

# Order, interpretation and architecture of state processes in hidden Markov movement models

Roland Langrock, Bielefeld University

This is my chance to talk about stuff that keeps bothering me!!

(but fear not, I'm not going to talk about the Lévy walk)

HMMs: motivating examples & introduction

Order selection (how many states?)

Interpretation of HMM states — the role of the temporal resolution

Complex state architectures for multi-scale inference

## Muskox movement

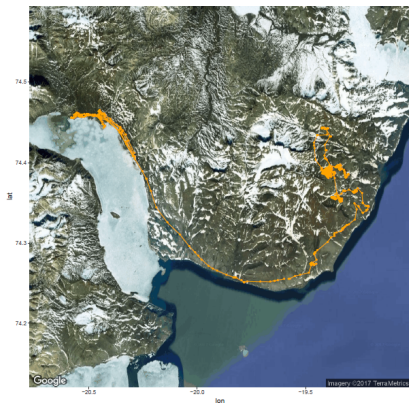


Figure: Movement track of a muskox in Greenland (hourly locations, one animal, 60 days).

## Muskox movement

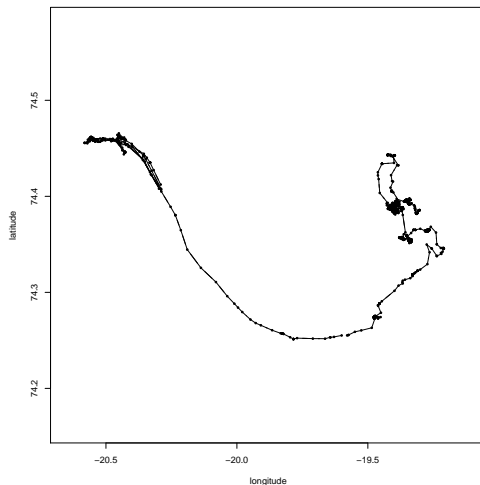


Figure: Movement track of a muskox in Greenland (hourly locations, one animal, 60 days).

## Muskox movement

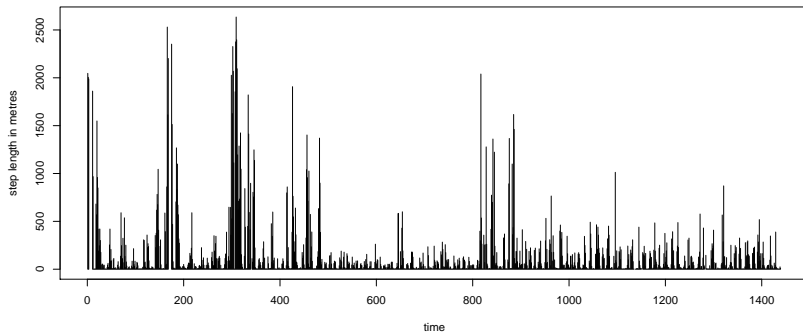
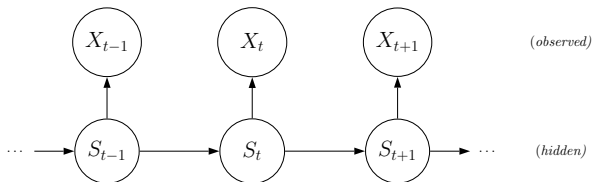


Figure: Time series of step lengths between successive locations.

## Hidden Markov models



- distribution of observations determined by underlying state
- hidden state process is an  $N$ -state Markov chain
- intuitive, but also a **model of convenience**, with simplifying assumptions

In animal movement:

- $X_t$ : step lengths at time  $t$
- $S_t$ : a **proxy** for behavioural state at time  $t$

## Building blocks of (discrete-time) HMMs

- $\{S_t\}_{t=1,2,\dots,T}$  is (usually) assumed to be an  $N$ -state **Markov chain**:

- state transition probabilities:  $\gamma_{ij} = \Pr(S_t = j \mid S_{t-1} = i)$
- transition probability matrix (t.p.m.):

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \dots & \gamma_{1N} \\ \vdots & \ddots & \vdots \\ \gamma_{N1} & \dots & \gamma_{NN} \end{pmatrix}$$

- initial state distribution:  $\delta = (\Pr(S_1 = 1), \dots, \Pr(S_1 = N))$

- **state-dependent distributions**  $f(x_t \mid s_t = j)$ :

- suitable family of distributions
- e.g. normal, Poisson, gamma, multivariate normal, von Mises, ...
- one set of parameters for each state



## Likelihood calculation using brute force

$$\begin{aligned}\mathcal{L}(\theta) &= f(x_1, \dots, x_T) \\ &= \sum_{s_1=1}^N \dots \sum_{s_T=1}^N f(x_1, \dots, x_T | s_1, \dots, s_T) f(s_1, \dots, s_T) \\ &= \sum_{s_1=1}^N \dots \sum_{s_T=1}^N \delta_{s_1} \prod_{t=1}^T f(x_t | s_t) \prod_{t=2}^T \gamma_{s_{t-1}, s_t}\end{aligned}$$

Simple form, but  $\mathcal{O}(TN^T) \rightsquigarrow$  numer. maximisation of *this* expression **infeasible**.

## Likelihood calculation via forward algorithm

Consider the so-called **forward variables**,

$$\alpha_t(j) = f(x_1, \dots, x_t, s_t = j)$$

These can be calculated using an **efficient recursive scheme**:

$$\alpha_1 = \delta \mathbf{Q}(x_1), \quad \alpha_t = \alpha_{t-1} \Gamma \mathbf{Q}(x_t)$$

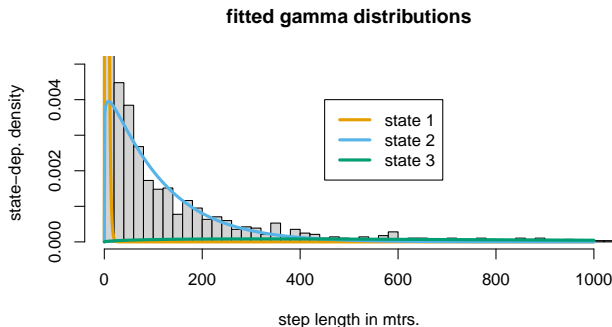
with  $\mathbf{Q}(x_t) = \text{diag}(f(x_t | s_t = 1), \dots, f(x_t | s_t = N))$

$$\rightsquigarrow \boxed{\mathcal{L}(\theta) = \sum_{j=1}^N \alpha_T(j) = \delta \mathbf{Q}(x_1) \Gamma \mathbf{Q}(x_2) \cdot \dots \cdot \Gamma \mathbf{Q}(x_T) \mathbf{1}}$$

Computational cost:  $\mathcal{O}(TN^2) \rightsquigarrow$  numerical max. is feasible even for very large  $T$ .

(this is implemented in both `moveHMM` and `momentuHMM`)

## Muskox example — fitted 3-state gamma HMM



$$\hat{\mu} = (4.27, 111.3, 849.5), \quad \hat{\sigma} = (2.83, 106.9, 689.6)$$

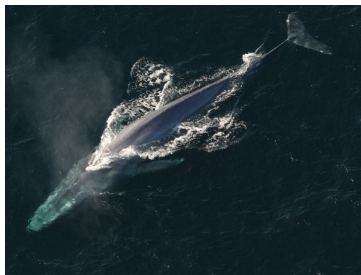
$$\hat{\Gamma} = \begin{pmatrix} 0.74 & 0.23 & 0.03 \\ 0.18 & 0.80 & 0.02 \\ 0.04 & 0.15 & 0.81 \end{pmatrix}, \quad \text{stat. dist. } \hat{\delta} = (0.38, 0.51, 0.11)$$

## A “bigger” example: blue whale response to sonar exposure

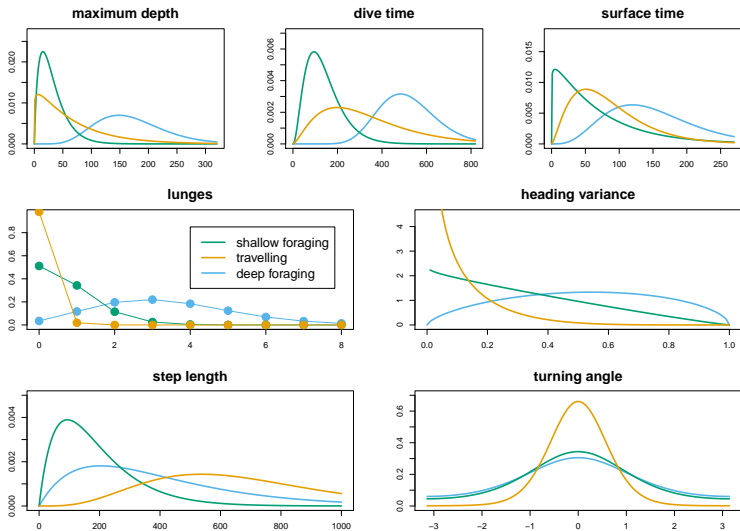
Variables observed on a **dive-by-dive** basis:

- maximum depth
- dive time
- post-dive surface duration
- number of feeding lunges
- variability of whale heading
- (horizontal) step length
- (horizontal) turning angle

- ~> 37 individual whales
- ~> 1054 dives in total
- ~> sonar exposure during 168 of the dives



## Blue whales — fitted 3-state model



## Blue whales — investigating response to sonar exposure

- substantial individual heterogeneity, captured using random effects
- transition prob. as function of sonar exposure (via row-wise multinom. logit)

Model without sonar covariate vs. model with sonar covariate:

$$\Delta AIC = 5.1 \text{ in favour of the model with sonar}$$
$$\text{p-value LR test} = 0.009$$

According to the model, when exposed to sonar, whales ...

**... are less likely to initiate “deep foraging”**

Order selection (how many states?)

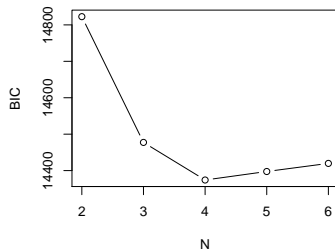
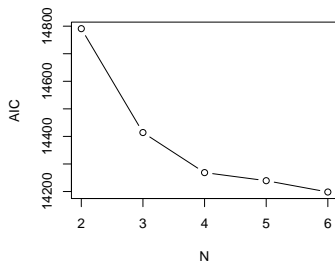
## Selecting the number of states

The **choice of the number of states**,  $N$ , is usually of primary interest.  
(unless of course  $N$  is known or simply fixed *a priori*)

In a maximum likelihood framework, this should be easy, shouldn't it?



## Selecting $N$ in the muskox example



- AIC selects the 6-state model
- BIC selects the 4-state model
- data owner thinks there should be 3 states ...
- when analysing the full data set, AIC & BIC select models with  $> 7$  states!

## Some remarks on model selection

Criteria like the AIC or the BIC are very appealing as they are:

- theoretically founded ...
- ... and at the same time very simple to use

As a consequence, they are often applied without any **critical reasoning**.

However, both AIC and BIC

- are only asymptotically exact solutions to their optim. problems
- make the very strong and **usually unrealistic assumption** that the true model is among the set of candidate models

What does this mean for HMMs?

## States mopping up structure

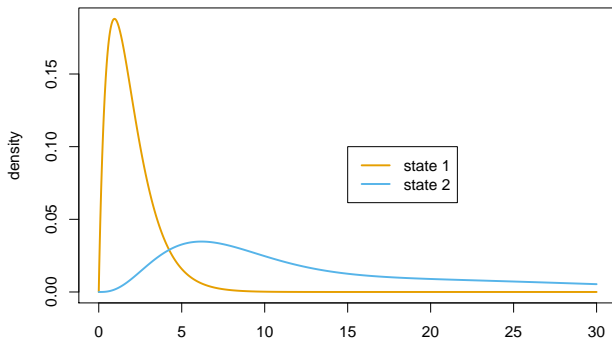
For sth. as complex as animal movement, models **will always be simplistic**.  
(and likewise in panel data analyses, multivariate medical time series, etc.)

For HMMs, both AIC and BIC **tend to favour models with “too many” states**:

- assumptions such as the Markov property, the conditional ind. assump. but also the form of the state-dep. distributions, are only approximately valid
- minor assumption violations accumulate to an overall substantial lack of fit
- thus, an HMM is often a **model of convenience**, but not actually realistic
- additional states may be able to **“mop up”** these otherwise ignored patterns

Resulting models may fit the data better, but are often not interpretable.

## States mopping up structure — illustrating example with simul. data



Observations were generated using the two state-dependent distributions displayed above and t.p.m.

$$\Gamma = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$$

If we fit gamma HMMs with  $N = 2, 3$ , then what  $N$  will be selected?

## States mopping up structure — illustrating example with simul. data

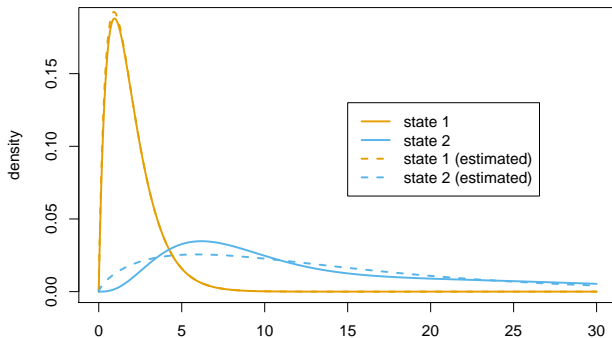


Figure: True state-dependent distributions and those of the fitted 2-state gamma HMM.

↪ lack of fit of the distribution in state 2

## States mopping up structure — illustrating example with simul. data

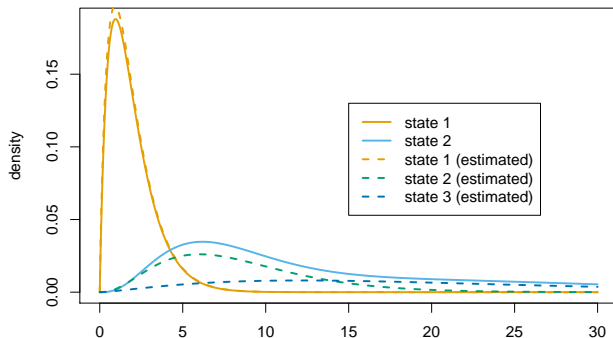


Figure: True state-dependent distributions and those of the fitted 3-state gamma HMM.

↪ this model is selected by both AIC and BIC

## Other causes of “too many” states

Other structure which, if neglected, often leads to additional states:

- outliers
- temporal variation
- individual heterogeneity
- violation of first-order Markov property
- violation of conditional independence assumption

How to deal with this?

- ~> ideally, improve model formulation to account for such structure
- ~> often not feasible for complex data!
- ~> we may need to accept lack of fit and **be pragmatic about choosing  $N$**

## Pragmatic selection of the number of states

- decide *a priori* on the the minimum and the maximum number of states
- closely inspect each of the fitted models
- use model checking methods to obtain a more detailed picture
- consider model selection criteria for guidance
- make **pragmatic choice** of  $N$
- if still in doubt, report results for different  $N$



## Pragmatic selection of $N$ in the muskox example (full data set)

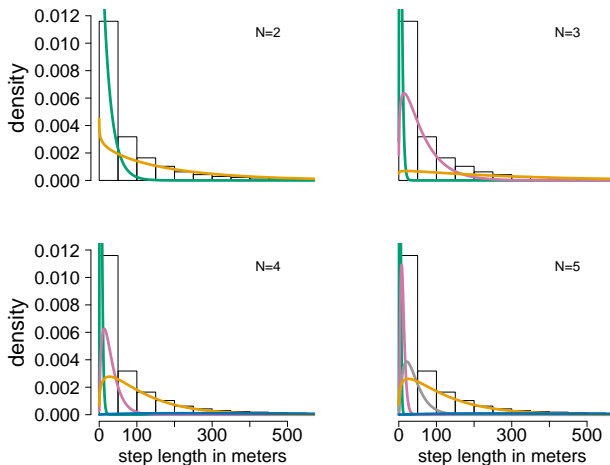


Figure: HMMs with 2–5 states fitted to the muskox data.

## Pragmatic selection of $N$ in the muskox example (full data set)

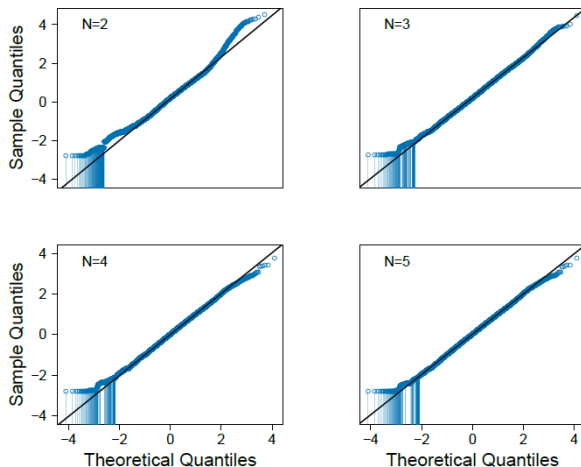


Figure: QQ-plots of pseudo-residuals obtained for models with 2–5 states.

## Pragmatic selection of $N$ in the muskox example (full data set)

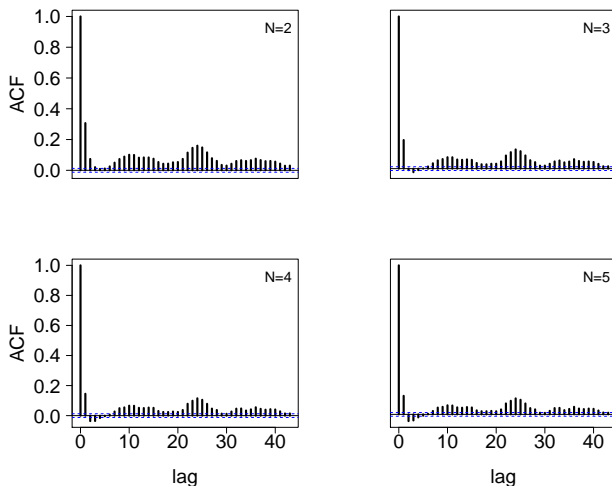


Figure: Sample ACFs of pseudo-residuals obtained for models with 2–5 states.

## Order selection — summary

- it is hard to formulate HMMs that capture all features of a complex data set
- when missing out on structure, then additional states can compensate this
- when focus lies on **forecasting**, then that's fine
- however, we are usually interested in **inference on state process**, which is hard if models are not interpretable anymore
- in such cases, a healthy dose of pragmatism is required

The choice of  $N$  may also depend on the temporal resolution of the observations!

## Interpretation of HMM states — the role of the temporal resolution

## Interpretation of HMM states

A trivial yet underappreciated HMM fact:

**HMM states do not map 1-to-1 to biologically meaningful behaviours**

Why?

1. features of HMM states are fully data-driven ( $\rightsquigarrow$  unsupervised learning)
2. temporal resolution determines what behaviours may be inferred at all

## An example of a coarse resolution

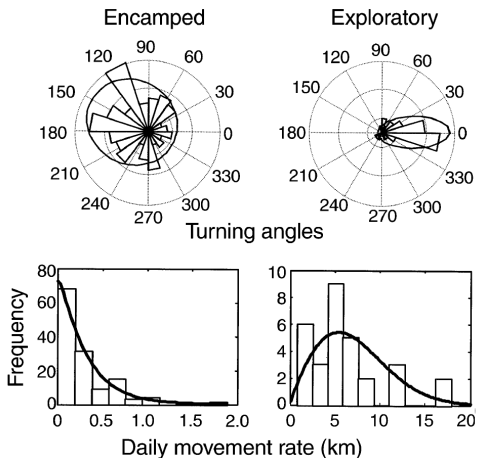


Figure: 2-state Weibull/wrapped Cauchy HMM fitted to elk data, with **one location observed per day** — example taken from Morales *et al.*, 2004.

## Morales' elk data re-analysed, not using informative priors

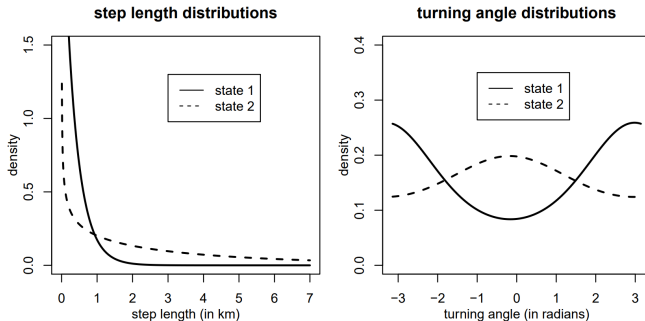


Figure: 2-state gamma/von Mises HMM fitted to the elk data from Morales *et al.*.



## A second example of a relatively coarse resolution

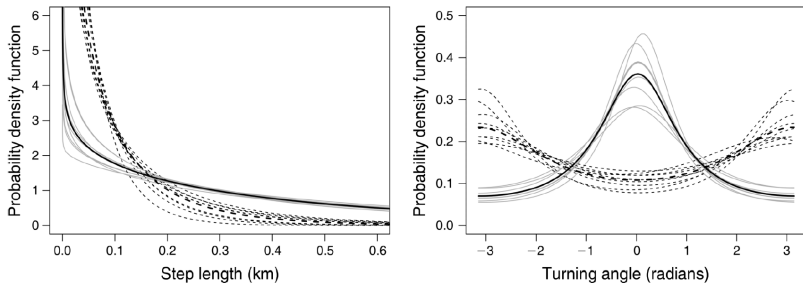


Figure: 2-state gamma/von Mises HMM fitted to bison data, with **one location observed every 3 hours** — example taken from Langrock *et al.*, 2012.

What to do if resolution is coarse?

**Don't expect HMMs or any other model to identify behaviours operating on a scale much finer than the data resolution.**

## What if resolution is much *finer* than necessary?

Nowadays, we're often faced with the other extreme:

- incredibly fine-scale data, e.g. collected at 16 Hz using accelerometers:

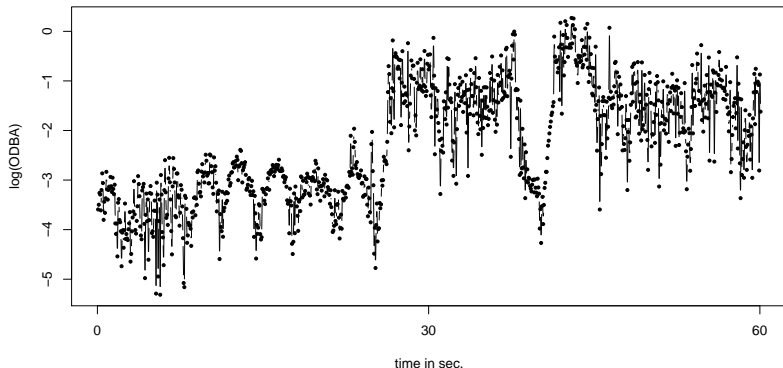


Figure: log (ODBA) values observed for an oceanic whitetip shark in the first minute of an 8-day observation period (~~~ what's displayed here is only 0.00009% of the data set!).

## What if resolution is much *finer* than necessary?

- biologically, such data are fantastic in terms of the information contained
- but they bring new statistical challenges:
  - ~> heavy computational cost
  - ~> much more complex dependence structures
  - ~> sometimes artifacts (e.g. zero-inflated turning angles)
- should we dare to **downsample**??

## 3-state HMM fitted to data downsampled by factor 48

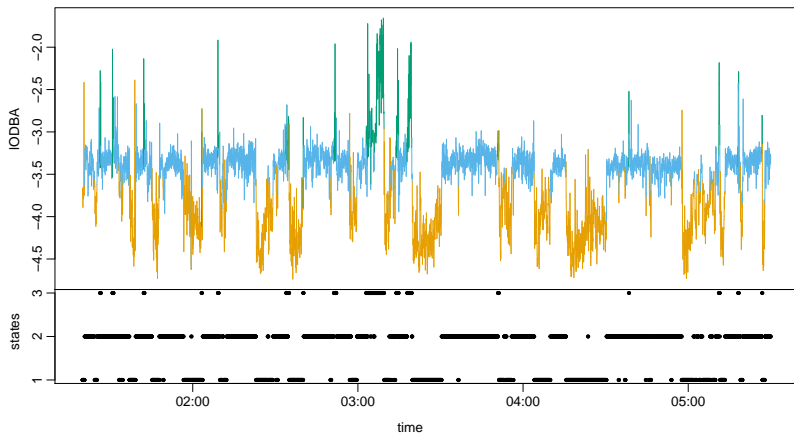


Figure: log (ODBA) values **averaged over 3-second intervals** and associated decoded states (displayed for a subset of the data).

## Complex state architectures for multi-scale inference

(work in progress!)

## Multi-scale inference: is this going to be a thing?

A few years ago, movement ecology was mostly about

- horizontal tracking data

Nowadays, we often deal with data sets combining

- horizontal tracking data
- accelerometer data
- information on vertical movements (depths or altitudes)
- ...

Consequences:

- more information  $\rightsquigarrow$  a more complete picture (yay!)
- mismatches between resolutions of the different data types:
  - opens up opportunities for multi-scale inference
  - biological relevance: different behaviours manifest themselves at different scales!
  - however, statistically, this is quite a challenge...

## Example: horizontal and vertical movement of Atlantic cod

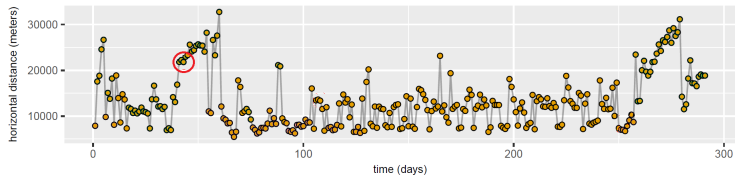


Figure: **Daily horizontal step lengths** observed for an Atlantic cod.

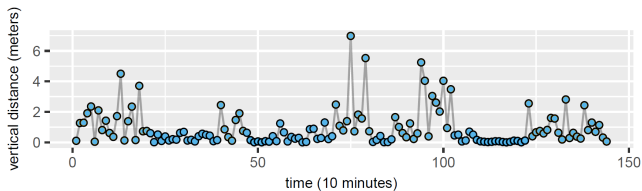
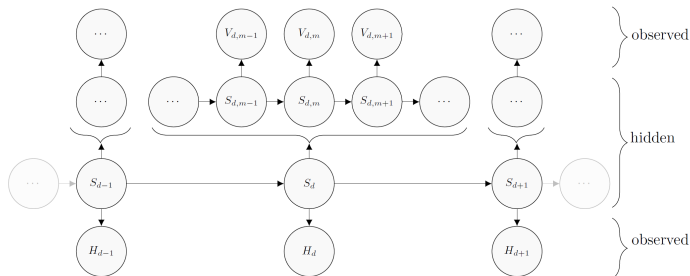


Figure: **Vertical distances** covered within **10-min. intervals** on day 43.

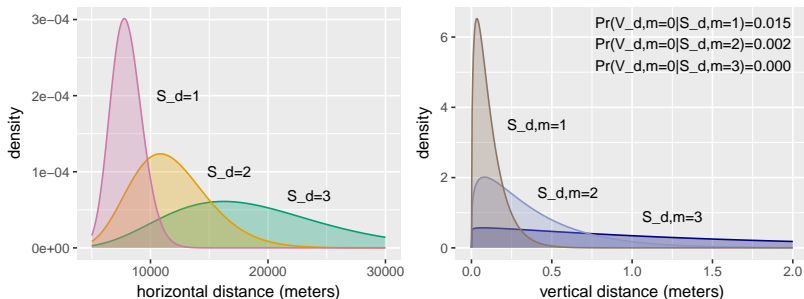


## Hierarchical HMM for multi-scale inference



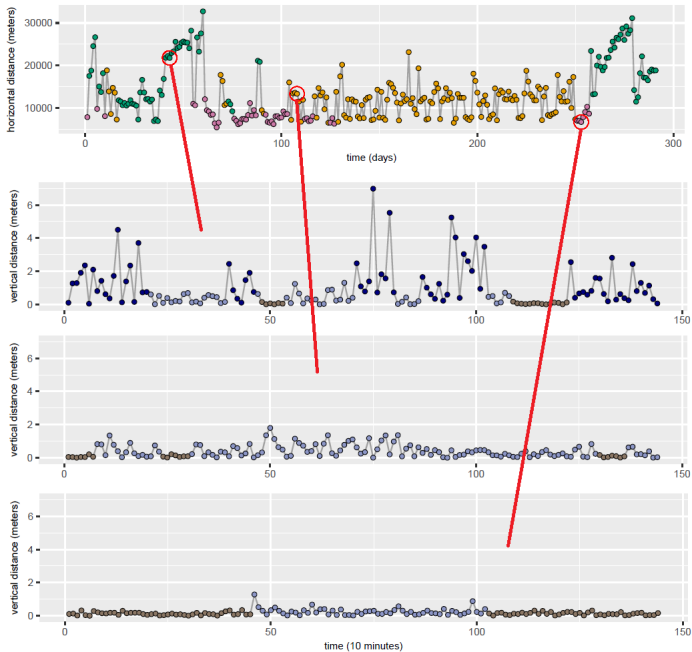
- crude-level HMM ( $S_d, H_d$ ), here operating on a daily scale
- fine-scale HMM ( $S_{d,m}, V_{d,m}$ ), here operating on 10-minute intervals
- for each state of  $S_d$ , there is a separate HMM ( $S_{d,m}, V_{d,m}$ )
- idea: if in migratory mode, then diving will be different than when foraging

## Hierarchical $3/3$ -state HMM fitted to Atlantic cod data



- states 1/2/3 of  $S_d$  roughly correspond to resident/foraging/migrating
- stationary distributions of the (within-day) state process  $S_{d,m}$ :

$$\delta_{S_d,m} = \begin{cases} (0.76, 0.22, 0.02) & \text{if } S_d = 1 \text{ (resident)} \\ (0.36, 0.62, 0.03) & \text{if } S_d = 2 \text{ (foraging)} \\ (0.25, 0.40, 0.35) & \text{if } S_d = 3 \text{ (migrating)} \end{cases}$$



## Remarks on hierarchical HMMs

Things we can potentially do using hierarchical HMMs:

- overcome limitations of HMMs, specifically w.r.t. the temporal resolution
- obtain a much more complete picture of animal behaviour
- find out at which scales behavioural decisions are made

Main problems:

- there are A LOT of modelling decisions to be made
- (formal) model selection is a can of worms
- numerical instability

## Concluding remarks

## Take-home messages

- HMMs are versatile and powerful tools for analysing behavioural time series
- while conceptually easy to understand, there tend to be many practical challenges, which require some caution & experience on the side of the user
- (a challenge that I didn't mention today: measurement error...)
- new types of data bring new challenges  $\rightsquigarrow$  the field will continue to grow
- to make use of these data, ecologists and statisticians ought to collaborate



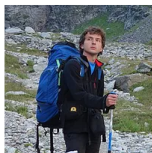
## Acknowledgements



Stacy DeRuiter



Vianey Leos-Barajas



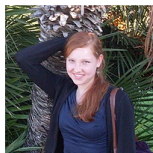
Théo Michelot



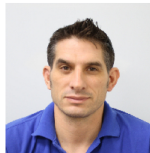
Toby Patterson



Timo Adam



Jennifer Pohle



Yannis Papastamatiou



Juan Morales

... and many others!

# References

## **HMMs fitted to blue whale data:**

DeRuiter et al. (2017). "A multivariate mixed hidden Markov model for blue whale behaviour [...]", *AoAS*

## **Order selection for HMMs:**

Pohle et al. (2017). "Selecting the number of states in hidden Markov models [...]", *JABES*

## **Hierarchical HMMs:**

Leos-Barajas et al. (2017). "Multi-scale modeling of animal movement and general behavior [...]", *JABES*

## **Analysing accelerometer data using HMMs:**

Leos-Barajas et al. (2017). "Analysis of animal accelerometer data using hidden Markov models", *MEE*

## **Accessible introduction to HMMs:**

Zucchini et al. (2016). *Hidden Markov Models for Time Series: An Introduction Using R*, Chapman & Hall

## **How HMMs compare to other models for animal movement:**

Patterson et al. (2017). "Statistical modelling of individual animal movement: an overview of [...]", *AStA*

## **R package for fitting standard animal movement HMMs:**

Michelot et al. (2016). "moveHMM: An R package for the statistical modelling of animal [...]", *MEE*