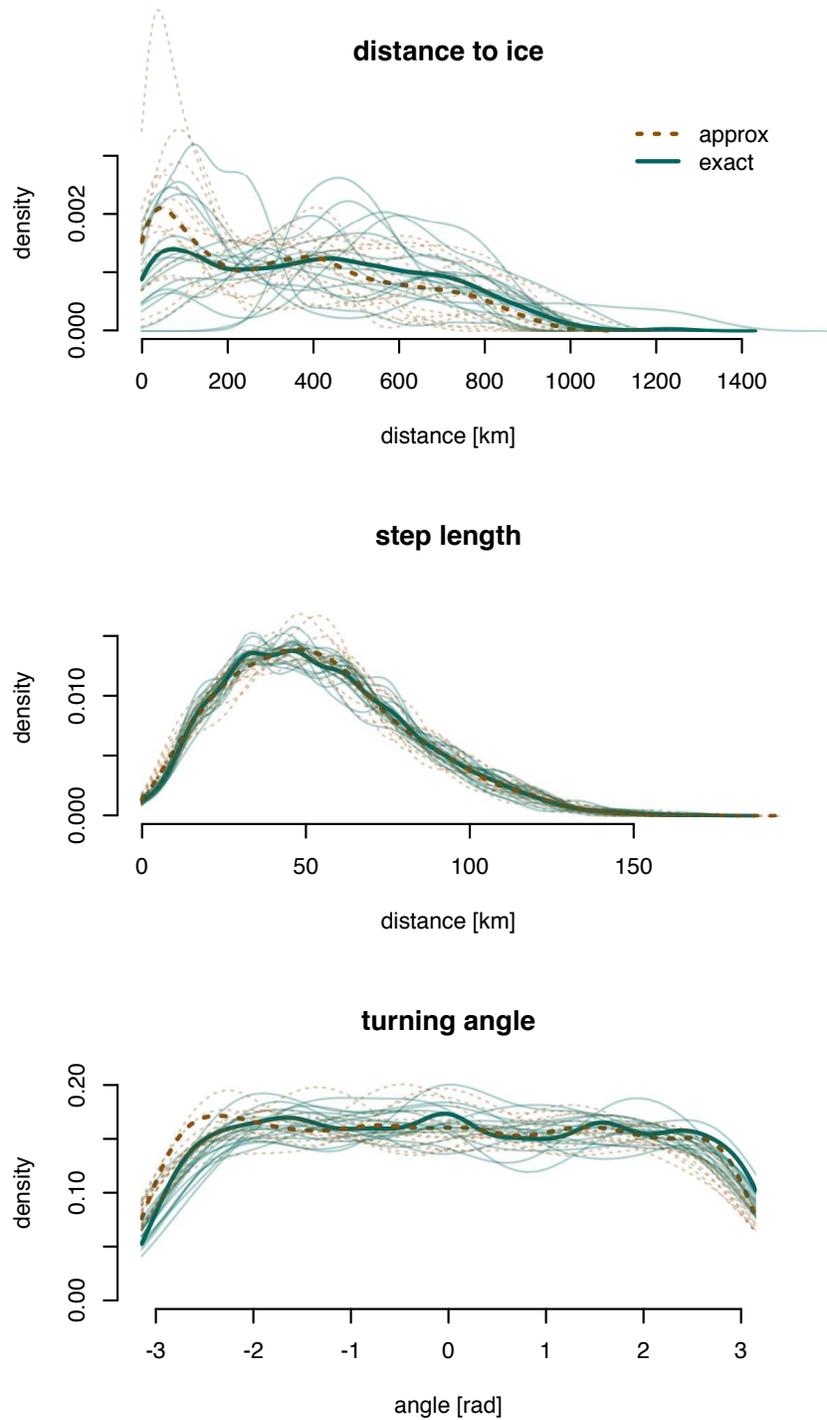


Figure 4.5: Distributions of three summary statistics for paths simulated from the exact (solid) and approximate (dotted) movement models. Thin, faint lines show the distribution of each statistic for each of 16 simulated individuals over 366 days. Thick, dark lines show the distribution of each statistic across all individuals and times.

4.5 Discussion

Our novel approximation method based on the linearization of a potentially time-varying spatial feature, $\mathcal{M}(t)$, reduces the computational burden of fitting models in the common RSF framework, allowing researchers increased flexibility in the types of RSFs they can specify in mechanistically-driven models for movement. We demonstrated our approach in an application involving the movement of polar bears as they respond to seasonal shifts in sea ice over the course of 2008–2016. There has been a recent interest in research focused on the so-called “greenwave” hypothesis (e.g., Aikens et al., 2017), which posits that herbivorous animals align their movement with bands of high-quality forage that shift throughout spring as different elevations and latitudes experience phenological changes. Our modeling approach represents a way to validate the hypothesis if information about the shape of the greenwave is known, or potentially estimate the location of the posited band of high quality forage based on the observed movement patterns of herbivores.

The linearization approximation methodology can also be extended to higher-dimensional spaces and features. For instance, in marine environments, RSFs based on two-dimensional features, such as isotherms, may be locally approximated using rank-deficient Gaussian distributions corresponding to infinite planes. One-dimensional features in three-dimensional spaces, such as wind or ocean currents, can also be approximated with improper distributions, where the rank of the covariance matrix is deficient by a degree of 2. In practice, researchers may need to evaluate the appropriateness of a linearization approximation; however, in many cases, our methodology offers a way to include complex drivers of movement that might otherwise have been computationally inaccessible.

Chapter 5

Conclusion

I have described two approaches for jointly modeling the trajectories of a population of interacting individuals, and a novel approximation technique that makes approximate inference available for RSF-based models incorporating complex environmental features. Several proposed extensions were discussed in the closing sections of Chapters 2-4. I conclude with additional ideas for future work not already addressed in previous chapters.

5.1 Models for dependent trajectories

The primary limitations of the models I developed to describe the trajectories of interacting individuals are computational. At present, implementation is limited to studies involving on the order of a dozen individuals on a temporal grid of hundreds of time points. Many data sets related to animal movement can include hundreds of individuals over time domains requiring much finer grids. To quote Lindgren et al. (2011), ‘Although the computational power today is at an all time high, the tendency seems to be that the dimension n is always set, or we want to set it, a little higher than the value that gives a reasonable computation time.’ The computational bottleneck in our estimation procedure involves the need to repeatedly invert the large, dense covariance matrix that describes the Gaussian process modeling the true continuous movement process. Two potentially fruitful directions for future research involve (1) approximating the covariance inversion using randomized matrix decomposition techniques (Halko et al., 2011), and (2) modifying the model such that we avoid the need to invert the covariance matrix at all. I present preliminary considerations related to each.

5.1.1 Randomized matrix decompositions

Frequently in spatio-temporal modeling and, in particular, for models presented in Chapters 2 and 3, covariance matrices are of the form $\Sigma = \sigma^2\mathbf{I} + \mathbf{C}$, where \mathbf{C} is a dense covariance matrix

that describes the dependence in an unobserved process, and σ^2 describes the variance of the measurement error process, which is independent in time. When Σ is dense and of rank n , the computational complexity of inverting Σ is $\mathcal{O}(n^3)$. However, if \mathbf{C} can be decomposed into the form $\mathbf{U}_{n \times m} \mathbf{D}_{m \times m} \mathbf{V}'_{n \times m}$, the complexity reduces to $\mathcal{O}(n^2 m)$ through use of the Sherman-Morrison-Woodbury matrix identity (Hager, 1989).

A common approach both to reducing the dimensionality and computational complexity of a statistical model is to replace large matrices with low-rank approximations. For the case of covariance matrices, the approximation produced using the first m eigen-vectors and -values is optimal under multiple matrix discrepancy norms by the Eckart-Young-Mirsky theorem (Eckart and Young, 1936). Moreover, it is well-established that matrices considerably less than full rank often provide high quality approximations. Thus, a natural approach to reducing the computational demand of fitting the models proposed in Chapters 2 and 3 involves replacing the exact covariance matrix, \mathbf{C} , with its optimal rank- m approximation, where m is chosen to be as large as computational resources allow.

The problem with an optimal low-rank approximation approach is that computing the exact eigen-decomposition of \mathbf{C} has the same complexity as computing the inverse. However, several algorithms have been developed that use random projections to rapidly approximate the first m eigen-vectors and values. Theoretical results have been derived that place probabilistic bounds on the discrepancy between these approximate decompositions and their exact counterparts (Halko et al., 2011). Randomized matrix decompositions have already been implemented successfully in the spatial statistics literature (e.g., Banerjee et al., 2013b), and may present an opportunity that makes approximate inference available for large data sets.

5.1.2 Modeling the precision matrix

Another way to address the computational burden of inverting large dense covariance matrices is to develop statistical models that directly construct the inverse, or precision matrix, thereby avoiding the computationally expensive calculation altogether. This philosophy underpins the suc-

successful modeling approach represented widely in the spatial and spatio-temporal literature of using Gaussian Markov random fields (GMRFs) to describe stochastic processes (e.g., Rue and Held, 2005b). GMRFs are models for Gaussian random vectors in which the precision matrix is specified to reflect conditional dependence based on a neighborhood structure and, hence, were originally designed for processes existing on a regular lattice of locations and/or time points. The Markovian assumption usually implies that the precision matrix is sparse, leading to further computational benefits. For the case when locations do not occur on a lattice, triangulations of the spatial and/or temporal domain can provide a useful notion of a neighborhood structure (Lindgren et al., 2011).

GMRFs have been successfully used to model dependence in time and space; however, we are not aware of any application of GMRFs to the case of multiple, dependent trajectories. Process convolution chains provide a way to account for dependence in time and across individuals separately without assuming a separable covariance structure. No such chaining procedure is immediately obvious for modeling precision matrices, but the development of analogous methodologies could provide valuable tools with excellent computational properties.

5.2 Resource Selection Functions

In the application to polar bears, we make the assumption that the parameter characterizing polar bear preference for space near ice edges and coastlines, τ^2 , is constant in time. This assumption ignores findings in previous work that concluded polar bears show strongest preference for space near the ice edge in the fall and summer (Durner et al., 2009), and could be relaxed by allowing τ^2 to vary in time. We also assume that the sub-population centers are constant in time. Auger-Méthé et al. (2016) recently showed that drifting pack ice, which can move several kilometers a day, can complicate estimation of individual activity centers. Further, changing conditions in the arctic due to climate change may result, or may have already resulted, in shifting sub-population boundaries. Future research could incorporate dynamic sub-population centers by allowing the activity center variances, σ_{ac}^2 to vary in time as well.

The primary limitation in the proposed linearization approximation for RSFs is the inflexibility of using a straight line at each timepoint to approximate a general surface. However, the conjugacy induced by the linearized surface requires that the RSF be represented by the density of a (possibly improper) Gaussian distribution, and this constraint cannot be readily relaxed. When the assumption of local linearity causes non-ignorable approximation error, our current approach cannot be used.

Another limitation of the linearization technique in Chapter 4 is the assumption of a Gaussian-like function form for both the availability and resource selection functions. While the form of the availability function we specify is common in the literature, the RSF is not, and there are many applications where the functional form would represent a clear misspecification. The symmetry of the Gaussian distributional form, for example, implies that we cannot use it to model an individual's preference for areas of the landscape near a coastline, but restricted to land.

One way to generalize our method would be to use mixtures of (possibly improper) Gaussian distribution functions to represent a feature of interest in the landscape (McLachlan and Peel, 2004). Mixtures could, in principle, contain components corresponding to points (using proper Gaussian densities), lines, and higher dimensional hyperplanes for movement in $d > 2$ dimensions, and would allow for greater flexibility in the form of the RSF that could accommodate multimodality and asymmetry.

5.3 Conclusion

The study of animal movement can reveal a wealth of useful information about ecology of a population, and the behavior of groups and individuals therein (Turchin 1998; Hooten et al. 2017a; Hooten et al. 2017b). As technological advances allow for increasingly sophisticated telemetry measurements to be made, we will be able to learn even more from remotely sensed data, provided we develop the methodological tools needed to analyze such dense data sets (Kays et al., 2015).

As research continues, collaboration among statisticians, ecologists, and wildlife biologists will remain a critical link. As the volume of data we hope to analyze continues to grow, it is

inevitable that computational demands will require researchers to make strong assumptions, simplifications, and approximations about the processes observed in nature. Only by leveraging the collective knowledge base spread among a wide range of related fields can those important aspects of statistical and mathematical modeling be adequately addressed.

Some of the work I have presented here might be described as a new solution to an old problem. Chapter 4 centers around a RSF-based movement model, for which the primary novel contribution addresses an intractable normalizing constant. Complementary to these types of advances, some of the work I present addresses more nascent scientific questions that are less clearly defined in the scientific literature. The models for dependent trajectories in Chapters 2 and 3 were originally motivated by asking what one *could* learn from movement data as much as by what one expected to learn from them. The data used in the applications within those chapters were not gathered with the intent that they would be scrutinized for evidence of dynamic social connections, yet the results of my analysis show that such knowledge is learnable from movement data. Part of the impact of this work may be that behavioral biologists now view remotely-sensed location observations as a more affordable, non-invasive way to study complex social behavior, which may, in turn, guide the types of questions they hope to answer. As discussed in Section 4.5, the findings in Chapter 4 may have a direct impact on the design of future deployments of telemetry devices.

Bibliography

- Aarts, G., Fieberg, J., and Matthiopoulos, J. (2012). Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods in Ecology and Evolution*, 3(1):177–187.
- Aikens, E. O., Kauffman, M. J., Merkle, J. A., Dwinnell, S. P., Fralick, G. L., and Monteith, K. L. (2017). The greenscape shapes surfing of resource waves in a large migratory herbivore. *Ecology Letters*, 20(6):741–750.
- Amstrup, S. C., Durner, G. M., Stirling, I., and McDonald, T. L. (2005). Allocating harvests among polar bear stocks in the Beaufort Sea. *Arctic*, 53(3):247–259.
- Andrews, R. D., Pitman, R. L., and 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.
- Arthur, S. M., Manly, B. F. J., McDonald, L. L., and Garner, G. W. (1996). Assessing habitat selection when availability changes. *Ecology*, 77(1):215–227.
- Atwood, T. C., Peacock, E., McKinney, M. A., Lillie, K., Wilson, R. R., Douglas, D. C., Miller, S., and Terletzky, P. (2016). Rapid environmental change drives increased land use by an Arctic marine predator. *PLoS ONE*, 11(6):e0155932.
- Auger-Méthé, M., Lewis, M. A., and Derocher, A. E. (2016). Home ranges in moving habitats: Polar bears and sea ice. *Ecography*, 39(1):26–35.
- Baird, R. W. and Whitehead, H. (2000). Social organization of mammal-eating killer whales: Group stability and dispersal patterns. *Canadian Journal of Zoology*, 78(12):2096–2105.
- Banerjee, A., Dunson, D. B., and Tokdar, S. T. (2013a). Efficient Gaussian process regression for large datasets. *Biometrika*, 100(1):75–89.

- Banerjee, A., Vogelstein, J., and Dunson, D. B. (2013b). Parallel inversion of huge covariance matrices. *arXiv preprint arXiv:1312.1869*.
- Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848.
- Barry, R. P. and Ver Hoef, J. M. (1996). Blackbox kriging: Spatial prediction without specifying variogram models. *Journal of Agricultural, Biological, and Environmental Statistics*, 1(3):297–322.
- Berliner, L. M. (1996). Hierarchical bayesian time series models. In Hanson, K. M. and Silver, R. N., editors, *Maximum entropy and Bayesian methods*, Fundamental Theories of Physics, pages 15–22. Springer Netherlands.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 36(2):192–236.
- Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82(4):733–746.
- Blackwell, P. (1997). Random diffusion models for animal movement. *Ecological Modelling*, 100(1):87–102.
- Bolin, D. and Lindgren, F. (2013). A comparison between Markov approximations and other methods for large spatial data sets. *Computational Statistics & Data Analysis*, 61:7–21.
- Brillinger, D. R. (2003). Simulating constrained animal motion using stochastic differential equations. *Lecture Notes-Monograph Series*, 41(2003):35–48.
- Brillinger, D. R. and Stewart, B. S. (1998). Elephant-seal movements: Modelling migration. *Canadian Journal of Statistics*, 26(3):431–443.

- Brost, B. M., Hooten, M. B., Hanks, E. M., and Small, R. J. (2015). Animal movement constraints improve resource selection inference in the presence of telemetry error. *Ecology*, 96(10):2590–2597.
- Brost, B. M., Hooten, M. B., and Small, R. J. (2016). Leveraging constraints and biotelemetry data to pinpoint repetitively used spatial features. *Ecology*, 98(1):12–20.
- Buderman, F. E., Hooten, M. B., Ivan, J. S., and Shenk, T. M. (2016). A functional model for characterizing long-distance movement behaviour. *Methods in Ecology and Evolution*, 7(3):264–273.
- Cagnacci, F., Boitani, L., Powell, R. A., and Boyce, M. S. (2010). Animal ecology meets GPS-based radiotelemetry: A perfect storm of opportunities and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1550):2157–2162.
- 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.
- Chetkiewicz, C.-L. B. and Boyce, M. S. (2009). Use of resource selection functions to identify conservation corridors. *Journal of Applied Ecology*, 46(5):1036–1047.
- Christ, A., Hoef, J. V., and Zimmerman, D. L. (2008). An animal movement model incorporating home range and habitat selection. *Environmental and Ecological Statistics*, 15(1):27–38.
- Codling, E. A. and Bode, N. W. (2014). Copycat dynamics in leaderless animal group navigation. *Movement Ecology*, 2(11).
- Costa, D. P., Robinson, P. W., Arnould, J. P. Y., Harrison, A. L., Simmons, S. E., Hassrick, J. L., Hoskins, A. J., Kirkman, S. P., Oosthuizen, H., Villegas-Amtmann, S., and Crocker, D. E.

- (2010). Accuracy of ARGOS locations of pinnipeds at-sea estimated using fastloc GPS. *PLoS ONE*, 5(1):e8677.
- Cressie, N. (1991). *Statistics for Spatial Data*. John Wiley & Sons, New York, New York, USA, first edition.
- Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):209–226.
- Cressie, N. and Pavlicová, M. (2002). Calibrated spatial moving average simulations. *Statistical Modelling*, 2(4):267–279.
- Croft, D. P., James, R., and Krause, J. (2008). *Exploring Animal Social Networks*. Princeton University Press.
- Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812.
- Davis, M. J., Thokala, S., Xing, X., Thompson Hobbs, N., Walsh, D. P., Han, R. Y., and Mishra, S. (2012). Developing a data-transfer model for a novel Wildlife-tracking network. *Wildlife Society Bulletin*, 36(4):820–827.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer-Verlag, New York.
- Doob, J. L. (1944). The elementary Gaussian processes. *Annals of Mathematical Statistics*, 15(3):229–282.
- Dunn, J. E. and Gipson, P. S. (1977). Analysis of radio telemetry data in studies of home range. *Biometrics*, 33(1):85.
- Durante, D. and Dunson, D. B. (2014). Nonparametric Bayes dynamic modeling of relational data. *Biometrika*, 101(4):1–16.

- Durante, D., Scarpa, B., and 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., and Pitman, R. L. (2017). Morphological and ecological evidence for two sympatric forms of Type B killer whale around the Antarctic Peninsula. *Polar Biology*, 40(1):231–236.
- Durban, J. W. and 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.
- Durner, G. M., Douglas, D. C., Albeke, S. E., Whiteman, J. P., Amstrup, S. C., Richardson, E., Wilson, R. R., and Ben-David, M. (2017). Increased Arctic sea ice drift alters adult female polar bear movements and energetics. *Global Change Biology*, 23(9):3460–3473.
- Durner, G. M., Douglas, D. C., Nielson, R. M., Amstrup, S. C., McDonald, T. L., Stirling, I., Mauritzen, M., Born, E. W., Wiig, Ø., DeWeaver, E., Serreze, M. C., Belikov, S. E., Holland, M. M., Maslanik, J., Aars, J., Bailey, D. A., and Derocher, A. E. (2009). Predicting 21st-century polar bear habitat distribution from global climate models. *Ecological Monographs*, 79(1):25–58.
- Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218.
- Fetterer, F., Savoie, M., Helfrich, S., and Clemente-Colón, P. (2010). Multisensor analyzed sea ice extent - Northern hemisphere (MASIE-NH), version 1, updated daily. *National Ice Center and National Snow and Ice Data Center*, 2008-2016:Boulder, Colorado USA. accessed August 13, 2017.
- Feynman, R. P., Leighton, R. B., and Sands, M. (1963). *The Feynman Lectures on Physics*. Addison-Wesley, Reading, Massachusetts.

- Finley, A. O., Banerjee, S., and Gelfand, A. E. (2015). spBayes for large univariate and multivariate point-referenced spatio-temporal data models. *Journal of Statistical Software*, 63(13).
- Fleming, C. H., Fagan, W. F., Mueller, T., Olson, K. A., Leimgruber, P., and 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.
- Forester, J. D., Ives, A. R., Turner, M. G., Anderson, D. P., Fortin, D., Beyer, H. L., Smith, D. W., and Boyce, M. S. (2007). State-space models link elk movement patterns to landscape characteristics in Yellowstone National Park. *Ecological Monographs*, 77(2):285–299.
- Franz, M., Mclean, E., Tung, J., Altmann, J., and Alberts, S. C. (2015). Self-organizing dominance hierarchies in a wild primate population. *Proceedings of the Royal Society B: Biological Sciences*, 282(1814):20151512.
- Furrer, R., Genton, M. G., and Nychka, D. W. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523.
- Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M. (2010). *Handbook of spatial statistics*. CRC Press Taylor & Francis Group, Boca Raton, Florida, USA.
- Gelfand, A. E., Kim, H.-J., Sirmans, C. F., and 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., and 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., and Wittemyer, G. (2014). Controlling for behavioural state reveals social dynamics among male African elephants, *Loxodonta africana*. *Animal Behaviour*, 95:111–119.
- Hager, W. W. (1989). Updating the inverse of a matrix. *SIAM Review*, 31(2):221–239.

- Halko, N., Martinsson, P. G., and Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288.
- Hanks, E. M., Hooten, M. B., and Alldredge, M. W. (2015a). Continuous-time discrete-space models for animal movement. *Annals of Applied Statistics*, 9(1):145–165.
- Hanks, E. M., Hooten, M. B., Johnson, D. S., and Sterling, J. T. (2011). Velocity-based movement modeling for individual and population level inference. *PLoS ONE*, 6(8):e22795.
- Hanks, E. M., Johnson, D. S., and Hooten, M. B. (2017). Reflected stochastic differential equation models for constrained animal movement. *Journal of Agricultural, Biological, and Environmental Statistics*, 22(3):353–372.
- Hanks, E. M., Schliep, E. M., Hooten, M. B., and Hoeting, J. A. (2015b). Restricted spatial regression in practice: Geostatistical models, confounding, and robustness under model misspecification. *Environmetrics*, 26(4):243–254.
- 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., and Fryxell, J. M. (2008). Socially informed random walks: Incorporating group dynamics into models of population spread and growth. *Proceedings of the Royal Society B: Biological Sciences*, 275(1638):1101–1109.
- Hays, G., Åkesson, S., Godley, B., Luschi, P., and Santidrian, P. (2001). The implications of location accuracy for the interpretation of satellite-tracking data. *Animal Behaviour*, 61(5):1035–1040.
- 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*, pages 37–56. Springer.
- Hjermann, D. O. (2000). Analyzing habitat selection in animals without well defined home range. *Ecology*, 81(5):1462–1468.
- Hobbs, N. T. and Hanley, T. A. (1990). Habitat evaluation: Do use/availability data reflect carrying capacity? *Journal of Wildlife Management*, 54(4):515–522.
- Hoff, P. D. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems*.
- Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.
- Hooten, M. B. and Johnson, D. S. (2017). Basis function models for animal movement. *Journal of the American Statistical Association*, 112(518):578–589.
- Hooten, M. B., Johnson, D. S., Hanks, E. M., and Lowry, J. H. (2010). Agent-based inference for animal movement and selection. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(4):523–538.
- Hooten, M. B., Johnson, D. S., McClintock, B. T., and Morales, J. M. (2017a). *Animal Movement: Statistical Models for Telemetry Data*. Chapman & Hall/CRC, Boca Raton, Florida, USA.
- Hooten, M. B., King, R., and Langrock, R. (2017b). Guest editor's introduction to the special issue on "animal movement modeling". *Journal of Agricultural, Biological, and Environmental Statistics*, 22(3):224–231.
- Johnson, D. S. (2016). crawl: Fit continuous-time correlated random walk models to animal movement data.
- Johnson, D. S., Hooten, M. B., and Kuhn, C. E. (2013). Estimating animal resource selection from telemetry data using point process models. *Journal of Animal Ecology*, 82(6):1155–1164.

- Johnson, D. S., London, J. M., and Kuhn, C. E. (2011). Bayesian inference for animal space use and other movement metrics. *Journal of Agricultural, Biological, and Environmental Statistics*, 16(3):357–370.
- Johnson, D. S., London, J. M., Lea, M.-A., and Durban, J. W. (2008a). Continuous-time correlated random walk model for animal telemetry data. *Ecology*, 89(5):1208–15.
- Johnson, D. S., Thomas, D. L., Ver Hoef, J. M., and Christ, A. (2008b). A general framework for the analysis of animal resource selection from telemetry data. *Biometrics*, 64(3):968–976.
- Jonsen, I. D., Flemming, J. M., and Myers, R. A. (2005). Robust state-space modeling of animal movement data. *Ecology*, 86(11):2874–2880.
- Kahle, D. and 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.
- Kays, R., Crofoot, M. C., Jetz, W., and Wikelski, M. (2015). Terrestrial animal tracking as an eye on life and planet. *Science*, 348(6240):aaa2478.
- Krause, J., Croft, D. P., and James, R. (2007). Social network theory in the behavioural sciences: Potential applications. *Behavioral Ecology and Sociobiology*, 62(1):15–27.
- 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(2):190–199.
- Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., and Morales, J. M. (2012). Flexible and practical modeling of animal telemetry data: Hidden Markov models and extensions. *Ecology*, 93(11):2336–2342.

- Lele, S. R. and Keim, J. L. (2006). Weighted distributions and estimation of resource selection probability functions. *Ecology*, 87(12):3021–3028.
- Lemasson, B. H., Anderson, J. J., and Goodwin, R. A. (2013). Motion-guided attention promotes adaptive communications during social navigation. *Proceedings of the Royal Society B: Biological Sciences*, 280(1754):20122003.
- Levin, I. I., Zonana, D. M., Burt, J. M., and Safran, R. J. (2015). Performance of encounter-net tags: Field tests of miniaturized proximity loggers for use on small birds. *PLOS ONE*, 10(9):e0137242.
- Lindgren, F., Rue, H., and Lindstrom, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.
- Manly, B. F. J., Mcdonald, L. L., and Thomas, D. L. (2002). *Resource Selection by Animals: Statistical Design and Analysis for Field Studies*. Springer, second edition.
- McClintock, B. T., Johnson, D. S., Hooten, M. B., Ver Hoef, J. M., and Morales, J. M. (2014). When to be discrete: The importance of time formulation in understanding animal movement. *Movement Ecology*, 2(21).
- McClintock, B. T., London, J. M., Cameron, M. F., and Boveng, P. L. (2015). Modelling animal movement using the Argos satellite telemetry location error ellipse. *Methods in Ecology and Evolution*, 6(3):266–277.
- McCullagh, P. and Nelder, J. (1983). *Generalized Linear Models*. Chapman and Hall, New York, NY, USA.
- McLachlan, G. and Peel, D. (2004). *Finite Mixture Models*. John Wiley & Sons.
- Miller, P. J. O. (2006). Diversity in sound pressure levels and active space of killer whale vocalizations. *Journal of Comparative Physiology A*, 192(5):449–459.

- Millspaugh, J. J. and Marzluff, J. M., editors (2001). *Radio Tracking and Animal Populations*. Academic Press, San Diego, CA.
- Morales, J. M., Moorcroft, P. R., Matthiopoulos, J., Frair, J. L., Kie, J. G., Powell, R. A., Merrill, E. H., and Haydon, D. T. (2010). Building the bridge between animal movement and population dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1550):2289–2301.
- Morin, P. A., Parsons, K. M., Archer, F. I., Ávila-Arcos, M. C., Barrett-Lennard, L. G., Dalla Rosa, L., Duchêne, S., Durban, J. W., Ellis, G. M., Ferguson, S. H., Ford, J. K. B., Ford, M. J., Garilao, C., Gilbert, M. T. P., Kaschner, K., Matkin, C. O., Petersen, S. D., Robertson, K. M., Visser, I. N., Wade, P. R., Ho, S. Y. W., and Foote, A. D. (2015). Geographical and temporal dynamics of a global radiation and diversification in the killer whale. *Molecular Ecology*, 24(15):3964–3979.
- Myers, D. E. (1982). Matrix formulation of co-kriging. *Journal of the International Association for Mathematical Geology*, 14(3):249–257.
- Niu, M., Blackwell, P. G., and Skarin, A. (2016). Modeling interdependent animal movement in continuous time. *Biometrics*, 72(2):315–324.
- Northrup, J. M., Hooten, M. B., Anderson, C. R. J., and Wittemyer, G. (2013). Practical guidance on characterizing availability in resource selection functions under a use-availability design. *Ecology*, 94(7):1456–1463.
- Nychka, D. W., Bandyopadhyay, S., Hammerling, D., Lindgren, F., and 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.
- Obbard, M. E., Thiemann, G. W., Peacock, E., and Debruyne, T. D. (2010). Proceedings of the 15th working meeting of the IUCN/SCC polar bear specialist group, 29 June - 3 July 2009, Copenhagen, Denmark. In *IUCN*, volume vii, page 235.

- Paciorek, C. J. (2003). *Nonstationary Gaussian processes for regression and spatial modelling*. PhD thesis, Carnegie Mellon University.
- Parsons, K. M., Balcomb, K. C., Ford, J. K. B., and Durban, J. W. (2009). The social dynamics of southern resident killer whales and conservation implications for this endangered population. *Animal Behaviour*, 77(4):963–971.
- Pinter-Wollman, N., Hobson, E. A., Smith, J. E., Edelman, A. J., Shizuka, D., De Silva, S., Waters, J. S., Prager, S. D., Sasaki, T., Wittemyer, G., Fewell, J., and McDonald, D. B. (2014). The dynamics of animal social networks: Analytical, conceptual, and theoretical advances. *Behavioral Ecology*, 25(2):242–255.
- Pitman, R. L. and Durban, J. W. (2010). Killer whale predation on penguins in Antarctica. *Polar Biology*, 33(11):1589–1594.
- Pitman, R. L. and Durban, J. W. (2012). Cooperative hunting behavior, prey selectivity and prey handling by pack ice killer whales (*Orcinus orca*), Type B, in Antarctic Peninsula waters. *Marine Mammal Science*, 28(1):16–36.
- Pitman, R. L. and Ensor, P. (2003). Three forms of killer whales (*Orcinus orca*) in Antarctic waters. *Journal of Cetacean Research and Management*, 5(2):131–140.
- Potts, J. R., Bastille-Rousseau, G., Murray, D. L., Schaefer, J. A., and 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.
- Prange, S., Jordan, T., Hunter, C., and Gehrt, S. D. (2006). New radiocollars for the detection of proximity among individuals. *Wildlife Society Bulletin*, 34(5):1333–1344.
- Rode, K. D., Regehr, E. V., Douglas, D. C., Durner, G. M., Derocher, A. E., Thiemann, G. W., and Budge, S. M. (2014). Variation in the response of an Arctic top predator experiencing habitat loss: Feeding and reproductive ecology of two polar bear populations. *Global Change Biology*, 20(1):76–88.

- Rode, K. D., Wilson, R. R., Regehr, E. V., Martin, M. S., Douglas, D. C., and Olson, J. W. (2015). Increased land use by Chukchi Sea polar bears in relation to changing sea ice conditions. *PLoS ONE*, 10(11):e0142213.
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, 91(434):473–489.
- Rue, H. and Held, L. (2005a). Continuous time random walks. In *Gaussian Markov Random Fields: Theory and Application*, chapter 3.5, pages 123–130. Chapman and Hall/CRC.
- Rue, H. and Held, L. (2005b). *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC.
- Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.
- 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(1):22–40.
- Rutz, C. and Hays, G. C. (2009). New frontiers in biologging science. *Biology Letters*, 5(3):289–292.
- Sarkar, P. and Moore, A. W. (2006). Dynamic social network analysis using latent space models. In *Advances in Neural Information Processing Systems*.
- Scharf, H. R., Hooten, M. B., Fosdick, B. K., Johnson, D. S., London, J. M., and Durban, J. W. (2016). Dynamic social networks based on movement. *Annals of Applied Statistics*, 10(4):2182–2202.

- Scharf, H. R., Hooten, M. B., and Johnson, D. S. (2017). Imputation approaches for animal movement modeling. *Journal of Agricultural, Biological, and Environmental Statistics*, 22(3):335–352.
- Service Argos (2016). Argos user’s manual. URL <http://www.argos-system.org/manual/>.
- Sewell, D. K. and Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512):1646–1657.
- Shepp, L. A. (1966). Radon-Nikodym derivatives of Gaussian measures. *Annals of Mathematical Statistics*, 37(2):321–354.
- Sih, A., Hanser, S. F., and McHugh, K. A. (2009). Social network theory: New insights and issues for behavioral ecologists. *Behavioral Ecology and Sociobiology*, 63(7):975–988.
- Steele, M., Dickinson, S., Zhang, J., and W. Lindsay, R. (2015). Seasonal ice loss in the Beaufort Sea: Toward synchrony and prediction. *Journal of Geophysical Research: Oceans*, 120(2):1118–1132.
- Turchin, P. (1998). *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*. Sinauer Associates, Sunderland, MA.
- U.S. Fish and Wildlife (2016). Polar bear (*Ursus maritimus*) conservation management plan, final. *U.S. Fish and Wildlife, Region 7, Anchorage, Alaska*, pages 1–104.
- Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 50(2):297–312.
- Ver Hoef, J. M. and Barry, R. P. (1998). Constructing and fitting models for cokriging and multi-variable spatial prediction. *Journal of Statistical Planning and Inference*, 69(907):275–294.
- Ver Hoef, J. M., Peterson, E., and Theobald, D. (2006). Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics*, 13(4):449–464.

- Ver Hoef, J. M. and 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.
- Ware, J. V., Rode, K. D., Bromaghin, J. F., Douglas, D. C., Wilson, R. R., Regehr, E. V., Amstrup, S. C., Durner, G. M., Pagano, A. M., Olson, J. W., Robbins, C. T., and Jansen, H. T. (2017). Habitat degradation affects the summer activity of polar bears. *Oecologia*, 184(1):87–99.
- Warton, D. I. and Shepherd, L. C. (2010). Poisson point process models solve the "pseudo-absence problem" for presence-only data in ecology. *Annals of Applied Statistics*, 4(3):1383–1402.
- Wecker, W. P. and Ansley, C. F. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *Journal of the American Statistical Association*, 78(381):81–89.
- Wey, T., Blumstein, D. T., Shen, W., and Jordán, F. (2008). Social network analysis of animal behaviour: A promising tool for the study of sociality. *Animal Behaviour*, 75(2):333–344.
- White, G. C. and Garrott, R. A. (1990). *Analysis of Wildlife Radio-Tracking Data*. Academic Press, San Diego, CA.
- Wikle, C. K. (2010). Low-rank representations for spatial processes. In *Handbook of Spatial Statistics*, chapter 8, pages 107–118. CRC Press Taylor & Francis Group, Boca Raton, FL.
- Williams, R. and Lusseau, D. (2006). A killer whale social network is vulnerable to targeted removals. *Biology Letters*, 2(4):497–500.
- Williams, R., Trites, A. W., and Bain, D. E. (2002). Behavioural responses of killer whales (*Orcinus orca*) to whale-watching boats: Opportunistic observations and experimental approaches. *Journal of Zoology*, 256(2):255–270.
- Wilson, R. R., Regehr, E. V., Rode, K. D., and St Martin, M. (2016). Invariant polar bear habitat selection during a period of sea ice loss. *Proceedings of the Royal Society B: Biological Sciences*, 283(1836):20160380.

Appendix A

Supplemental Material for Chapter 2

For all fits, we acquire 200,000 iterations on a single computing node, discard the first 100,000 draws as burnin, and thin by a factor of 5 to yield a sample of size 20,000.

We employ diffuse priors for most parameters in both the simulation and application.

A.1 Priors

Prior distributions are selected to be conjugate where possible.

$$\sigma^2 \sim \text{IG}(a_\sigma, b_\sigma) \tag{A.1}$$

$$[\alpha] \propto (1 + \alpha)^{\alpha_\alpha - 1} (1 - \alpha)^{\beta_\alpha - 1} \mathbf{1}_{(-1,1)}(\alpha) \tag{A.2}$$

This is the kernel for the density of a random variable $X = 2Y - 1$ where $Y \sim \text{Beta}(\alpha_\alpha, \beta_\alpha)$. It is the same shape in a sense as the Beta, but spread over the support $[-1, 1]$, so that we allow for the possibility of anti-alignment. See Section A.1 for details.

$$\beta \sim \text{N}(0, \sigma_\beta^2) \tag{A.3}$$

$$c \sim \text{IG}(a_c, b_c) \tag{A.4}$$

$$\phi \sim \text{Beta}(\alpha_\phi, \beta_\phi) \tag{A.5}$$

$$p_1 \sim \text{Beta}(\alpha_{p_1}, \beta_{p_1}) \tag{A.6}$$

Prior for the strength of the alignment effect

The relevant support for α is the real interval $(-1, 1)$. Negative values for α correspond to the case when connected individuals tend to move in parallel, but opposite, directions. We generally expect α to be positive, but to allow for the possibility of anti-alignment behavior we specify a prior on the full support. To specify a flexible family of prior distributions with support $(-1, 1)$,

we shift and scale a Beta distribution to align with $(-1, 1)$. In fitting our model with an MCMC algorithm, α is sampled with a Metropolis-Hastings step, therefore we only need the kernel of the full-conditional

$$[\alpha | \alpha_\alpha, \beta_\alpha] \propto (1 + \alpha)^{\alpha_\alpha - 1} (1 - \alpha)^{\beta_\alpha - 1} I_{(-1,1)}(\alpha). \quad (\text{A.7})$$

A.2 Full conditionals

Define the following variables as:

$$n \equiv \text{number of individuals in study population} \quad (\text{A.8})$$

$$T \equiv \text{number of discrete time steps in mean position process} \quad (\text{A.9})$$

It is convenient to also define

$$\mathbf{K}(t) \equiv (\mathbf{W}_+^c(t) - \alpha \mathbf{W}(t)) \otimes \mathbf{I}_2 \quad (\text{A.10})$$

so that

$$\mathbf{Q}(t) = \sigma^{-2} \mathbf{K}(t). \quad (\text{A.11})$$

σ : variability of mean step process

Gibbs step.

$$\sigma^2 | \cdot \sim \text{IG}(a^*, b^*) \quad (\text{A.12})$$

$$a^* \equiv a_\sigma + nT \quad (\text{A.13})$$

$$b^* \equiv b_\sigma + \frac{1}{2} \sum_{t=2}^T (\boldsymbol{\mu}(t) - \boldsymbol{\mu}(t-1) - \beta \tilde{\boldsymbol{\mu}}(t-1))' \mathbf{K}(t) (\boldsymbol{\mu}(t) - \boldsymbol{\mu}(t-1) - \beta \tilde{\boldsymbol{\mu}}(t-1)) \quad (\text{A.14})$$

W: network

Gibbs step.

$$\Pr(w_{ij}(1) = 1|\cdot) \propto [\boldsymbol{\mu}_i(2)|\dots, w_{ij}(1) = 1] [\boldsymbol{\mu}_j(2)|\dots, w_{ij}(1) = 1] \quad (\text{A.15})$$

$$\times p_{1|1}^{w_{ij}(2)} p_{0|1}^{1-w_{ij}(2)} p_1 \quad (\text{A.16})$$

$$\Pr(w_{ij}(1) = 0|\cdot) \propto [\boldsymbol{\mu}_i(2)|\dots, w_{ij}(1) = 0] [\boldsymbol{\mu}_j(2)|\dots, w_{ij}(1) = 0] \quad (\text{A.17})$$

$$\times p_{1|0}^{w_{ij}(2)} p_{0|0}^{1-w_{ij}(2)} (1 - p_1) \quad (\text{A.18})$$

$$(\text{A.19})$$

$$\Pr(w_{ij}(t) = 1|\cdot) \propto [\boldsymbol{\mu}_i(t)|\dots, w_{ij}(t) = 1] [\boldsymbol{\mu}_j(t)|\dots, w_{ij}(t) = 1] \quad (\text{A.20})$$

$$\times [\boldsymbol{\mu}_i(t+1)|\dots, w_{ij}(t) = 1] [\boldsymbol{\mu}_j(t+1)|\dots, w_{ij}(t) = 1] \quad (\text{A.21})$$

$$\times p_{1|1}^{w_{ij}(t+1)} p_{0|1}^{1-w_{ij}(t+1)} p_{1|0}^{1-w_{ij}(t-1)} p_{1|1}^{w_{ij}(t-1)} \quad (\text{A.22})$$

$$\Pr(w_{ij}(t) = 0|\cdot) \propto [\boldsymbol{\mu}_i(t)|\dots, w_{ij}(t) = 0] [\boldsymbol{\mu}_j(t)|\dots, w_{ij}(t) = 0] \quad (\text{A.23})$$

$$\times [\boldsymbol{\mu}_i(t+1)|\dots, w_{ij}(t) = 0] [\boldsymbol{\mu}_j(t+1)|\dots, w_{ij}(t) = 0] \quad (\text{A.24})$$

$$\times p_{1|0}^{w_{ij}(t+1)} p_{0|0}^{1-w_{ij}(t+1)} p_{0|1}^{w_{ij}(t-1)} p_{0|0}^{1-w_{ij}(t-1)} \quad (\text{A.25})$$

$$(\text{A.26})$$

$$\Pr(w_{ij}(T) = 1|\cdot) \propto [\boldsymbol{\mu}_i(T)|\dots, w_{ij}(T) = 1] [\boldsymbol{\mu}_j(T)|\dots, w_{ij}(T) = 1] \quad (\text{A.27})$$

$$\times p_{1|0}^{1-w_{ij}(T-1)} p_{1|1}^{w_{ij}(T-1)} \quad (\text{A.28})$$

$$\Pr(w_{ij}(T) = 0|\cdot) \propto [\boldsymbol{\mu}_i(T)|\dots, w_{ij}(T) = 0] [\boldsymbol{\mu}_j(T)|\dots, w_{ij}(T) = 0] \quad (\text{A.29})$$

$$\times p_{0|1}^{w_{ij}(T-1)} p_{0|0}^{1-w_{ij}(T-1)} \quad (\text{A.30})$$

α : alignment/anti-alignment

M-H step. Kernel of full conditional density is

$$[\alpha|\cdot] \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n \sum_{t=2}^T w_{i+}(t) (\mathbf{h}_i(t) - \alpha \bar{\mathbf{h}}_i(t))' (\mathbf{h}_i(t) - \alpha \bar{\mathbf{h}}_i(t)) \right\} \quad (\text{A.31})$$

$$\times (1 + \alpha)^{\alpha\alpha-1} (1 - \alpha)^{\beta\alpha-1} \mathbf{1}_{(-1,1)}(\alpha) \quad (\text{A.32})$$

where

$$\mathbf{h}_i(t) \equiv \boldsymbol{\mu}_j(t) - \boldsymbol{\mu}_j(t-1) - \beta \tilde{\boldsymbol{\mu}}_j(t-1) \quad (\text{A.33})$$

$$\bar{\mathbf{h}}_i(t) \equiv \sum_{j \neq i} \frac{w_{ij}(t)}{w_{i+}^c(t)} \mathbf{h}_j(t) \quad (\text{A.34})$$

and $\bar{\mathbf{h}}_i(t) = 0$ if animal i is completely unconnected at time t . Proposal distribution is normal:

$$\text{N}(\alpha_{\text{iter}-1}, \sigma_{\alpha\text{-tune}}^2) \quad (\text{A.35})$$

β : attraction/repulsion

Gibbs step.

$$\beta | \cdot \sim \text{N}(\mu_\beta^*, \sigma_\beta^{*2}) \quad (\text{A.36})$$

$$\sigma_\beta^{*2} \equiv \left(\sum_{t=2}^T \tilde{\boldsymbol{\mu}}(t-1)' \mathbf{Q}(t) \tilde{\boldsymbol{\mu}}(t-1) + \frac{1}{\sigma_\beta^2} \right)^{-1} \quad (\text{A.37})$$

$$\mu_\beta^* \equiv \sigma_\beta^{*2} \left(\sum_{t=2}^T \tilde{\boldsymbol{\mu}}(t-1)' \mathbf{Q}(t) (\boldsymbol{\mu}(t) - \boldsymbol{\mu}(t-1)) + \mu_\beta \right) \quad (\text{A.38})$$

ϕ : network stability

M-H step. Kernel of full conditional density is

$$[\phi|\cdot] \propto (p_{0|0})^{g_{0|0}}(p_{1|0})^{g_{1|0}}(p_{0|1})^{g_{0|1}}(p_{1|1})^{g_{1|1}}(\phi)^{\alpha_\phi-1}(1-\phi)^{\beta_\phi-1} \quad (\text{A.39})$$

$$g_{0|0} \equiv \sum_{i<j} \sum_{t=2}^T (1-w_{ij}(t-1))(1-w_{ij}(t)) \quad (\text{A.40})$$

$$g_{1|0} \equiv \sum_{i<j} \sum_{t=2}^T (1-w_{ij}(t-1))w_{ij}(t) \quad (\text{A.41})$$

$$g_{0|1} \equiv \sum_{i<j} \sum_{t=2}^T w_{ij}(t-1)(1-w_{ij}(t)) \quad (\text{A.42})$$

$$g_{1|1} \equiv \sum_{i<j} \sum_{t=2}^T w_{ij}(t-1)w_{ij}(t) \quad (\text{A.43})$$

with proposals coming from a Beta distribution centered on ϕ_{iter-1} :

$$\text{Beta} \left(\frac{\beta_{\phi\text{-tune}}\phi_{iter-1}}{1-\phi_{iter-1}}, \beta_{\phi\text{-tune}} \right) \quad (\text{A.44})$$

p_1 : network density

M-H step. Kernel of full conditional density is

$$[p_1|\cdot] \propto (p_1)^{\sum_{i<j} w_{ij}(1)}(1-p_1)^{\sum_{i<j} (1-w_{ij}(1))} (p_{0|0})^{g_{0|0}}(p_{1|0})^{g_{1|0}}(p_{0|1})^{g_{0|1}}(p_{1|1})^{g_{1|1}} \\ \times (p_1)^{\alpha_{p_1}-1}(1-p_1)^{\beta_{p_1}-1} \quad (\text{A.45})$$

with proposals coming from a Beta distribution centered on $p_{1iter-1}$:

$$\text{Beta} \left(\frac{\beta_{p_1\text{-tune}}p_{1iter-1}}{1-p_{1iter-1}}, \beta_{p_1\text{-tune}} \right) \quad (\text{A.46})$$

c : precision for unconnected animals

Gibbs step.

$$c|\cdot \sim \text{IG} \left(a_c + \sum_{i=1}^n \sum_{t=2}^T \mathbf{1}_{\{w_{i+}(t)=0\}}, b_c + \frac{1}{2} \sum_{i=1}^n \sum_{t=2}^T \mathbf{1}_{\{w_{i+}(t)=0\}} \mathbf{h}_i(t)' \mathbf{h}_i(t) \right) \quad (\text{A.47})$$

A.3 Hyper-/tuning-parameters

Table A.1: Simulation

parameter	hyper parameters		tuning
α	$\alpha_\alpha = 1$	$\beta_\alpha = 1$	$\sigma_{\alpha-tune}^2 = 0.1^2$
β	$\mu_\beta = 0$	$\sigma_\beta^2 = 1000$	conjugate
p_1	$\alpha_{p_1} = 1$	$\beta_{p_1} = 1$	$\beta_{p_1-tune} = 24$
ϕ	$\alpha_\phi = 100$	$\beta_\phi = \frac{100}{9}$	$\beta_{\phi-tune} = 7$
c	$a_c = 3.5$	$b_c = 1.5$	conjugate
σ	$a_\sigma = 0.1$	$b_\sigma = 0.001$	conjugate

parameter	hyper parameters		tuning
α	$\alpha_\alpha = 1$	$\beta_\alpha = 1$	$\sigma_{\alpha-tune}^2 = 0.01^2$
β	$\mu_\beta = 0$	$\sigma_\beta^2 = 1000$	conjugate
p_1	$\alpha_{p_1} = 1$	$\beta_{p_1} = 1$	$\beta_{p_1-tune} = 10$
ϕ	$\alpha_\phi = 100$	$\beta_\phi = \frac{100}{9}$	$\beta_{\phi-tune} = 5$
c	$a_c = 3.5$	$b_c = 1.5$	conjugate
σ	$a_\sigma = 0.1$	$b_\sigma = 0.001$	conjugate

Table A.2: Killer whales

A.4 Dynamic Social Network between killer whale types

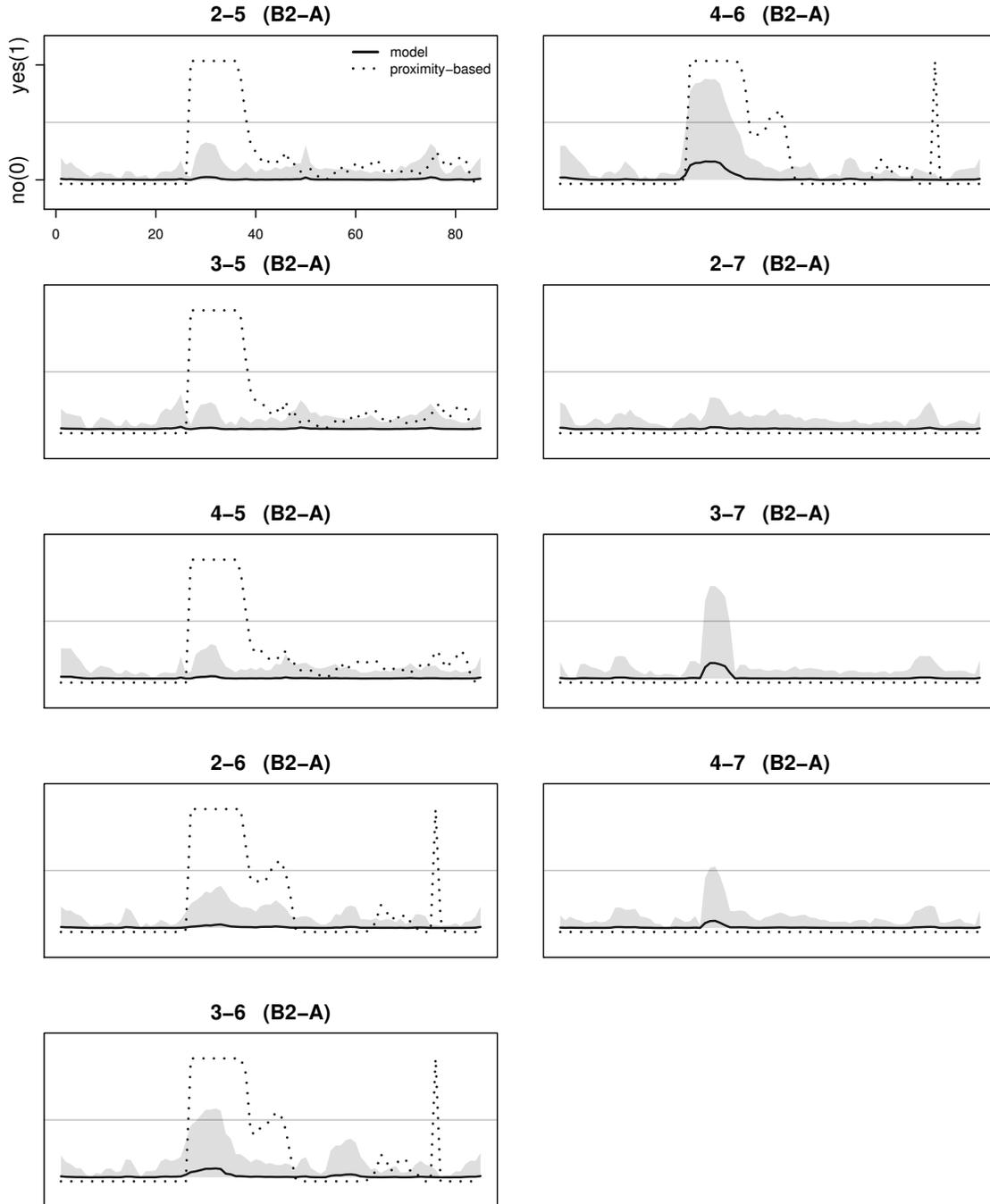


Figure A.1: A selection of the $\binom{7}{2} = 21$ possible pairs of individuals in the killer whale study sample. The plots displayed are for all inter-type pairs of killer whales of type B2 (labeled 2, 3, 4) and A (labeled 5, 6, 7). The solid line in each plot shows the posterior mean for w_{ij} and the gray region represents one standard deviation above and below the posterior mean. The dotted line shows the network defined by \mathbf{W}^R , where individuals are deemed connected whenever they are separated by a distance less than R . No posterior means above 0.5 were predicted for inter-type connections. (Note: The lines are offset slightly near 0 and 1 for visual clarity.)

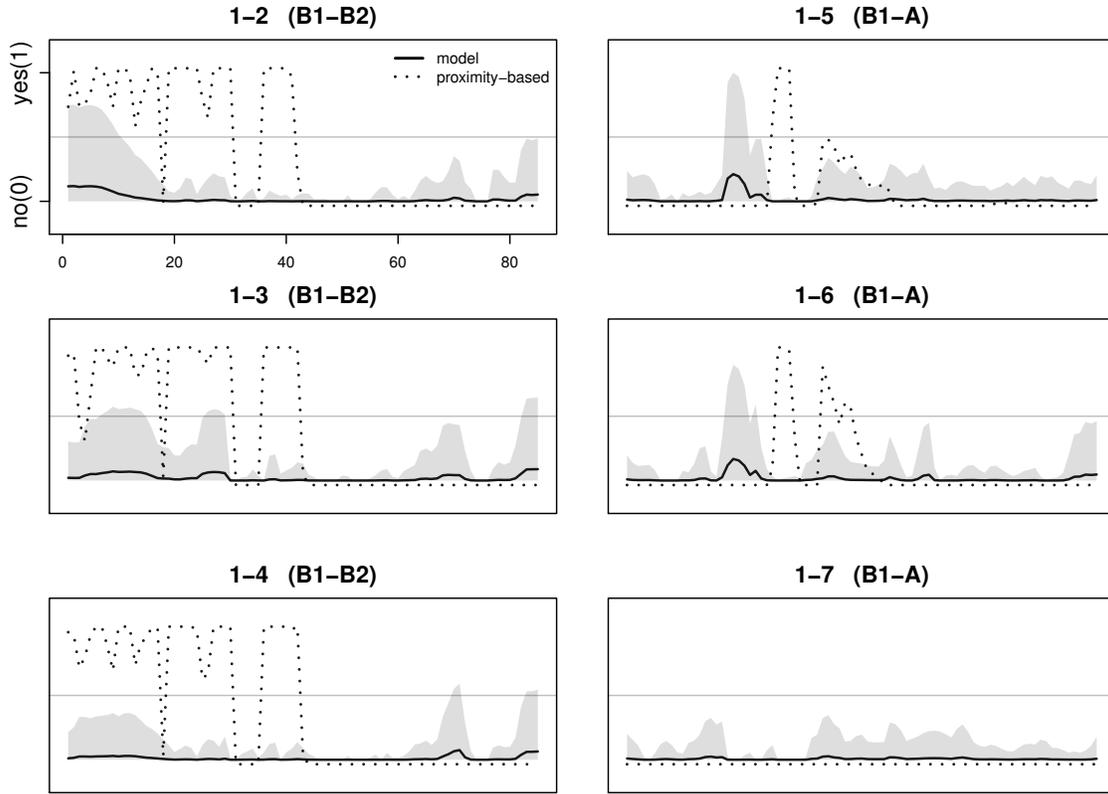


Figure A.2: A selection of the $\binom{7}{2} = 21$ possible pairs of individuals in the killer whale study sample. The plots displayed are for all inter-type pairs of killer whales between the sole individual of type B1 (labeled 1) and those of type B2 (labeled 2, 3, 4) and A (labeled 5, 6, 7). The solid line in each plot shows the posterior mean for w_{ij} and the gray region represents one standard deviation above and below the posterior mean. The dotted line shows the network defined by \mathbf{W}^R , where individuals are deemed connected whenever they are separated by a distance less than R . No posterior means above 0.5 were predicted for inter-type connections. (Note: The lines are offset slightly near 0 and 1 for visual clarity.)

Appendix B

Supplemental Material for Chapter 3

B.1 Model specification

Data model

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

$$\begin{aligned} \mathbf{s} &\sim \mathbf{N}\left(\mathbf{0}, \sigma_s^2 \left(\mathbf{I} + \sigma_{\mu/s}^2 \Delta\tau \tilde{\mathbf{H}}' \Sigma_{dB} \tilde{\mathbf{H}}\right)\right) \\ \tilde{\mathbf{H}} &= \mathbf{H}^{(inl)}(\phi_{inl}) \mathbf{H}^{(soc)}(dB_w, \phi_w, \sigma_w^2) \mathbf{H}^{(bm)} \\ \Sigma_{dB}(t, \tau, i, j) &= \mathbf{1}_{\{t=\tau\}} \left(\sigma_0^2 \mathbf{1}_{\{t=0\}} + \mathbf{1}_{\{t>0\}}\right) \mathbf{1}_{\{i=j\}} \\ \mathbf{H}^{(bm)}(t, \tau, i, j) &= \mathbf{1}_{\{\tau < t\}} \mathbf{1}_{\{i=j\}} \\ \mathbf{H}^{(soc)}(t, \tau, i, j) &= \mathbf{1}_{\{\tau=t\}} \frac{w_{ij}(t)}{|w_i(t)|} \\ \mathbf{H}^{(inl)}(t, \tau_{soc}, i, j) &\equiv \frac{|t - \tau_{soc}|}{\phi_{inl}} K_1(|t - \tau_{soc}|/\phi_{inl}) \mathbf{1}_{\{i=j\}} \end{aligned}$$

Process model (integrated out for model fitting)

$$\begin{aligned} \boldsymbol{\tau} &\equiv (\tau_1 = 0, \tau_2, \dots, \tau_{m-1}, \tau_m = 1)' \\ \Delta\tau_i &\equiv \tau_i - \tau_{i-1}, \quad 1 < i \leq m \\ dB(\tau_i) &\sim \begin{cases} \mathbf{N}(0, \sigma_0^2); & i = 1 \\ \mathbf{N}(0, \Delta\tau_i); & i > 1 \end{cases} \\ dB &= (dB(\tau_1), \dots, dB(\tau_m))' \\ \tilde{\boldsymbol{\mu}} &= \mathbf{H}^{(inl)} \mathbf{H}^{(soc)} \mathbf{H}^{(bm)} dB = \tilde{\mathbf{H}} dB \end{aligned}$$

Prior model

$$\begin{aligned}\phi_{inl} &\sim \text{Gamma}(\alpha_s, \beta_s) & \sigma_0^2 &\sim \text{IG}(a_0, b_0) \\ \sigma_{\mu/s}^2 &\sim \text{IG}(a_{\mu/s}, b_{\mu/s}) & \sigma_s^2 &\sim \text{IG}(a_s, b_s)\end{aligned}$$

Application only:

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

Fixed parameters

Simulation study:

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

Application:

$$\phi_w$$

B.2 Model fitting details

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 B.1. For each fit, we obtained 4,000 iterations and discard the first 3,000 as burnin. The entire simulation study was parallelized across 6 3GHz cores and required approximately one week of computation time.

Table B.1: True values and prior distributions used in 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})

Killer whales

We acquired 100,000 iterations on a single computing node, and discarded the first 50,000 as burnin 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 hours of computing time.

Table B.2: Posterior credible intervals and prior distributions for the IP-DEP model in killer whale application.

parameter	posterior (IP-DEP)		posterior (IP-IND)		prior density
	median	(2.5%, 97.5%)	median	(2.5%, 97.5%)	
ϕ_{inl}	0.00848	(0.00756, 0.00954)	0.00850	(0.00762, 0.00951)	Gamma(2, 100)
σ_0^2	705	(292, 2040)	1440	(627, 4510)	IG(1, 10)
$\sigma_{\mu/s}^2$	3710	(2600, 5490)	1700	(1380, 2400)	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)

Appendix C

Supplemental Material for Chapter 4

C.1 MCMC Implementation details

Stage 1

We used the Ornstein-Uhlenbeck movement model of Johnson et al. (2008a) as the process model for the imputation distribution in the process imputation framework (Scharf et al., 2017). For each individual in the study, we define the first-stage model by pairing the process model in Johnson et al. (2008a) and the measurement error model described in Section 3.4 that matched the individual’s device type. We then used the R package `crawl` (Johnson, 2016) to fit the hierarchical model to the individual’s telemetry observations, and drew $K = 24$ realizations from the posterior distribution of the continuous movement process. Each realized draw was evaluated on a daily time scale corresponding to the rate at which we have observations of sea ice extent. Figure C.1 shows realizations from the imputation distribution for two example polar bears in 2008.

The approximate time required to fit the imputation distributions and sample 24 paths for each of the 279 individuals on a computer with a 3GHz processor was 3 hours. Pre-computing the values of $\mathbf{m}_i(t)$ across all $K = 24$ paths for $i = 279$, and all times t required approximately 12 hours (see Section 4.3.5).

Stage 2

We drew realizations from the posterior distribution of the parameters $\boldsymbol{\theta} \equiv (\sigma_\mu^2, \sigma_{ac}^2, \tau^2, \mathbf{z}, \mathbf{p}, \boldsymbol{\mu}_{SB}, \boldsymbol{\mu}_{CS})'$ using Metropolis within Gibbs updates in an MCMC procedure. At each stage of the MCMC algorithm, we draw one of the 24 paths with equal probability from the process imputation distribution generated in stage 1. We ran the MCMC algorithm for a total of 10^4 iterations, discarding the first half as burnin. Updating the MCMC algorithm for 10^4 iterations required approximately 8 days.

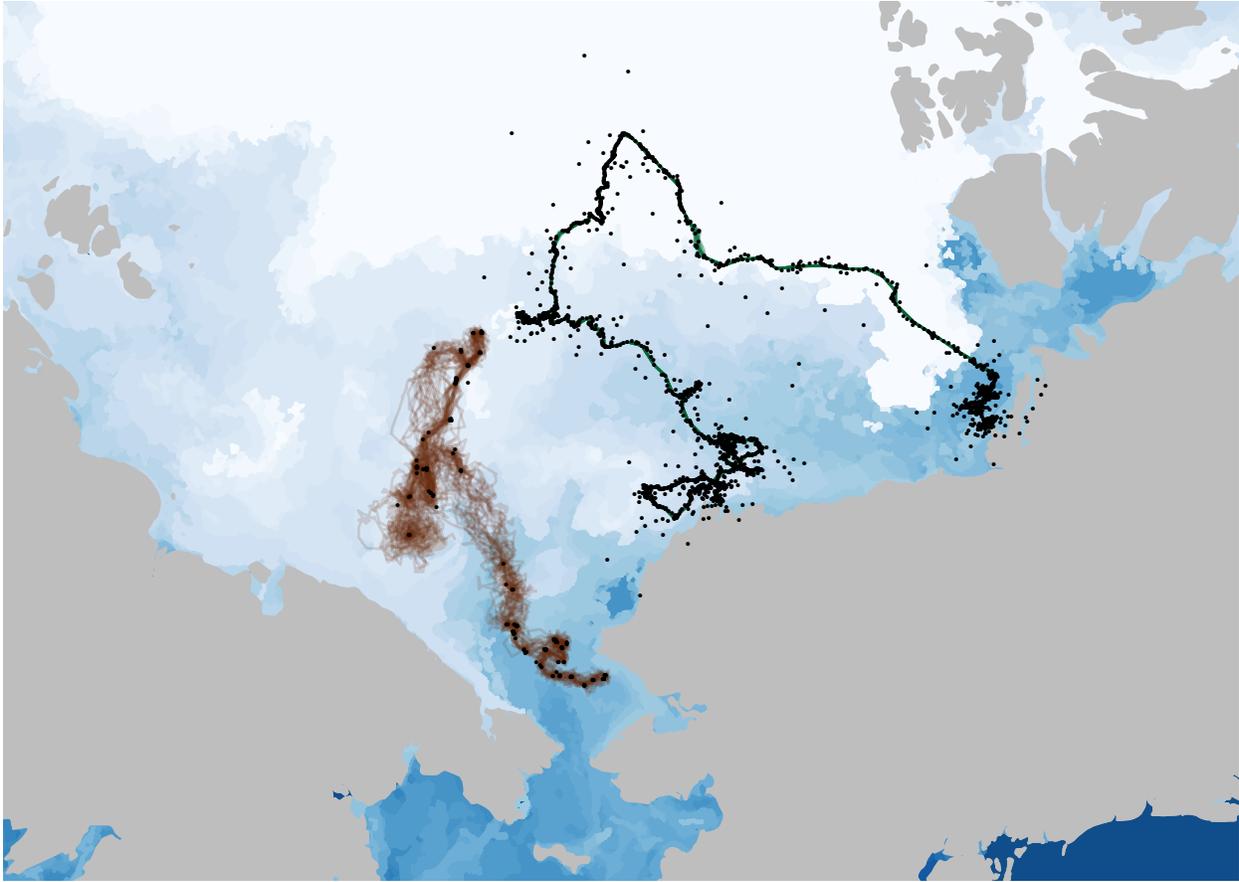


Figure C.1: Realizations from the process imputation distribution for two polar bears in 2008. The polar bear to the right has fewer observations with greater measurement uncertainty, leading to greater variability among the sampled paths. The polar bear on the right has dense observations from both GPS and Argos devices. Sampled paths are much less variable.

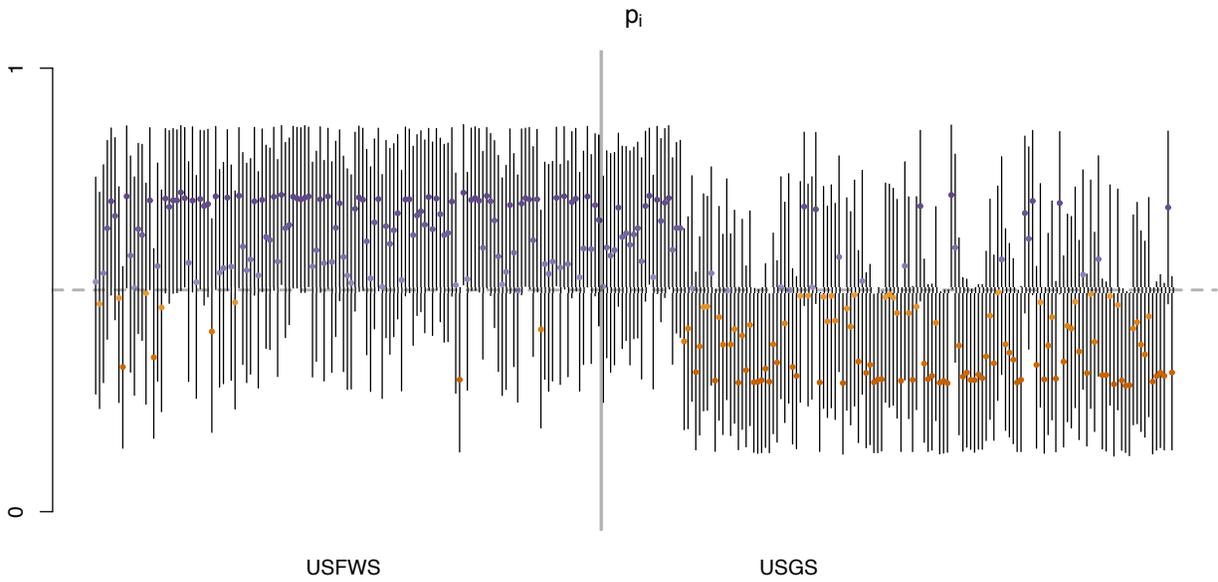


Figure C.2: Posterior medians for each p_i , organized by the agency responsible for tagging the individual. Darker points reflect posterior medians closer to 0 (SB, orange) and 1 (CS, purple). Vertical lines represent posterior 1st and 3rd quartiles marginally for each p_i .