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

Elizabeth Eisenhauer & Ephraim Hanks

Penn State University

bit.ly/isec-liz

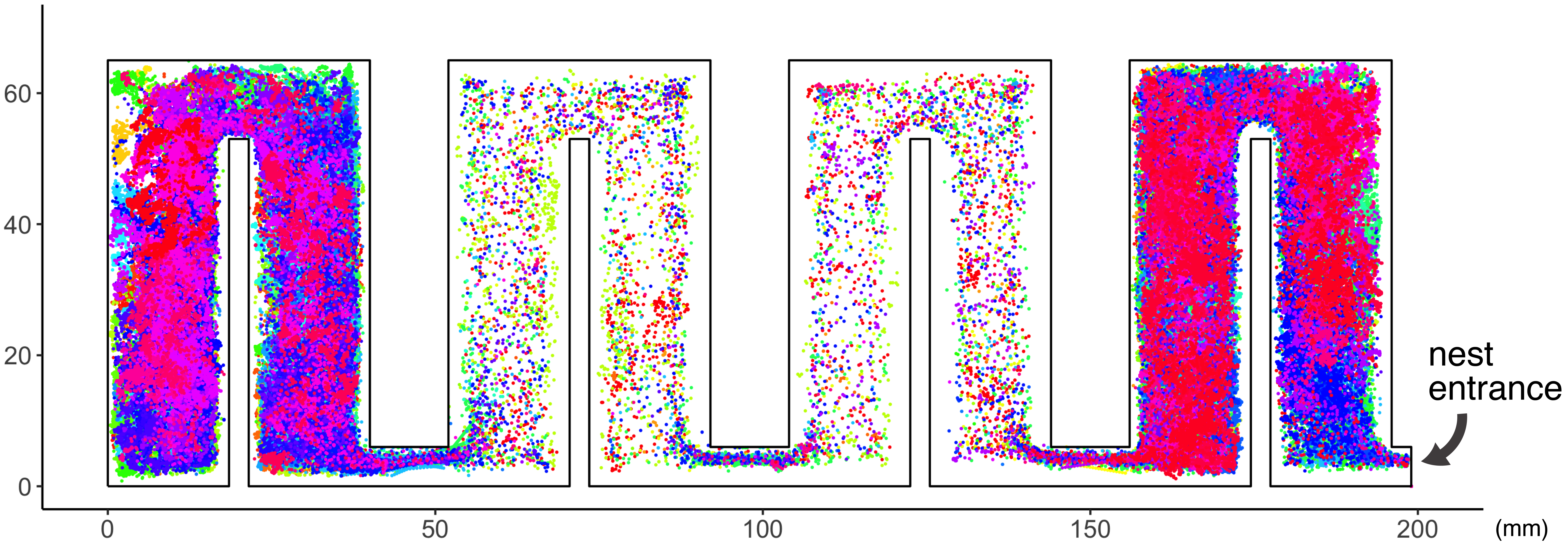
eisenhauer@psu.edu

@LizStats

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).



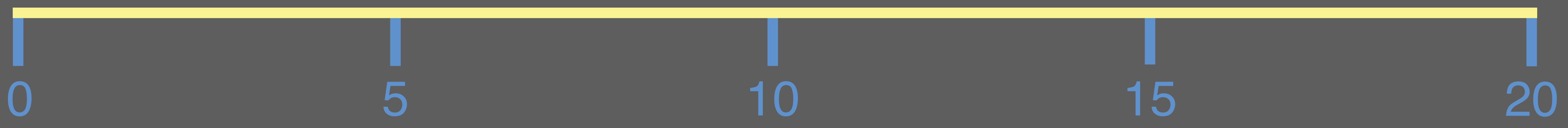
We were approached by the researchers with the **scenario**:

- Next time, we will collect **lower resolution** data.
- How should we do this to **minimize the loss of information** about movement behavior?

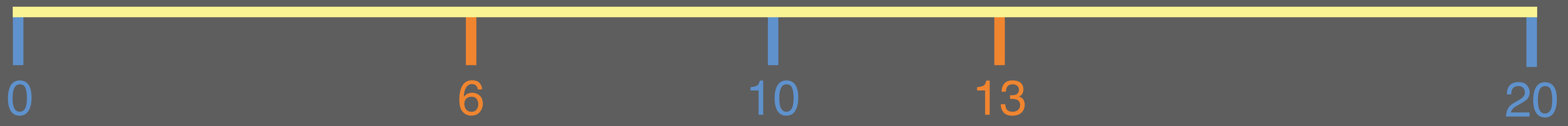
This question is **relevant to many researchers** collecting animal movement data.

2 sampling designs:

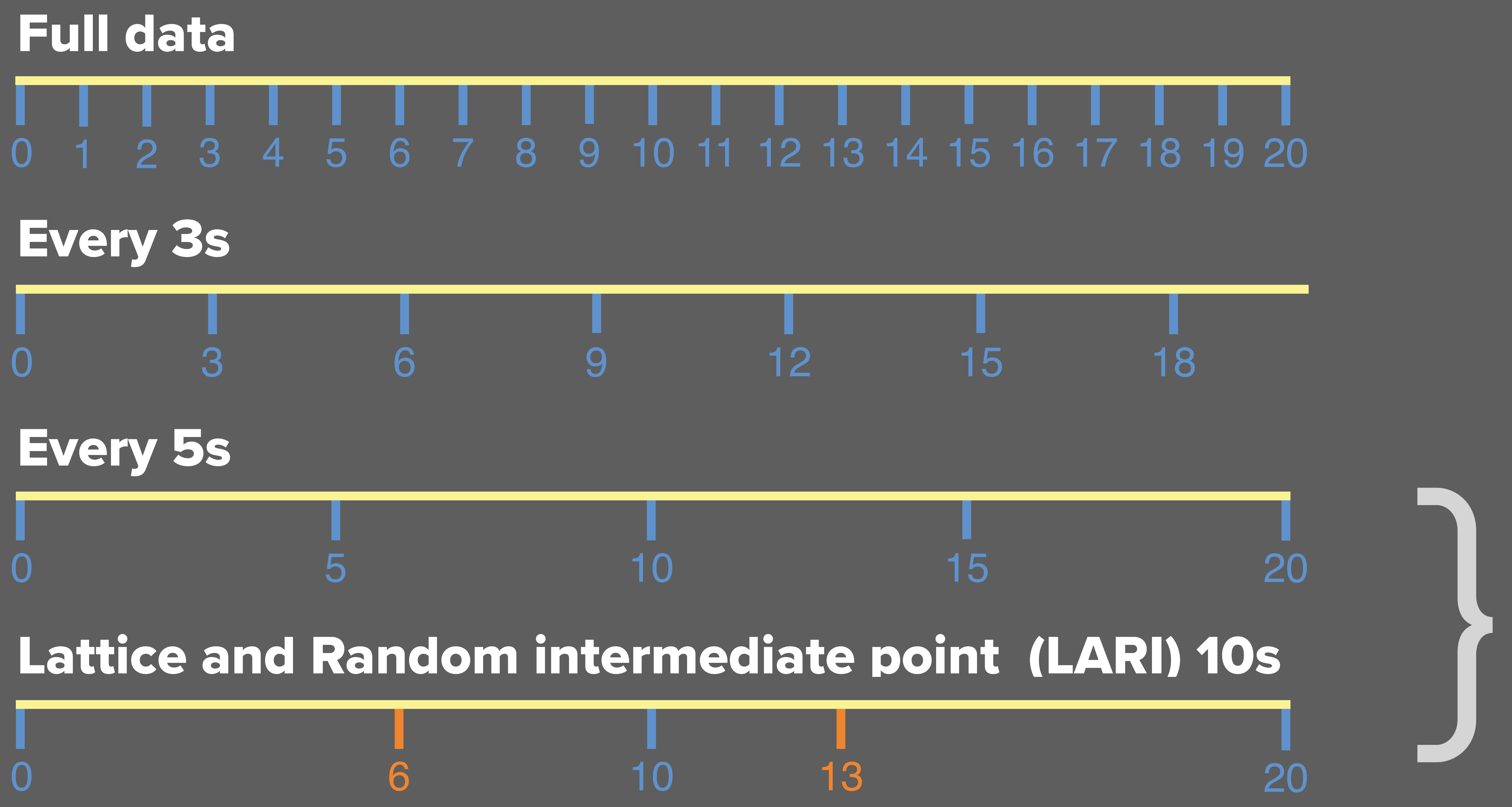
REGULAR



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

Stochastic differential equation (SDE) model for animal movement

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

SDE model framework:

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

Utilizing motility and potential surfaces, define:

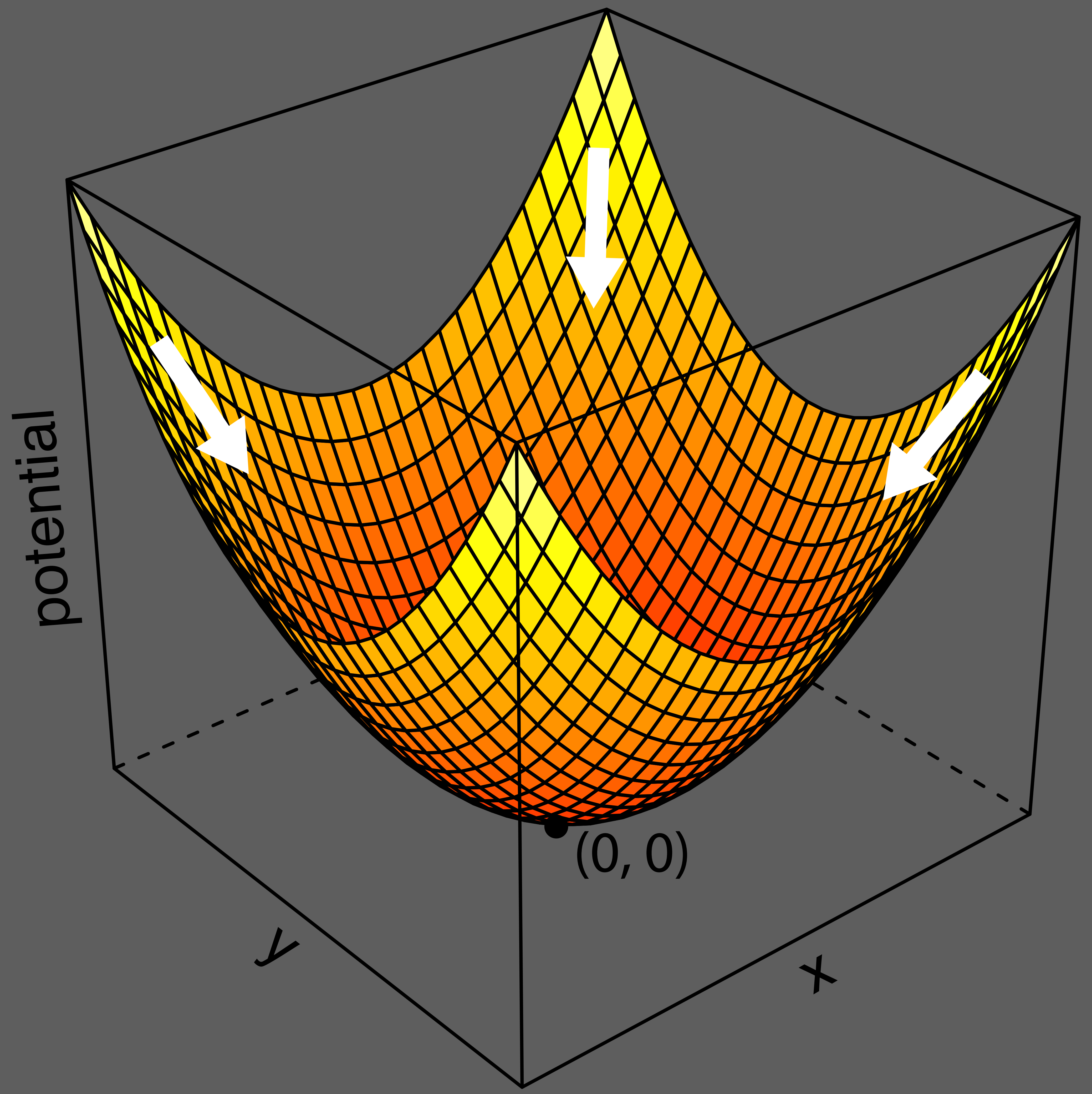
$$\boldsymbol{\mu}(\mathbf{x}_t) = m(\mathbf{x}_t)\ [-\nabla p(\mathbf{x}_t)]$$ (mean drift)

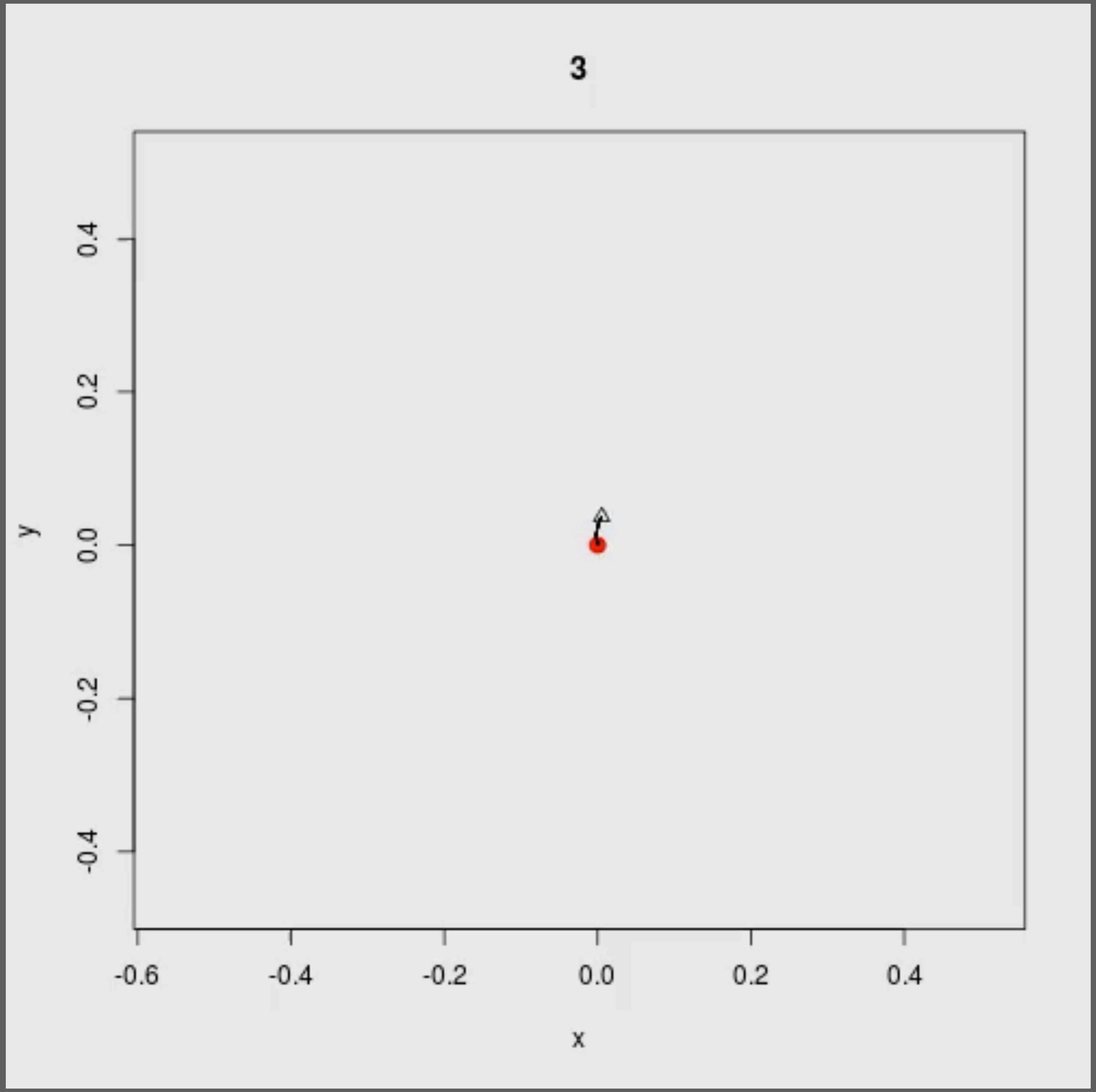
$$c(\mathbf{x}_t) = \sigma\ m(\mathbf{x}_t)$$ (magnitude of stochasticity)

(James Russell et al., 2018)

We describe animal movement using a stochastic differential equation model with 2 parameters:

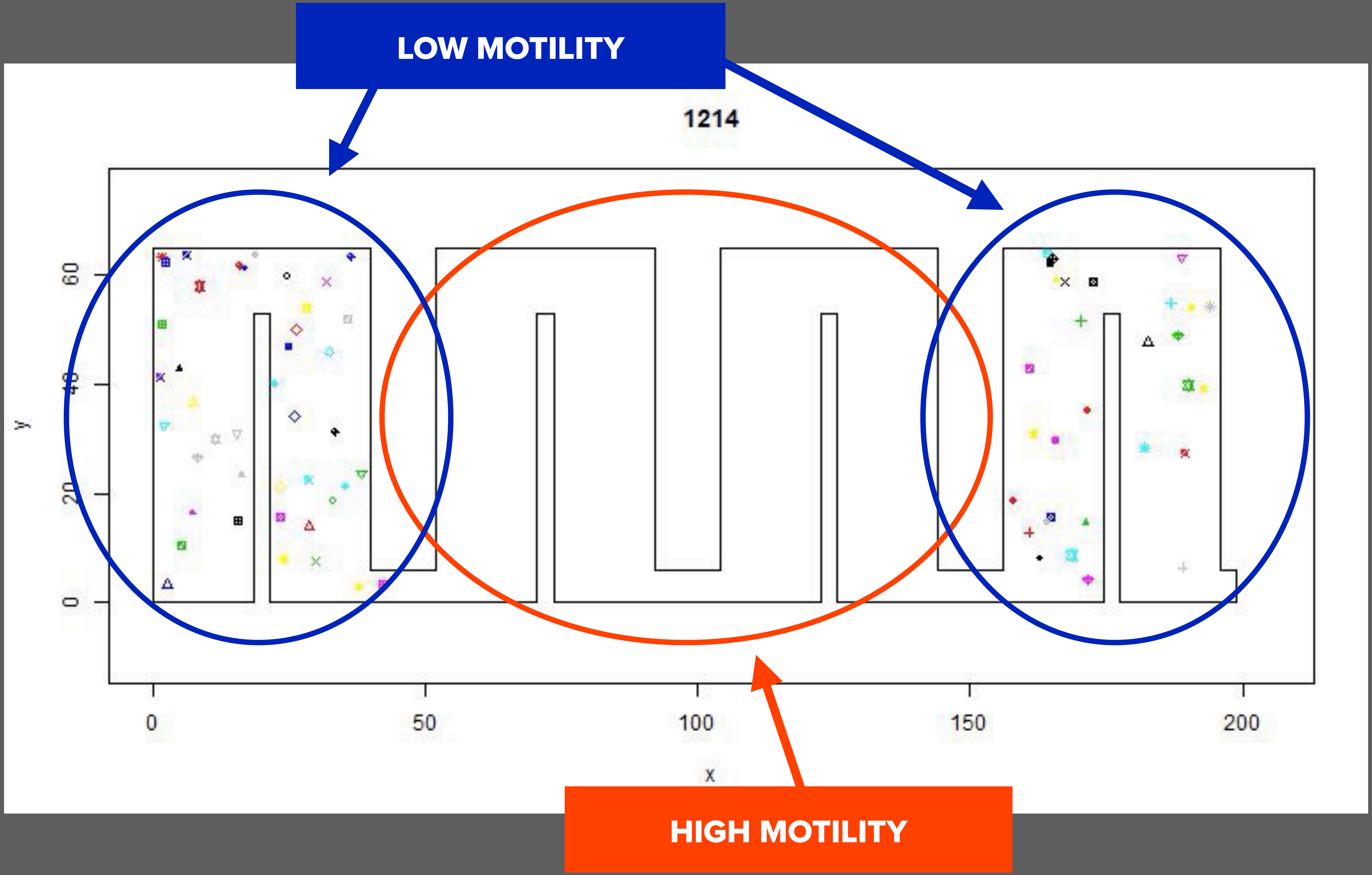
1. POTENTIAL SURFACE





We describe animal movement using a stochastic differential equation model with 2 parameters:

2. MOTILITY SURFACE



Since we don't observe animal movement in continuous time, we **numerically approximate derivatives**.

$$\frac{d\mathbf{x}_\tau}{dt} \approx \frac{\mathbf{x}_{\tau+1} - \mathbf{x}_\tau}{h_\tau}$$

$$\frac{d\mathbf{v}_\tau}{dt} \approx \frac{\mathbf{v}_{\tau+1} - \mathbf{v}_\tau}{h_\tau} \approx \frac{\mathbf{x}_{\tau+2} - \mathbf{x}_{\tau+1}}{h_\tau h_{\tau+1}} - \frac{\mathbf{x}_{\tau+1} - \mathbf{x}_\tau}{h_\tau^2}$$

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

where

- $\mathbf{x}_\tau$ 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**

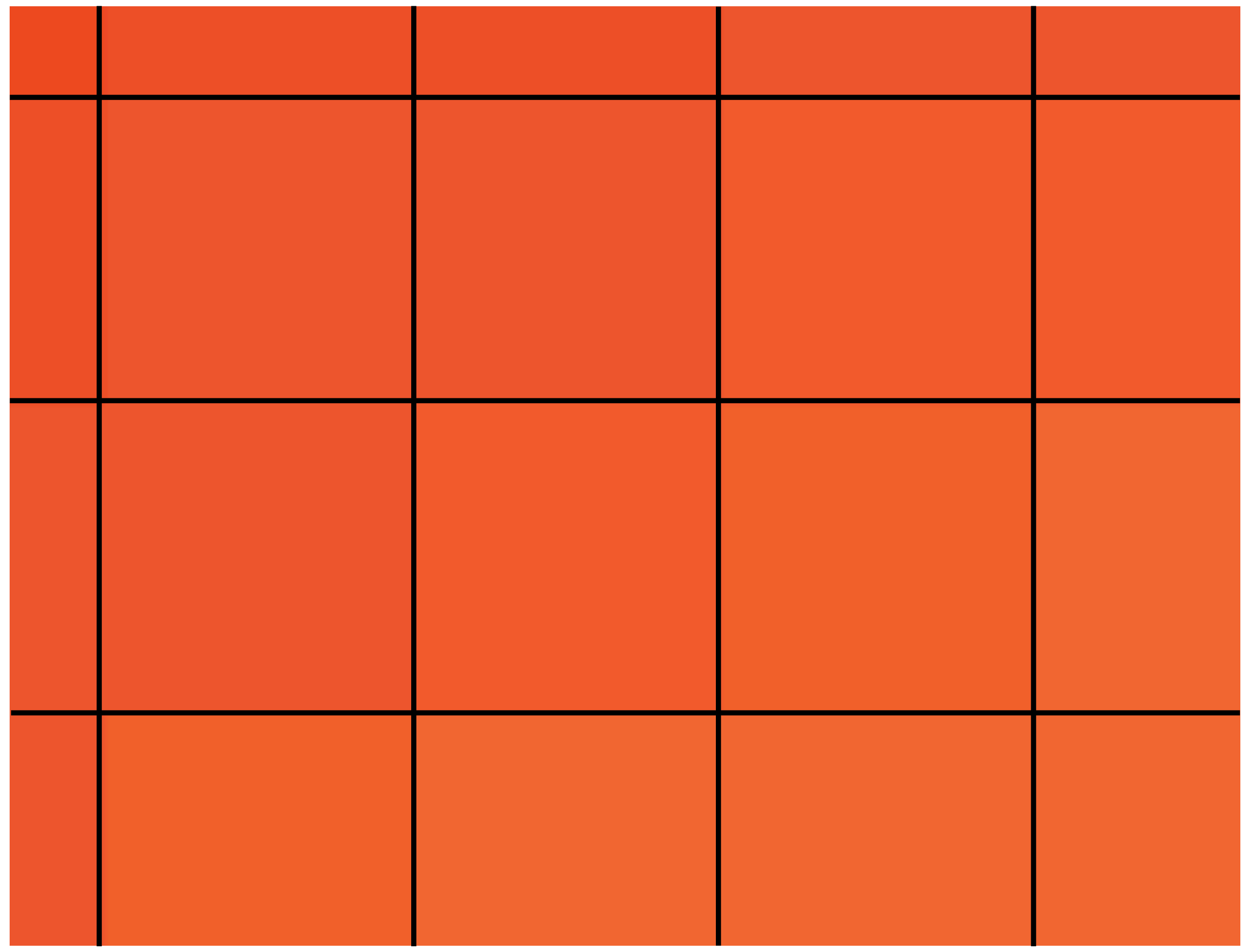
$$\mathbf{x}_{\tau+2} = \left(1 + \frac{h_{\tau+1}}{h_\tau} - \beta h_{\tau+1}\right)\mathbf{x}_{\tau+1} + \left(\beta h_{\tau+1} - \frac{h_{\tau+1}}{h_\tau}\right)\mathbf{x}_\tau + \beta\ h_\tau h_{\tau+1}\ m(\mathbf{x}_\tau)\ [-\triangledown p(\mathbf{x}_\tau)] + \sigma\ m(\mathbf{x}_\tau)\ h_{\tau+1}h_\tau^2\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

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

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

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

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



Estimate motility surface (***m***) and potential surface (***p***)

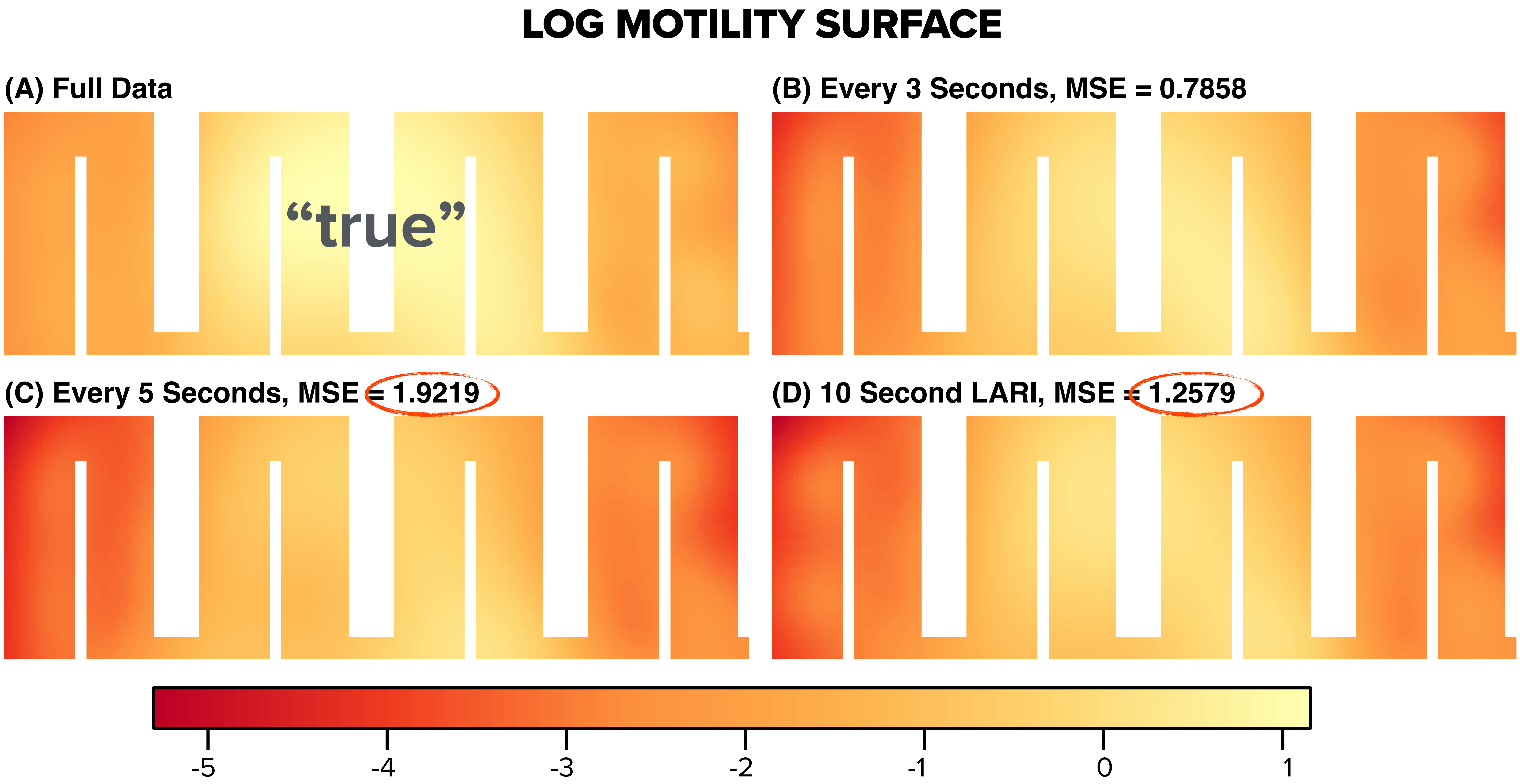
Penalize the roughness of ***m*** and ***p***

Smoothness parameters are chosen with a holdout set.

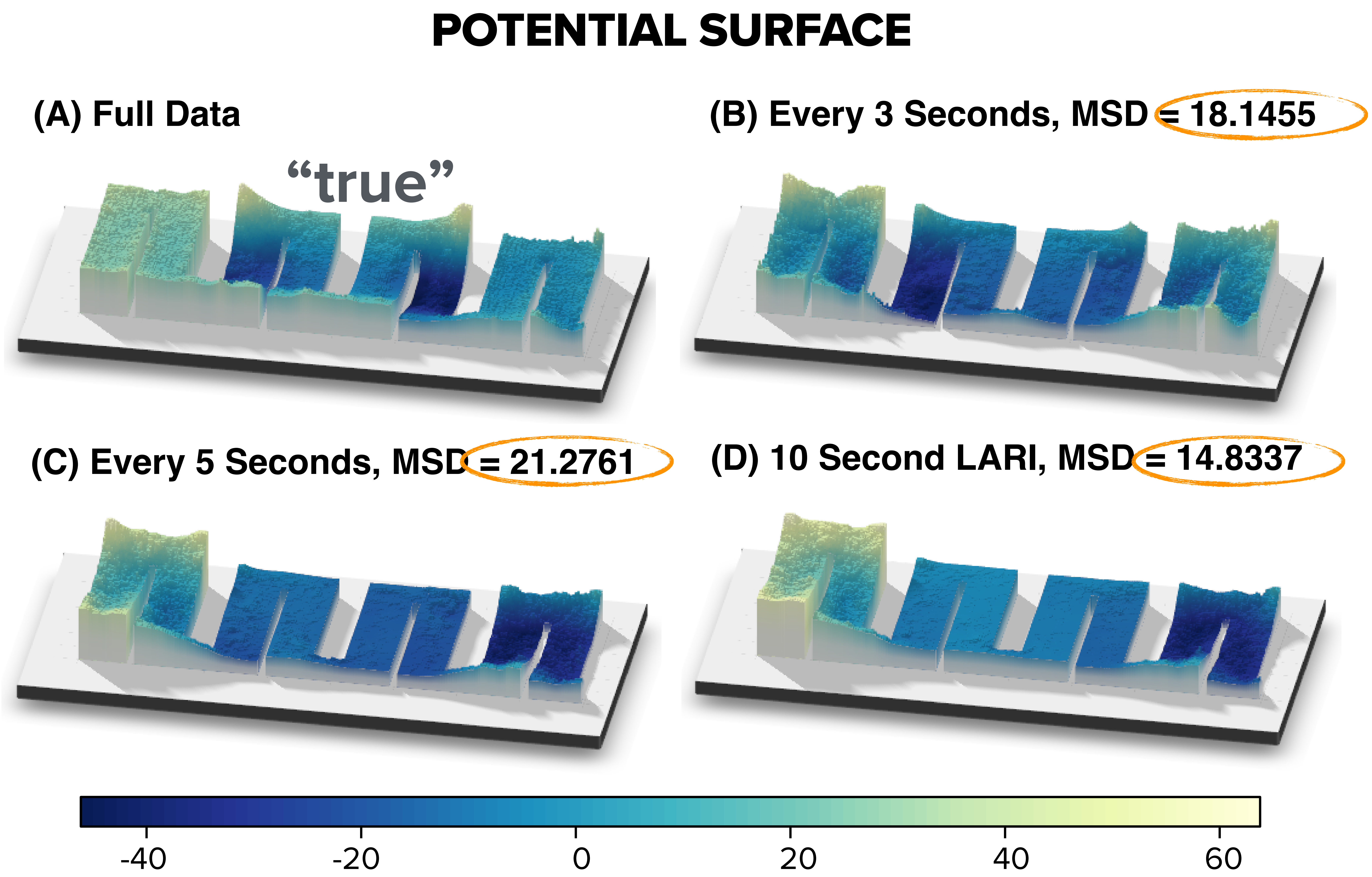
REML-style approximation
- First **estimate variance terms** to get ***m***
- Then **estimate mean** drift to get ***p***

Computing time **~20 minutes** (single core)
- 14,401 x 78 data points

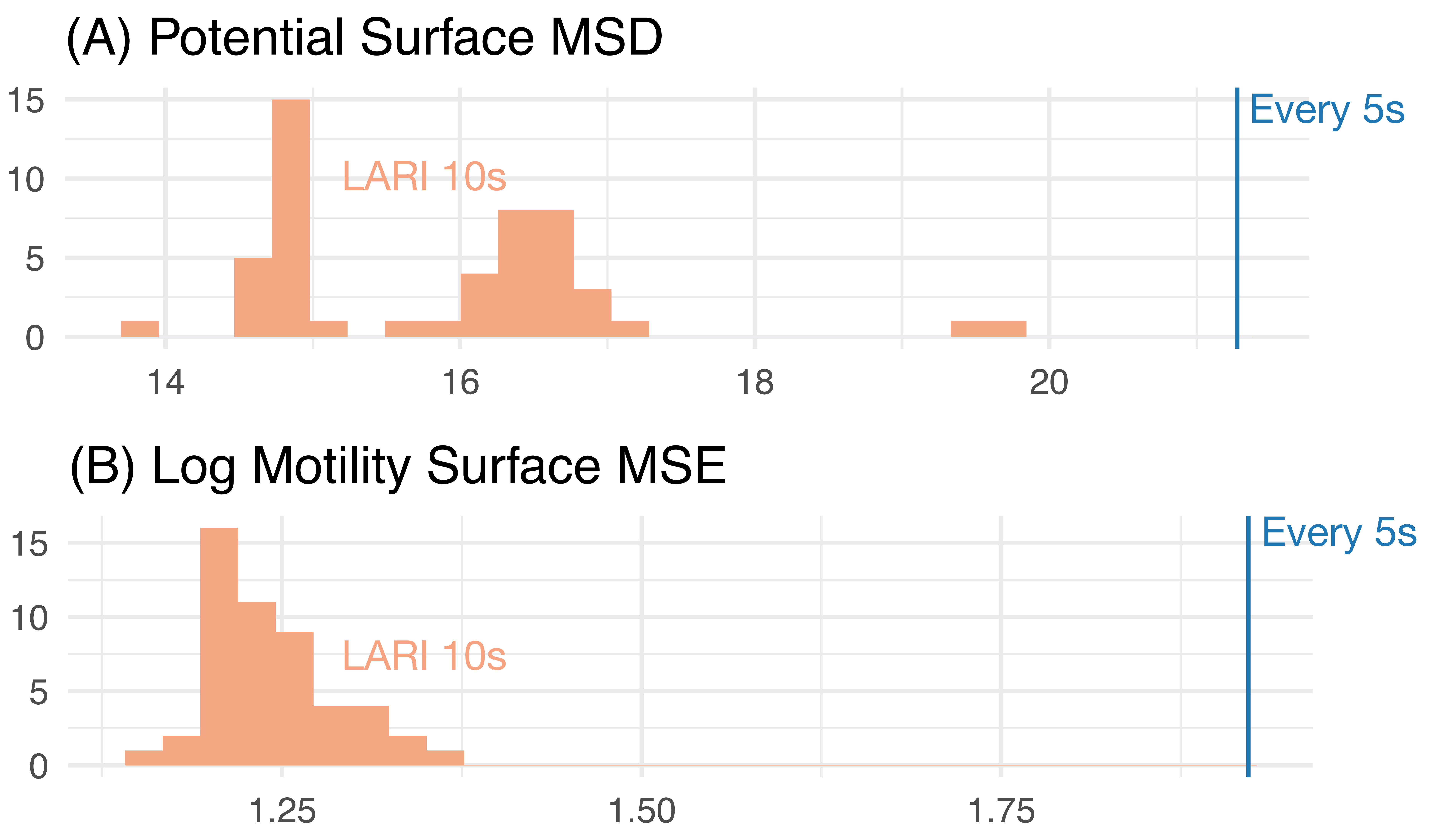
We **compare motility and potential** surfaces estimated with the 4 subsamples using multiple metrics.



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



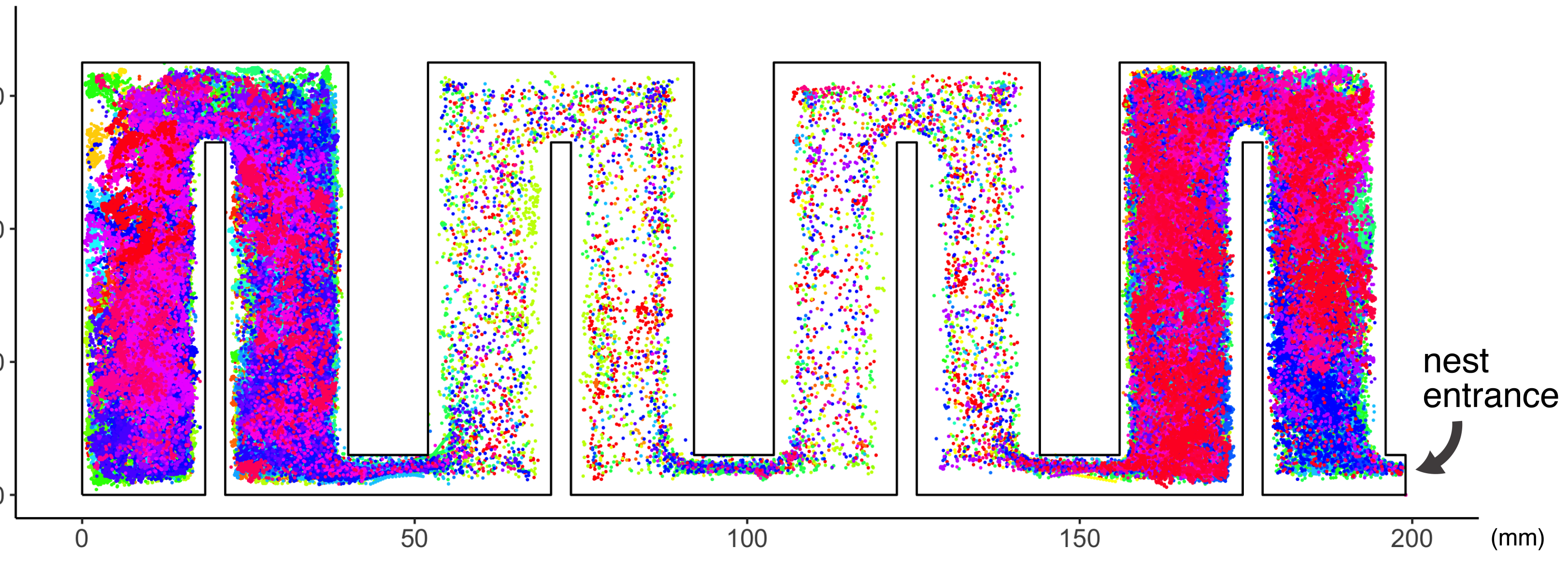
We fit 50 different LARI subsamples to understand random variation.



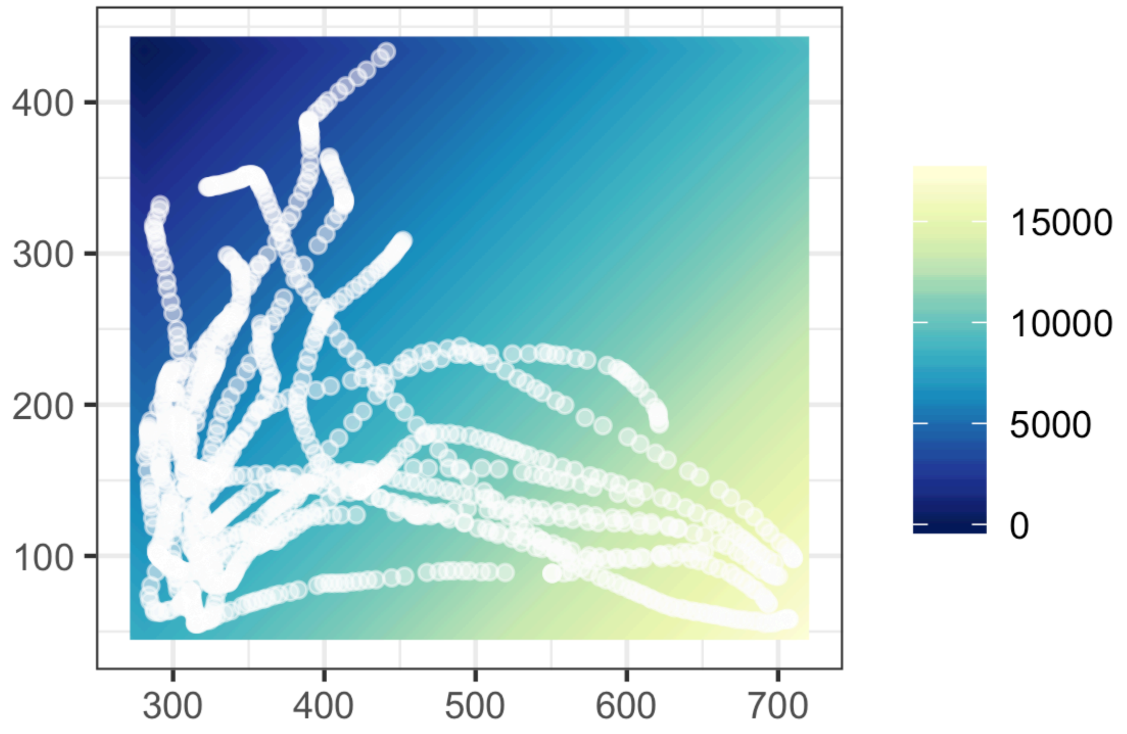
Result: **LARI sampling was better** than regular sampling overall for understanding movement behavior. A simulation study and additional data example support this conclusion. It may also be better for estimating missing data.

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

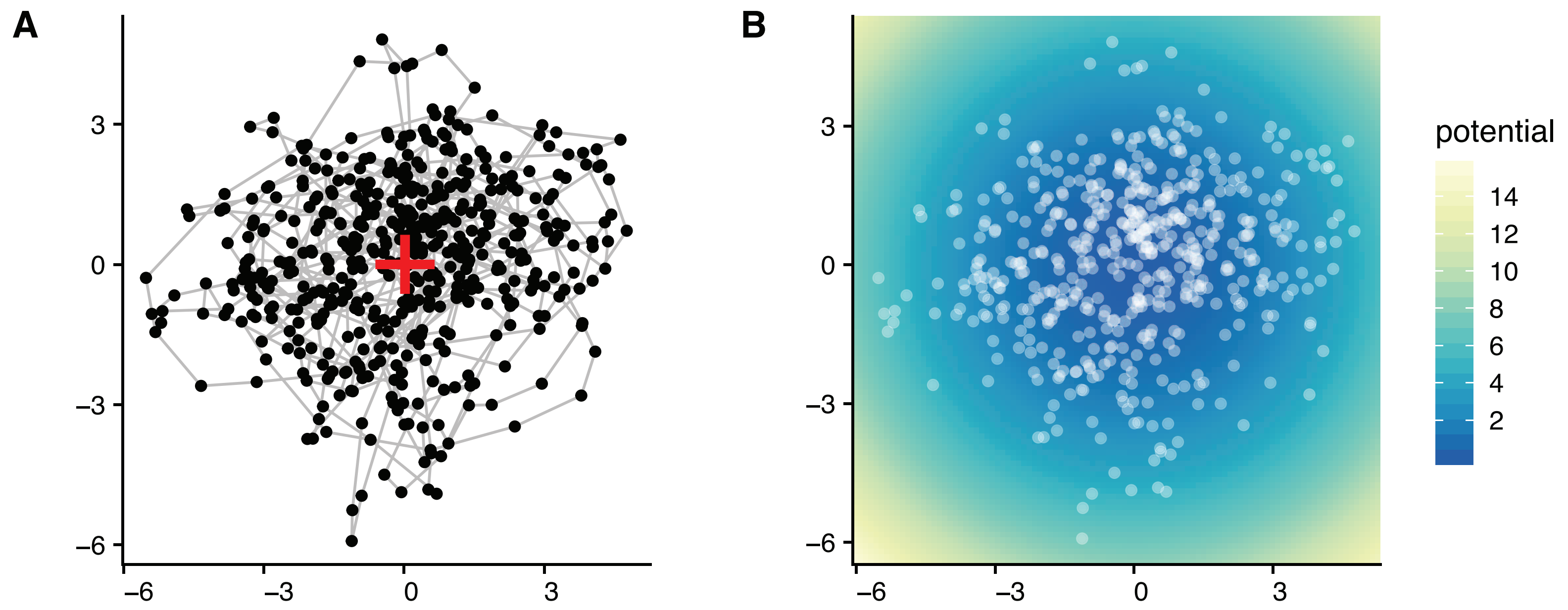
Ant data



Guppy data



Simulated data



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