The Pennsylvania State University
The Graduate School

ADVANCES IN STOCHASTIC MODELS FOR ANIMAL MOVEMENT AND ASSESSMENT OF ATTITUDES TOWARD PROBABILITY

A Dissertation in
Statistics
by
Elizabeth Eisenhauer

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2022

The dissertation of Elizabeth Eisenhauer was reviewed and approved by the following:

Ephraim Hanks
Associate Professor of Statistics
Chair of Graduate Studies
Dissertation Co-Advisor
Co-Chair of Committee

Matthew Beckman
Assistant Research Professor of Statistics
Director of Undergraduate Programs
Dissertation Co-Advisor
Co-Chair of Committee

Stephen Berg
Assistant Professor of Statistics

Frances Buderman
Assistant Professor of Quantitative Wildlife Ecology

Abstract

In this dissertation, I will discuss three research projects: two in the field of environmental statistics and one in the field of statistics education. In Chapter 2, we propose a novel sampling design for animal movement studies which combines samples at regular and random time intervals. We compare our novel sampling design, called lattice and random intermediate point (LARI), to regular sampling designs in two data examples and one simulation example. In these cases, LARI sampling leads to more accurate and precise parameter estimation compared to regular sampling. In Chapter 3, we describe a flexible model for golden eagles (*Aquila chrysaetos*) and other partially migrating species. We compare our proposed approach using varying coefficients to latent-state models, and we show that our approach better describes dispersal, migration, and resident behaviors. The flexibility of our approach also allows us to model behavior of less stereotypical individuals. In Chapter 4, we describe the development, deployment, and analyses of the Survey of Probability Attitudes (SPA), which we adapted from the existing Survey of Attitudes Toward Statistics (SATS-36). We present validity evidence for the SPA and used mixed effects models to model the gain in attitude component scores from the beginning to the end of a probability course. We illustrate the effect of regression to the mean in an attitude study and discuss the implications for future work.

Table of Contents

List of Figures vii

List of Tables xiii

Acknowledgments xiv

Chapter 1
Introduction 1
1.1 Overview of Research 1

Chapter 2
A Lattice and Random Intermediate Point Sampling Design for Animal Movement 3
2.1 Introduction 3
2.2 A Lattice and Random Intermediate Points Sampling Scheme 5
2.3 Stochastic Differential Equation Model for Animal Movement 6
2.3.1 Numerical Approximations 8
2.4 Simulation Example 9
2.4.1 Simulation from an SDE Model with Quadratic Potential Function 9
2.4.2 Parameter Identifiability 11
2.4.3 Estimation of Model Parameters and Missing Values 12
2.4.4 Simulation Example Results 13
2.5 Guppy Data Example 17
2.5.1 SDE Model 18
2.5.2 Results 18
2.6 Carpenter Ant Example 20
2.6.1 Ant Movement Model 21
2.6.2 Results 22
2.7 Discussion 27
2.8 Data availability 29
S2 Supplement to a Lattice and Random Intermediate Point Sampling Design for Animal Movement 29
S2.1 Examples of Model Capabilities 29
S2.2 Simulation Example without Infill 31

S2.3 Bayesian framework . . . 32
S2.4 Estimation . . . 33
S2.5 Simulation study of the estimation procedure . . . 37

Chapter 3
A Flexible Movement Model for Partially Migrating Species **41**
3.1 Introduction . . . 41
3.2 Golden Eagle Telemetry Data . . . 43
3.3 A Stochastic Differential Equation Model Framework for Animal Movement 46
3.4 Flexible Model for Partially Migratory Species . . . 47
3.5 Resident Example . . . 47
3.5.1 Single-State Residence Model for Comparison . . . 48
3.5.2 Resident Results . . . 49
3.6 Dispersal Example . . . 50
3.6.1 Latent-State Dispersal Model for Comparison . . . 50
3.6.2 Dispersal Results . . . 51
3.7 Migrant Example . . . 53
3.7.1 Latent-State Migrant Model for Comparison . . . 54
3.7.2 Migrant Results . . . 56
3.8 Fitting Boundary Individuals . . . 58
3.9 Discussion and Future Work . . . 58
S3 Supplement to A Flexible Movement Model for Partially Migrating Species 60
S3.1 Justification for Daily Observations . . . 60
S3.2 Additional Plots for Single-State Resident Model . . . 64
S3.3 Additional Plots for Latent-State Dispersal Model . . . 65
S3.4 Additional Plots for Latent-State Migrant Model . . . 66
S3.5 Varying Coefficients . . . 68

Chapter 4
Survey of Probability Attitudes **72**
4.1 Introduction . . . 72
4.2 Survey of Attitudes Toward Statistics-36 . . . 73
4.3 Methods . . . 74
4.3.1 Development of the Survey of Probability Attitudes . . . 74
4.3.2 SPA Distribution and Scoring . . . 76
4.3.3 Validity Approach . . . 76
4.3.4 Analysis Approach . . . 77
4.4 Results . . . 78
4.4.1 Validity Evidence for the SPA . . . 80
4.4.1.1 Scoring . . . 81
4.4.1.2 Generalization . . . 81
4.4.1.3 Extrapolation . . . 84
4.4.1.4 Implications . . . 86
4.4.2 Analysis of Attitude Changes from Pre- to Post-SPA . . . 86

4.4.2.1 t-Procedures 86
4.4.2.2 Linear Model for Probability Students 86
4.4.2.3 Linear Model for Probability and Non-probability Students 89
4.5 Discussion 90
4.5.1 Limitations 91
4.5.2 Implications for Teaching 92
4.5.3 Implications for Research 92
S4 Supplement to Survey of Probability Attitudes 92
S4.1 Pre-Survey of Probability Attitudes for Probability Students . . . 92
S4.2 Post-Survey of Probability Attitudes for Probability Students . . 94
S4.3 Alternative Pre- and Post-Survey of Probability Attitudes for Non-Probability Students 96
S4.4 Scoring Guide for SPA and alternative SPA 98
S4.5 Effort Plot 98

Chapter 5
Conclusion and Future Work **100**

Bibliography **102**

List of Figures

2.1 Quadratic potential surface with a single attraction point at the origin. . 7

2.2 (A) Simulated data for an individual with grey lines connecting those data points that are adjacent in time. The single attraction point is displayed as a red "+". (B) The quadratic potential surface with simulated positions for a single individual depicted in white. 11

2.3 95% equal-tailed credible intervals for each subsample. 13

2.4 95% equal-tailed credible intervals for the "best case" subset. 14

2.5 Stacked histograms using all converged simulations (the absolute values of Geweke Z-scores for α, β, and σ^2 were all less then 3). The dotted line delineates the mean of the LARI subsamples, and the solid line portrays the mean of the regular subsamples. 16

2.6 Stacked histograms from the "best case" simulations. The dotted line is positioned at the mean of the LARI subsamples and the solid line at the mean of the regular subsamples. 16

2.7 The first 50 time points with the true simulated values connected by a black line and 95% credible intervals depicted as orange bands. Grey dashed lines indicate the time points that were observed in the sample. The panels on the left (A and C) depict the results for the regular sample and the panels on the right (B and D) depict the results for the LARI sample. 17

2.8 The estimated potential surface using the full data. Observations are shown in white. 19

2.9 The estimated values of the parameters k, β and σ^2 from a regular subsample and 100 corresponding LARI subsamples (orange empirical density). 19

2.10 $14,401$ positions over 4 hours for each of the 73 ants, color coded by individual. 20

2.11 $14,401$ positions over 4 hours for a single ant, color coded by observation time. 20

2.12 The natural log of the motility surfaces estimated using the full data (A) and 3 subsamples (B)-(D). 24

2.13 Differences between the estimated log motility surfaces from each of the subsamples and the full data (calculated by grid cell). Negative values indicate underestimation of the motility surface. 25

2.14 Potential surfaces estimated with the four samples. The same potential surfaces are plotted in three dimensions on the left (using the `rayshader` R package) and in two dimensions on the right. 26

2.15 Statistics calculated for the motility and potential surfaces fit using 50 different 10 second LARI subsamples (orange) are compared to statistics from the every 5 second subsample (blue). 27

2.16 Panels (A)-(B): 20 locations simulated from (10) with parameters (2.24)–(2.25). The simulations are shown with the potential surface (A) and motility surface (B). The arrow points in the direction the simulated individual is heading. Panel (C): 1,000 locations simulated from (10) with parameters (2.24)–(2.25). Panel (C) displays histograms of the distances from $\mathbf{r}_{\tau-1}$ to $\mathbf{r}_{\tau}$ for each possible value of $m(\mathbf{r}_{\tau-2})$. Black vertical lines represent the means for each group. 30

2.17 Plots are as in Figure 2.16 using parameters (2.24) and (2.26). 30

2.18 Plots are as in Figure 2.16 using parameters (2.27) and (2.25). 31

2.19 MCMC draws from the posterior distribution conditioned on a regular subsample with every other timepoint. The MCMC draws are plotted in blue, the red lines bound equal-tailed 95% credible intervals, the black dashed line is the estimated posterior mean and the black solid line is the true parameter value. 31

2.20 True potential and motility surfaces with one example simulation shown in white. Black gradient vectors on the potential surface depict the negative gradient scaled by 5. 39

2.21 Density estimates for errors in the model fit on simulated data. The black vertical line is positioned at 0. 39

2.22 For one randomly selected simulation, the estimated potential (A) and motility (B) surfaces are compared to the true potential (B) and motility (D) surfaces. 40

3.1 Year-long movement paths for 3 golden eagles in the western United States, where path color differentiates between individuals. 42

3.2 Paths for all eagles in the dataset. Color indicates individual bird. 45

3.3 Plot replicated with permission from Spitz et al. (2017). 45

3.4 (A) Movement path for the resident in 2015. (B) Potential surface for the varying coefficient model. Attractors are numbered. (C) Potential surface for the single-state residence model. The label "1" is located at the single attractor. Bounds are the same as in B. The original path is plotted in blue with the simulation in red from (D) the single-state model and (E) the varying coefficient model. (F) Density plots of average distance from the simulations to the true path for varying coefficient and single-state models. Vertical lines indicate the means for each model. 48

3.5 (A) Movement path for the individual showing dispersal in 2018. (B) Potential surfaces for the two states in the latent-state model for dispersal. The attractors are identified with the numbers 1 and 2. (C) Potential surface for the varying coefficient model with same bounds as B. Attractors are numbered. The original path is plotted in blue with the simulation in red from (D) the latent-state model and (E) the varying coefficient model. (F) Density plots of average distance from the simulations to the true path for varying coefficient and latent-state models. Vertical lines indicate the means for each model. 52

3.6 (A) Movement path for the migrant in 2012. (B) Potential surfaces for the five states in the latent-state model for the migratory strategy. The attractors are identified with the numbers. (C) Potential surface for the varying coefficient model with same bounds as B. The original path is plotted in blue with the simulation in red from (D) the latent-state model and (E) the varying coefficient model. (F) Density plots of average distance from the simulations to the true path for varying coefficient and latent-state models. Vertical lines indicate the means for each model. . . 55

3.7 In (A), (C), and (E), the original paths are plotted in blue with the simulations from the varying coefficient model plotted in red. The attractors are labeled. In (B), (D), and (F), the varying coefficients corresponding to the attractors in (A), (C), and (E), respectively, are shown over time. For example, k1 in (B) is the varying coefficient corresponding to attractor 1 in (A). 57

3.8 Potential surfaces for the varying coefficient model for the resident with (A) daily observations and (B) observations every 2 days. Black dots indicate attractors. 61

3.9 Potential surfaces for the varying coefficient model for the migratory path with (A) daily observations and (B) observations every 2 days. Black dots indicate attractors. 62

3.10 Potential surfaces for the varying coefficient model for the path displaying dispersal with (A) daily observations and (B) observations every 2 days. Black dots indicate attractors. 63

3.11 Histograms of MCMC samples from the marginal posterior distributions. 64

3.12 Traceplots of MCMC samples from the marginal posterior distributions. . 64

3.13 Histograms of MCMC samples from the marginal posterior distributions for the latent-state model for dispersal. 65

3.14 Traceplots of MCMC samples from the marginal posterior distributions for the latent-state model for dispersal. 65

3.15 States for the dispersal model. In the top plot, the line width is proportional to the posterior probabilities of being in each state across time. In the bottom plot, the y axis is the y-coordinate of location (in meters). The observations are colored by most likely state. 66

3.16 Histograms of MCMC samples from the marginal posterior distributions for the latent-state model for the migratory strategy. 66

3.17 Traceplots of MCMC samples from the marginal posterior distributions for the latent-state model for the migratory strategy. 67

3.18 On the top, we see the latitude of the individual changing over time, and on the bottom, we see the covariate, daily change in day length, changing over time. All data is for NM Tredwell's location in 2012. 67

3.19 States for the migrant model. In the top plot, the line width is proportional to the posterior probabilities of being in each state across time. In the bottom plot, the y axis is the y-coordinate of location. The observations are colored by most likely state. 68

3.20 The smooth estimates of βk_{it} where k_{it} is the coefficient of attraction to attractor i at time t are plotted as solid black lines. Dotted lines are drawn two standard errors above and below the estimate of the smooth. The x-axis is the day of the year. 69

3.21 As in Figure 3.20, smooth estimates are plotted with dotted lines two standard errors above and below. 70

3.22 As in Figure 3.20, smooth estimates are plotted with dotted lines two standard errors above and below. 71

4.1 Distribution of majors for the 427 probability students who completed both pre- and post-surveys. Students with double majors were instructed to select the major which matched more closely with their interests. . . . 80

4.2 Sankey diagrams comparing the modified 6-factor structure we hypothesize to the 6-factor structure resulting from EFA. EFA results used (A) pre-survey data from probability students and (B) post-survey data from probability students. 83

4.3 Box plots of inter-individual response variability for the SPA. Sample size is 20 for all pre-survey scores except *Affect* (n=18) and *Interest* (n=19). Sample size is 14 for all post-survey scores. 83

4.4 Correlations (r) between factor scores on the PAI (y-axis) and post-SPA (x-axis) are shown in blue. Black dots depict scored for the 216 students who volunteered to complete the additional survey. 85

4.5 In (A), changes in mean pre- and post- SATS-36 scores for each component are shown for 101 introductory statistics sections (used with permission from Schau and Emmioğlu, 2012). In (B), changes in mean pre- and post-SPA scores for each component are shown for 13 probability sections. 85

4.6 Scatter plots of pre SPA scores versus gain from pre to post are displayed in A-E, with jittered points for ease of visibility, for 5 attitude components. Colored lines depict the section mean lines from the mixed effects models in Table 4.7B. Marginal densities are displayed in the margins. A shared legend is shown in F. 88

4.7 QQ-plot for residuals after fitting model (4.1) for the gain in *Effort* scores. 99

List of Tables

2.1 Potential surface error statistics for the three subsamples. 24

4.1 Bonferroni-corrected p values for testing the difference between probability students who did and did not complete the post-survey of those that completed the pre-survey. 79

4.2 Demographic make-up of the 427 probability students who completed both pre- and post-surveys. 80

4.3 Fit indices for the 6-factor SPA from probability students in the spring of 2021 81

4.4 Estimated latent factor correlations for the original and modified 6-factor SPA in the spring of 2021 82

4.5 Standard errors for factor scores based on students who completed the pre- or post-SPA twice. 84

4.6 Cronbach's alpha values for the SPA. 84

4.7 Linear mixed models fit with REML for each attitude component. Fit model (4.1) for probability students in (A) and fit model (4.2) for both probability and non-probability students in (B). *Affect* and *Cognitive Competence* scores were calculated as the mean of 4 items each, instead of 6 and 5 respectively, because some items could not easily be adapted for non-probability students. 87

4.8 Bonferroni-corrected p values for testing the difference between probability students and non-probability students. 89

4.9 Results of ratio t tests comparing mean gains for probability and non-probability students given specific values for pre score. 90

Acknowledgments

I would like to express my deepest gratitude to my co-advisors Dr. Ephraim Hanks and Dr. Matthew Beckman for their unwavering encouragement, guidance, and support. They have shown me by example the type of mentor I would like to be.

I would like to acknowledge my collaborators Dr. Robert Murphy, Dr. Todd Katzner and Dr. Tricia Miller for sharing their golden eagle expertise. Their contributions were essential for Chapter 3.

I would like to acknowledge the many faculty members and graduate students inside and outside of the statistics department who worked closely with me to allow me to collect data from their students for Chapter 4. I am grateful to Huy Dang for contributing his time to conduct think-alouds with students for Chapter 4 (only one of many reasons I am grateful to have him as a friend).

I am thankful to Dr. Frances Buderman and Dr. Stephen Berg for agreeing to serve on my committee and for their invaluable input along the way.

This dissertation research was supported by the NSF and NIH under Award No. NSF EEID 1414296, NSF DMS-2015273 and NIH GM116927-01. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the NSF or NIH.

Lastly, I would like to thank my family and friends who picked up the phone or otherwise gave me their time when I needed it most. I am eternally grateful.

Chapter 1 Introduction

This dissertation makes advances to two scientific fields: environmental statistics and statistics education. These topics are usually not studied together, and they will not be discussed within the same chapter outside of the introduction and conclusion. However, both environmental statistics and statistics education present a range of applied problems that benefit from collaborative efforts between statisticians and non-statisticians. This work contributes to the field of environmental statistics by building and comparing sampling and modeling approaches for animal movement. This work contributes to the field of statistics education by developing a tool for assessing students' attitudes toward probability. Chapters 2 and 3 contain two projects related to animal movement modeling, and Chapter 4 contains one statistics education project.

1.1 Overview of Research

In Chapter 2, we propose an irregular sampling design which could lead to greater efficiency and information gain in animal movement studies. Our novel sampling design, called lattice and random intermediate point (LARI), combines samples at regular and random time intervals. We compare the LARI sampling design to regular sampling designs in an example with common black carpenter ant location data, an example with guppy location data, and a simulation study of movement with a point of attraction. We modify a general stochastic differential equation model to allow for irregular time intervals and use this framework to compare sampling designs. When parameters are estimated reasonably well, regular sampling results in greater precision and accuracy in prediction of missing data. However, in each of the data and simulation examples explored in this paper, LARI sampling results in more accurate and precise parameter estimation, and thus better prediction of missing data as well. This result suggests that

researchers might gain greater insight into underlying animal movement processes by choosing LARI sampling over regular sampling.

In Chapter 3, we propose a flexible model for a partially migrating species, which we demonstrate using yearly paths for golden eagles (Aquila chrysaetos). Our model relies on a smoothly time-varying potential surface defined by a number of attractors. We compare our proposed approach using varying coefficients to a latent-state model, which we define differently for migrating, dispersing, and local individuals. While latent-state models are more common in the existing animal movement literature, varying coefficient models have various benefits including the ability to fit a wide range of movement strategies without the need for major model adjustments. We compare simulations from the models for three individuals to illustrate the ability of our model to better describe movement behavior for specific movement strategies. We also demonstrate the flexibility of our model by fitting several individuals whose movement behavior is less stereotypical.

Elementary probability courses are a common component of the statistics and data science curriculum. In Chapter 4, we adapt an existing survey called the Survey of Attitudes Toward Statistics (SATS-36) to assess students' attitudes toward probability. While the SATS-36 and other tools have become popular for assessing students' attitudes toward statistics, very few studies have assessed attitudes of students in probability courses toward the subject of probability. We distributed our Survey of Probability Attitudes (SPA) virtually among students at the beginning and end of 13 probability and 7 non-probability courses at Penn State University Park and Behrend campuses in the spring semester of 2021. The study includes 427 probability students and 159 students who were not enrolled in a probability course, all of whom completed both pre and post versions of the SPA. We gathered validity evidence which supports the SPA for its intended use case. We used mixed effects models to examine the change in five attitude components for students who were and were not enrolled in a probability course, and we compare this analysis approach to less nuanced approaches. The results suggest a large effect of regression to the mean and a lack of evidence of differences between probability attitudes in students who were and were not enrolled in a probability course.

Chapter 2
A Lattice and Random Intermediate Point Sampling Design for Animal Movement

This chapter was published in Environmetrics on January 3, 2020 (Eisenhauer and Hanks, 2020).

2.1 Introduction

Animal movement studies advance scientific knowledge of animal behavior in space and time. Insight from animal movement models helps researchers understand how animals interact with human and environmental factors. For example, researchers have conducted analyses of wildlife telemetry data to predict the affect of climate change on species range (Schloss et al., 2012) and to assess the impact of roadways on gene flow in terrestrial vertebrate populations (Shepard et al., 2008). Furthermore, understanding the relationships between animals and their surroundings can benefit conservation efforts (Festa-Bianchet and Apollonio, 2003; Chester, 2012; Berger, 2004) and provide insight into disease dynamics (Wijeyakulasuriya et al., 2019; Conner and Miller, 2004).

Researchers often record wildlife telemetry data at regular intervals (Weimerskirch et al., 2002; Forester et al., 2007; Kareiva and Shigesada, 1983; Parlin et al., 2018; Roeleke et al., 2018; McDuie et al., 2019) and occasionally at higher frequencies at times when finer movement behavior is expected (Richardson et al., 2018). There is evidence that increasing the frequency of regular samples greatly improves estimates of movement distance and territory size (Mills et al., 2006), but resource limitations often lead to difficulties in consistently obtaining samples at high frequencies without

reducing the overall length of the study. In this work, we show that the use of sampling designs other than regular sampling can lead to better inference on parameters in animal movement models without requiring additional samples or reducing study duration. While Millspaugh and Marzluff (2001) mention the application of a range of sampling designs with stochastic components for wildlife telemetry studies, these designs have not been thoroughly compared in context and are rarely implemented.

There is a widely accepted view in geostatistics that samples at regular intervals in space lead to better interpolation of data, while clustered samples lead to better estimation of spatial covariance parameters (Zimmerman, 2006). To compromise between parameter estimation and prediction at unobserved locations, Zimmerman (2006) suggests inclusion of samples with regular spacing as well as groups or pairs of points that are close together. One example of this is a lattice plus close pairs approach in which at least half of the locations form a regular lattice in 2D space, while the remaining points are randomly assigned within a disc centered at randomly selected lattice locations (Diggle and Lophaven, 2006). Diggle and Lophaven (2006) found that a lattice plus close pairs design performed better than a lattice alone or a lattice and infill approach in regard to spatial prediction.

Theoretical support for sampling at different scales is lacking in the geostatistical literature. On the contrary, systematic or regular sampling has been shown to be optimal within some subclasses of two-dimensional sampling designs in regard to minimizing variance of the sample mean (Bellhouse, 1977). However, few problems remain in animal movement modeling in which the only goal is precise estimation of a population mean. Instead, we look to the experimental literature which evidences the superiority of irregular sampling for detection of spatial patterns (Fortin et al., 1990; Oliver and Webster, 1986).

In this paper, we propose a sampling scheme for animal telemetry data inspired by the lattice plus close pairs geostatistical design. Our proposed approach, which we call a lattice and random intermediate point (LARI) design, requires data collection at regular time intervals coupled with one randomly selected time point in between each adjacent pair of regular samples. We conjecture that the regular time intervals will result in suitable temporal coverage while the random intermediate points will capture behavior at short time lags. We suspect that capturing behavior at different time scales will correspond with improved estimation of movement parameters.

This LARI sampling design was motivated by a problem that arose in collection of wood nesting carpenter ant, *Camponotus pennsylvanicus*, movement data at the Pennsylvania State University. Members of the Hughes laboratory captured video footage

of ants in a wooden nest over a 4 hour time frame and recorded coordinate locations of the ants at 1 second intervals (Modlmeier et al., 2019). The data collection procedure was manually expensive, requiring the recruitment, training, and labor of seventeen undergraduate students (Modlmeier et al., 2019). As a new experiment was planned involving a large number of nests over a longer time frame, it became apparent that the data collection strategy previously employed would not be feasible at the necessary scale. Thus we set out to develop a sampling design that would result in similar model inference while reducing the manual cost. Of course, this motivation is not limited to the ant example. Restrictions on data collection frequency and magnitude are commonplace in animal movement studies, especially those that employ tracking devices (Tomkiewicz et al., 2010).

We describe the LARI sampling scheme in detail in Section 2.2. In Section 3.3 we outline a stochastic differential equation (SDE) model for movement similar to that of Russell et al. (2018). In Section 2.4, we compare parameter estimation and prediction accuracy between sampling designs via a simulated example. In Section 2.5, we compare parameter estimates between sampling designs using subsamples of guppy movement data. In Section 2.6, we present a novel modeling framework which we apply to the high resolution carpenter ant movement data and implement to compare sampling designs.

2.2 A Lattice and Random Intermediate Points Sampling Scheme

In a given animal movement study, assume data collection is set to begin at time 0 and end at time T. Assume resources are limited and only n samples will be collected in this time frame for a single individual. Sampling the animal's position at regular time intervals of length $h = \frac{T}{n-1}$ results in the data matrix

$$\mathbf{D}_{\text{Regular}} \equiv \begin{bmatrix} 0 & h & 2h & \dots & T-h & T \\ \mathbf{r}_0 & \mathbf{r}_h & \mathbf{r}_{2h} & \dots & \mathbf{r}_{T-h} & \mathbf{r}_T \end{bmatrix}' \tag{2.1}$$

where $\mathbf{r}_t \equiv \begin{bmatrix} x_t & y_t \end{bmatrix}'$ is the x- and y-coordinate vector of the animal's position at time $t \in \{0, h, 2h, \dots, T-h, T\}$. While regular sampling minimizes the maximum time between observations, movement behavior occurring at finer time scales than those sampled is not captured in the observed data.

We propose a lattice and random intermediate points (LARI) sampling scheme, which

produces the data matrix

$$\mathbf{D}_{\text{LARI}} \equiv \begin{bmatrix} 0 & t_0^* & 2h & t_1^* & 4h & \dots & T-2h & t_{\frac{n}{2}}^* & T \\ \mathbf{r}_0 & \mathbf{r}_{t_0^*} & \mathbf{r}_{2h} & \mathbf{r}_{t_1^*} & \mathbf{r}_{4h} & \dots & \mathbf{r}_{T-2h} & \mathbf{r}_{t_{\frac{n}{2}}^*} & \mathbf{r}_T \end{bmatrix}' \tag{2.2}$$

where

$$t_i^* \sim \text{Uniform}\left(2hi, 2h(i+1)\right), \quad i \in \left\{0, 1, 2, \dots, \frac{n}{2}\right\}.$$

In practice, it may be more realistic to choose t_i^* from a Discrete Uniform distribution depending on the sampling resolution.

Both data matrices $\mathbf{D}_{\text{Regular}}$ and $\mathbf{D}_{\text{LARI}}$ contain n observations for a single individual. To collect data for multiple individuals over multiple time frames, repeat this procedure as necessary.

2.3 Stochastic Differential Equation Model for Animal Movement

We follow Russell et al. (2018) and Hanks et al. (2017) and consider a flexible stochastic differential equation (SDE) model for an animal's position $\mathbf{r}_t$ at time t

$$d\mathbf{r}_t = \mathbf{v}_t dt \tag{2.3}$$

$$d\mathbf{v}_t = -\beta(\mathbf{v}_t - \boldsymbol{\mu}(\mathbf{r}_t))dt + c(\mathbf{r}_t)\mathbf{I}d\mathbf{w}_t \tag{2.4}$$

where $\mathbf{v}_t$ is the velocity of the animal at time t, β is the coefficient of friction (Nelson, 1967) which controls autocorrelation in movement, $\boldsymbol{\mu}(\mathbf{r}_t)$ is the mean drift in the direction of movement, $c(\mathbf{r}_t)$ is a scalar that controls the variance in the stochastic term, $\mathbf{I}$ is a 2×2 identity matrix, and $\mathbf{w}_t$ is independent Brownian motion in $\mathbb{R}^2$. This SDE framework is attractive because of the wide range of movement behavior which can be modeled. For example, the right hand side of (2.4) can be viewed as the sum of forces acting on the animal at time t and at position $\mathbf{r}_t$. For instance, there could be a force toward the center of the animal's home range, toward the nearest food source, toward breeding grounds, toward higher or lower elevation, away from the nearest predator, or away from cooler temperatures. Depending on the time frame and study species, these forces could vary over time or space.

Brillinger et al. (2012) used a similar SDE framework and adopted potential functions from particle and planetary movement models to model elk movement by setting $\boldsymbol{\mu}(\mathbf{r}_t) = -\nabla p(\mathbf{r}_t)$, the negative gradient of a potential surface $p(\mathbf{r}_t)$. The potential surface is a continuous surface or grid with the highest values on the surface at repulsive locations, lowest values at attractive locations, and relatively central values in areas where the force is neutral. Under this model, the average animal in a population moves around the space avoiding those points of repulsion or areas with high potential and moving toward points of attraction or areas with low potential. A simple example of a potential surface is the quadratic function $k(x^2 + y^2)$ which will be used in a simulation example in Section 2.4. This quadratic potential surface has a single point of attraction at the origin, as shown in Figure 2.1 with the parameter $k = 1$. The white arrows displayed in Figure 2.1 point down the gradient of the potential surface, in the direction of mean drift. One might utilize this potential surface in a model for movement of a central place forager, with movement centered around $\begin{bmatrix} 0 & 0 \end{bmatrix}'$ and k controlling the strength of attraction to this central location. Potential surfaces can be much more complex than this example, as we will see in Section 2.6. For further detail on the use of potential surfaces to model animal movement, see Preisler et al. (2013).

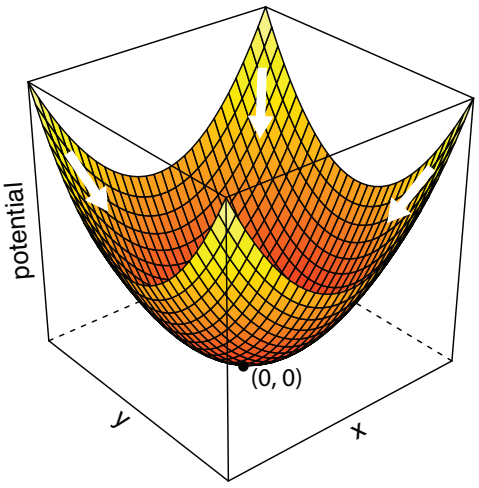

Figure 2.1. Quadratic potential surface with a single attraction point at the origin.

Russell et al. (2017, 2018) expanded the SDE framework of Brillinger et al. (2012) to include motility surfaces, which describe overall speed independent of direction as a function of location. The motility surface is a surface or grid of values assigned on the space inhabited by the animal. High motility values are indicative of fast movement or high speed in the corresponding location. Low motility is indicative of slow movement.

The SDE model we define in this section is similar to that of Russell et al. (2018) with

zero measurement error and assuming the motility surface is smooth. As in Russell et al. (2018), we define the mean drift and magnitude of stochasticity with spatially-varying motility and potential surfaces. The potential surface $p(\mathbf{r}_t)$ captures spatially-varying directional bias (drift) through $-\nabla p(\mathbf{r}_t)$, while the motility surface $m(\mathbf{r}_t)$ models spatial variation in speed without directional bias by compressing and dilating time. The mean drift $\boldsymbol{\mu}(\mathbf{r}_t)$ and magnitude of stochasticity $c(\mathbf{r}_t)$ are defined

$$\boldsymbol{\mu}(\mathbf{r}_t) \equiv m(\mathbf{r}_t)[-\nabla p(\mathbf{r}_t)] \tag{2.5}$$

$$c(\mathbf{r}_t) \equiv \sigma m(\mathbf{r}_t) \tag{2.6}$$

where σ controls the magnitude of the random forces acting on the animal. We chose to ignore measurement error because the measurement error in our ant data is negligible. As sophisticated technology allows for greater accuracy in animal tracking, we expect the need for measurement error specification for animal location to diminish. When movement error is not negligible, state-space models can be used with the SDE model (2.3)–(2.6) being a model for the true, but latent, animal position over time.

2.3.1 Numerical Approximations

A closed-form solution to (2.3)–(2.6) only exists for very simple choices of $m(\cdot)$ and $p(\cdot)$. There is no closed-form solution whenever spatial constraints are present (e.g., Hanks et al. (2017); Russell et al. (2018)). In this section, we present a general numerical approximation to the SDE which is applicable in a broad range of settings including those where there is no closed-form solution.

Hanks et al. (2017) and Russell et al. (2018, 2017) describe numerical approximations using samples at regular time intervals and do not consider irregular time lags between samples. Our discrete-time approximation approach is similar to that of Russell et al. (2018), but we extend their framework to the case where the intervals between observation times can vary. Developing numerical methods for irregular time intervals will make inference more straightforward when data are missing or irregularly sampled. To simplify notation for irregular samples, we now change the subscript in equations from continuous time t to ordered observation number τ. Henceforth, $\mathbf{r}_\tau$ is the vector of elements in column τ, row 2 of a data matrix of the form (2.1) or (2.2).

Euler-Maruyama approximations are derived from Taylor series expansions (Kloeden and Platen, 2013) and are commonly used to numerically approximate SDE models because they are easy to calculate. The Euler-Maruyama method approximates (2.3)–(2.4)

by

$$\mathbf{r}_{\tau+1} = \mathbf{r}_\tau + \mathbf{v}_\tau h_\tau \tag{2.7}$$

$$\mathbf{v}_{\tau+1} = \mathbf{v}_\tau - \beta(\mathbf{v}_\tau - \boldsymbol{\mu}(\mathbf{r}_\tau))h_\tau + c(\mathbf{r}_\tau)\mathbf{I}d\mathbf{w}_\tau \tag{2.8}$$

where h_τ is the change in time from observation τ to observation $\tau+1$. Here our approach differs from the SDE model of Russell et al. (2018) where h_τ was constant with respect to τ. Substituting (2.7) into (2.8) following Hanks et al. (2017) results in

$$\frac{\mathbf{r}_{\tau+2} - \mathbf{r}_{\tau+1}}{h_{\tau+1}} - \frac{\mathbf{r}_{\tau+1} - \mathbf{r}_\tau}{h_\tau} = \beta h_\tau \left(\boldsymbol{\mu}(\mathbf{r}_\tau) - \frac{\mathbf{r}_{\tau+1} - \mathbf{r}_\tau}{h_\tau}\right) + c(\mathbf{r}_\tau)h_\tau^{1/2}\boldsymbol{\epsilon}_\tau \tag{2.9}$$

where $\boldsymbol{\epsilon}_\tau \overset{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{I})$ and $\mathbf{0}$ is the zero vector in $\mathbb{R}^2$. This can be re-expressed as

$$\mathbf{r}_{\tau+2} = \mathbf{r}_{\tau+1} + \frac{h_{\tau+1}}{h_\tau}(\mathbf{r}_{\tau+1} - \mathbf{r}_\tau) + \beta h_\tau h_{\tau+1}\left(\boldsymbol{\mu}(\mathbf{r}_\tau) - \frac{\mathbf{r}_{\tau+1} - \mathbf{r}_\tau}{h_\tau}\right) + c(\mathbf{r}_\tau)h_\tau^{1/2}h_{\tau+1}\boldsymbol{\epsilon}_\tau, \tag{2.10}$$

an equation in which the ant's position is a function of the two previous observed positions.

In Supplemental Material S2.1, we provide examples of potential and motility surfaces, which we simulate from using (2.10). These examples illustrate how changing the motility and potential surfaces effects the movement behavior described by the model.

2.4 Simulation Example

2.4.1 Simulation from an SDE Model with Quadratic Potential Function

We conducted a simulation example to compare the sampling schemes in (2.1) and (2.2). We simulated data at a fine temporal scale from an SDE model with a quadratic potential function and constant motility surface. The quadratic function biases movement toward a single attraction point at $\begin{bmatrix}0 & 0\end{bmatrix}'$. This approximates real movement behavior exhibited by central place foragers such as white-tailed deer (Tierson et al., 1985). In this example,

$$m(\mathbf{r}_\tau) \equiv 1$$

$$p(\mathbf{r}_\tau) \equiv k\mathbf{r}_\tau'\mathbf{r}_\tau$$

where $k \in \mathbb{R}$ controls the strength of attraction to the central location $\begin{bmatrix} 0 & 0 \end{bmatrix}'$. Consequently,

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{r}_\tau) &= -\nabla p(\mathbf{r}_\tau) = -2k\mathbf{r}_\tau \\ c(\mathbf{r}_\tau) &= \sigma. \end{aligned}$$

The set of SDEs (2.3) and (2.4) become

$$d\mathbf{r}_t = \mathbf{v}_t dt \tag{2.11}$$

$$d\mathbf{v}_t = -\beta[\mathbf{v}_t - (-2k\mathbf{r}_\tau)]dt + \sigma \mathbf{I} d\mathbf{w}_t \tag{2.12}$$

and the numerical approximation (2.9) becomes

$$\frac{\mathbf{r}_{\tau+2} - \mathbf{r}_{\tau+1}}{h_{\tau+1}} - \frac{\mathbf{r}_{\tau+1} - \mathbf{r}_\tau}{h_\tau} = \beta h_\tau \left(-2k\mathbf{r}_\tau - \frac{\mathbf{r}_{\tau+1} - \mathbf{r}_\tau}{h_\tau} \right) + h_\tau^{1/2} \sigma \boldsymbol{\epsilon}_\tau. \tag{2.13}$$

Since the simulated data is generated at regular time steps, we set $h_\tau = h$ for all observations τ and solve for $\mathbf{r}_{\tau+2}$ to get

$$\mathbf{r}_{\tau+2} = \mathbf{r}_{\tau+1}(2 - \beta h) + \mathbf{r}_\tau(\beta h - 1 - 2\beta k h^2) + h^{3/2} \sigma \boldsymbol{\epsilon}_\tau \tag{2.14}$$

which is an autoregressive model of order 2.

We simulated movement data for one individual over $n = 500$ time points with time step $h = 1$ and model parameters $\beta = 0.4$, $\alpha \equiv k\beta = 0.08$, and $\sigma = 0.5$. Since simulation of observation τ requires observations $\tau - 1$ and $\tau - 2$ as input, we fixed the positions at the first two time points near the point of attraction $\begin{bmatrix} 0 & 0 \end{bmatrix}'$. Specifically, $\mathbf{r}_1 = \mathbf{r}_2 = \begin{bmatrix} 1 & 1 \end{bmatrix}'$. The next 498 time points were simulated recursively from (2.14). Figure 2.2 depicts one path simulated with this procedure. We will simulate 150 paths and consider each path separately. We will compare regular and LARI sampling schemes by subsampling each simulated path using (2.1)–(2.2) with $h = 5$ and comparing subsamples.

We refer to the full simulation containing all x- and y-coordinates by $\{\mathbf{r}\}$. The general notation for the positions included in the LARI or regular subsample is $\{\mathbf{r}\}_{\text{obs}}$ and the positions removed by the subsampling procedure are $\{\mathbf{r}\}_{\text{unobs}}$. Thus, the subsampling procedure is represented by

$$\{\mathbf{r}\}_{\text{obs}} = \{\mathbf{r}\} \setminus \{\mathbf{r}\}_{\text{unobs}}.$$

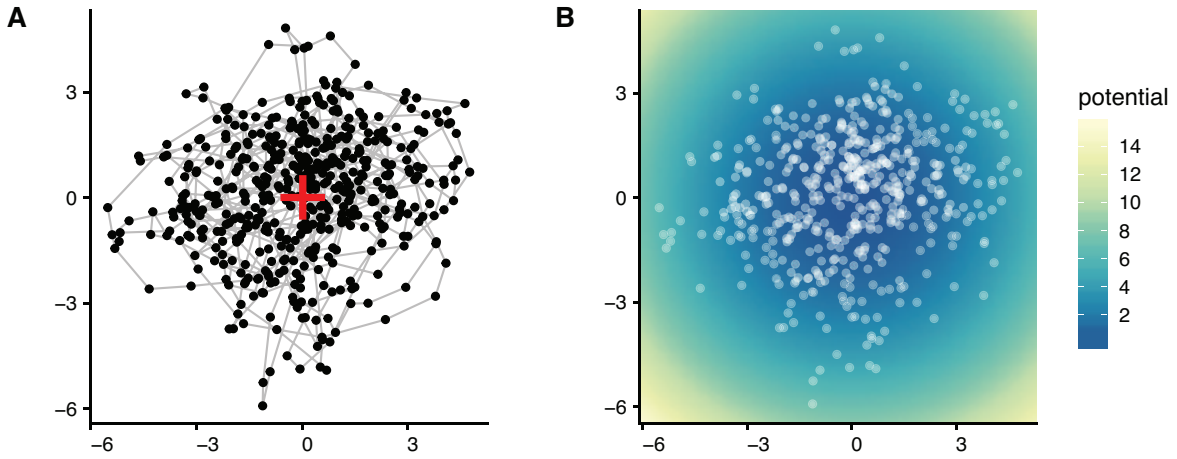

Figure 2.2. (A) Simulated data for an individual with grey lines connecting those data points that are adjacent in time. The single attraction point is displayed as a red "+". (B) The quadratic potential surface with simulated positions for a single individual depicted in white.

2.4.2 Parameter Identifiability

In Kloeden and Platen (2013), the authors derive vector and matrix ordinary differential equations for the vector mean and second moment of a general vector linear SDE. In this subsection, we will interpret this derivation in the context of (2.11) and (2.12). For simplicity, we describe this result in the x direction only, where $r_x(t)$ is the x component of the coordinate vector $\mathbf{r}_t$, $v_x(t)$ is the x component of the coordinate vector $\mathbf{v}_t$, and $w_x(t)$ is independent Brownian motion in $\mathbb{R}^1$.

We are interested in determining whether the parameters we intend to estimate (β, σ, and α) are identifiable as we approach the stationary distribution (i.e., as $t \to \infty$). By combining (2.11) and (2.12), we obtain the vector SDE

$$d\begin{bmatrix} r_x(t) \\ v_x(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -2\alpha & -\beta \end{bmatrix} \begin{bmatrix} r_x \\ v_x \end{bmatrix}_t dt + \begin{bmatrix} 0 \\ \sigma \end{bmatrix} dw_x(t), \tag{2.15}$$

which allows us to derive the vector mean

$$\mathbf{n}(t) = \mathrm{E}\left(\begin{bmatrix} r_x(t) \\ v_x(t) \end{bmatrix}\right) = \begin{bmatrix} 0 & 1 \\ -2\alpha & -\beta \end{bmatrix}^{-1} \frac{d\mathbf{n}(t)}{dt}. \tag{2.16}$$

As we approach the stationary distribution and $\frac{d\mathbf{n}(t)}{dt} = 0$, the mean vector

$$\mathbf{n}(t) = 0.$$

Therefore, the second moment of the stationary distribution

$$S(t) = \mathrm{E}\left(\begin{bmatrix} r_x(t) \\ v_x(t) \end{bmatrix}\begin{bmatrix} r_x(t) \\ v_x(t) \end{bmatrix}'\right) = \mathrm{Var}\left(\begin{bmatrix} r_x(t) \\ v_x(t) \end{bmatrix}\right) \tag{2.17}$$

is found by solving the system of equations

$$\frac{dS(t)}{dt} = \begin{bmatrix} 0 & 1 \\ -2\alpha & -\beta \end{bmatrix} S(t) + S(t) \begin{bmatrix} 0 & 1 \\ -2\alpha & -\beta \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma \end{bmatrix}\begin{bmatrix} 0 \\ \sigma \end{bmatrix}'. \tag{2.18}$$

The stationarity of the distribution implies $\frac{dS(t)}{dt} = 0$, which along with (2.18) yeilds

$$S(t) = \begin{bmatrix} \frac{\sigma^2}{2\beta} & 0 \\ 0 & \frac{\sigma^2}{4\beta\alpha} \end{bmatrix}. \tag{2.19}$$

Thus, we have 2 equations and 3 unknowns, rendering σ, β, and α unidentifiable.

This result highlights the value of having samples at short time lags. When telemetry data are sampled regularly at large time lags, the transient distribution will be well-approximated by the stationary distribution, and parameters in the model may become unidentifiable, or only weakly identifiable. Thus, we expect the regular subsample will lead to unidentifiability in parameter estimation. In Section 2.4.3, we outline the model-fitting procedure applied to the LARI and regular subsamples.

2.4.3 Estimation of Model Parameters and Missing Values

Initially, we attempted posterior approximation of the model parameters ignoring the missing data. However, this approach led to poor parameter inference (see Supporting Material S2.2), which could be due to the large reduction in movement variability that occurs when data are subsampled. These results led to our decision to estimate the positions at unobserved time points, thus reintroducing an appropriate amount of variability into the movement paths.

To estimate β, α, σ, and $\{\mathbf{r}\}_{\text{unobs}}$, we took a Bayesian approach and constructed an MCMC algorithm to sample from the joint posterior $\pi(\alpha, \beta, \sigma, \{\mathbf{r}\}_{\text{unobs}}|\{\mathbf{r}\}_{\text{obs}})$. For details on the posterior distribution and MCMC sampler, see Supporting Material S2.3.

2.4.4 Simulation Example Results

We simulated 150 paths, subsampled the paths using both a LARI and regular design, and individually fit each subsample using the estimation approach described in Section 2.4.3. We assessed convergence of the MCMC algorithm in each case using the Geweke diagnostic (Geweke, 1991). The Geweke convergence diagnostic quantifies the dissimilarity of the means of the first 10% and last 50% of the Markov chain iterations. In the Geweke diagnostic, the test statistic for variable η is a z-score

$$z = \frac{\bar{\eta}^{(\text{first } 10\%)} - \bar{\eta}^{(\text{last } 50\%)}}{\widehat{SE}}$$

where $\bar{\eta}^{(\text{first } 10\%)}$ is the sample mean of the first 10% of the Markov chain, $\bar{\eta}^{(\text{last } 50\%)}$ is the sample mean of the last 50% of the chain, and $\widehat{SE}$ is the asymptotic standard error of the difference, computed using spectral density estimates for the two sections of the chain.

For each subsample and each simulated path, the Geweke diagnostic was computed for the three parameters α, β, and σ. We labelled a subsample "converged" if the absolute values of the Geweke diagnostics for all three chains were less than 3. By this definition, of the 300 subsamples, 76.3% converged. Specifically, 84% of the LARI subsamples and 68.7% of the regular subsamples converged. We removed all simulations where at least one subsample (regular or LARI) did not converge, leaving 87 of the 150 simulations for analysis.

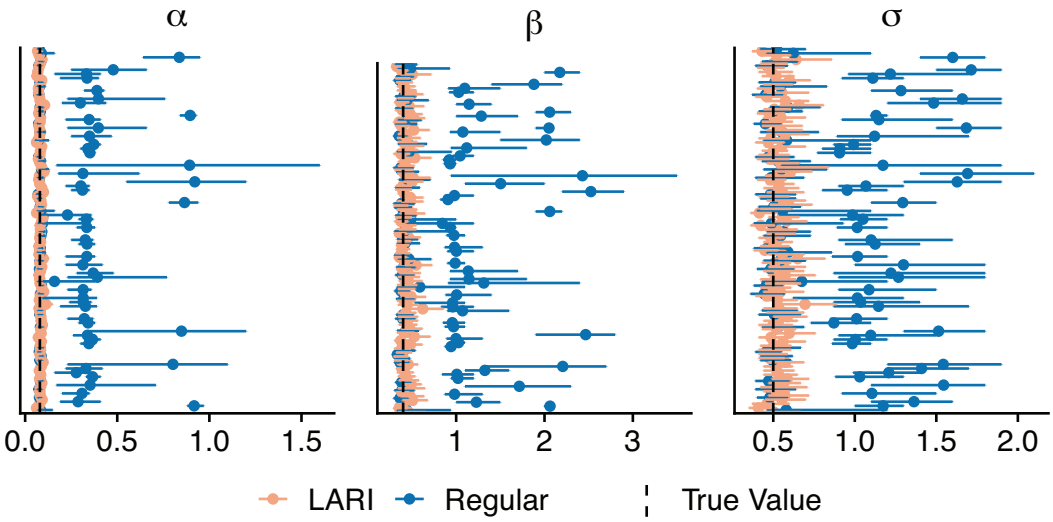

Figure 2.3. 95% equal-tailed credible intervals for each subsample.

95% equal-tailed credible intervals for the model parameters are shown for the 174 remaining subsamples in Figure 2.3. As depicted in Figure 2.3, many of the regular subsamples result in poor estimation of the model parameters. Only 46.6% of the 87 regular subsamples resulted in credible intervals containing all three true parameter values, compared to 78.4% of the 87 LARI subsamples. This result is consistent with the theoretical justification in Section 2.4.2, which suggests unidentifiability when we use a regular subsample at large enough time steps.

Although we have already obtained evidence in favor of LARI sampling for parameter estimation, we are also interested in the "best case" scenario where both subsamples capture the true parameter values in their 95% credible intervals. There are 34 remaining simulations in this "best case" subset. The model parameter 95% credible intervals for the "best case" subsamples are shown in Figure 2.4.

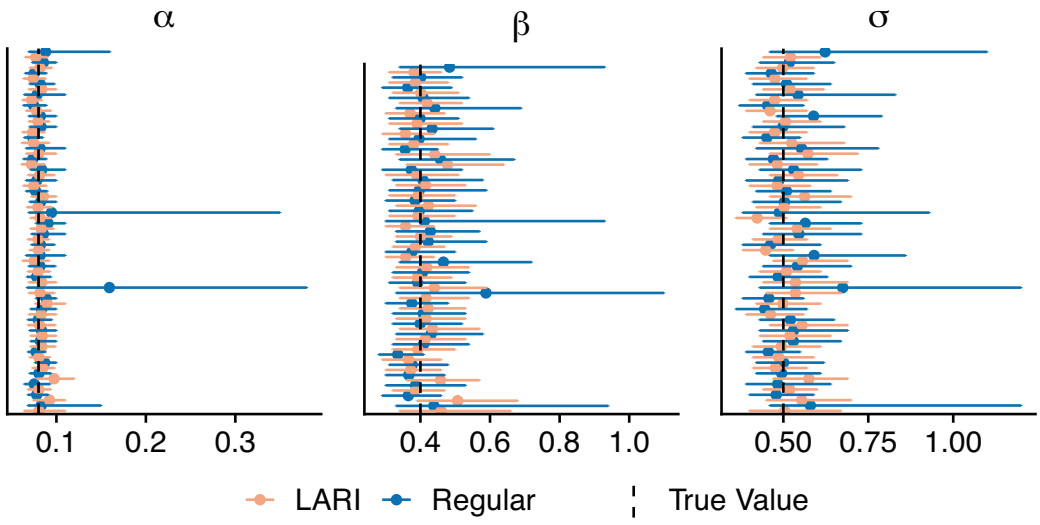

Figure 2.4. 95% equal-tailed credible intervals for the "best case" subset.

We are now limited to 2 subsets; the subsamples which led to convergence by our definition and the "best case" subsamples. For each of these subsets, we will compare parameter estimation and prediction of missing values between the LARI and regular sampling designs with 8 metrics. We define these metrics in the following passages and display our results in Figures 2.5 and 2.6.

To assess parameter estimation accuracy, we found the posterior mean squared error (PMSE) for each of the model parameters, α, β, and σ^2. The PMSE of variable η is the mean squared difference between the MCMC draws $\eta^{(i)}$ and the true parameter value

η_{true}

$$\begin{aligned}\text{PMSE}(\eta) &= \int (\eta - \eta_{\text{true}})^2 \pi(\eta|\{\mathbf{r}\}_{\text{obs}}) d\eta \\ &\approx \frac{1}{100,000} \sum_{i=1}^{100,000} (\eta^{(i)} - \eta_{\text{true}})^2.\end{aligned}$$

We constructed 95% equal-tailed credible intervals for α, β, and σ^2 and recorded credible interval width to assess parameter estimation precision.

To assess prediction accuracy for missing time points, mean squared predictive errors (MSPE) were found for each subsample, where

$$\text{MSPE}(\{\mathbf{r}\}_{\text{unobs}}) = \sum_{k \in \{\tau:\ \mathbf{r}_\tau \in \{\mathbf{r}\}_{\text{unobs}}\}} \left[\bar{\mathbf{r}}_k^{(i)} - \mathbf{r}_k\right]' \left[\bar{\mathbf{r}}_k^{(i)} - \mathbf{r}_k\right],$$

$\mathbf{r}_k$ is the true location of observation k, $\mathbf{r}_k^{(i)}$ is the i^{th} sample from the posterior distribution of $\mathbf{r}_k$, and $\bar{\mathbf{r}}_k^{(i)} = \frac{\sum_{i=1}^{100,000} \mathbf{r}_k^{(i)}}{100,000}$. We found the mean width of 95% equal-tailed credible intervals for missing values $\{\mathbf{r}\}_{\text{unobs}}$ to assess precision of the predictions.

Figure 2.5 depicts the statistics for all converged simulations (that is, all with Geweke Z-scores for α, β, and σ^2 less then 3), and Figure 2.6 looks at a further subset, the "best case" simulations (those which included the true values of α, β, and σ^2 in their equal-tailed 95% credible intervals). The results displayed in Figure 2.5 indicate the LARI subsamples outperform the regular subsamples with respect to 95% credible interval width and PMSE for α, β, and σ^2. The LARI subsamples also outperform the regular subsamples on average when we compare them based on the metrics for predicting missing data. However, in the "best case" subset where LARI and regular subsamples both estimate the parameters well, the regular subsamples more accurately predict missing data. The missing data MSPEs from all but one of the LARI subsamples are lower than all MSPEs from the regular subsamples.

Thus, for the simulations that converged, the LARI sampling design led to better estimation of the model parameters α, β, and σ as well as better prediction of missing locations; but when both LARI and regular subsamples estimated the parameters well, the regular sampling design led to greater accuracy and precision in prediction of missing data points. This result is consistent with the hypothesis that the variability in time intervals between observations in the LARI design leads to a better understanding of movement behavior through greater accuracy in model parameter estimation.

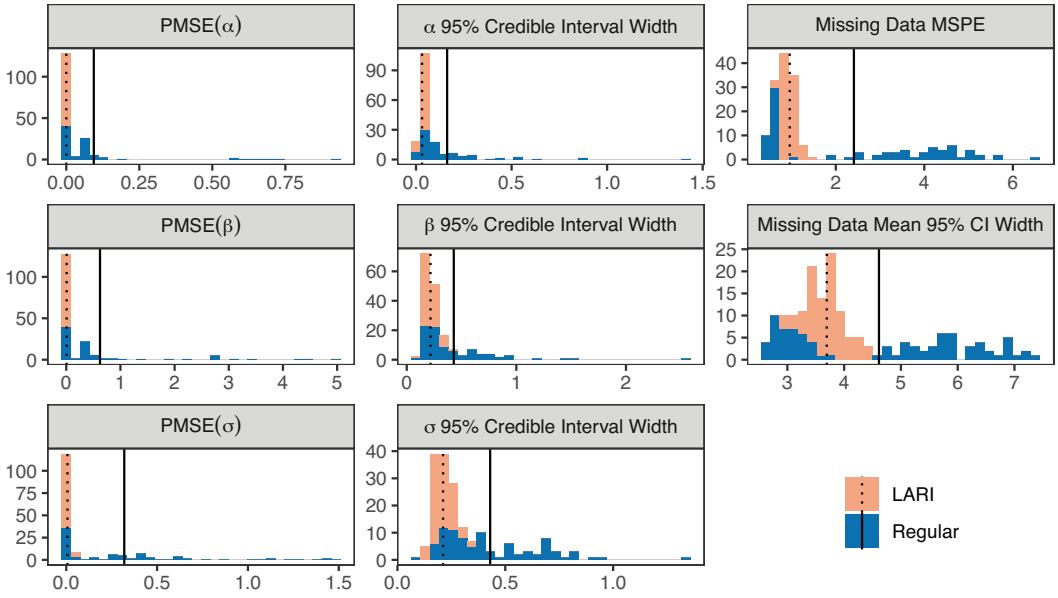

Figure 2.5. Stacked histograms using all converged simulations (the absolute values of Geweke Z-scores for α, β, and σ^2 were all less then 3). The dotted line delineates the mean of the LARI subsamples, and the solid line portrays the mean of the regular subsamples.

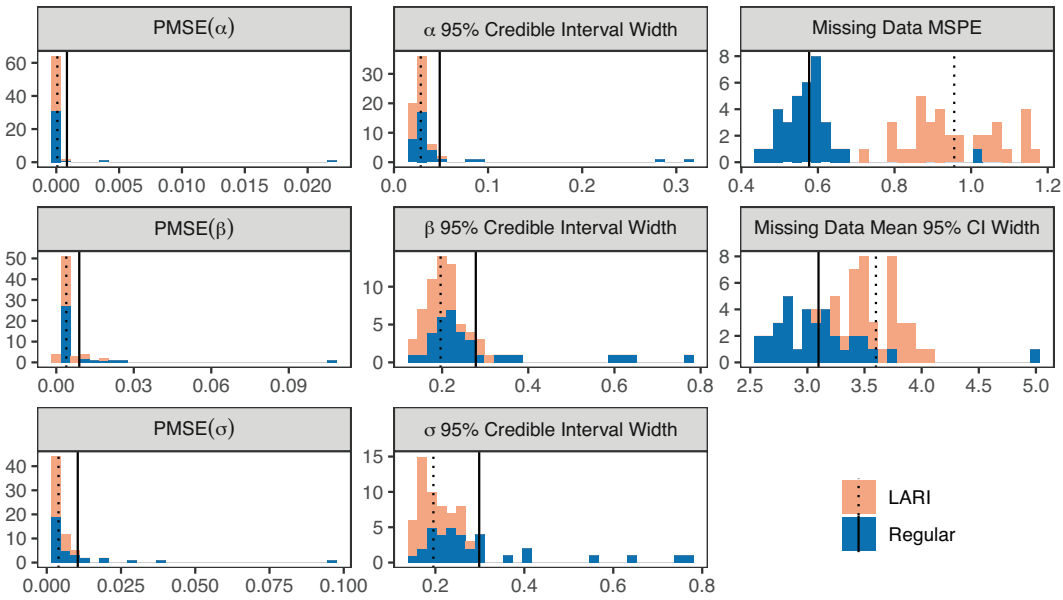

Figure 2.6. Stacked histograms from the "best case" simulations. The dotted line is positioned at the mean of the LARI subsamples and the solid line at the mean of the regular subsamples.

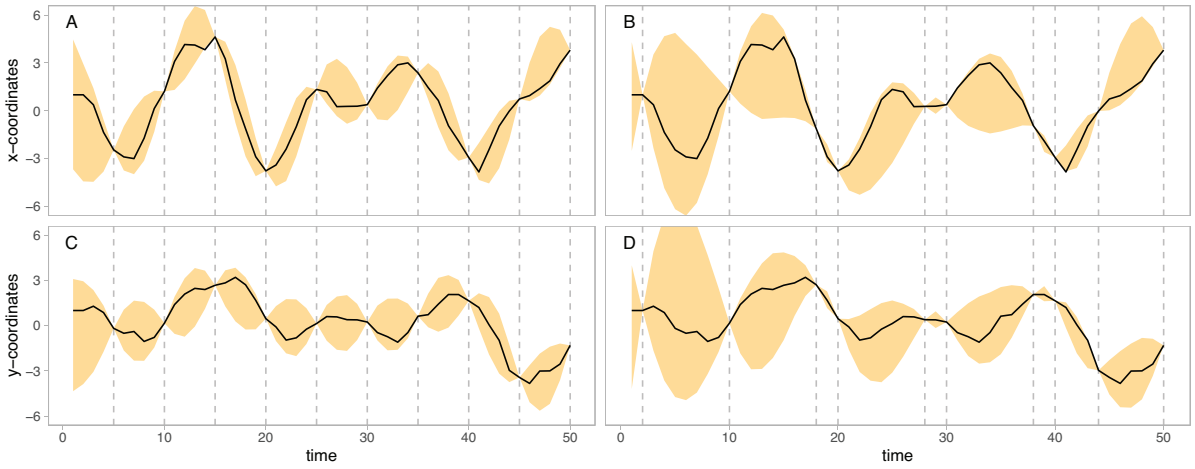

Figure 2.7. The first 50 time points with the true simulated values connected by a black line and 95% credible intervals depicted as orange bands. Grey dashed lines indicate the time points that were observed in the sample. The panels on the left (A and C) depict the results for the regular sample and the panels on the right (B and D) depict the results for the LARI sample.

To better understand the missing data prediction results, we explore one of the "best case" simulations. In Figure 2.7, we plot the true x- and y-coordinates for the first 50 time points with corresponding 95% credible intervals. As shown in Figure 2.7, sampling at regular intervals often results in smaller credible intervals for unobserved values. We suspect this is because the LARI design includes larger time gaps than regular sampling, which disproportionately affects the mean of credible interval widths.

2.5 Guppy Data Example

In our first data example, we will use movement data from a captive population of guppies (*Poecilia reticulata*). The group of guppies were released in the bottom right corner of a flat-bottomed square tank and swam toward a sheltered area in the opposite corner of the tank. The data consists of 360 observations recorded at 0.1 second intervals for each of 10 guppies. For more information regarding data collection, see Bode et al. (2012). After fitting a SDE to the full data, we will compare the sampling schemes in (2.1) and (2.2) by subsampling the data and comparing the resulting model fits.

2.5.1 SDE Model

As in the simulation example, we can represent the movement of individual guppies with a set of SDEs. In this example, we define motility and potential surfaces

$$\begin{aligned} m(\mathbf{r}_\tau) &\equiv 1 \\ p(\mathbf{r}_\tau) &\equiv k|\mathbf{r}_\tau - \mathbf{a}| \end{aligned}$$

where $k \in \mathbb{R}$ controls the strength of the drift toward a known point of attraction $\mathbf{a} = \begin{bmatrix} 281 & 434 \end{bmatrix}'$ in the sheltered corner of the tank. The potential surface is defined in this way to elicit a constant force toward the point of attraction. This specification of potential and motility surfaces results in mean drift and magnitude of stochasticity

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{r}_\tau) &= -\nabla p(\mathbf{r}_\tau) = -k \times \text{sign}(\mathbf{r}_\tau - \mathbf{a}) \\ c(\mathbf{r}_\tau) &= \sigma. \end{aligned}$$

The set of SDEs (2.3) and (2.4) become

$$d\mathbf{r}_t = \mathbf{v}_t dt \tag{2.20}$$

$$d\mathbf{v}_t = -\beta[\mathbf{v}_t - (-k \times \text{sign}(\mathbf{r}_\tau - \mathbf{a}))]dt + \sigma \mathbf{I} d\mathbf{w}_t \tag{2.21}$$

and the numerical approximation (2.9) becomes

$$\frac{\mathbf{r}_{\tau+2} - \mathbf{r}_{\tau+1}}{h_{\tau+1}} - \frac{\mathbf{r}_{\tau+1} - \mathbf{r}_\tau}{h_\tau} = \beta h_\tau \left(-k \times \text{sign}(\mathbf{r}_\tau - \mathbf{a}) - \frac{\mathbf{r}_{\tau+1} - \mathbf{r}_\tau}{h_\tau} \right) + h_\tau^{1/2} \sigma \boldsymbol{\epsilon}_\tau. \tag{2.22}$$

Linear regression was implemented to estimate k, β, and σ^2.

2.5.2 Results

We implemented a data subsampling procedure which resulted in 3 regular subsamples and 300 LARI subsamples. The regular subsamples were recorded at 0.3 second intervals beginning at each of the first 3 time points. For each of the regular subsamples, 100 corresponding LARI subsamples were collected which consisted of every other sample from the regular data (i.e., regular samples at 0.6 second intervals) coupled with a random time point selected from the observations in each subinterval. This resulted in 3 groups of subsamples, each consisting of one regular subsample and 100 corresponding LARI subsamples.

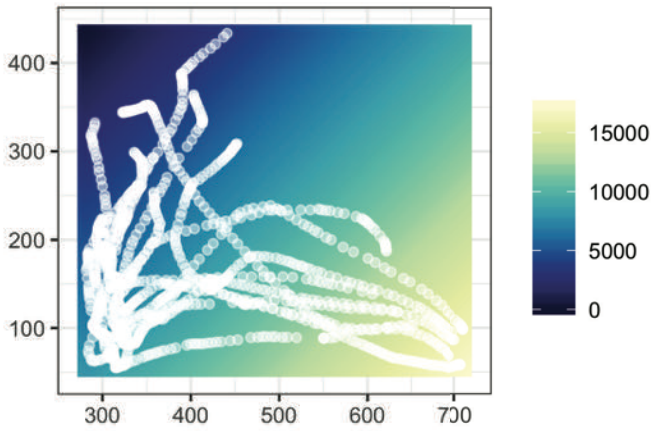

Figure 2.8. The estimated potential surface using the full data. Observations are shown in white.

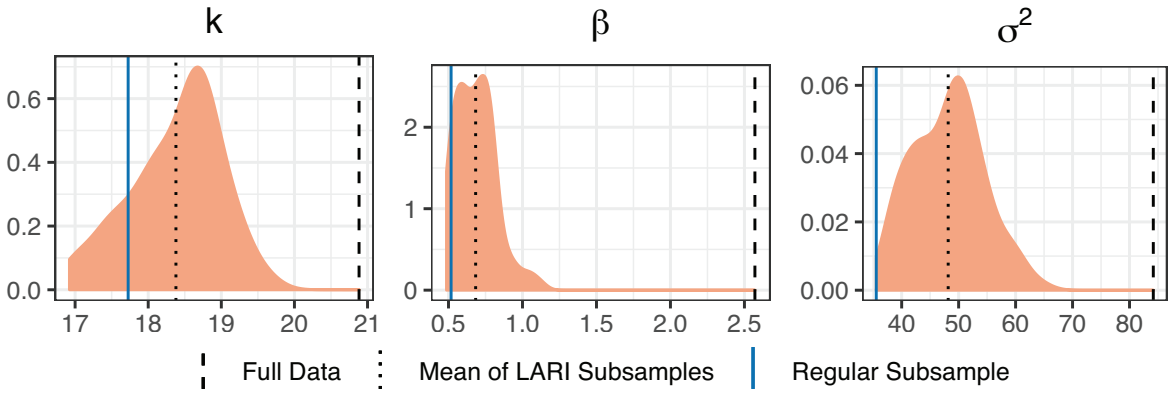

Figure 2.9. The estimated values of the parameters k, β and σ^2 from a regular subsample and 100 corresponding LARI subsamples (orange empirical density).

The potential surface estimated with the full data is shown in Figure 2.8 with observations depicted in white. The potential surfaces estimated with subsampled data are similar in appearance, so we analyze them by comparing estimates of k, β, and σ^2. The 3 groups of subsamples resulted in identical conclusions, so we chose a random group to plot in Figure 2.9. In Figure 2.9, we display the parameter estimation results for k, β, and σ^2. While all subsamples led to underestimation of the model parameters compared to the full data, the LARI subsamples are closer to the full data. In particular, when estimating σ^2, the LARI subsamples always outperform the regular subsample.

2.6 Carpenter Ant Example

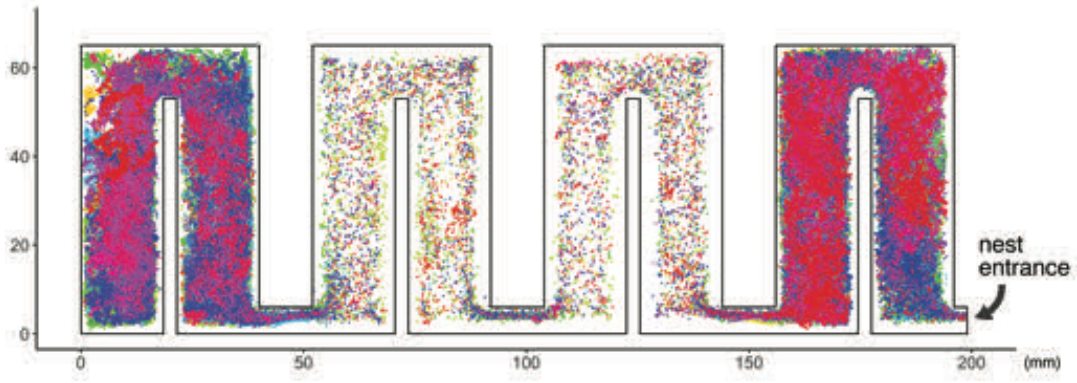

Figure 2.10. 14, 401 positions over 4 hours for each of the 73 ants, color coded by individual.

We now turn to the dataset introduced in Section 2.1, which consists of the positions of 78 ants at 1 second intervals over a 4 hour time frame (14,401 total observations per ant). Researchers observed the ants in a $200 \times 65 \times 6$ mm wooden nest, shown in Figure 2.10 along with the positions of all ants at all time points. The ants could enter or exit the nest at any time to utilize a separate foraging area. The data collection procedure is described in further detail by Modlmeier et al. (2019). In Figure 2.11, we illustrate movement observed for one individual who stayed inside the nest throughout the 4 hour time frame.

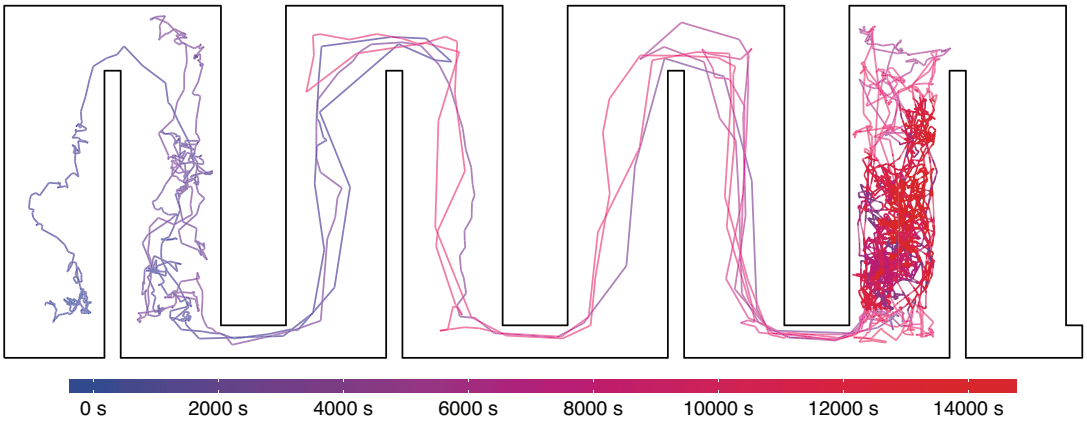

Figure 2.11. 14, 401 positions over 4 hours for a single ant, color coded by observation time.

From this data we obtained four datasets for comparison, one of which is the full data (1 second intervals). The three additional datasets are subsamples from the full data; we produced one with the regular sampling design using $h = 3$ seconds, one dataset

with the LARI sampling design using $h = 5$ seconds, and one with the regular sampling design using $h = 5$ seconds. Ants display stop and start behavior, but the SDE model alone cannot handle state switching. Thus, we removed observations where the ants were stationary within each dataset. After removal of stationary observations, a total of $232,571$ observations remained in the full dataset, the largest of the four datasets. Similar to Russell et al. (2018), we only consider modeling ant movement when ants are moving.

2.6.1 Ant Movement Model

In this example, we apply the framework from (2.9) to model ant movement behavior. We represent the surface of the ant nest using $J = 9,998$ grid cells (1×1 mm). Following Russell et al. (2018), we specify spatially-varying motility and potential surfaces to capture spatial heterogeneity in ant movement. The zeroth order spline representations of the potential and motility surfaces evaluated at position $\mathbf{r}_\tau$ are

$$p(\mathbf{r}_\tau) \equiv \sum_{j=1}^{J} p_j s_j(\mathbf{r}_\tau)$$
$$m(\mathbf{r}_\tau) \equiv \sum_{j=1}^{J} m_j s_j(\mathbf{r}_\tau)$$

where

$$s_j(\mathbf{r}_\tau) \equiv \begin{cases} 1, & \mathbf{r}_\tau \text{ in } j^{th} \text{ grid cell} \\ 0, & \text{otherwise} \end{cases}$$

and p_j and m_j are the potential and motility surfaces respectively, evaluated in grid cell j. Of course, there are other basis functions we could use to build potential and motility surfaces, such as thin plate splines. However, thin plate splines and other bases are more difficult work with in the constrained geometry of the ant nest.

The model equation (2.9) has infinitely many solutions if constraints are not imposed. To obtain identifiability, we fix $\sigma = 1$ as in Russell et al. (2018). Russell et al. (2018) took a Bayesian approach to parameter estimation with a similar model. However, our novel approximation in (2.10) allows for a direct evaluation of the likelihood of animal locations observed at irregular time intervals. We propose an algorithm for estimation of model parameters based on maximizing the likelihood (2.10) while penalizing the roughness of the potential and motility surfaces. Modlmeier et al. (2019) used a related algorithm, but only allowing for regularly sampled data. Modlmeier et al. (2019) also do

not provide full mathematical details of the algorithm, which we provide in summary here and in detail in Supplemental Materials S2.4.

We estimate $\mathbf{p} \equiv \begin{bmatrix} p_1 & \dots & p_J \end{bmatrix}'$ and $\mathbf{m} \equiv \begin{bmatrix} m_1 & \dots & m_J \end{bmatrix}'$ with an iterative procedure beginning with the model equation (2.9). Rewriting (2.9), we have

$$\frac{\mathbf{r}_{\tau+2} - \mathbf{r}_{\tau+1}}{h_{\tau+1}} - \frac{\mathbf{r}_{\tau+1} - \mathbf{r}_{\tau}}{h_{\tau}} \sim \mathrm{N}\left(\beta h_\tau \left(\boldsymbol{\mu}(\mathbf{r}_\tau) - \frac{\mathbf{r}_{\tau+1} - \mathbf{r}_\tau}{h_\tau}\right), \mathrm{diag}(c^2(\mathbf{r}_\tau)h_\tau)\right) \tag{2.23}$$

where $\boldsymbol{\mu}(\mathbf{r}_\tau) = m(\mathbf{r}_\tau)[-\nabla p(\mathbf{r}_\tau)]$ is estimated as a function of $\mathbf{m}$ and $\mathbf{p}$ and $c(\mathbf{r}_\tau) = m(\mathbf{r}_\tau)$ is estimated as a function of $\mathbf{m}$. Thus, we refer to $\mathbf{p}$ as a mean parameter, and $\mathbf{m}$ could be considered both a mean and a variance parameter. However, we will estimate $\mathbf{m}$ using the variance and thus we refer to it as a variance parameter.

We hold out 20% of the data to use when choosing the tuning parameter λ, which controls the smoothness of the surfaces, later in the procedure. The remaining 80% of the data are fit simultaneously for all ants, assuming ants move independently and there is no correlation in the x- and y-components of movement. The procedure is similar to restricted maximum likelihood (REML) approaches common in mixed models, as we use residuals to estimate covariance parameters, which are then used to estimate mean parameters. Our proposed approach is as follows:

1. Obtain a preliminary estimate of mean parameters (β and $\mathbf{p}$) assuming the motility surface is constant (model errors are independent and identically distributed).

2. Estimate variance parameters ($\mathbf{m}$) using residuals from step 1.

3. Estimate mean parameters (β and $\mathbf{p}$) conditioned on the variance estimates from step 2.

For details about the estimation approach, refer to Supplemental Material S2.4.

2.6.2 Results

Section 2.6.1 describes a computationally efficient method of fitting spatially-varying coefficients in SDE movement models. We completed the 3-step procedure for 17 values of the tuning parameter using the full data (232,571 total observations) in less than 25 minutes. We completed the procedure in the programming language R (version 3.5.2) on a MacBook Pro with a 2.9 GHz Intel Core i5 processor and 8 GB of 2133 MHz LPDDR3 RAM.

We estimated motility and potential surfaces for four samples: the full data with observations at 1 second intervals, a subsample with observations every 3 seconds, a subsample with observations every 5 seconds, and a LARI subsample with regular samples every 10 seconds coupled with a random point in between each pair of regular samples. The 10 second LARI sample and every 5 second regular sample have an equal number of observations, so comparison of the results from these two datasets amounts to a direct comparison of the LARI sampling scheme to the regular sampling scheme for this data.

For each of the four samples, we chose the optimal value of λ separately based on prediction accuracy on the holdout set. We chose $\log(\lambda) = 0$ for the full data, $\log(\lambda) = 2$ for the every 3 seconds sample, $\log(\lambda) = 3$ for the every 5 seconds sample, and $\log(\lambda) = 4$ for the 10 second LARI sample.

Figure 2.12 displays the estimated log motility surfaces for the four datasets. Since high motility indicates high activity, it is evident from the plots that the ants moved more quickly in the center chambers (Modlmeier et al., 2019). Assuming the motility surface generated with the full data is closest to the truth, we found mean squared errors (MSE) of the log motility surfaces in Figure 2.12(B–D) by summing squared differences between those surfaces and the surface in Figure 2.12(A) over all grid cells. As shown in the text in Figure 2.12, the MSE for the 10 second LARI sampling scheme is smaller than the every 5 seconds sampling scheme, which suggests that the motility surface was better estimated with the LARI subsample than the regular subsample of the same size. Each of the three subsamples underestimated the motility surface compared to the full data, a sign that regardless of sampling design, we lost information about the motility surface by subsampling (Figure 2.13).

In Figure 2.14, we show the estimated potential surfaces with gradient vectors pointing down the gradient of potential surface. The gradient vectors depict the negative gradient of the potential surface scaled by 5 to improve visibility. We chose to plot the gradient vectors in every third grid cell for visual clarity. Since the potential surface is identifiable only up to an additive constant, we subtracted the mean from each potential surface to view them on roughly the same scale. We carried out comparisons between potential surfaces through mean squared distance (MSD) between the ends of the gradient vectors generated from the full data and those generated from the subsample. Of the three subsamples, the potential surface estimated with the 10 second LARI subsample had the smallest MSD (Table 2.1). We then decomposed the MSD into two additional metrics: mean error in magnitude and angle of the gradient vectors. The 10 second LARI design resulted in a smaller error in the angle of the gradient vectors compared to the every

	Every 3 Seconds	Every 5 Seconds	10s LARI
Mean Error in Gradient Vector Magnitude	-0.2216	-0.5045	-1.624
Mean Error in Gradient Vector Angle	-0.0317	0.0565	0.0164
MSD between Gradient Vectors	18.1455	21.2761	14.8337

Table 2.1. Potential surface error statistics for the three subsamples.

5 second design, but the 5 second design resulted in a smaller mean error in gradient vector magnitude (Table 2.1). On average, all three potential surfaces estimated with subsamples of the data are too smooth, i.e., the gradient vectors are biased toward zero (Table 2.1).

Since the LARI sampling scheme requires random samples in each 10 second time interval, there are $9^{1,439}$ possible 10 second LARI subsamples. To evaluate the variability attributed to this random component, we took 50 different 10 second LARI subsamples and fit each subsample separately. We found that the 10 second LARI subsample consistently outperformed the every 5 second subsample in regard to all the metrics we looked at except magnitude of potential surface gradient vectors (Figure 2.15).

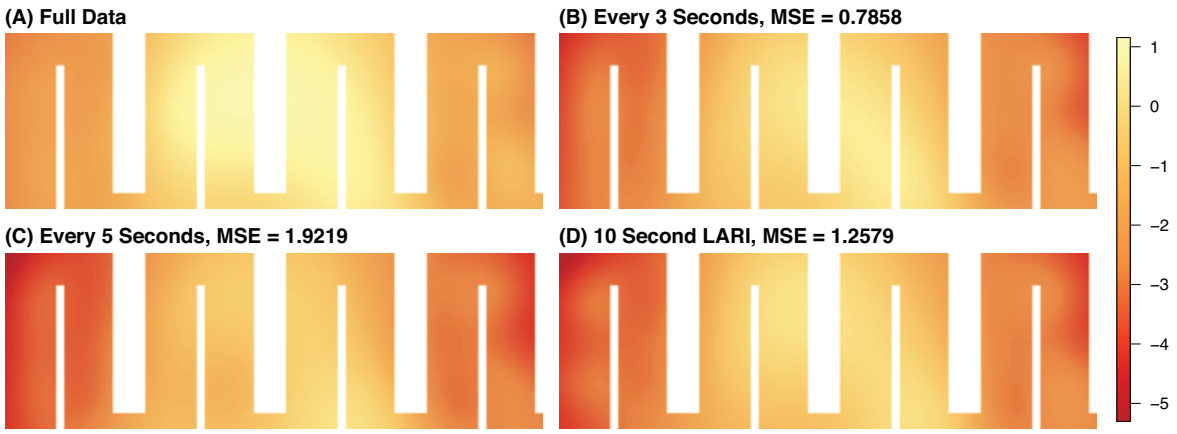

Figure 2.12. The natural log of the motility surfaces estimated using the full data (A) and 3 subsamples (B)-(D).

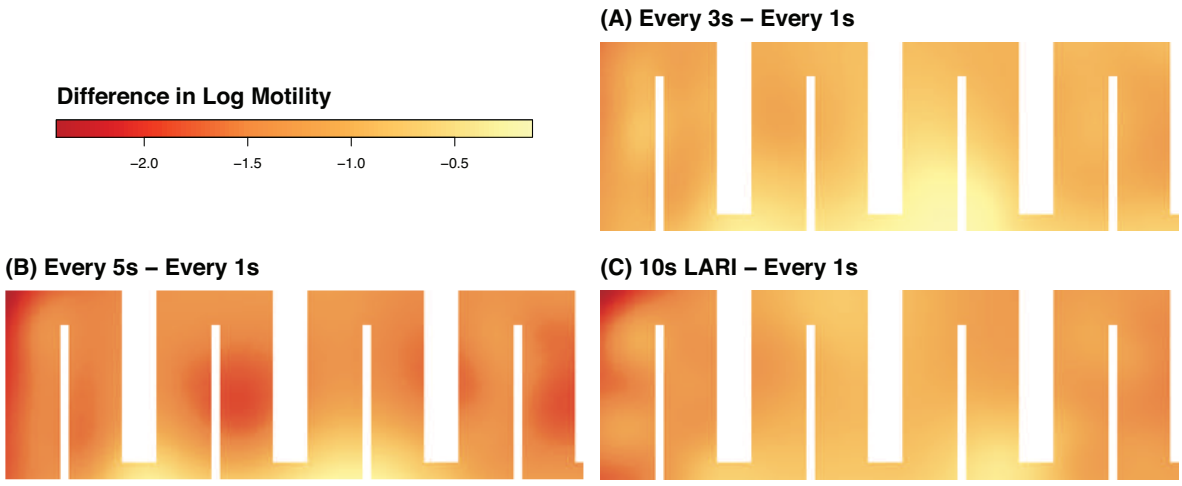

Figure 2.13. Differences between the estimated log motility surfaces from each of the subsamples and the full data (calculated by grid cell). Negative values indicate underestimation of the motility surface.

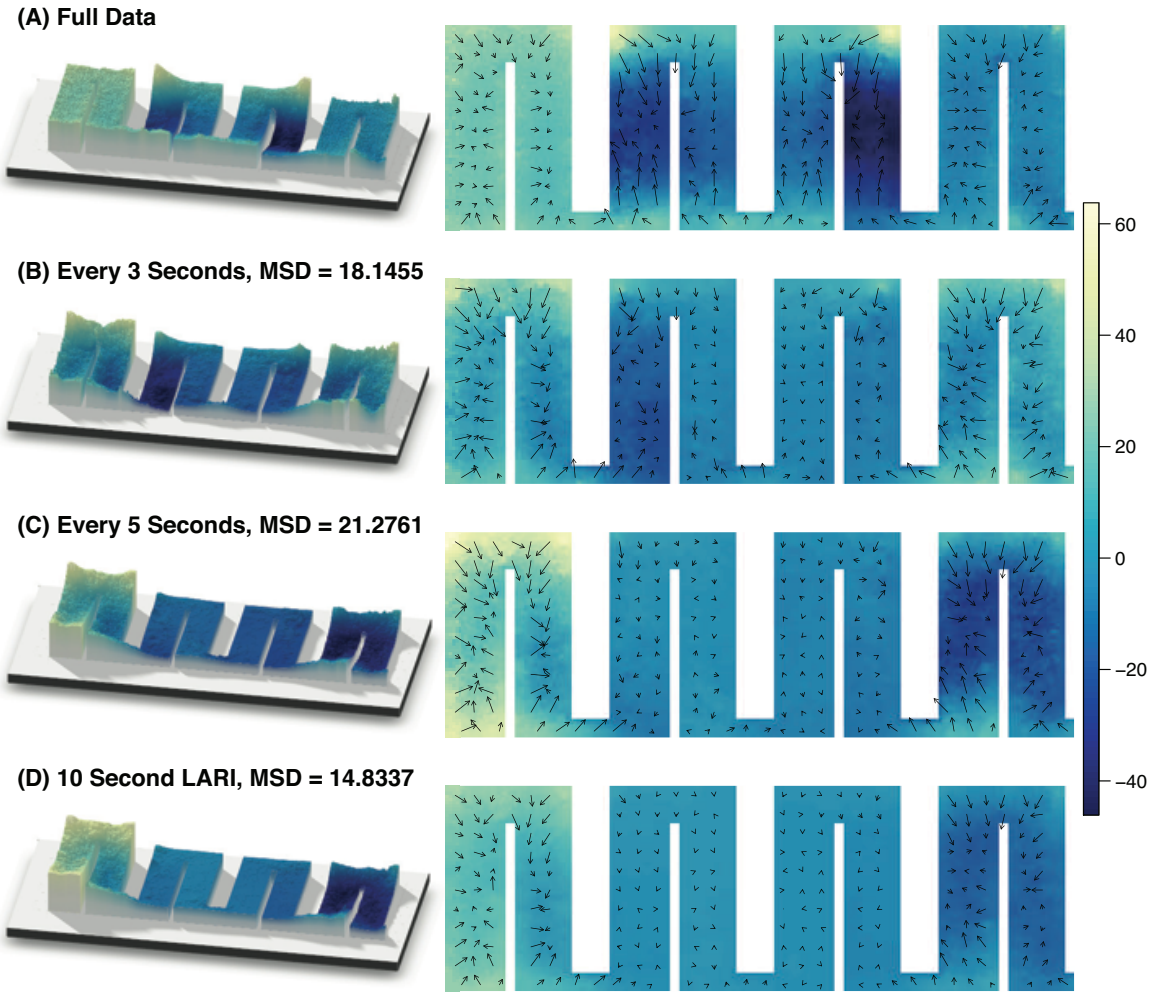

Figure 2.14. Potential surfaces estimated with the four samples. The same potential surfaces are plotted in three dimensions on the left (using the `rayshader` R package) and in two dimensions on the right.

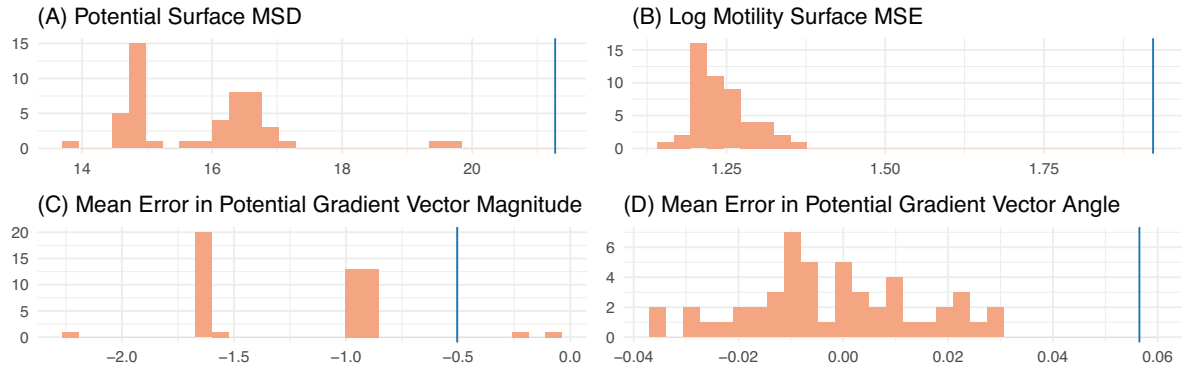

Figure 2.15. Statistics calculated for the motility and potential surfaces fit using 50 different 10 second LARI subsamples (orange) are compared to statistics from the every 5 second subsample (blue).

2.7 Discussion

The simulation, guppy data, and ant data examples describe vastly different systems, but the sampling procedures laid out in Section 2.2 and general model framework described in Section 3.3 were applicable in all three cases. We chose to highlight these three examples to emphasize the generalizability of the SDE framework and our proposed sampling approach. In all of the examples, the LARI sampling design led to greater accuracy in parameter estimation compared to samples at regular time intervals. As shown in the simulation example in Section 2.4, the LARI subsample also resulted in better prediction of missing values compared to the regular subsample. We conclude that when conducting animal movement research on data similar to that examined in this paper, a LARI sample is preferable to a regular sample of the same sample size and duration.

We determined that predicting finer scale movement (infill) was greatly useful for parameter estimation in the simulation example (see Supporting Materials S2.2 for details). This result implies that imputation of missing data at a finer scale might improve parameter estimation in the guppy and ant data examples as well. In the ant example, we found that the motility surfaces were underestimated using both regular and LARI sampling schemes. Augmenting observations with additional latent infill points in the ant example would introduce more variation in the movement paths, potentially reducing the underestimation of the motility surface. In the guppy example, all three parameters were underestimated when the data were subsampled, suggestion that augmenting observations might be useful here as well. However, in the simulation example,

predicting missing data with the Metropolis-within-Gibbs algorithm was computationally intensive. We needed 90 hours of computational time on a high-performance computing cluster to simulate, subsample, and fit each of the 150 datasets.

While computational complexity was an issue in the simulation example, the multi-step model-fitting procedure in the ant example was extremely computationally efficient. For each sample, less than 25 minutes on a laptop computer were required to fit the model with a range of 17 different tuning parameter values. The scalability and computational feasibility of adding components to this model are huge assets.

In this paper, we presented a general SDE modeling framework along with three model-fitting procedures. The SDE framework has a wide range of possible extensions, including the addition of known seasonal variation, asynchronous movement across individuals, and state switching. The model framework as presented here assumes the animals are in constant motion throughout the study period, allowing us to describe movement behavior while an individual is in motion but not when the individual is stationary. In the ant example, we met this restriction by removing all data points where the ants were not moving. The addition of state switching would allow us to capture the start and stop behavior in the ant data and could be used to predict finer scale movement.

In this paper, we have compared two sampling designs in the context of animal movement. Thus, we have barely scraped the surface of a research area which is largely unexplored: optimal sampling for animal movement. As of present, there is no comprehensive guide to sampling design for animal movement. A thorough examination of sampling design for animal movement would allow researchers to allocate resources more efficiently and discover details of animal movement behavior that might otherwise be lost.

2.8 Data availability

The raw carpenter ant data is available through Dryad (DOI: 10.5061/dryad.sh4m4s6), as is the guppy data (DOI: 10.5061/dryad.kt3109v7).

S2 Supplement to a Lattice and Random Intermediate Point Sampling Design for Animal Movement

S2.1 Examples of Model Capabilities

We will briefly examine data simulated from (10) in the main text with 3 different sets of parameters. In all 3 simulations, we started the simulation at $\begin{bmatrix}0 & 0\end{bmatrix}'$ for the first 2 time steps and used $\beta = 0.4$ and $\sigma = 0.5$. In the first simulation,

$$p(\mathbf{r}_\tau) = x_\tau \tag{2.24}$$

$$m(\mathbf{r}_\tau) = \begin{cases} 5, & y_\tau \leq 0 \\ 20, & y_\tau > 0 \end{cases} \tag{2.25}$$

result in movement with average drift in the negative x direction and where the average speed is higher in the space where $y > 0$ than where $y \leq 0$. Since in (10) each simulated value $\mathbf{r}_\tau$ is a function of the motility evaluated two observations prior, the average speed at $\mathbf{r}_\tau$ is higher (and step size smaller) when $y_{\tau-2} > 0$ than when $y_{\tau-2} \leq 0$. The first 20 simulated time points are plotted in Figure 2.16A-B with the potential and motility surfaces which generated them. Figure 2.16C depicts the difference in average step size attributed to the motility surface in a simulation of 1,000 time steps.

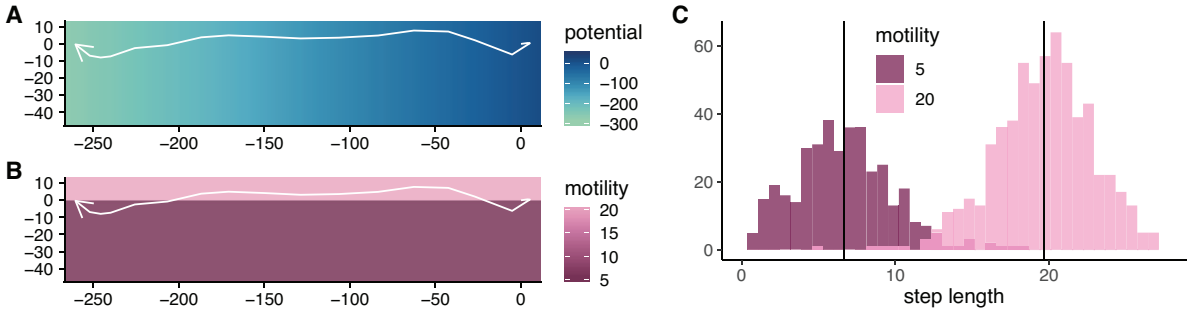

Figure 2.16. Panels (A)-(B): 20 locations simulated from (10) with parameters (2.24)–(2.25). The simulations are shown with the potential surface (A) and motility surface (B). The arrow points in the direction the simulated individual is heading. Panel (C): 1,000 locations simulated from (10) with parameters (2.24)–(2.25). Panel (C) displays histograms of the distances from $\mathbf{r}_{\tau-1}$ to $\mathbf{r}_\tau$ for each possible value of $m(\mathbf{r}_{\tau-2})$. Black vertical lines represent the means for each group.

We reran the simulation with

$$m(\mathbf{r}_\tau) = \begin{cases} 5, & y_\tau \leq 0, \\ 10, & y_\tau > 0. \end{cases} \tag{2.26}$$

This change left the average step size in the low motility area unchanged but led to a reduction in the average step size in the high motility area (Figure 2.17).

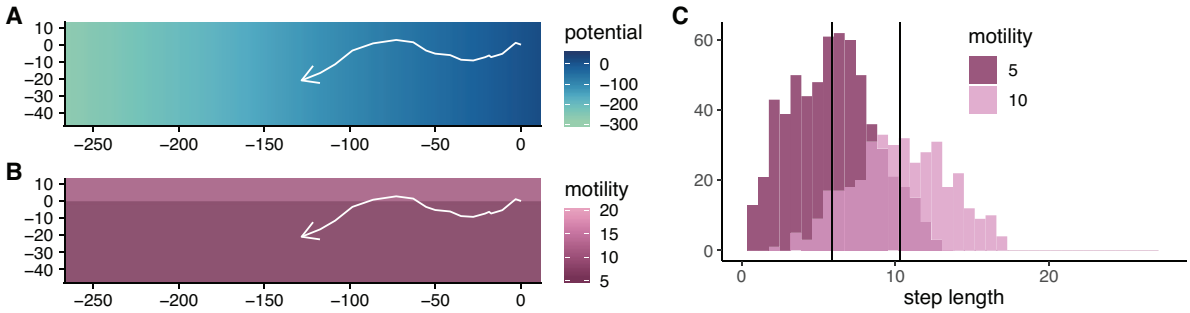

Figure 2.17. Plots are as in Figure 2.16 using parameters (2.24) and (2.26).

Lastly, we used the original motility surface (2.25) and reduced the severity of the gradient of the potential surface by letting

$$p(\mathbf{r}_\tau) = 0.5x_\tau \tag{2.27}$$

which led to a reduced average step size in both groups (Figure 2.18C) as well as less direct movement down the gradient of the potential surface (Figure 2.18A-B).

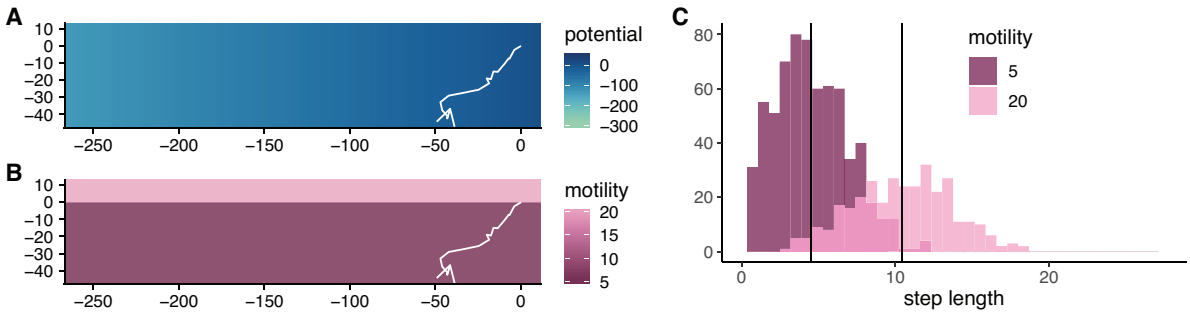

Figure 2.18. Plots are as in Figure 2.16 using parameters (2.27) and (2.25).

S2.2 Simulation Example without Infill

Dividing equation (2.13) in the main text by $h_\tau^{1/2}$ results in

$$\frac{\mathbf{r}_{\tau+2} - \mathbf{r}_{\tau+1}}{h_{\tau+1} h_\tau^{1/2}} - \frac{\mathbf{r}_{\tau+1} - \mathbf{r}_\tau}{h_\tau^{3/2}} = \alpha \left(-2 h_\tau^{1/2} \mathbf{r}_\tau\right) - \beta \left(\frac{\mathbf{r}_{\tau+1} - \mathbf{r}_\tau}{h_\tau^{1/2}}\right) + \sigma \boldsymbol{\epsilon}_\tau, \tag{2.28}$$

which is a linear model with independent and identically distributed errors. We could then fit (2.28) with simple linear regression. Assessment of linear regression estimation accuracy was carried out on a subsample including every other simulated data value. Thus, the subsample is very similar to the true data. However, of the 95% confidence intervals constructed for β, α, and σ, only the confidence interval for β captured the true parameter value. Adoption of a Bayesian approach would allow us to place priors on the parameters that may lead to better approximations.

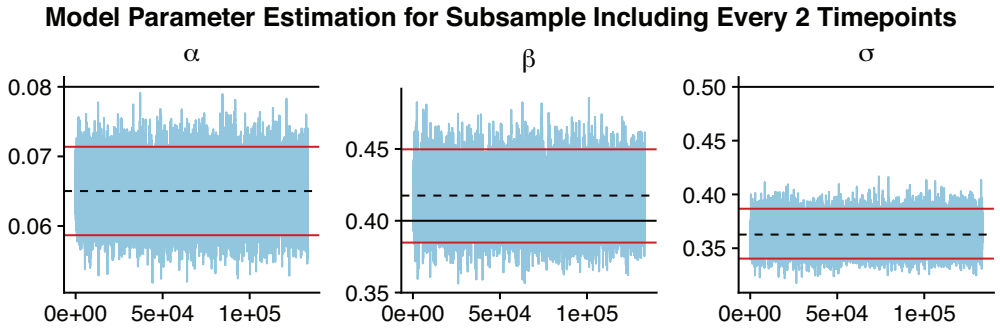

Figure 2.19. MCMC draws from the posterior distribution conditioned on a regular subsample with every other timepoint. The MCMC draws are plotted in blue, the red lines bound equal-tailed 95% credible intervals, the black dashed line is the estimated posterior mean and the black solid line is the true parameter value.

Vague priors

$$\pi(\sigma) = \text{InverseGamma}\,(1, 1) \tag{2.29}$$
$$\pi(\beta) = \text{Exponential}\,(1) \tag{2.30}$$
$$\pi(\alpha) = \text{Normal}\left(0, 10^2\right) \tag{2.31}$$

were placed on the parameters and Metropolis-Hastings samples were drawn from the posterior distribution. As in the linear regression model, we subset the data to include every other time point. Despite this subset being very close to the original simulation, only β was captured in its equal-tailed 95% credible interval, as shown in Figure 2.19.

S2.3 Bayesian framework

To estimate β, α, σ, and $\{\mathbf{r}\}_{\text{unobs}}$ as alluded to in Section 4.3, we took a Bayesian approach and constructed an MCMC algorithm to sample from the joint posterior $\pi(\alpha, \beta, \sigma, \{\mathbf{r}\}_{\text{unobs}}|\{\mathbf{r}\}_{\text{obs}})$ where

$$\pi(\alpha, \beta, \sigma, \{\mathbf{r}\}_{\text{unobs}}|\{\mathbf{r}\}_{\text{obs}}) \propto P(\{\mathbf{r}\}|\beta, \sigma, \alpha)\pi(\beta)\pi(\sigma)\pi(\alpha).$$

Vague priors (2.29)–(2.31) were placed on the model parameters. We also tried increasing the variance of these prior distributions by a factor of 10, but we saw no meaningful difference in the resulting empirical posterior distributions.

The initial positions $\mathbf{r}_\tau = \begin{bmatrix} x_\tau & y_\tau \end{bmatrix}'$, $\tau = 1, 2$ were assigned independent uniform priors

$$\begin{aligned} x_1 &\sim \text{Uniform}\,(\min\{\mathbf{x}\}_{\text{obs}}, \max\{\mathbf{x}\}_{\text{obs}}) \\ y_1 &\sim \text{Uniform}\,(\min\{\mathbf{y}\}_{\text{obs}}, \max\{\mathbf{y}\}_{\text{obs}}) \\ x_2 &\sim \text{Uniform}\,(\min\{\mathbf{x}\}_{\text{obs}}, \max\{\mathbf{x}\}_{\text{obs}}) \\ y_2 &\sim \text{Uniform}\,(\min\{\mathbf{y}\}_{\text{obs}}, \max\{\mathbf{y}\}_{\text{obs}}) \end{aligned}$$

resulting in the joint distribution of the observed and unobserved positions

$$P(\{\mathbf{r}\}|\beta, \sigma, \alpha) = P(x_1)P(y_1)P(x_2)P(y_2)\prod_{\tau=1}^{n-2} P(\mathbf{r}_{\tau+2}|\mathbf{r}_{\tau+1}, \mathbf{r}_\tau, \beta, \sigma, \alpha).$$

We employed a Metropolis-within-Gibbs sampler where α, β, σ, and the elements of $\{\mathbf{r}\}_{\text{unobs}}$ are updated in turn. We performed updates using random walk Metropolis steps

for all parameters with adaptive tuning (Haario et al., 2001; Roberts and Rosenthal, 2009) (also see Craiu and Rosenthal (2014)) to improve mixing. We ran the adaptive algorithm for $100,000$ iterations, which were subsequently discarded as burn-in, to tune the proposal covariance matrix. We drew the next $100,000$ MCMC samples using the tuned proposal covariance matrix from the first $100,000$ iterations. Finally, we used these $100,000$ iterations for posterior estimation. This procedure took about 90 hours of computational time on a high-performance computing cluster. While not all chains converged in this many iterations, running all chains until convergence would have been infeasible in this framework. Thus, we chose to assess convergence at an individual level and remove the simulated paths which failed to meet the convergence criterion.

S2.4 Estimation

The model equation based on (9) in the main text is

$$g_{x\tau} = v_{x\tau}\beta + m(\mathbf{r}_\tau)\mathbf{a}'_{x\tau}\boldsymbol{\gamma} + \epsilon_\tau \tag{2.32}$$

where

$$g_{x\tau} \equiv \frac{x_{\tau+2} - x_{\tau+1}}{h_{\tau+1}} - \frac{x_{\tau+1} - x_\tau}{h_\tau} \tag{2.33}$$

$$v_{x\tau} \equiv x_\tau - x_{\tau+1} \tag{2.34}$$

$$\boldsymbol{\gamma} \equiv \begin{bmatrix} -\beta p_1 & -\beta p_2 & \dots & -\beta p_J \end{bmatrix}' \tag{2.35}$$

$$\epsilon_\tau \sim \mathrm{N}(0, h_\tau \sigma^2 m^2(\mathbf{r}_\tau)). \tag{2.36}$$

Element $u = 1, \dots, J$ of column vector $\mathbf{a}_{x\tau}$ is $\frac{1}{2}h_\tau$ if grid cell u contains $\begin{bmatrix} x_\tau + 1 & y_\tau \end{bmatrix}'$, $-\frac{1}{2}h_\tau$ if grid cell u contains $\begin{bmatrix} x_\tau - 1 & y_\tau \end{bmatrix}'$, and 0 otherwise. This formulation comes from the use of a raster-based centered difference equation to approximate the gradient of the potential surface. We include h_τ in $\mathbf{a}_{x\tau}$ instead of keeping the two separate for ease of vector notation in (2.32). If we multiply $\mathbf{a}'_{x\tau}$ by a vector of grid cell values, we get the product of h_τ and the x component of the estimated gradient of the gridded surface at $\mathbf{r}_\tau$. Thus, we approximate

$$\beta h_\tau m(\mathbf{r}_\tau) \left[-\frac{\partial}{\partial x} p(\mathbf{r}_\tau) \right] \approx m(\mathbf{r}_\tau)\mathbf{a}'_{x\tau}\boldsymbol{\gamma}$$

$$= -\beta h_\tau m(\mathbf{r}_\tau) \frac{p\left(\begin{bmatrix} x_\tau + 1 & y_\tau \end{bmatrix}'\right) - p\left(\begin{bmatrix} x_\tau - 1 & y_\tau \end{bmatrix}'\right)}{2}.$$

Similarly, we could construct (2.32)–(2.36) in the y direction by replacing each x with y. Element $u = 1, \ldots, J$ of $\mathbf{a}_{y\tau}$ is $\frac{1}{2}h_\tau$ if grid cell u contains $\begin{bmatrix} x_\tau & y_\tau + 1 \end{bmatrix}'$, $-\frac{1}{2}h_\tau$ if grid cell u contains $\begin{bmatrix} x_\tau & y_\tau - 1 \end{bmatrix}'$, and 0 otherwise.

We estimate $\mathbf{p} \equiv \begin{bmatrix} p_1 & \ldots & p_J \end{bmatrix}'$ and $\mathbf{m} \equiv \begin{bmatrix} m_1 & \ldots & m_J \end{bmatrix}'$ with an iterative procedure. Our proposed approach is as follows:

1. Obtain a preliminary estimate of mean parameters (β and $\mathbf{p}$) assuming the motility surface is constant (model errors in (2.36) are independent and identically distributed).

2. Estimate variance parameters ($\mathbf{m}$) using residuals from step 1.

3. Estimate mean parameters (β and $\mathbf{p}$) conditioned on the variance estimates from step 2.

In step 1, we assume the motility surface is approximately constant (i.e., m_j is similar to the motility surface evaluated in grid cells adjacent to cell j) so that the error variance in (2.36) is constant and we can absorb $m(\mathbf{r}_\tau)$ in the estimate of $\boldsymbol{\gamma}$ when we fit the model (2.32). Let $\mathbf{E} \equiv \begin{bmatrix} \mathbf{v} & \mathbf{A} \end{bmatrix}$ where $\mathbf{v}$ is a column vector of all $v_{x\tau}$ and $v_{y\tau}$ that are not in the holdout set and $\mathbf{A}$ is a matrix with rows consisting of all $\mathbf{a}'_{x\tau}$ and $\mathbf{a}'_{y\tau}$ that are not in the holdout set. Let $\mathbf{g}$ be the column vector containing all $g_{x\tau}$ and $g_{y\tau}$ that are not in the holdout set. When combining over the indices $x\tau$ and $y\tau$, we ensure the combined values are ordered in the same way each time by ant, time, and direction.

We obtain preliminary estimates of β and $\boldsymbol{\gamma}$ using penalized least squares estimation

$$\begin{aligned} \begin{bmatrix} \hat{\beta} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} &= \text{argmin}_{\beta,\boldsymbol{\gamma}} \left\{ \left(\mathbf{g} - \mathbf{E} \begin{bmatrix} \beta \\ \boldsymbol{\gamma} \end{bmatrix} \right)' \left(\mathbf{g} - \mathbf{E} \begin{bmatrix} \beta \\ \boldsymbol{\gamma} \end{bmatrix} \right) + \lambda \begin{bmatrix} \beta \\ \boldsymbol{\gamma} \end{bmatrix}' \mathbf{Q} \begin{bmatrix} \beta \\ \boldsymbol{\gamma} \end{bmatrix} \right\} \\ &= (\mathbf{E}'\mathbf{E} + \lambda \mathbf{Q})^{-1} \mathbf{E}'\mathbf{g} \end{aligned}$$

where the $(J+1) \times (J+1)$ penalty matrix $\mathbf{Q}$ ensures the potential surface is smooth. To bypass penalization in estimation of β, we let the first row and column of $\mathbf{Q}$ be $\mathbf{0}$ vectors. If we ignore the first row and column of $\mathbf{Q}$, the elements of the $J \times J$ submatrix are $\{\mathbf{Q}_{ij} : i, j = 1, \ldots, J\}$. $\mathbf{Q}$ penalizes the sum of squared first differences between the

estimated potential surface in neighboring grid cells, i.e.,

$$\begin{aligned}\begin{bmatrix}\beta\\ \boldsymbol{\gamma}\end{bmatrix}' \mathbf{Q} \begin{bmatrix}\beta\\ \boldsymbol{\gamma}\end{bmatrix} &= \sum_{i,j \text{ adjacent grid cells}} (\gamma_i - \gamma_j)^2 \\ &= \sum_{i,j \text{ adjacent grid cells}} (\beta p_j - \beta p_i)^2 .\end{aligned}$$

This is accomplished by letting $\mathbf{Q}_{ii}$ be the count of grid cells adjacent to grid cell i. The off-diagonal elements $\{\mathbf{Q}_{ij} : i \neq j\}$ are defined

$$\mathbf{Q}_{ij} = \begin{cases} -1, & \text{grid cells } i \text{ and } j \text{ are adjacent} \\ 0, & \text{otherwise.} \end{cases}$$

As the tuning parameter λ increases, the estimated potential surface becomes smoother. This is similar to putting a conditional autoregressive prior on $\boldsymbol{\gamma}$.

In step 2, we use the residual term

$$\hat{\boldsymbol{\epsilon}} \equiv \mathbf{g} - \mathbf{E} \begin{bmatrix}\hat{\beta}\\ \hat{\boldsymbol{\gamma}}\end{bmatrix}$$

from the preliminary estimation procedure in step 1 to estimate the motility surface. Similar to restricted maximum likelihood estimation, we estimate error variance using $\hat{\boldsymbol{\epsilon}}^2$. In (2.36), $\text{Var}(\epsilon_\tau) = \text{E}(\epsilon_\tau^2) = h_\tau \sigma^2 m^2(\mathbf{r}_\tau)$. We want to estimate $m^2(\mathbf{r}_\tau) = \text{E}(\epsilon_\tau^2 h_\tau^{-1} \sigma^{-2}) = \text{E}(\epsilon_\tau^2 h_\tau^{-1})$ (recall $\sigma = 1$). To avoid negative estimates of $m^2(\mathbf{r}_\tau)$, we estimate $\log(m^2(\mathbf{r}_\tau))$ instead. We fit a generalized additive model with mean

$$\text{E}\left[\log(\hat{\epsilon}_\tau^2 h_\tau^{-1})\right] = \nu + f(x_\tau, y_\tau) \tag{2.37}$$

where f is a non-linear smooth function of location and ν is the intercept term. We fit this generalized additive model using the `mgcv` R package (Wood, 2017). While alternative methods including maximum likelihood estimation of the residuals could be implemented, the generalized additive model is fast and results in accurate estimation when the data size is large. We exponentiate the fitted values to obtain estimates of $m^2(\mathbf{r}_\tau)$ for all observations.

We now have the components to easily obtain an estimate $\hat{\mathbf{m}} = \begin{bmatrix}\hat{m}_1 & \dots & \hat{m}_J\end{bmatrix}'$ of

the vector $\mathbf{m}$ of motility grid cell values. For $j = 1, \ldots, J$,

$$\hat{m}_j \equiv \hat{\nu} + \hat{f}(\mathbf{c}_j)$$

where $\hat{\nu}$ and $\hat{f}$ are estimates of ν and f from the generalized additive model fit and $\mathbf{c}_j$ contains the coordinates for the center of grid cell j.

In step 3, we use the original model equation (2.32), but now we replace $m(\mathbf{r}_\tau)$ with $\hat{m}(\mathbf{r}_\tau)$, the estimated motility surface evaluated at $\mathbf{r}_\tau$. Then we divide by $\hat{m}(\mathbf{r}_\tau)h_\tau^{1/2}$ to produce a model with independent and identically distributed errors. The resulting model equation is

$$\tilde{g}_{x\tau} = \tilde{v}_{x\tau}\beta + \tilde{\mathbf{a}}'_{x\tau}\boldsymbol{\gamma} + \tilde{\epsilon}_{x\tau} \tag{2.38}$$

where

$$\tilde{g}_{x\tau} \equiv \frac{g_{x\tau}}{\hat{m}(\mathbf{r}_\tau)h_\tau^{1/2}} \tag{2.39}$$

$$\tilde{v}_{x\tau} \equiv \frac{v_{x\tau}}{\hat{m}(\mathbf{r}_\tau)h_\tau^{1/2}} \tag{2.40}$$

$$\tilde{\mathbf{a}}'_{x\tau} \equiv h_\tau^{-1/2}\mathbf{a}'_{x\tau} \tag{2.41}$$

$$\boldsymbol{\gamma} \equiv \begin{bmatrix} -\beta p_1 & -\beta p_2 & \ldots & -\beta p_J \end{bmatrix}' \tag{2.42}$$

$$\tilde{\epsilon}_{x\tau} \sim \mathrm{N}(0, 1). \tag{2.43}$$

As in the original model equation, we construct (2.38)–(2.43) similarly in the y direction. Let $\tilde{\mathbf{E}} \equiv \begin{bmatrix} \tilde{\mathbf{v}} & \tilde{\mathbf{A}} \end{bmatrix}$ where $\tilde{\mathbf{v}}$ is a column vector of all $\tilde{v}_{x\tau}$ and $\tilde{v}_{y\tau}$ and $\tilde{\mathbf{A}}$ is a matrix with rows of all $\tilde{\mathbf{a}}'_{x\tau}$ and $\tilde{\mathbf{a}}'_{y\tau}$. Let $\tilde{\mathbf{g}}$ be the column vector containing all $\tilde{g}_{x\tau}$ and $\tilde{g}_{y\tau}$. This time, we obtain final estimates of β and $\boldsymbol{\gamma}$

$$\begin{bmatrix} \hat{\beta}^* \\ \hat{\boldsymbol{\gamma}}^* \end{bmatrix} \equiv (\tilde{\mathbf{E}}'\tilde{\mathbf{E}} + \lambda\mathbf{Q})^{-1}\tilde{\mathbf{E}}'\tilde{\mathbf{g}}.$$

Finally, we estimate $\mathbf{p}$, the vector of potential grid cell values, with

$$\hat{\mathbf{p}} \equiv -\frac{\hat{\boldsymbol{\gamma}}^*}{\hat{\beta}^*}.$$

As explained in step 1, the penalty matrix $\mathbf{Q}$ ensures the estimated potential surface is smooth. Since the potential surface only enters the model through its gradient, the

potential surface is only identifiable up to an unknown additive constant.

We repeated this 3-step procedure for $\log(\lambda) = -8, -7, \ldots, 7, 8$. We chose the $\log(\lambda)$ that resulted in the lowest mean squared prediction error

$$\left[\mathbf{g}_{\text{holdout}} - \hat{\mathbf{g}}_{\text{holdout}}\right]' \left[\mathbf{g}_{\text{holdout}} - \hat{\mathbf{g}}_{\text{holdout}}\right],$$

where

$$\hat{\mathbf{g}}_{\text{holdout}} \equiv \begin{bmatrix}\mathbf{v}_{\text{holdout}} & \text{diag}[\hat{m}(\mathbf{r}_{\text{holdout}})]\mathbf{A}_{\text{holdout}}\end{bmatrix} \begin{bmatrix}\hat{\beta}^* \\ \hat{\boldsymbol{\gamma}}^*\end{bmatrix}$$

and $\mathbf{g}_{\text{holdout}}$, $\mathbf{v}_{\text{holdout}}$, and $\mathbf{A}_{\text{holdout}}$ are identical to $\mathbf{g}$, $\mathbf{v}$, and $\mathbf{A}$ in step 1 except now they consist of only the holdout data. $\hat{m}(\mathbf{r}_{\text{holdout}})$ is a vector of estimated motility values at the positions in the holdout data.

S2.5 Simulation study of the estimation procedure

As outlined in Supplementary Materials S2.4, step 2 of the estimation procedure entails estimation of the squared motility surface evaluated at each observation, i.e., $m^2(\mathbf{r}_\tau) = \text{E}(\epsilon_\tau^2 h_\tau^{-1})$. We achieve this by estimating $\text{E}[\log(m^2(\mathbf{r}_\tau))]$ and exponentiating the fitted values. We recognize that this procedure may introduce bias, and thus we have conducted a simulation study to quantify this bias. We chose to simulate movement from simpler potential and motility surfaces than those fit with the ant data to avoid dealing with nest boundaries. Instead, we constructed a large surface of which only the center is utilized and analyzed our results in the areas where simulated data is present.

The data were simulated from an SDE model (3)-(4) with a linear motility and quadratic potential surface

$$\begin{aligned} m(x_\tau, y_\tau) &\equiv 0.02y_\tau + 2 \\ p(\mathbf{r}_\tau) &\equiv 0.02(\mathbf{r}_\tau - 50)'(\mathbf{r}_\tau - 50). \end{aligned}$$

The numerical approximation (9) becomes

$$\begin{aligned} \frac{\mathbf{r}_{\tau+2} - \mathbf{r}_{\tau+1}}{h_{\tau+1}} - \frac{\mathbf{r}_{\tau+1} - \mathbf{r}_\tau}{h_\tau} = \beta h_\tau \left([0.02y_\tau + 2][-0.04(\mathbf{r}_\tau - 50)] - \frac{\mathbf{r}_{\tau+1} - \mathbf{r}_\tau}{h_\tau}\right) \\ + h_\tau^{1/2}(0.02y_\tau + 2)\boldsymbol{\epsilon}_\tau \end{aligned}$$

which, after setting $h_\tau = 1$, we rearrange to produce the simulation generation equation

$$\mathbf{r}_{\tau+2} = 2\mathbf{r}_{\tau+1} - \mathbf{r}_\tau + \beta\left[(0.02y_\tau + 2)(-0.04(\mathbf{r}_\tau - 50)) - (\mathbf{r}_{\tau+1} - \mathbf{r}_\tau)\right] + (0.02y_\tau + 2)\boldsymbol{\epsilon}_\tau.$$

We performed 100 simulation runs, each of which generated 5 individual paths for 2000 time steps. For illustrative purposes, we plotted the simulated path of one individual on the potential and motility surfaces which generated the simulations in Figure 2.20. We fit each of these simulations using the estimation procedure in Section 4.3 to produce 100 pairs of estimated motility and potential surfaces. Each surface is a 50×50 grid of values, where each grid cell is 2×2. When comparing surfaces, we subset the surfaces to where the majority of the data were located: within a radius of 23 from the center $(50, 50)$. On average, 77.95% of the simulated observations fell within the radius of 23 (standard deviation = 1.21%).

We assessed motility and potential surface estimation using three metrics: mean potential gradient vector angle, the average angle between the gradient vectors for the estimated and true potential surfaces; mean potential gradient vector length error, the average of the magnitudes of the estimated potential surface gradient vector minus those of the true potential surface gradient vector, and mean motility surface error, the average of the grid cell values of the estimated motility surface minus those of the true motility surface. The empirical density of the metrics for the 100 simulations are displayed in Figure 2.21. As shown in Figure 2.21(A), the average angle of the potential surface gradient vectors is unbiased, as we would expect. As in 2.21(B), the average error in the magnitude of the potential surface gradient vectors is always negative, which is also unsurprising since the estimation procedure involves smoothing. In 2.21(C), we see that the motility surface is estimated well in simulations. Although the motility surface is underestimated on average in 84 of the 100 simulations, the interpretation of the results is not impacted by this slight underestimation. In Figure 2.22, we randomly chose one simulation to display next to the true potential and motility surfaces. The potential surfaces are centered at zero since they are unidentifiable up to an additive constant.

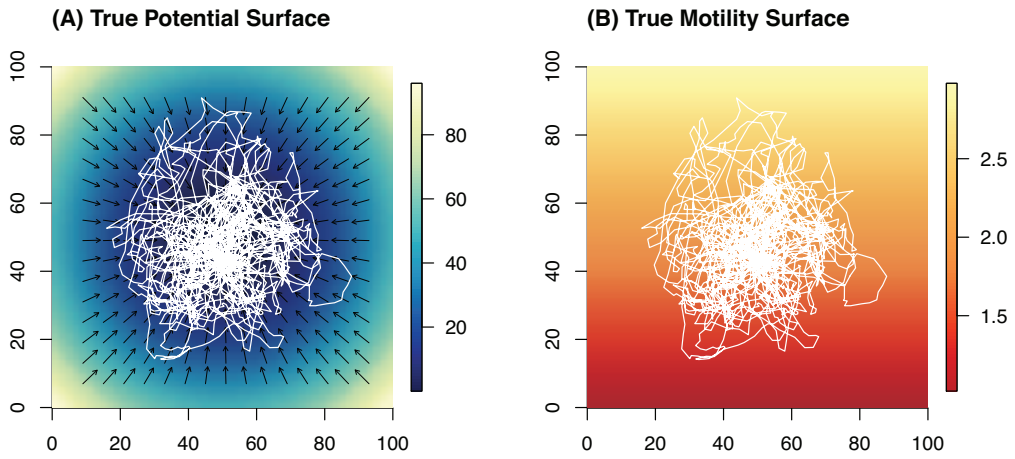

Figure 2.20. True potential and motility surfaces with one example simulation shown in white. Black gradient vectors on the potential surface depict the negative gradient scaled by 5.

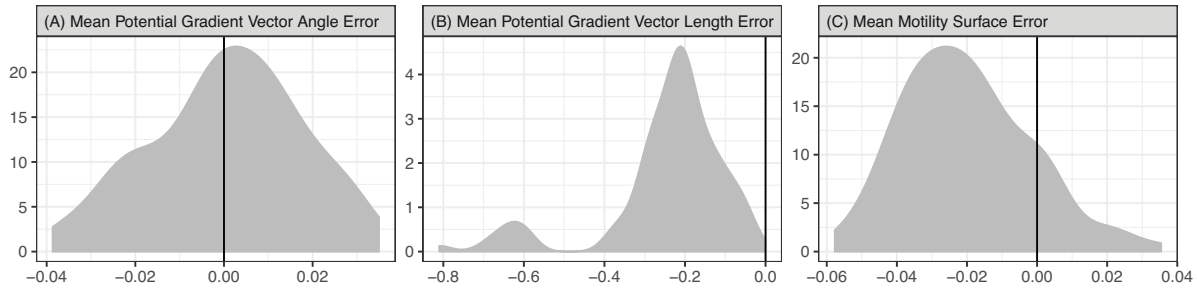

Figure 2.21. Density estimates for errors in the model fit on simulated data. The black vertical line is positioned at 0.

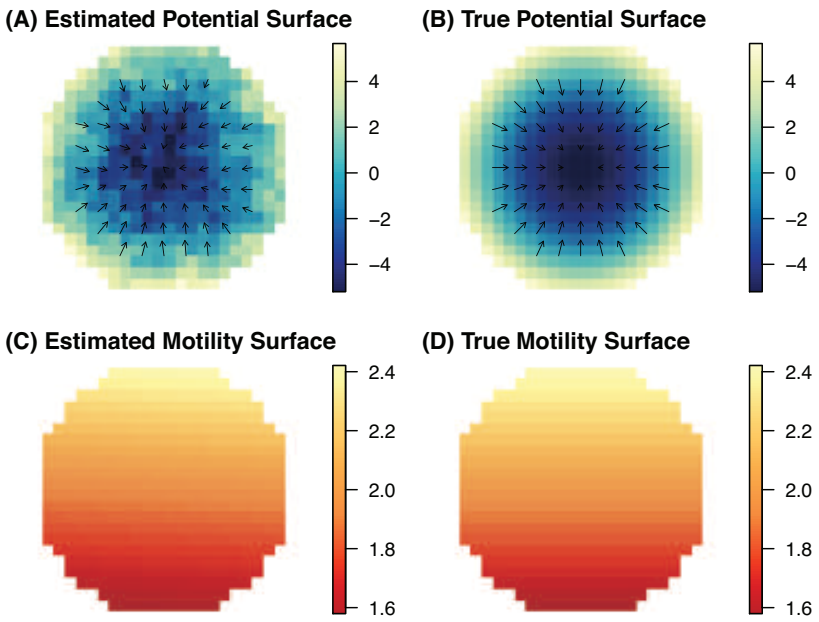

Figure 2.22. For one randomly selected simulation, the estimated potential (A) and motility (B) surfaces are compared to the true potential (B) and motility (D) surfaces.

Chapter 3
A Flexible Movement Model for Partially Migrating Species

This chapter was published in Spatial Statistics on February 24, 2022 (Eisenhauer et al., 2022).

3.1 Introduction

Movement behavior within species is often highly variable across individuals and years. While some animal populations follow similar migratory trajectories or travel in groups, many display partial migration, meaning seasonal migration is observed only in a fraction of individuals in the population (Chapman et al., 2011). Non-migratory strategies include residential (i.e., sedentary), nomadic, and dispersal behaviors (Mueller and Fagan, 2008). Current inference frameworks for partially migrating species require researchers to first define explicit movement strategies exhibited by the species (e.g., Fullman et al., 2021; Poessel et al., 2016). Researchers then classify individual paths as one of the defined strategies using methods based on spatially explicit measures or model selection (Cagnacci et al., 2016). Classification is often followed by descriptive statistics for each movement strategy or interpretation of statistical models formulated for each movement strategy (Fullman et al., 2021).

Golden eagles (*Aquila chrysaetos*) display partial migration (Poessel et al., 2016). Understanding the yearly movement strategies of golden eagles is important for conservation and management of the species. Golden eagles' high mobility, for example, carries individuals across political boundaries, forcing management efforts for the same individuals to be shared by multiple governing bodies (Brown et al., 2017).

Morales et al. (2010) argue for the importance of understanding the links between

Figure 3.1. Year-long movement paths for 3 golden eagles in the western United States, where path color differentiates between individuals.

movement and population dynamics. Population-level inference using a hierarchical structure depends on individual-level models (e.g., Hooten et al., 2016), so it is essential to develop individual-level models that describe behavior well. The task of developing realistic individual-level models becomes more difficult the more heterogeneous the population.

In Figure 3.1, we display year-long paths for three individuals that used three movement strategies, which could be described as residence, migration, and dispersal. We define residence as attraction to a single location throughout the year. We define migration as a path where the individual spends a portion or all of the summer season in a single location and a portion or all of winter in a more southern location. We define dispersal as a path where the individual is attracted to one location for a period starting in the beginning of the year and switches to a new location for the remainder of the year.

Partitioning groups of golden eagles based on movement strategy can be a challenging task due to the presence of 'less-stereotyped' or 'mixed' cases (Cagnacci et al., 2016). Some authors have suggested movement strategies in partially migrating populations would be better described as existing on a continuum, which would better accomodate those less-stereotyped cases (Ball et al., 2001; Cagnacci et al., 2016). Thus there is a need for flexible models that are capable of fitting multiple movement strategies, without

predefining those strategies.

Varying coefficient models that allow behavior to transition smoothly in time have recently received attention in the animal movement literature for being a more flexible and realistic alternative to the latent-state model (Michelot et al., 2020; Russell et al., 2018, 2017). In this work, we describe a single varying coefficient model which utilizes a stochastic differential equation (SDE) framework similar to that of Eisenhauer and Hanks (2020). We fit the varying coefficient model for a variety of movement paths including those displaying residential, migratory, and dispersal behavior. The advantage of this varying coefficient framework is that the same model can easily be used to provide insight into movement behavior that fits one of these three categories, as well as behavior that does not clearly fit into only one of these categories. Our proposed model can produce realistic simulated paths for a range of movement strategies.

We compare our approach to a latent-state model within the same SDE framework, and we show that our varying coefficient model better describes movement behavior. Latent-state models are commonly used in animal movement modeling (Pirotta et al., 2018; Patterson et al., 2017), and there exist popular R packages that can be used to easily fit these types of models for animal telemetry data (Michelot et al., 2016; McClintock and Michelot, 2018). The latent-state models are not as flexible as our varying coefficient model and need to be specified differently depending on the movement strategy. We defined different sets of states for residential, dispersal, and migratory movement strategies.

In Section 3.2, we describe the golden eagle data and motivate the selection of the subset we focus on in this paper. In Section 3.3, we describe the SDE model framework which is common to all models we consider in this paper. In Section 3.4, we present our varying coefficient model. In Sections 3.5-3.7, we fit the varying coefficient and alternative models to three paths we selected to illustrate three stereotypic movement strategies: residence, dispersal, and migration. We then illustrate how the varying coefficient model can be used to fit a wider range of movement behaviors in Section 3.8. Lastly, we summarize the results and suggest areas for future work in Section 3.9.

3.2 Golden Eagle Telemetry Data

We obtained satellite telemetry data for 68 individuals, each of which was tracked for at least 1 year in the western United States. Tagging of eagles and collection of data was funded by the National Raptor Program of the U.S. Fish and Wildlife Service

(USFWS), and we accessed the data through collaboration with the USFWS and one of its contractors, Eagle Environmental, Inc. Movement paths for all individuals, based on hourly GPS locations accurate to within 19 meters, are shown in Figure 3.2. Most of the eagles in this dataset were fledgelings when tagged. The eagles utilized a wide range of habitats in the western United States from desert, semi-arid plains, shrub-steppe, and mountains to arctic tundra. We subdivided individual movement paths by year (Jan 1 – Dec 31) and removed paths with a span of observations shorter than 290 days. We were left with 194 yearly paths for a total of 67 unique individuals. An exploratory analysis of each yearly path using the `migrateR` R package (Spitz et al., 2017) identified 18 dispersers or nomads, 161 migrants or mixed migrants, and 15 residents using net squared displacement. The meaning of these terms is shown graphically in Figure 3.3 (Spitz et al., 2017). Model selection using AIC favors the more complex model, and inspection revealed that many of the paths classified as migration or mixed migration appeared closer to the dispersal, residential, or nomadic strategies. Our varying coefficient modeling approach removes the need to classify each path into only one category (e.g., migrant, resident, disperser) and allows for a more nuanced and realistic representation of bird behavior.

These data display a wide range of movement behaviors, most of which do not clearly belong to a single movement strategy. This may be due the young age of many of the birds and much of the data collection taking place in a desert ecosystem. Thus, we selected datasets from three individuals which were visually identified as clearly displaying one of each of the three movement strategies: migration, residence, and dispersal. We focused this analysis on comparing models fit for these three records.

The latent-state models require regular time intervals, but the data are irregular in time. To resolve this issue, we thinned the data to only one observation per day and linearly interpolated missing intervals that were all shorter than 30 days. We chose to use daily observations because we are interested in movement behaviors that happen over the course of five days at the least. See Appendix S3.1 for details.

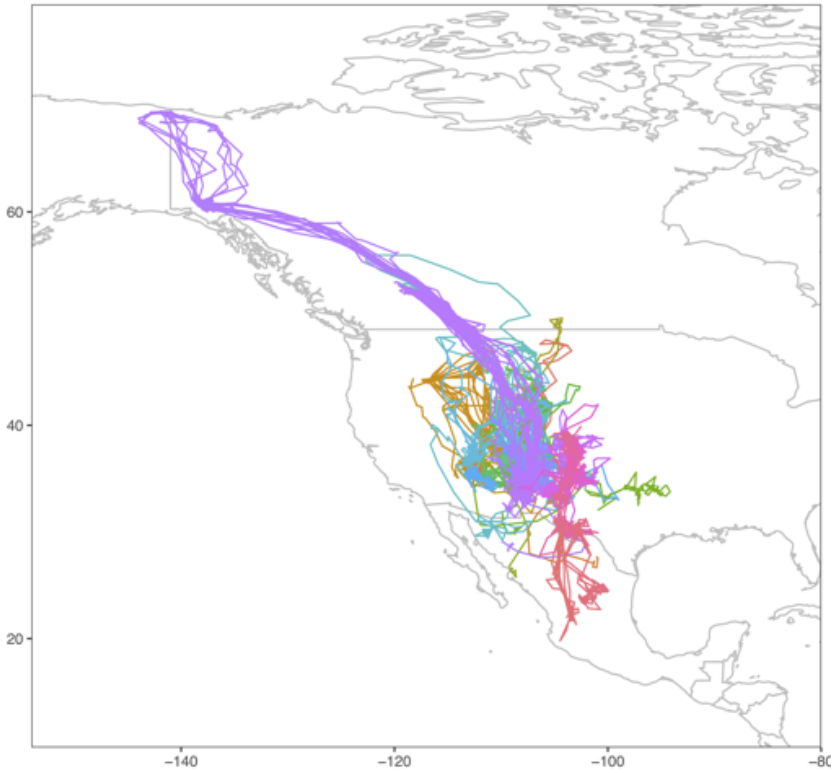

Figure 3.2. Paths for all eagles in the dataset. Color indicates individual bird.

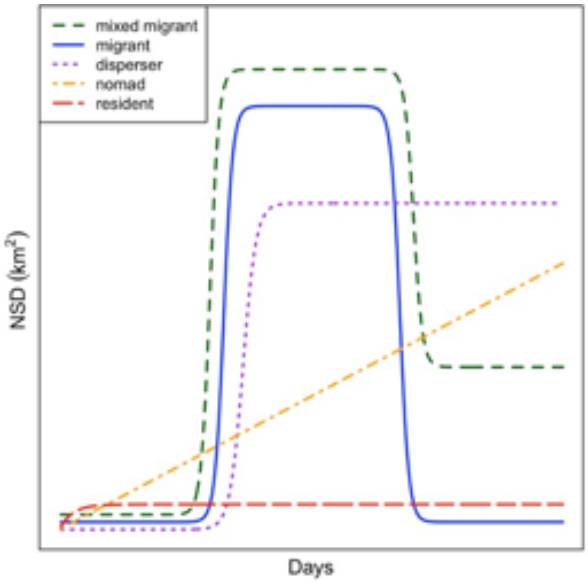

Figure 3.3. Plot replicated with permission from Spitz et al. (2017).

3.3 A Stochastic Differential Equation Model Framework for Animal Movement

We considered a flexible SDE model framework following Russell et al. (2018) and Hanks et al. (2017). We adopt the notation of Eisenhauer and Hanks (2020). The continuous time model for an animal's position $\mathbf{r}_t$ at time t is

$$d\mathbf{r}_t = \mathbf{v}_t dt \tag{3.1}$$

$$d\mathbf{v}_t = -\beta(\mathbf{v}_t - \boldsymbol{\mu}(\mathbf{r}_t))dt + \sigma \mathbf{I} d\mathbf{w}_t \tag{3.2}$$

where $\mathbf{v}_t$ is the velocity of the animal at time t, β is the coefficient of friction (Nelson, 1967) which controls autocorrelation in movement, $\boldsymbol{\mu}(\mathbf{r}_t)$ is the mean drift in the direction of movement, σ is a scalar that controls the variance in the stochastic term, $\mathbf{I}$ is a 2×2 identity matrix, and $\mathbf{w}_t$ is independent Brownian motion in $\mathbb{R}^2$.

We adopt the additional simplification of a constant motility surface and regular time intervals of 1 day. Thus an Euler-Maruyama scheme (Kloeden and Platen, 1992) approximates (3.1)–(3.2) by

$$\mathbf{r}_{t+1} = \mathbf{r}_t + \mathbf{v}_t \tag{3.3}$$

$$\mathbf{v}_{t+1} = \mathbf{v}_t - \beta(\mathbf{v}_t - \boldsymbol{\mu}(\mathbf{r}_t)) + \sigma \mathbf{I} d\mathbf{w}_t \tag{3.4}$$

and substituting (3.3) into (3.4) results in

$$\mathbf{r}_{t+2} - 2\mathbf{r}_{t+1} + \mathbf{r}_t = \beta\left(-\nabla p(\mathbf{r}_t) - \mathbf{r}_{t+1} + \mathbf{r}_t\right) + \sigma \boldsymbol{\epsilon}_t \tag{3.5}$$

where we have modeled the mean drift $\boldsymbol{\mu}(\mathbf{r}_t)$ as the negative gradient of a potential function $p(\mathbf{r}_t)$ (Brillinger et al., 2012; Preisler et al., 2013; Eisenhauer and Hanks, 2020; Russell et al., 2018). We define this potential function $p(\mathbf{r}_t)$ as a weighted sum of distances to m fixed attractors with x-coordinates $a_{x1}, a_{x2}, \ldots, a_{xm}$ and y-coordinates $a_{y1}, a_{y2}, \ldots, a_{ym}$, i.e.,

$$p(\mathbf{r}_t) = \sum_{i=1}^{m} k_{it} \sqrt{(x_t - a_{xi})^2 + (y_t - a_{yi})^2} \tag{3.6}$$

where x_t and y_t are the x- and y-coordinates of $\mathbf{r}_t$ and k_{it} is the coefficient of attraction to the i^{th} attractor. The models considered in this paper all follow this framework but

have varying number of attractors m and coefficients of attraction k_{it} that may or may not change over time t.

3.4 Flexible Model for Partially Migratory Species

The varying coefficient model we considered fixes the number of attractors $m = 8$ and allows k_{it} for $i = 1, 2, \ldots, m$ to change smoothly over time. We chose to use $m = 8$ as an overestimate of the number of attractors, and we used shrinkage methods to effectively select a subset of the attractors (Marra and Wood, 2011). It is clear that at least two attractors are needed for a migrant or disperser model, and the additional attractors might capture stopover sites or other irregular behavior. We chose the eight attractors with k medians clustering of the daily locations with eight clusters, but they could be chosen with any method of the researcher's choice or considered latent variables to be estimated jointly with other parameters. Each coefficient of attraction k_{it} is a weighted sum of cyclic cubic regression spline basis functions. Thus, the potential function becomes

$$p(\mathbf{r}_t) = \sum_{i=1}^{m} \sum_{j=1}^{J} \alpha_{ij} B_j(t) \sqrt{(x_t - a_{xi})^2 + (y_t - a_{yi})^2} \tag{3.7}$$

where α_{ij} is the coefficient of each cyclic cubic basis function $B_j(t)$ for attractor i, and the number of basis functions J is bounded above by 30. We fit the model expressed by plugging the gradient of (3.7) into (3.5) using the `gam` function in the `mgcv` R package. We also penalized the null space of the basis functions to implement selection of the attractor coefficients via regularization (Marra and Wood, 2011).

3.5 Resident Example

We define a resident as an individual attracted to the same location throughout the year. In this section, we compare our varying coefficient model from Section 3.4 to a single-state model formulated specifically for the residential movement strategy. For this comparison, we chose a path consisting of a single year of data for one individual. We visually determined that this path, shown in Figure 3.4A, displayed a residential movement strategy.

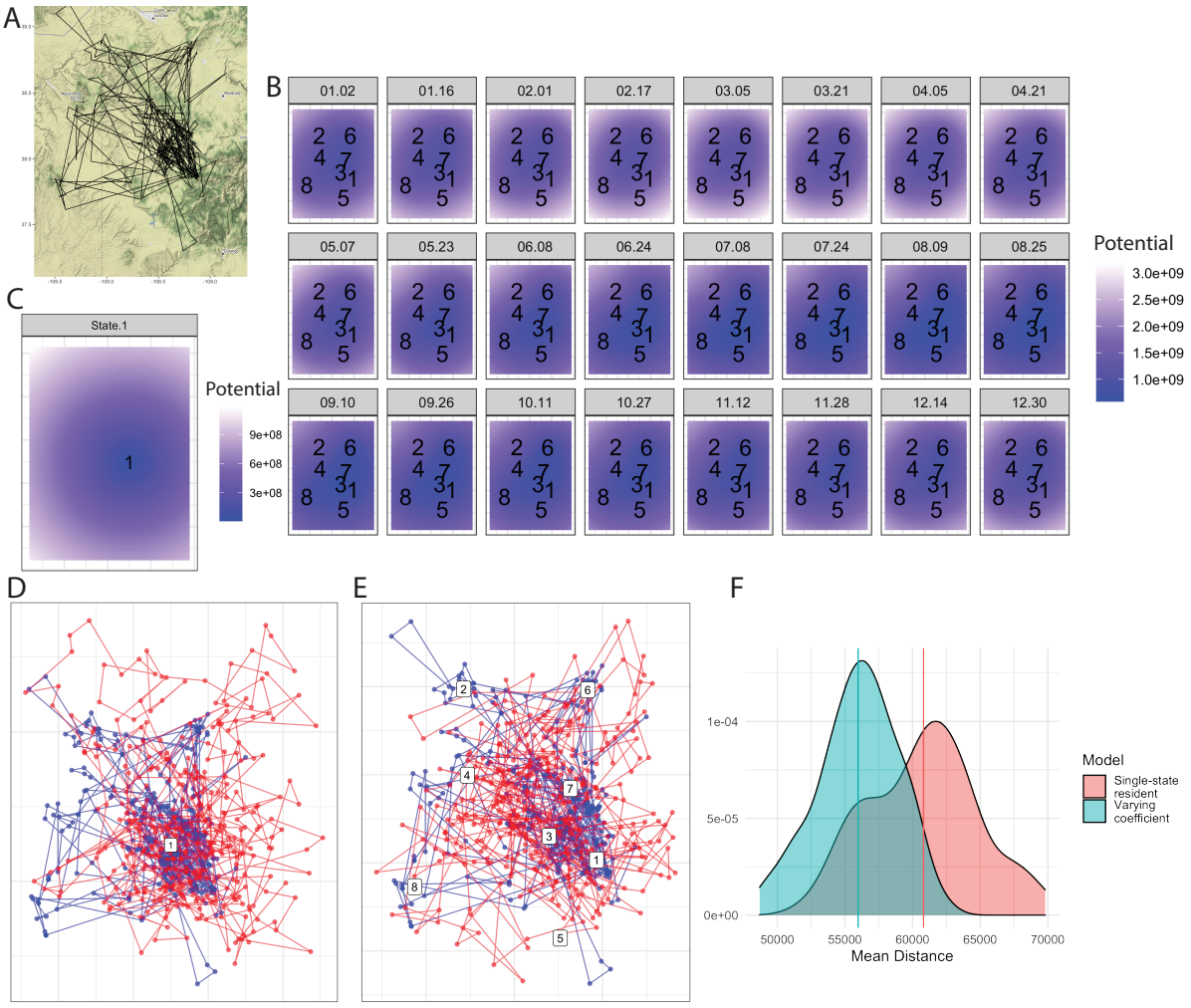

Figure 3.4. (A) Movement path for the resident in 2015. (B) Potential surface for the varying coefficient model. Attractors are numbered. (C) Potential surface for the single-state residence model. The label "1" is located at the single attractor. Bounds are the same as in B. The original path is plotted in blue with the simulation in red from (D) the single-state model and (E) the varying coefficient model. (F) Density plots of average distance from the simulations to the true path for varying coefficient and single-state models. Vertical lines indicate the means for each model.

3.5.1 Single-State Residence Model for Comparison

We formulate a model in the SDE framework which is specific to residential behavior. Since we assume only one type of attraction, e.g., a nest, occurs throughout the year for a resident individual, $m = 1$ and we estimated the single attractor as the median of all data points. The potential function becomes

$$p(\mathbf{r}_t) = k\sqrt{(x_t - a_{x1})^2 + (y_t - a_{y1})^2} \tag{3.8}$$

where a_{x1} and a_{y1} are the x- and y-coordinates of the single attractor and k is the single coefficient of attraction. We restricted k to be positive so that the attractor is forced to have a positive attraction as opposed to a negative attraction or repulsion. The unknown parameters of this simple model are β, k, and σ.

We fit this model in a Bayesian framework with the no-U-turn sampler (NUTS) implemented in Stan for consistency with the latent-state models for the migrant and dispersal, but it was not necessary to fit this model in a Bayesian framework. We also fit the same model using the `lm` function in the `stats` R package and observed similar results. Two Markov chains of length 10,000 were assessed visually to assess convergence. The first 5000 samples were discarded as burn-in, and the last 5000 iterations from the first chain were used for inference.

3.5.2 Resident Results

The attractors for the single-state residence model from Section 3.5.1 and our varying coefficient model from Section 3.4 are labeled on top of daily locations in blue in Figure 3.4D and Figure 3.4E, respectively. For the single-state residence model, the posterior mean of the coefficient of attraction, k, is 6773.0. We used this estimate to construct the potential surface for the single-state model (3.8) in Figure 3.4B. More information about the posterior samples can be found in Appendix S3.2.

For our varying coefficient model described in Section 3.4, model diagnostics tests – e.g., **mgcv:: gam.check( )** in R – did not reveal evidence of a substantive departure from the model assumptions. This includes verification that the imposed maximum of 30 basis functions had been sufficient (i.e., not restrictively low). We estimated $\beta k_{it} = \beta \sum_{j=1}^{J} \alpha_{ij} B_j(t)$ for $i \in \{1, 2, \ldots, 8\}$ in the potential function (3.7). The smooth estimates of βk_{it} over time t are shown in Appendix S3.5 with standard errors. Values of βk_{it} above 0 indicate attractor i has a positive attraction at time t, and values of βk_{it} below 0 indicate attractor i has a negative attraction or a repulsion at time t. We used these estimates to construct a potential surface which changes smoothly over time. A series of snapshots of this potential surface throughout the year is shown in Figure 3.4C. The degree of attraction to different attractors is relatively constant throughout the year for most attractors, with the possible exceptions of attractors 2, 4, and 5.

One way to assess how well a model fits a movement path is to simulate from the model and measure how close the simulation is to the original path. To compare the single-state and varying coefficient model, we simulated from both models fit with the same path 100 times. We simulated 365 days of data, the same length as the original

path. One representative simulation from each model is shown in Figure 3.4D-E. For each model, we calculated the distance between each simulated location and the original path on the same day of the year. Mean distances for each simulation from both models are shown in Figure 3.4F. On average, the varying coefficient model simulations more closely resemble the original path, indicating that our varying coefficient model better describes the movement behavior (Figure 3.4F).

All computing was performed on a laptop computer (2.9 GHz Dual-Core Intel Core i5). The computing time for the varying coefficient model is less than 1 second while the computing time for the single-state model is 27 seconds to run 2 chains for 10,000 iterations.

3.6 Dispersal Example

We define dispersal as an individual that switches from being attracted to one location to being attracted to a second location at some point in the year and remains attracted to the second location for the rest of the year. The path we analyzed as a path displaying dispersal was collected in the year 2018. We compare our varying coefficient model described in Section 3.4 to a latent-state model formulated specifically for dispersal.

3.6.1 Latent-State Dispersal Model for Comparison

We formulated a model in the SDE framework which is specific to dispersal. In this model, we estimated two attractors using k means. To capture the shift from attractor 1 to attractor 2, we used a framework similar to a Hidden Markov Model (HMM) framework consisting of 2 states. We chose to use the term latent-state model instead of HMM following Zucchini et al. (2008) due to the dependence between observations after accounting for the latent states. Thus, the Markov assumption is violated. However, the distinction is relatively unimportant since the methods are functionally almost identical to those used for a HMM, including the forward algorithm and Viterbi algorithm (Zucchini et al., 2017).

In state 1, the individual is attracted to attractor 1, and in state 2, the individual is attracted to attractor 2. Thus we estimated β and σ as in Section 3.5.1 but now we also estimated 2 different values for k, called k_1 and k_2, corresponding to the strength of attraction to attractor 1 while in state 1 and the strength of attraction to attractor 2 while in state 2, respectively. We also estimated the probability of transitioning from

state 1 to state 2 on any particular day. We characterized this probability of transitioning with a vector of length 2 called **g** where the first element is the probability of staying in state 1 and the second is the probability of transitioning. Once in state 2, the probability of transitioning back to state 1 was set equal to 0 to ensure only one transition.

As for the single-state resident model, we fit this model in a Bayesian framework with NUTS implemented in Stan. Two Markov chains of length 10,000 were assessed visually to assess convergence. The first 5000 samples were discarded as burn-in, and the last 5000 iterations from the first chain were used for inference.

3.6.2 Dispersal Results

Attractors chosen with k means for each of the latent-state and varying coefficient model are shown in Figure 3.5D-E along with the true path in blue. The posterior mean estimates from the latent-state dispersal model for k_1 and k_2 are 9228.7 and 7067.3, respectively. We used these values to construct the potential surfaces for the two states in the latent-state model as shown in Figure 3.5B. More information about the posterior samples can be found in Appendix S3.3.

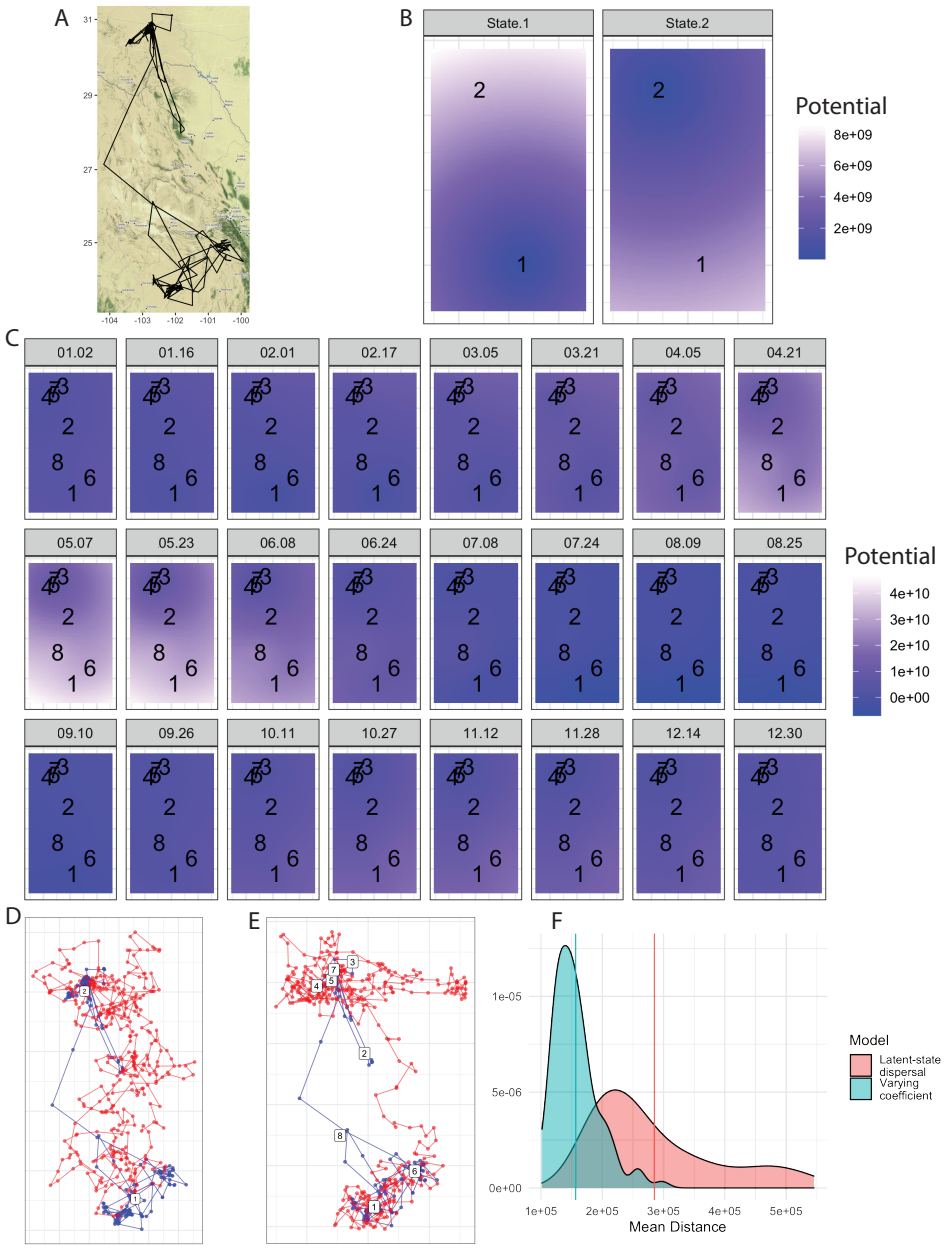

Figure 3.5. (A) Movement path for the individual showing dispersal in 2018. (B) Potential surfaces for the two states in the latent-state model for dispersal. The attractors are identified with the numbers 1 and 2. (C) Potential surface for the varying coefficient model with same bounds as B. Attractors are numbered. The original path is plotted in blue with the simulation in red from (D) the latent-state model and (E) the varying coefficient model. (F) Density plots of average distance from the simulations to the true path for varying coefficient and latent-state models. Vertical lines indicate the means for each model.

We estimated the posterior probability of being in each state on each day using the Viterbi algorithm (see Zucchini et al. (2017) for a description of the algorithm). In Appendix S3.3 the most probable state sequence is depicted using different colors for each state. The state switches when the individual begins dispersal.

For our varying coefficient model described in Section 3.4, model assumptions were satisfied including the maximum of 30 basis functions not being restrictively low. We estimated $\beta k_{it} = \beta \sum_{j=1}^{J} \alpha_{ij} B_j(t)$ for $i \in \{1, 2, \ldots, 8\}$ in the potential function (3.7). The smooth estimates of βk_{it} over time t are shown in Appendix S3.5 with standard errors. The smooth estimates were used to construct a potential surface which changes smoothly over time (see Figure 3.5C). Around the end of April, a shift can be seen which corresponds to the time of the dispersal from one location to another.

We compared 100 365-day-long simulations from the varying coefficient and the latent-state dispersal model. One representative simulation from each model is shown in Figure 3.5D-E. For both models, we calculated the distance between each simulated location and the original path on the same day of the year. Mean distances for each simulation from both models are shown in Figure 3.5F. As was the case for the resident strategy, the varying coefficient model simulations more closely resemble the original path on average (Figure 3.5F).

All computing was performed on a laptop computer (2.9 GHz Dual-Core Intel Core i5). The computing time for the varying coefficient model is less than 1 second while the computing time for the latent-state dispersal model is 110 seconds to run 2 chains for 10,000 iterations each. Convergence of the chains was assessed visually (see Appendix S3.3).

3.7 Migrant Example

We defined a migrant as an individual who switches seasonally from being attracted to a southern location to a northern location and back to the original southern location throughout the year. The path we analyzed was collected from an eagle in the year 2012. We compare our varying coefficient model described in Section 3.4 to a latent-state model formulated specifically for migration.

3.7.1 Latent-State Migrant Model for Comparison

We formulated a model in the SDE framework which is specific to migration. To fit this model, we began by estimating two attractors using k means. We used a latent-state model framework, this time with five states which are analogous to northern migration, northern residence, southern migration, and two states for southern residence. States 1 (northern migration) and 2 (northern residence) both have an attraction to attractor 2, which is the northern-most attractor, but the coefficient of attraction k in (3.8) for state 1, which we call k_1, was restricted to be larger than the k for state 2, which we call k_2. This means the attraction is stronger in state 1 than in state 2. We defined states 3 (southern migration) and 4 (southern residence) similarly with the same southern attractor and restricting $k_3 > k_4$. State 5 is identical to state 4 in all ways except for the transition probabilities. Thus, the coefficient of attraction for state 5 is $k_5 = k_4$.

Since the year starts in January, we assumed the path to begin in a state of attraction to the southern residence, which we call state 4. We then allowed a positive transition out of state 4 so the individual may transition into northern migration. We also assumed the individual ends the year, in December, attracted to the southern residence again. We restricted the probabilities to only allow one cycle through the states according to a prescribed sequence. This restriction prevents the use of these states to characterize brief excursions by quickly cycling through the states repeatedly. In order to specify this rule, we created the fifth state, state 5, which is identical to state 4 but which cannot be transitioned out of. Thus the required order of transitions is 4 to 1 to 2 to 3 to 5 (i.e., southern residence starting on Jan 1, migration, northern residence, migration, southern residence ending on Dec 31).

Instead of directly estimating the transition probabilities, we allowed the transition probabilities to change over time by letting them be functions of a covariate, daily change in daylight length. To illustrate why we chose to use this covariate, we've plotted the latitude of the daily observations and the covariate, which both change in time, in Appendix S3.4. It appears that the covariate should be high during transitions from states 4 to 1 and from states 1 to 2. Similarly, the covariate is low during transitions from states 2 to 3 and from states 3 to 5. Thus, we intended to model this relationship.

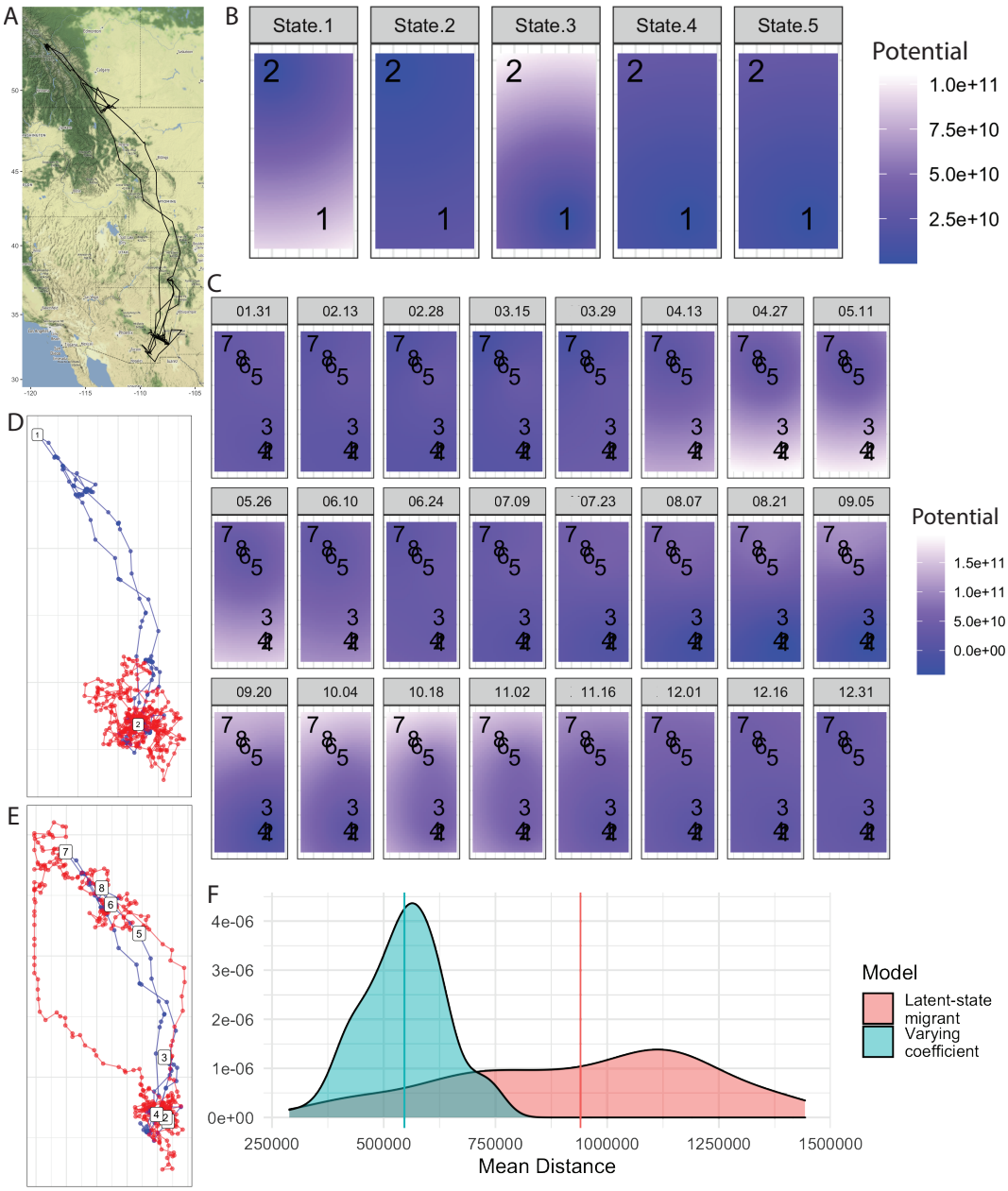

Figure 3.6. (A) Movement path for the migrant in 2012. (B) Potential surfaces for the five states in the latent-state model for the migratory strategy. The attractors are identified with the numbers. (C) Potential surface for the varying coefficient model with same bounds as B. The original path is plotted in blue with the simulation in red from (D) the latent-state model and (E) the varying coefficient model. (F) Density plots of average distance from the simulations to the true path for varying coefficient and latent-state models. Vertical lines indicate the means for each model.

3.7.2 Migrant Results

We compared the latent-state migrant model to our varying coefficient model described in Section 3.4. Attractors chosen with k means for each of models are shown in Figure 3.6D-E on top of true paths in blue. For the latent-state migrant model, the posterior means of $k_1, k_2, \ldots, k_5$ are as follows: $\hat{k}_1 = 33826.8$, $\hat{k}_2 = 10022.8$, $\hat{k}_3 = 39865.8$, and $\hat{k}_4 = \hat{k}_5 = 12950.6$. We used these values to construct the potential surfaces for the five states in the latent-state model as shown in Figure 3.6B. More information about the posterior samples can be found in Appendix S3.4.

We estimated the posterior probability of being in each state on each day using the Viterbi algorithm. In Appendix S3.4 the most likely state sequence is depicted using different colors for each state. The latent-state migrant model is reasonably subdividing the path (see Figure 3.19 in Appendix S3.4).

For our varying coefficient model described in Section 3.4, model assumptions were satisfied including the maximum of 30 basis functions not being restrictively low. We estimated $\beta k_{it} = \beta \sum_{j=1}^{J} \alpha_{ij} B_j(t)$ for $i \in \{1, 2, 3, 4\}$ in the potential function (3.7). The smooth estimates of βk_{it} over time t are shown in Appendix S3.5 with standard errors. The smooth estimates were used to construct a potential surface which changes smoothly over time (see Figure 3.6C). The snapshots of the potential surface throughout the year indicate that southern migration occurred around April and May, and northern Migration occurred in August and September.

We compared 100 337-day-long simulations from the varying coefficient and the latent-state migrant model. One representative simulation from each model is shown in Figure 3.6D-E. For both models, we calculated the distance between each simulated location and the original path on the same day of the year. Mean distances for simulations from both models are shown in Figure 3.6F. As was the case for the resident and dispersal strategies, the varying coefficient model simulations more closely resemble the original path on average (Figure 3.6F).

All computing was performed on a laptop computer (2.9 GHz Dual-Core Intel Core i5). The computing time for the varying coefficient model was less than 1 second, while the computing time for the latent-state migrant model was >10 hours to run 2 chains for 5000 iterations each. Convergence of the chains was assessed visually (see Appendix S3.4).

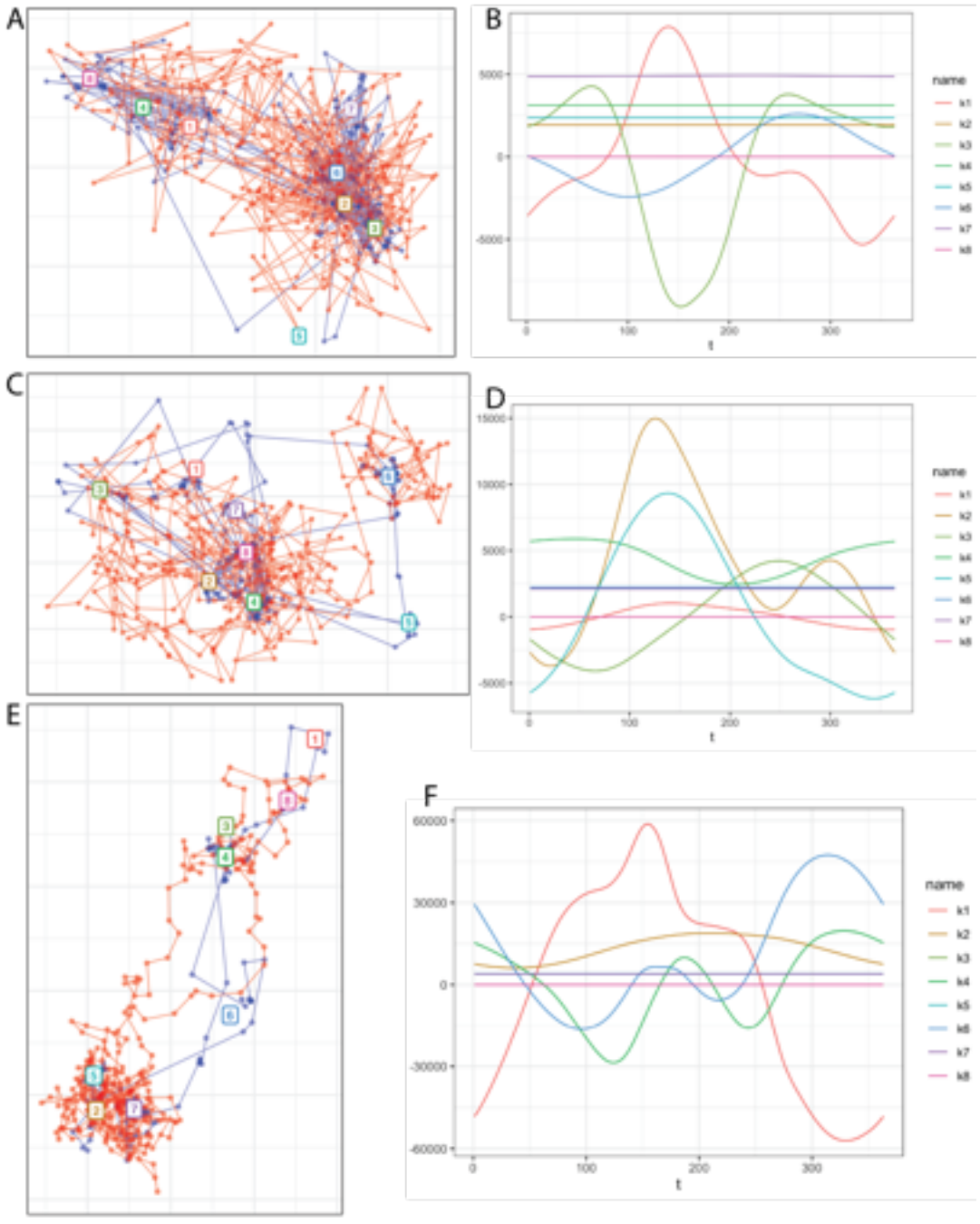

Figure 3.7. In (A), (C), and (E), the original paths are plotted in blue with the simulations from the varying coefficient model plotted in red. The attractors are labeled. In (B), (D), and (F), the varying coefficients corresponding to the attractors in (A), (C), and (E), respectively, are shown over time. For example, k1 in (B) is the varying coefficient corresponding to attractor 1 in (A).

3.8 Fitting Boundary Individuals

An important benefit of the flexible movement model described in Section 3.4 is its ability to fit a wide range of movement behaviors, including those that do not clearly fit the migrant, resident, or dispersal stereotypes. There are many such individuals in the golden eagle dataset since they are a partially migrating species. Three examples of less-stereotyped paths are shown in Figure 3.7 along with simulations from the varying coefficient model fit to each path and the varying coefficients. Simulations from the varying coefficient model are reasonable even for more irregular movement behavior such as these examples.

3.9 Discussion and Future Work

We have described a flexible model using varying coefficients for fitting individual movement paths for a partially migrating species. We compared our varying coefficient model described in Section 3.4 to latent-state models within the same SDE model framework for three individual golden eagles. For these three individuals displaying migration, residence, and dispersal, simulations from our varying coefficient model more closely resembled the true paths. We also illustrated the ability of our varying coefficient model to fit boundary individuals which do not clearly exhibit migration, residence, or dispersal. The latent-state model is also restricted by needing discrete time steps. In our examples, we fit the varying coefficient model to paths with observations at discrete daily time steps for the sake of model comparison, but our varying coefficient model could be fit to irregular time steps as well.

Ecologists could use our flexible modeling framework described in this paper to better understand movement behavior in less-stereotyped individuals. Ultimately, however, it is important to make inference at the population level to inform conservation efforts. We have taken the first step toward population level inference by defining a model that is flexible enough to fit all individuals in the population including individuals that do not easily conform to pre-specified movement strategies. The next step is a structured population level model, which would require a model for the locations of attractors and could incorporate covariates. One could also build a model using individual random effects with population level means to make population level inferences.

Attractive points were fixed in our model, but the attractors could be modeled at the population level by implementing clustering approaches or spatial point process

models. Clustering approaches such as k means clustering of all attractors across all individuals could identify popular attractive regions. Incorporating covariates in the model is important to ecologists who want to understand why movement decisions are made by individuals. Covariates could be included in a model for the attractors by utilizing a spatial point process model for the attractors (Warton and Shepherd, 2010). To model speed as a function of spatial covariates, a motility function could be included as in Russell et al. (2018) but allowing the motility function to depend on covariates.

While latent-state models are more popular in the animal movement literature, our work shows that our varying coefficient model could help researchers better understand the nuances inherent in animal telemetry data. While the latent-state models must be tailored to each movement strategy, our varying coefficient model does not need such adjustments. The implementation of a general SDE framework using potential functions with varying coefficients controlling the degree of attraction or repulsion to given attractors allows for a reasonable model fit for many movement strategies.

S3 Supplement to A Flexible Movement Model for Partially Migrating Species

S3.1 Justification for Daily Observations

We chose to use daily observations in our analysis since the patterns we were interested in modeling (e.g., migration, dispersal) took place over several days at the least. We did not use a finer resolution because of the inconsistent resolution of the original data and since we were not interested in more fine scale behaviors. The choice of sampling resolution is an important one and can be explored by fitting the same model to the data at different resolutions.

We fit the varying coefficient model for the three paths described in Sections 3.5-3.7 at two different resolutions: daily and every 2 days. Resulting potential surfaces are shown in Figures 3.8-3.10. In each case, attractors chosen with k means were located in very similar locations at both resolutions. The potential surfaces are centered at zero to facilitate a fairer comparison. For the resident and migrant, potential surfaces are very similar when we use daily observations or observations every 2 days (Figures 3.8-3.9). For dispersal, the move from one attractor to the other occurs over about 5 days, so we would lose that event by using one observation every 2 days (Figure 3.10).

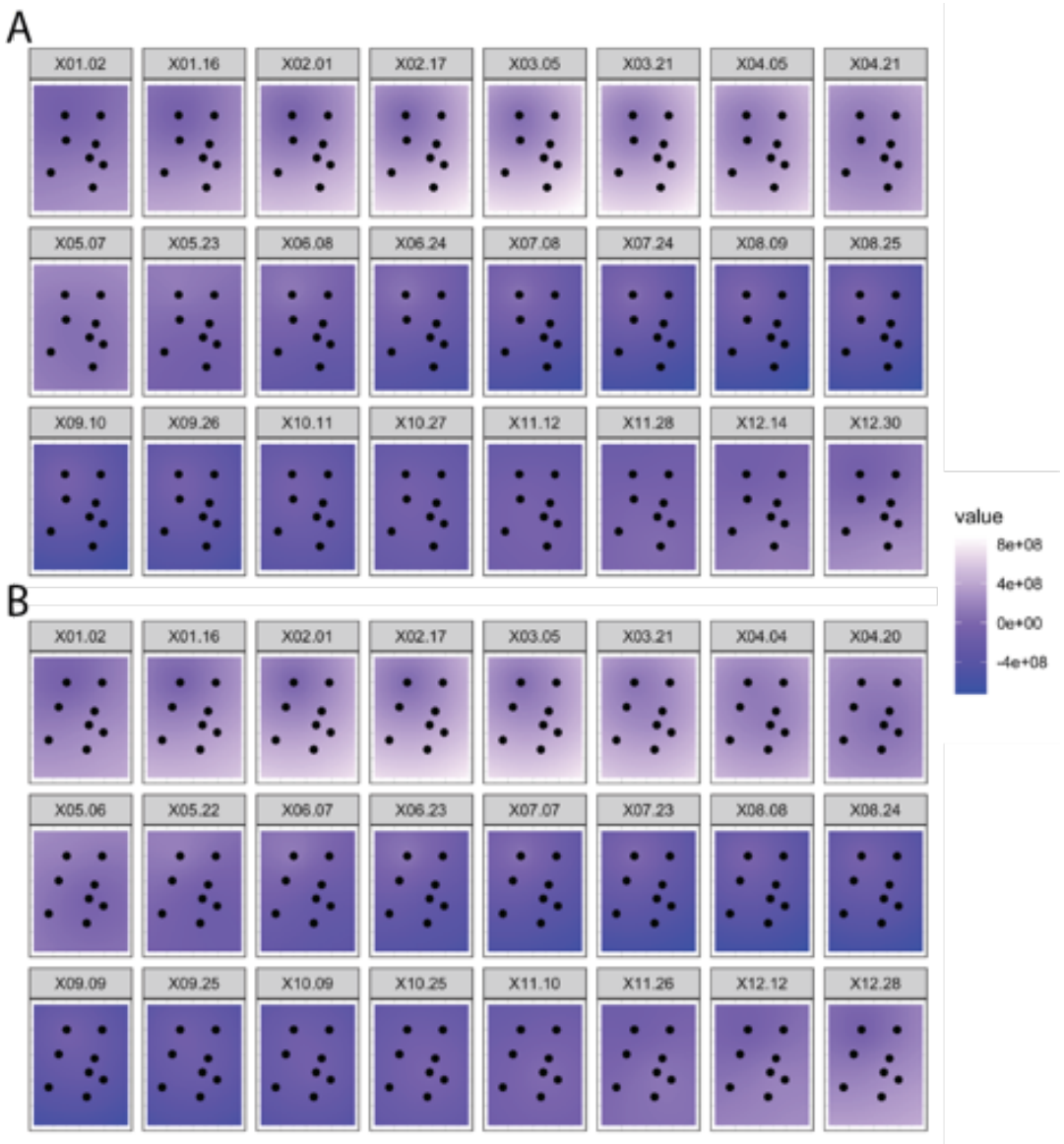

Figure 3.8. Potential surfaces for the varying coefficient model for the resident with (A) daily observations and (B) observations every 2 days. Black dots indicate attractors.

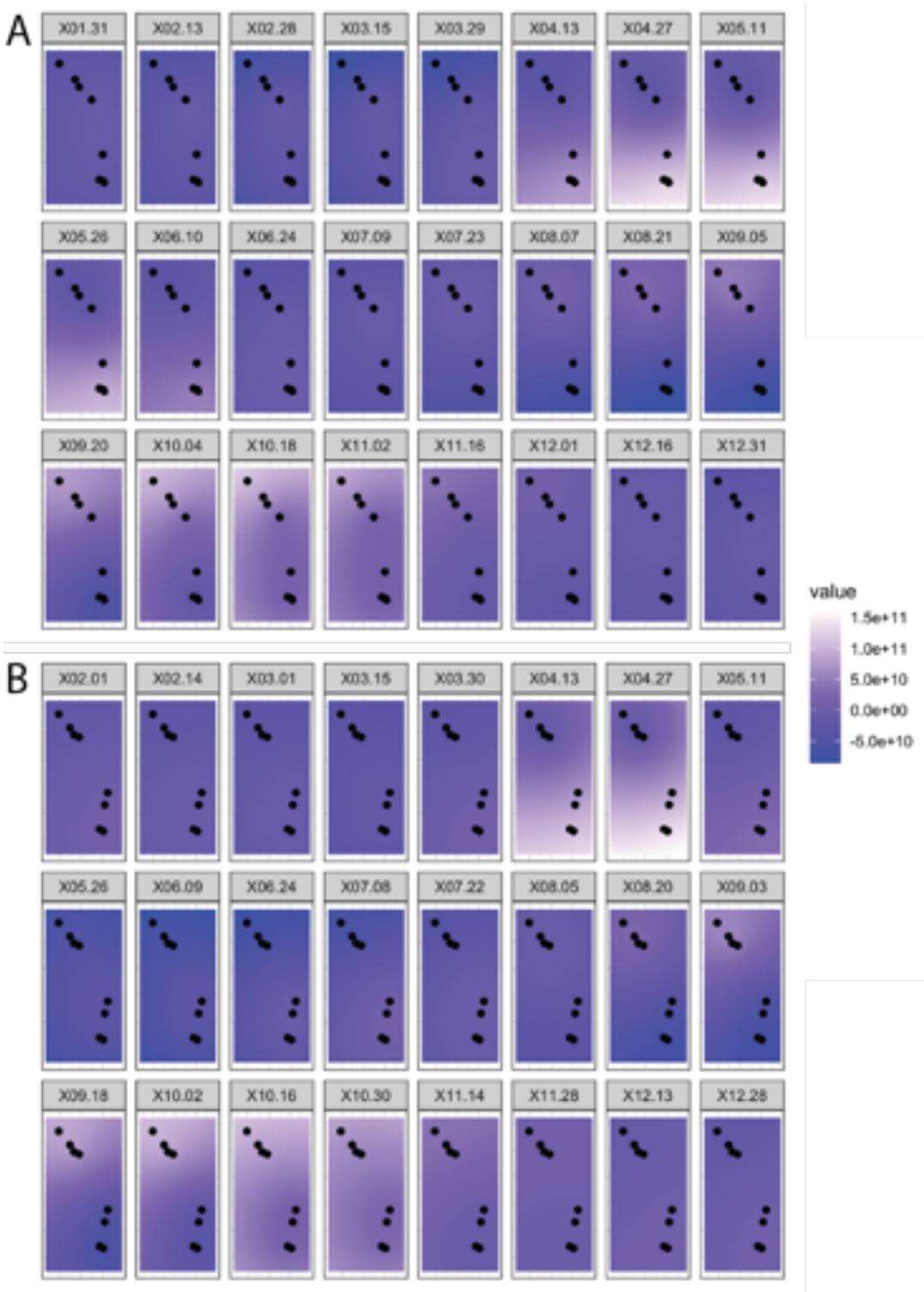

Figure 3.9. Potential surfaces for the varying coefficient model for the migratory path with (A) daily observations and (B) observations every 2 days. Black dots indicate attractors.

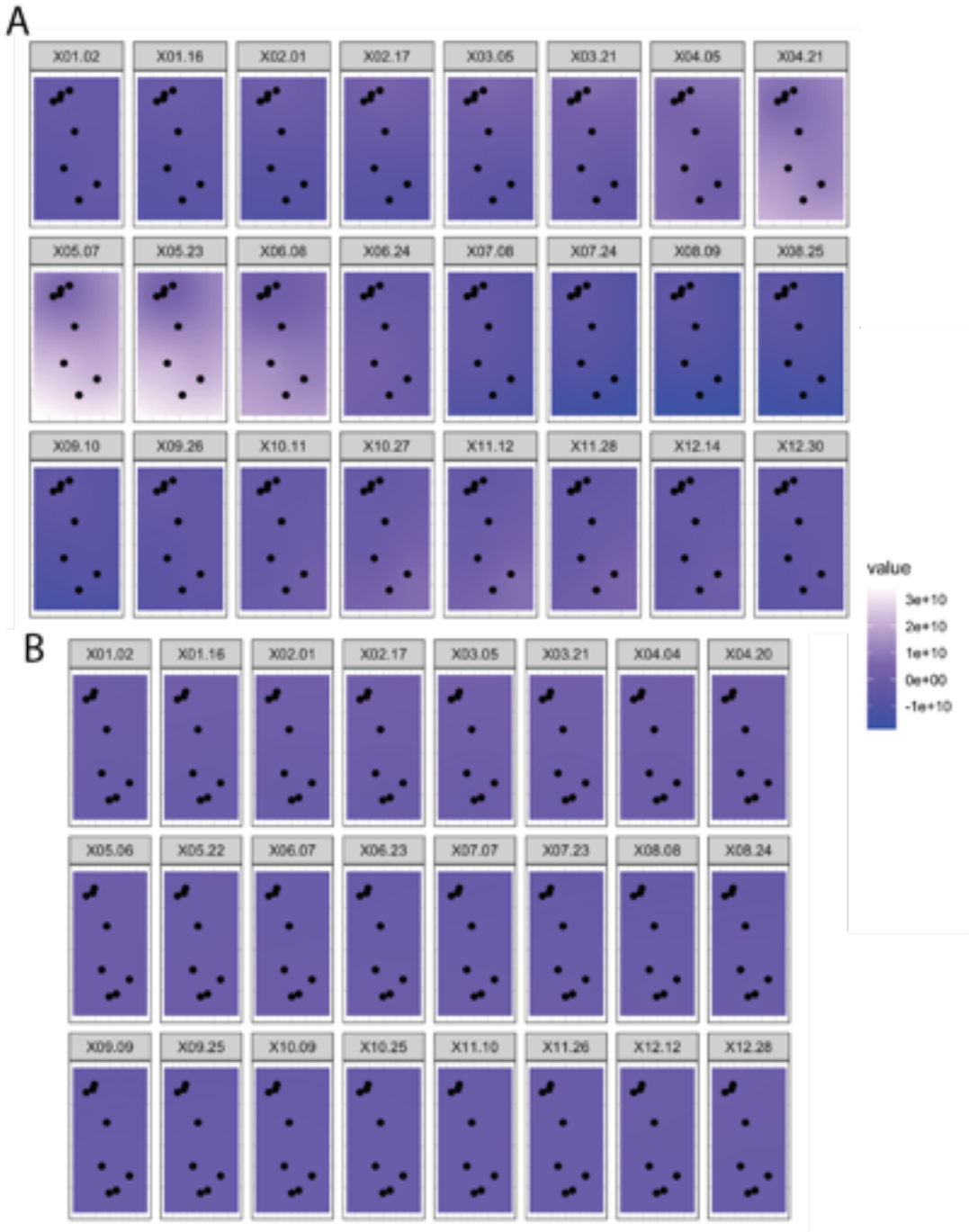

Figure 3.10. Potential surfaces for the varying coefficient model for the path displaying dispersal with (A) daily observations and (B) observations every 2 days. Black dots indicate attractors.

S3.2 Additional Plots for Single-State Resident Model

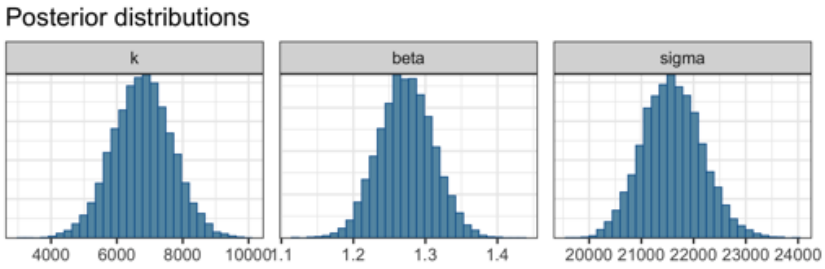

Figure 3.11. Histograms of MCMC samples from the marginal posterior distributions.

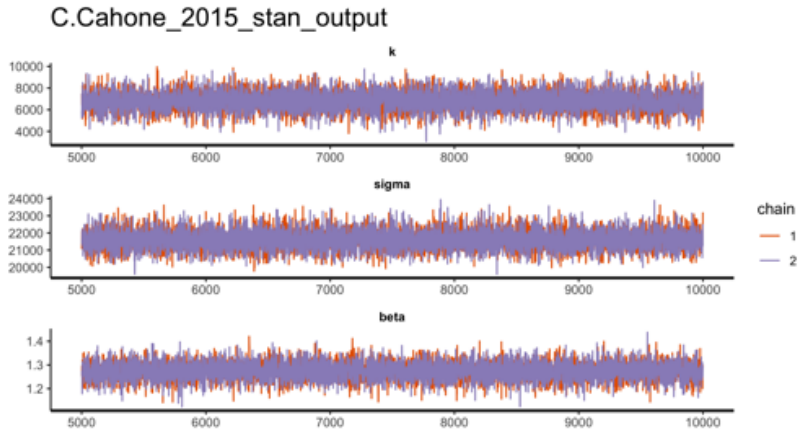

Figure 3.12. Traceplots of MCMC samples from the marginal posterior distributions.

S3.3 Additional Plots for Latent-State Dispersal Model

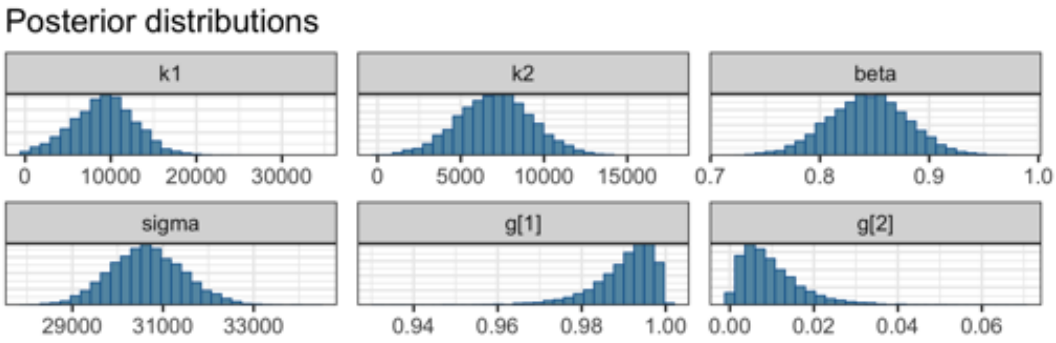

Figure 3.13. Histograms of MCMC samples from the marginal posterior distributions for the latent-state model for dispersal.

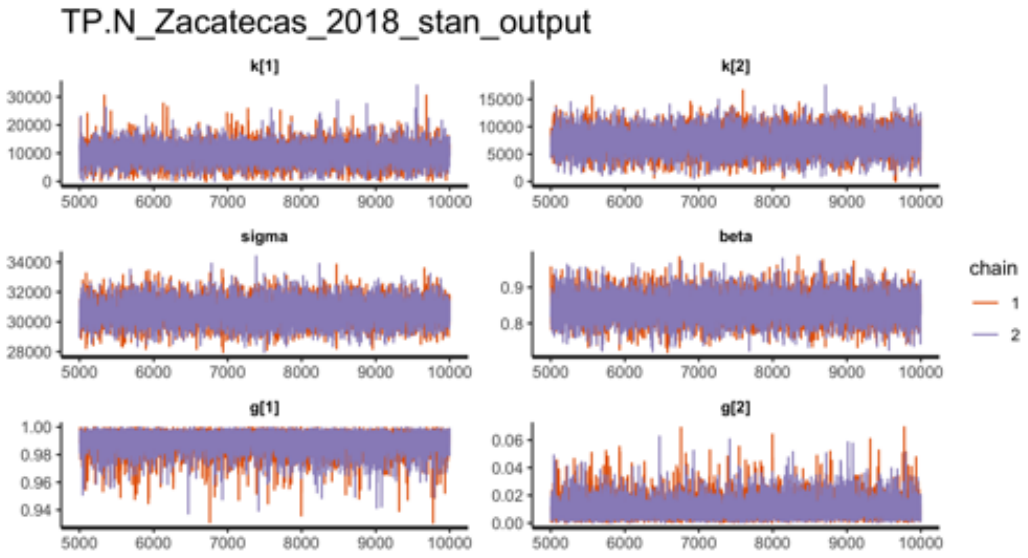

Figure 3.14. Traceplots of MCMC samples from the marginal posterior distributions for the latent-state model for dispersal.

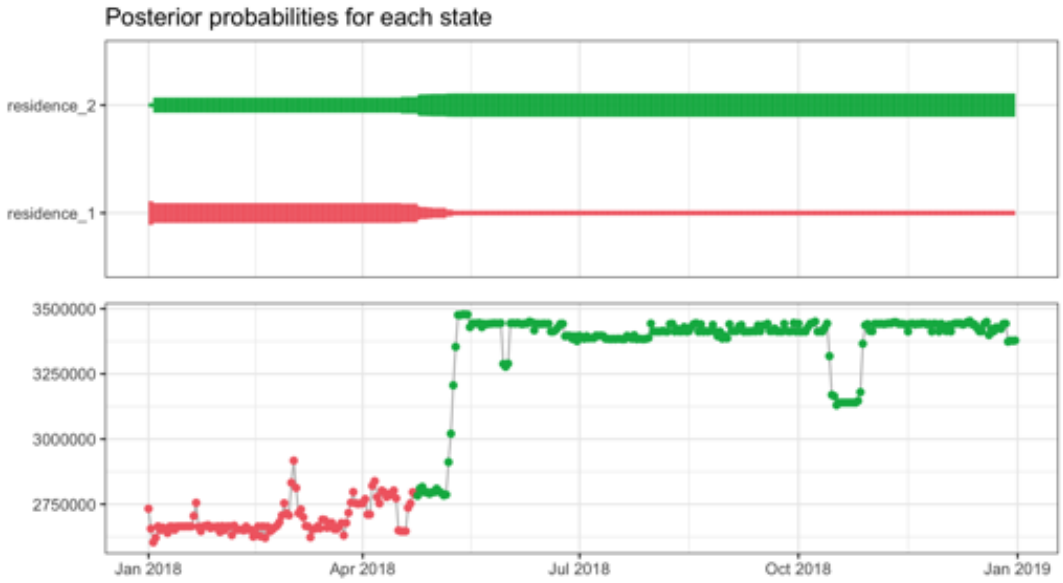

Figure 3.15. States for the dispersal model. In the top plot, the line width is proportional to the posterior probabilities of being in each state across time. In the bottom plot, the y axis is the y-coordinate of location (in meters). The observations are colored by most likely state.

S3.4 Additional Plots for Latent-State Migrant Model

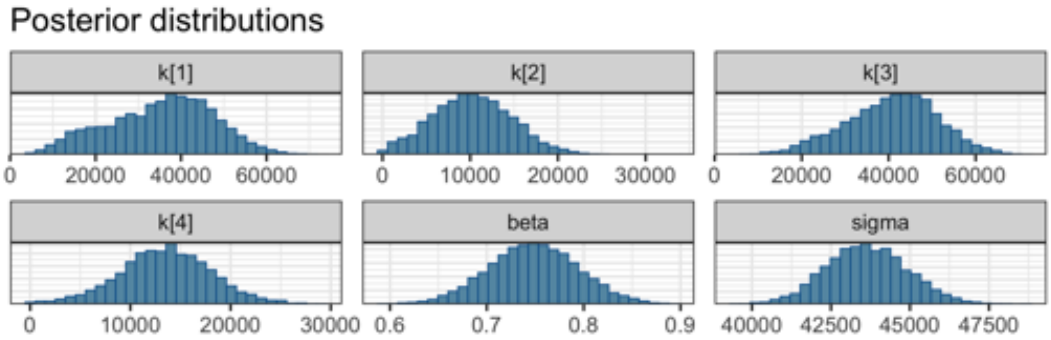

Figure 3.16. Histograms of MCMC samples from the marginal posterior distributions for the latent-state model for the migratory strategy.

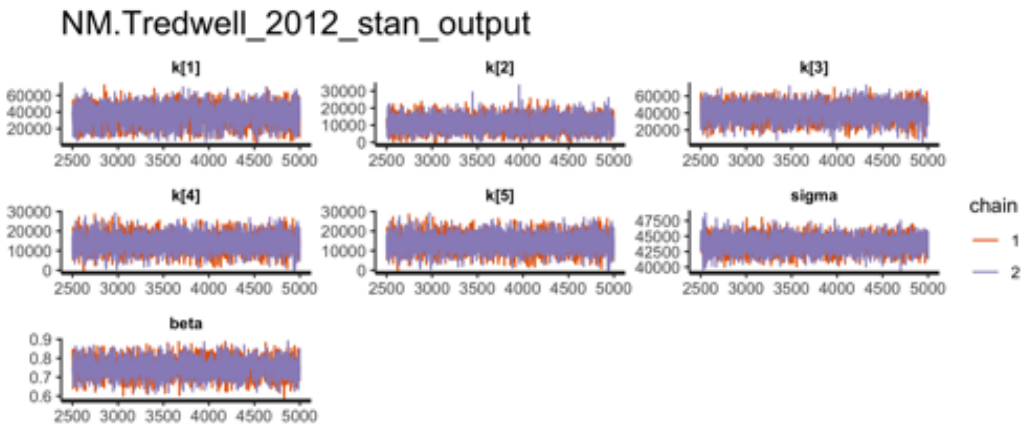

Figure 3.17. Traceplots of MCMC samples from the marginal posterior distributions for the latent-state model for the migratory strategy.

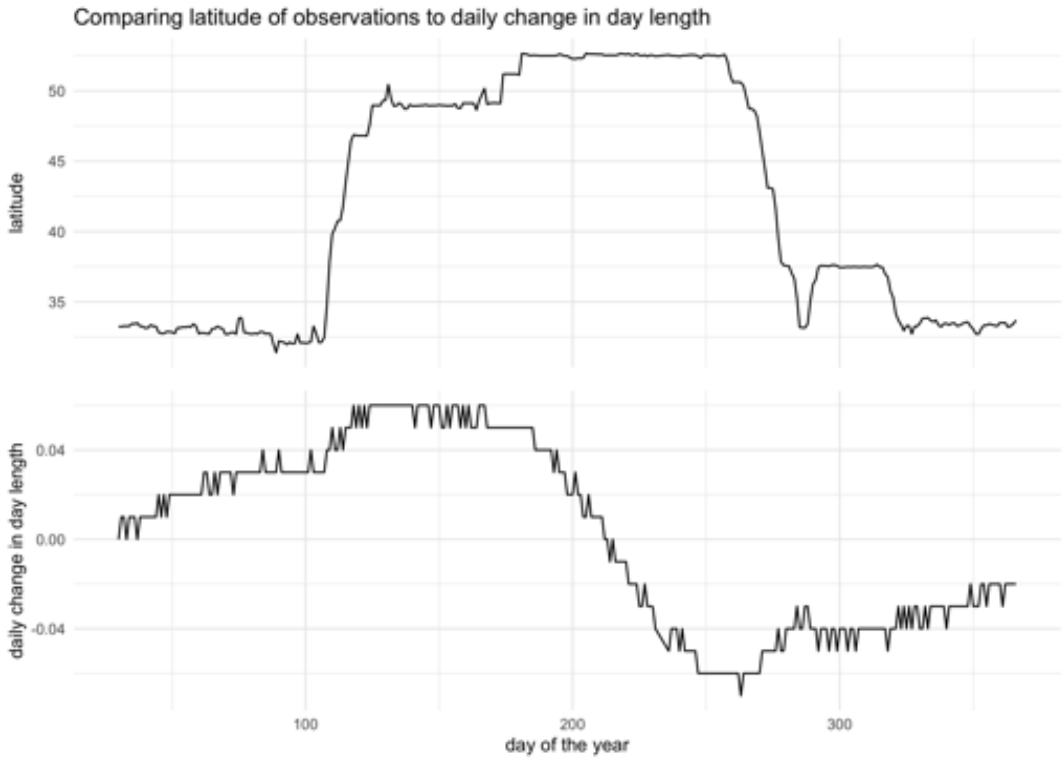

Figure 3.18. On the top, we see the latitude of the individual changing over time, and on the bottom, we see the covariate, daily change in day length, changing over time. All data is for NM Tredwell's location in 2012.

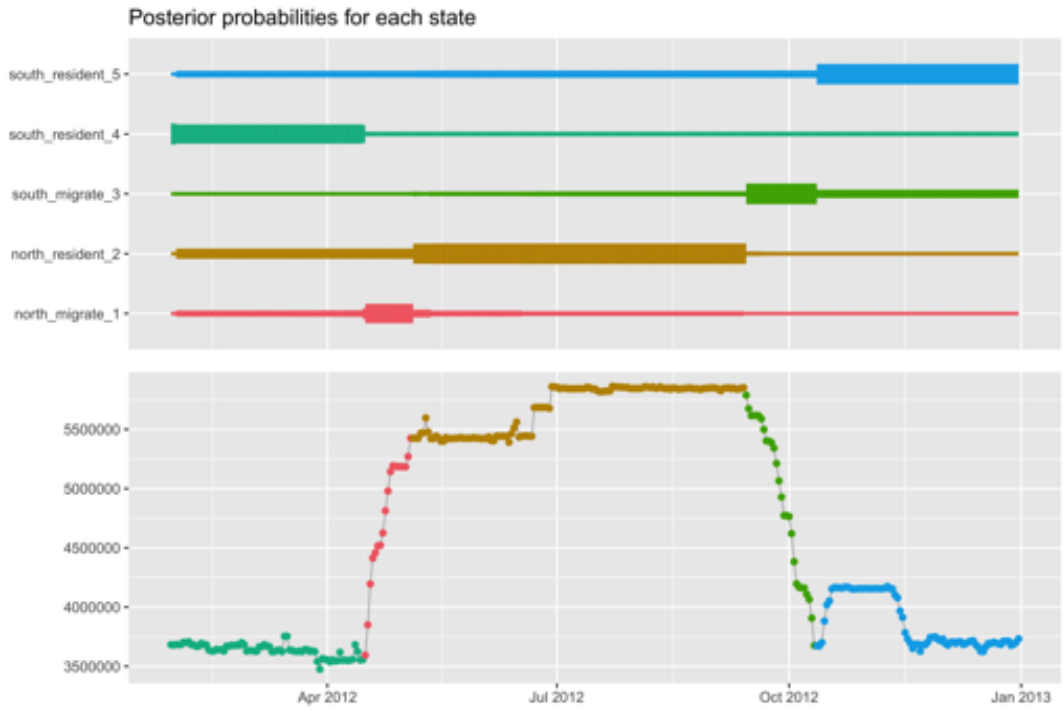

Figure 3.19. States for the migrant model. In the top plot, the line width is proportional to the posterior probabilities of being in each state across time. In the bottom plot, the y axis is the y-coordinate of location. The observations are colored by most likely state.

S3.5 Varying Coefficients

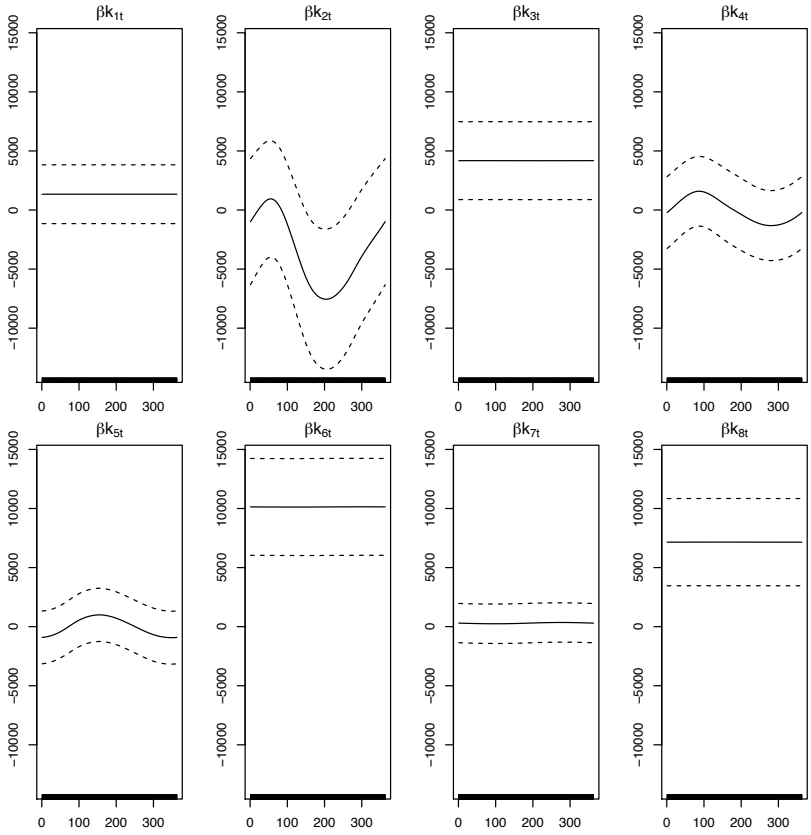

Figure 3.20. The smooth estimates of βk_{it} where k_{it} is the coefficient of attraction to attractor i at time t are plotted as solid black lines. Dotted lines are drawn two standard errors above and below the estimate of the smooth. The x-axis is the day of the year.

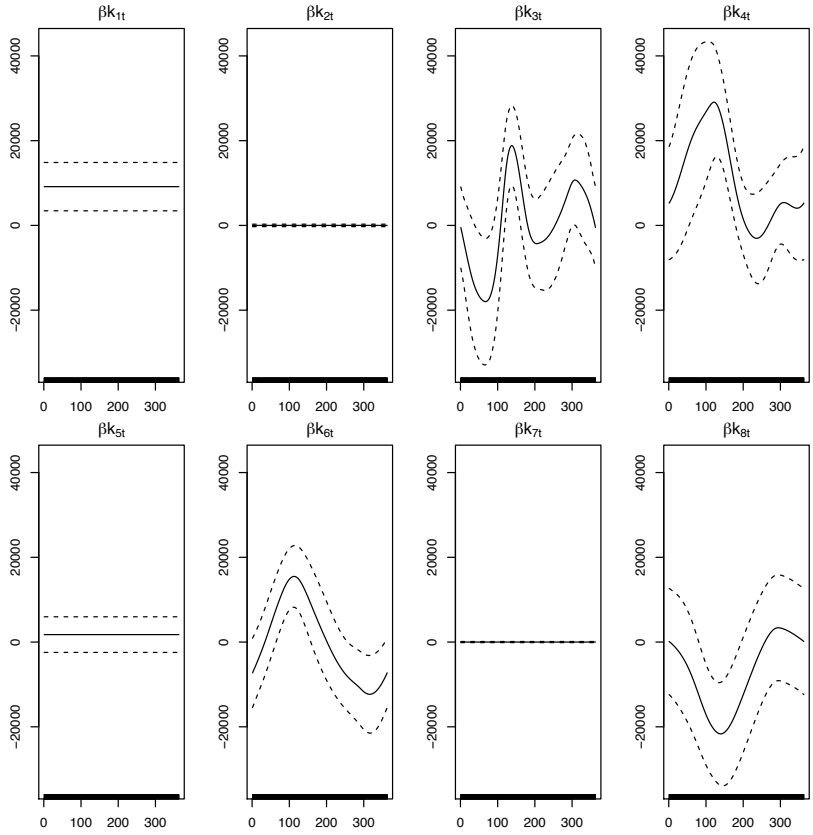

Figure 3.21. As in Figure 3.20, smooth estimates are plotted with dotted lines two standard errors above and below.

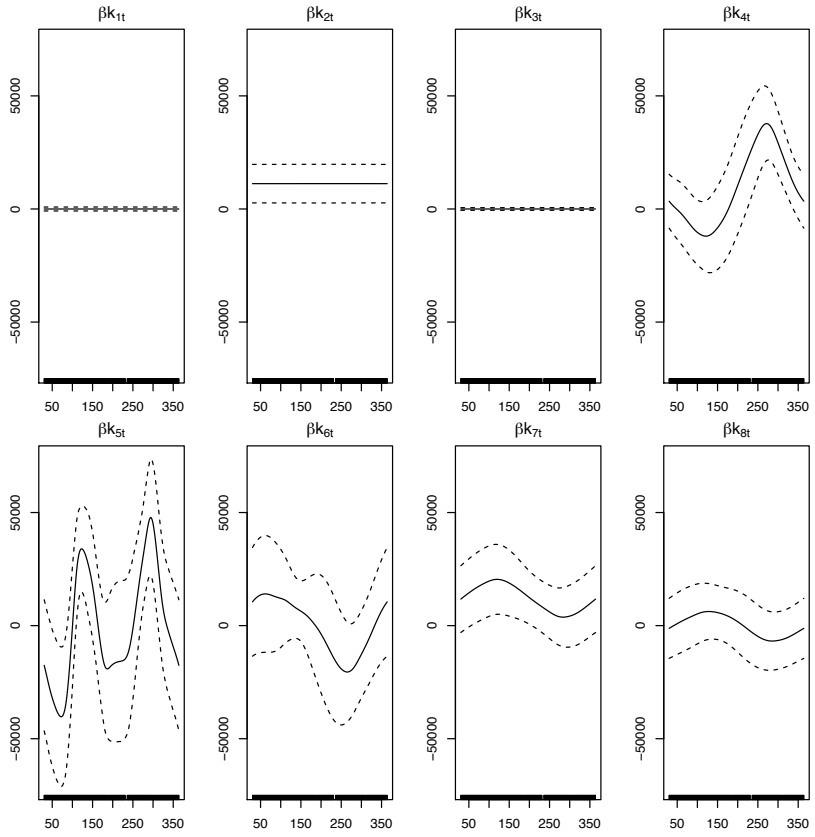

Figure 3.22. As in Figure 3.20, smooth estimates are plotted with dotted lines two standard errors above and below.

Chapter 4
Survey of Probability Attitudes

4.1 Introduction

"People forget what they do not use. But attitudes 'stick"' (Ramirez et al., 2012). The American Statistical Association consistently cites probability as a necessary topic in undergraduate statistics programs (Horton et al., 2014; American Statistical Association, 2000), and Pearl et al. (2012) argue for the importance of developing instruments which measure a variety of affective constructs within the context of statistics curricula. Few peer-reviewed studies have examined probability attitudes, so we appeal to existing work on attitudes toward statistics.

The first tools for assessing students' attitudes toward statistics were developed and frequently used by researchers in the 1980s (Roberts and Bilderback, 1980; Wise, 1985). In total, at least 15 surveys purporting to assess students' attitudes toward statistics have been developed (Ramirez et al., 2012), one of the most popular of which is the *Survey of Attitudes Toward Statistics-36* (SATS-36; Schau, 2020). The SATS-36 has been an important tool in the effort to improve statistics courses (Chance et al., 2016; Gundlach et al., 2015). While the traditional probability/inference sequence is a core component of the statistics curriculum and the American Statistical Association and others have suggested meaningful changes to these courses, there has been relatively little effort to systematically improve these courses in recent years (Horton et al., 2014).

We have only found one tool for measuring students' attitudes toward probability, the *Probability Attitude Inventory* (PAI) (Tan et al., 2011). The PAI was adapted from the *Mathematics and Science Attitude Inventory* (MSAI) under the "Project EDGE" of Rochester Institute of Technology for the purpose of assessing the impact of the graphing calculator on students' attitudes toward probability (Tan et al., 2011; Paciorek, 1997). The only adaptation was changing the word "mathematics" to "probability" and

removing all questions relating to science.

Because of the compelling record of careful development and validity evidence for the SATS-36 and the utility of the resulting component scores, we decided to study whether the tool could be adapted, extended, and validated for studying attitudes toward probability. We changed words related to statistics to words related to probability, and we added open-ended questions. We administered a preliminary version of our survey to 15 undergraduate probability course sections in the fall semester of 2020. After updating the instrument, we administered the final version of the *Survey of Probability Attitudes* (SPA) to 13 undergraduate probability course sections and 7 mathematics or biostatistics course sections at Penn State in the spring semester of 2021.

We describe the SATS-36 in Section 4.2. In Section 4.3, we describe the development, distribution, and scoring of the SPA along with the validity and analysis methods. In Section 4.4, we describe the resulting dataset, evidence of the validity of the instrument, and analyses of the gain in mean attitude component scores between pre and post SPA. Lastly, in Section 4.5, we summarize our work thus far and describe plans for future work.

4.2 Survey of Attitudes Toward Statistics-36

The SATS-36 is an updated version of the *Survey of Attitudes Toward Statistics-28* (SATS-28), which added eight new items measuring two additional components (Schau et al., 1995). The resulting SATS-36 assesses six components: *Affect, Cognitive Competence, Value, Difficulty, Interest,* and *Effort*. The process of developing the two surveys included initial examination of existing tools and students' open-ended statements, organization of a focus group, pilot testing and revision, validation of the four-component structure using Confirmatory Factor Analysis and based on its relationship to other measures, and validation of the final six-component structure using Confirmatory Factor Analysis (Ramirez et al., 2012). However, some improvements to the SATS-36 have been suggested, including changing the number of components and removing some items that perform poorly (Vanhoof et al., 2011).

The SATS-36 has been used by researchers to explore questions such as how students' attitudes relate to teaching methods (Bateiha et al., 2020; Budé et al., 2012; DeVaney, 2010) and whether attitudes differ between different student populations (Griffith et al., 2012). The SATS surveys have also been adapted for students who speak languages other than English (Sarikaya et al., 2018; Khavenson et al., 2012; Vanhoof et al., 2011) and,

importantly for our purposes, to assess attitudes toward subjects other than statistics (Wisecup, 2017; Cole, 2010; Tempelaar and Nijhuis, 2007).

The 36 survey questions comprising the six components on the SATS-36 are Likert-type questions with responses ranging from 1 to 7, where 1 is "strongly disagree", 4 is "neither disagree or agree", and 7 is "strongly agree". There are also some demographic questions and questions related to mathematics experience. The pre-survey version of the SATS-36 along with the scoring guidelines can be found at `www.evaluationandstatistics.com`.

4.3 Methods

The SPA was developed and revised in 2020 and 2021, and the final version was distributed to students enrolled in courses at Pennsylvania State University with the help of their instructors in the spring of 2021. The development process and final survey data provided evidence of the validity of the instrument. We also analyzed the pre- and post-survey scores using mixed effects models and t-tests.

4.3.1 Development of the Survey of Probability Attitudes

In our first version of the SPA administered in the fall of 2020, we adapted the SATS-36 by updating the demographic questions to be more specific to our audience, replacing words related to statistics with words related to probability, and adding six open-ended questions to the end of the survey. One item asking whether the student planned to attend all class sessions was removed due to its ambiguous meaning for asynchronous courses. Prior to sending the survey to students, we assessed the validity of the survey using think-alouds. We recruited three students who had recently taken one of the probability courses we would be surveying. The students were incentivized by a gift card for 10 US dollars. A graduate student who was not the students' previous instructor nor one of the researchers met with each of the students virtually and asked the students to read through all of the survey questions, explaining what they thought the meaning of each question was.

In our second version of the SPA administered in the spring of 2021, we updated several aspects of the survey based on the results from the previous semester. We moved one of the open-ended questions, "Describe the difference between probability and statistics in your own words," from the end of the survey to the position immediately prior to the first Likert-type items. From the think-alouds, it had been clear that even the

most high-achieving students sometimes confuse statistics with probability. By moving this open-ended question to earlier in the survey, students are forced to articulate the difference before answering any questions about attitudes toward probability, clarifying their position for us and for themselves. After consulting with experts in the area of data ethics, we also split the open-ended question, "How would you compare the ethical aspect of probability to the ethical aspect of statistics?" into two questions: "How do you feel about ethics in the context of probability?" and "How do you feel about ethics in the context of statistics?". The 36 Likert-type items on the pre- and post-SPA are listed in Supplemental Material S4.1 and S4.2 (full surveys provided upon request).

The components of the SPA along with their definitions were adapted from the six components of the SATS-36 (Schau and Emmioğlu, 2012). The components of the SPA with definitions and examples of pre survey items are

- *Affect* (6 items) – students' positive and negative feelings concerning probability – "I am scared by probability."
- *Cognitive Competence* (6 items) – students' attitudes about their intellectual knowledge and skills when applied to probability – "I can learn probability."
- *Value* (9 items) – students' attitudes about the usefulness, relevance, and worth of probability in personal and professional life – "I use probability in my everyday life."
- *Difficulty* (7 items) – students' attitudes about the difficulty of probability as a subject – "Most people have to learn a new way of thinking to do probability."
- *Interest* (4 items) – students' level of individual interest in probability – "I am interested in using probability."
- *Effort* (4 items) – amount of work the student expends to learn probability – "I plan to work hard in my probability course."

In the spring of 2021 implementation of the SPA, we also developed an alternative SPA for students who were not currently enrolled in a probability course (see Supplemental Material S4.3; full survey provided upon request). The questions were changed slightly to ask more generally about the field of probability as opposed to a specific course, and some irrelevant questions were removed. This version of the survey given to non-probability students serves as a control and to quantify the variability in these measures due to external factors.

4.3.2 SPA Distribution and Scoring

In this paper, we will present results from the spring semester of 2021. These survey participants were primarily undergraduate students at Penn State University Park campus who were enrolled in an introductory-level probability course in the Statistics Department. The SPA was distributed to two sections of MATH/STAT 318 (Elementary Probability), four sections of STAT 401 (Experimental Methods), five sections of MATH/STAT 414 (Introduction to Probability Theory), and two sections of STAT 418 (Introduction to Probability and Stochastic Processes for Engineering). The alternative SPA for non-probability students was distributed to two sections of MATH 41 at Penn State Behrend (Trigonometry and Analytic Geometry) and one section each of MATH 38 (Elementary Linear Algebra), MATH 141H (Honors Calculus with Analytic Geometry II), STAT 250 (Introduction to Biostatistics), MATH 22 at Penn State Behrend (College Algebra II and Analytic Geometry), and MATH 34 at Penn State Behrend (The Mathematics of Money).

A total of 14 instructors were recruited to distribute the survey among their students. In all sections except MATH 141H, students received extra credit of some form in return for submission of both the pre- and post-surveys. The amount of extra credit was left up to the instructors but was communicated to students prior to them gaining access to either survey. Students had the option to accept or decline consent to allow use of their data in this study. If they declined consent, they were directed to a read and respond activity. Students received the extra credit whether they consented or not as long as they completed the survey or the activity. In return for their help administering the survey, instructors received de-identified data for their students after the semester was completed and will be acknowledged in any resulting publications if they wish.

The pre- and post-SPA scores for each student for each attitude component were determined by switching the scale for the negatively-worded questions and taking the mean of all items within a component. Details about scoring can be found in Supplemental Information S4.4.

4.3.3 Validity Approach

When discussing the evidence supporting the validity of the SPA, we follow the validation inferences validity framework (see Cook and Hatala, 2016 and Kane, 2006). The validation inferences validity framework identifies four key inferences: scoring, generalization, extrapolation, and implications/decisions.

We examined the scoring structure of the SPA with confirmatory factor analysis, comparing several hypothesized models, similarly to the comparisons of hypothesized models for the SATS-36 in Vanhoof et al. (2011). The development process, including the think aloud exercises and expert review, also support scoring inferences.

Evidence to support generalization of SPA results would ideally show consistency of scores if the same individual were to take the SPA more than once, and similarity between item scores within the same construct. We plotted inter-individual variability and calculated standard errors for a subset of students who completed either the pre- or post-survey twice. This occurred for students who completed the survey to get extra credit in two course sections or students who were only enrolled in one section but voluntarily completed the survey twice. We calculated the reliability of scores using Cronbach's alpha.

Evidence to support extrapolation from SPA scores to true student attitudes should show a comparison between the SPA and other measures of student attitudes. Immediately after taking the post-SPA at the end of the spring 2021 semester, 216 probability students volunteered to complete an additional survey. These 216 students completed the PAI (Tan et al., 2011) in addition to the SPA. We calculated and plotted correlations between constructs from both instruments for these individuals. We also visually compare change scores between the SATS-36 and the SPA.

Lastly, we examined evidence that the SPA provides a rational basis for decisions related to probability teaching methods. This entailed examination of the impacts of the SPA on the students and instructors. To inform teaching methods, the SPA should also exhibit changes when major changes are made in the teaching method. While we have not used the SPA to assess changes in teaching methods, we compare students enrolled in a probability course to students not enrolled in a probability course in Section 4.4.2.3. We also examined the practical considerations when implementing the SPA.

4.3.4 Analysis Approach

We used mixed effects models and t-tests to model the difference between pre- and post-SPA scores. When comparing pre- and post- SATS-36 component scores, researchers often perform direct 2-sample hypothesis testing using, for example, a t-test or non-parametric Wilcoxon signed rank test (e.g., Swanson et al., 2014; Kiekkas et al., 2015). While t-procedures and their non-parametric counterparts are easy to implement and interpret, they have a number of negative characteristics when applied to assess student attitudes (Millar and Schau, 2010). For example, the assumption that each observation

is independent of the others is violated, since students share instructors and classrooms. They also cannot account for the relationship between the pre-survey scores and the gain. Despite these caveats, we performed t-tests for each attitude component in order to facilitate comparisons with similar studies in the literature.

A more thorough approach to comparing pre- and post-attitude scores would be to model the gain as a function of the pre-attitude score (e.g., Millar and Schau, 2010). We adopted this approach, including a random intercept for course section. For each attitude construct, we estimated the model:

$$\text{gain}_{ij} = \beta_0 + \beta_1(\text{pre}_{ij}) + \gamma_j + \epsilon_{ij} \tag{4.1}$$

where $\gamma_j \sim N(0, \sigma^2_\gamma)$ and $\epsilon_{ij} \sim N(0, \sigma^2_\epsilon)$. Response variable gain_{ij} is the difference between the pre- and post-survey factor score for individual i in course section j, and pre_{ij} is the pre-survey factor score for individual i.

As Millar and Schau (2010) point out, the model defined in (4.1) confounds the true relationship between the gain and pre-attitude and regression to the mean. Regression to the mean occurs for Likert-type items because a low pre-score allows for a larger gain than a high pre-score allows, and a high pre-score allows for a larger negative gain than a low pre-score allows. Thus, we would expect $\beta_1 < 0$ in (4.1) even absent of a true relationship between gain and pre-attitudes. To address this issue, we added two terms which are estimated using data from students who were not enrolled in a probability course. For each attitude construct, we estimated the revised model:

$$\text{gain}_{ij} = \alpha_0 + \beta_0(\text{prob}_{ij}) + \alpha_1(\text{pre}_{ij}) + \beta_1(\text{prob}_{ij})(\text{pre}_{ij}) + \gamma_j + \epsilon_{ij} \tag{4.2}$$

where prob_{ij} is 1 if individual i was currently enrolled in a probability course and 0 if individual i was not enrolled in a probability course.Under (4.2), $\alpha_0 + \alpha_1(\text{pre})$ reflects the expected gain from the beginning to the end of the semester attributed to regression to the mean. Then $\beta_0 + \beta_1(\text{pre})$ reflects the expected gain attributed to the probability course.

4.4 Results

A total of 615 probability students (>78% of enrollment) completed the pre-survey and 525 probability students (>67% of enrollment, some of which did not complete the pre-survey) completed the post-survey from 13 sections in the spring of 2020. We removed

Demographic variable from pre-survey	Bonferroni-corrected p value
Taken a statistics class before	0.68
Gender	1.00
Pursuing bachelors degree	1.00
Of Hispanic, Latino, or Spanish origin	1.00
Race	1.00
Grade expected in this course	1.00
Math ability (average of 2 questions)	0.07
Months since most recent college-level mathematics or statistics course	1.00

Table 4.1. Bonferroni-corrected p values for testing the difference between probability students who did and did not complete the post-survey of those that completed the pre-survey.

surveys from individuals who repeatedly entered the same number, completed the survey in less than 5 minutes, or completed less than 80% of the main Likert items, leaving 595 pre-surveys (>76% of enrollment) and 501 post-surveys (>64% of enrollment) from probability students. Remaining missing Likert items were imputed using section means; imputation was implemented for 4.1% of probability students. Probability students who completed the pre-survey but not the post-survey were not significantly different than students who completed both surveys as measured for a range of demographic variables (see Table 4.1).

427 probability students completed both the pre- and post-survey. These students identified as 63% white or Asian males and 29% white or Asian females (see Table 4.2). 5% identified as being of Hispanic, Latino, or of Spanish origin. 90% identified as seeking a Bachelor's degree, with the remainder seeking joint Bachelor's and graduate degrees (6%), solely graduate degrees (3%), or something else (1%). Of those seeking Bachelor's degrees, the most common cohort was Sophomores (41%), followed by Juniors (30%), Seniors (14%), Freshmen (4%), and other (<1%). The most common majors fell under Engineering (29%), followed by Computer Science (19%) and Statistics/Data Science (15%) (see Figure 4.1). 61% of the students reported having taken at least one college-level probability, statistics, or data science course prior to the semester in which the study took place.

	Male (n=289)	Female (n=136)	Other (n=3)
White (n=220)	151	69	0
Asian (n=166)	115	51	0
Black or African American (n=9)	2	7	0
White and Asian (n=7)	4	2	1
White and (Black or African American) (n=4)	3	1	0
Middle Eastern (n=3)	2	1	0
Other race (n=18)	12	4	2

Table 4.2. Demographic make-up of the 427 probability students who completed both pre- and post-surveys.

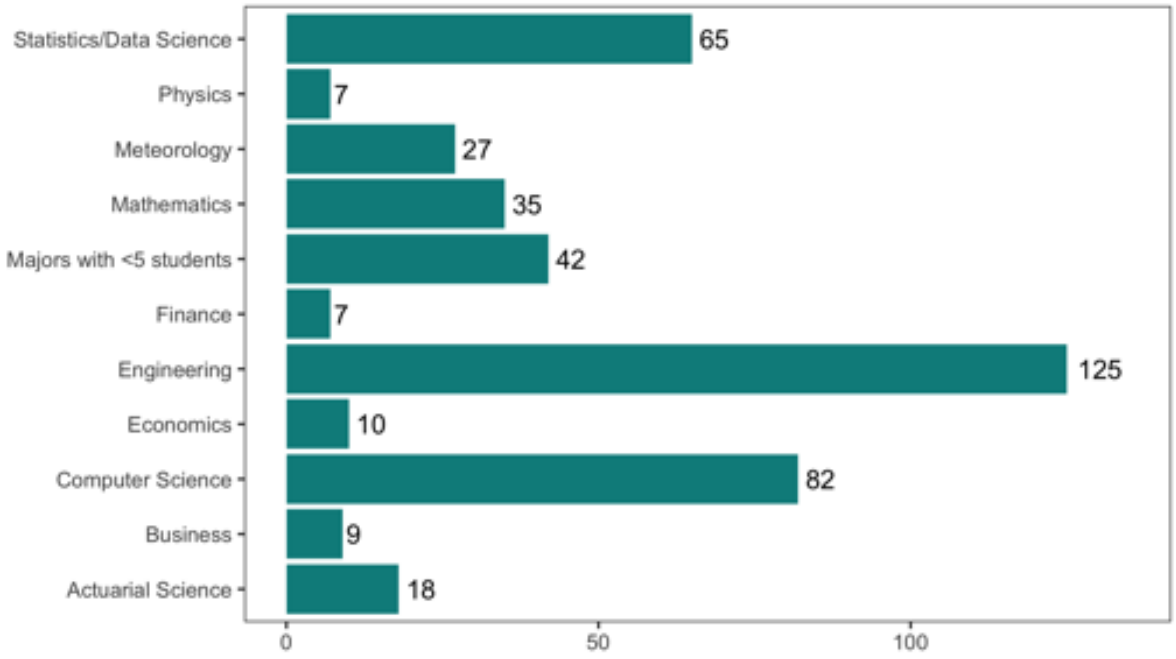

Figure 4.1. Distribution of majors for the 427 probability students who completed both pre- and post-surveys. Students with double majors were instructed to select the major which matched more closely with their interests.

4.4.1 Validity Evidence for the SPA

The SPA is intended to be used by educators for A-B testing of methods for teaching probability. The survey tool is intended to measure students' feelings about their probability class and the subject of probability. Following the guidance of Cook and Hatala (2016), we present the evidence supporting the validity of the SPA scores, generalization of scores within the testing context, extrapolation to real attitudes, and implications for decision-making.

Model	Data	n	RMSEA	NNFI	CFI
Original	Pre-survey	595	0.119	0.851	0.863
Original	Post-survey	501	0.115	0.860	0.872
Modified	Pre-survey	595	0.110	0.885	0.895
Modified	Post-survey	501	0.105	0.896	0.905

Table 4.3. Fit indices for the 6-factor SPA from probability students in the spring of 2021

4.4.1.1 Scoring

Use of this instrument for the intended purpose requires the assumption that students' attitudes are correctly transformed into a consistent numeric score for each of the six components. The development process as described in Section 4.3.1 evidences the relevance and interpretability of the questions.

Confirmatory factor analyses confirm the transferability of the structure of the SATS-36 for assessment of attitudes toward probability. As shown in Table 4.3, reasonable fit indices were obtained for the original 6-factor structure adopted from the SATS-36. Associations among the factors are shown in Table 4.4. While there is a high correlation between the *Cognitive Competence* and *Affect* factors in both the pre-survey ($r = 0.955$) and post-survey ($r = 0.933$), the 6-factor structure was found to fit the data significantly better than an otherwise equivalent model that combines *Cognitive Competence* and *Affect* ($\triangle\chi^2 = 108.73$, $p < 0.001$ for pre-survey; $\triangle\chi^2 = 68.075$, $p < 0.001$ for post-survey).

As in Vanhoof et al. (2011), items with factor loadings below 0.4 were considered for deletion, since they may not relate well to the underlying construct. In the resulting modified 6-factor model, items 6 ("Probability formulas are easy to understand", meant to assess *Difficulty* construct), 22 ("Probability is a subject quickly learned by most people", meant to assess *Difficulty* construct), and 26 ("I make a lot of math errors in probability", meant to assess *Cognitive Competence* construct) were removed. Table 4.3 displays reasonable fit indices for the modified structure, and Table 4.4 displays associations between factors. Exploratory factor analysis results also align well with the modified structure (Figure 4.2). We will adopt the modified structure for the SPA throughout the remainder of this paper.

4.4.1.2 Generalization

Some students completed the pre- or post-SPA twice due to their enrollment in multiple courses taking part in the study or of their own accord. Twenty individuals took the pre-survey twice (n=18 completed all items), and 14 individuals completed the post-survey

Original Pre-survey	*Affect*	*Cognitive Competence*	*Value*	*Difficulty*	*Effort*
Cognitive Competence	0.955				
Value	0.472	0.606			
Interest	0.634	0.582	0.752		
Difficulty	0.405	0.309	-0.261	-0.211	
Effort	0.248	0.402	0.520	0.466	-0.464
Original Post-survey	*Affect*	*Cognitive Competence*	*Value*	*Difficulty*	*Effort*
Cognitive Competence	0.933				
Value	0.553	0.624			
Interest	0.635	0.524	0.692		
Difficulty	0.512	0.502	-0.027	-0.073	
Effort	0.230	0.362	0.333	0.335	-0.260

Modified Pre-survey	*Affect*	*Cognitive Competence*	*Value*	*Difficulty*	*Effort*
Cognitive Competence	0.939				
Value	0.480	0.635			
Interest	0.648	0.617	0.753		
Difficulty	0.274	0.167	-0.293	-0.315	
Effort	0.258	0.433	0.520	0.467	-0.482
Modified Post-survey	*Affect*	*Cognitive Competence*	*Value*	*Difficulty*	*Effort*
Cognitive Competence	0.923				
Value	0.557	0.637			
Interest	0.646	0.544	0.693		
Difficulty	0.298	0.267	-0.127	-0.252	
Effort	0.236	0.379	0.334	0.336	-0.350

Table 4.4. Estimated latent factor correlations for the original and modified 6-factor SPA in the spring of 2021

twice. The inter-individual variability of factor scores for these individuals is displayed in Figure 4.3, and corresponding standard errors are shown in Table 4.5.

We calculated reliability of the SPA using Cronbach's alpha (Table 4.6). As was the case for the SATS-36 in Schau and Emmioğlu (2012), alpha values for each factor in the SPA exceeded the 0.7 heuristic for a low stakes assessment . This supports the internal consistency of the items making up each factor score. While Cronbach's alpha is the most popular tool to estimate internal consistency reliability, the procedure has several well-documented limitations, including the assumptions of uncorrelated errors, tau-equivalence, and normality (Trizano-Hermosilla and Alvarado, 2016). For this reason,

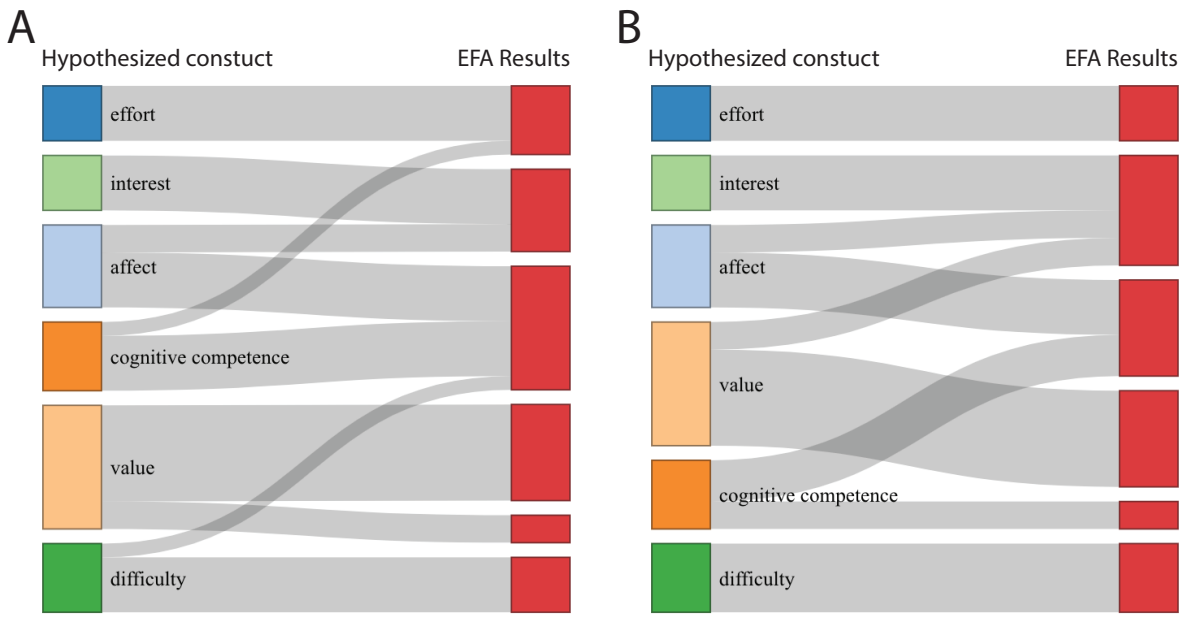

Figure 4.2. Sankey diagrams comparing the modified 6-factor structure we hypothesize to the 6-factor structure resulting from EFA. EFA results used (A) pre-survey data from probability students and (B) post-survey data from probability students.

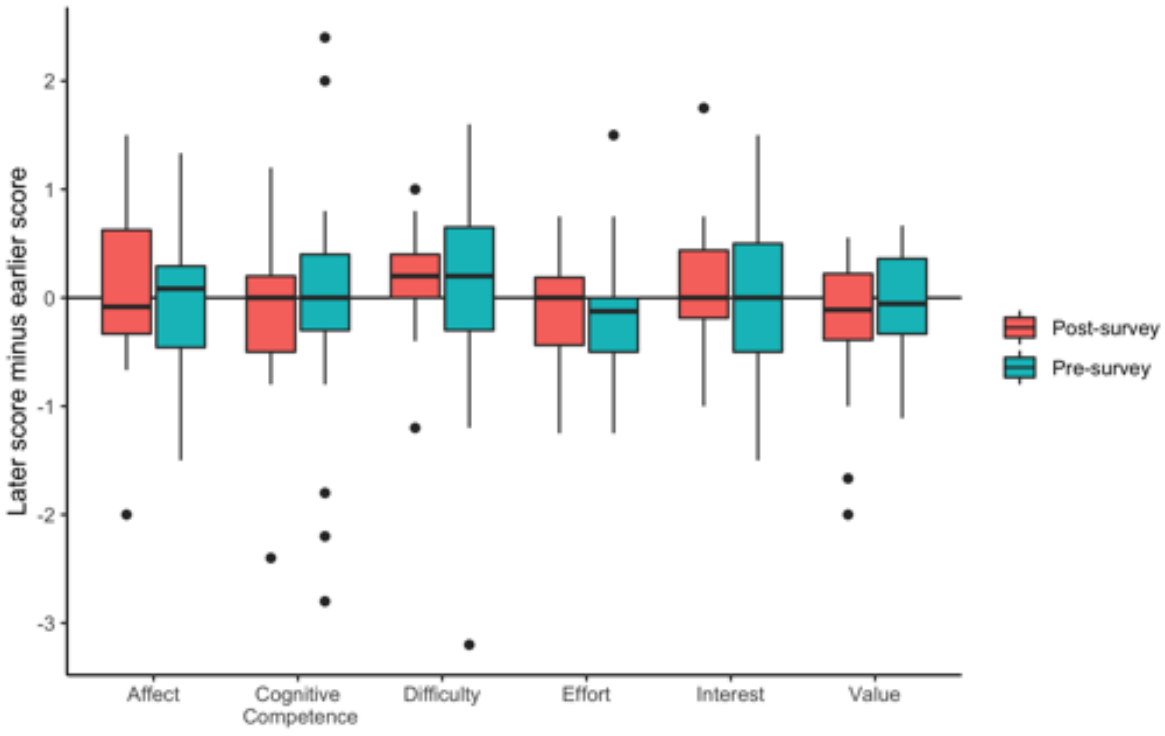

Figure 4.3. Box plots of inter-individual response variability for the SPA. Sample size is 20 for all pre-survey scores except *Affect* (n=18) and *Interest* (n=19). Sample size is 14 for all post-survey scores.

	Affect	Cognitive Competence	*Value*	*Interest*	*Difficulty*	*Effort*
Pre-survey	0.173	0.276	0.119	0.170	0.248	0.136
Post-survey	0.242	0.224	0.203	0.211	0.152	0.169

Table 4.5. Standard errors for factor scores based on students who completed the pre- or post-SPA twice.

Attitude Component	SPA Pre-survey	SPA Post-survey
Affect	.79	.79
Cognitive Competence	.78	.80
Value	.85	.87
Difficulty	.73	.74
Interest	.86	.91
Effort	.81	.72

Table 4.6. Cronbach's alpha values for the SPA.

we also looked at McDonald's omega coefficient, but results were very similar in our case (Trizano-Hermosilla and Alvarado, 2016).

4.4.1.3 Extrapolation

We computed correlations and plotted the relationship between factor scores for the modified post-SPA described in Section 4.4.1.1 and the PAI (see Figure 4.4). We also computed p-values for each pair using Pearson's product moment correlation coefficient. There is a strong positive correlation between the *Interest* scale on the SPA and the interest scale on the PAI ($p < 0.001$). The construct assessing students' self concept in probability in the PAI is strongly correlated with the *Affect* and *Cognitive Competence* factors in the SPA ($p < 0.001$). Students' attitudes toward the usefulness of probability as measured by the PAI is also strongly correlated with *Value* in the SPA ($p < 0.001$). Thus, the *Affect*, *Cognitive Competence*, *Interest*, and *Value* scales on the SPA are strongly related to at least one factor in the PAI. *Difficulty* and *Effort*, however, do not have counterparts on the PAI that they relate well with.

Figure 4.5 shows average course section change scores from pre- to post-survey for the SPA and the SATS-36 (Schau and Emmioğlu, 2012). Median change scores are in a similar range of 0 for both the SATS-36 and SPA. In both SPA and SATS-36 results, mean section change scores rarely surpass the threshold for importance of ± 0.5 specified in Schau and Emmioğlu (2012).

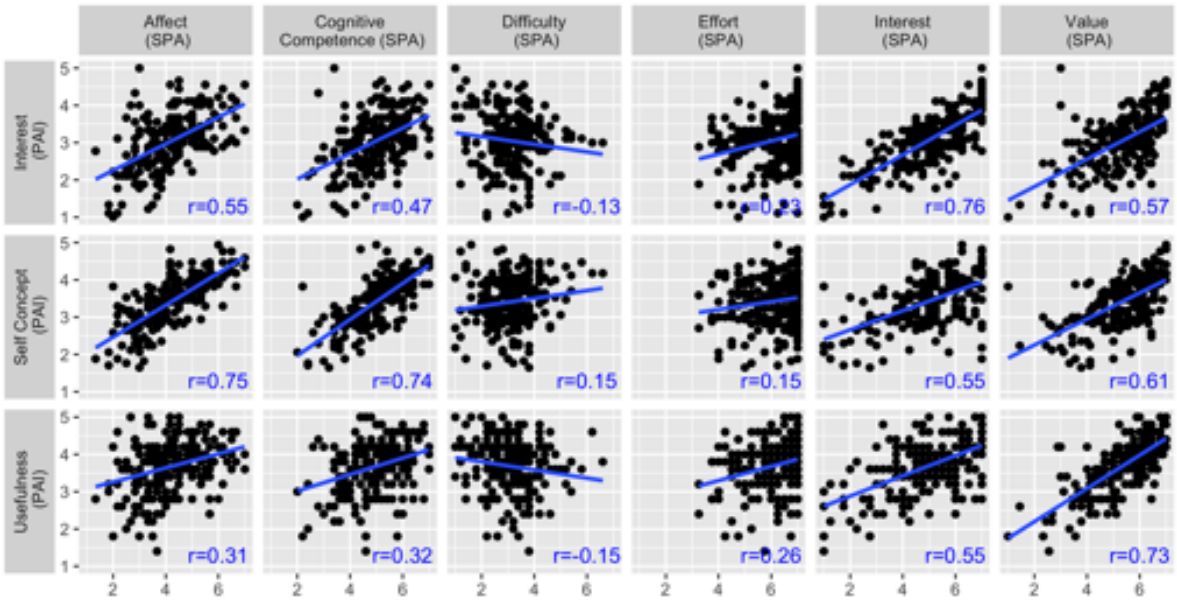

Figure 4.4. Correlations (r) between factor scores on the PAI (y-axis) and post-SPA (x-axis) are shown in blue. Black dots depict scored for the 216 students who volunteered to complete the additional survey.

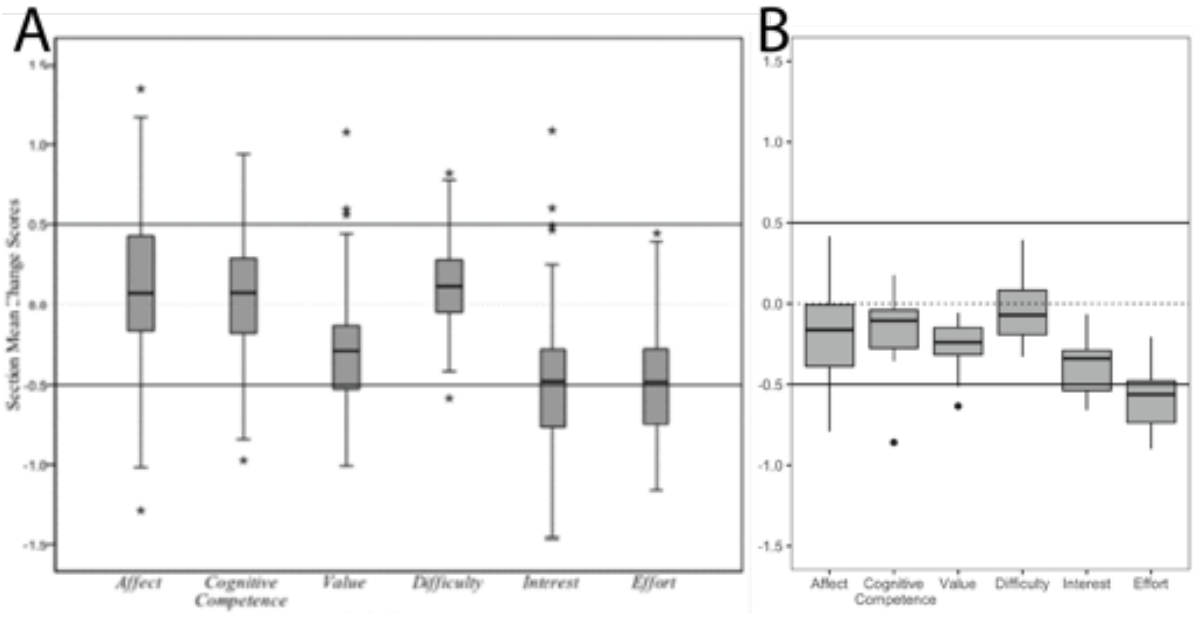

Figure 4.5. In (A), changes in mean pre- and post- SATS-36 scores for each component are shown for 101 introductory statistics sections (used with permission from Schau and Emmioğlu, 2012). In (B), changes in mean pre- and post-SPA scores for each component are shown for 13 probability sections.

4.4.1.4 Implications

We examined the impacts of completing the SPA on students and instructors and determined they were minimal. The median time students required to complete the pre-SPA was 12 minutes and 44 seconds. The median time for the post-SPA was 10 minutes and 51 seconds. Since students completed the survey online in their own time, they could take as much time as they needed. Students could also leave the survey partway through and return at a later time to complete the questions. Instructors needed only to send out the link to their students and provide extra credit after the deadline to those who participated.

If the SPA is to be used to inform teaching methodology, the SPA should exhibit change corresponding with significant course changes. In Section 4.4.2.3, we compare SPA scores for students enrolled in a probability course and students not enrolled in a probability course during the same semester. This illustrates an extreme case of comparing SPA scores for sections implementing different methods of teaching probability; namely, teaching probability is compared to not teaching probability (i.e., the control group). The lack of significant differences between these two groups could indicate an area of weakness with regard to the validity of the instrument.

4.4.2 Analysis of Attitude Changes from Pre- to Post-SPA

4.4.2.1 t-Procedures

We performed paired t-tests to test whether each mean attitude component score was different at the beginning and end of a probability course. There was a significant decrease in mean *Affect* ($p = 0.0044$; Cohen's $d = -0.14$), *Cognitive Competence* ($p = 0.0013$; Cohen's $d = -0.16$), *Effort* ($p < 0.0001$; Cohen's $d = -0.54$), *Interest* ($p < 0.0001$; Cohen's $d = -0.33$), and *Value* ($p < 0.0001$; Cohen's $d = -0.30$). The change in *Difficulty* was not significant ($p = 0.3730$; Cohen's $d = -0.04$).

4.4.2.2 Linear Model for Probability Students

We fit model (4.1) via restricted maximum likelihood (REML) estimation for each attitude component. Parameter estimates for each of the attitude component models are shown in Table 4.7A. The normality assumption was not satisfied for the *Effort* model, so *Effort* is excluded from this analysis (see Supplemental Information S4.5). For each of the five other attitude components, there is statistical evidence of a negative linear

A. Linear Mixed Effects Regression for 427 Probability Students

	Estimated mean gain conditional on pre-score, averaged across course sections	Estimated standard deviation of course section random effect
Affect	Mean Gain = 1.70 – 0.42(Pre)	0.25
Cognitive Competence	Mean Gain = 1.94 – 0.42(Pre)	0.16
Value	Mean Gain = 1.08 – 0.24(Pre)	0.05
Interest	Mean Gain = 0.88 – 0.24(Pre)	0.09
Difficulty	Mean Gain = 1.63 – 0.48(Pre)	0.23

B. Linear Mixed Effects Regression for 427 Probability Students and 159 Non-Probability Students

	Estimated mean gain conditional on pre-score and whether or not the student is enrolled in a probability course, averaged across course sections	Estimated standard deviation of course section random effect
Affect	Mean Gain = 0.92 – 0.21(Pre) + 0.65(Prob) – 0.17(Pre)(Prob)	0.19
Cognitive Competence	Mean Gain = 2.30 – 0.44(Pre) – 0.31(Prob)	0.16
Value	Mean Gain = 1.93 – 0.42(Pre) – 0.84(Prob) + 0.18(Pre)(Prob)	0.09
Interest	Mean Gain = 1.76 – 0.43(Pre) – 0.87(Prob) + 0.19(Pre)(Prob)	0.09
Difficulty	Mean Gain = 1.87 – 0.46(Pre) – 0.30(Prob)	0.26

Table 4.7. Linear mixed models fit with REML for each attitude component. Fit model (4.1) for probability students in (A) and fit model (4.2) for both probability and non-probability students in (B). *Affect* and *Cognitive Competence* scores were calculated as the mean of 4 items each, instead of 6 and 5 respectively, because some items could not easily be adapted for non-probability students.

relationship between pre-score and gain (all p-values were < 0.0001). Thus, the attitudes of individuals who have smaller pre-scores are expected to improve more at the end of the course than for those who have larger pre-scores, as measured by each of the 5 attitude components.

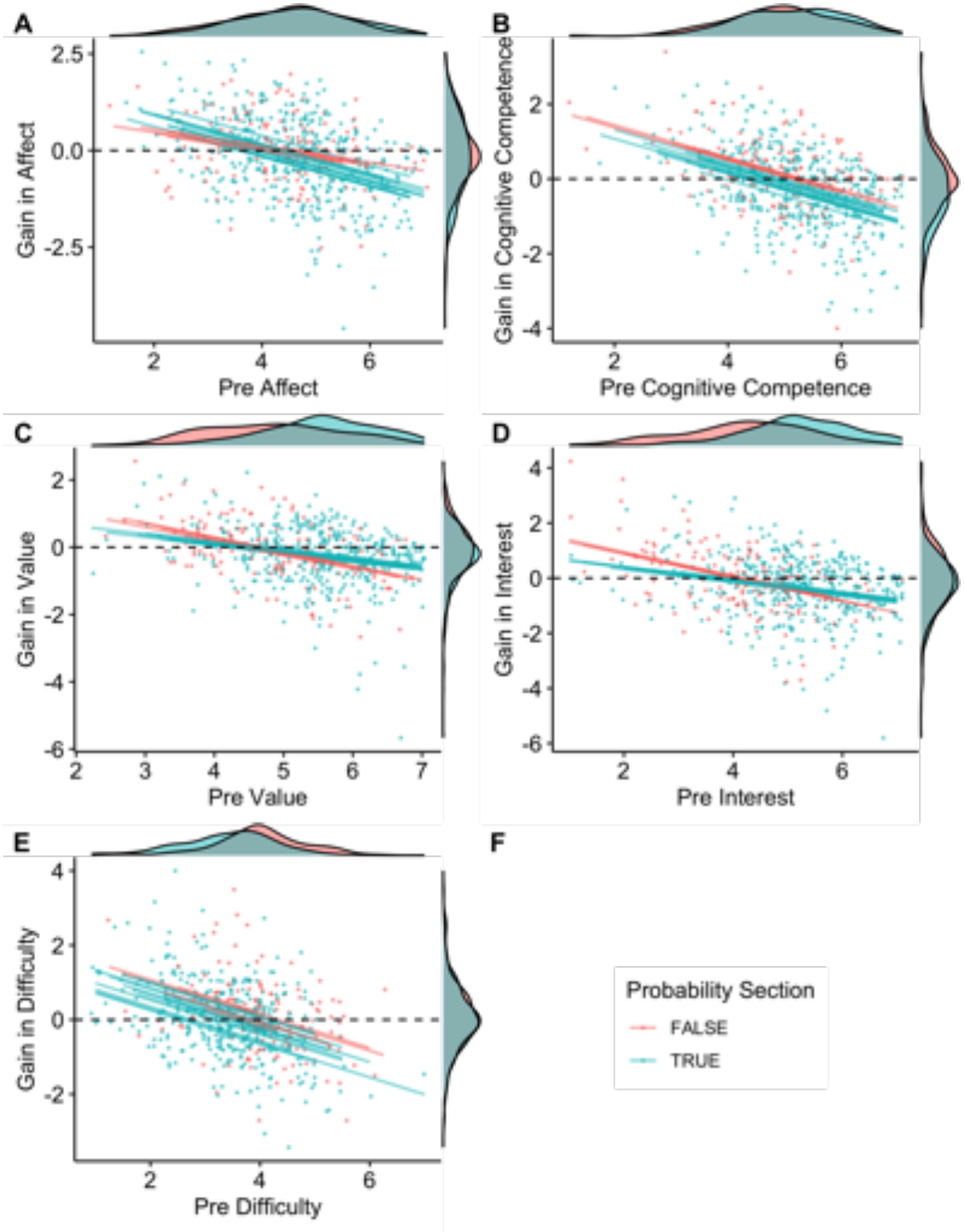

Figure 4.6. Scatter plots of pre SPA scores versus gain from pre to post are displayed in A-E, with jittered points for ease of visibility, for 5 attitude components. Colored lines depict the section mean lines from the mixed effects models in Table 4.7B. Marginal densities are displayed in the margins. A shared legend is shown in F.

Demographic variable from pre-survey	Bonferroni-corrected p value
Taken a statistics class before	0.07
Gender	2.81×10^{-23}
Pursuing bachelors degree	1.00
Of Hispanic, Latino, or Spanish origin	1.00
Race	2.66×10^{-13}
Grade expected in this course	0.04
Math ability (average of 2 questions)	5.17×10^{-5}
Months since most recent college-level mathematics or statistics course	1.55×10^{-5}

Table 4.8. Bonferroni-corrected p values for testing the difference between probability students and non-probability students.

4.4.2.3 Linear Model for Probability and Non-probability Students

Scatter plots of the relationship between pre-survey scores and gain for probability students are shown in red in Figure 4.6. While the negative relationship between pre-survey scores and gain is clearly shown in each case, it is impossible to know whether it is due to true dependence between pre-survey scores and gain or due to regression to the mean using only data from the 427 probability students. To estimate the additional parameters in model (4.2), we added results from an alternative SPA completed by 159 students enrolled in non-probability courses which are listed in Section 4.3.2. The demographic make-up was significantly different between probability and non-probability students, as tested using Bonferroni-corrected p values (see Table 4.8).

We fit model (4.2) via restricted maximum likelihood (REML) estimation for each attitude component. Parameter estimates for each of the attitude component models are shown in Table 4.7B. *Effort* was excluded from this analysis as well since the items within the *Effort* component could not be easily adapted for non-probability students, so *Effort* was not included on the alternative SPA.

These data support evidence of an interaction effect between pre score and whether the student is enrolled in a probability section for each of *Affect* ($p = 0.02493$), *Interest* ($p = 0.02371$), and *Value* ($p = 0.02689$). Thus, the difference between probability and non-probability students is different depending on the score on the pre survey. However, when we tested the differences in mean gain between probability and non-probability students at the 10^{th} and 90^{th} percentile of pre scores, none of these differences were significant at $\alpha = 0.05$ (see Table 4.9).

The interaction effect was removed from the models for *Cognitive Competence* and *Difficulty* (see Table 4.7B). The gain in *Cognitive Competence* was significantly smaller

	Pre Score	Mean Difference in Gain (Probability – Non-Probability)	p Value
Affect	10th Percentile: 3.12	0.115	0.5119
Affect	90th Percentile: 6	-0.374	0.0521
Value	10th Percentile: 4	-0.134	0.3176
Value	90th Percentile: 6.56	0.318	0.0555
Interest	10th Percentile: 3.25	-0.270	0.1179
Interest	90th Percentile: 6.38	0.312	0.1353

Table 4.9. Results of ratio t tests comparing mean gains for probability and non-probability students given specific values for pre score.

for probability students that for non-probability students ($p = 0.01818$), although the mean difference of 0.31 would not be considered an important difference according to the standard of 0.5 points described for the SATS-36 by Schau and Emmioğlu (2012). The mean gain in *Difficulty* was smaller for probability students than non-probability students, but the difference was not significant using $\alpha = 0.05$ ($p = 0.07082$). Recall that high scores for *Difficulty* imply less perceived difficulty, and low scores imply the student perceives more difficulty. Thus, a smaller gain in *Difficulty* would imply the students saw less improvement in viewing probability as not difficult.

4.5 Discussion

Probability attitudes have received little attention in the statistics education community despite the general understanding of the importance of attitudes and the subject of probability in statistics curricula. We found only one instrument for measurement of probability attitudes, the PAI, which has not gone through the same careful development and validation process as the SATS-36. In this paper, we describe the process of adapting the SATS-36 to develop the SPA, we present evidence of the validity of the instrument, and we illustrate analyses of pre- and post-SPA results for probability and non-probability students. Paired t-tests alone would lead the researcher to conclude that there was a significant decrease in five of the six attitude components from the beginning to the end of a college-level probability course at Pennsylvania State University. However, further analyses lead to a more nuanced conclusion. We modeled mean gain in attitude component score as a function of pre-score and with a random effect for course section, which illustrated the difference in the trends for probability students with high and low pre-scores. We then added survey scores from an adapted version of the SPA for non-probability students, and we fit a mixed effects model to the scores from all probability

and non-probability student participants. The results suggest few differences between probability and non-probability student scores.

4.5.1 Limitations

There are some important caveats regarding the similarities and differences between the probability and non-probability groups participating in this study. The data was collected during a global pandemic, during which all classes were held online. There is evidence that students' attitudes were impacted by the shift to remote learning during the COVID-19 pandemic (Wester et al., 2021), so the study should be repeated during a less tumultuous period for students. The demographic make-up of the probability and non-probability students also differs, and the non-response rate is much higher in the non-probability sections (>37% of non-probability students enrolled completed both the pre- and post-SPA versus >54% of probability students enrolled completed both the pre- and post-SPA). Future studies should also include students from a variety of colleges and universities.

In the data collection procedure, we had to balance careful research with student privacy. If we weren't subject to IRB & FERPA considerations, we ideally would have recorded additional variables including students' academic history, grades, and assignments within the class. We also needed to balance several factors in the timing of the pre- and post-surveys. We asked students to complete the surveys in the first and last three weeks of the semester to allow students time to complete the assignment in their own time, to allow time for students who add the course late to complete the pre-survey, and to allow instructors time to allot extra credit at the end of the semester. First impressions likely fluctuate in those first days, and students that "late drop" are lost to follow up which almost certainly biases the data to some extent.

One assumption inherent in this study is the assumption that students know what probability is, and that they know how it differs from statistics. Further studies could examine the impact of a probability course on students attitudes toward statistics. There is an open-ended question on the SPA which prompts students for their interpretation of how probability and statistics differ. Further analyses could make use of the answers to this question, possibly giving higher weight to students with a better idea of what probability is. It should also be noted that 66% of the non-probability students were enrolled in introductory biostatistics, which although not primarily a probability course, may cover basic probability concepts. However, the analyses in Sections 4.4.2.2 and 4.4.2.3 were repeated without this cohort, and the main conclusions did not differ.

4.5.2 Implications for Teaching

Despite the limitations of this study, we have provided evidence that students' attitudes toward probability did not improve on average from the beginning to the end of the probability courses in our study. This suggests that probability teachers may want to consider positive attitudes toward probability as a desired course outcome, and this outcome should inform teaching methods. Instructors may choose to implement the SPA over the PAI due to the validity evidence supporting the SPA or because of the analogous structure of the SPA and SATS-36. Instructors who choose to implement the SPA in their own classrooms should not expect small modifications in teaching methods to translate to large differences in SPA scores between treatment and control groups. Alternatively, instructors may decide to target specific attitude components that are especially low in their students at the beginning of the semester, or the SPA could be used to motivate in-class discussions with students to allow students to examine their own attitudes toward probability.

4.5.3 Implications for Research

The lack of differences between probability and non-probability students' attitudes toward probability suggest the import of using a control group for attitude studies in general. The effect of the pre score on the gain also cannot be ignored when conducting similar analyses. While we found strong evidence of the validity of the SPA for assessing students' attitudes toward probability, future iterations could improve the ability of the SPA to identify small changes in students' attitudes. The motivations behind students attitudes should also be explored. Several open-ended questions on the SPA prompt students to list experiences they believe most positively or negatively influenced their current attitude toward statistics or probability. While assessment of these answers was outside of the scope of this study, further research could use similar prompts to identify areas for improvement.

S4 Supplement to Survey of Probability Attitudes

S4.1 Pre-Survey of Probability Attitudes for Probability Students

The statements below are designed to identify your attitudes about probability. Each item has 7 possible responses. The responses range from 1 (strongly disagree) through 4

(neither disagree nor agree) to 7 (strongly agree). If you have no opinion, choose response 4. Please read each statement. Mark the one response that most clearly represents your degree of agreement or disagreement with that statement. Try not to think too deeply about each response. Record your answer and move quickly to the next item. Please respond to all of the statements.

1. I plan to complete all of my probability assignments.
2. I plan to work hard in my probability course.
3. I will like probability.
4. I will feel insecure when I have to do probability problems.
5. I will have trouble understanding probability because of how I think.
6. Probability formulas are easy to understand.
7. Probability is worthless.
8. Probability is a complicated subject.
9. Probability should be a required part of my professional training.
10. Probability skills will make me more employable.
11. I will have no idea of what's going on in this probability course.
12. I am interested in being able to communicate probabilistic arguments to others.
13. Probability is not useful to the typical professional.
14. I plan to study hard for every probability test.
15. I will get frustrated going over probability tests in class.
16. Probabilistic thinking is not applicable in my life outside my job.
17. I use probability in my everyday life.
18. I will be under stress during probability class.
19. I will enjoy taking probability courses.

20. I am interested in using probability.
21. Probability conclusions are rarely presented in everyday life.
22. Probability is a subject quickly learned by most people.
23. I am interested in understanding probabilistic arguments.
24. Learning probability requires a great deal of discipline.
25. I will have no application for probability in my profession.
26. I will make a lot of math errors in probability.
27. I plan to attend every probability class session.
28. I am scared by probability.
29. I am interested in learning probability.
30. Probability involves massive computations.
31. I can learn probability.
32. I will understand probability equations.
33. Probability is irrelevant in my life.
34. Probability is highly technical.
35. I will find it difficult to understand probability concepts.
36. Most people have to learn a new way of thinking to do probability.

S4.2 Post-Survey of Probability Attitudes for Probability Students

The statements below are designed to identify your attitudes about probability. Each item has 7 possible responses. The responses range from 1 (strongly disagree) through 4 (neither disagree nor agree) to 7 (strongly agree). If you have no opinion, choose response 4. Please read each statement. Mark the one response that most clearly represents your degree of agreement or disagreement with that statement. Try not to think too deeply about each response. Record your answer and move quickly to the next item. Please respond to all of the statements.

1. I tried to complete all of my probability assignments.
2. I worked hard in my probability course.
3. I like probability.
4. I feel insecure when I have to do probability problems.
5. I have trouble understanding probability because of how I think.
6. Probability formulas are easy to understand.
7. Probability is worthless.
8. Probability is a complicated subject.
9. Probability should be a required part of my professional training.
10. Probability skills will make me more employable.
11. I have no idea of what's going on in this probability course.
12. I am interested in being able to communicate probabilistic arguments to others.
13. Probability is not useful to the typical professional.
14. I tried to study hard for every probability test.
15. I get frustrated going over probability tests in class.
16. Probabilistic thinking is not applicable in my life outside my job.
17. I use probability in my everyday life.
18. I am under stress during probability class.
19. I enjoy taking probability courses.
20. I am interested in using probability.
21. Probability conclusions are rarely presented in everyday life.
22. Probability is a subject quickly learned by most people.
23. I am interested in understanding probabilistic arguments.

24. Learning probability requires a great deal of discipline.

25. I will have no application for probability in my profession.

26. I make a lot of math errors in probability.

27. I tried to attend every probability class session.

28. I am scared by probability.

29. I am interested in learning probability.

30. Probability involves massive computations.

31. I can learn probability.

32. I understand probability equations.

33. Probability is irrelevant in my life.

34. Probability is highly technical.

35. I find it difficult to understand probability concepts.

36. Most people have to learn a new way of thinking to do probability.

S4.3 Alternative Pre- and Post-Survey of Probability Attitudes for Non-Probability Students

The statements below are designed to identify your attitudes about probability. Each item has 7 possible responses. The responses range from 1 (strongly disagree) through 4 (neither disagree nor agree) to 7 (strongly agree). If you have no opinion, choose response 4. Please read each statement. Mark the one response that most clearly represents your degree of agreement or disagreement with that statement. Try not to think too deeply about each response. Record your answer and move quickly to the next item. Please respond to all of the statements.

3. I would like probability.

4. I would feel insecure if I had to do probability problems.

5. I would have trouble understanding probability because of how I think.

6. Probability formulas are easy to understand.
7. Probability is worthless.
8. Probability is a complicated subject.
9. Probability should be a required part of my professional training.
10. Probability skills would make me more employable.
12. I am interested in being able to communicate probabilistic arguments to others.
13. Probability is not useful to the typical professional.
16. Probabilistic thinking is not applicable in my life outside my job.
17. I use probability in my everyday life.
19. I would enjoy taking probability courses.
20. I am interested in using probability.
21. Probability conclusions are rarely presented in everyday life.
22. Probability is a subject quickly learned by most people.
23. I am interested in understanding probabilistic arguments.
24. Learning probability requires a great deal of discipline.
25. I will have no application for probability in my profession.
26. I would make a lot of math errors when doing probability.
28. I am scared by probability.
29. I am interested in learning probability.
30. Probability involves massive computations.
31. I can learn probability.
32. I would understand probability equations.
33. Probability is irrelevant in my life.

34. Probability is highly technical.

35. I would find it difficult to understand probability concepts.

36. Most people have to learn a new way of thinking to do probability.

S4.4 Scoring Guide for SPA and alternative SPA

Component (subscale) scores on the SPA are formed by

1. Reversing the responses to the negatively worded items indicated with an asterisk* (1 becomes 7, 2 becomes 6, etc.),

2. Summing the item responses within each component, and

3. Dividing by the number of items within each component.

The following lists the attitude components followed by the item numbers within each component. Item numbers are the same in the pre and the post versions in Supplemental Material S4.1, S4.2, and S4.3. Negatively worded items are indicated with an asterisk*, and items which are not included in the alternative SPA are indicated with a dagger†.

- *Affect*: 3, 4*, 15*†, 18*†, 19, 28
- *Cognitive Competence*: 5*, 11*†, 26*, 31, 32, 35*
- *Value*: 7*, 9, 10, 13*, 16*, 17, 21*, 25*, 33*
- *Difficulty*: 6, 8*, 22, 24*, 30*, 34*, 36*
- *Interest*: 12, 20, 23, 29
- *Effort*: 1†, 2†, 14†, 27†

S4.5 Effort Plot

As shown in Figure 4.7, the residuals for the *Effort* model fail the normality assumption.

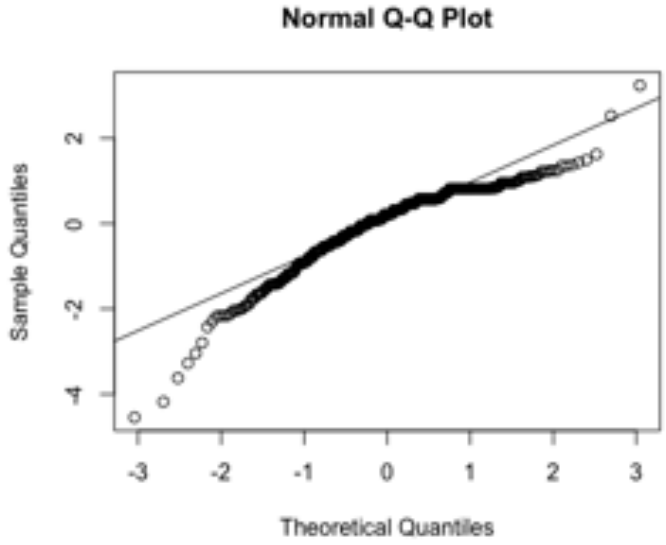

Figure 4.7. QQ-plot for residuals after fitting model (4.1) for the gain in *Effort* scores.

Chapter 5
Conclusion and Future Work

This dissertation describes three projects where statistical models, designs, or surveys were developed or adapted to address research questions related to animal movement or statistics education. While animal movement modeling and statistics education research are distinct in many ways, they both have increased in importance in recent years. Rapid advances in data availability due to technological progress has impacted both fields: statistical methods for analyzing animal movement data have become more relevant as it becomes easier to collect and store location data, and statistics education is becoming essential to address the growing need to analyze data in a range of industries.

Wildlife telemetry data is typically recorded at regular time intervals, and although those time intervals are getting shorter, telemetry units are being developed for increasingly smaller animals. Battery life is still a limitation, as is computational complexity when fitting models to extremely high resolution movement data. Thus, sampling and sub-sampling will always be relevant in animal movement studies. In Chapter 2, we proposed a novel sampling design called lattice and random intermediate point (LARI) which combines regular and irregular time intervals. We showed through two data examples and one simulation example that LARI sampling is superior to regular sampling in at least some cases.

While the results described in Chapter 2 imply that regular sampling may not always be the appropriate choice for animal movement studies, we do not suggest that LARI sampling is always the best choice. Future work in this area should examine optimal sampling for animal movement which should also be informed by input from wildlife experts regarding species behavior and the research questions. Tracking units are also becoming more diverse with regard to the data they collect, and the addition of covariates such as proper acceleration and temperature should also be considered.

Many animal populations display partial migration, meaning seasonal migration is

only present in a subset of the population. This presents challenges for conventional modeling methods which require categorization of movement paths into distinct groups. In Chapter 3, we presented a more flexible approach which can account for less stereotypical individuals. We also showed that our approach using varying coefficients produces more realistic simulations of stereotypical behavior than a latent-state modeling approach.

While the flexible model presented in Chapter 3 is useful at the individual level, it is important to make inference at the population level to inform species conservation management efforts. Since the flexible model we describe can be fit to a wide range of movement behaviors without significant model adjustments, it could be included in a structured population level model which incorporates covariates to model the attractor locations. The flexibility of this approach would be especially useful, for example, for species whose behavior is changing within a subgroup or adapting to a changing climate.

Students' attitudes toward the subject being taught have been identified by researchers as an important metric to inform teaching practice. While attitudes toward statistics have received enough attention in statistics education that many survey tools purporting to assess attitudes toward statistics have been developed, many students learn elementary probability before introductory statistics. In Chapter 4, we develop the Survey of Probability Attitudes (SPA) by adapting an existing tool for assessment of attitudes toward statistics. We distributed the SPA at the beginning and end of the semester to students in a probability course, and we distributed a version of the SPA to students who were not enrolled in a probability course. We used the resulting dataset to assess the validity of the instrument and analyzed the scores from both groups of students. We found a large effect of regression to the mean which was present in both groups.

In Chapter 4, we found strong evidence of the validity of the SPA, but changes in attitudes were difficult to identify. Despite the caveats we pointed out including the timing of the study taking place during a global pandemic and remote learning, we expect there were some changes to students' attitudes toward probability as a result of their probability course. Future iterations should be developed to improve the ability to identify changes in students' attitudes, and further research should also examine the motivations and beliefs influencing students attitudes.

My work thus far has tackled applied problems from the fields of ecology and statistics education, and I expect my future work to originate from a variety of sources as well. The projects described in this dissertation were collaborative efforts with statisticians, wildlife biologists, and educators. I anticipate working collaboratively with statisticians and non-statisticians throughout my career.

Bibliography

American Statistical Association (2000). 2000 curriculum guidelines for undergraduate programs in statistical science.

Ball, J. P., Nordengren, C., Wallin, K. (2001). Partial migration by large ungulates: characteristics of seasonal moose alces alces ranges in northern sweden. *Wildlife Biology* 7 (3), 39–47.

Bateiha, S., Marchionda, H., Autin, M. (2020). Teaching style and attitudes: a comparison of two collegiate introductory statistics classes. *Journal of Statistics Education* 28 (2), 154–164.

Bellhouse, D. (1977). Some optimal designs for sampling in two dimensions. *Biometrika* 64 (3), 605–611.

Berger, J. (2004). The last mile: how to sustain long-distance migration in mammals. *Conservation Biology* 18 (2), 320–331.

Bode, N. W., Franks, D. W., Wood, A. J., Piercy, J. J., Croft, D. P., Codling, E. A. (2012). Distinguishing social from nonsocial navigation in moving animal groups. *The American Naturalist* 179 (5), 621–632.

Brillinger, D. R., Preisler, H. K., Ager, A. A., Kie, J. (2012). The use of potential functions in modelling animal movement. In: Selected Works of David Brillinger. Springer, pp. 385–409.

Brown, J. L., Bedrosian, B., Bell, D. A., Braham, M. A., Cooper, J., Crandall, R. H., DiDonato, J., Domenech, R., Duerr, A. E., Katzner, T. E., et al. (2017). Patterns of spatial distribution of golden eagles across north america: how do they fit into existing landscape-scale mapping systems? *Journal of Raptor Research* 51 (3), 197–215.

Budé, L., van de Wiel, M. W., Imbos, T., Berger, M. P. (2012). The effect of guiding questions on students' performance and attitude towards statistics. *British journal of educational psychology* 82 (2), 340–359.

Cagnacci, F., Focardi, S., Ghisla, A., Van Moorter, B., Merrill, E. H., Gurarie, E., Heurich, M., Mysterud, A., Linnell, J., Panzacchi, M., et al. (2016). How many routes lead to migration? comparison of methods to assess and characterize migratory movements. *Journal of Animal Ecology* 85 (1), 54–68.

Chance, B., Wong, J., Tintle, N. (2016). Student performance in curricula centered on simulation-based inference: A preliminary report. *Journal of Statistics Education* 24 (3), 114–126.

Chapman, B. B., Brönmark, C., Nilsson, J.-Å., Hansson, L.-A. (2011). The ecology and evolution of partial migration. *Oikos* 120 (12), 1764–1775.

Chester, C. C. (2012). *Conservation across borders: biodiversity in an interdependent world*. Island Press.

Cole, J. D. (2010). *The impact of interactive whiteboards on students' attitudes toward applied mathematics*. Nipissing University, Faculty of Education.

Conner, M. M., Miller, M. W. (2004). Movement patterns and spatial epidemiology of a prion disease in mule deer population units. *Ecological Applications* 14 (6), 1870–1881.

Cook, D. A., Hatala, R. (2016). Validation of educational assessments: a primer for simulation and beyond. *Advances in Simulation* 1 (1), 1–12.

Craiu, R. V., Rosenthal, J. S. (2014). Bayesian computation via markov chain monte carlo. *Annual Review of Statistics and Its Application* 1 (1), 179–201.
URL `https://doi.org/10.1146/annurev-statistics-022513-115540`

DeVaney, T. A. (2010). Anxiety and attitude of graduate students in on-campus vs. online statistics courses. *Journal of Statistics Education* 18 (1).

Diggle, P., Lophaven, S. (2006). Bayesian geostatistical design. *Scandinavian Journal of Statistics* 33 (1), 53–64.
URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9469.2005.00469.x`

Eisenhauer, E., Hanks, E. (2020). A lattice and random intermediate point sampling design for animal movement. *Environmetrics*, e2618.

Eisenhauer, E., Hanks, E., Beckman, M., Murphy, R., Miller, T., Katzner, T. (2022). A flexible movement model for partially migrating species. *Spatial Statistics*, 100637.
URL `https://www.sciencedirect.com/science/article/pii/S2211675322000252`

Festa-Bianchet, M., Apollonio, M. (Eds.) (2003). *Animal Behavior and Wildlife Conservation*. Island Press.

Forester, J. D., Ives, A. R., Turner, M. G., Anderson, D. P., Fortin, D., Beyer, H. L., Smith, D. W., Boyce, M. S. (2007). State-space models link elk movement patterns to landscape characteristics in yellowstone national park. *Ecological Monographs* 77 (2), 285–299.

Fortin, M.-J., Drapeau, P., Legendre, P. (1990). Spatial autocorrelation and sampling design in plant ecology. In: Progress in theoretical vegetation science. Springer, pp. 209–222.

Fullman, T. J., Person, B. T., Prichard, A. K., Parrett, L. S. (2021). Variation in winter site fidelity within and among individuals influences movement behavior in a partially migratory ungulate. *PloS one* 16 (9), e0258128.

Geweke, J. F. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Staff Report 148, Federal Reserve Bank of Minneapolis. URL `https://ideas.repec.org/p/fip/fedmsr/148.html`

Griffith, J. D., Adams, L. T., Gu, L. L., Hart, C. L., Nichols-Whitehead, P. (2012). Students'attitudes toward statistics across the disciplines: A mixed-methods approach. *Statistics Education Research Journal* 11 (2).

Gundlach, E., Richards, K. A. R., Nelson, D., Levesque-Bristol, C. (2015). A comparison of student attitudes, statistical reasoning, performance, and perceptions for web-augmented traditional, fully online, and flipped sections of a statistical literacy class. *Journal of Statistics Education* 23 (1).

Haario, H., Saksman, E., Tamminen, J. (04 2001). An adaptive metropolis algorithm. *Bernoulli* 7 (2), 223–242.
URL `https://projecteuclid.org:443/euclid.bj/1080222083`

Hanks, E. M., Johnson, D. S., Hooten, M. B. (2017). Reflected stochastic differential equation models for constrained animal movement. *Journal of Agricultural, Biological, and Environmental Statistics* 22 (3).

Hooten, M. B., Buderman, F. E., Brost, B. M., Hanks, E. M., Ivan, J. S. (2016). Hierarchical animal movement models for population-level inference. *Environmetrics* 27 (6), 322–333.

Horton, N., Chance, B., Cohen, S., Grimshaw, S., Hardin, J., Hesterberg, T., Hoerl, R., Malone, C., Nichols, R., Nolan, D. (2014). Curriculum guidelines for undergraduate programs in statistical science. *American Statistical Association. Available online at http://www. amstat. org/education/curriculumguidelines. cfm*.

Kane, M. T. (2006). Validation. *Educational measurement* 4 (2), 17–64.

Kareiva, P., Shigesada, N. (1983). Analyzing insect movement as a correlated random walk. *Oecologia* 56 (2-3), 234–238.

Khavenson, T., Orel, E., Tryakshina, M. (2012). Adaptation of survey of attitudes towards statistics (sats 36) for russian sample. *Procedia-Social and Behavioral Sciences* 46, 2126–2129.

Kiekkas, P., Panagiotarou, A., Malja, A., Tahirai, D., Zykai, R., Bakalis, N., Stefanopoulos, N. (2015). Nursing students' attitudes toward statistics: Effect of a biostatistics course and association with examination performance. *Nurse Education Today* 35 (12), 1283–1288.

Kloeden, P. E., Platen, E. (1992). Stochastic differential equations. In: Numerical Solution of Stochastic Differential Equations. Springer, pp. 103–160.

Kloeden, P. E., Platen, E. (2013). *Numerical solution of stochastic differential equations*. Vol. 23. Springer Science & Business Media.

Marra, G., Wood, S. N. (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis* 55 (7), 2372–2387.

McClintock, B. T., Michelot, T. (2018). momentuhmm: R package for generalized hidden markov models of animal movement. *Methods in Ecology and Evolution* 9 (6), 1518–1530.

McDuie, F., Casazza, M. L., Overton, C. T., Herzog, M. P., Hartman, C. A., Peterson, S. H., Feldheim, C. L., Ackerman, J. T. (2019). Gps tracking data reveals daily spatio-temporal movement patterns of waterfowl. *Movement ecology* 7 (1), 6.

Michelot, T., Glennie, R., Harris, C., Thomas, L. (2020). Varying-coefficient stochastic differential equations with applications in ecology. *arXiv preprint arXiv:2008.09111*.

Michelot, T., Langrock, R., Patterson, T. A. (2016). movehmm: an r package for the statistical modelling of animal movement data using hidden markov models. *Methods in Ecology and Evolution* 7 (11), 1308–1315.

Millar, A. M., Schau, C. (2010). Assessing students' attitudes: the good, the bad, and the ugly. In: Proceedings of the Joint Statistical Meetings. Citeseer, pp. 1133–1143.

Mills, K. J., Patterson, B. R., Murray, D. L. (2006). Effects of variable sampling frequencies on gps transmitter efficiency and estimated wolf home range size and movement distance. *Wildlife Society Bulletin* 34 (5), 1463–1469.

Millspaugh, J., Marzluff, J. M. (2001). *Radio tracking and animal populations*. Academic Press.

Modlmeier, A. P., Colman, E., Hanks, E. M., Bringenberg, R., Bansal, S., Hughes, D. P. (2019). Ant colonies maintain social homeostasis in the face of decreased density. *eLife* 8, e38473.

Morales, J. M., Moorcroft, P. R., Matthiopoulos, J., Frair, J. L., Kie, J. G., Powell, R. A., Merrill, E. H., Haydon, D. T. (2010). Building the bridge between animal movement and population dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1550), 2289–2301.

Mueller, T., Fagan, W. F. (2008). Search and navigation in dynamic environments–from individual behaviors to population distributions. *Oikos* 117 (5), 654–664.

Nelson, E. (1967). The ornstein-uhlenbeck theory of brownian motion. In: Dynamical Theories of Brownian Motion. Princeton University Press, Ch. 9, pp. 45–52.

Oliver, M. A., Webster, R. (1986). Combining nested and linear sampling for determining the scale and form of spatial variation of regionalized variables. *Geographical analysis* 18 (3), 227–242.

Paciorek, E. (1997). Mathematics and science attitude inventory.
URL `https://oerl.sri.com/instruments/up/studsurv/instr129.html`

Parlin, A. F., Nardone, J. A., Dougherty, J. K., Rebein, M., Safi, K., Schaeffer, P. J. (07 2018). Activity and movement of free-living box turtles are largely independent of ambient and thermal conditions. *Movement Ecology*.

Patterson, T. A., Parton, A., Langrock, R., Blackwell, P. G., Thomas, L., King, R. (2017). Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis* 101 (4), 399–438.

Pearl, D., Garfield, J., delMas, R., Groth, R., Kaplan, J., McGowan, H., Lee, H. (2012). Connecting research to practice in a culture of assessment for introductory college-level statistics. *Recuperado de: www. causeweb. org/research/guidelines/ResearchReport_Dec_2012. pdf*.

Pirotta, E., Katzner, T., Miller, T. A., Duerr, A. E., Braham, M. A., New, L. (2018). State-space modelling of the flight behaviour of a soaring bird provides new insights to migratory strategies. *Functional ecology* 32 (9), 2205–2215.

Poessel, S. A., Bloom, P. H., Braham, M. A., Katzner, T. E. (2016). Age-and season-specific variation in local and long-distance movement behavior of golden eagles. *European Journal of Wildlife Research* 62 (4), 377–393.

Preisler, H. K., Ager, A. A., Wisdom, M. J. (2013). Analyzing animal movement patterns using potential functions. *Ecosphere* 4 (3), 1–13.

Ramirez, C., Schau, C., Emmioglu, E. (2012). The importance of attitudes in statistics education. *Statistics education research journal* 11 (2).

Richardson, P. L., Wakefield, E. D., Phillips, R. A. (Mar 2018). Flight speed and performance of the wandering albatross with respect to wind. *Movement Ecology* 6 (1), 3.
URL `https://doi.org/10.1186/s40462-018-0121-9`

Roberts, D. M., Bilderback, E. W. (1980). Reliability and validity of a statistics attitude survey. *Educational and psychological measurement* 40 (1), 235–238.

Roberts, G. O., Rosenthal, J. S. (2009). Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics* 18 (2), 349–367.
URL https://doi.org/10.1198/jcgs.2009.06134

Roeleke, M., Teige, T., Hoffmeister, U., Klingler, F., Voigt, C. C. (2018). Aerial-hawking bats adjust their use of space to the lunar cycle. *Movement ecology* 6 (1), 11.

Russell, J. C., Hanks, E. M., Haran, M., Hughes, D. (06 2018). A spatially varying stochastic differential equation model for animal movement. *Ann. Appl. Stat.* 12 (2), 1312–1331.
URL https://doi.org/10.1214/17-AOAS1113

Russell, J. C., Hanks, E. M., Modlmeier, A. P., Hughes, D. P. (2017). Modeling collective animal movement through interactions in behavioral states. *Journal of Agricultural, Biological and Environmental Statistics* 22 (3), 313–334.

Sarikaya, E. E., Ok, A., Aydin, Y. C., Schau, C. (2018). Turkish version of the survey of attitudes toward statistics: Factorial structure invariance by gender. *International Journal of Higher Education* 7 (2), 121–127.

Schau, C. (2020). Student attitude surveys and online educational consulting.
URL https://www.evaluationandstatistics.com/

Schau, C., Emmioğlu, E. (2012). Do introductory statistics courses in the united states improve students'attitudes? *Statistics Education Research Journal* 11 (2).

Schau, C., Stevens, J., Dauphinee, T. L., Vecchio, A. D. (1995). The development and validation of the survey of antitudes toward statistics. *Educational and psychological measurement* 55 (5), 868–875.

Schloss, C. A., Nuñez, T. A., Lawler, J. J. (2012). Dispersal will limit ability of mammals to track climate change in the western hemisphere. *Proceedings of the National Academy of Sciences* 109 (22), 8606–8611.
URL http://www.pnas.org/content/109/22/8606

Shepard, D. B., Kuhns, A. R., Dreslik, M., Phillips, C. A. (2008). Roads as barriers to animal movement in fragmented landscapes. *Animal Conservation* 11 (4), 288–296.

Spitz, D. B., Hebblewhite, M., Stephenson, T. R. (2017). 'migrater': extending model-driven methods for classifying and quantifying animal movement behavior. *Ecography* 40 (6), 788–799.

Swanson, T., VanderStoep, J., Tintle, N. (2014). Student attitudes toward statistics from a randomization-based curriculum. In: International Conference on Teaching Statistics, Flagstaff, AZ. http://icots. info/icots/9/proceedings/pdfs/ICOTS9_1F1_SWANSON. pdf.

Tan, C.-K., Harji, M. B., Lau, S.-H. (2011). Fostering positive attitude in probability learning using graphing calculator. *Computers & Education* 57 (3).

Tempelaar, D., Nijhuis, J. (2007). Commonalities in attitudes and beliefs toward different academic subjects. In: The challenges of educating people to lead in a challenging world. Springer, pp. 225–249.

Tierson, W. C., Mattfeld, G. F., Sage, R. W., Behrend, D. F. (1985). Seasonal movements and home ranges of white-tailed deer in the adirondacks. *The Journal of Wildlife Management* 49 (3), 760–769.
URL `http://www.jstor.org/stable/3801708`

Tomkiewicz, S. M., Fuller, M. R., Kie, J. G., Bates, K. K. (2010). Global positioning system and associated technologies in animal behaviour and ecological research. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1550), 2163–2176.

Trizano-Hermosilla, I., Alvarado, J. M. (2016). Best alternatives to cronbach's alpha reliability in realistic conditions: congeneric and asymmetrical measurements. *Frontiers in psychology* 7, 769.

Vanhoof, S., Kuppens, S., Castro Sotos, A. E., Verschaffel, L., Onghena, P. (2011). Measuring statistics attitudes: Structure of the survey of attitudes toward statistics. *Statistics Education Research Journal* 10 (1).

Warton, D. I., Shepherd, L. C. (2010). Poisson point process models solve the" pseudo-absence problem" for presence-only data in ecology. *The Annals of Applied Statistics*, 1383–1402.

Weimerskirch, H., Bonadonna, F., Bailleul, F., Mabille, G., Dell'Omo, G., Lipp, H.-P. (2002). Gps tracking of foraging albatrosses. *Science* 295 (5558), 1259–1259.

Wester, E. R., Walsh, L. L., Arango-Caro, S., Callis-Duehl, K. L. (2021). Student engagement declines in stem undergraduates during covid-19–driven remote learning. *Journal of microbiology & biology education* 22 (1), ev22i1–2385.

Wijeyakulasuriya, D. A., Hanks, E. M., Shaby, B. A., Cross, P. C. (2019). Extreme value-based methods for modeling elk yearly movements. *Journal of Agricultural, Biological and Environmental Statistics* 24 (1), 73–91.

Wise, S. L. (1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and psychological measurement* 45 (2), 401–405.

Wisecup, A. K. (2017). Take it or leave it: Students' attitudes about research methods. *Teaching Sociology* 45 (1), 73–79.

Wood, S. (2017). *Generalized Additive Models: An Introduction with R*, 2nd Edition. Chapman and Hall/CRC.

Zimmerman, D. L. (2006). Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics* 17 (6), 635–652. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/env.769`

Zucchini, W., MacDonald, I. L., Langrock, R. (2017). *Hidden Markov models for time series: an introduction using R*. CRC press.

Zucchini, W., Raubenheimer, D., MacDonald, I. L. (2008). Modeling time series of animal behavior by means of a latent-state model with feedback. *Biometrics* 64 (3), 807–815.

Vita

Elizabeth W. **Eisenhauer**

EDUCATION

PENNSYLVANIA STATE UNIVERSITY **University Park, PA**
Ph.D. in Statistics, GPA 3.84 / 4.00 2017 – August 2022 (Expected)

- **Dissertation:** Advances in Stochastic Models for Animal Movement and Assessment of Probability Attitudes
- **Co-Advisors:** Ephraim Hanks and Matthew Beckman

THE COLLEGE OF NEW JERSEY **Ewing, NJ**
B.A. in Mathematics with a Statistics specialization, GPA 3.77 / 4.00 2013 – 2017

- **Honors Thesis:** Structural Equation Modeling of Signaling Networks in Head and Neck Squamous Cell Carcinoma
- **Advisor:** Michael Ochs
- **Honors:** *magna cum laude*, Departmental Honors, Phi Beta Kappa, Pi Mu Epsilon Mathematics Honor Society

PUBLICATIONS

- **Eisenhauer, Elizabeth**, Ephraim Hanks, Matthew Beckman, Robert Murphy, Tricia Miller, and Todd Katzner. "A Flexible Movement Model for Partially Migrating Species. " *Spatial Statistics* (2022): 100637.
- Wijeyakulasuriya, Dhanushi A., **Elizabeth W. Eisenhauer**, Benjamin A. Shaby, and Ephraim M. Hanks. "Machine learning for modeling animal movement." *PloS one* 15.7 (2020): e0235750.
- **Eisenhauer, Elizabeth**, and Ephraim Hanks. "A lattice and random intermediate point sampling design for animal movement." *Environmetrics* (2020): e2618.

SELECTED AWARDS & HONORS

WINNER OF HARKNESS AWARD 2021
For showing excellence in teaching | Pennsylvania State University Department of Statistics

STUDENT AWARD FOR ORAL PRESENTATION (2ND PLACE) 2021
Modeling Yearly Patterns in Golden Eagle Movement | EURING Analytical Meeting & Workshop

RUNNER-UP FOR HARKNESS AWARD 2020
For showing excellence in teaching | Pennsylvania State University Department of Statistics

STUDENT PRIZE FOR CONTRIBUTED TALK 2020
A Lattice and Random Intermediate Point Sampling Design for Animal Movement | vISEC

VOLLMER-KLECKNER SCHOLARSHIP IN SCIENCE 2018 – 2019
Pennsylvania State University

DISTINGUISHED GRADUATE FELLOWSHIP 2017 – 2018
Pennsylvania State University

TEACHING EXPERIENCE

PENNSYLVANIA STATE UNIVERSITY **University Park, PA**
Instructor of Record

- STAT 401: Experimental Methods Spring & Fall 2021
- STAT 200: Elementary Statistics Summer 2020 & 2021
- MATH/STAT 318: Elementary Probability Fall 2019, Spring & Fall 2020