Advances in Stochastic Models for Animal Movement and Assessment of Attitudes Toward Probability

link to slides

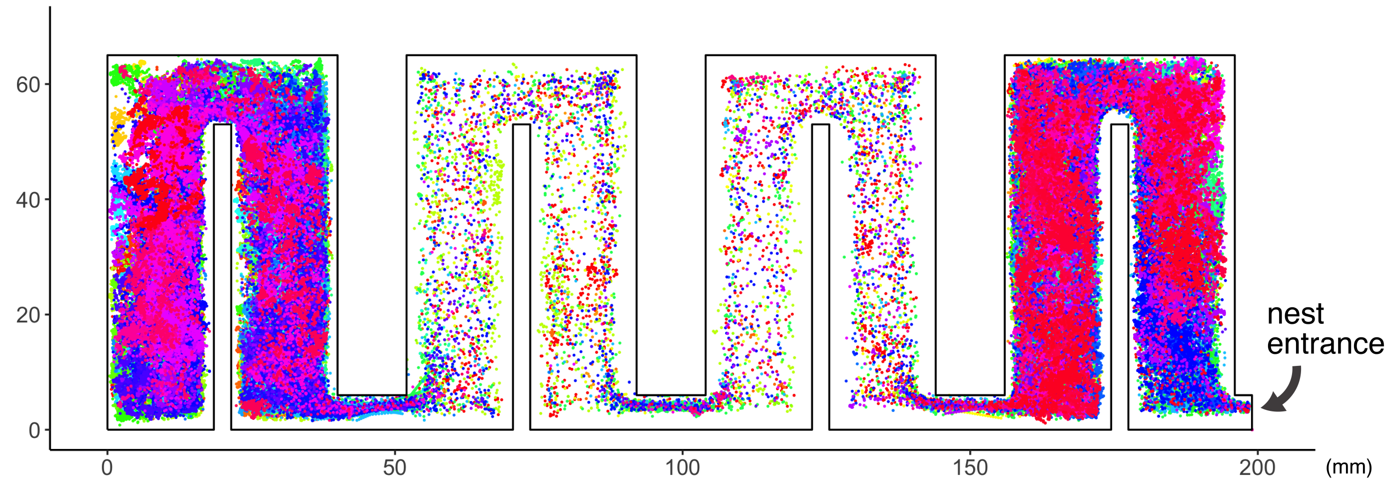
Elizabeth Eisenhauer

Co-advisors: Ephraim Hanks & Matthew Beckman

Outline

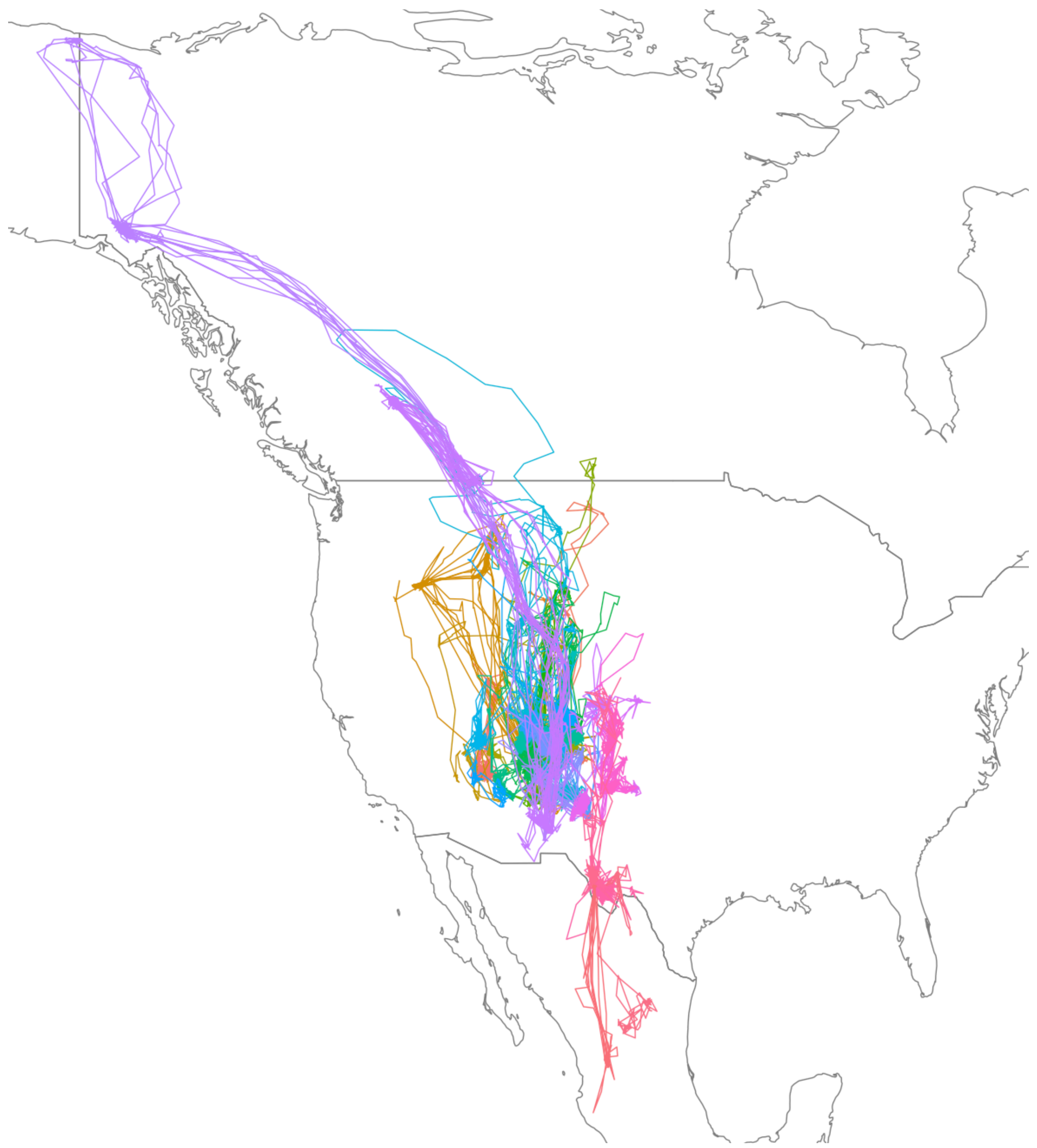
1

A Lattice and Random Intermediate Point Sampling Design for Animal Movement



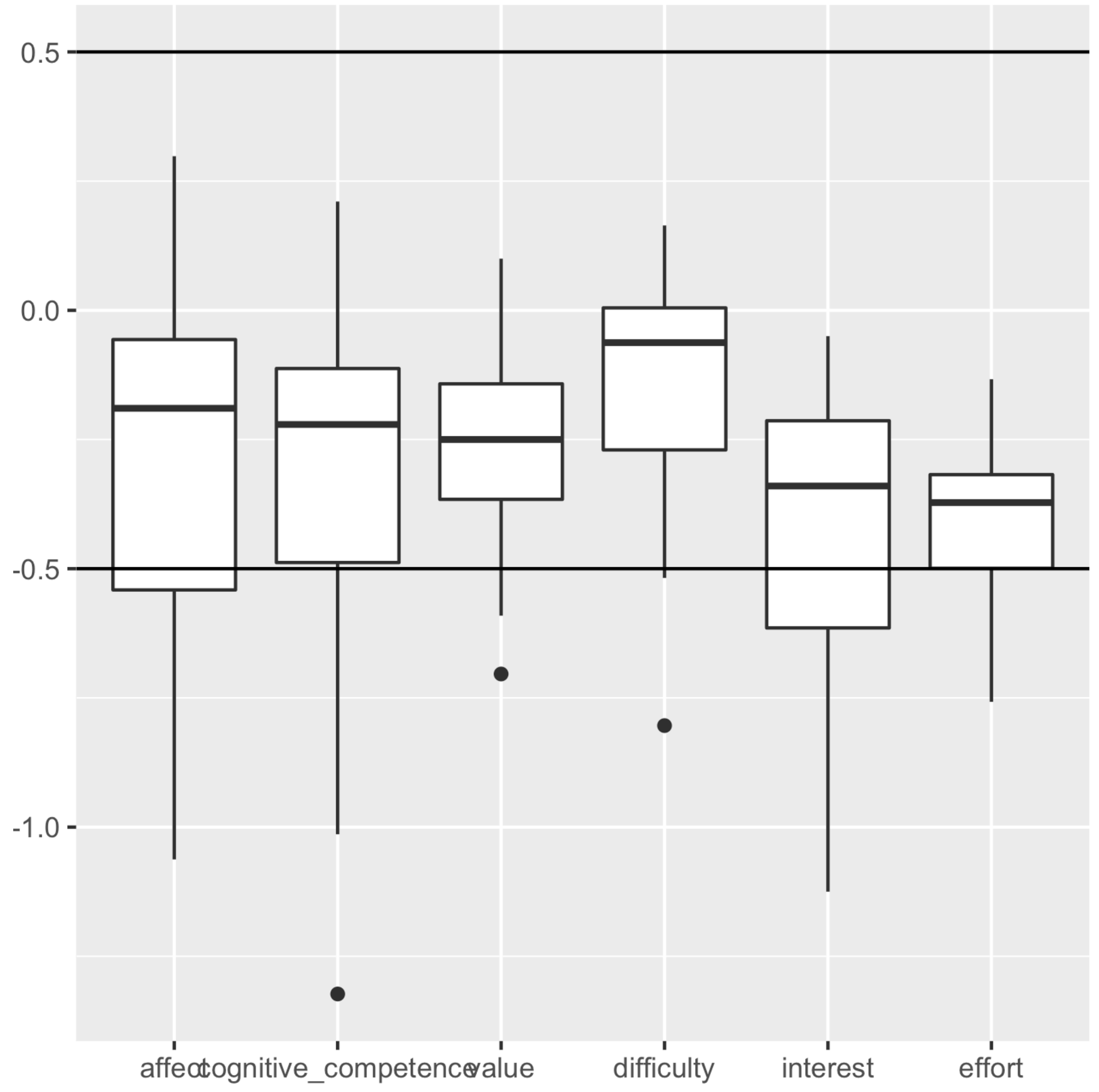
2

Modeling Yearly Patterns in Golden Eagle Movement



3

Survey of Attitudes toward Probability

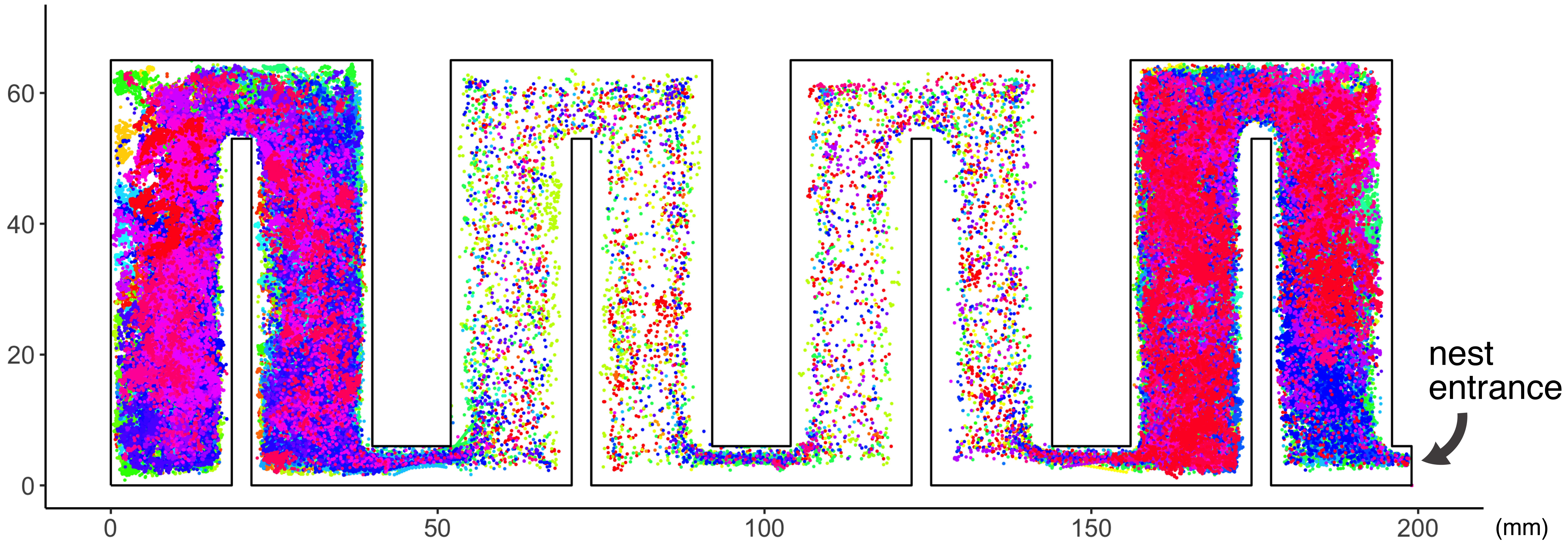


A Lattice and Random Intermediate Point (LARI) Sampling Design for Animal Movement

Hughes lab at Penn State collected **high resolution ant data,** but data collection was **time-consuming** (1000s of student-hours).

4 hours of movement data
78 ants
1 second intervals

The resulting dataset consists of **4 hours** of movement data for **78 ants** at 1 second intervals (14,401 observations per ant).



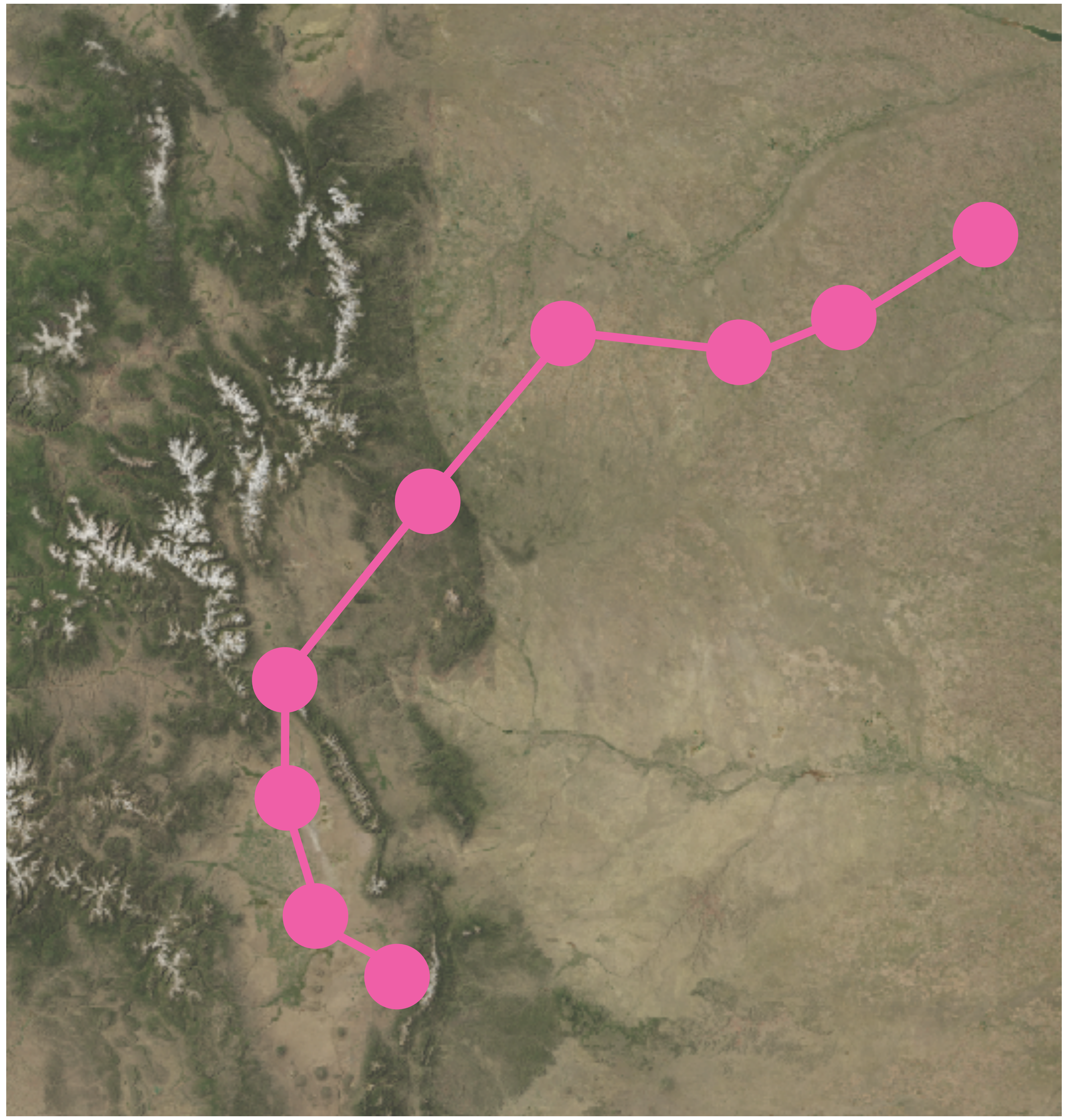
We were approached by the researchers with the **scenario**:

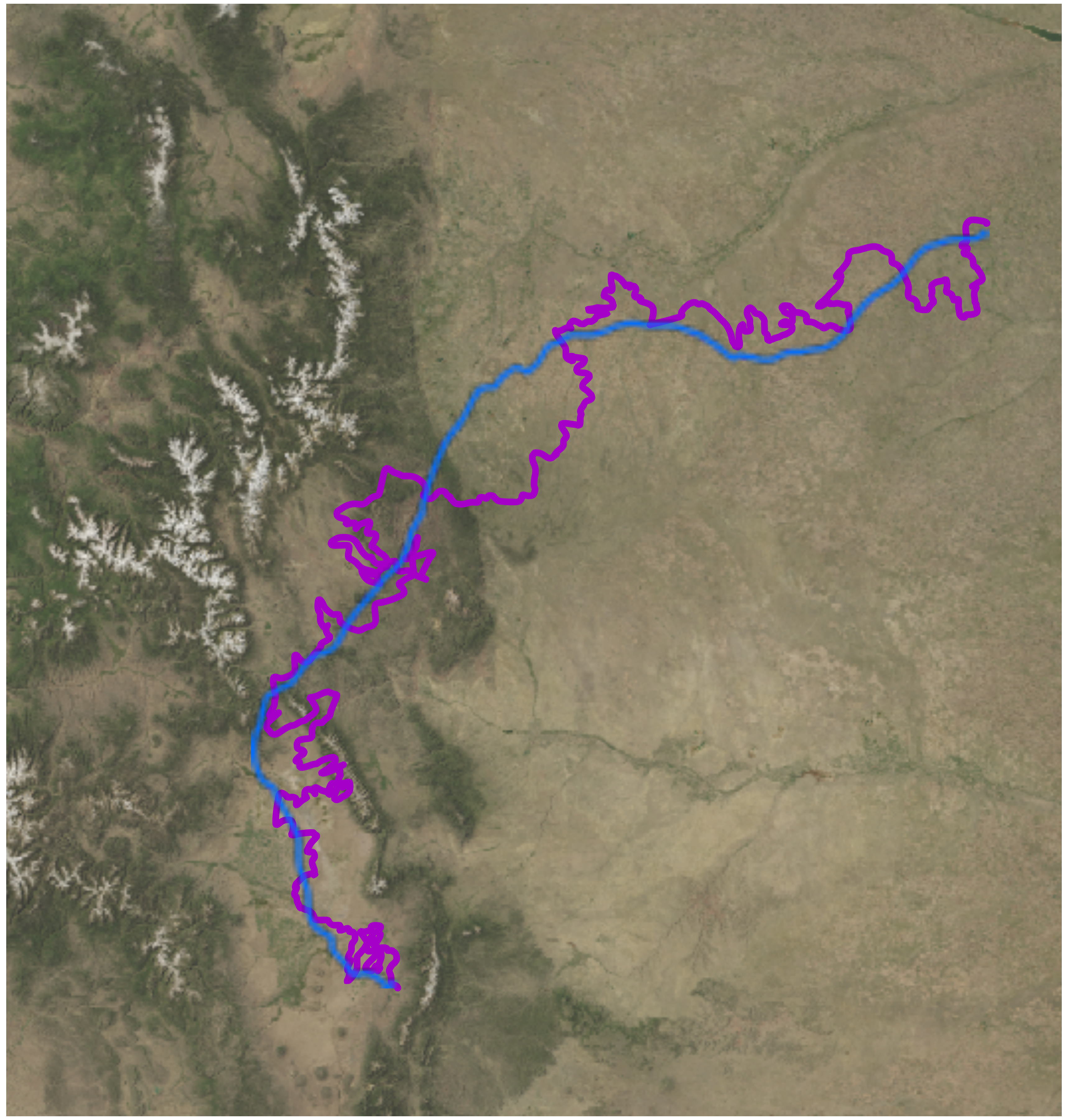
Next time, we will collect **lower resolution** data.

How should we do this to **minimize the loss of information** about movement behavior?

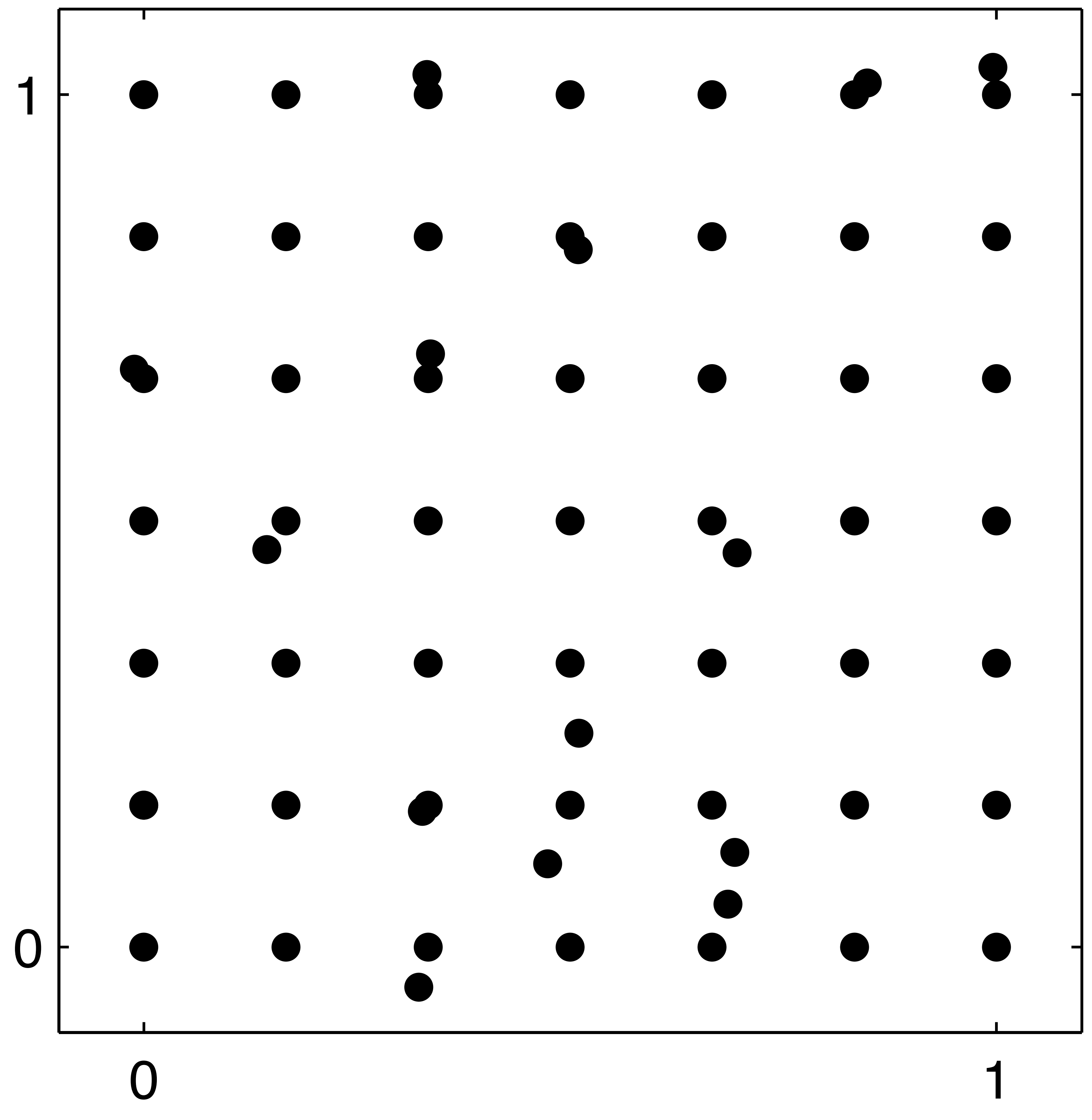
This question is **relevant to many researchers** collecting animal movement data.

Sampling at **regular time intervals can hide important information** about the speed and tortuosity of the path.





In geostatistics, researchers often adopt a **lattice plus close pairs** design over a lattice alone or a lattice and infill approach.

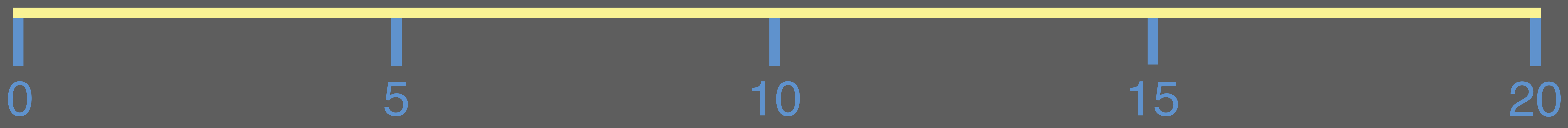


(Diggle and Lophaven, 2006)

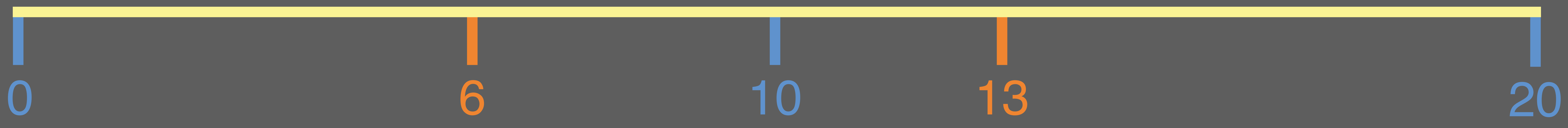
We **propose a sampling scheme** for animal telemetry data inspired by the lattice plus close pairs geostatistical design.

2 sampling designs:

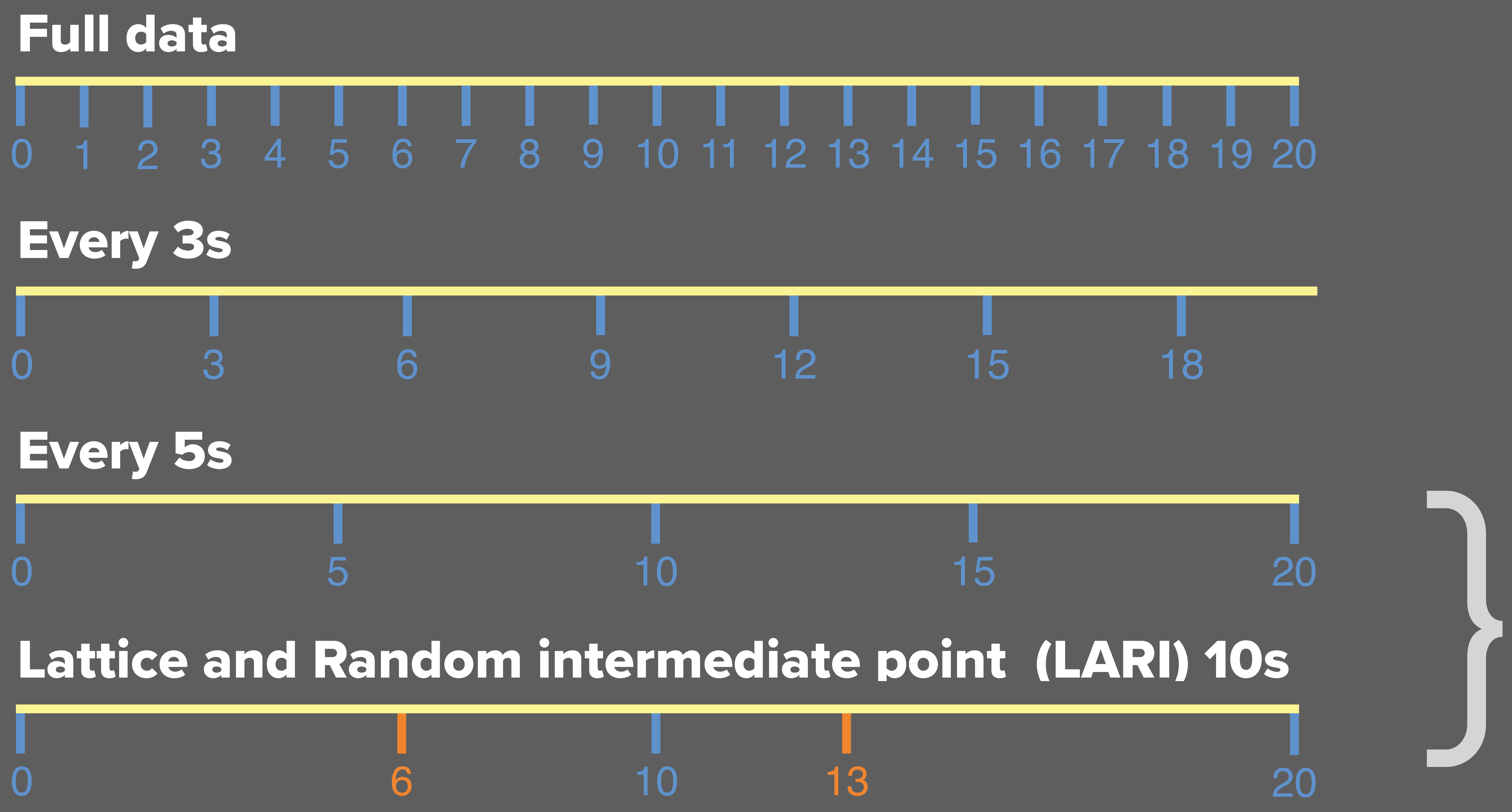
REGULAR



LATTICE AND RANDOM INTERMEDIATE POINT (LARI)



To compare regular and LARI sampling designs, we look at **4 subsamples** of the ant data.



Every 5s and LARI 10s have the same number of data points

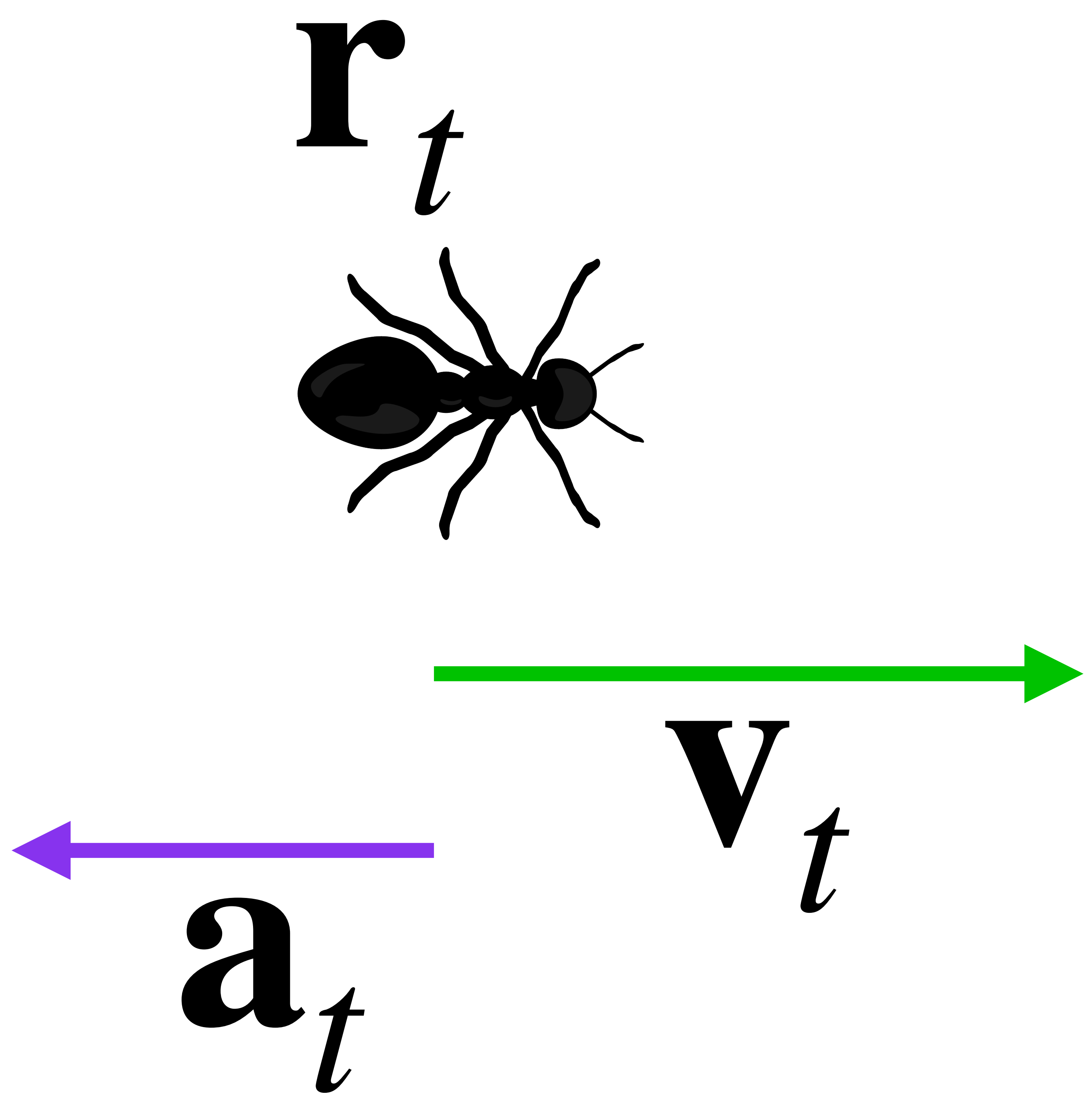
To **model movement at an individual level**, we use basic concepts from physics.

$\mathbf{r}_t =$ position

$\mathbf{v}_t =$ velocity

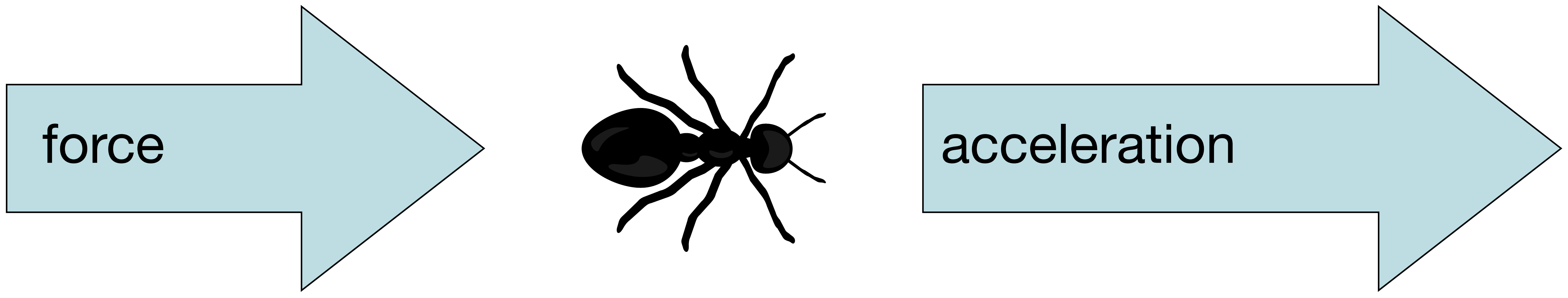
$\mathbf{a}_t =$ acceleration

This ant is slowing down



$$\mathbf{F}_t = m\mathbf{a}_t$$

So modeling acceleration is the same as **modeling "force"** acting on an animal.



The **2 main equations** for this model

The derivative of position with respect to time is velocity.

$$\boxed{\frac{d\mathbf{r}_t}{dt} = \mathbf{v}_t} \longleftrightarrow \boxed{d\mathbf{r}_t = \mathbf{v}_t dt}$$

The derivative of velocity with respect to time is acceleration.

$$\boxed{\frac{d\mathbf{v}_t}{dt} = \mathbf{a}_t} \longleftrightarrow \boxed{d\mathbf{v}_t = \mathbf{a}_t dt}$$

The **2 main equations** for this model

To model animal movement, we use

$$\boxed{d\mathbf{r}_t = \mathbf{v}_t \, dt}$$

and rewrite acceleration as a sum of forces

$$d\mathbf{v}_t = \underbrace{\boxed{\beta \, (\boldsymbol{\mu}(\mathbf{r}_t) - \mathbf{v}_t) \, dt}}_{\text{mean-reverting force}} + \underbrace{\boxed{c(\mathbf{r}_t) \, \mathbf{I} d\mathbf{w}_t}}_{\text{random force}}$$

Stochastic differential equation (SDE) model for animal movement

Data: $\mathbf{r}_t,\ t = 1, 2, \ldots, 14401$ for each ant

SDE model framework:

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

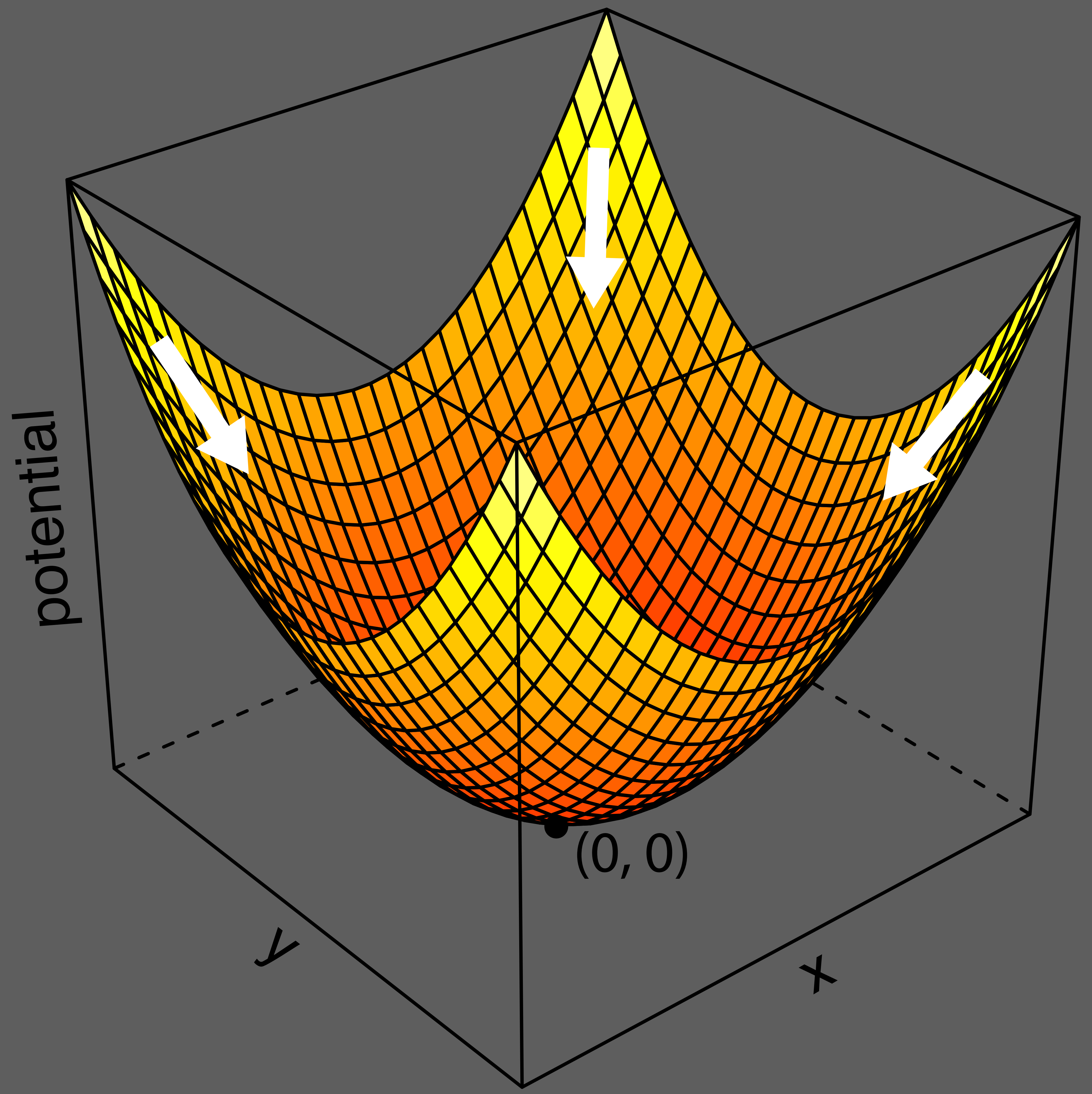
Utilizing motility and potential surfaces, define:

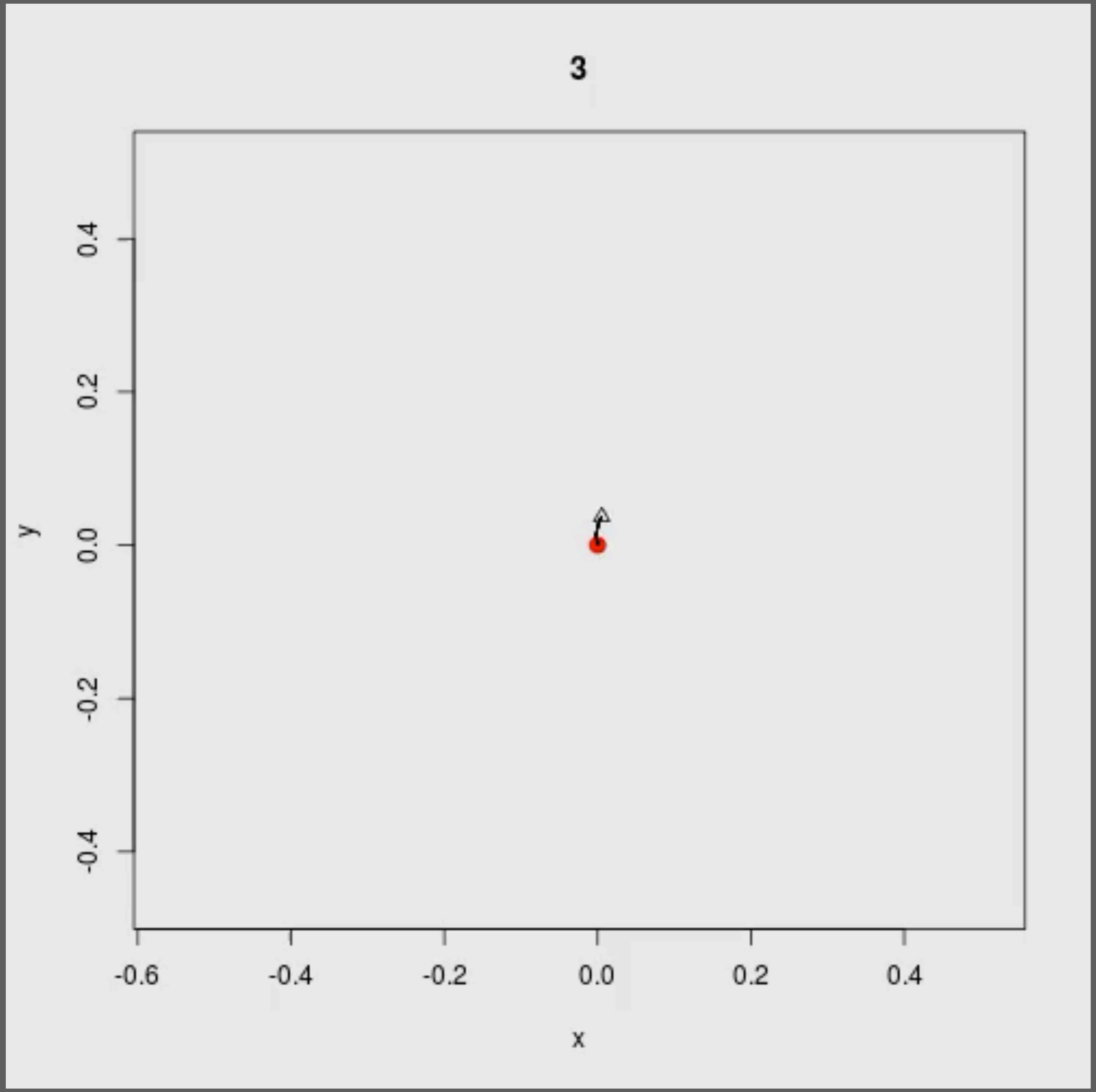
$\boldsymbol{\mu}(\mathbf{r}_t) = m(\mathbf{r}_t) \, [- \nabla p(\mathbf{r}_t)]$ (mean drift)

$c(\mathbf{r}_t) = \sigma \, m(\mathbf{r}_t)$ (magnitude of stochasticity)

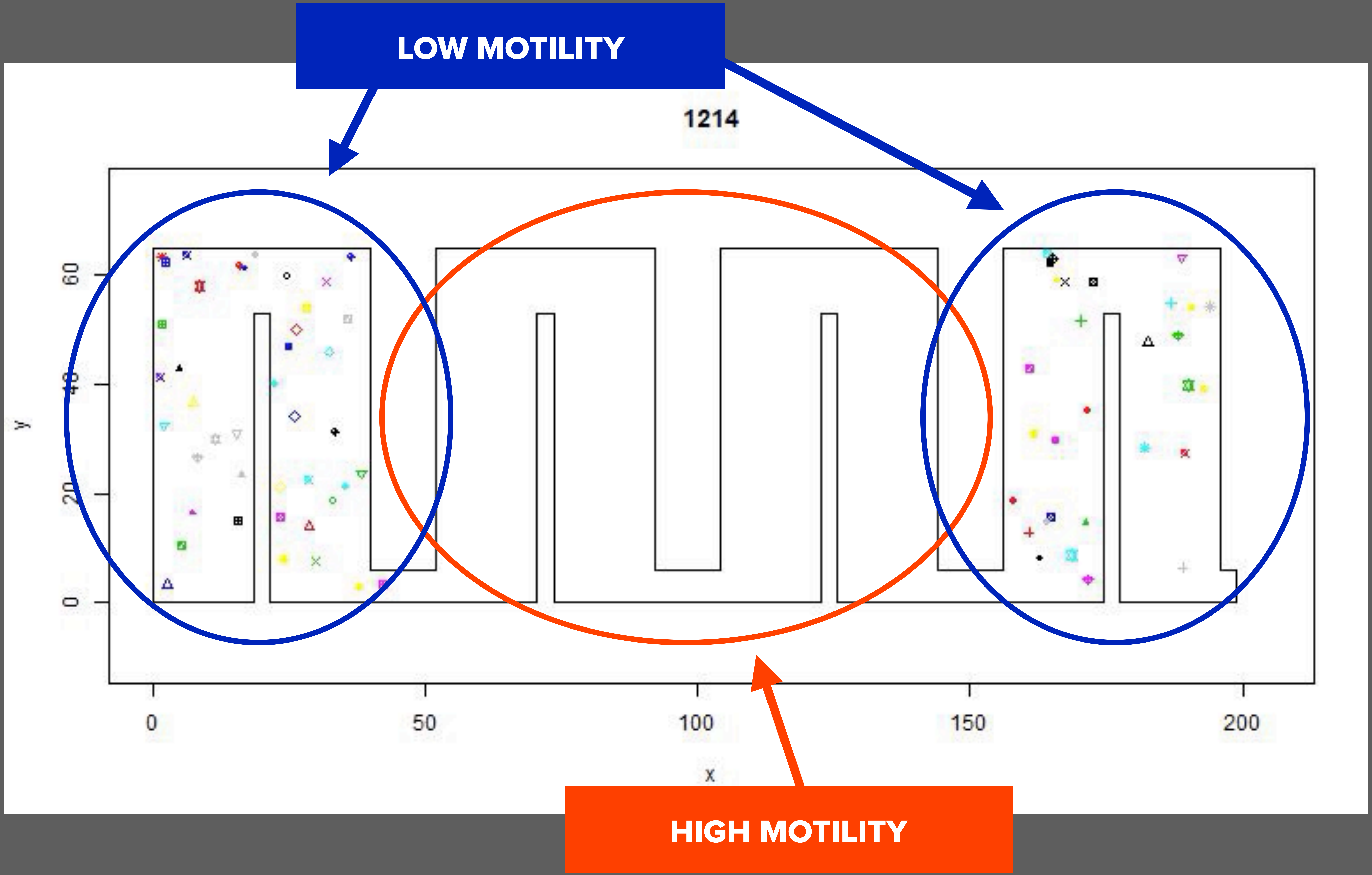
(James Russell et al., 2018)

We describe animal movement using a stochastic differential equation model with 2 parameters: **potential** and motility





We describe animal movement using a stochastic differential equation model with 2 parameters: potential and **motility**

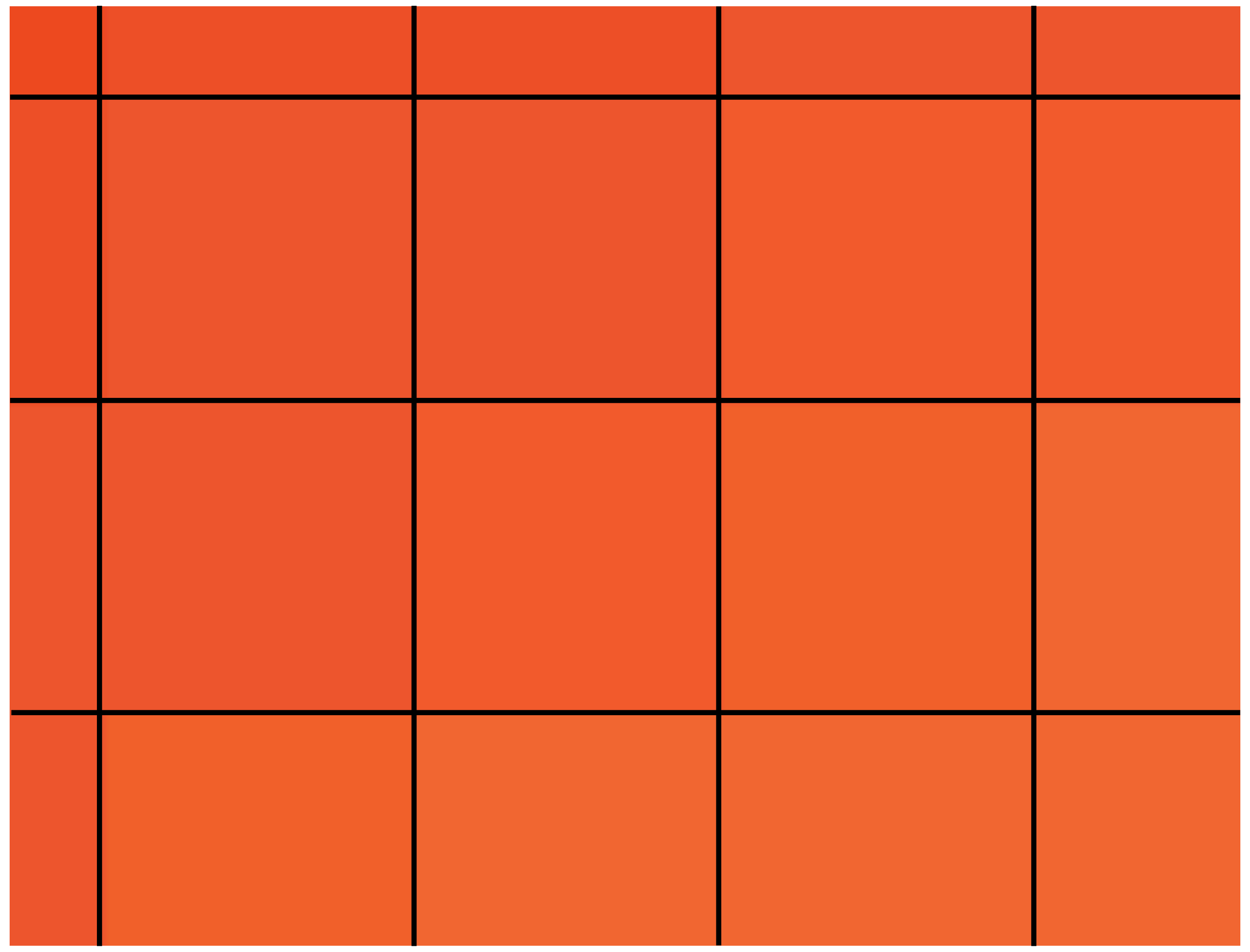


Spline expansion (degree 0, piecewise constant) of the motility and potential surfaces

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

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

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



Penalize the roughness of ***m*** and ***p***

Smoothness parameters are chosen with a holdout set.

Since we don't observe animal movement in continuous time, we **numerically approximate derivatives** (Euler-Maruyama method).

$$\left.\begin{aligned} \frac{d\mathbf{r}_\tau}{dt} &\approx \frac{\mathbf{r}_{\tau+1} - \mathbf{r}_\tau}{h_\tau} \\ \frac{d\mathbf{v}_\tau}{dt} &\approx \frac{\mathbf{v}_{\tau+1} - \mathbf{v}_\tau}{h_\tau} \approx \frac{\mathbf{r}_{\tau+2} - \mathbf{r}_{\tau+1}}{h_\tau h_{\tau+1}} - \frac{\mathbf{r}_{\tau+1} - \mathbf{r}_\tau}{h_\tau^2} \end{aligned}\right\}$$

Note that this works for data that is irregular in time.

where

- $\mathbf{r}_\tau = [x_\tau \ y_\tau]'$ is the position of ordered observation τ
- $\mathbf{v}_\tau$ is the (unobserved) velocity of observation τ
- h_τ is the change in time from observation τ to $\tau + 1$

Resulting in the **model equation**

$$\mathbf{r}_{\tau+2} = \left(1 + \frac{h_{\tau+1}}{h_\tau} - \beta h_{\tau+1}\right)\mathbf{r}_{\tau+1} + \left(\beta h_{\tau+1} - \frac{h_{\tau+1}}{h_\tau}\right)\mathbf{r}_\tau + \beta\ h_\tau h_{\tau+1}\ m(\mathbf{r}_\tau)\ [-\triangledown\, p(\mathbf{r}_\tau)] + \sigma\ m(\mathbf{r}_\tau)\ h_{\tau+1}h_\tau^2\ N(\mathbf{0}, I)$$

Recall:

- $\mathbf{r}_\tau = [x_\tau\ y_\tau]'$ is the position of ordered observation τ
- $\mathbf{v}_\tau$ is the (unobserved) velocity of observation τ
- h_τ is the change in time from observation τ to $\tau + 1$

$$\mathbf{r}_{\tau+2} = \left(1 + \frac{h_{\tau+1}}{h_\tau} - \beta h_{\tau+1}\right)\mathbf{r}_{\tau+1} + \left(\beta h_{\tau+1} - \frac{h_{\tau+1}}{h_\tau}\right)\mathbf{r}_\tau + \beta\ h_\tau h_{\tau+1}\ m(\mathbf{r}_\tau)\ [-\triangledown p(\mathbf{r}_\tau)] + \sigma\ m(\mathbf{r}_\tau)\ h_{\tau+1}h_\tau^2\ N(\mathbf{0}, I)$$

Goal – Estimate $\mathbf{m} \equiv [m_1 \ \ldots \ m_J]'$ and $\mathbf{p} \equiv [p_1 \ \ldots \ p_J]'$ for J grid cells with an iterative approach:

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

Computing time **~20 minutes** (single core)

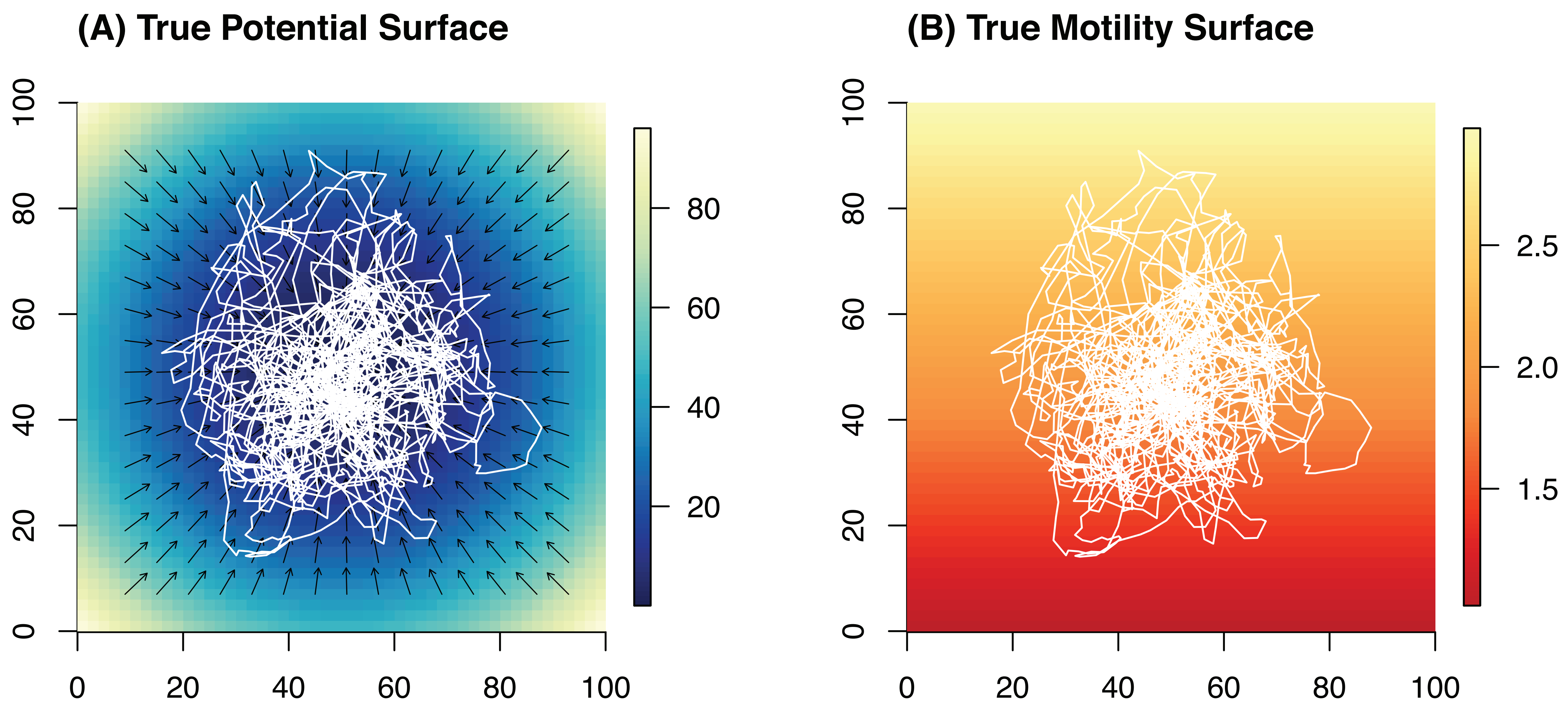
- 14,401 x 78 data points

Simulation study of the estimation procedure

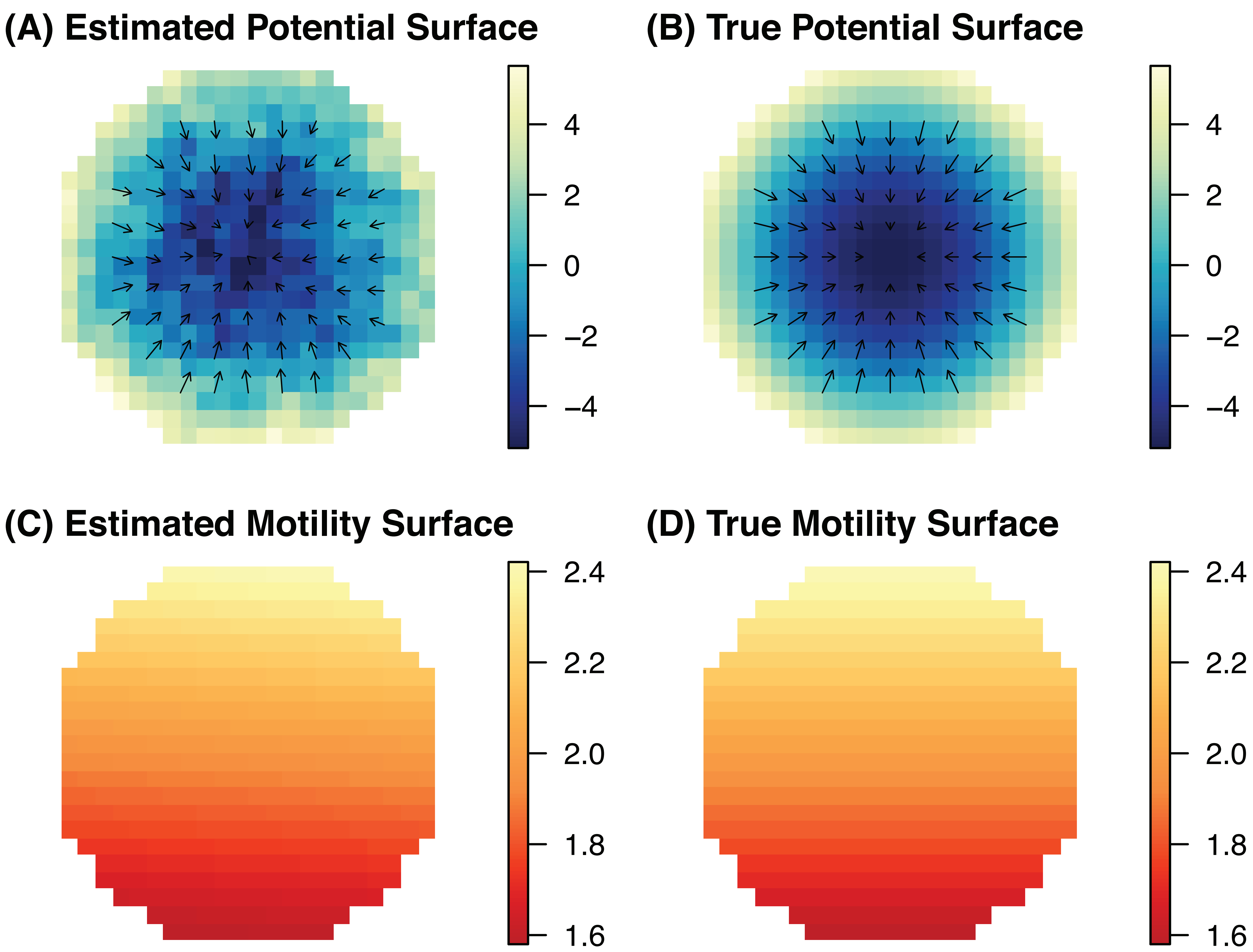
Step 2 of the estimation procedure entails estimation of the squared motility surface evaluated at each observation, i.e., $m^2(\mathbf{r}_\tau) = \mathrm{E}(\epsilon_\tau^2 h_\tau^{-1})$. We achieve this by estimating $\mathrm{E}[\log(m^2(\mathrm{r}_\tau))]$ and exponentiating the fitted values.

We recognize this may introduce bias and have conducted a simulation study **to quantify this bias**.

One of 500 simulations

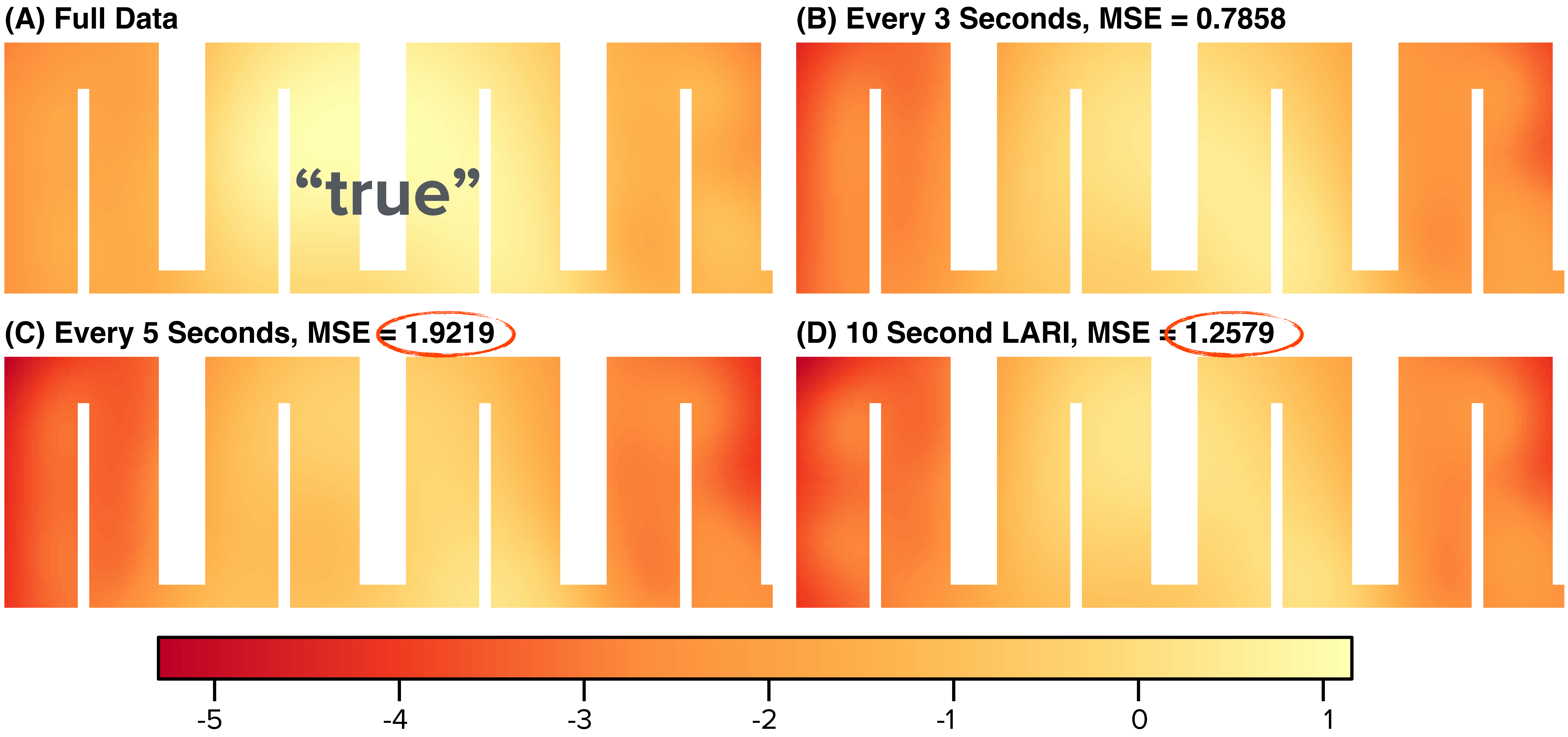


True and estimated surfaces for one randomly selected simulation

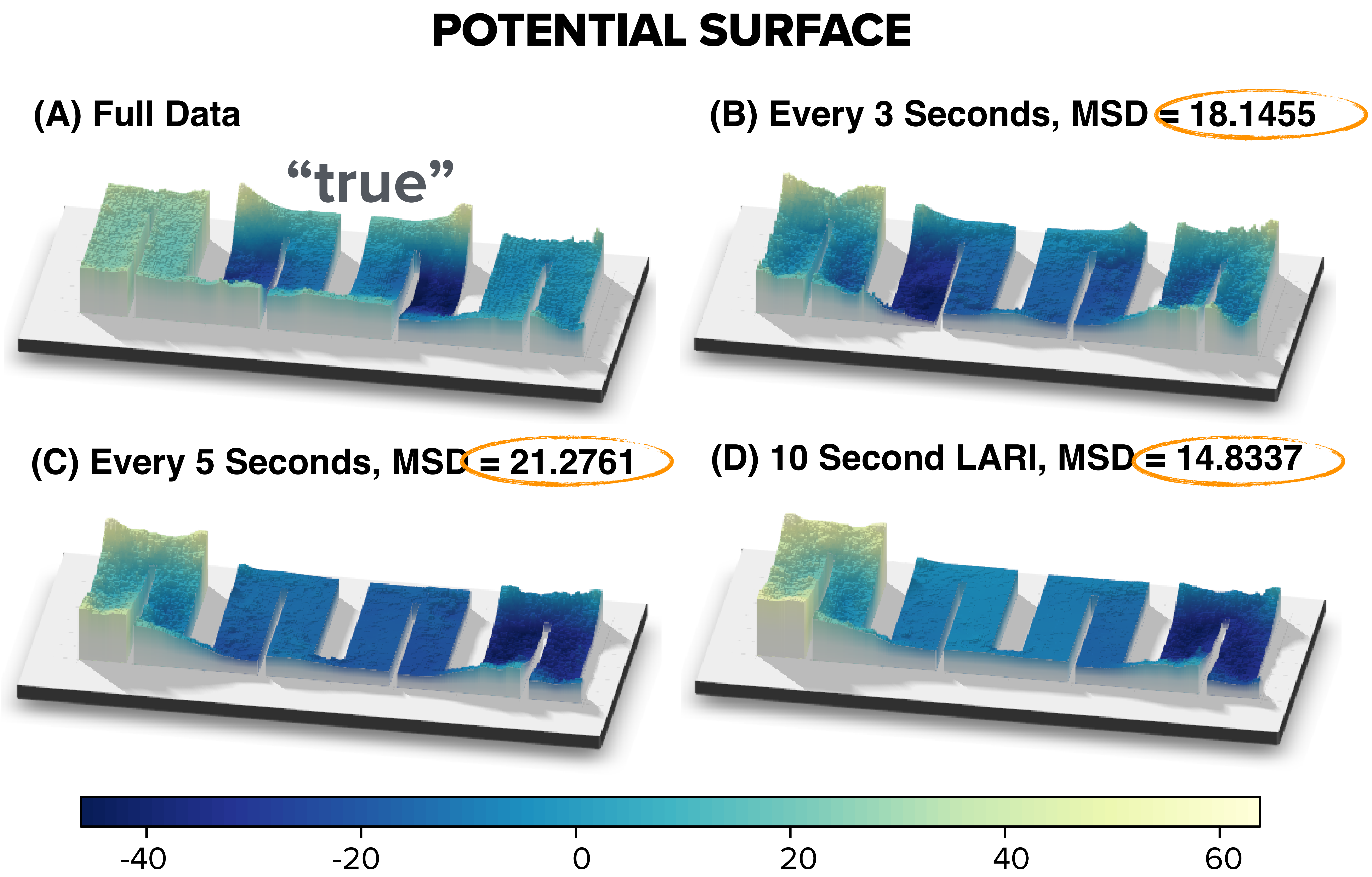


We **compared true and estimated motility and potential** surfaces using multiple metrics.

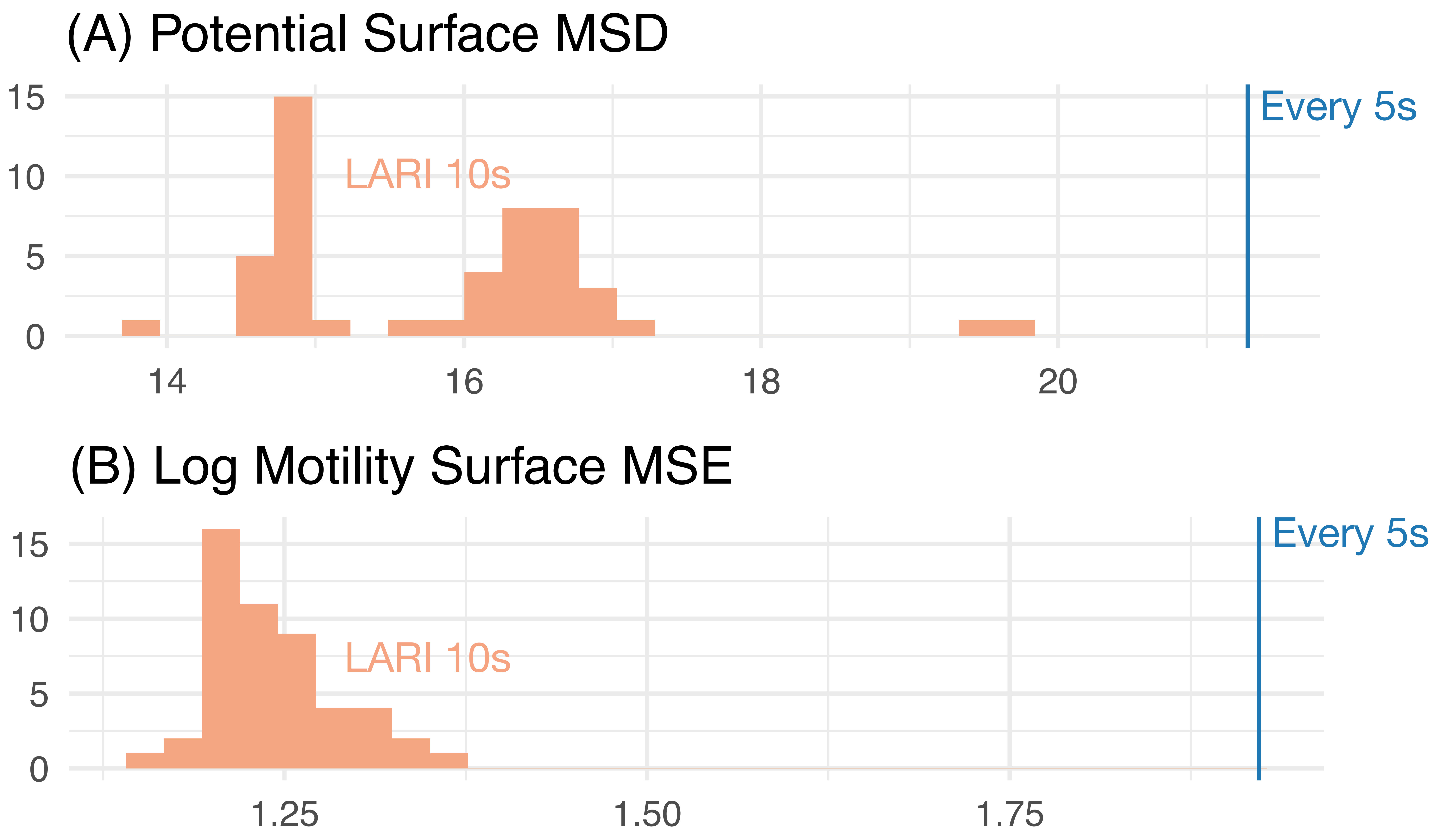
We **compare motility and potential** surfaces estimated with the 4 subsamples using multiple metrics.



We **compare motility and potential** surfaces estimated with the 4 subsamples using multiple metrics.



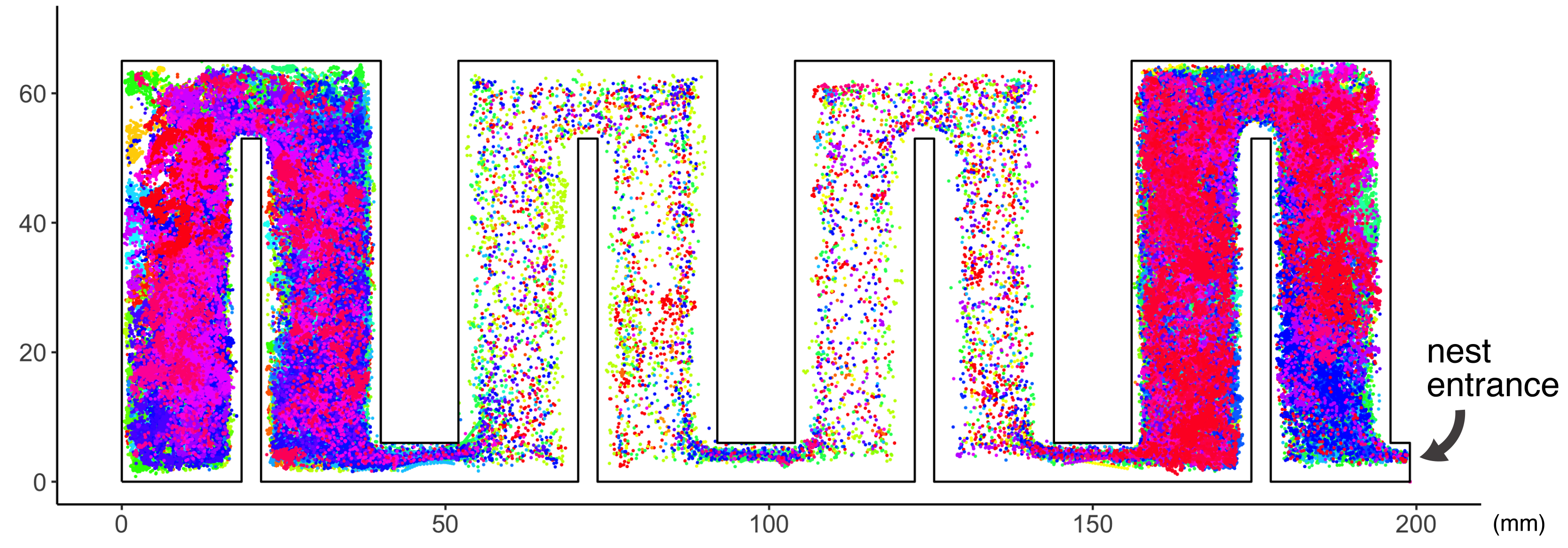
We fit 50 different LARI subsamples to understand random variation.



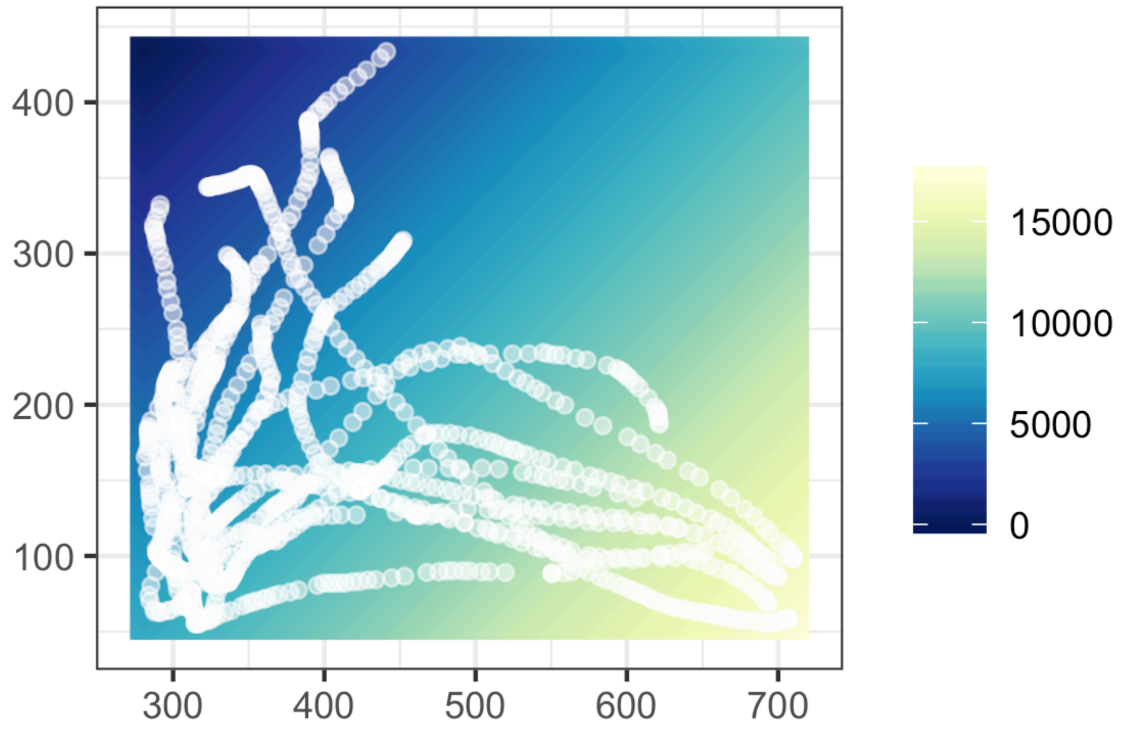
Result: **LARI sampling was better** than regular sampling overall for understanding movement behavior. A simulation study and additional data example support this conclusion. It may also be better for estimating missing data.

Conclusion: Regular sampling may not always be the best choice.

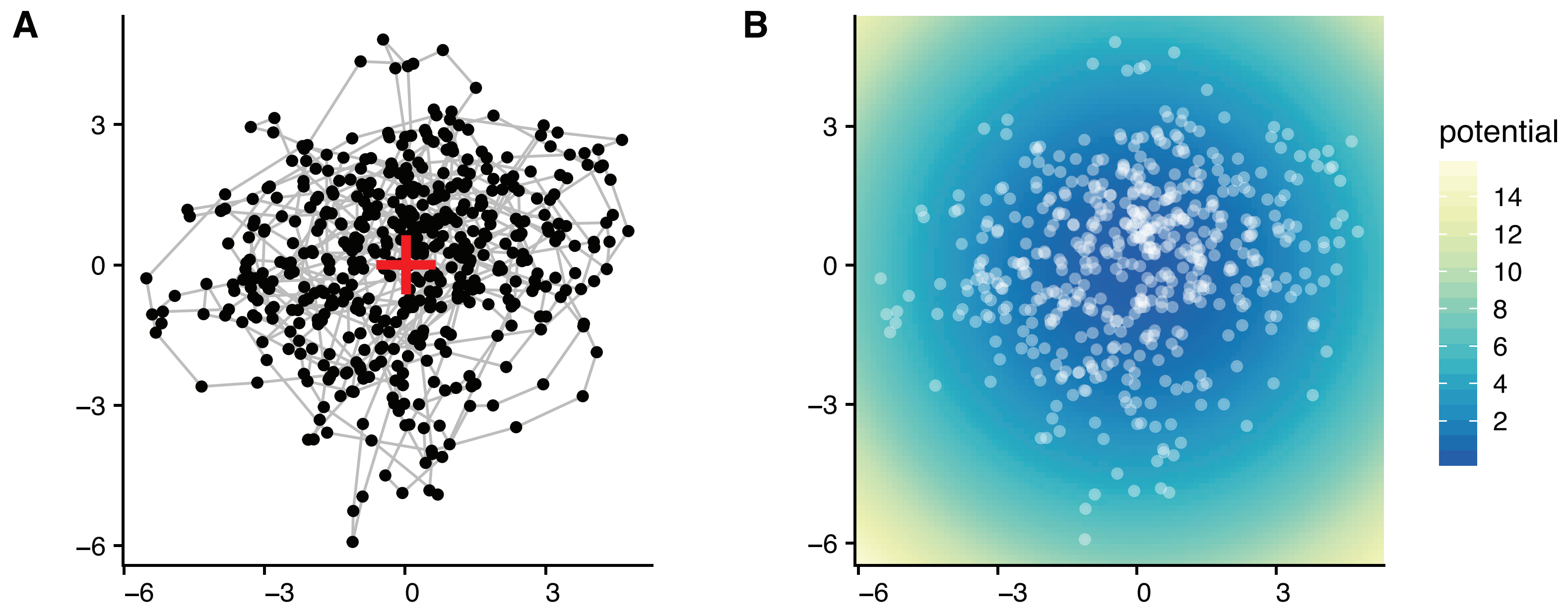
Ant data



Guppy data



Simulated data



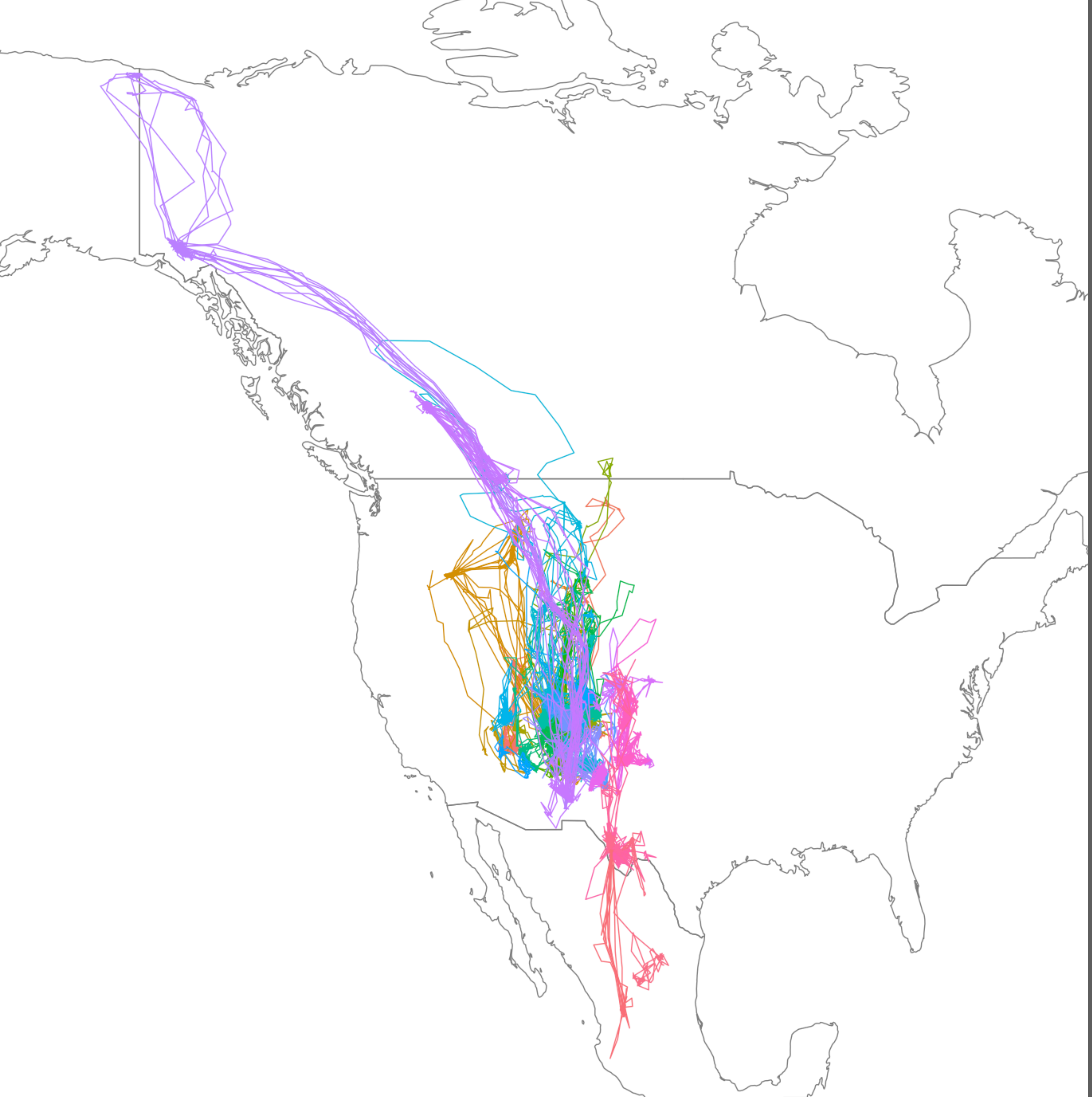
Eisenhauer, Elizabeth, and Ephraim **Hanks**. "A lattice and random intermediate point sampling design for animal movement." *Environmetrics* (2020): e2618.


Modeling Yearly Patterns in
Golden Eagle Movement

Golden eagles, like many species, display **partial migration**, meaning only some individuals in the population migrate.

Due to climate change, we can expect migration to change or become less common. Thus it is important that we can **identify** and **quantify** different movement strategies.

Current methods to classify movement strategies work best on the most stereotypical cases and are often in disagreement (Cagnacci et al., 2016).



Data collection funded by the USFWS

68 eagles with at least 1 year of data

Large **variability** in individual movement behaviors

Each color is one individual

Data cleaning:

1. Subset **1 observation per day**
2. **Linearly interpolate** missing intervals smaller than 30 days
3. Each **“path”** is one year for one individual and is analyzed separately

Big picture goals:

1. Relatively **simple** model (few parameters)
2. Capture the full range of movement behavior from **resident** to **migrant** to **dispersal**.
3. Use to **classify individuals** and better understand boundary individuals

We describe animal movement using a **stochastic differential equation (SDE)** model with a constant motility surface.

Data: $\mathbf{r}_t,\ t = 1, 2, \ldots, T$ for each eagle

SDE model framework:

$$d\mathbf{r}_t = \mathbf{v}_t\ dt$$
$$d\mathbf{v}_t = \beta\ (\boldsymbol{\mu}(\mathbf{r}_t, t) - \mathbf{v}_t)\ dt + \sigma\ d\mathbf{W}_t$$

Utilizing a potential function, define:

$$\boldsymbol{\mu}(\mathbf{r}_t, t) = -\nabla p(\mathbf{r}_t, t) \qquad \text{(mean drift)}$$

(James Russell et al., 2018)

We again use **Euler-Maruyama** approximations.
We assume **regular time steps**.

Resulting model equation:

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

where

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

a weighted sum of distances to m fixed attractors with x-coordinates $a_{x1}, a_{x2}, \ldots, a_{xm}$ and y-coordinates $a_{y1}, a_{y2}, \ldots, a_{ym}$

We want the coefficient of attraction k_{it} for attractor i to **change over time.**

Methods:

1. **Varying coefficient model** allows k_{it} to change smoothly over time.
2. **Latent-state model** allows k_{it} to switch between discrete values over time.
   - Note: The term latent-state model is used over HMM because of the feedback inherent in the model, i.e., $\mathbf{r}_t$ depends on the previous 2 time points as well as the state at time t

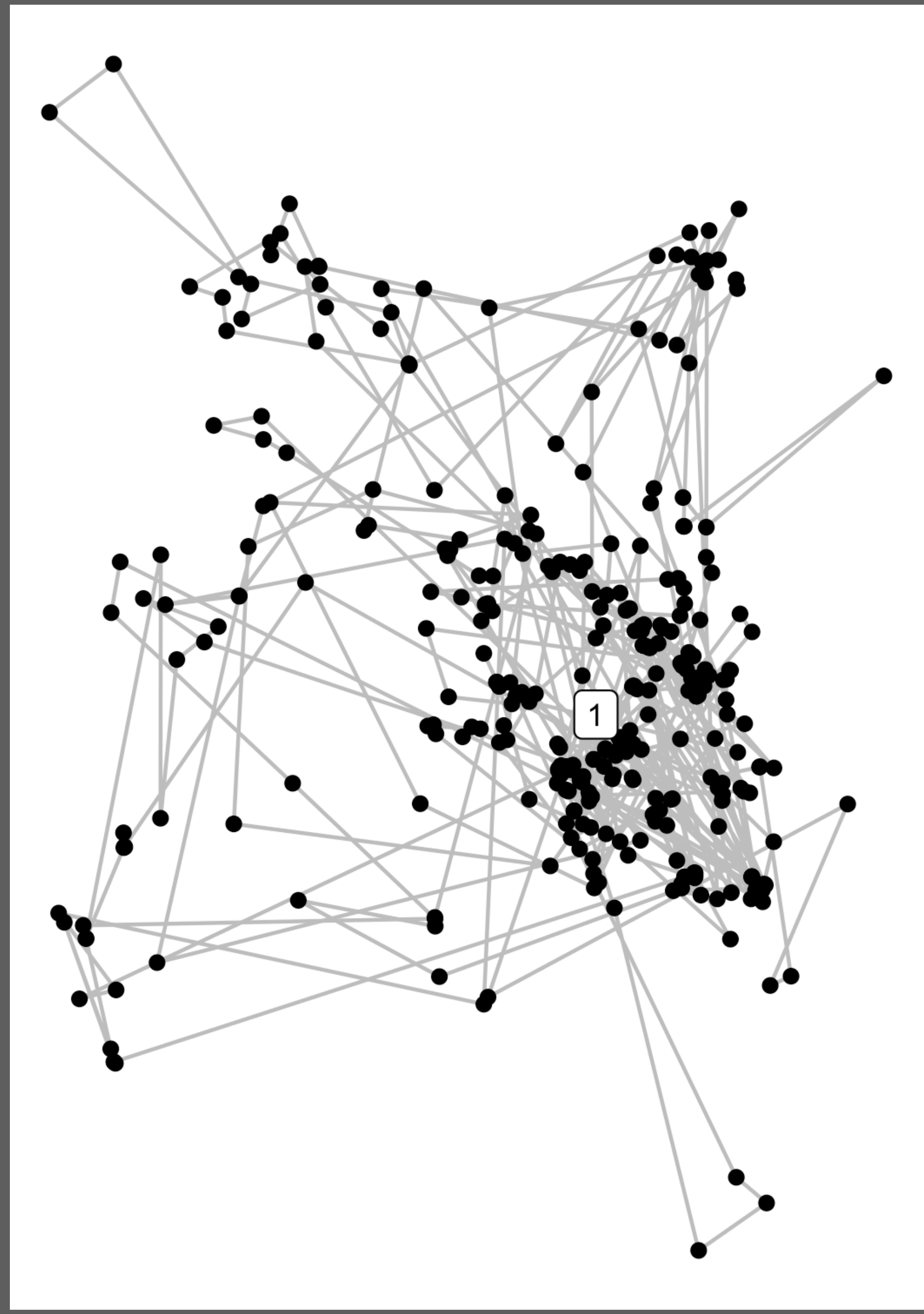
Latent-state model:

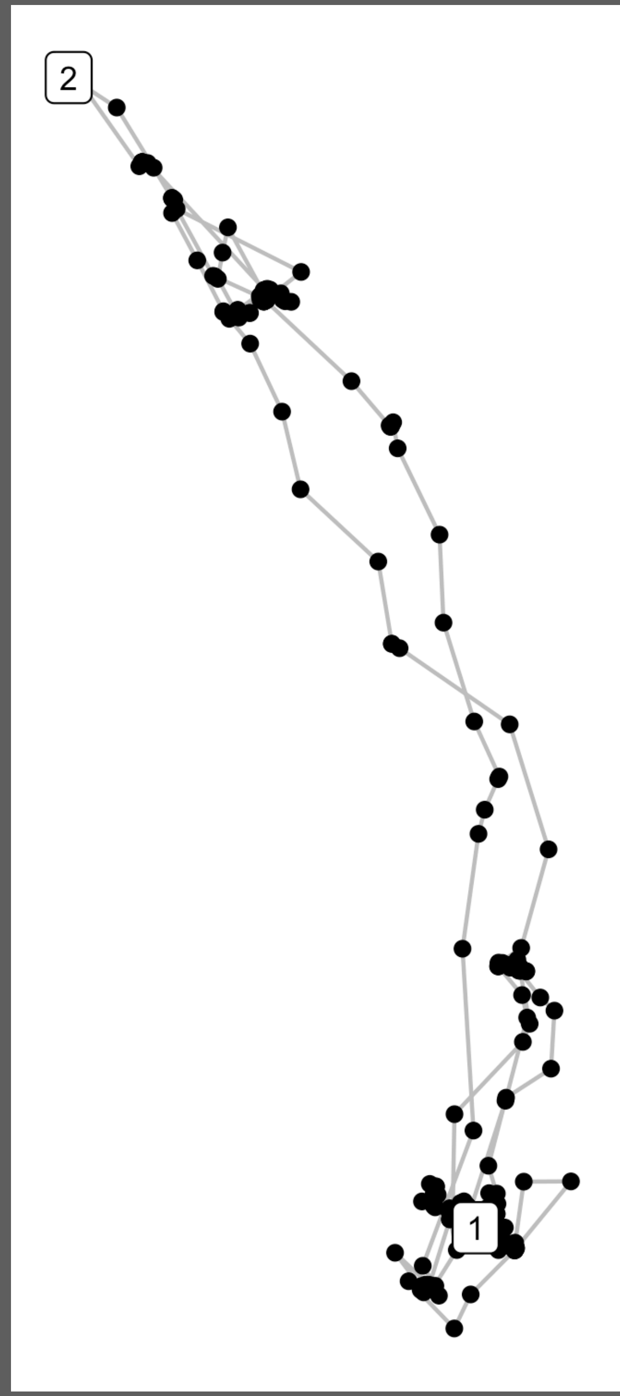
1. Each state has different k_{it}
2. We fit each model in a **Bayesian framework**
   - We sample using the No U-Turn Sampler implemented in Stan

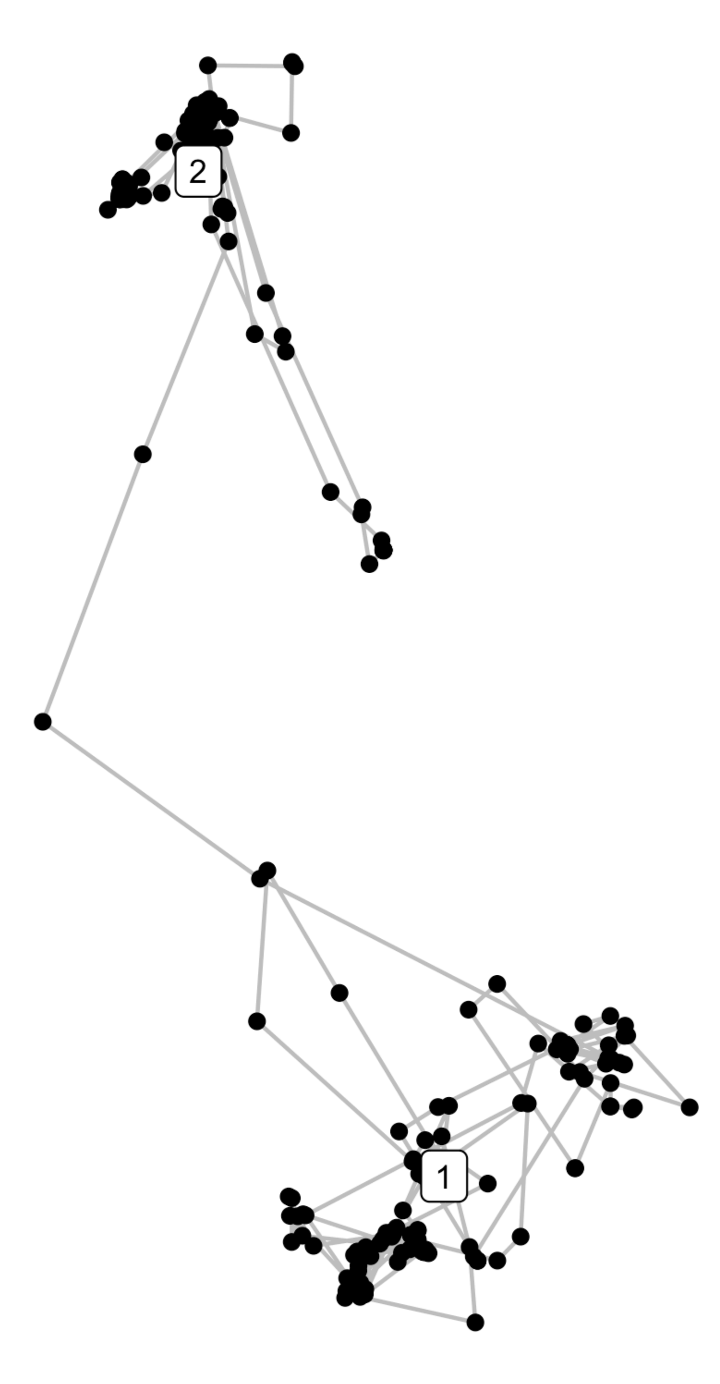
Varying coefficient model:

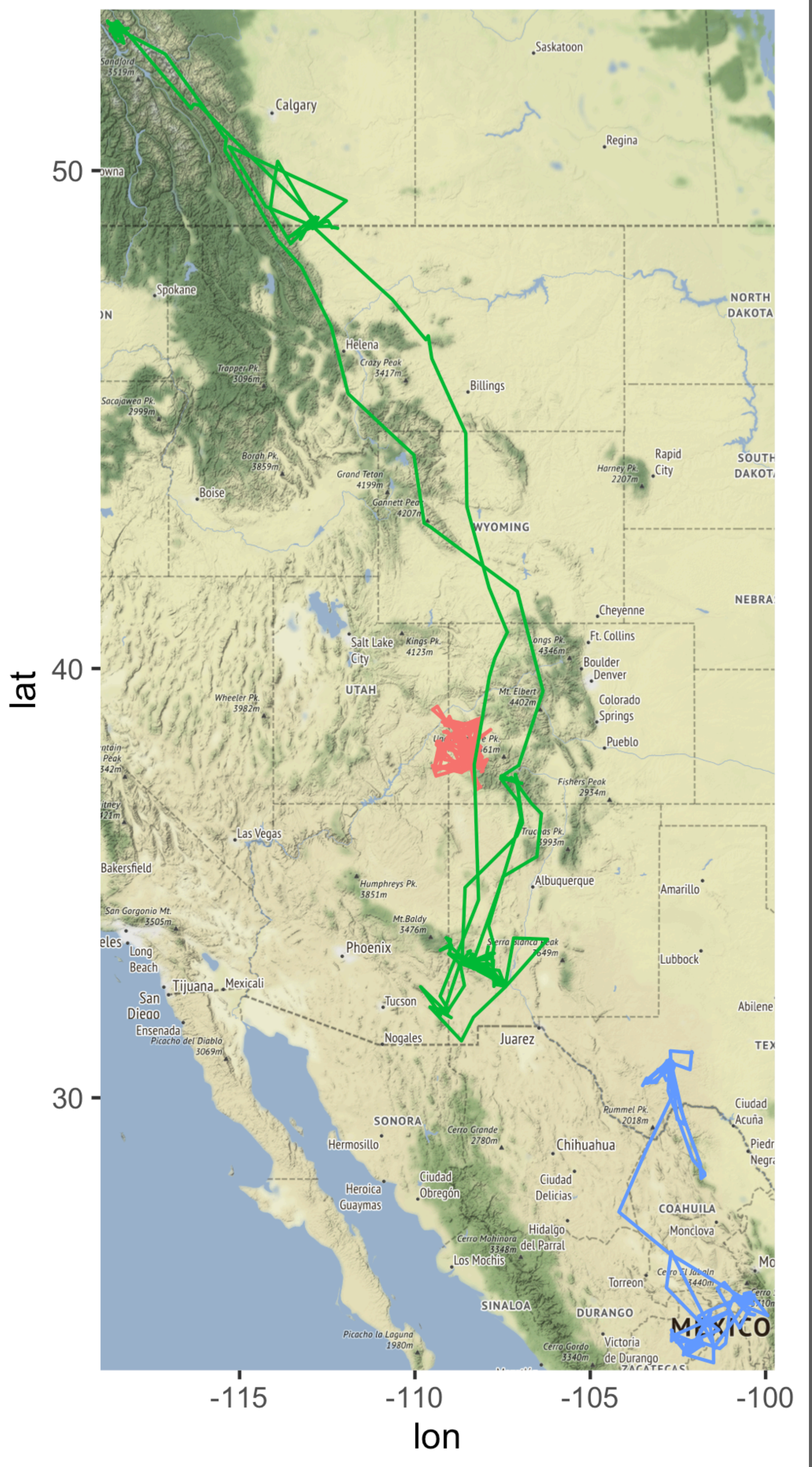
1. k_{it} changes smoothly over time
2. Fit using **GAM** function in R
3. The **same model** for resident, migrant, and disperser

We will fit these models for **3 representative individuals**

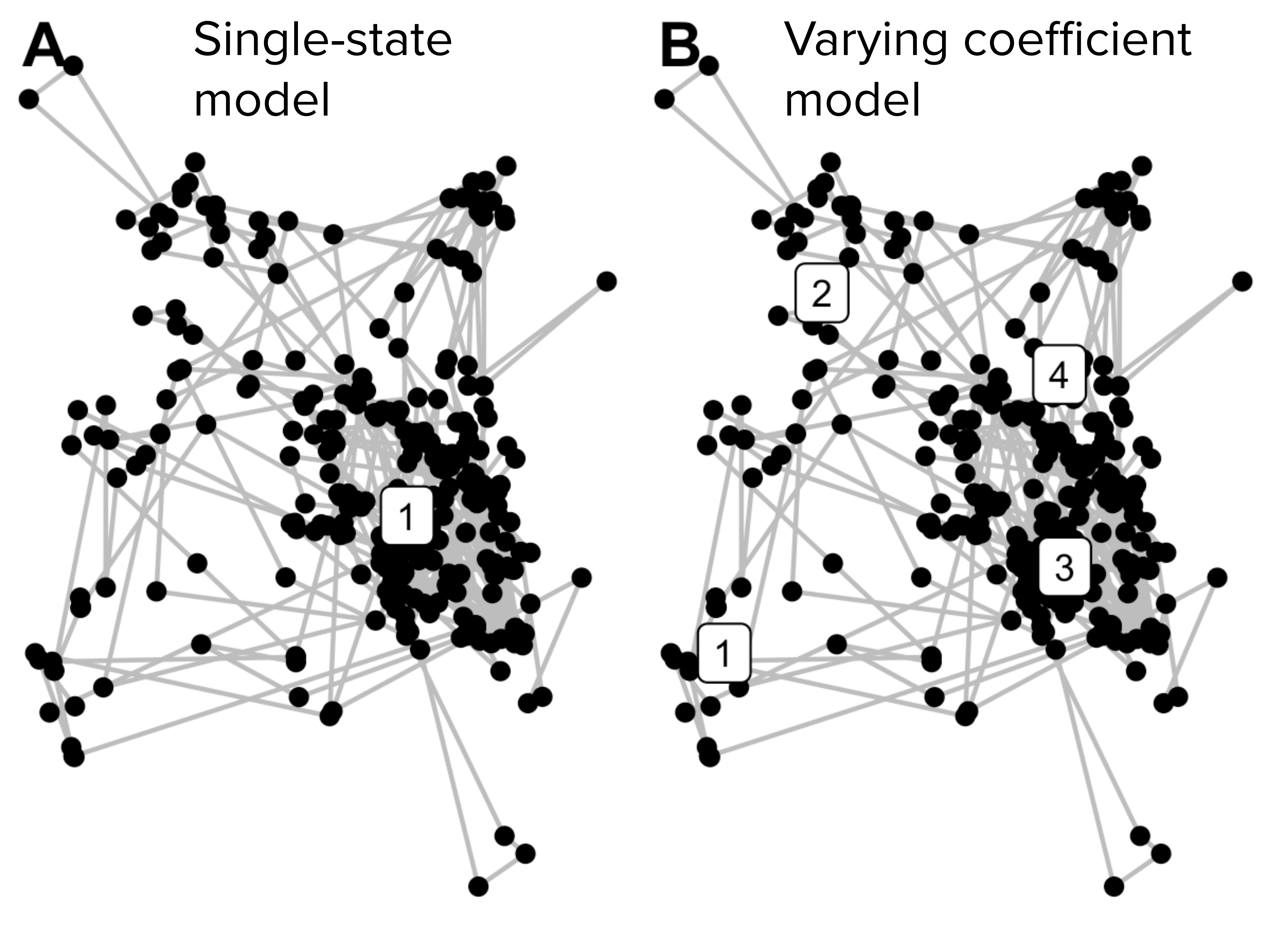








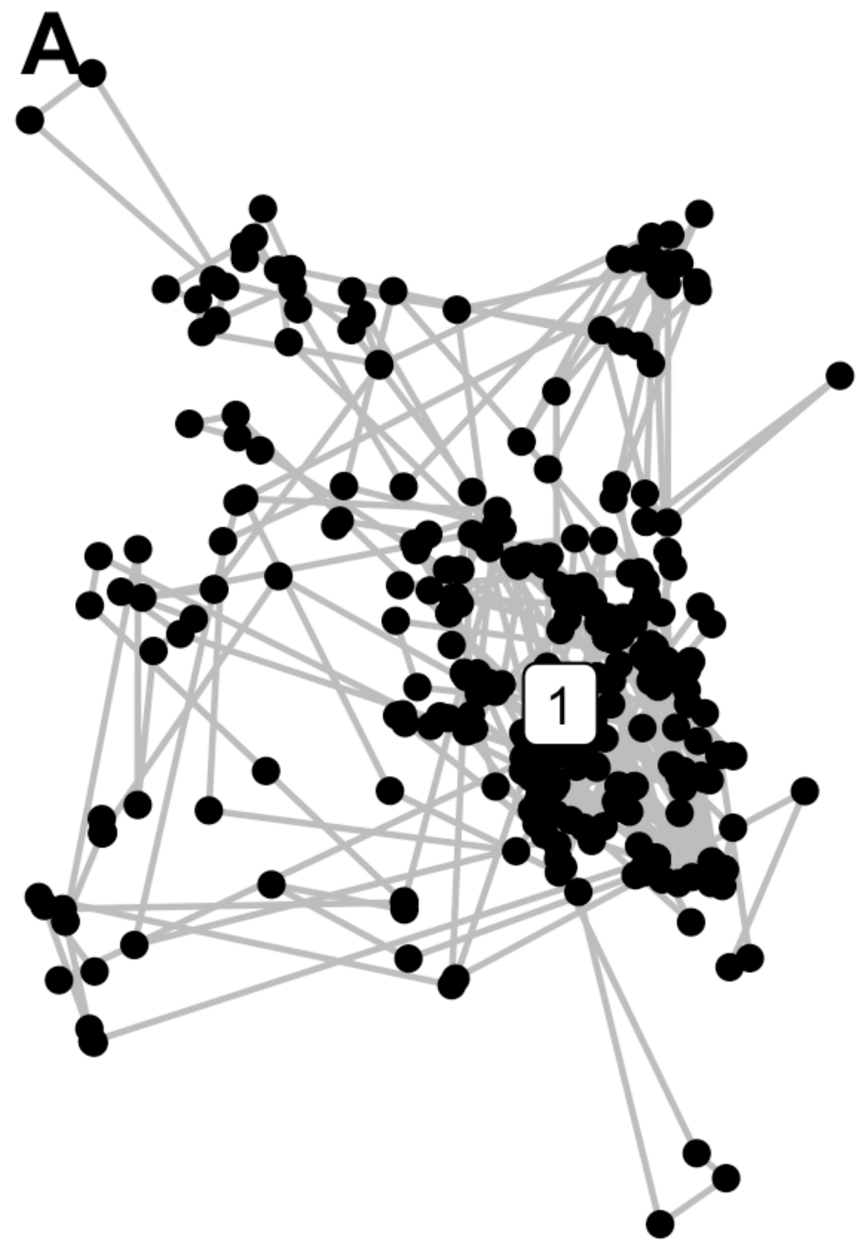
Resident: 4C.Cahone in 2015

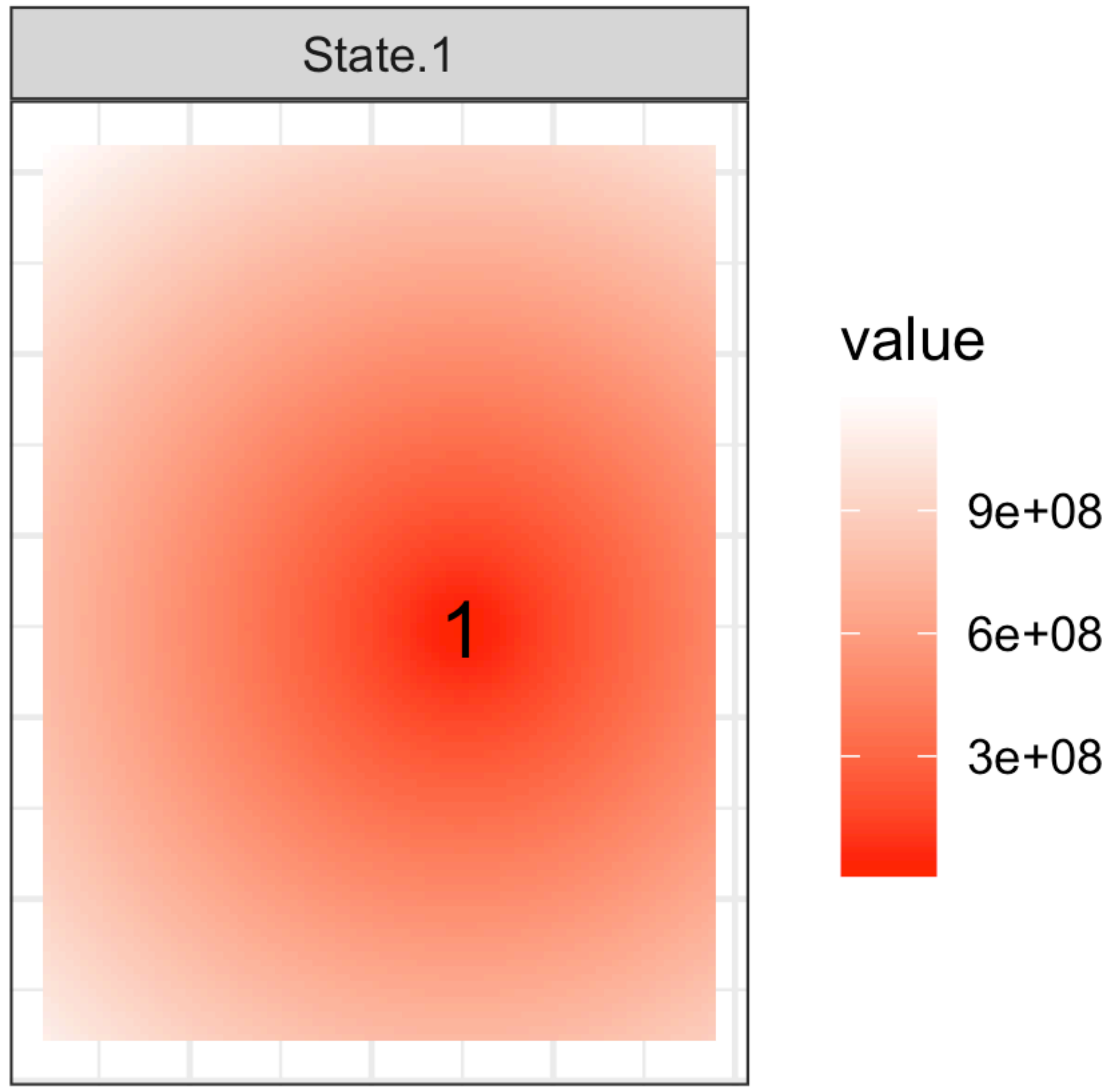


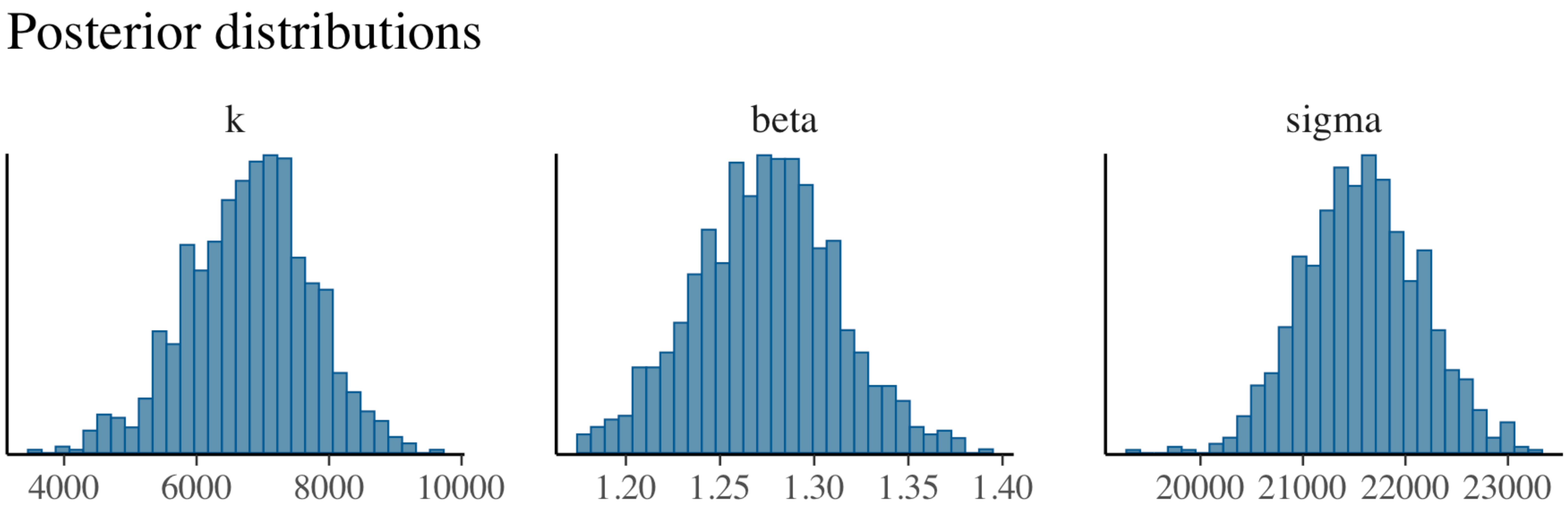
Single-State Model for Resident

$m = 1$ attractor with coefficient of attraction $k_{1t} = k$

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





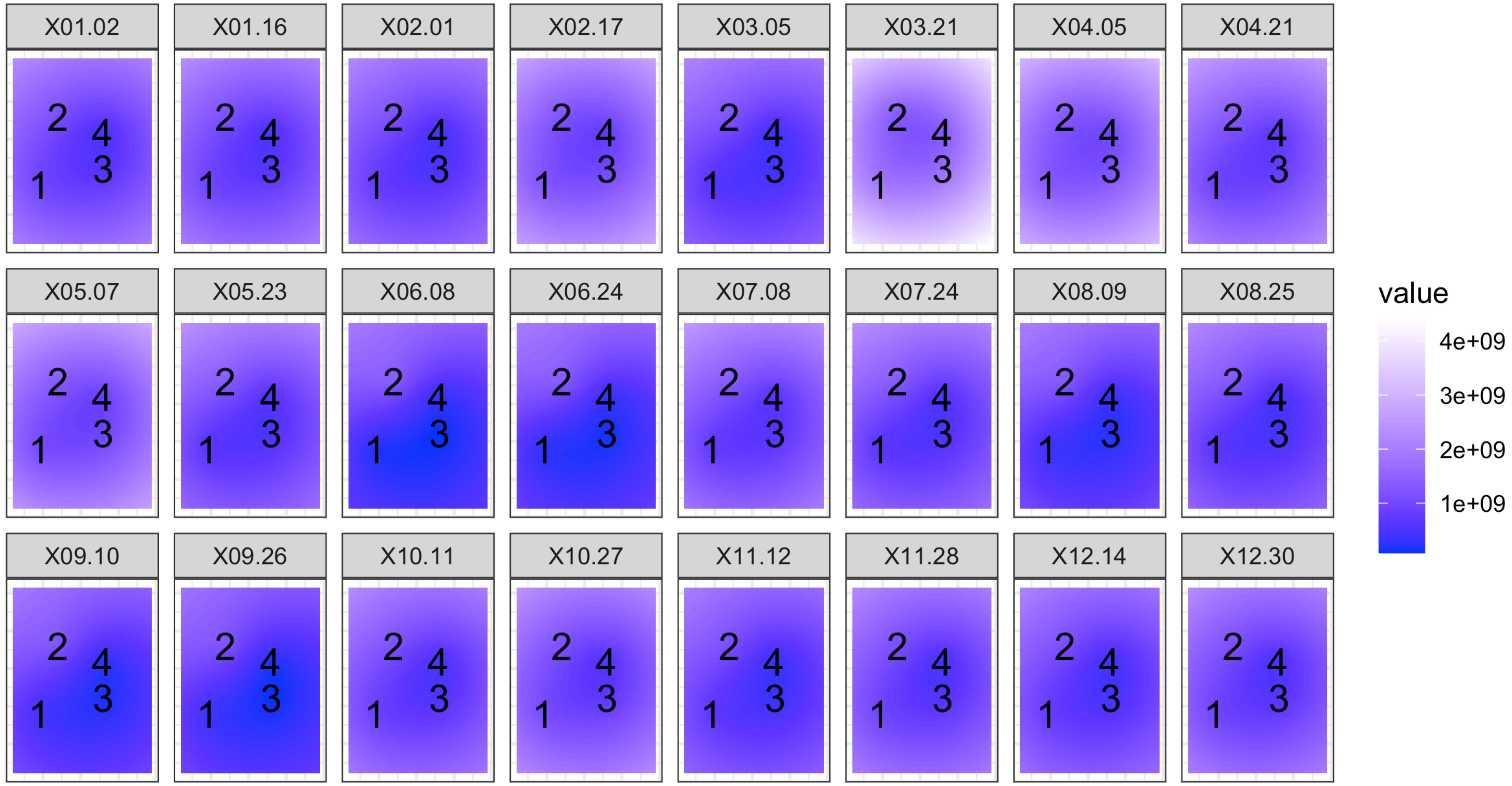


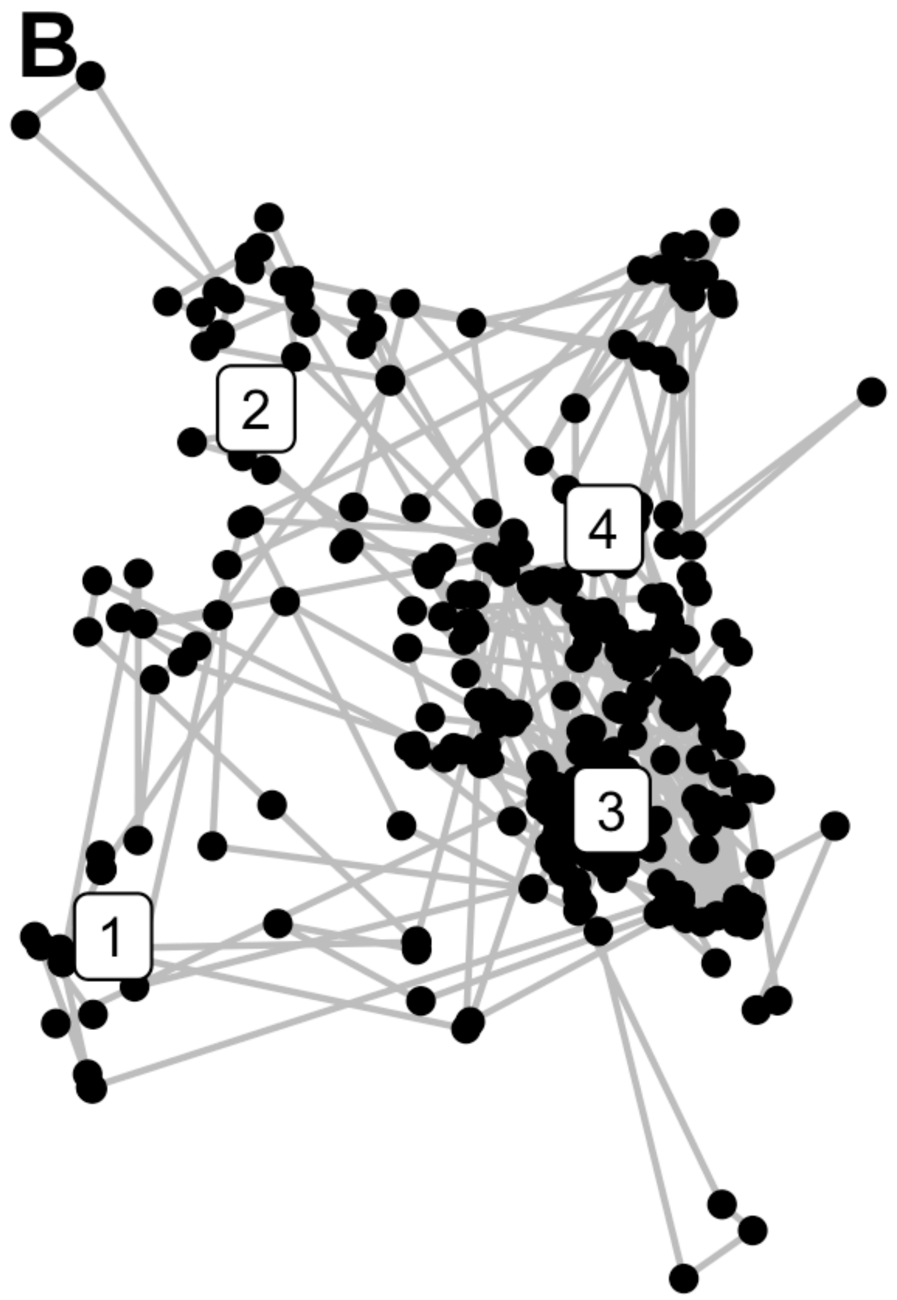
Varying Coefficient Model for Resident

Fixed $m = 4$ attractors and k_{it} for $i = 1,2,3,4$ changes smoothly over time.

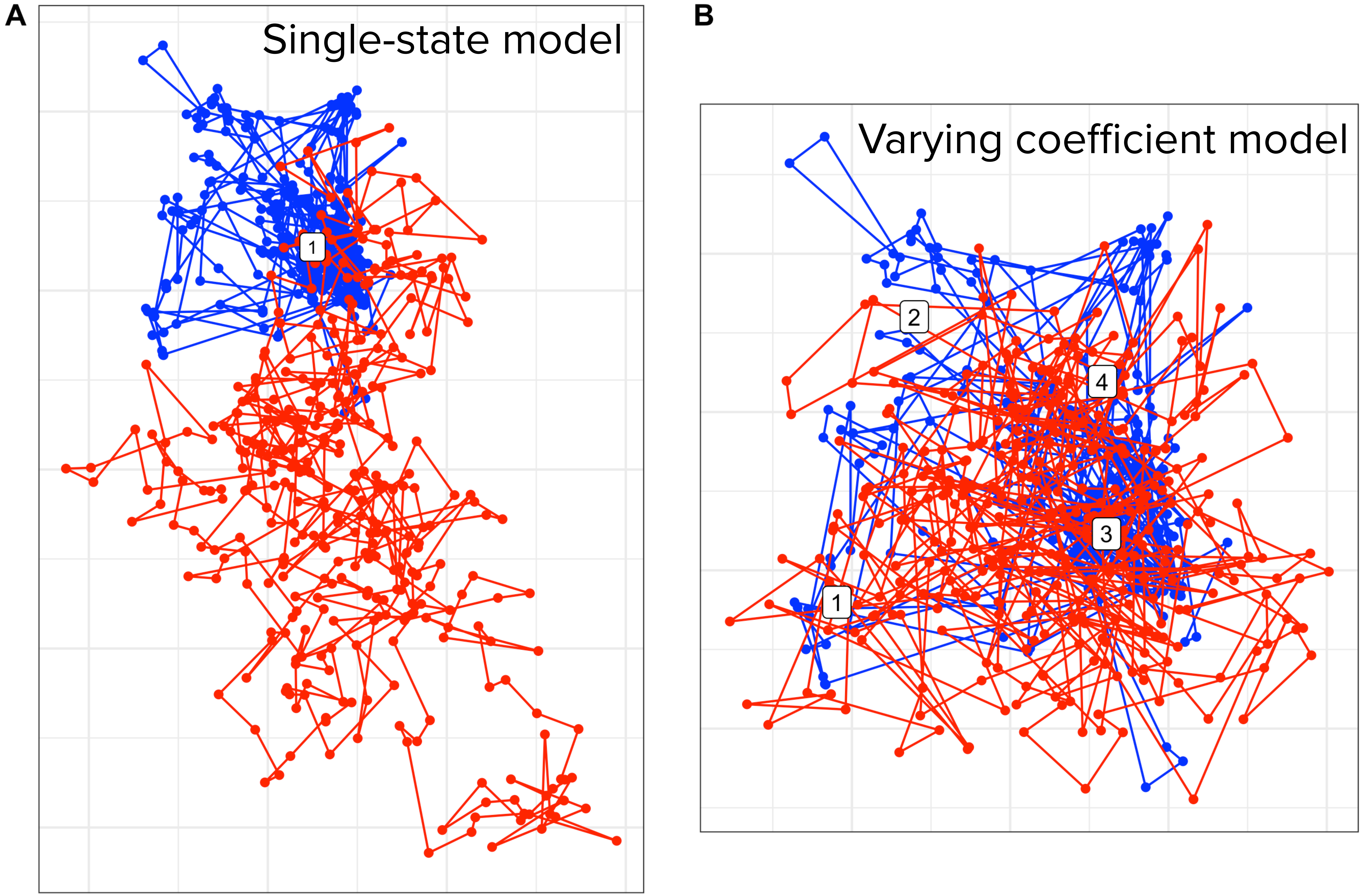
$$p(\mathbf{r}_t) = \sum_{i=1}^{m} \sum_{j=1}^{J} \alpha_{ij} B_j(t) \sqrt{(x_t - a_{xi})^2 + (y_t - a_{yi})^2}$$

where α_{ij} is the coefficient of the j^{th} cyclic cubic basis function $B_j(t)$ for attractor i

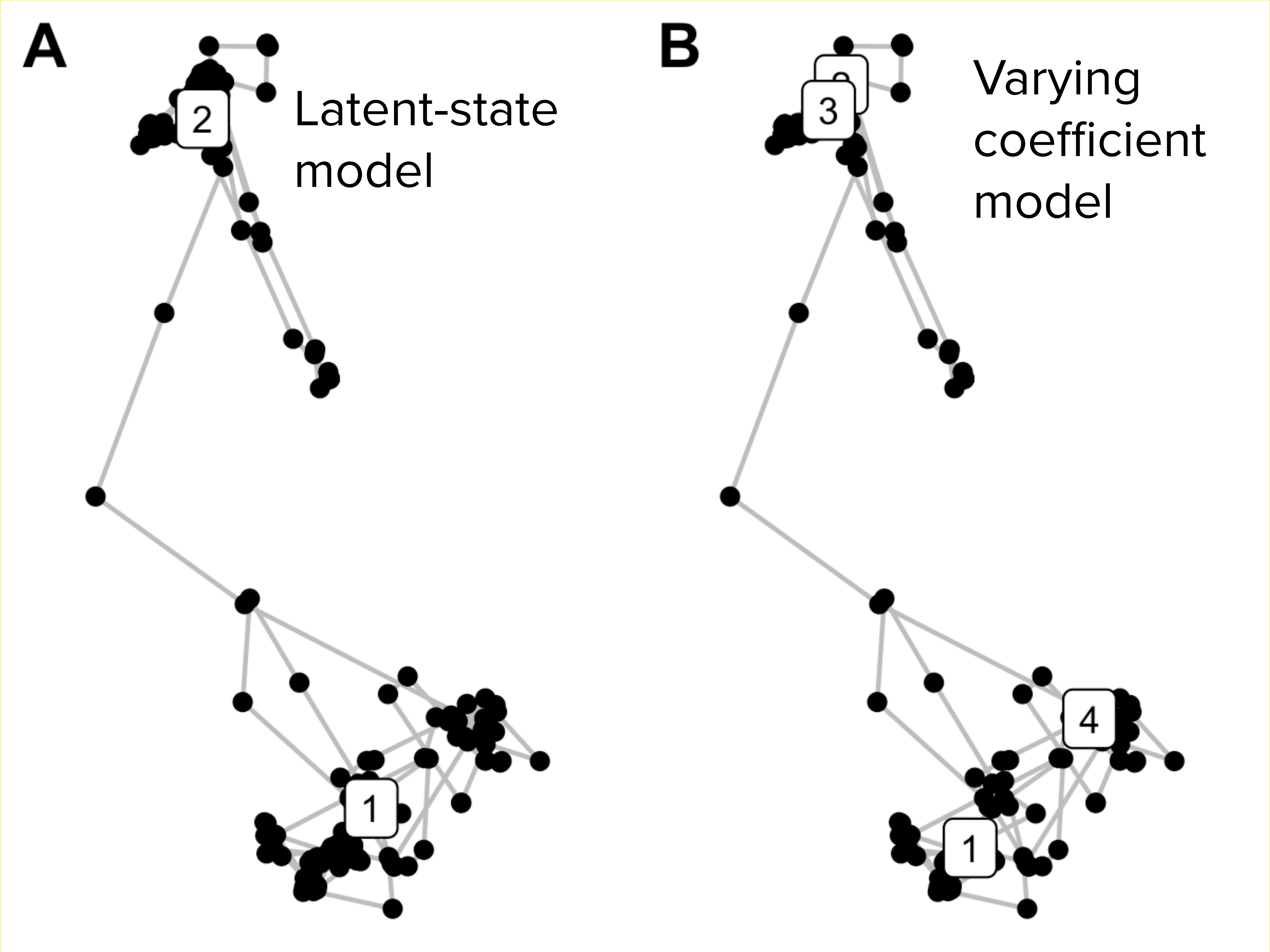




Simulations



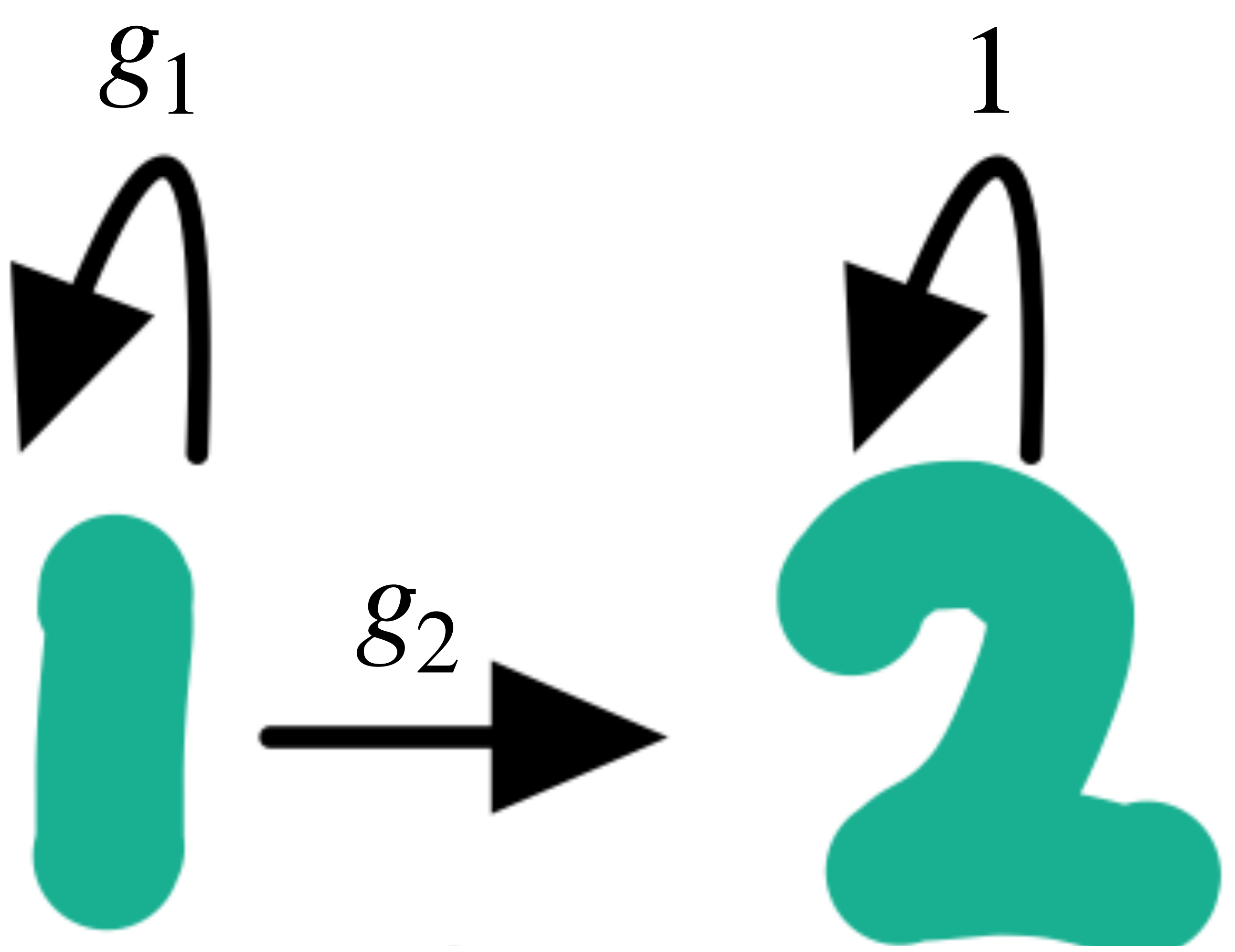
Dispersal: TP.N Zacatecas in 2018



Latent-State Model for Dispersal

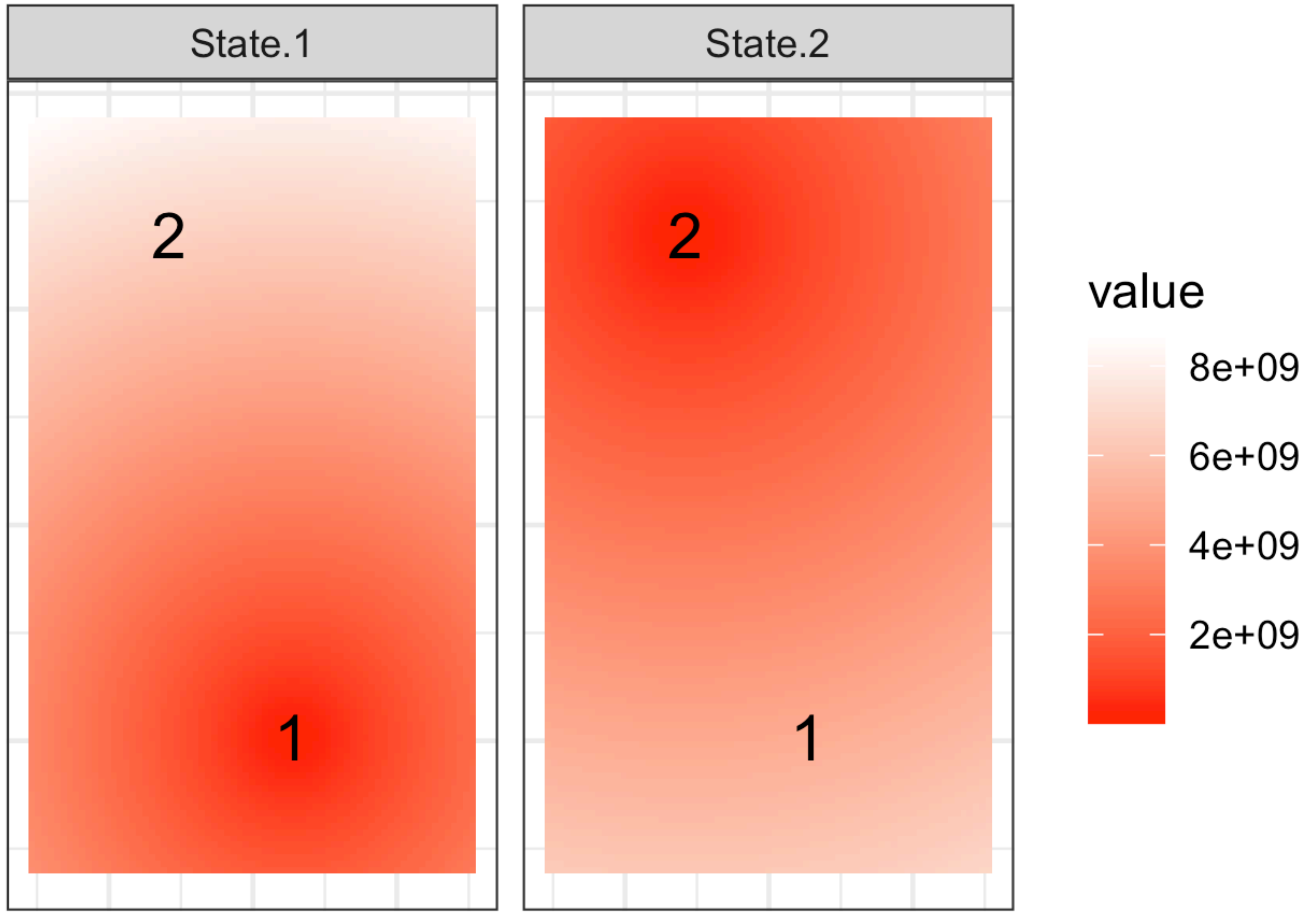
$$p(\mathbf{r}_t, s_t) = k_{s_t}\sqrt{(x_t - a_{xs_t})^2 + (y_t - a_{ys_t})^2}$$

Latent states:

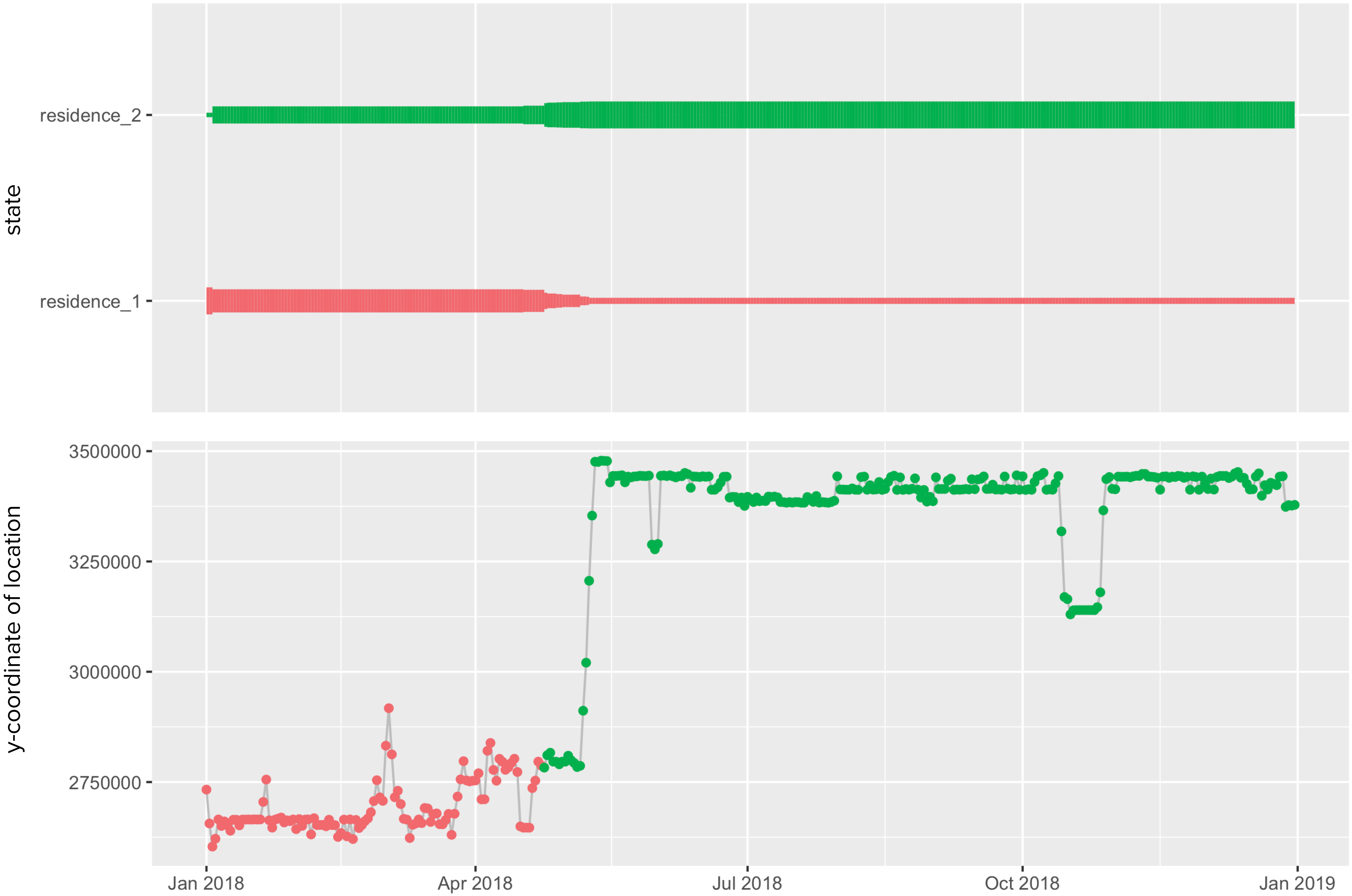


Attractor 1 and coefficient k_1

Attractor 2 and coefficient k_2



Posterior probabilities for each state

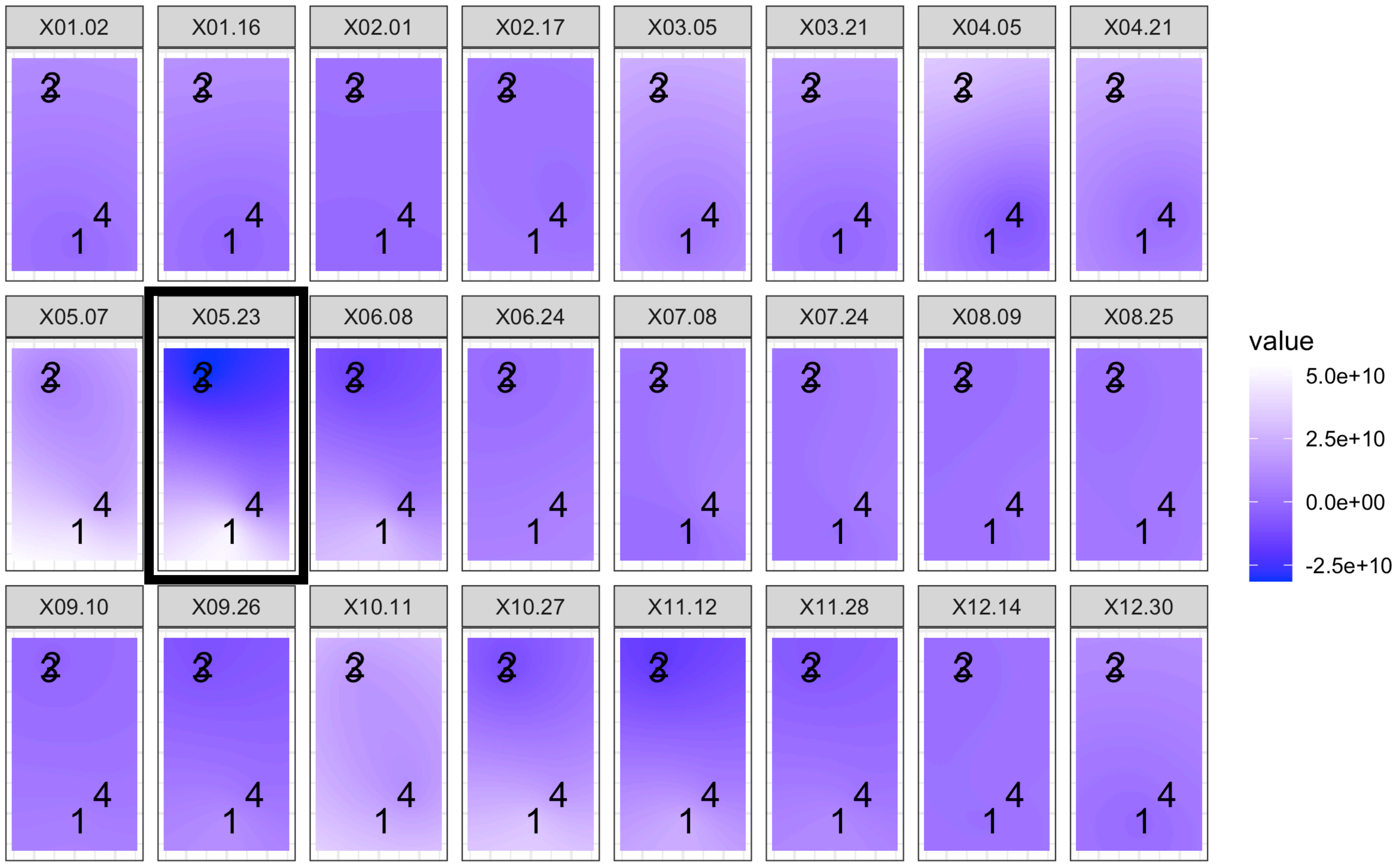


Varying Coefficient Model for Dispersal
(identical to varying coefficient model for resident)

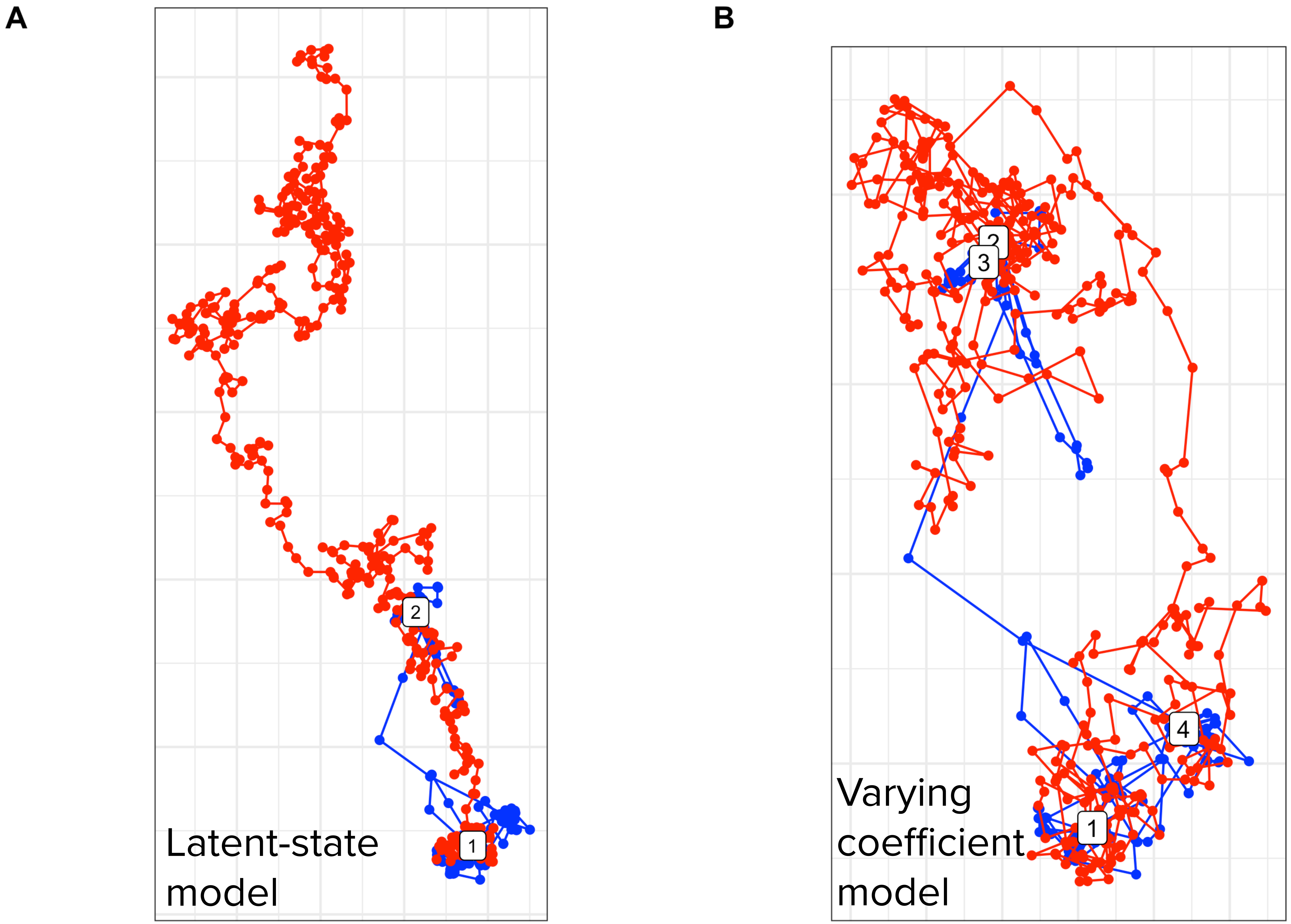
Fixed $m = 4$ attractors and k_{it} for $i = 1,2,3,4$ changes smoothly over time.

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

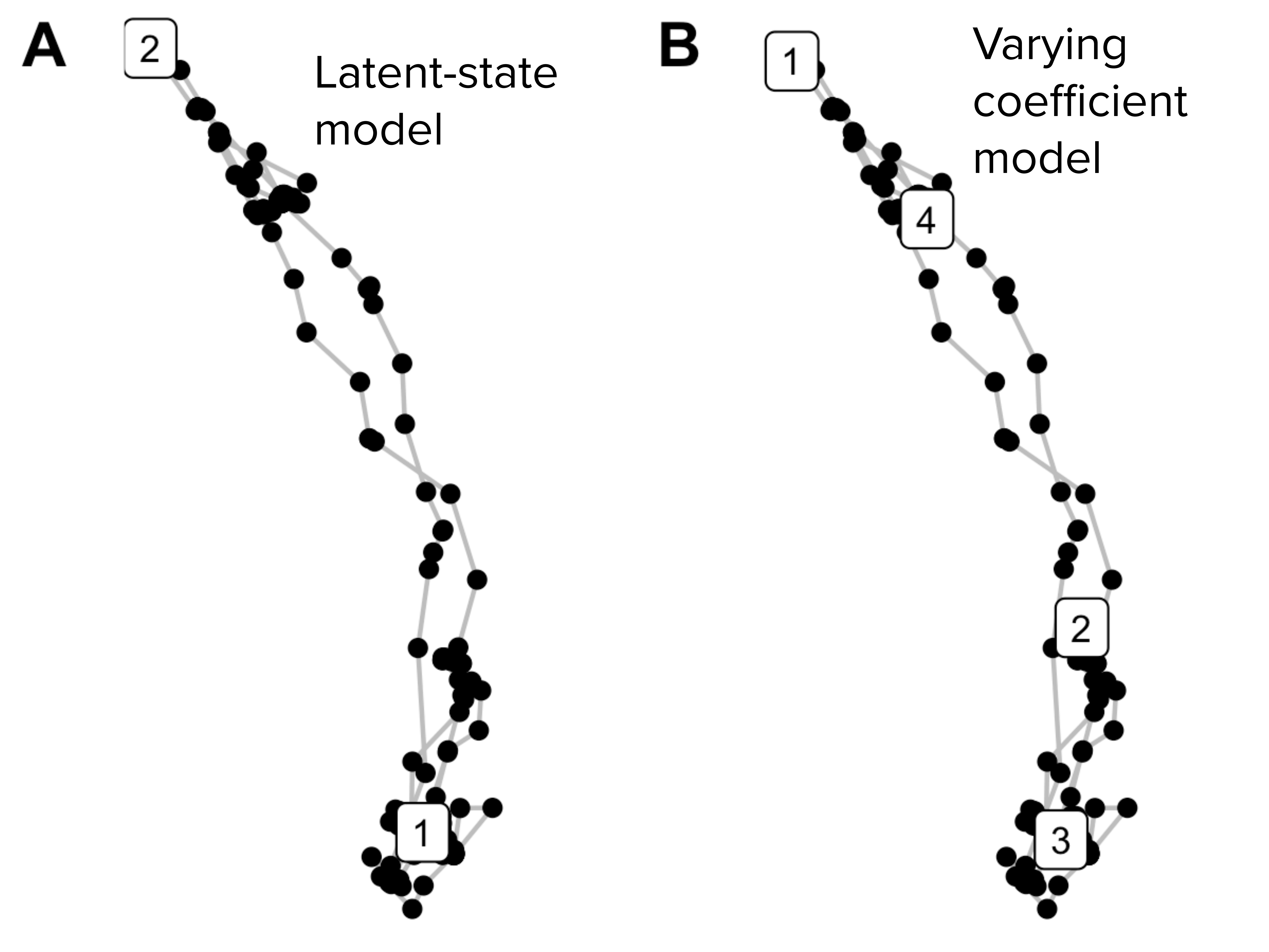
where α_{ij} is the coefficient of the j^{th} cyclic cubic basis function $B_j(t)$ for attractor i



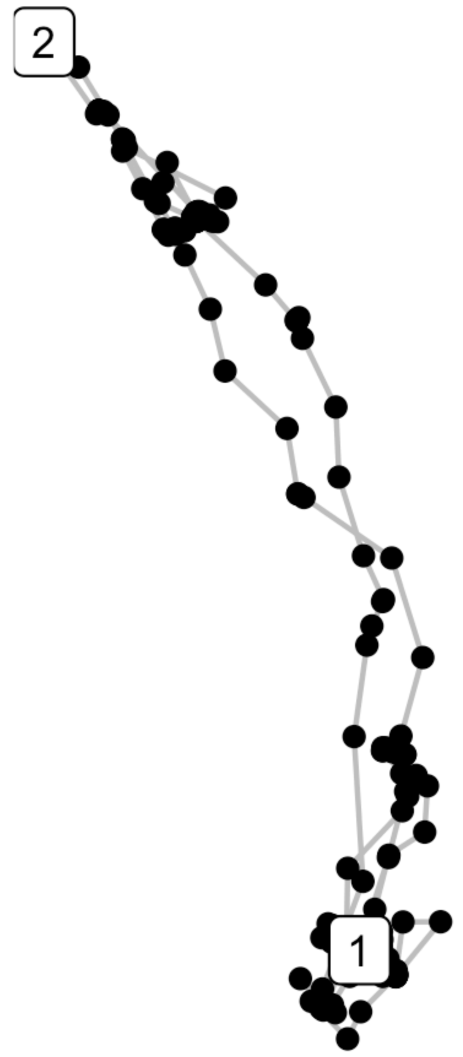
Simulations



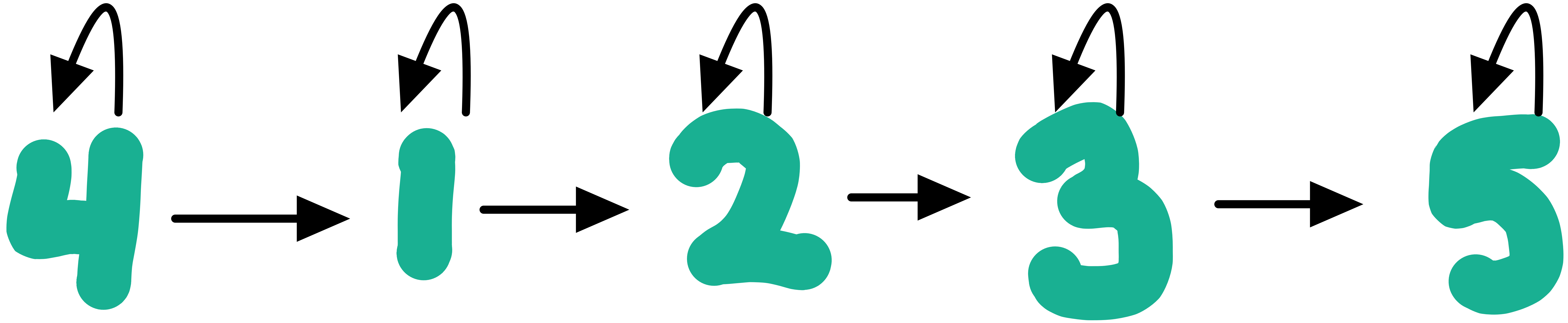
Migrant: NM.Tredwell in 2012



Latent-State Model for Migration



Latent states:



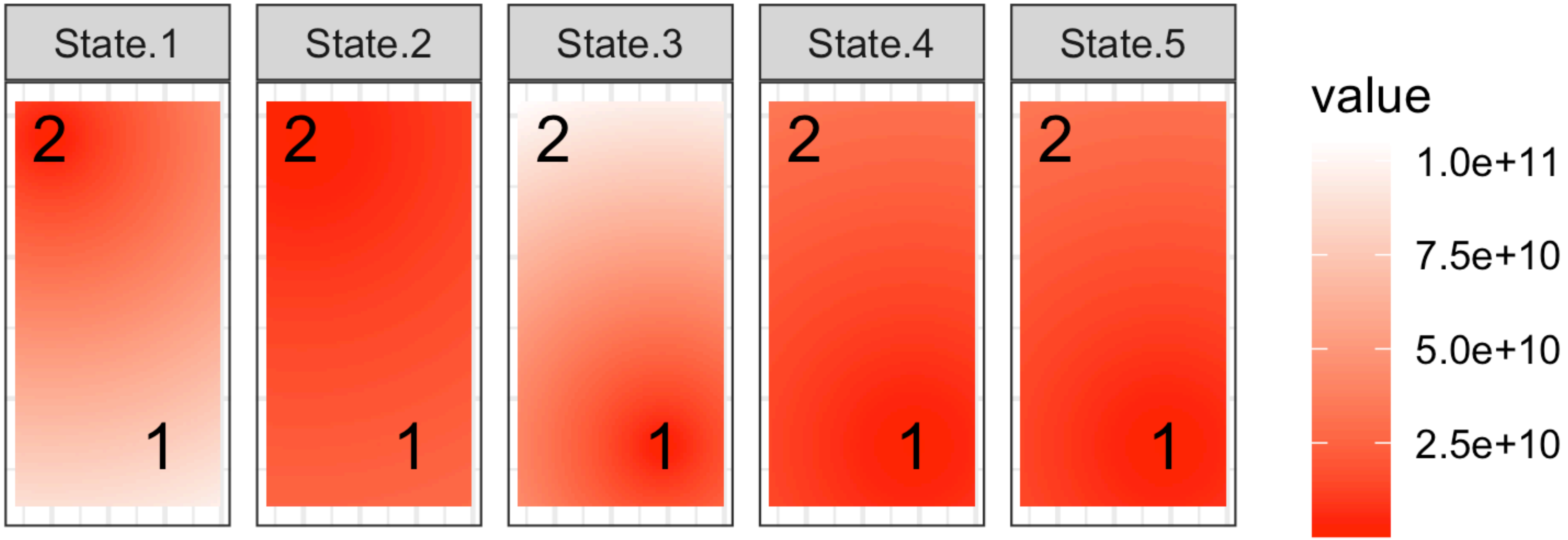
Southern Residence
Attractor 1 and coefficient k_4

Northern Migration
Attractor 2 and coefficient k_1 $(k_1 > k_2)$

Northern Residence
Attractor 2 and coefficient k_2

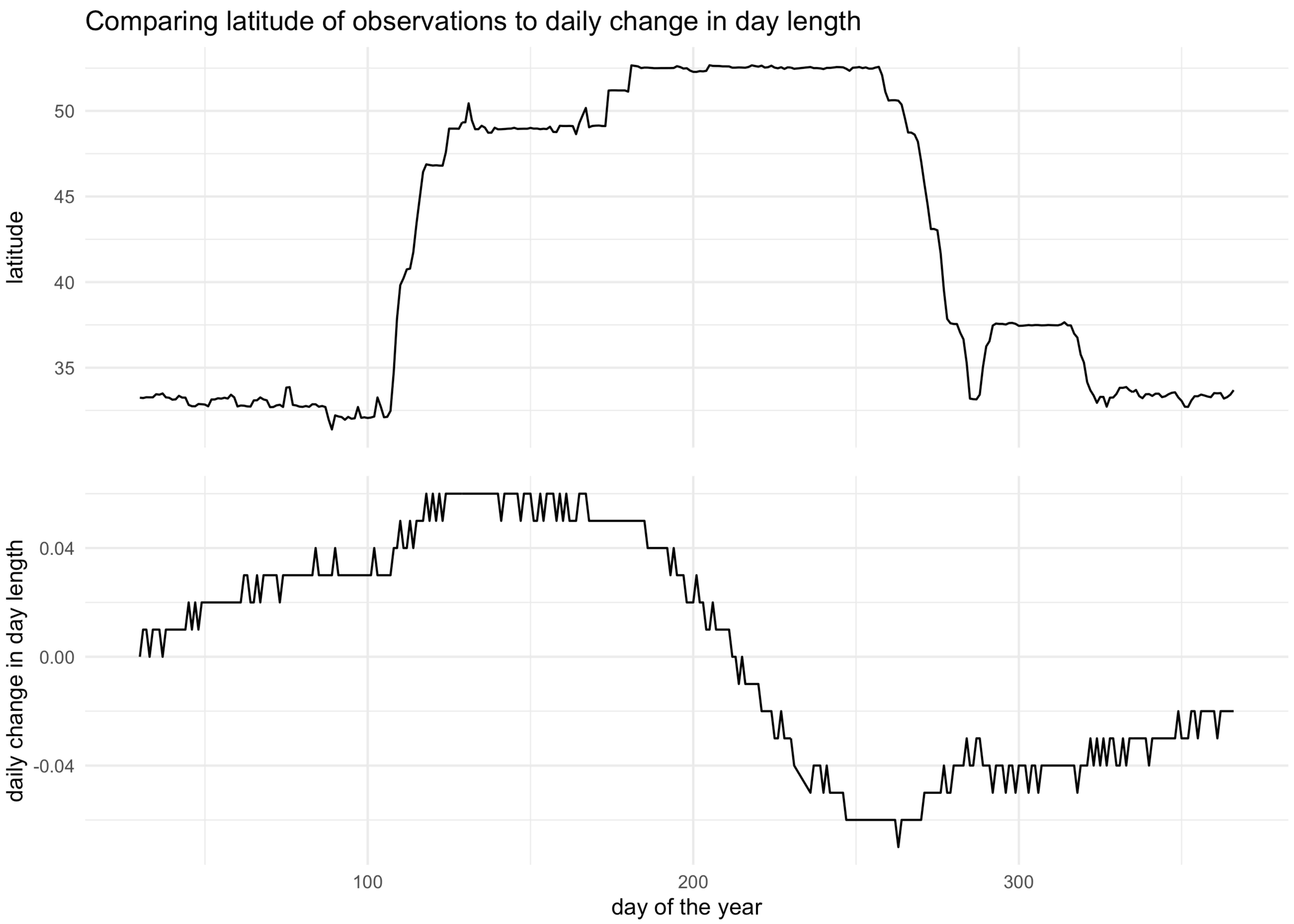
Southern Migration
Attractor 1 and coefficient k_3 $(k_3 > k_4)$

Southern Residence
Attractor 1 and coefficient $k_5 = k_4$

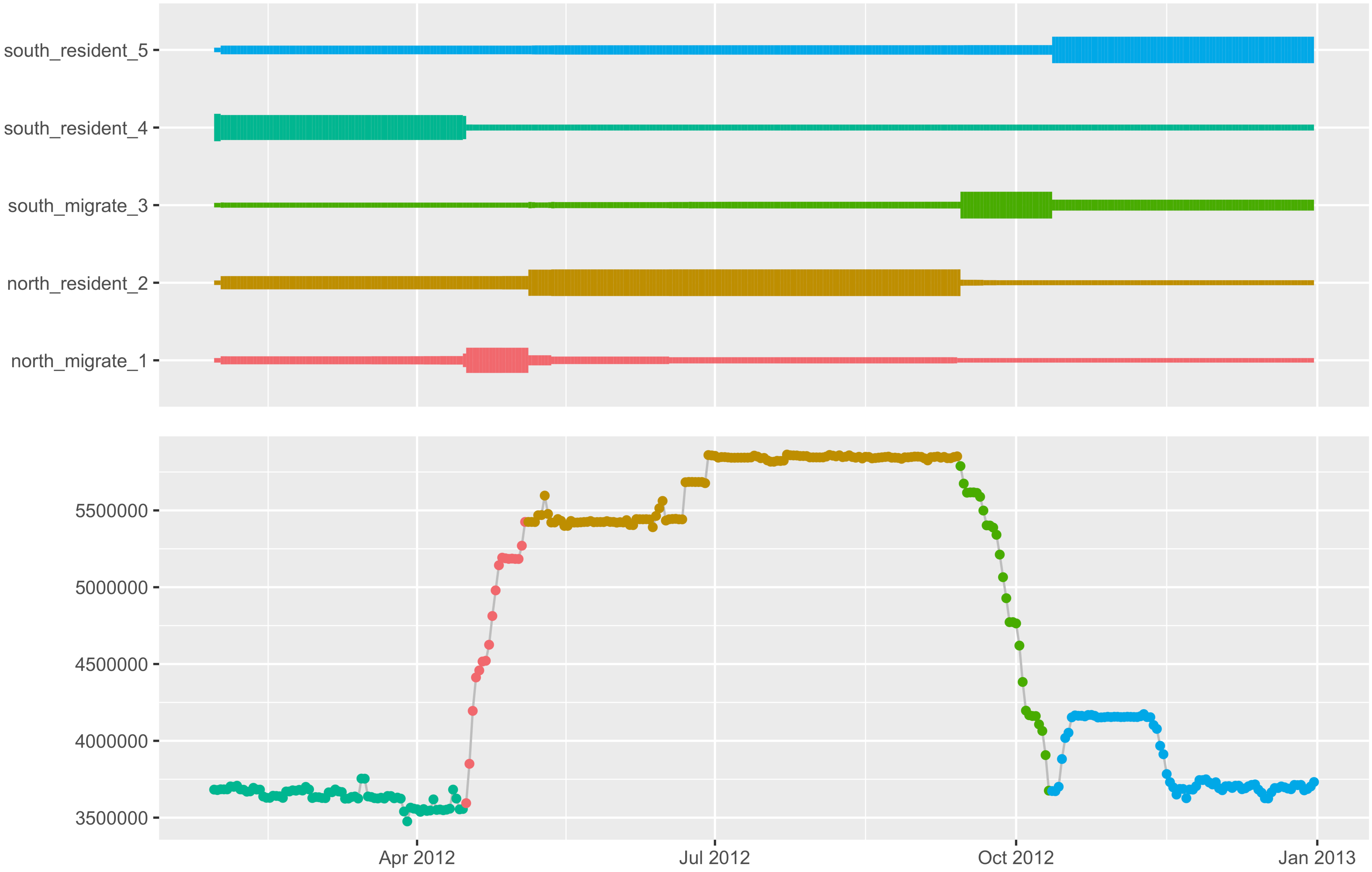


$$p(\mathbf{r}_t, s_t) = \begin{cases} k_{s_t}\sqrt{(x_t - a_{x1})^2 + (y_t - a_{y1})^2}, & s_t \in \{3,4,5\} \\ k_{s_t}\sqrt{(x_t - a_{x2})^2 + (y_t - a_{y2})^2}, & s_t \in \{1,2\} \end{cases}$$

Transition probabilities are functions of a covariate.



Posterior probabilities for each state

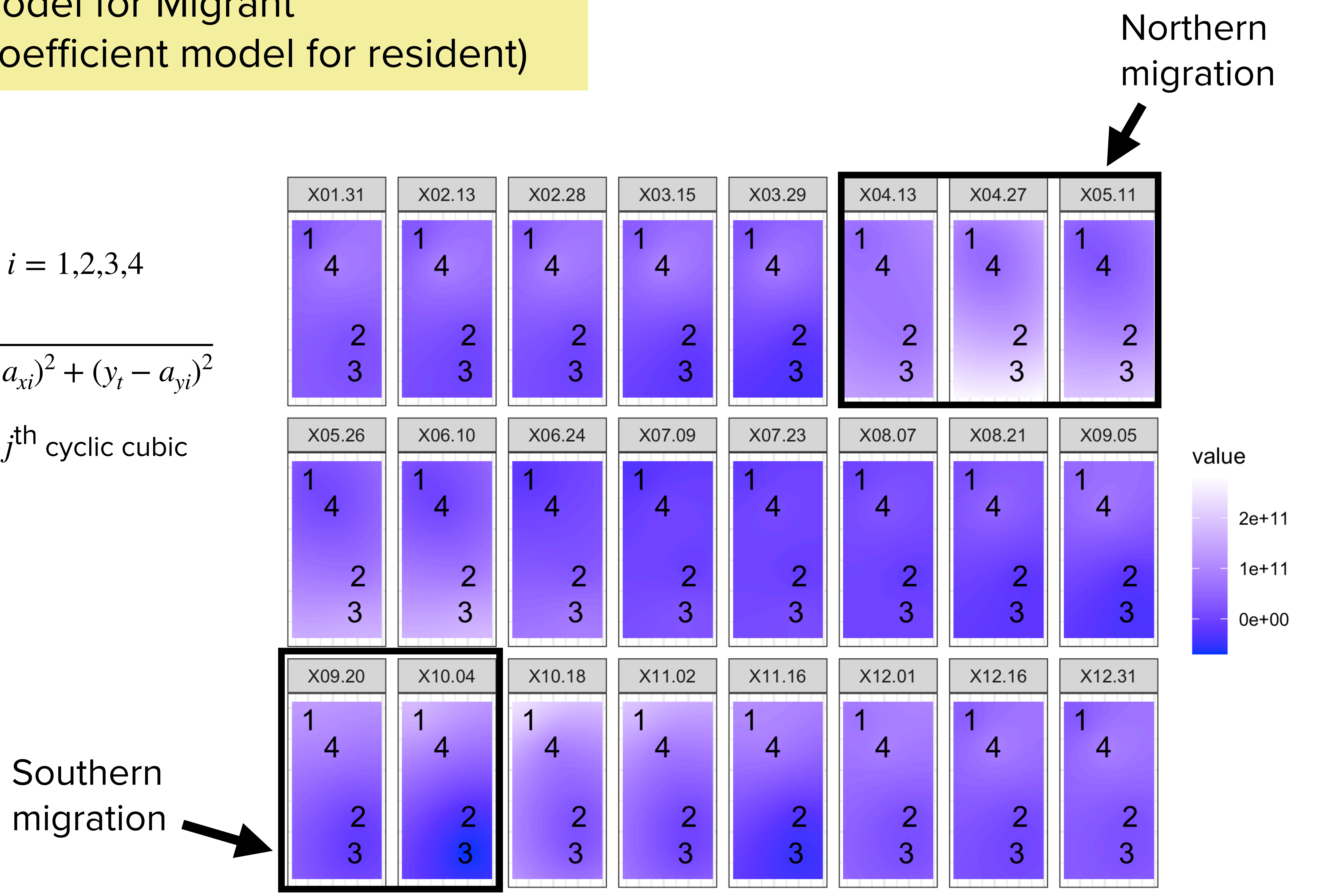


Varying Coefficient Model for Migrant
(identical to varying coefficient model for resident)

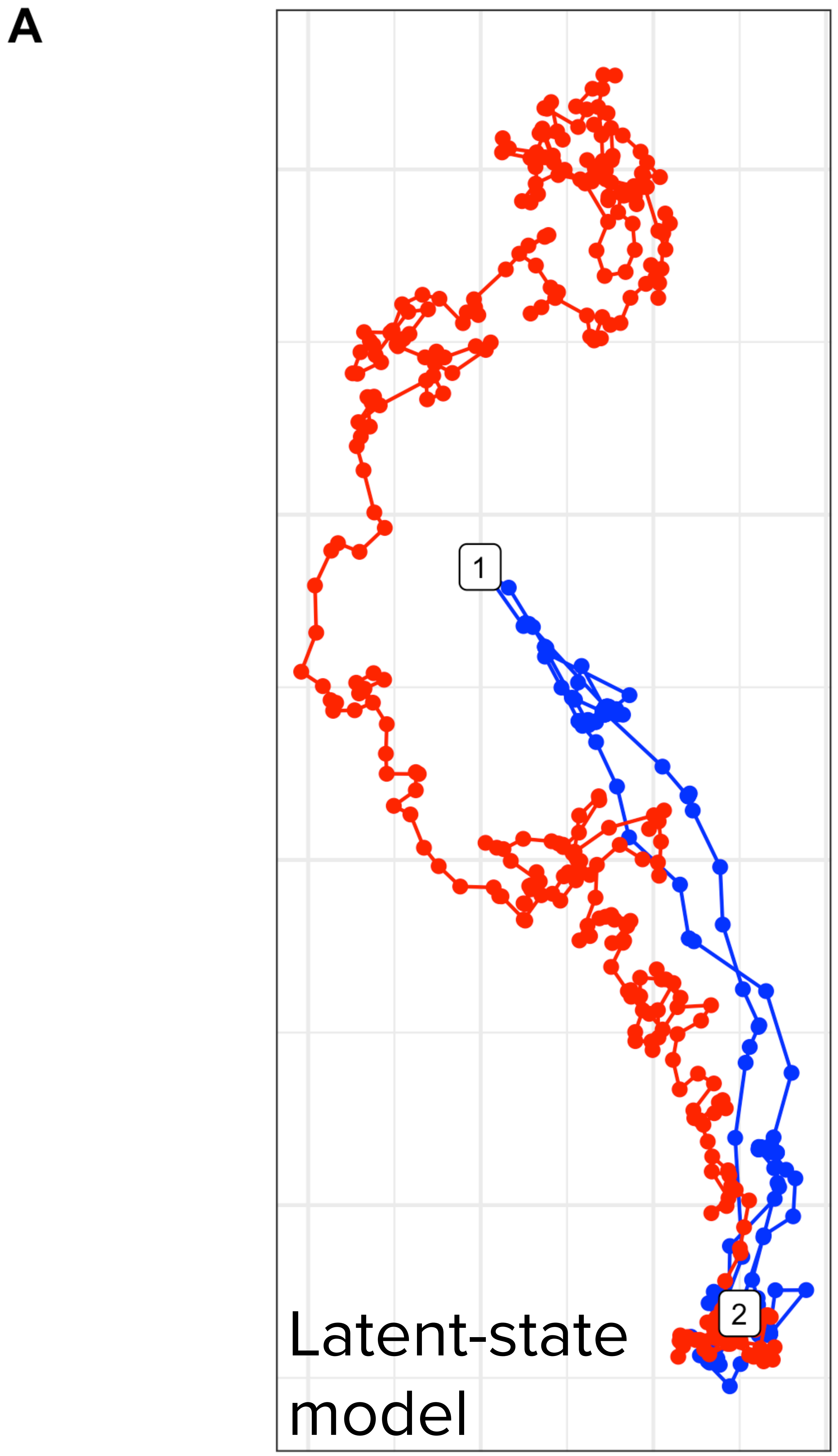
Fixed $m = 4$ attractors and k_{it} for $i = 1,2,3,4$ changes smoothly over time.

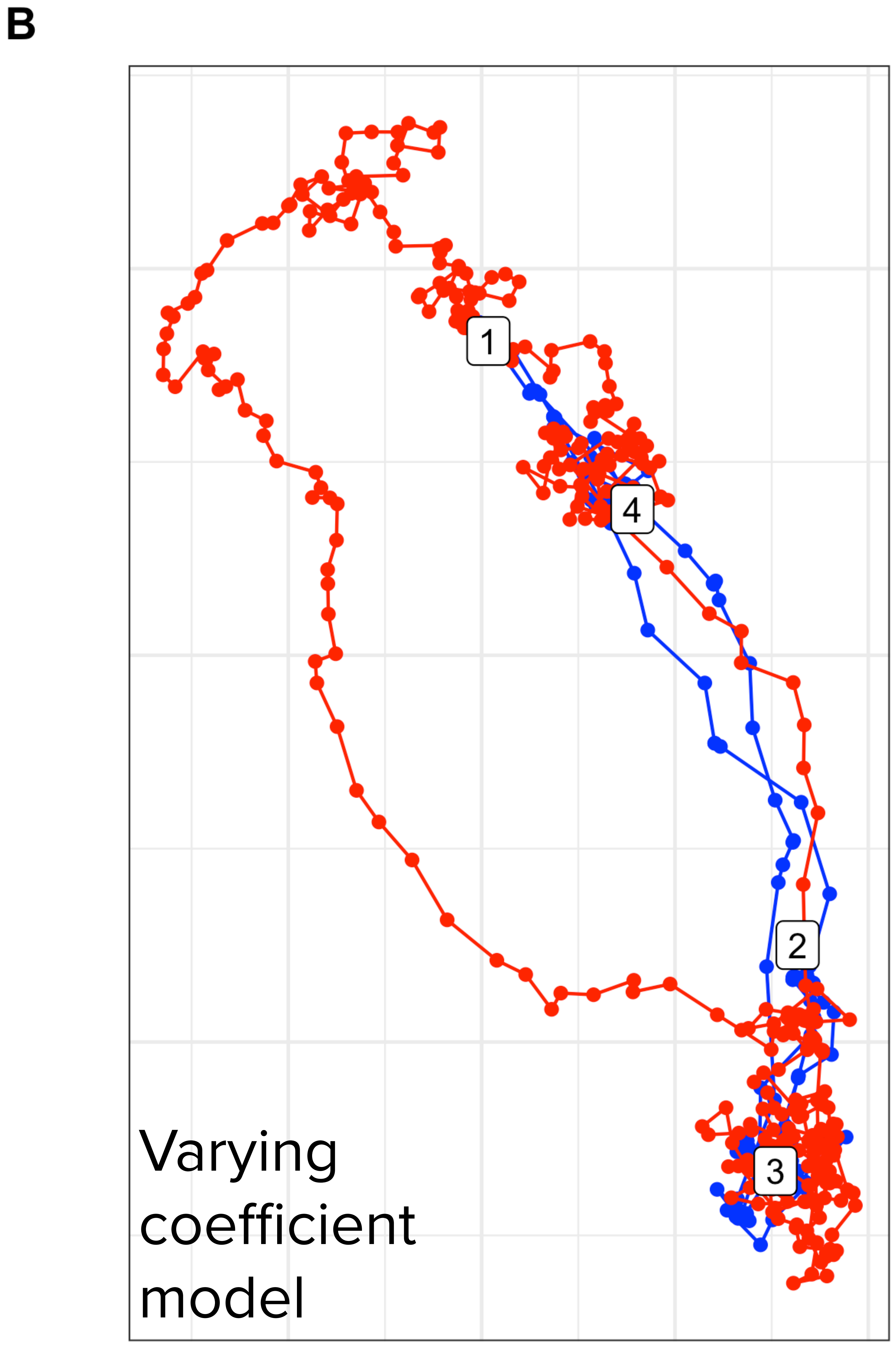
$$p(\mathbf{r}_t) = \sum_{i=1}^{m} \sum_{j=1}^{J} \alpha_{ij} B_j(t) \sqrt{(x_t - a_{xi})^2 + (y_t - a_{yi})^2}$$

where α_{ij} is the coefficient of the j^{th} cyclic cubic basis function $B_j(t)$ for attractor i



Simulations





Conclusions:

1. We described **2 classes of models** (varying coefficient and latent-state) for fitting resident, migrant, and dispersal movement strategies.
2. The varying coefficient model is more **flexible** and seems to **better explain movement behavior**.

Future work:

1. **Classify** paths as migrant, resident, or disperser.
2. Make varying coefficient model more interpretable by **restricting attractor coefficients** to be positive
3. Fit varying coefficient model in a **Bayesian framework**
4. Fit **more individuals**, including some boundary individuals

Survey of Attitudes Toward Probability

The first tools to assess students' attitudes toward statistics were implemented in classrooms in the 1980s.

In present-day, the **Survey of Attitudes Toward Statistics-36 (SATS-36)** is the most popular of the available tools and has been used in the effort to improve the introductory statistics course.

The traditional probability/inference sequence is a core component of the statistics curriculum, but the prospect of improving these courses has received relatively little attention.

We **adapt the SATS-36** to assess attitudes toward probability.

Our survey tool, the **Survey of Attitudes toward Probability (SAP)**, was administered to 15 probability sections in the fall semester of 2020 with the help of the instructors of 318, 401, 414, and 418.

We replace words related to statistics with words related to probability.

I will like **statistics**. → I will like **probability**.

I am interested in understanding **statistical information**. → I am interested in understanding **probabilistic arguments**.

We updated demographic questions and added open ended questions including:

- Describe the **difference between probability and statistics** in your own words.

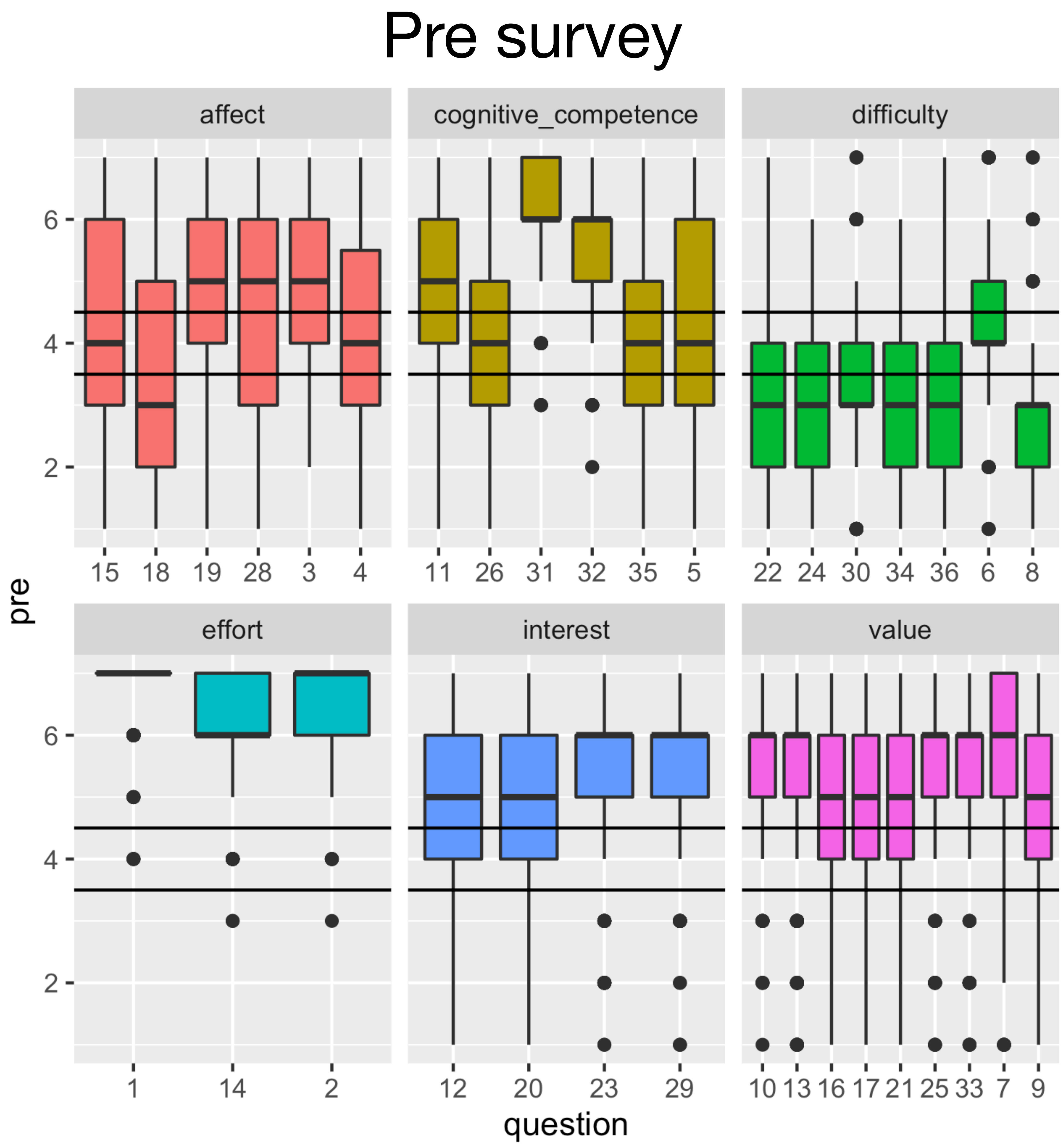
- What experiences do you believe most *positively* **influenced your current attitude** toward **probability**?

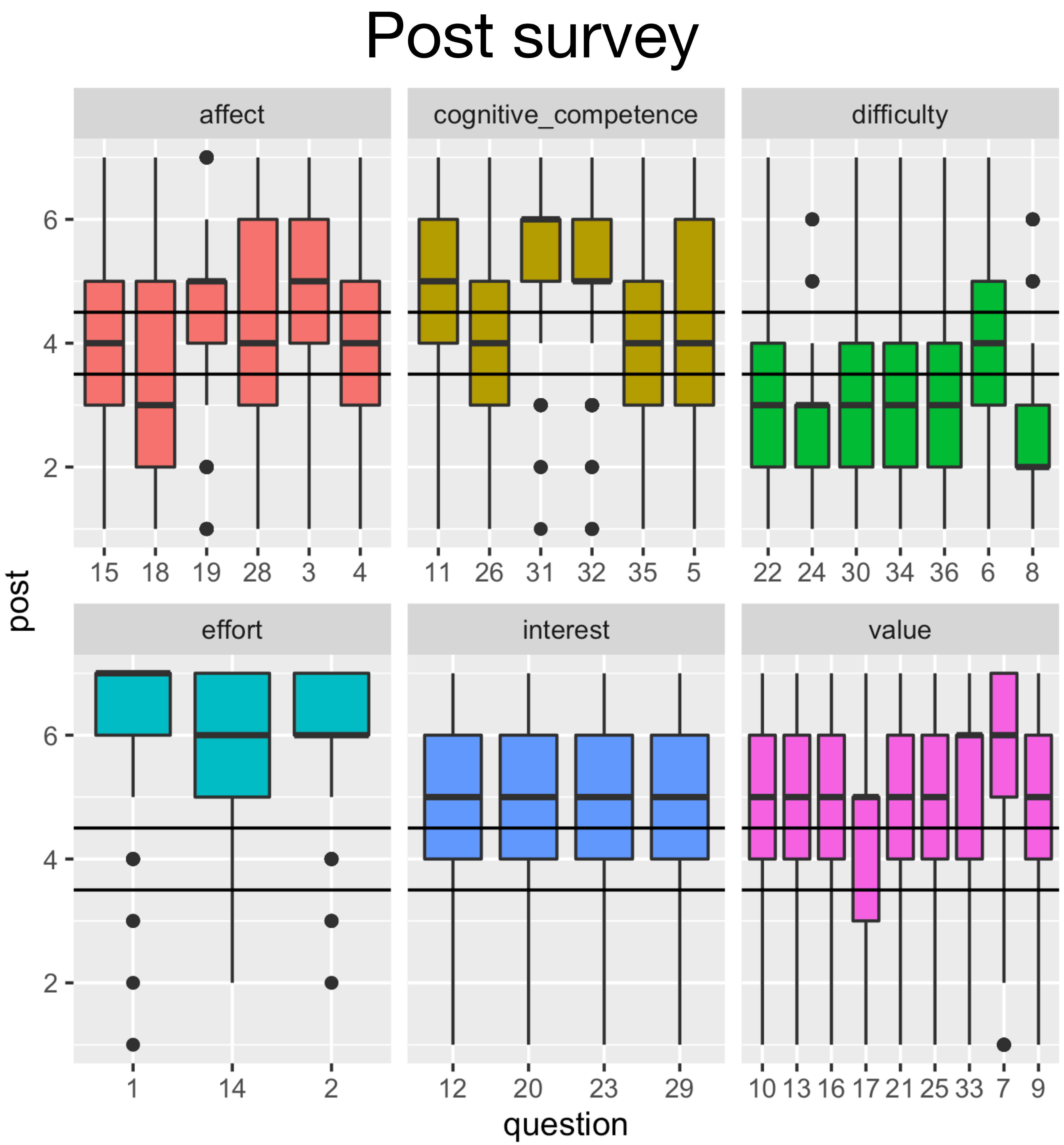
Each Likert survey question is associated with an attitude component, following Candice Schau's SATS-36.

Attitude components for us (and Schau):

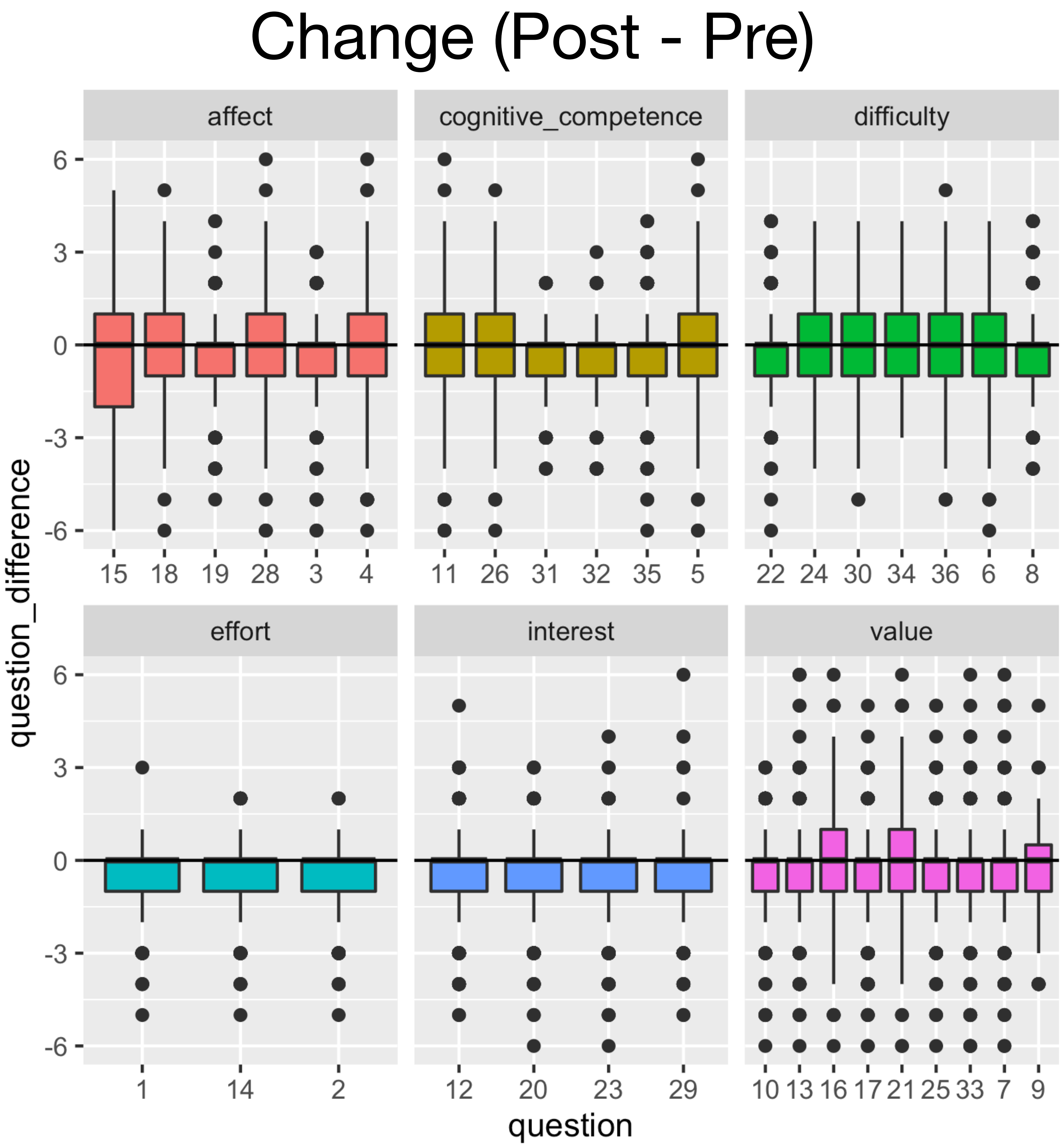
- **Affect** – students' feelings concerning probability (statistics)
- **Cognitive Competence** – students' attitudes about their intellectual knowledge and skills when applied to probability (statistics)
- **Value** – students' attitudes about the usefulness, relevance, and worth of probability (statistics) in personal and professional life
- **Difficulty** – students' attitudes about the difficulty of probability (statistics) as a subject
- **Interest** – students' level of individual interest in probability (statistics)
- **Effort** - amount of work the student expends to learn probability (statistics)

Each Likert question was associated with an attitude component (each data point is one student response)





Each question was associated with an attitude component (each data point is one student response)



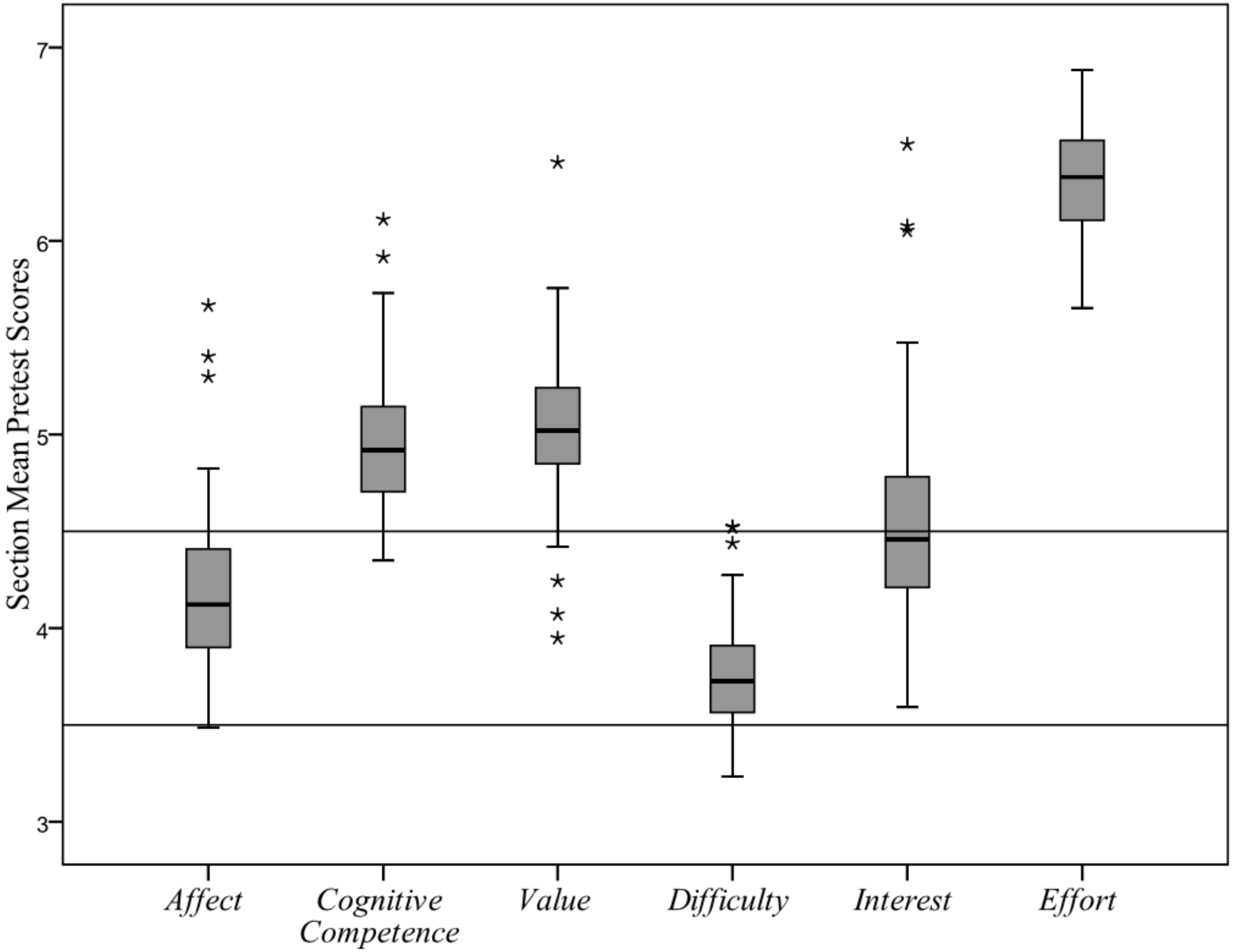
We **compare our results to the Schau (2012)** analysis of the SATS-36.

(2200 students enrolled in 101 sections of post-secondary introductory statistics service courses)

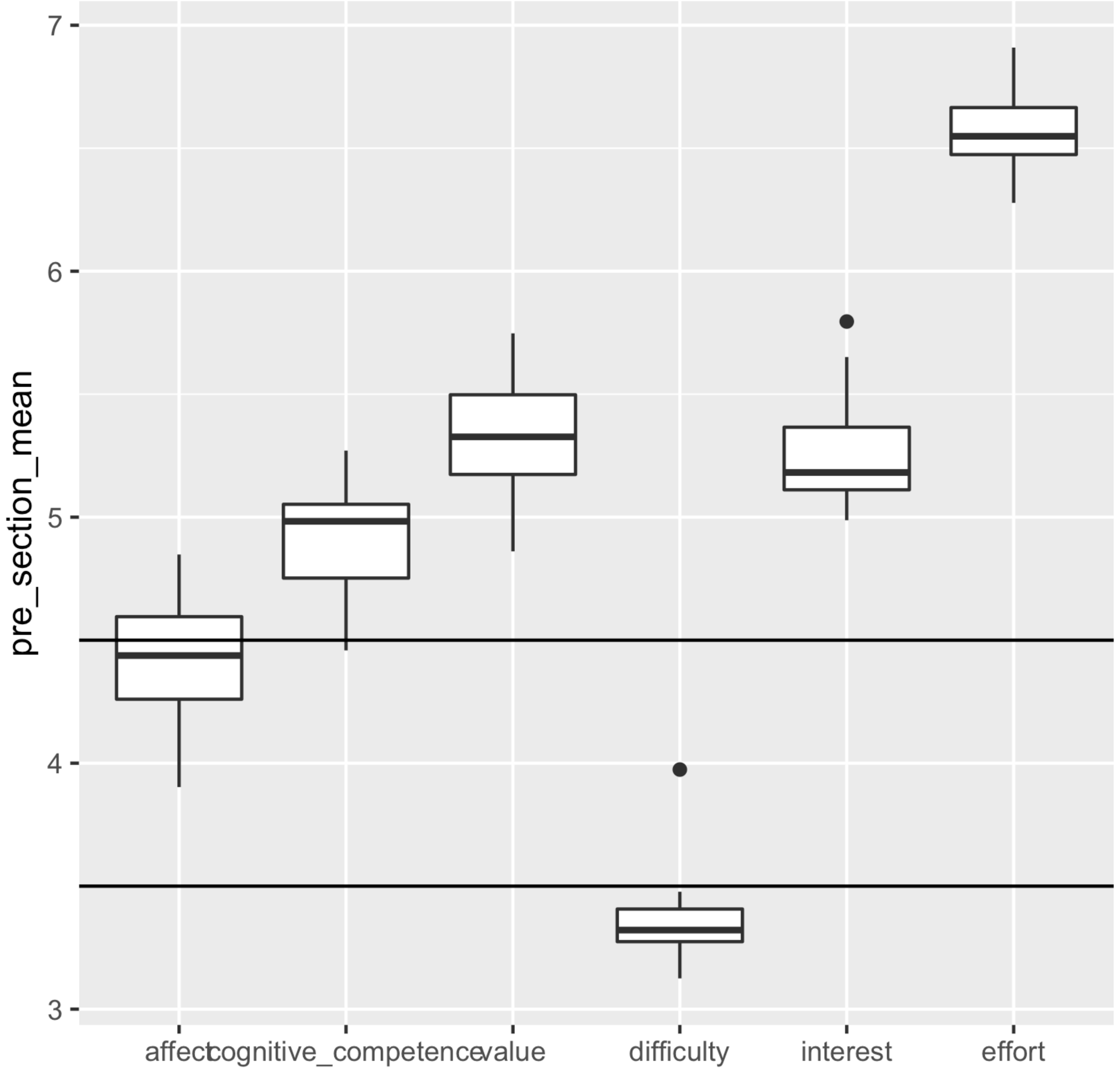
Our total sample size: 343 students in 15 sections completed both pre and post surveys.

Pre scores (section means are plotted)

Schau (2012) - Attitudes toward statistics



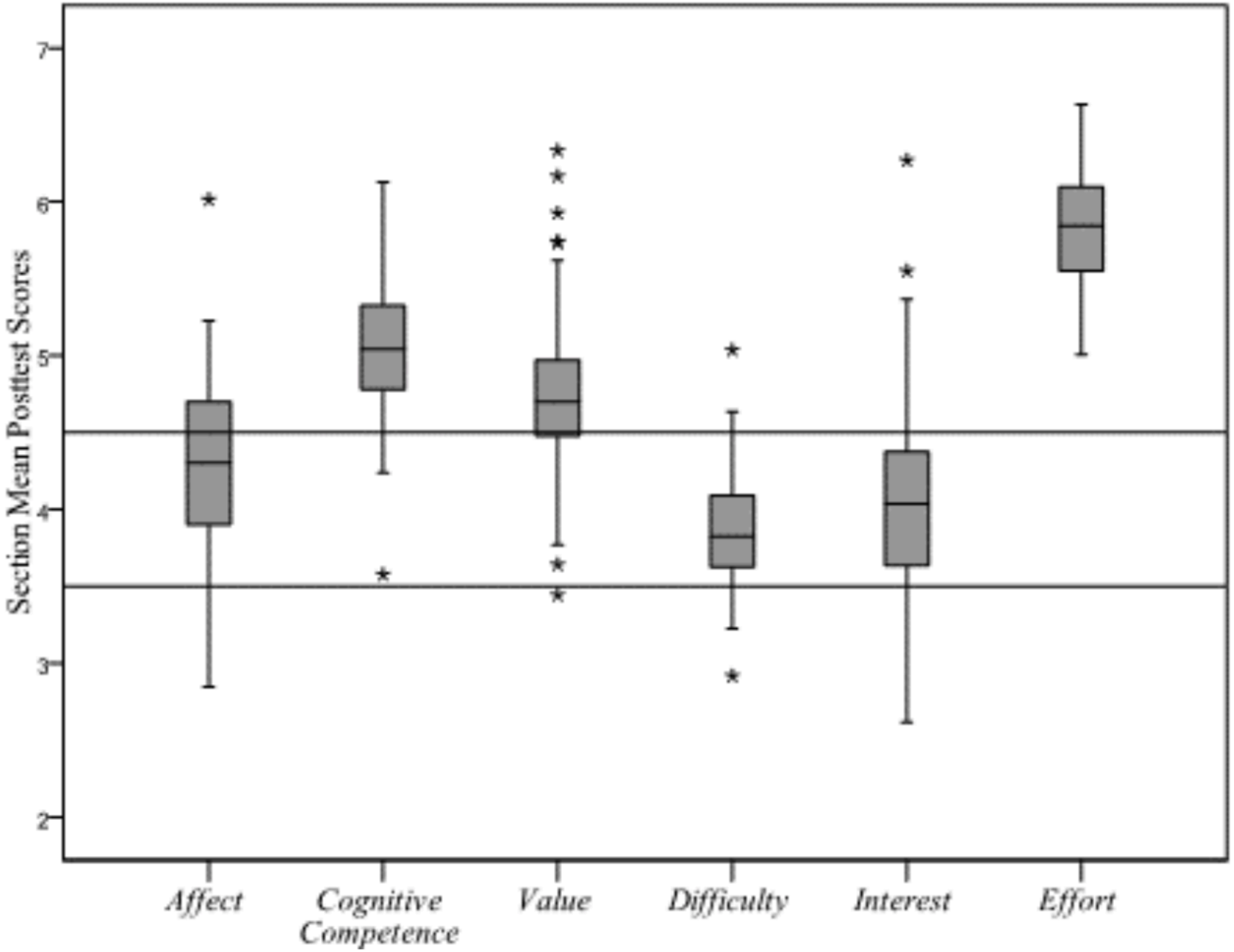
Us - Attitudes toward probability



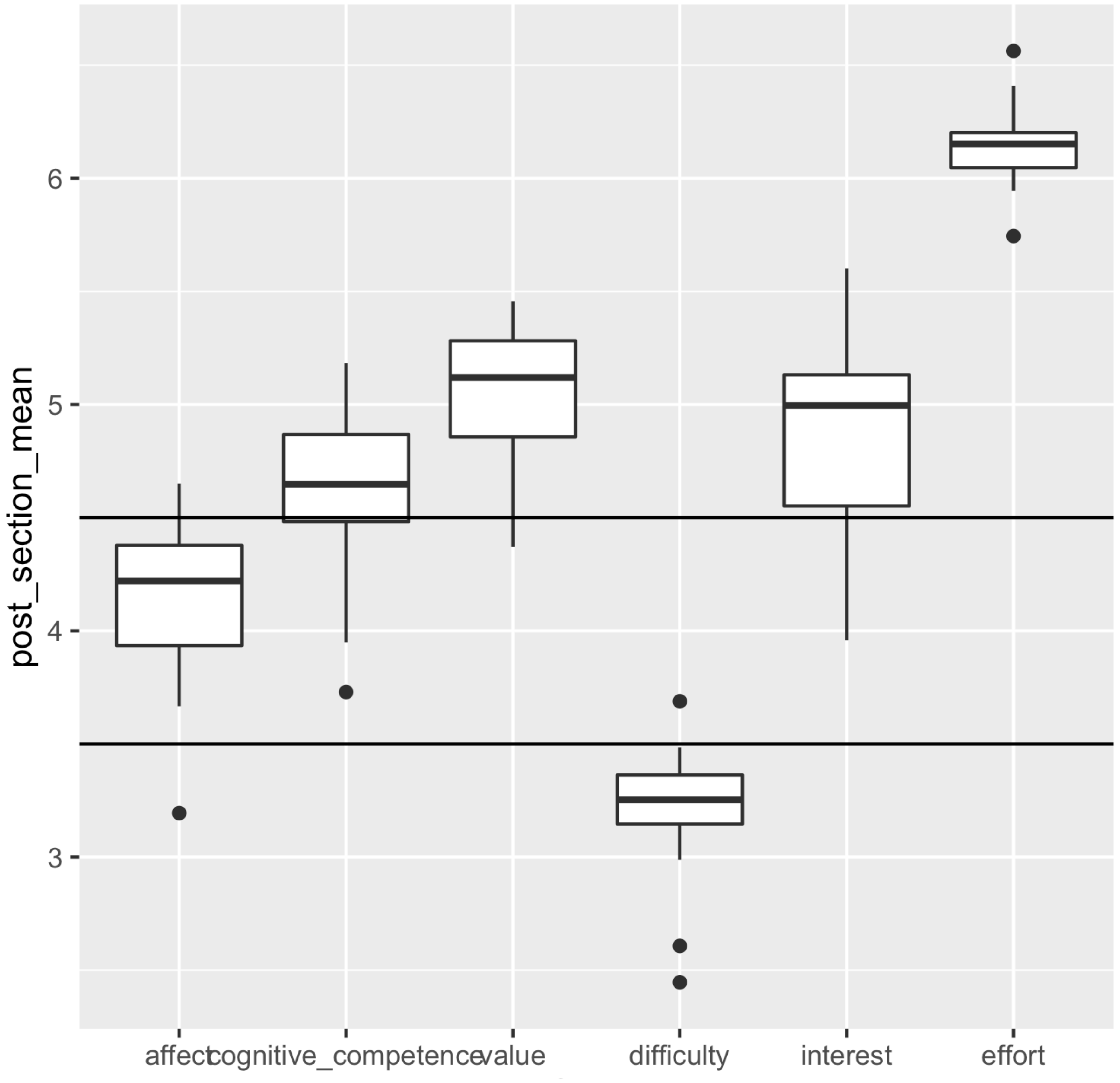
Students who took our pre survey thought probability was **more difficult** and were **more interested** in the subject. This is compared to the introductory statistics students Schau surveyed on their attitudes toward statistics.

Post scores (section means are plotted)

Schau (2012)



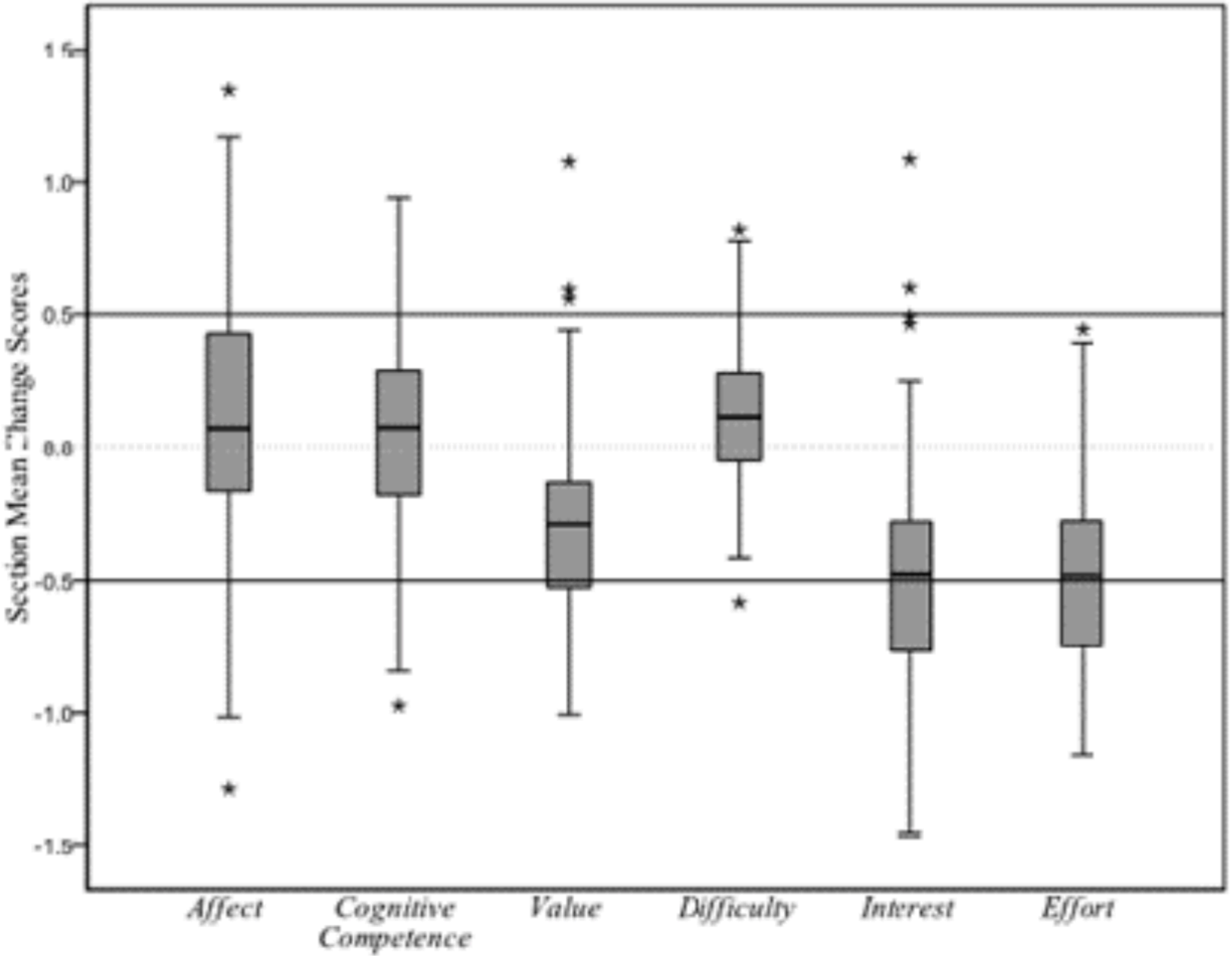
Us



The trend described on the previous slide is **also present in the post survey**.

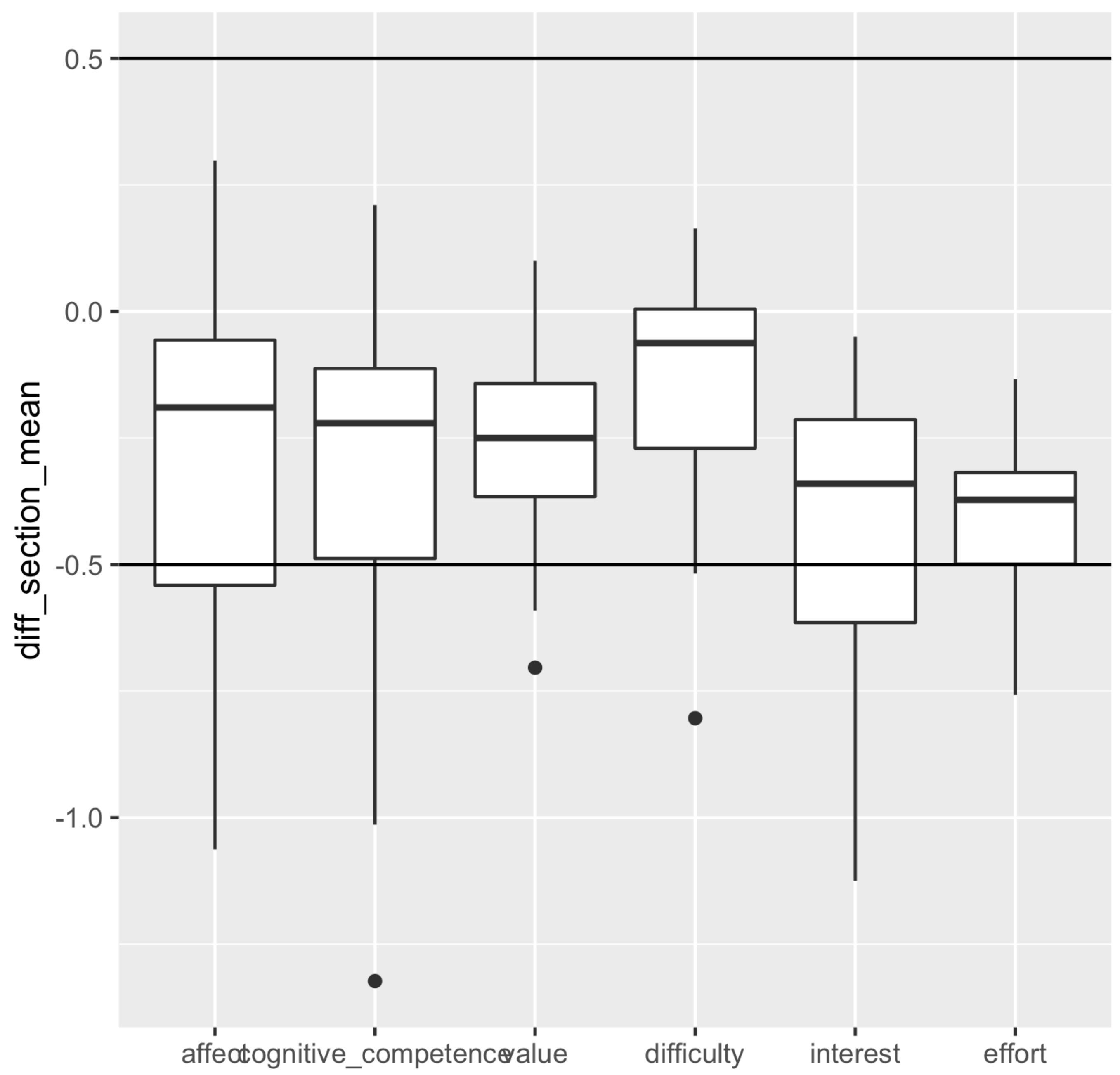
Change in scores (section means are plotted)

Schau (2012)



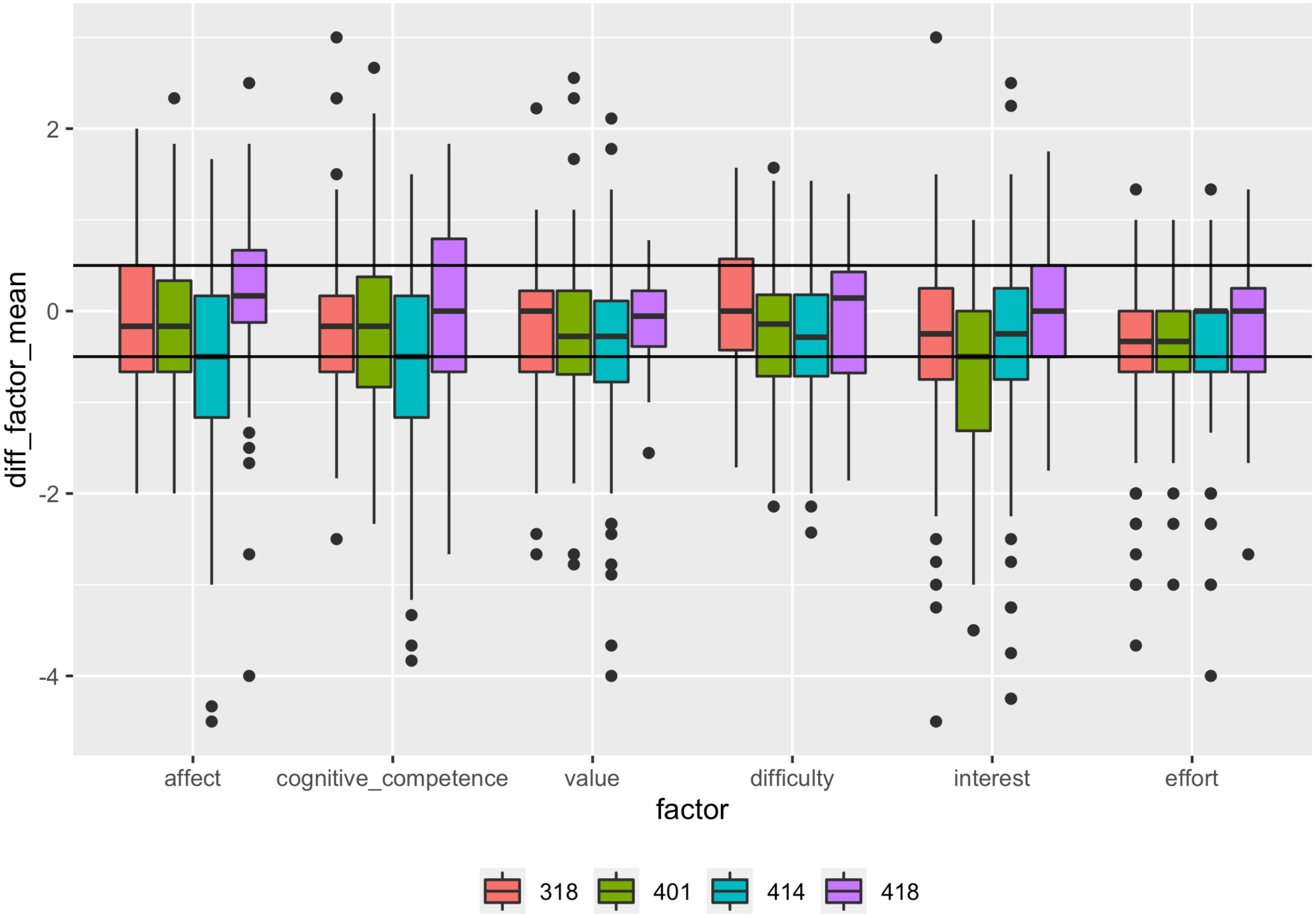
All attitude components had a slight mean **negative shift** from pre to post surveys in our data. This was true even for components with mean positive shifts in Schau's data.

Us



Change in scores by class (all students separately)

418 often has a more positive change (might have to do with type of student)



Conclusions:

1. We **adapted the SATS-36** to assess attitudes toward probability.
2. In our **preliminary analyses**, we compared SAP results and SATS-36 results.
3. The SAP could be a **tool for researchers interested in improving probability courses** in the same way that the SATS-36 has been for statistics courses.

Future work:

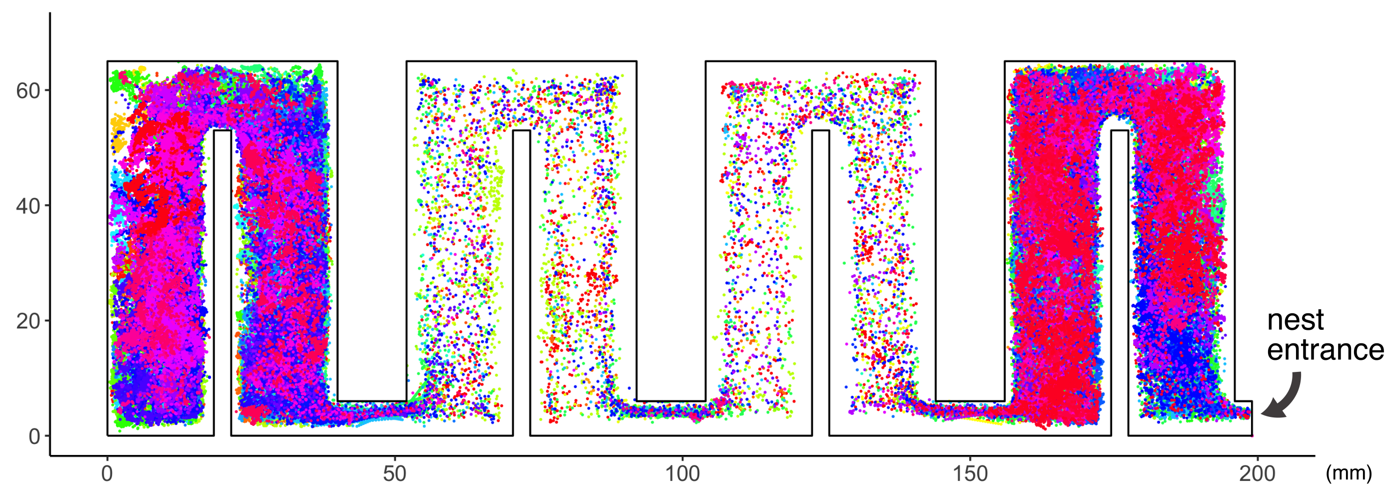
1. **More detailed analyses** of this data.
   - **Demographic information** and math background could help explain variability.
   - Examine **open-ended questions**.
   - Assess the **internal consistency** of the SAP.
2. We are collecting more data using an updated survey in the **Spring of 2021** from 21 sections, including some non-probability courses.
   - Over 600 responses to the pre SAP from probability sections

Thanks to all the instructors for their help!

Projects

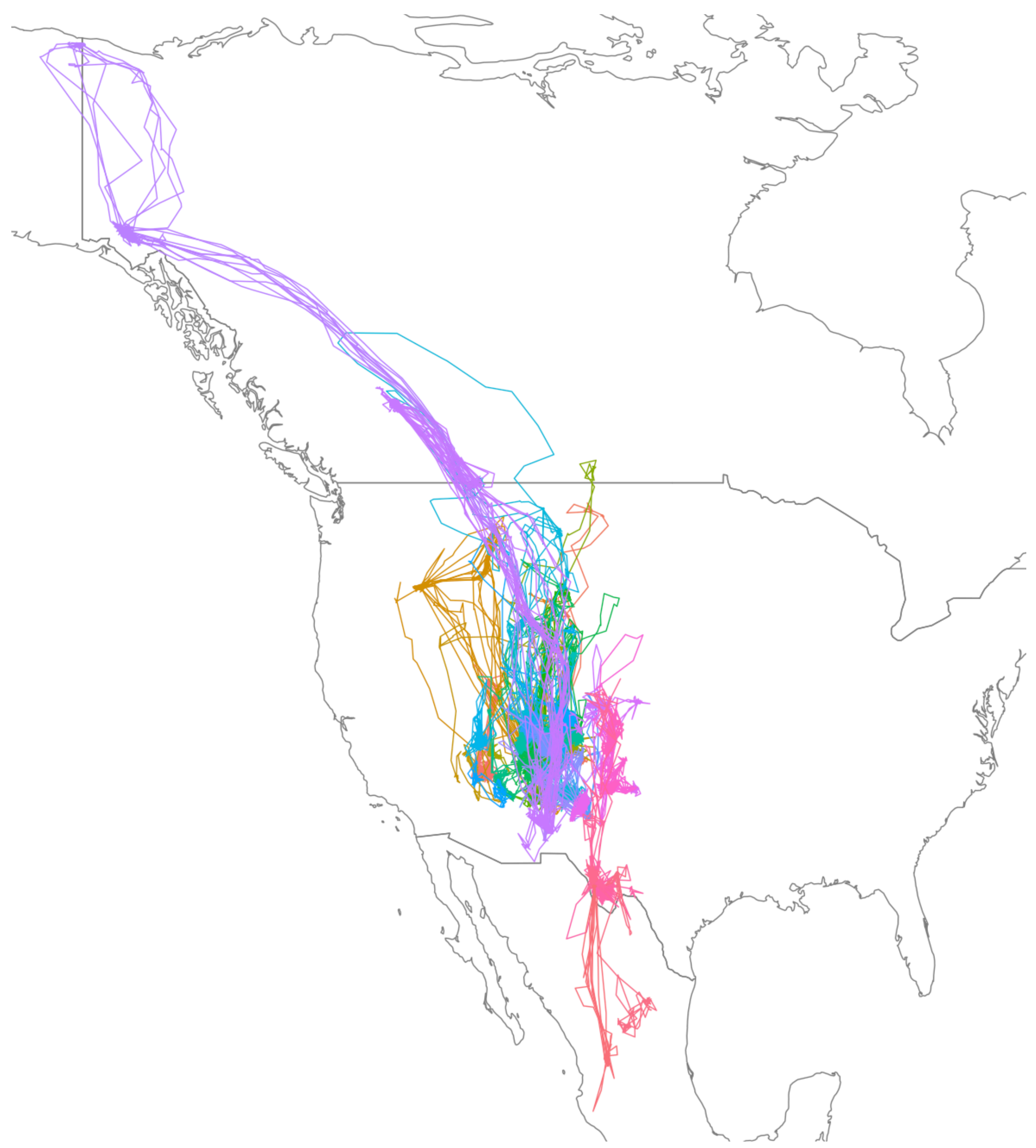
1

A Lattice and Random Intermediate Point Sampling Design for Animal Movement



2

Modeling Yearly Patterns in Golden Eagle Movement



3

Survey of Attitudes toward Probability