

Hidden Markov Models in Marine Sciences

model selection, covariates, random effects, hierarchical structures,
incorporating state labels

Vianey Leos Barajas & Marco Gallegos Herrada

Model Selection

In Search of the “Right” HMM

- Now that we have covered how to fit an HMM to data, we can construct various candidate models that vary in terms of the:
 - numbers of states
 - forms of the state-dependent distributions
- Picking the ‘best’ model amongst the candidate models is called **model selection**.
- We can also simultaneously do **model checking** (via pseudo-residuals).

Using Information Criterion to Choose

- There are two general information criteria that are used to compare models for HMMs in momentuHMM:
 - **AIC (Akaike information criterion)**: tends to pick the ‘best’ model in terms of its predictive capacity
 - **BIC (Bayesian information criterion)**: tends to pick the ‘best’ model in terms of its goodness-of-fit
- However, the models we choose amongst depends on those we construct.

Selecting the Number of States

- In building HMMs in practice, choosing the number of states is often of interest.
- However, in HMMs, **AIC/BIC tend to favour models with ‘too many’ states:**
 - assumptions such as the Markov property, the conditional independence assumption, but also the form of the state-dependent distributions are only approximately valid
 - HMM is often useful but not completely realistic (it lacks detailed mechanisms of the animal system)
 - additional states may be able to pick up these patterns

Practical Guidance to Selecting # of States

- AIC/BIC can be used as a guidance – but not a rule
- Start with $m=2$, and build slowly, identifying what each additional state contributes to the model
- Balancing model fit, complexity, and (importantly) interpretability
- **OR** there may also be unexplained individual variability across the states that can be accounted for factors like: **size, age of individuals**

Model Selection in HMMs

- Use AIC/BIC as guidance — or other information criteria as a guidance, not a rule.
- We choose the ‘best’ model amongst those we construct — we should be careful to keep this in mind. It’s unclear whether the ‘true’ data model is amongst these.
- What if our state decodings don’t change across different HMMs (with the same number of states)? Even when including covariates (we’ll see this in a bit) — state decodings may not change as they can be quite robust to model misspecification.

Covariates

Incorporating drivers of movement

- One of the key reasons HMMs are so popular is that they not only cluster observations into different states that can serve as proxies for behavior, BUT we can also directly incorporate covariates that may drive the observed behaviors!
- We can include covariates in the state process and/or the observation process:

covariates in the transition probability matrix

covariates in the state-dependent distributions

Covariates in the State Process

- Including covariates in the transition probability matrix, i.e. the mechanism that describes the evolution of the states over time, allows us to answer the following question(s):

“Do animals vary their behavior under different environmental conditions/time of day/conditions?”

- The transition probability matrix tells us how animals switch between states, but also how long they remain in a state. Including covariates such as time of day provides a nice summary of what behaviours are likely to be exhibited at different times of day.

Covariates in the State Process

- Recall that the t.p.m. is a matrix, Γ , with entries $0 < \gamma_{ij} < 1$ and $\sum_{i=1}^m \gamma_{ij} = 1$.
- For a 2-state HMM, using the constraints given above, we can write our t.p.m. as follows: $\Gamma(t) = \begin{pmatrix} 1 - \gamma_{12}(t) & \gamma_{12}(t) \\ \gamma_{21}(t) & 1 - \gamma_{21}(t) \end{pmatrix}$
- In order to incorporate covariates, we need to transform $\gamma_{12}(t)$ and $\gamma_{21}(t)$ so that they span the entire real line. (like the constrained to unconstrained transformation)

$$\text{logit}(\gamma_{12}(t)) = \beta_0^{12} + \sum_{k=1}^K \beta_k^{12} x_{t,k}$$

Covariates in the State Process

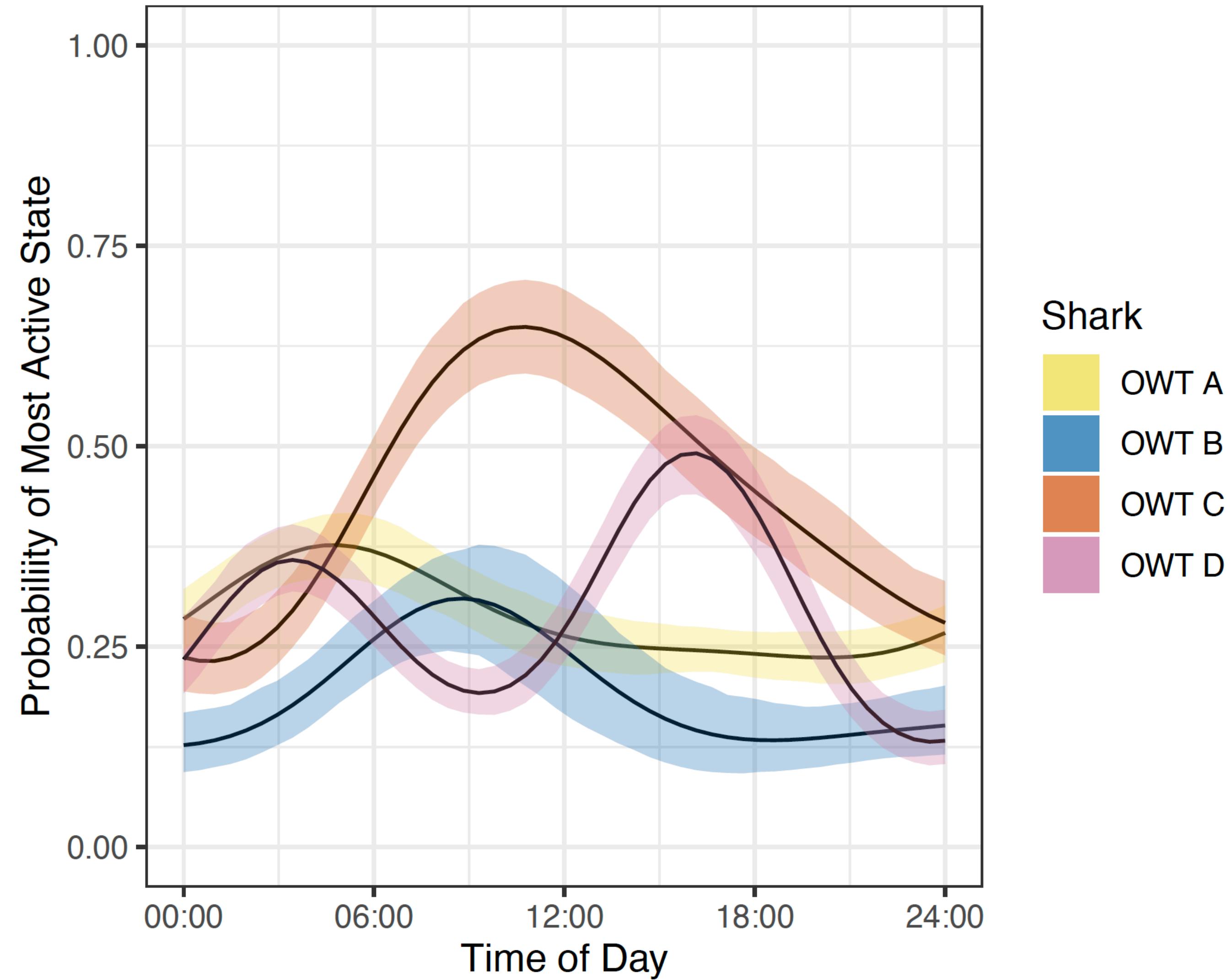
- Incorporating covariates is then just like a generalized linear model (with a few more components)!
- We can extend this to $m > 2$ states, this is what momentuHMM does for us! And as before, our goal is to estimate the values of the β 's, and from that we derive estimates of $\gamma(t)$'s.
- We can also assess whether values of β deviate from 0, i.e. if the covariates we've included are factors that affect the movement process.

Marginal Stationary Distributions

- If we recall, the interpretation for γ_{12} is the probability of switching to state 2 at time $t + 1$, given that we are in state 1 at time t .
- Incorporating covariates, our probabilities vary over time. However, we can also look at the so-called ‘marginal stationary distribution’ of the t.p.m. at time t , given covariates \mathbf{x}_t , which is the vector $\delta(\mathbf{x}_t) = \boldsymbol{\delta}(\mathbf{x}_t)\boldsymbol{\Gamma}(\mathbf{x}_t)$.
- The entries $\delta_i(\mathbf{x}_t) = \Pr(C_t = i)$ for $i = 1, \dots, m$, give us the marginal probabilities of the states over time, under the conditions observed at the time as given by \mathbf{x}_t .

Marginal Stationary Distributions

- Used overall dynamic body acceleration derived from accelerometers placed on four oceanic whitetip (OWT) sharks in the Bahamas.
- Constructed a 3-state HMM to identify 3 different activity levels.



Covariates in the state-dependent distributions

- One of the reasons to incorporate covariates in the transition probability matrix, is to keep the state-dependent distributions ‘fixed’ across all individuals in the model. In that way, all individuals share the exact same distribution for all states.
- However, there are reasons to allow for individual (or seasonal) specific state-dependent distributions. For example, if we care about measures like ‘step length’, i.e. distance travelled between consecutive time points, we need to account for variation in step lengths across individuals of different sizes, or that some behaviours, like traveling, may manifest differently in the summer vs the winter.

Covariates in the state-dependent distributions

- How we incorporate covariates often depends on the distributional form.
- Often, we care to vary the mean of the distribution by individual. In particular, we can have:

$$\mu_i = \tau_0 + \tau_1 * size_{indiv}$$

- Keep in mind, allowing for state-dependent distributions to vary across individuals can affect interpretation — i.e. are all state 1's across individuals comparable?

Random Effects

Accounting for Variability Across Individuals

- Suppose that there is variability across individuals not accounted for by covariates at hand.
- Random effects are one way of accounting for extra heterogeneity in the state process or the observation process.
- **Pooling:**
 - Complete pooling
 - Partial pooling
 - no pooling

Continuous-Valued Random Effects

- One option for the ‘partial pooling’ route is to include continuous-valued random effects. Often, this is done by assuming that some of the parameters differ by individual, but come from the same distributional form.

e.g. $\mu_k \sim N(\nu, \eta)$

- We can also do similarly in the t.p.m., so that each individual has their own specific baseline state-switching dynamics.

e.g. $\beta_0^{ij} \sim N(\nu_{\beta_0^{ij}}, \eta_{\beta_0^{ij}})$

- In a maximum likelihood framework, this can be computationally challenging to implement.

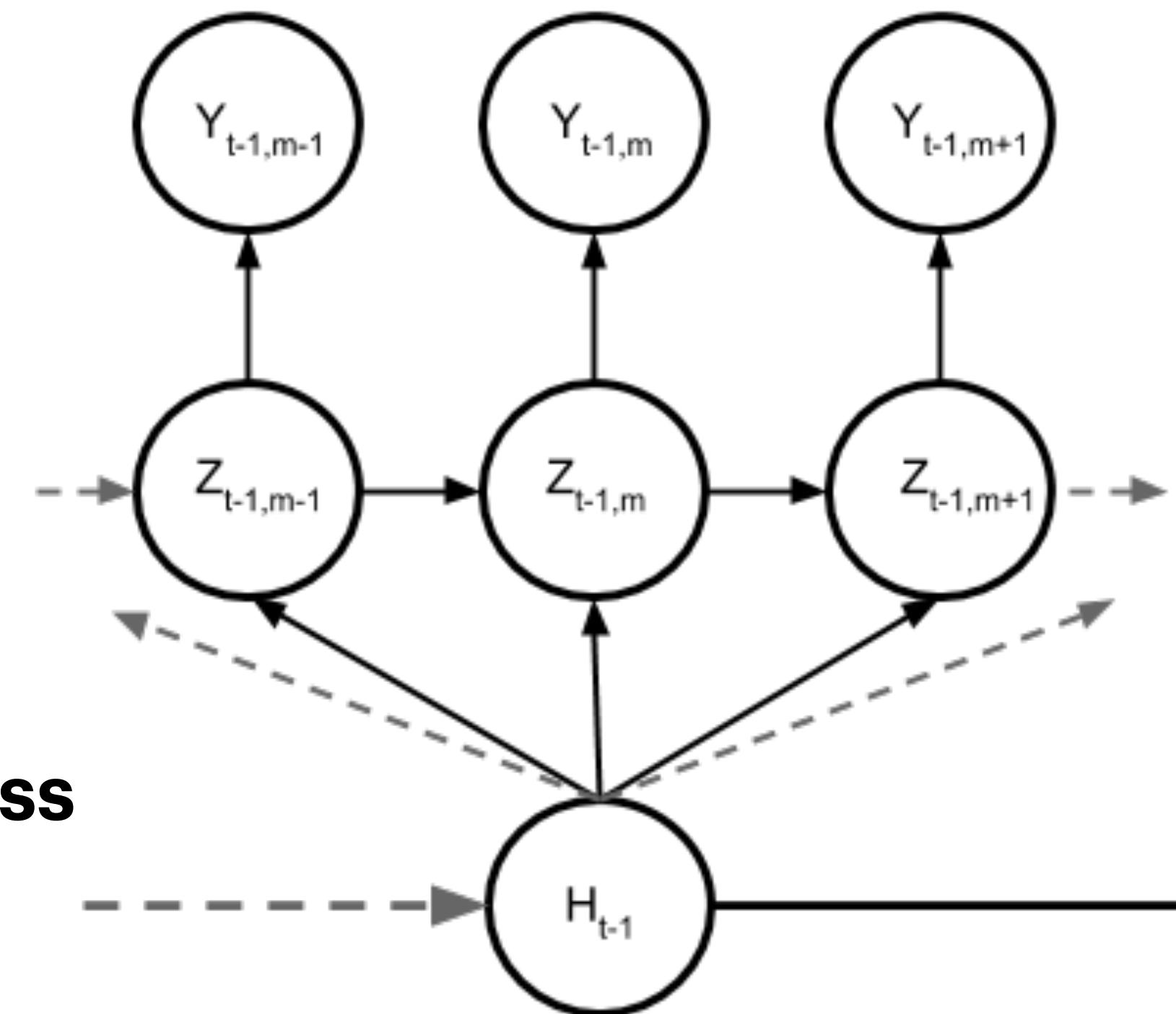
Discrete-Valued Random Effects

- Another approach is the discrete-valued random effects structure, typically assumed for the state-switching dynamics (the t.p.m.)
- It assumes that there G possible t.p.m.s that account for all of the differences exhibited across individuals. So, it can also help cluster individuals who have similar state-switching dynamics, while allowing differences for others.
- It is computationally easier to implement, but we introduce the challenge of estimating G .

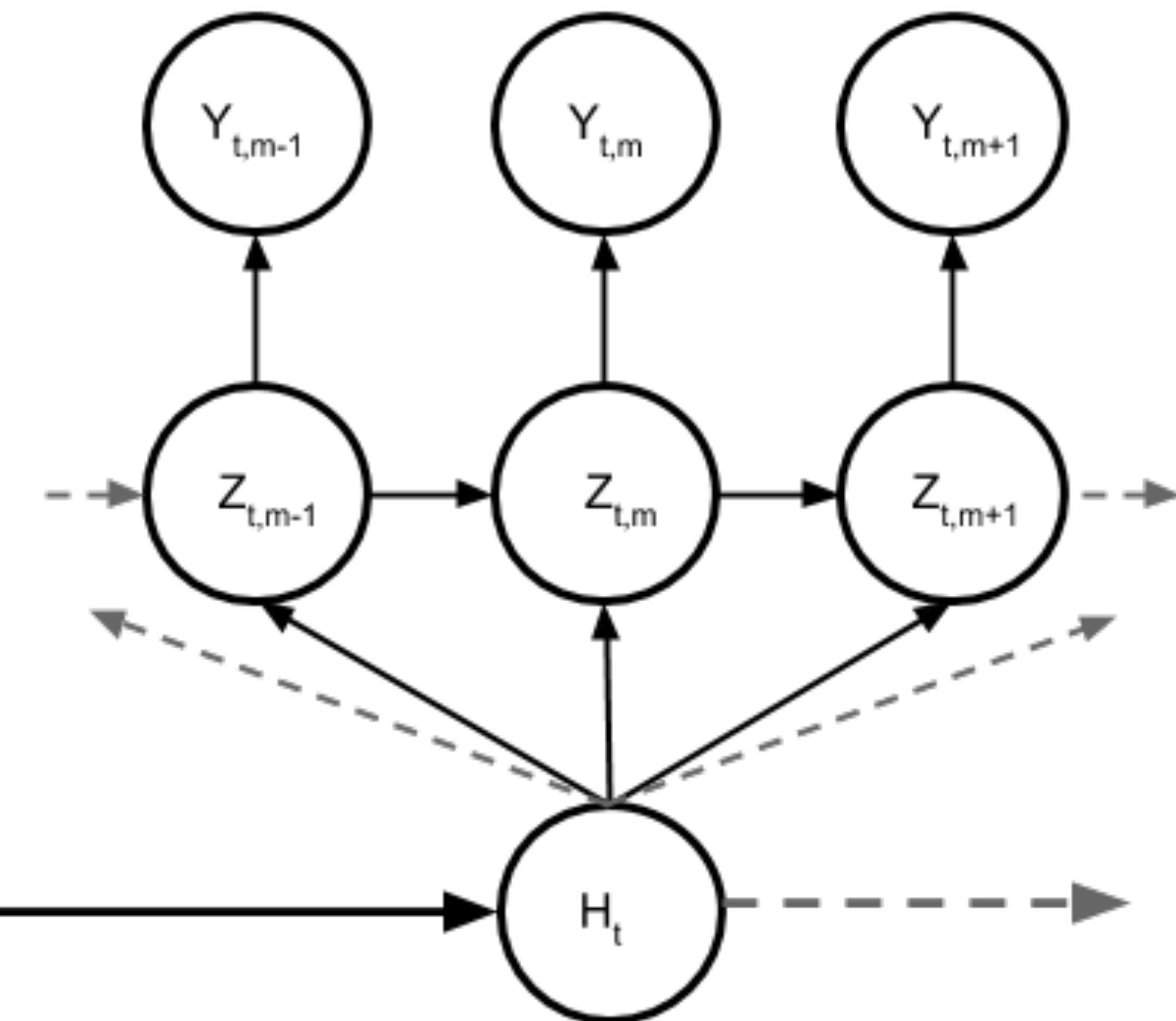
Hierarchical HMMs (i.e. multi-scale HMMs)

Multi-Scale HMMs (Hierarchical HMMs)

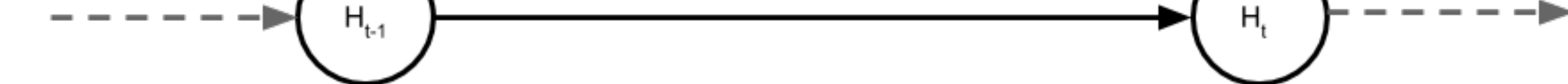
Observation process



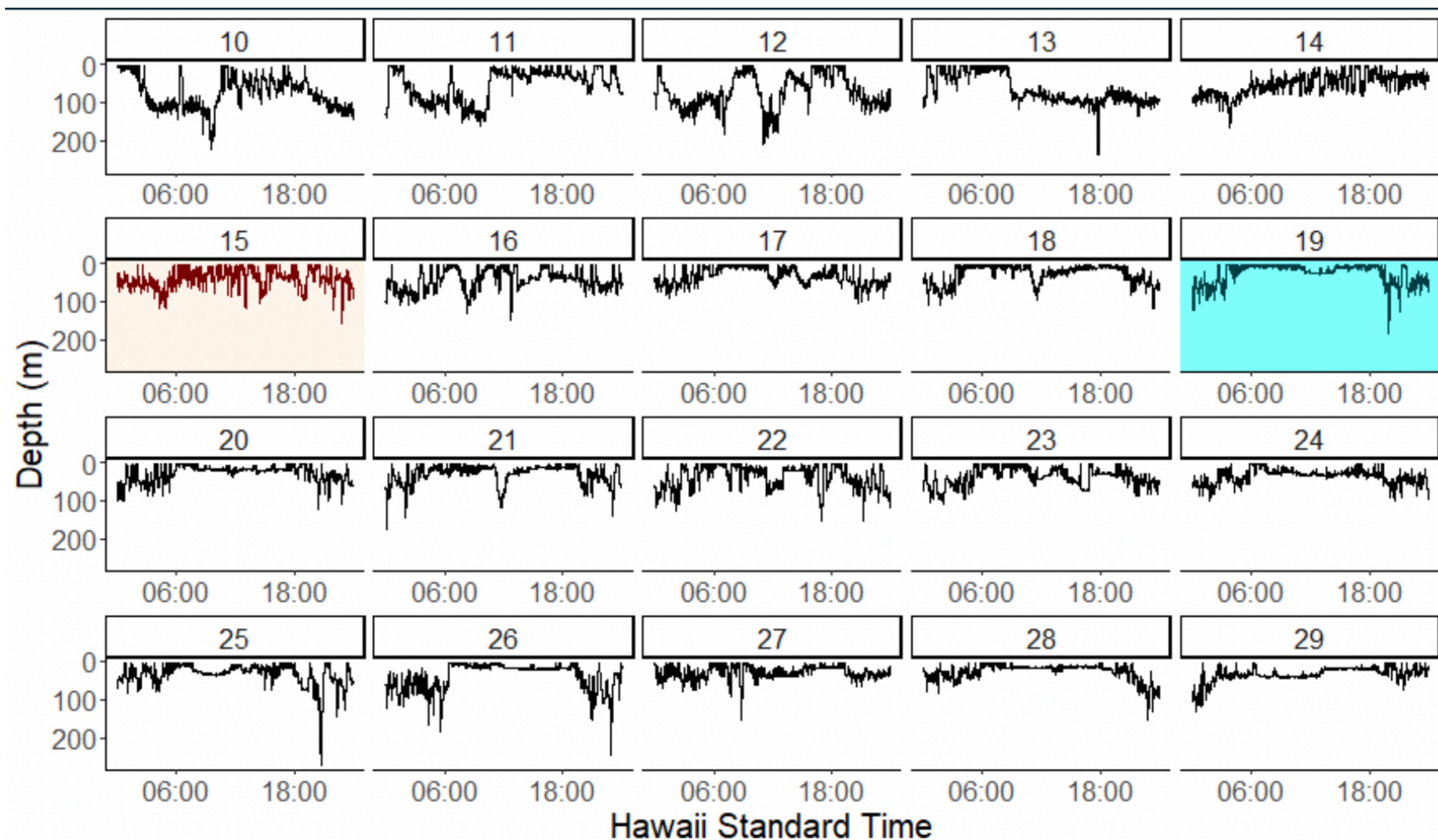
Fine-scale state process
(production state)



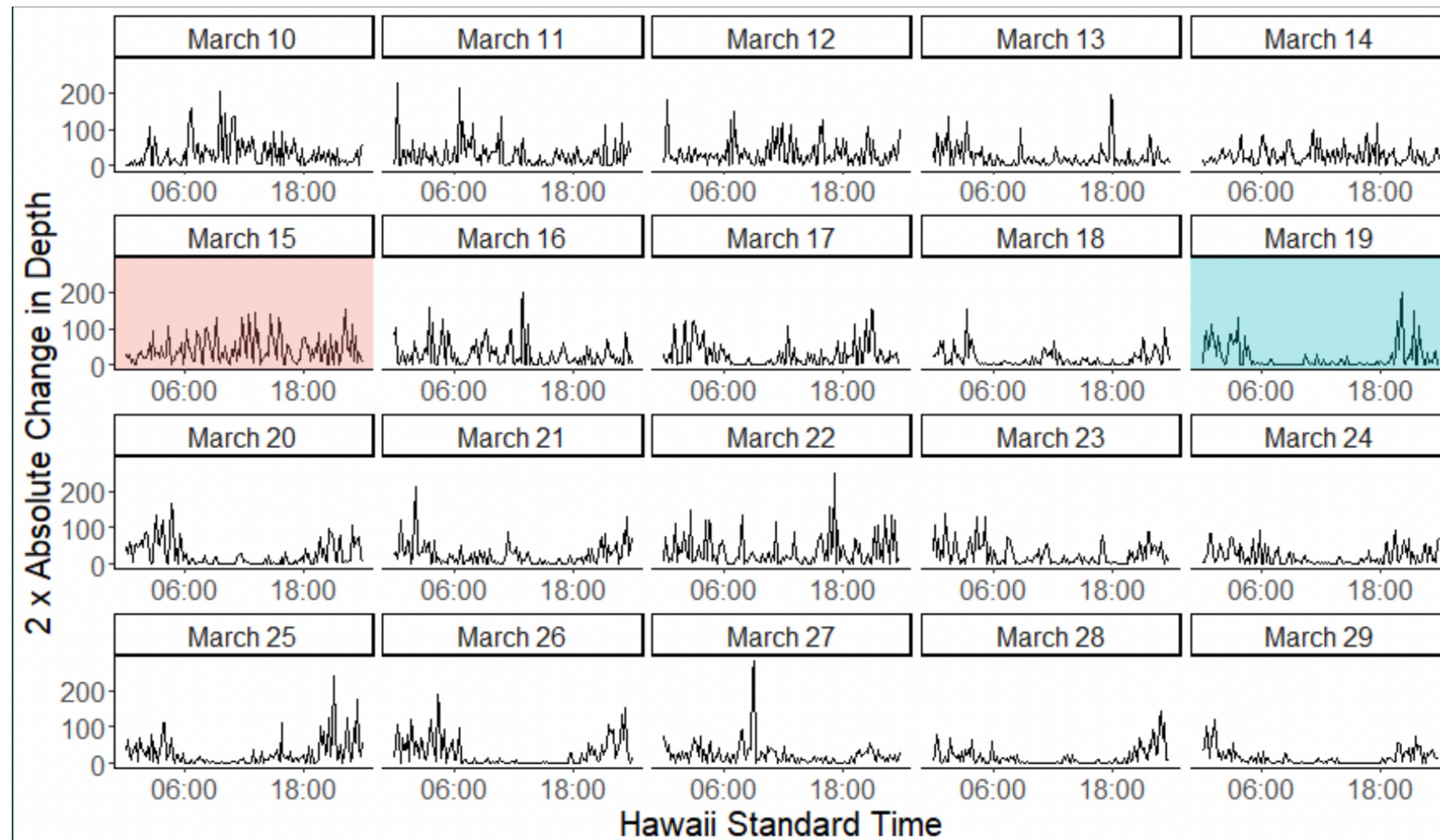
Coarse-scale state process
(internal state)



Tiger Shark Swimming Patterns

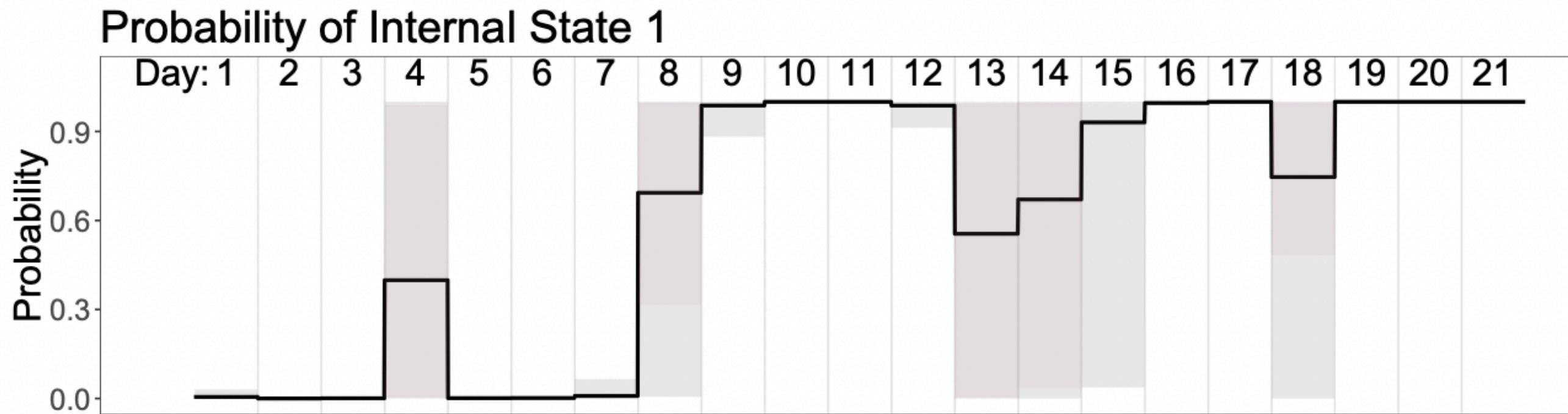


Tiger Shark Swimming Patterns

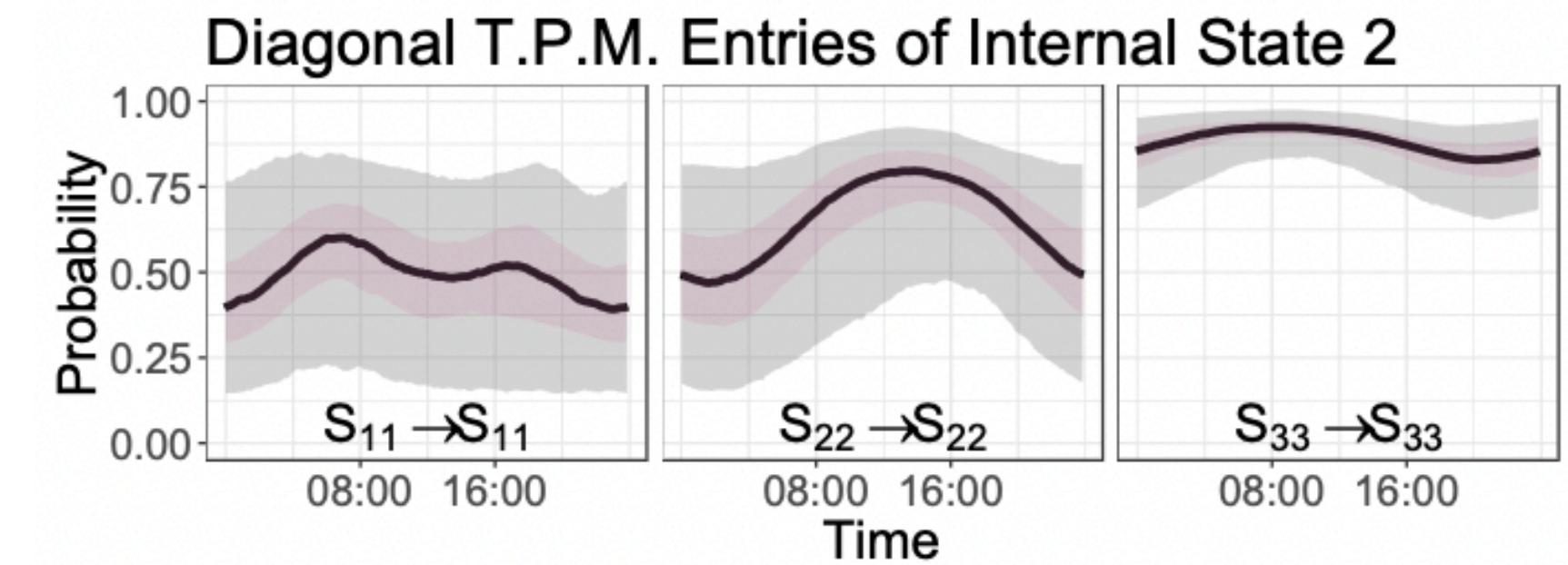
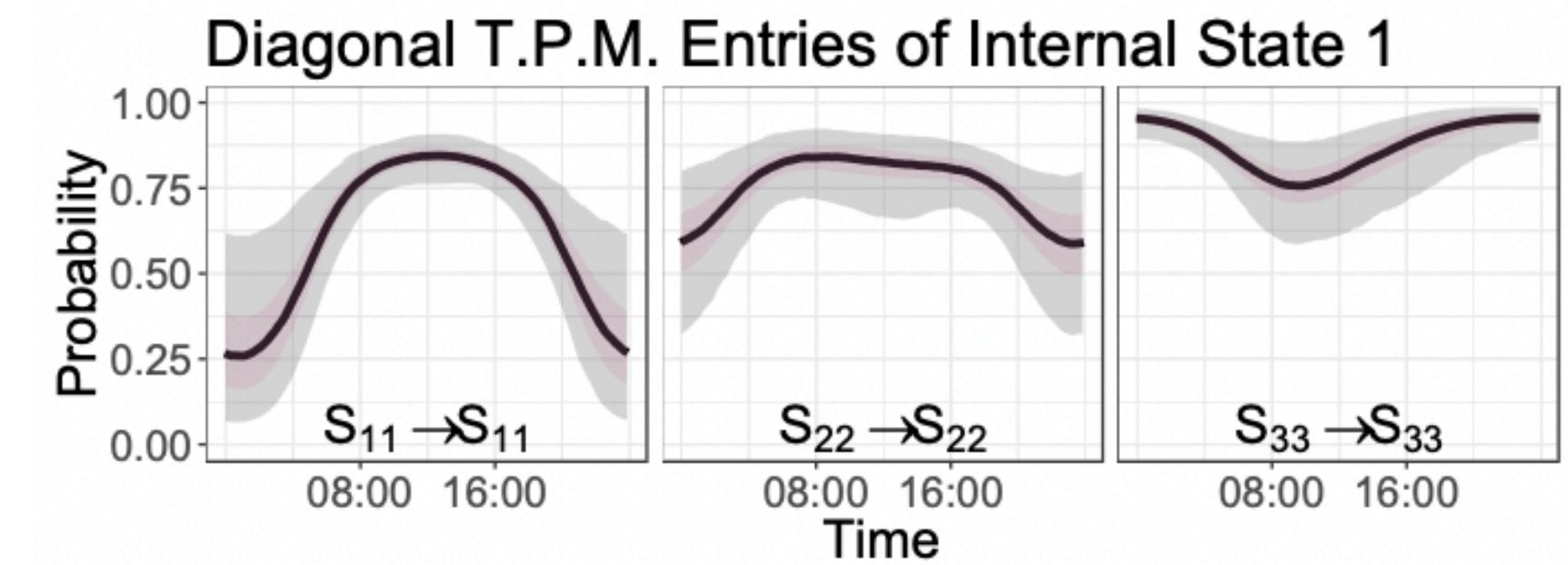
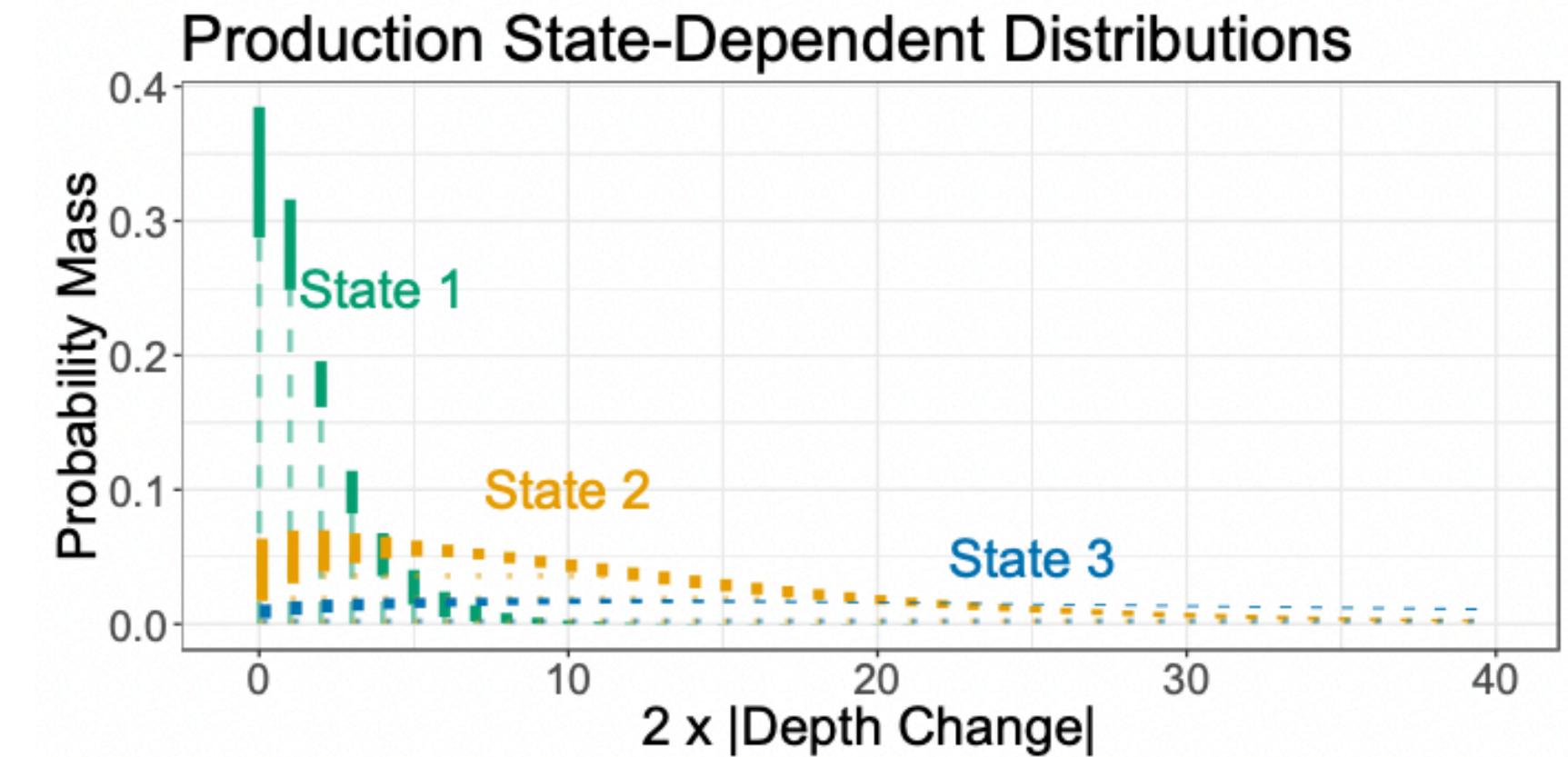


Some Results

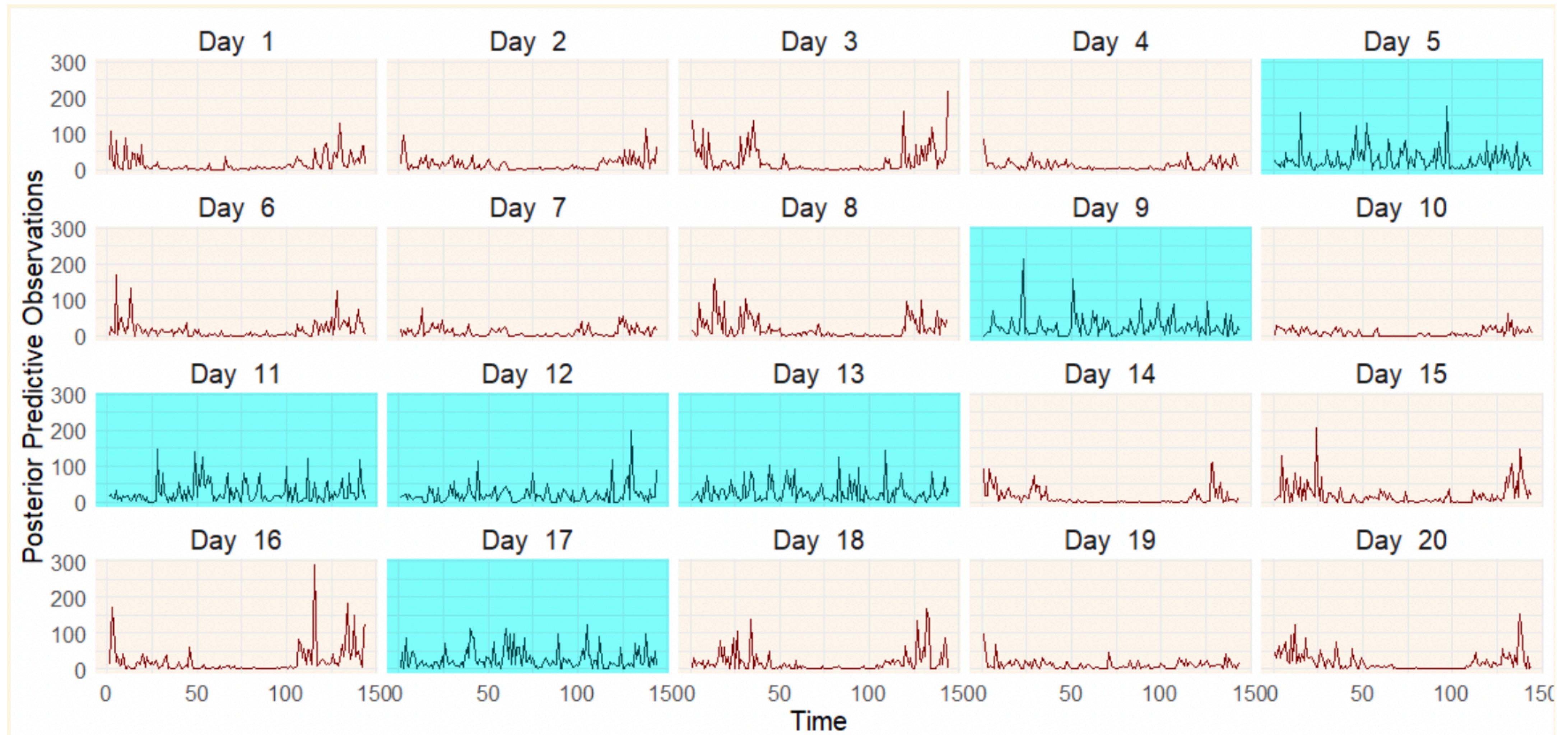
Coarse-scale results



Fine-Scale Patterns



Simulating from the Fitted Model



Multi-scale HMMs for time series collected at multiple temporal resolutions

Received: 18 November 2018

Accepted: 23 May 2019

DOI: 10.1111/2041-210X.13241

RESEARCH ARTICLE

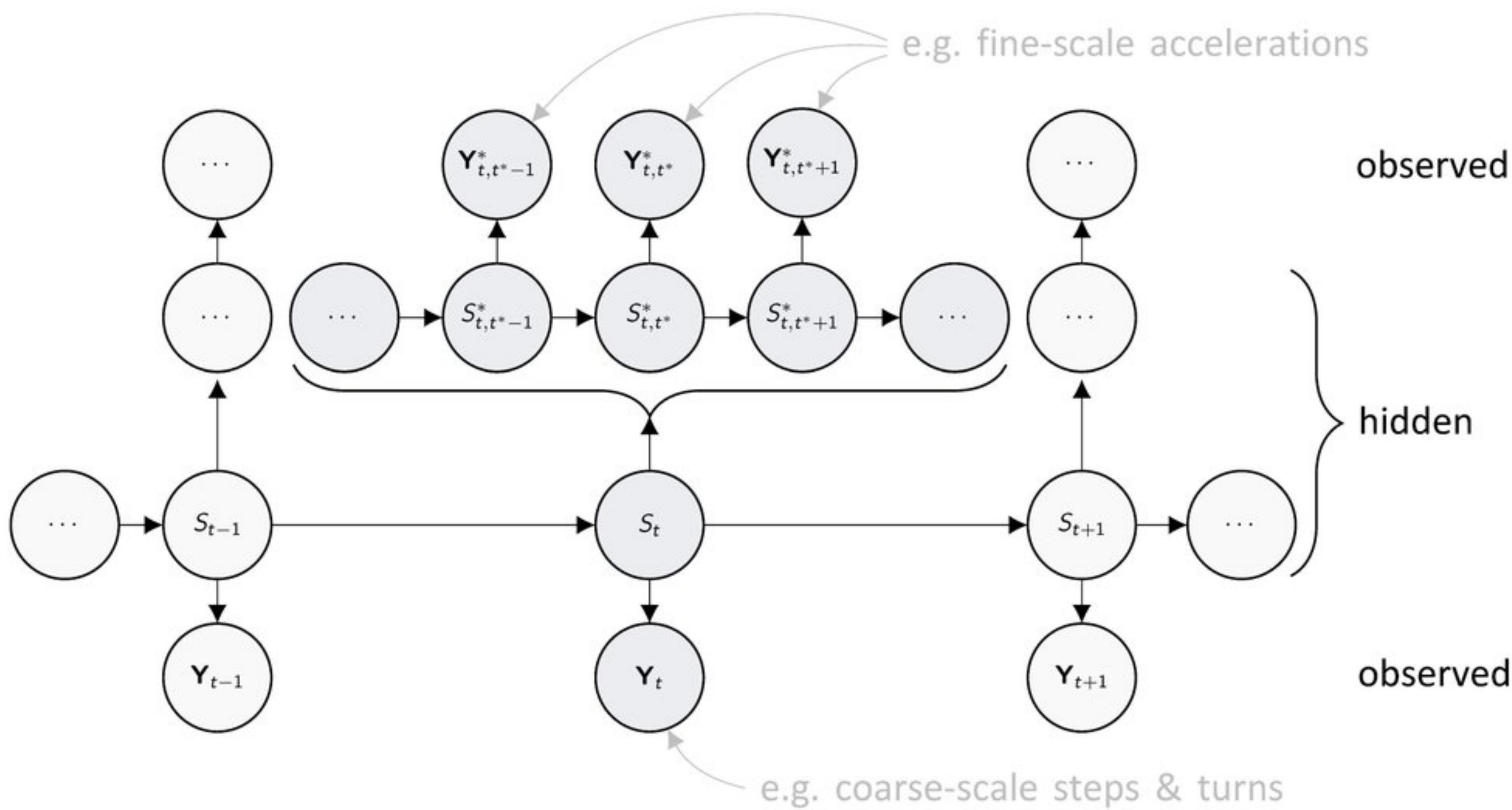
Methods in Ecology and Evolution



Joint modelling of multi-scale animal movement data using hierarchical hidden Markov models

Timo Adam¹ | Christopher A. Griffiths^{2,5} | Vianey Leos-Barajas^{1,3} |
Emily N. Meese⁴ | Christopher G. Lowe⁴ | Paul G. Blackwell² | David Righton⁵ |
Roland Langrock¹

Combining data streams at different temporal resolutions



Horn Sharks

Data collected by the shark lab at California State University, Long Beach

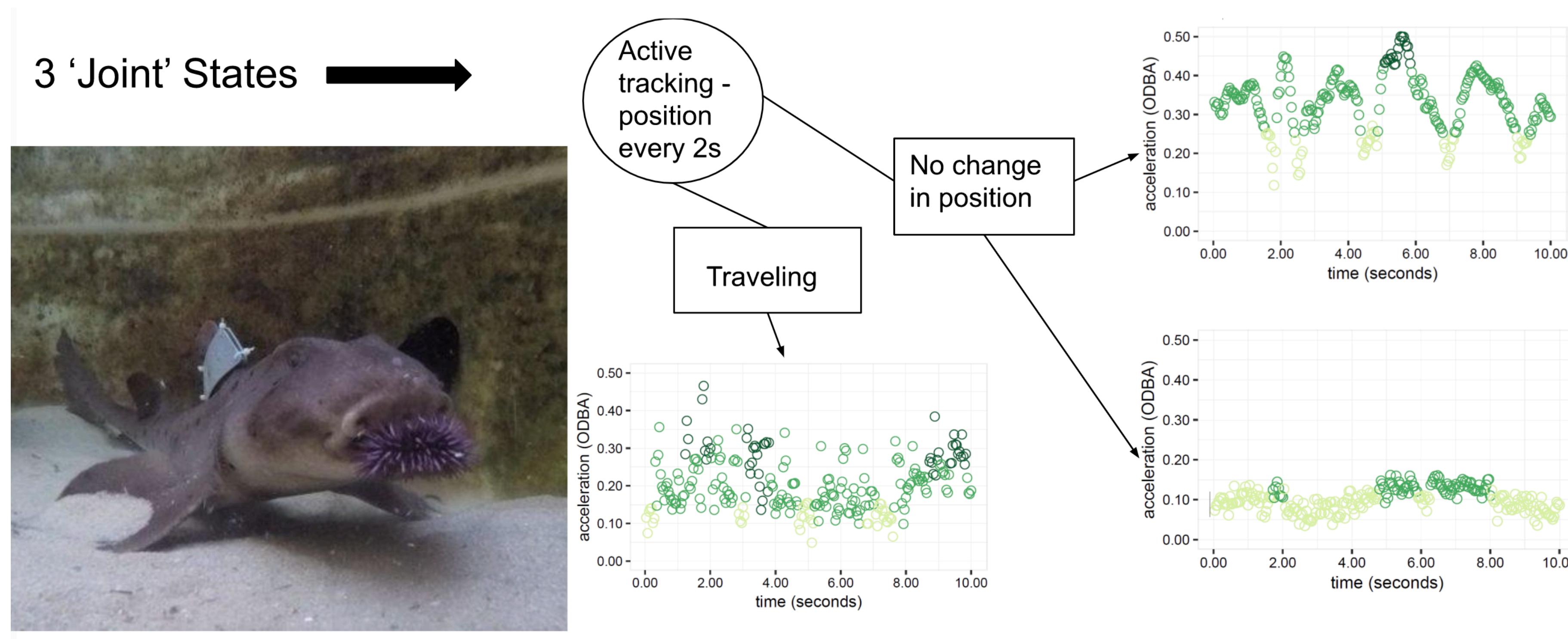


Photo and study by Emily Meese

The Three HMMs

What if there are labels available?

- In practice, there are three manners of applying an HMM to movement data:
 - In an **unsupervised learning** framework: no labeled data
 - In a **semi-supervised learning** framework: some labeled data
 - In a **supervised learning** framework: all data has a behavioral label
- momentuHMM can accommodate all three frameworks.

Model Checking Changes

- In a fully unsupervised learning framework, we can not assess if the states reflect meaningful behaviors in any capacity.

Keep in mind, a lot of people will emphasize the state decoding results — but this can only be validated with their domain expertise, it is not validated in any statistical way.
- In a semi-supervised or supervised framework, we can start to assess whether our states **do** actually capture the behaviours of interest.
- We can also assess model fit! Although often supervised frameworks (with labeled time series data) often focus on developing models for state prediction.

In Summary

HMMs are ***very*** useful for marine movement data

- HMMs are highly flexible frameworks for data collected over time, or in sequence, that can accommodate a lot of structures that are useful to learn about marine systems.
- Their formulation matches the intuition that what we observe is a manifestation of the animal's underlying behavioural state.
- With care, HMMs can reveal novel insights into the behavioral dynamics of marine animals.