Ecological applications of hidden Markov models and related doubly stochastic processes

ROLAND LANGROCK

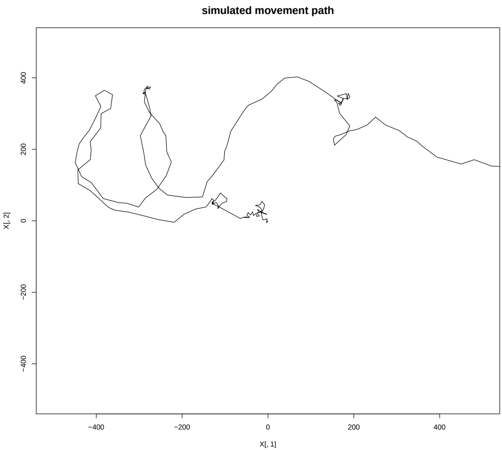
University of St Andrews

Motivating example

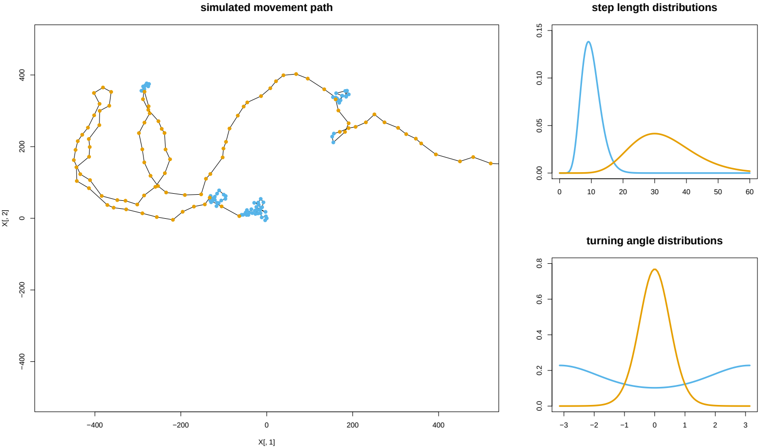
Hidden Markov models for animal movement

Figure: *Dux magnus gentis venteris saginati* (Wild Haggis).

Hidden Markov models for animal movement



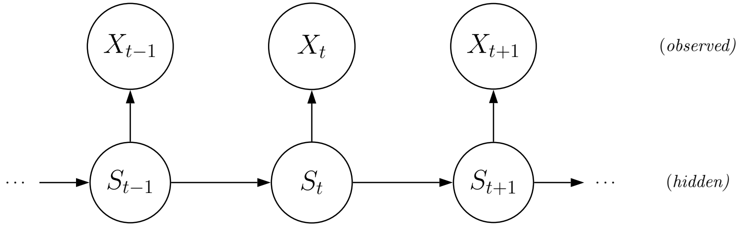
Hidden Markov models for animal movement



$$\Pr(S_t = j \mid S_{t-1} = j) = 0.95 \quad \text{for } j = 1, 2,$$

where S_t: state at time t

HMMs — summary/definition



- two (discrete-time) stochastic processes, one of them hidden
- distribution of observations determined by underlying state
- hidden state process is an N-state Markov chain

HMM machinery

HMMs – likelihood calculation using brute force

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

Simple form, but N^T summands, numerical maximization of *this* expression thus infeasible.

HMMs – likelihood calculation via forward algorithm

Consider instead the so-called **forward probabilities**,

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

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

$$\boldsymbol{\alpha}_1 = \boldsymbol{\delta}\mathbf{P}(x_1)$$
$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t \boldsymbol{\Gamma}\mathbf{P}(x_{t+1})$$

with $\mathbf{P}(x_t) = \text{diag}\big(f(x_t|s_t = 1), \ldots, f(x_t|s_t = N)\big)$ and t.p.m. $\boldsymbol{\Gamma}$.

$$\Rightarrow \quad \boxed{\mathcal{L}^{\text{HMM}} = \sum_{j=1}^{N} \alpha_T(j) = \boldsymbol{\delta}\mathbf{P}(x_1)\boldsymbol{\Gamma}\mathbf{P}(x_2) \cdot \ldots \cdot \boldsymbol{\Gamma}\mathbf{P}(x_T)\mathbf{1}}$$

Computational effort **linear** in T!

HMMs – likelihood calculation via forward algorithm

Consider instead the so-called **forward probabilities**,

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

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

$$\begin{aligned} \boldsymbol{\alpha}_1 &= \boldsymbol{\delta} \mathbf{P}(x_1) \\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{\alpha}_t \boldsymbol{\Gamma} \mathbf{P}(x_{t+1}) \end{aligned}$$

with $\mathbf{P}(x_t) = \text{diag}\big(f(x_t|s_t = 1), \ldots, f(x_t|s_t = N)\big)$ and t.p.m. $\boldsymbol{\Gamma}$.

$$\Rightarrow \quad \boxed{\mathcal{L}^{\text{HMM}} = \sum_{j=1}^{N} \alpha_T(j) = \boldsymbol{\delta} \mathbf{P}(x_1) \boldsymbol{\Gamma} \mathbf{P}(x_2) \cdot \ldots \cdot \boldsymbol{\Gamma} \mathbf{P}(x_T) \mathbf{1}}$$

Computational effort **linear** in T!

Related model classes

- **state-space models** $\longrightarrow$ can be approximated arbitrarily accurately by HMMs by finely discretizing the state space
- **Markov-modulated Poisson processes** $\longrightarrow$ can be regarded as HMMs (with slightly modified dependence structure)
- **Markov-switching regression models** $\longrightarrow$ these are HMMs (with slightly modified dependence structure)

Ecological applications, Part I: (Horizontal) animal movement

Animal movement modelling: woodpecker example

- A standard (hidden Markov) model for movement:
  - N "behavioural" states, switching governed by Markov chain
  - e.g., von Mises and gamma state-dependent distributions for turning angles and step lengths, respectively

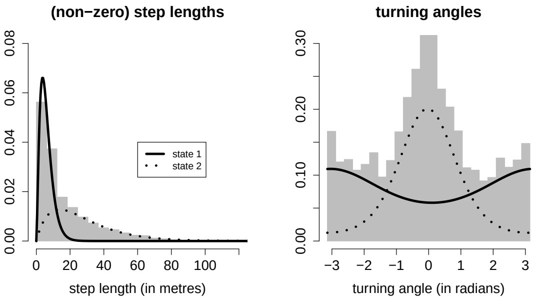


Figure: Turning angle and step length distributions for **woodpeckers** in two behavioural states (30 obs. per hour).

Animal movement modelling: some challenges

- the **sampling frequency** is crucial
  - should be driven by biological question of interest...
  - ...but more often than not this isn't the case

- moreover, if sampling is irregular, HMM-type time series models are usually not readily applicable
  - models such as **state-switching SDEs** are being developed...
  - ...but are hardly used by ecologists

- in general, most movement models are VERY simplistic relative to the actual process

Ecological applications, Part II:
General animal behaviour

General animal behaviour: blue whale example

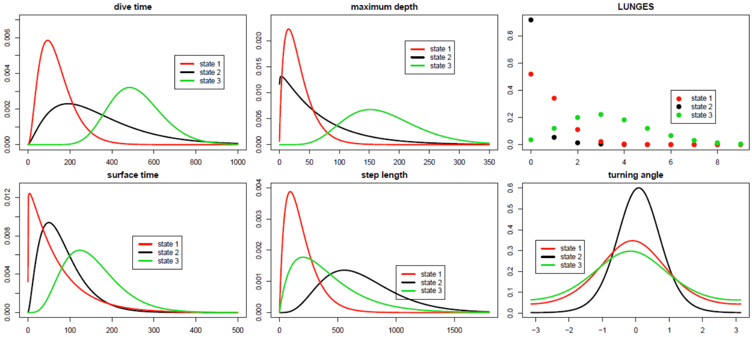
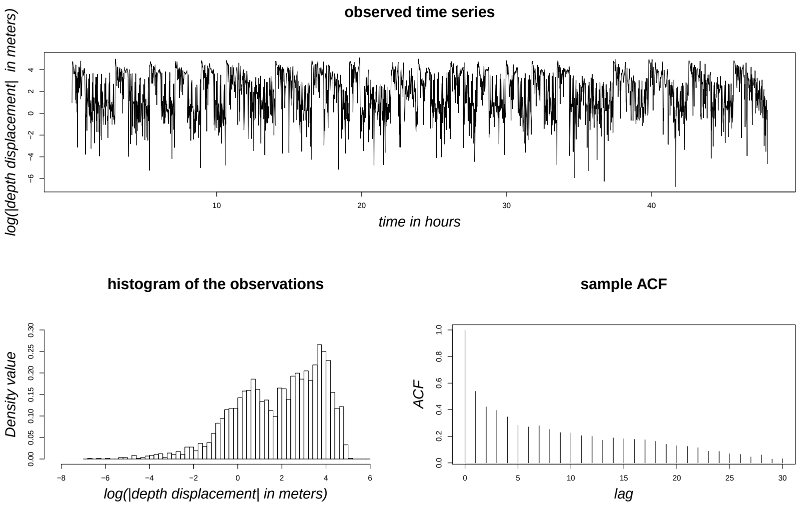


Figure: State-dependent distributions of an HMM fitted to 37 multivariate **dive-by-dive data on blue whale behaviour**.

- large individual heterogeneity ($\rightarrow$ discrete random effects)
- under exposure to sonar, whales were less likely to occupy the "deep foraging dive" and the "travelling" state (states 2 & 3)

Ecological applications, Part III: Modelling vertical speeds of a diving whale

Blainville's beaked whale dive data

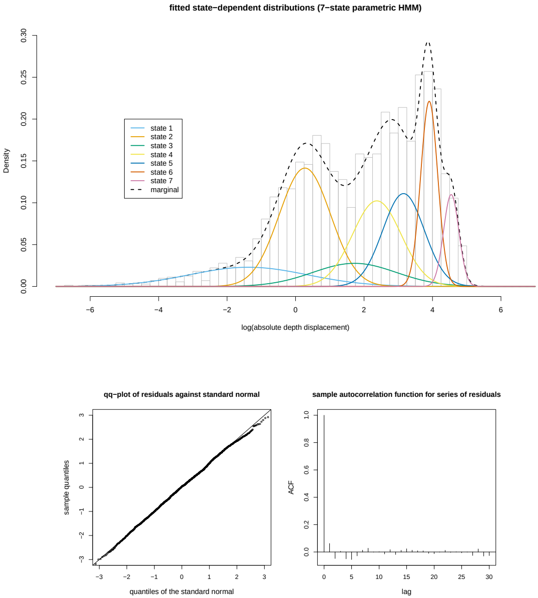


Blainville's beaked whale – parametric HMMs

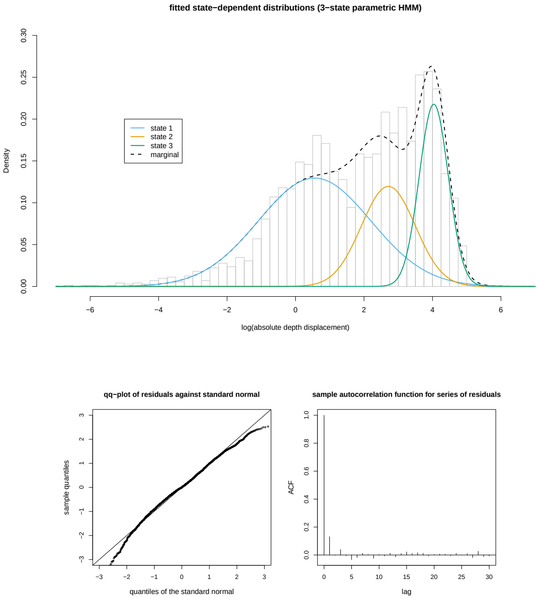
Table: Results of fitting **HMMs with normal state-dependent distributions** to the beaked whale data.

#states	p	AIC	BIC
3	12	9784.00	9855.59
4	20	9498.16	9617.47
5	30	9400.30	9579.27
6	42	9294.88	9545.43
7	56	9208.04	**9542.11**
8	72	9129.15	9558.67
9	90	9090.98	9627.87
10	110	**9064.53**	9720.74

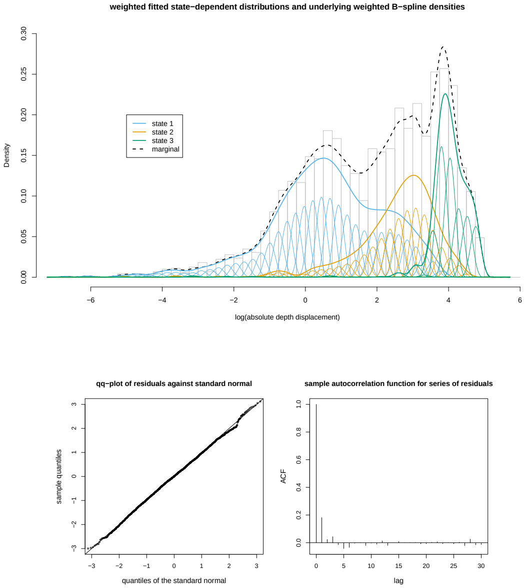
Blainville's beaked whale – parametric HMM, $N = 7$



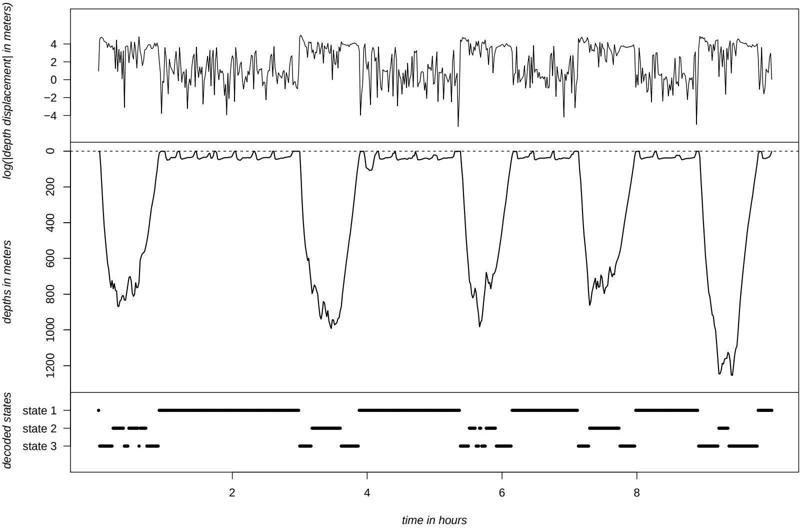
Blainville's beaked whale – parametric HMM, $N = 3$



Blainville's beaked whale – nonparametric HMM with $N = 3$



Blainville's beaked whale – Viterbi for nonparametric HMM with $N = 3$



Ecological applications, Part IV: capture-recapture

Capture-recapture: basic model formulation

- capture-recapture: mark animals at a site of interest, visit site at regular time intervals and record re-sightings of animals
- aim: usually either abundance or survival rate estimation
- a capture-recapture **encounter history** such as

  1 1 0 1 0 0 0
  (0: not seen; 1: seen alive)

  can be regarded as the **outcome of an HMM**, with
  - the states corresponding to the animal's survival state, so that

    $$\mathbf{\Gamma} = \begin{pmatrix} \phi & 1-\phi \\ 0 & 1 \end{pmatrix},$$

  - Bernoulli state-dep. distribution for state 1
    (the probability of success being the recapture probability)

Capture-recapture: extensions

- capture-recapture-**recovery** data
- multi-state (Arnason-Schwarz) models
- semi-Markov state processes

- example of a more complex capture-recapture model:
  - **encounter history**:

    1 1 0 0 1 0 0 0

  - associated **survival history**:

    A A A A A ? ? ?

    individual time-varying covariate (e.g. body mass):

    16.7 17.8 ? ? 15.3 ? ? ?

$\longrightarrow$ HMM with bivariate state process (after discretization of covariate space)

Capture-recapture — example Soay sheep

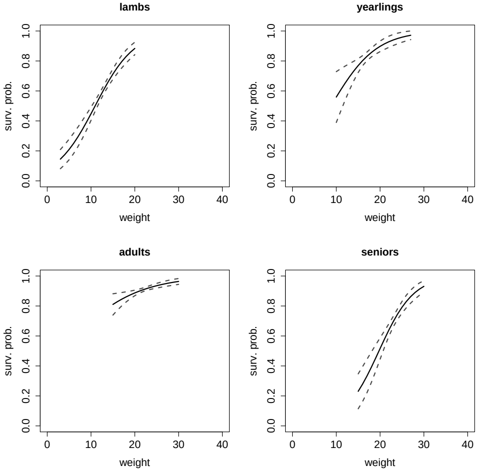


Figure: Estimated yearly survival probabilities of Soay sheep as a function of body mass (and 95% pointwise CIs indicated by dashed lines).

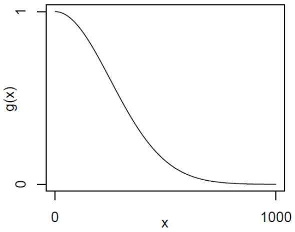
Ecological applications, Part V: estimating abundance of marine mammals

Conventional distance sampling

1.) record sightings of animals within predetermined strips
2.) quantify number of animals missed within strips
3.) scale corresponding estimate up to the area of interest

Details/assumptions involved in 2.):

- model detection probability $g(x)$ as a function of perpendicular distance x to trackline



- crucial assumption: $g(0) = 1$ (certain detection on the line)
- $g(0) \neq 1 \Rightarrow$ abundance estimates (negatively) biased
- for marine mammals typically $g(0) \neq 1$

Incorporate availability process in distance sampling

Two main reasons for missing whales in surveys:

1.) animal is underwater

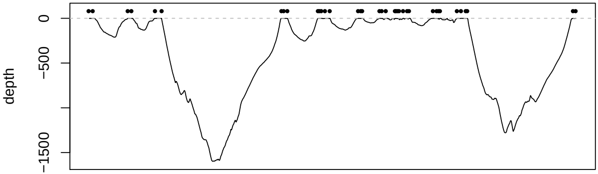
2.) animal is at surface but not detected

Aim: formulate models for 1.) & 2.) and put these pieces together

Strategy:

- quantify availability via stochastic process
- model conditional detection probability, given the animal is available, as a function of distance from animal to observer

MMPP for surfacing/availability events



- surfacings are continuous-time events that tend to occur in clusters (due to behavioural switches)
- **Markov-modulated Poisson process** (MMPP):
  - continuous-time Markov chain $\lambda(I)$
  - determined by state transition rate matrix $\mathbf{Q} = \left(\mu_{ij}\right)$ and rate of signals (surfacings) in the different states, $\lambda_1, \ldots, \lambda_N$

Embedding an MMPP in a distance sampling survey

- MMPP can be used to describe surfacings
- but not all surfacings are associated with a detection...
- assume prob. of detections of surfacings to depend on x and y
- e.g., hazard probability:

$$h(x, y) = \mu \exp\left(-\frac{x^\gamma + y^\gamma}{\sigma^\gamma}\right)$$

- corresponds to a thinning of the MMPP
- detections are then realizations of a nonhomogeneous MMPP, with rates $\lambda_i h(x, y)$, $i = 1, \ldots, N$

→ given a fitted nonhomogeneous MMPP, standard techniques can be used to obtain abundance estimates

Minke whale application

- Data:
  - seven sequences of minke whale surfacing events
  - 870 sightings of whales in the Northeastern Atlantic, made during shipboard surveys
- Analysis:
  - MMPPs were fitted to the series of surfacings
  - fixing the MMPP parameters, we estimated the detection probability (given an availability event):

$$h(y,x) = \frac{\theta_1 \theta_2^{\theta_3}}{(\theta_2^2 + x^2 + y^2)^{\theta_3/2}} \, ,$$

$$\text{logit}(\theta_2) = \beta_0 + \beta_1 \, I\big(B \in \{0,1\}\big) + \beta_2 \, I\big(B \in \{3,4\}\big) \, ,$$

with B giving the sea state in Beaufort (categorical variable)

Table: Estimated probability of detecting an animal that is on the trackline, $g(0)$, for the three different covariate levels (and 95% confidence intervals obtained through bootstrapping).

	estimate	95% *C.I.*
$g(0)$ for $B \in \{0, 1\}$	0.75	*(0.49,0.80)*
$g(0)$ for $B = 2$	0.68	*(0.42,0.73)*
$g(0)$ for $B \in \{3, 4\}$	0.61	*(0.35,0.66)*

Concluding remarks

Concluding remarks

- Statistical ecology is fun!
  - new tracking technology, camera traps etc. have led to an explosion in the amount of data being collected by ecologists
  - **huge demand for statistical expertise!**
- HMM-type models are becoming increasingly popular in the ecological community
  - **versatile general-purpose models** for ecological (and other) time series data
  - key features: flexibility, mathematical simplicity & computational tractability

References

- DeRuiter, S., Langrock, R., Skirbutas, T., Goldbogen, J., Chalambokidis, J., Friedlaender, A.S., Southall, B.L. (2015). Analyzing the effect of sonar exposure on the behavioural state-switching dynamics of blue whales. *Preprint*
- King, R., Langrock, R., Semi-Markov Arnason-Schwarz models. *arXiv:1505.04827*
- McKellar, A., Langrock, R., Walters, J., Kesler, D. (2015). Using mixed hidden Markov models to examine behavioural states in a cooperatively breeding bird. *Behavioural Ecology*
- Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., Morales, J. (2012). Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*
- Langrock, R., King, R., Maximum likelihood estimation of mark-recapture-recovery models in the presence of continuous covariates. *Annals of Applied Statistics*
- Langrock, R., Kneib, T., Sohn, A., DeRuiter, S. (2015). Nonparametric inference in hidden Markov models using P-splines. *Biometrics*
- Langrock, R., Borchers, D., Skaug, H. (2013). Markov-modulated nonhomogeneous Poisson processes for modeling detections in distance sampling surveys. *Journal of the American Statistical Association*
- Pradel, R. (2005). Multievent: an extension of multistate capture-recapture models to uncertain states. *Biometrics*